Title
On Working Memory: Its organization and capacity limits

Author
Lara, Antonio Homero

Publication Date
2010

Peer reviewed|Thesis/dissertation

On Working Memory: Its organization and capacity limits

by

Antonio Homero Lara

Dissertation presented in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jonathan D. Wallis
Professor Jose M. Carmena
Professor Robert T. Knight
Professor Robert W. Levenson

Fall 2010

Abstract

On Working Memory: Its organization and capacity Limits

Antonio Homero Lara

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Jonathan D. Wallis, Chair

This work examines two questions in the working memory field: the organization of working memory function within the prefrontal cortex (PFC) and the nature of the limited capacity of visual working memory. To study the organization of working memory function within the PFC, we recorded the activity of single neurons from different areas within the PFC and from the gustatory cortex (GUS) of two subjects while they performed a gustatory delayed-match-to-sample task with intervening gustatory distraction. Neurons that encoded the gustatory stimulus across the delay, consistent with a role in gustatory working memory, were most prevalent in the orbitofrontal cortex (OFC) (the main recipient of gustatory inputs within the PFC) and GUS compared with dorsolateral PFC and ventrolateral PFC. Gustatory information in OFC was more resilient to intervening distraction, paralleling previous findings regarding visual working memory processes in PFC and posterior sensory cortex. Our findings are consistent with a model of working memory organization in which different PFC areas encode different types of information in working memory depending on their afferent connections with different sensory brain areas.

In the second part of this work, we investigate the nature of the limited capacity in visual working memory by training two subjects on a multiple-item color change detection task. Our results show that when subjects have to store multiple items in visual working memory, the fidelity of the memory traces decreases as more items are loaded into memory. Thus, visual working memory can be described as a limited resource that must be shared in the representation of multiple items. By recording activity from single neurons in the ventral PFC while subjects held multiple colors in memory, we show that for a subset of color-tuned neurons, the sharpness of tuning decreases when two items are being held in working memory. Thus the amount of information encoded by the neurons decreases as more items are loaded in memory, analogous to the loss in precision of the internal memory representations. This finding, however, only applies when we take into account which color the subjects were covertly attending to, suggesting that attention plays a significant role in the allocation of the limited memory resources.

A mis padres

Table of Contents

Acknowledgements

First and foremost I am eternally grateful of my family, they have been an invaluable source of support an encouragement through the long years. My father has been the greatest source of inspiration to pursue my dreams despite the all the difficulties and even when seemingly better and easier options appeared. I could have never gotten this far without the unwavering support of my mother, it is because of her constant support and motivation that I was able to get through the 'final stretch' three times, twice during my undergraduate years and finally in this true 'final stretch' in graduate school. And finally I owe great thanks to my brothers for always believing in me and constantly reminding me of how great working in science can be despite my occasional skepticism.

I am really grateful to all the many great friends I made through the years, Denia Djokic, Hania Köver, Brad Voytek, Erica and Richard Warp, Ryan Canolty, Matar Davis, Brian Miller, Jeremy Mardon, Westgate Haynes, amongst many others. A few people deserve special thanks, Mr. Josh Hoffman for his unwavering friendship, in the good times but especially during the difficult times; my friend, I owe much to all of your helpful words, but I owe more to all your helpful silences, sometimes no words at all spoke louder than anything else. I owe much to Alona Roded for all the life advice and support through the many life changing events and for always listening to all my elaborate plans and theories about life. I would also like to thank Missa Thumm, I will greatly miss having morning coffee and evening pints at the triple, she always provided a much appreciated break from the academic world. I am extremely grateful to Wendy de Heer, above all, she has been an incredible friend and her patience and support in those late night writing sessions were crucial. A very special thanks goes out to Kati Markowitz for all her invaluable help in wading through the administrative hurdles and for making sure I didn't end up on the streets.

In the world of science I am very grateful to Professor Harry Swinney for giving me my first opportunity to work in the Center for Nonlinear Dynamics and for cajoling Dan Goldman to take me on as an undergraduate research assistant; and to Dan for his patience and for introducing me to the thrilling world of experimental science. My graduate career would have been significantly less rewarding without all the long conversations and discussions with Dr. Steve Kennerly about how the brain does what ever it is it does, his constant challenging and interest in my work were invaluable.

Lastly and most importantly I am eternally grateful to Jon Wallis. I know it was a big gamble to take as a first student a transplant from the experimental physics world whose experience with neurophysiology was minimal at best, not to mention his complete ignorance about cognitive neuroscience. It was a very challenging six years and without his mentoring and support I would have never made it. I feel truly fortunate to have had such a great advisor, but mostly, I am really thankful that he did not give me the boot when I arrived to that first surgery at 8 in the morning, completely groggy and half asleep.

List of figures

List of Tables

Chapter 1

Introduction

Imagine driving down the road on your way from work. It has been a long day and you are very hungry. Suddenly you hear your phone ring, it's your spouse and he needs a few things for dinner. Before you take your eyes off the road to find the phone, you take a good look ahead and make a note of the conditions. While keeping the traffic situation in memory, you quickly reach for your phone. With your eyes back on the road, you move your thumb across the keys on the phone; the answer key is bigger than the rest so you have to feel each key until you find the right one. Graham crackers, butter, cream cheese, and strawberries… your spouse is making your favorite cheesecake. You conjure up such vivid images of the delicious dessert you can almost taste it. The list is too long to remember so you make a mental note of traffic again and fumble around looking a pen and paper, all the while the list keeps growing and you cannot stop thinking of the cheesecake. You wisely realize the tremendous danger of the situation and ask your spouse to send you a text message with the list and you diligently go back to driving with your eyes fully on the road.

1.1.1 Two open questions in Working Memory

The example above illustrates some of the many different ways in which we use working memory in our everyday lives. When you take your eyes of the road, you have to keep in mind a representation of the visual world in front of you. As you are feeling the keys of the phone with your thumb, you have to keep in mind the size of each button and compare it to the previous one until you find the biggest. Imagining the future cheesecake involves bringing to mind its taste and smell. Furthermore, the story illustrates an important constraint of working memory: it is highly limited in the amount of information it can hold. As the list of items for the cheesecake recipe starts to grow, at some point it overloads your capacity to put them in working memory and you find yourself having to write them down.

As early as the 1930's it was known that lesions to the frontal part of the brain, the prefrontal cortex (PFC), affected subjects ability to keep information in "immediate memory" or working memory as it s now known (Jacobsen 1935). About forty years later it was discovered that neurons in this area of the brain have the peculiar property of remaining active even after a stimulus is no longer present in the environment (Fuster and Alexander 1971; Fuster 1973). This was the first evidence that PFC neurons are critically involved in the active maintenance of information in working memory (Fuster 1973; Miller and Cohen 2001; Fuster 2008). However, PFC is an anatomically complex region, comprising at least 18 cytoarchitectonically distinct areas. It is not clear how these different anatomical areas differentially contribute to the neuronal instantiation of working memory. Some theories have argued that different PFC areas are responsible for

maintaining different types of information in working memory (Goldman-Rakic 1987). Other theories have argued that different PFC areas are responsible for different cognitive processes that collectively comprise working memory (Petrides 1996). Still other theories argue that anatomy confers few constraints on PFC function with individual PFC neurons flexibly altering their selectivity in response to ongoing task demands (Miller and Cohen 2001).

At the same time that the neurophysiological basis of working memory was being unraveled, cognitive psychologists were trying to work out the limits of working memory in humans. As early as the 1950's there was ample evidence that subjects can maintain a very small number of objects in working memory (Miller 1956). The majority of the work in this field has focused on trying to estimate the amount of information that can be held (Sperling 1960; Sanders 1968; Luck and Vogel 1997) or on developing theoretical models as to why the limit exists (Daneman and Carpenter 1980; Baddeley 1986; Lisman and Idiart 1995; Bays and Husain 2008; Zhang and Luck 2008) if they exist at all (Meyer and Kieras 1997; Bays, Catalao et al. 2009). It has not been until recently that neurophysiologists have turned their attention to the neuronal mechanisms underlying this limit (Warden and Miller 2007; Siegel, Warden et al. 2009).

Thus, despite decades of investigation, two questions remain open: 1) what is the role of different PFC cytoarchitectonic areas in working memory? 2) How do PFC neurons encode multiple pieces of information in working memory and what happens to those mechanisms at the limits of memory capacity?

The work in this dissertation aims to provide answers these two important questions. The first part of the dissertation will describe experiments in which we sought to find evidence for functional segregation within the PFC based on the nature of the sensory information being maintained in working memory. The main experiment in Part I deals with working memory in the gustatory modality, which has previously received very little attention, yet it is very well situated to help unravel the functional organization of working memory within PFC. In Part II we tackle the question of what happens as we push PFC neurons to the limits of how much information they can maintain. By taking a cue from the human psychophysics work, we adapted a multi-item visual working memory paradigm to be used with non-human primates and used it to investigate the properties of PFC neurons at the limits of performance.

1.1.2 Overview of the dissertation

The remainder of this chapter will introduce the two problems in more detail as well as outline our experiments. Before diving into the two questions at hand, however, we will take a brief detour to review the anatomy of the primate PFC and the characteristics of PFC neurons. With this foundation, Chapter 3 will deal with the question of the organization of working memory within PFC. Chapters 4 and 5 will describe our efforts to elucidate the neuronal mechanisms of the capacity limit in visual working memory. Chapter 2 will describe the experimental apparatus and techniques we used in the course of the experiments as well as general analytical techniques used to

analyze the data and Chapter 6 will conclude.

1.2 Structural and functional anatomy of the monkey prefrontal cortex

The frontal lobe is defined as the area of the brain directly rostral to the central sulcus (Figure 1). In rheshsus macaque monkeys, the prefrontal cortex (PFC) lies anterior to the arcuate sulcus and it is the recipient of projections from the mediodorsal nucleus of the thalamus (Barbas and Pandya 1989). The PFC has the feature that it receives inputs from all sensory modalities. These inputs project to distinct regions within the PFC and in turn, the PFC projects to distinct premotor, sensory and subcortical structures. The PFC can be subdivided into various distinct sub-regions: the lateral PFC (LPFC), which itself is subdivided into dorsal and ventral aspects; the medial prefrontal cortex (mPFC); the orbital frontal cortex (OFC), and the mid-dorsal prefrontal cortex. These areas have very distinct cytoarchitectonic characteristics (Walker 1940; Pandya and Yeterian 1985; Morecraft, Geula et al. 1992), intercortical connections (Barbas and Pandya 1989; Barbas and Pandya 1991) as well as different inputs from the sensory cortices (Barbas and Mesulam 1985; Goldman-Rakic 1987; Barbas and Pandya 1989). In what follows, we will briefly review the different cytoarchitectectonical subdivisions of the different PFC areas, the distinct sensory inputs that each area receives and finally the interconnections between prefrontal areas.

1.2.1 Anatomical organization of prefrontal areas

In primates, PFC largely consists of six-layered cortex that has a conspicuous layer 4 with a granular appearance. In other mammals, and in some parts of the primate PFC, layer 4 is missing and is termed agranular. Finally, some of the cortical areas have a subtle layer 4 and are termed dysgranular. Agranular prefrontal regions are considered the phylogenetically oldest areas, granular areas are considered the phylogenetically newest and dysgranular regions lie somewhere between these two extremes. The lateral PFC consists almost exclusively of granular cortex. It is dominated by the principal sulcus, which contains area 46, and the arcuate sulcus, which contains area 8 (Figure 1). Area 9 lies dorsal to the principal sulcus and is the only lateral area that is dysgranular. Areas 47/12 and 45 lie ventral to the principal sulcus. Area 10 is anterior to the principal sulcus and comprises the frontal pole.

The orbitofrontal cortex (OFC) is organized along an anterior-posterior axis. The posterior part of OFC contains five agranular areas that are recognized as the rostral extension of the agranular insula (Iam, Iai, Ial, Iapm and Iapl). Anterior to these regions is area 13, which comprises both agranular and dysgranular cortex, and anterior to area 13 is area 11, which consists of granular cortex. The medial extent of OFC consists of dysgranular area 14 while the lateral extent consists of granular area 12.

Most areas on the medial wall of the PFC are agranular. It is dominated by area 32, along with area 24 dorsally. Area 25 is located underneath the corpus callosum. Other

areas on the medial aspect of the prefrontal cortex include the supplementary eye field (F7) and medial aspects of areas 9, 10 and 14.

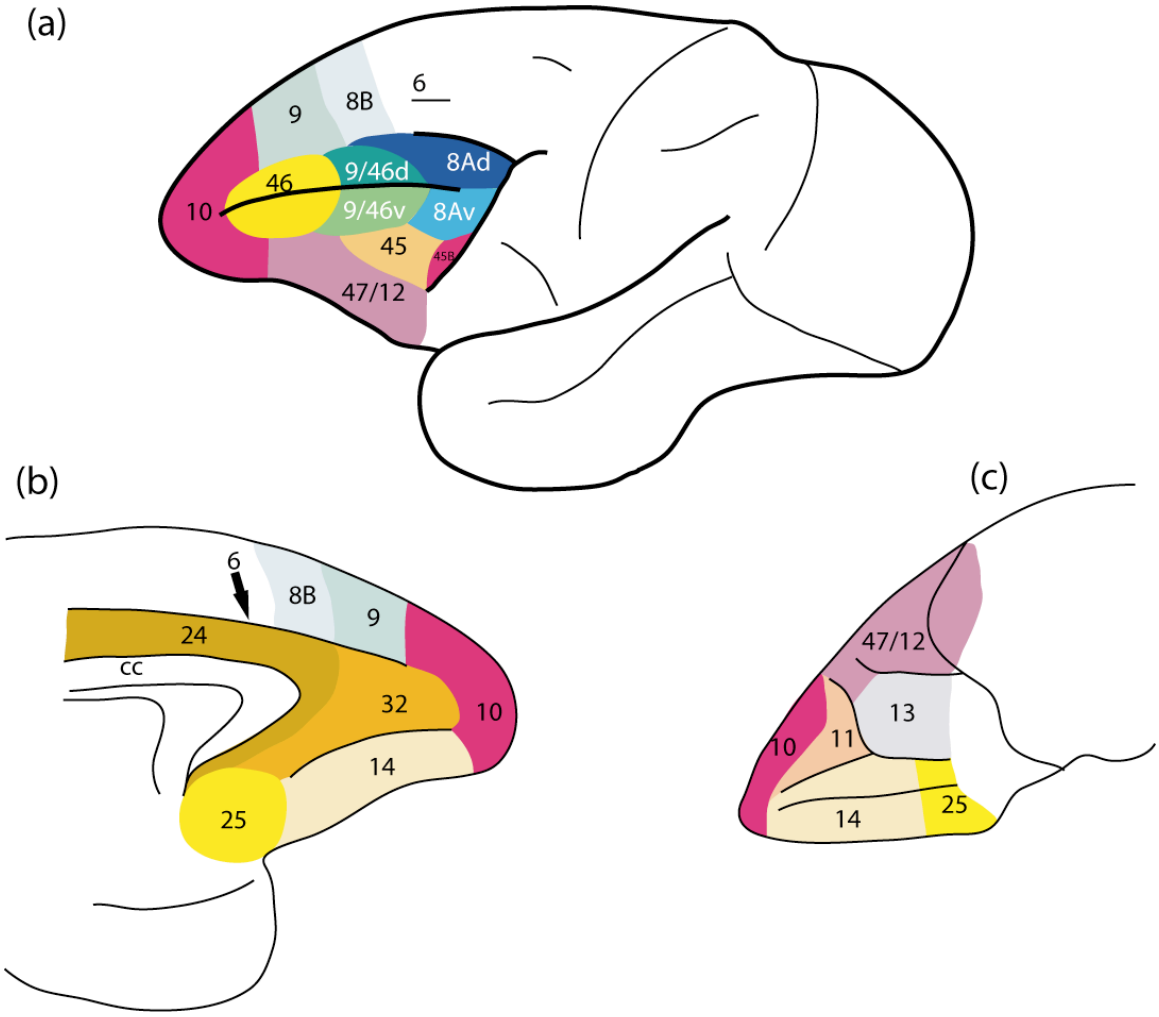
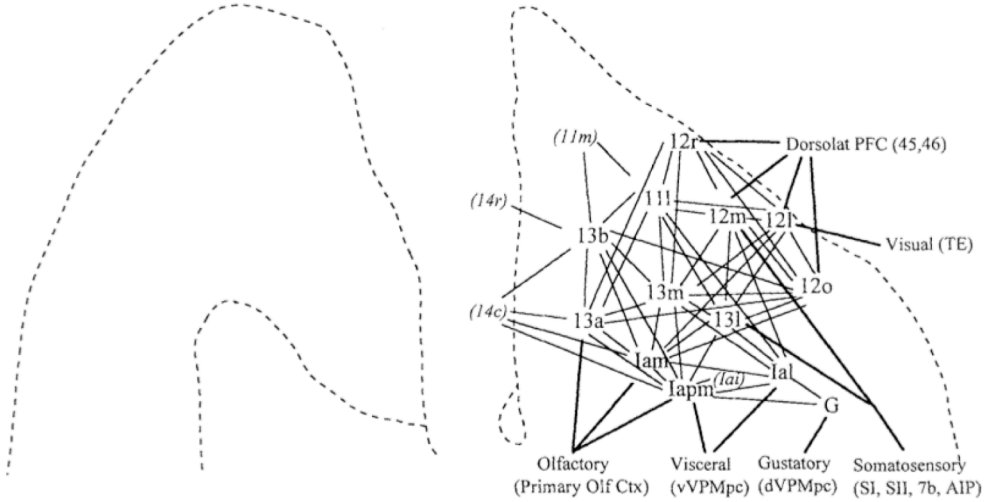


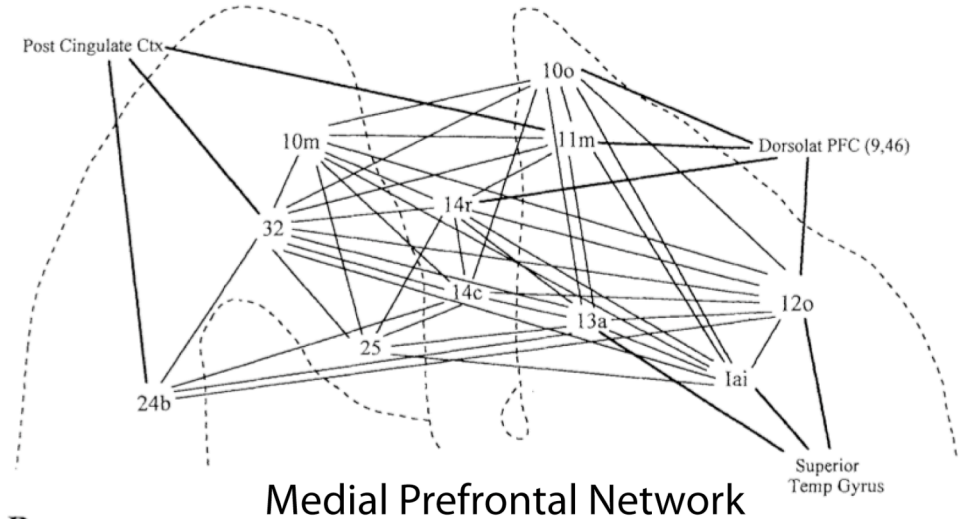
Figure 1.1 Cytoarchitectonic subdivisions of the macaque prefrontal cortex. Lateral surface (a), medial surface (b) and inferior surface (c). the arrows show the extension of area 32 in the upper part of the cingulate gyurs. Adapted from Petrides and Pandya (1994)

The different PFC areas are massively interconnected. There are reciprocal connections between all major subdivisions (LPFC, OFC and MPFC) and between their constituent areas (Pandya and Barnes 1987; Barbas and Pandya 1989). However, it would be a mistake to assume that every region is connected to every other region. Instead there appear to be distinct networks. Within a network, all areas connect to one another, but they only weakly connect to areas in other networks. Perhaps the best delineated networks are those in OFC and MPFC identified by Carmichael and Price (1996). Most of the posterior, central and lateral orbital surfaces (insular areas Ial, Iam, Iapl, and Iapm,

and orbital areas 13b, 13l, 13m, 11l, 12r, 12m and 12l) are interconnected in what they termed the 'orbital prefrontal network' (Figure 1.2). A separate network within the mPFC and OFC includes all the areas on the medial wall (areas 25, 32, 14r, 14c, 24a and 24b) and related areas on the orbital surface (areas 11m, 13a, Iai and 12o), together these areas comprise the 'medial prefrontal network' (Carmichael and Price 1996).



A. Orbital Prefrontal Network



B. Medial Prefrontal Network

Figure 1.2. Orbital and Medial Prefrontal Networks within the PFC. From Carmichael and Price (1996)

1.2.2 Sensory connections of PFC

The aim of this work is to understand the role that PFC plays in holding sensory information in working memory. Given that, we will spend some time reviewing the route by which that sensory information arrives in PFC.

Visual information enters PFC through two primary pathways, reflecting specialization in posterior sensory cortex. The dorsal visual stream includes areas in parietal cortex and is particularly involved in the analysis of spatial information and motion. The parietal cortex projects to the PFC in a systematic manner: projections from area LIP terminate mostly in the frontal eye field (area 8a) whereas area 7a projects mostly to the dorsal part of the principal sulcus, area 46 (Cavada & GR, 1989). The ventral visual stream includes areas in temporal cortex. Neurons in this pathway are highly selective for complex features of visual stimuli and sometimes respond only to specific images or objects (Desimone, et al 1984; Fujita et al, 1992; Gross et al, 1972; Tanaka et al 1991). Neurons in temporal cortex are reciprocally connected to the ventral part of the PFC. These connections typically maintain their anterior-posterior organization such that area TEO projects mostly to the inferior limb of the arcuate sulcus (area 45), whereas area TE targets primarily area 12 of the inferior convexity (Distler et al, 1993; Kawamura & Naito, 1984; Markowitsch et al, 1985; Webster et al, 1994)

Other sensory modalities also project largely to LPFC. Somatosensory information enters the PFC primarily through the ventral rim of area 46 which is interconnected with the parietal lobe regions, particularly area 7a.(Cavada and Goldman-Rakic 1989). The rostral belt of the auditory cortex, which processes the identity of auditory stimuli, projects to the inferior convexity (areas 12 and 45), whereas the caudal belt cortex, which is responsible for localizing sounds in space, projects to the caudal region of area 46 and area 8 (Hackett et al., 1999; Romanski et al., 1999; Romanski, Tian, et al. 1999).

Unlike the sensory modalities described above, gustatory and olfactory inputs to the PFC do not directly project to lateral regions of the PFC, but rather enter the frontal lobe through the orbitofrontal cortex. Gustatory inputs arise from the primary gustatory cortex (GUS), located in the post-central gyrus and secondary gustatory areas, including the insula and precentral opercular areas. They project primarily to areas 12 and 13 of the orbitofrontal cortex (OFC) (Cavada et al, 2000). The OFC also receives olfactory information from the pyriform cortex..

1.2.3 Physiological properties of PFC neurons related to working memory

Neurons in the PFC have been known to show sustained activity during the delay period of a visual delay response task (Fuster and Alexander 1971; Fuster 1973). In these now classical experiments, monkeys were shown a morsel of food being hidden in one of two wells placed side by side. After a delay of up to several seconds the monkey was allowed to uncover one well. In order to uncover the right one and obtain the food

reward, the monkey needed to maintain the location of the well that had the morsel of food in working memory. Fuster and Alexander (1971) showed that neurons in the principal sulcus showed a sustained firing thought the delay period. Subsequent studies confirmed this finding and showed that this sustained increase in activity was specific to the side the animal was holding in working memory (Kubota and Niki 1971; Fuster 1973; Niki 1974; Niki 1974). As research progressed, PFC proved to be more versatile than just showing delay period activity in delayed response tasks (Niki and Watanabe 1979; Fuster, Bauer et al. 1982) with a variety of neuronal responses being reported even in the earliest studies of Fuster and Alexander (1973). For example, in their seminal study Fuster and Alexander classified neuronal responses into six different categories such as those that showed "[a] transient excitatory reaction to the lifting of the screen" (p 64.) and neurons that "were inhibited during most or all of the cue period ... [and] at the beginning of the delay, these units showed a reversal of impulse frequency to a level higher than the intertrial baseline." (p. 65).

Goldman-Rakic and colleagues developed a more controlled version of the delayed response task. She required monkeys to remember the locations of visual stimuli shown on a screen while keeping their eye fixated on the center of the screen in order to maintain constant the visual input to the brain (Funahashi, Bruce et al. 1989). Using this task, work in her laboratory showed that neurons in the DLPFC display a sustained increased firing that is selective for the spatial position the animal is remembering (Figure 1.3). Subsequent work in her laboratory and others reveled that this sustained activity was not related to the upcoming response (Funahashi, Chafee et al. 1993). In other words, DLPFC neurons showed *spatial* memory fields.

Subsequent work showed that when animals were trained to remember the identity of a visual stimuli while ignoring its location in space, neurons in the inferior convexity of the PFC showed selective increase in firing (Wilson, Scalaidhe et al. 1993). Areas in the inferior convexity of the PFC are the major recipients of afferents from the ventral visual stream (see section 1.2.2), thus it appeared that there was a functional and anatomical relationship between the nature of inputs a PFC area receives and the information neurons within that area maintained in working memory (Wilson, Scalaidhe et al. 1993) .

Subsequently, similar types of memory responses were seen for tasks that involved remembering other kinds of non-spatial stimuli. When sounds (Bodner, Kroger et al. 1996) or tactile stimuli (Romo, Brody et al. 1999) were used as memoranda, neurons in distinct areas of the PFC showed selective delay period activity related to maintaining those types of stimuli in working memory.

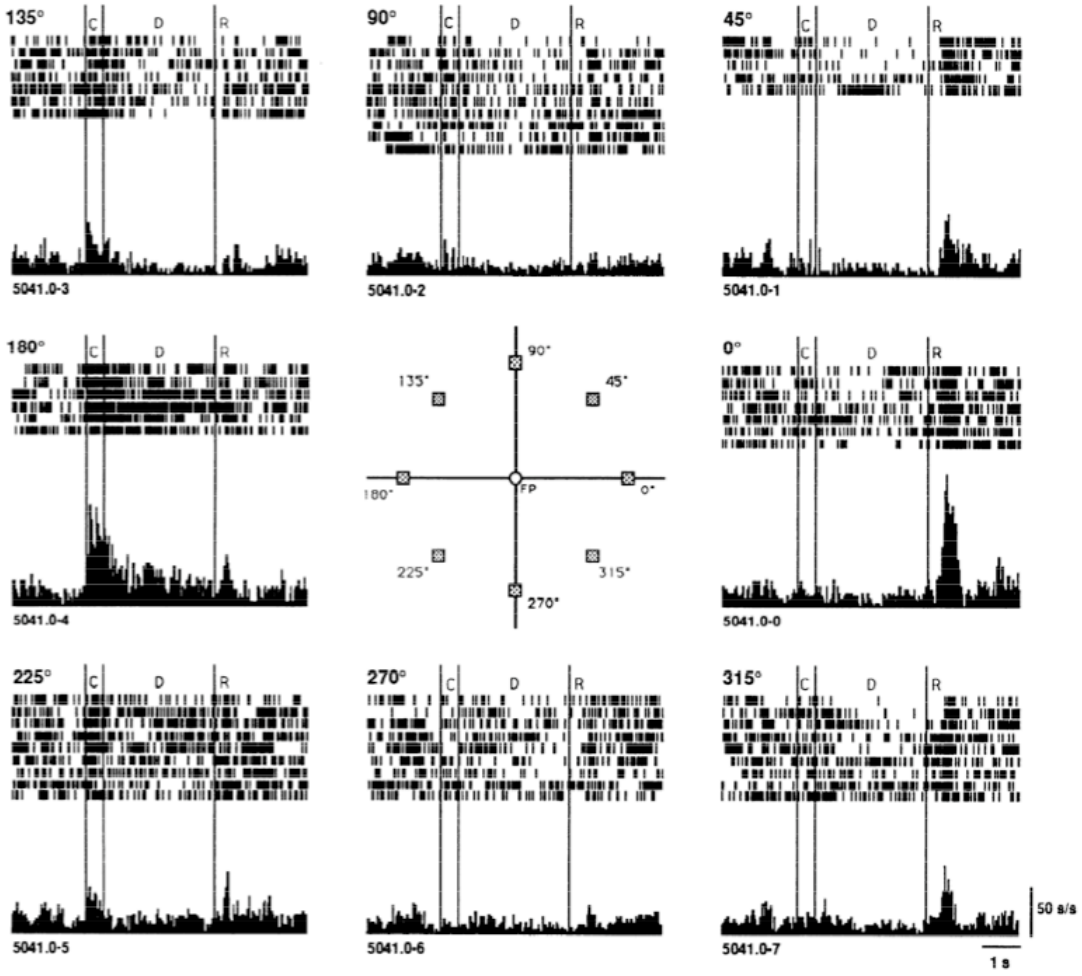


Figure 1.3. Delay period activity of a neuron from the prefrontal cortex while a subject was performing an oculomotor delayed match to sample task. This neuron shows the highest increase in activity for stimuli presented 180º from fixation. Taken from Funahashi *et al.* (1990).

Work by Miller and colleagues has since identified two further features of PFC neurons that would be useful for a system involved in working memory. In an experiment where monkeys had to maintain a picture in working memory though in intervening sequence of distracter images, Miller and colleagues (Miller, Erickson et al. 1996) showed that neurons in the PFC encoded the target image in working memory and that this activity was resistant to the intervening stimuli of the sequence. This is different than what was seen in the inferior temporal (IT) cortex, a visual area that involved visual object encoding. Neurons from this area also encoded a the target image during the intervening delay between images. However, as soon as another image appeared on the screen, IT neurons switched their response pattern to encode the current image (Miller and Desimone 1994). Thus, while neurons in lower sensory areas show sustained delay period activity, this activity was primarily related to the current image in the sequence.

PFC neurons maintained a trace of the target image that was resistant to the intervening images, which supports its role in maintaining information in working memory.

In addition, individual PFC neurons are often able to encode more than one type of information in working memory. In an experiment where monkeys were trained to remember first the identity of a cue presented at the center of the screen over a brief delay ('what' delay). After the delay the target image reappeared paired with a distractor image in one of four different locations. Monkeys where required to remember the spatial location of the target image for a second delay ('where' delay). After the second delay, monkeys had to saccade to the target location to obtain reward. Miller and colleagues (Rao, Rainer et al. 1997) found that over half of the neurons from DLPFC and VLPFC showed working memory activity that reflected both spatial location and object identity. Moreover, neuronal activity mirrored the demands of the task: when monkeys had to remember the identity of the target, some neurons showed high object selectivity, and when the monkeys then had to remember the location of the cue these same neurons were able to switch the modality of information they encoded and showed only location selectivity.

1.3 Organization of WM in PFC

1.3.1 The content-model of PFC organization

Exactly how information is organized within the PFC has been the subject of great debate (Goldman-Rakic 1987; Owen, Evans et al. 1996; Petrides 1996; Owen, Stern et al. 1998; Levy and Goldman-Rakic 2000). Based largely on lesion data, neurophysiological recordings, and the distinct anatomical patterns of inputs into PFC (section 1.2.2), Goldman-Rakic put forward an influential model in which she proposed that working memory is organized in a domain specific manner within the prefrontal cortex (Goldman-Rakic 1987).

This model of working memory functional organization postulates an organization within PFC based on the content of the information being maintained in working memory. For example, given that the DLPFC is the primary recipient of visuo-spatial information from the dorsal visual pathway, the theory posits that this area will maintain spatial location in working memory. Goldman-Rakic, took this observation further and predicted that this functional segregation of working memory would apply to all other sensory modalities (Goldman-Rakic 1987; Goldman-Rakic 1996; Levy and Goldman-Rakic 2000). Thus, the content model states that different PFC areas encode different types of information in working memory depending on their afferent connections with other brain areas. Support for this model came from studies showing that lesions of dorsolateral PFC (DLPFC), which strongly connects with parietal cortex (Cavada and Goldman-Rakic 1989), impaired spatial working memory (Funahashi, Chafee et al. 1993), whereas lesions of ventrolateral PFC (VLPFC), which strongly connects with inferior temporal cortex, impaired working memory for objects (Mishkin and Manning 1978). Early neurophysiological studies also supported this distinction:

neurons maintaining spatial information in working memory occurred in DLPFC (Funahashi, Bruce et al. 1989), whereas those maintaining object information were in VLPFC (Wilson, Scalaidhe et al. 1993).

1.3.2 PFC flexibility

Recent findings have argued against the content model. For example, when monkeys have to remember both the identity and spatial location of an object, neurons encoding both attributes occur throughout lateral PFC, including both DLPFC and VLPFC subregions (Rao, Rainer et al. 1997). Furthermore, PFC neurons can switch whether they encode information about an object's identity or location depending on which aspect is currently relevant to the task. These findings suggest that PFC organization does not depend on the nature of the information encoded by the neurons.

Additionally, PFC can flexibly encode information about categories of stimuli within the same modality. Freedman and colleagues (Freedman, Riesenhuber et al. 2001) trained monkeys to categorize computer generated stimuli as cats and dogs. They systematically varied the shape of the stimuli to look more like cats or more like dogs, with the category boundary being defined arbitrarily. They found that activity in the PFC reflected the category of visual stimuli, even when the category boundary was changed and monkeys had to re-learn the categories. This provided further evidence that PFC activity is malleable.

In addition, evidence has increased for integration of spatial and object information in posterior sensory cortex. For example, neurons in parietal cortex, traditionally associated with encoding spatial information, will encode the color of objects if color is behaviorally relevant (Toth and Assad 2002). Neurons in the lateral intra parietal area can also encode both spatial and category information in a motion categorization task (Freedman and Assad 2009). If integration of spatial and object information occurs in posterior sensory cortex, it is not surprising that the activity of PFC neurons reflects this integration.

1.3.3 Testing the content-model

In the first part of this thesis we intend to test the content model of working memory organization by testing other sensory modalities. Researchers have also studied working memory for auditory and somatosensory information at the single neuron level (Romo, Brody et al. 1999; Romanski and Goldman-Rakic 2002). However, information from these modalities predominantly enters the frontal lobe via VLPFC (Carmichael and Price 1995; Romanski, Bates et al. 1999) and so consequently they do not provide a stronger test of the content model than has already been established by studying the visual modality. In the first part of this dissertation we provide evidence in support of the content model of functional organization by turning to the gustatory modality. Gustatory information enters PFC via the OFC (Morecraft, Geula et al. 1992; Barbas 1993), thus the content model argues that OFC but not VLPFC neurons would encode olfactory and gustatory information in working memory. To test this, we simultaneously recorded

neuronal activity from multiple PFC subregions (DLPFC, VLPFC and OFC) and primary gustatory cortex (GUS), while subjects performed a gustatory delayed match-to-sample (DMS) task. This experiment and our results are discussed in detail in Chapter 3.

In the final part of this Introduction we will focus on a second question involving the working memory system. Namely, we will turn to the limits in the amount of information that can be simultaneously maintained in working memory.

1.4 Capacity limits of working memory

The ability to simultaneously hold multiple pieces of information in working memory is paramount to complex cognitive abilities such as learning, problem solving and language comprehension (Cowan 2001; Cowan, Elliott et al. 2005) and it is highly correlated with general intelligence (Conway, Kane et al. 2003). This ability, however, is extremely limited in capacity. In the mid 1950's Miller summarized a large body of work dealing with the matter of making discrimination judgments among multiple stimuli (Miller 1956). In most of the experiments he reviewed subjects were presented with a stimulus, a tone for example, and were asked to assign a label to each stimulus (e.g. is this tone 1 or tone 5). The evidence reviewed by Miller suggested that subjects could handle at most 6 or 7 categories simultaneously. Miller took these results and linked the limit in absolute judgment to a limit in memory span. He performed an information theoretical analysis and concluded that on average there is a limit to the amount of information that can be held in memory. Specifically, he proposed that subjects could keep 7±2 items in memory. Miller called this number the "magical number 7" and his article became one of the most influential articles in modern psychology (Baddeley 1992).

1.4.1 Magical number four

One of the biggest difficulties in trying to measure a capacity limit has been the phenomenon of chunking, where subjects group several pieces of information in to a single chunk, which have lead to over estimates of capacity. Another difficulty has been trying to disentangle the passage of time from estimates of capacity. Several researchers have argued that the observed limit is not due to a limited capacity store *per se*, but rather there is a limit in the amount of information that can be maintained in memory without rehearsal (Baddeley 1986).

Experiments made after Miller's seminal paper took care to control for these and other confounds by requiring subjects to mentally rehearse a single word while trying to remember a set of tones, for example or by overloading the system at the time of stimulus presentation such that there is more information than the subject can rehearse or encode (Cowan 2001) . Henderson (1972) reviewed an extensive set of such experiments and proposed a "new magic number 4 ± 1" (p. 486). The majority of these experiments were done in the verbal storage system. It was not until the late 1990's that researchers turned to the visual system (Luck and Vogel 1997). In their paper Luck and Vogel provided an

estimate for visual short-term memory capacity using a set of experiments that asked subjects to detect a change in color, orientation or both in an array of up to 12 colors. Their experiments varied in complexity and controlled for various mnemonic strategies subjects might adapt. They concluded that subjects can retain an average four visual features and also proposed that the capacity of visual working memory should be understood in terms of integrated objects rather than individual stimulus features.

1.4.2 Visual working memory capacity

The majority of studies that have investigated capacity limits in visual short-term memory have used modified versions of Luck and Vogel's color change detection task. The prevailing view that emerged at the time from this type of experiments was that there is high-level memory store that is limited to about four memory 'slots' (Alvarez and Cavanagh 2004). This view arose from change detection experiments that used multi-item displays with highly discriminable stimuli. Researchers found that an increase in the number of elements in the display lead to a monotonic decrease in the ability of observers to detect a change in the display, and this decrease was typically seen only after the number of items in memory was about three to four (Pashler 1988; Luck and Vogel 1997; Vogel, Woodman et al. 2001). However, in a series of experiments where subjects were asked to detect a change in color, orientation, or spatial frequency in an array containing 8-12 stimuli and report their confidence of the change, Wilken and Ma (2004) found that the apparent limit in capacity of visual working memory can be explained if one assumes that the brain stores a noisy representations, and that the noise increases as a the memory load increases.

Recently, this assumption was tested by two different groups, which came to two very different conclusions. One grouped incorporated noise into the slots model (Zhang and Luck 2008). In their experiments, Zhang and Luck asked subjects to report the color that had changed in a multi-item array by pointing to a region in a color wheel. By looking at the difference between reported color and the actual color that had changed, they were able to estimate the precision with which subjects remembered the color. They concluded that memories are stored in slots with fixed representational noise (i.e. with fixed precision) and that there is a limited number of fixed precision slots which give rise to the capacity limit.

Another group proposed a completely different model, the shared resources model (Bays and Husain 2008). This model states that there is a limited resource pool that is shared amongst the items in working memory. When there is a single item in memory, all the resources will go towards representing it, which will in turn produce a very high precision (low noise) memory trace. When the number of items increases the resources are split between the items at the expense of representational precision. This model states that there is no capacity limit *per se*; rather, when the number of items in memory is large the representations become so imprecise that information about the items in memory is lost. They tested this model by asking subjects to report whether a single item in a multi-item display had shifted position to the right of to the left. By varying the amount of displacement they were able to estimate the precision of the memory for that object's

location. They found that as the number of objects in memory increased, subjects were less able to detect small displacements, and crucially, this impairment was seen even in the transition between one to two objects with no evidence of a discontinuity in the region of four items. In order to rule out the possibility that their results were due to a differences in the behavioral paradigm they used, Husain and colleagues repeated their experiments using a color change detection task and were able to replicate their results (Bays, Catalao et al. 2009).

In summary, within cognitive psychology there is still controversy regarding the nature of the capacity limit in visual working memory. Understanding the neural underpinnings of the capacity limit could help to validate one or other psychological model. To date, only a small number of researchers have begun to probe these neural underpinnings. The two prevailing models described above, make different predictions. The fixed-precision 'slots' model, predicts that a fixed number of neurons in the brain will maintain a memory trace with a fixed precision. As the number of items in memory increases, a separate population of fixed-precision neurons will get activated to represent the second item and so forth. The capacity limit arises when there are no more slots to form a memory trace. On the other hand, the shared resources model makes the prediction that there might be a fixed total number of neurons that will be shared amongst the items in working memory. When there is a single item in memory, all the neuronal resources will go towards it, which will in turn produce a very high precision representation. When the number of items increases to two, half the neuronal resources go to one item and half to the other at the expense of representational precision. Alternatively, the number of neurons representing an item might remain fixed, but the neurons will change the precision of their memory encoding as the number of items increases. Multiple items are thus represented simultaneously by all the available neurons but each with a decreased precision.

1.4.3 Neurophysiological basis of the capacity limit

Meanwhile, Vogel and Machizawa (2004) showed that the magnitude of the event related potential (ERP) recorded over posterior parietal and lateral-occipital sites was strongly modulated by the number of objects being held in working memory. These studies, although an important first step, leave many questions open. Both studies found changes in activity in parietal and occipital sites which form part of the dorsal stream of visual information processing (Goodale and Milner 1992), and they did not see evidence of modulation of activity in PFC sites. It is unclear why this might be the case, one possible reason why these studies failed to detect PFC involvement might be the low spatial and temporal resolution of the respective methods. Recent work by Voytek et al., however, did find that in patients with frontal lobe damage, the PFC plays a critical role in this task (Voytek, Davis et al. 2010). Neurophysiological experiments in animal models, however, have indeed found that PFC neurons are involved in maintaining multiple items in working memory (Warden and Miller 2007; Siegel, Warden et al. 2009). Miller and colleagues have showed that when animal have to remember a sequence of two images, the activity of PFC neurons encodes information about both images. Although their experiments provided an important first step in understanding the

encoding of multiple items in working memory, the neuronal mechanisms of the limit in capacity in visual working memory remain largely unexplored. The second part of this thesis aims to elucidate the neuronal mechanisms of this limit. We will do this by using a version of the color change detection task adapted from Luck and Vogel (1997) that allowed us to investigate how the information encoded by PFC neurons changes as the number of items increases. In Chapter 4 we will describe the behavioral results we obtained from two subjects trained in the color change detection task. In Chapter 5 we will discuss experiments we performed where we recorded the activity of single neurons in the ventral PFC, an area that receives afferents from posterior color areas (Webster, Bachevalier et al. 1994), while subjects were maintaining multiple colors in working memory. We will show that neurons the PFC selectively encode the color of the stimuli in working memory, and moreover, that as the memory load increases, neurons loose their color tuning and thus encode less able to encode information which might form the basis for the limited capacity in working memory.

To date, no work has been done to test directly the predictions of the two models at the single neuron level. However, several studies have used variants of the color change detection task to begin to elucidate the neuronal mechanisms of the capacity limit. Notably, two recent studies have begun to explore the neuronal correlates of the capacity limit of visual working memory in humans (Todd and Marois 2004; Vogel and Machizawa 2004). These studies used a version of the color change detection task in which subjects were presented with an array of colored squares in one side of the visual field. Subjects were required to report a change in one of the items after an intervening delay. Todd and Marois (2004) showed that the strength of the blood oxygenation level-dependent (BOLD) signal in parietal and occipital cortex was correlated with the number of items being held in working memory.

Chapter 2

General Methods

This chapter will provide a description of the methods and techniques used in the course of the experiments. What is covered here applies to all the experiments described in future chapters. Methods that are particular to a certain experiment will be described in detail in the chapter in which they appear.

2.1 Behavioral training materials and methods

2.1.1 Subjects

We used three male rhesus monkeys (Maccaca mulatta) for the experiments. Subject C took part in the gustatory working memory experiment, was 5 years old and weighed 13 kg at the time of recording. Subject I took part in the working memory capacity experiment, was 6 years old and weighed 14 kg at the time of recording. Subject G participated in both experiments; was 4 years old and weighed 9 kg at the time of recording for the gustatory working memory experiment and was 5 years old and weighed 9 kg at the time of recording for the working memory capacity experiment. All subjects were pair housed in a primate colony with a total population of 14 monkeys. The subjects were kept in a 12hrs/12hrs light/dark cycle and were fed twice a day. Subjects' fluid intake was regulated in order to maintain motivation to perform the task.

2.1.2 Behavioral Training

We first trained subjects to sit quietly in a primate chair and to move a lever to obtain rewards in response to cues presented in a computer monitor (see below for a description of the apparatus used for training).

2.1.3 Surgery

All subjects underwent three operations. The first was to implant a custom-made titanium head-positioning post using titanium orthopedic screws. This post ensured that the subject's head could be immobilized during recording. The second surgery was to implant titanium recording chambers, using a combination of bone cement and screws. The third was to perform craniotomies to enable access to the brain.

For all surgeries, subjects were anesthetized (ketamine induction, 10 mg/kg i.m. with anaesthetic maintenance isofuorane [1±3%] in balance oxygen). Full sterile precautions are used during surgery. Veterinary personnel maintain anesthesia and records during the surgical procedures. The anesthesia level was monitored continuously, and the level was assessed by heartbeat (90-150 beats/min), respiration rate (17-23 breaths/min), temperature (36-39 °C) and absence of responses to stimuli such as toe pinch. Blood oxygen saturation was also monitored with a pulse-oxymeter. A lactated ringer IV drip (2-4 ml/kg/hr) was used during surgery to ensure that the animal received sufficient hydration, and a heating pad and towels were used to maintain body temperature. At the end of the surgical procedure gas anesthesia was discontinued, and once the animal showed signs of coming out of anesthesia, the intubation tube was removed and buprenorphine (0.01-0.03 mg/kg SC or IM) or morphine (1-2 mg/kg SC) were injected for relief of post-operative pain. In the hours following any surgery, the surgical team monitored the animal closely, to ensure normal recovery from anesthesia and appropriate level of analgesia. After initial recovery, the animal was checked several times per day (as necessary) at which time appropriate analgesics and antibiotics were administered. All procedures were in accord with the National Institute of Health guidelines and the recommendations of the U.C. Berkeley Subject Care and Use Committee.

2.1.4 Behavioral Training Apparatus

For all tasks, subjects sat in a primate chair facing a 19-inch LCD computer screen placed at a distance of 32 cm from the chair. The apparatus used to control behavioral events is diagramed in Figure 2.1. A pair of computers running the NIMH Cortex program for data acquisition and experimental control (http://www.cortex.salk.edu/) controlled the timing and presentation of stimuli. In this program, the 'receive' computer is used to present visual stimuli on the LCD screen and is controlled by the 'send' computer via the COM port. A video splitter is used to split the video output of the 'receive' computer to allow for monitoring of the visual stimuli by the experimenter. The 'send' computer is used to compile and run timing routines and interface with various external devices using a PCI-DAS1602/12 data acquisition (DAQ) card (Measurement Computing, Norton, MA). Each behavioral event in the trial is marked with a code that is sent to the MAP system (see below) using an 8-bit 'word' from the DAQ card. The MAP system reads in this word and records a time stamp for each event in the trial. The 'send' computer runs with a single interrupt routine that triggers every 1 millisecond and both updates a software clock and samples all data lines. Thus, the control of the behavioral contingencies, the presentation of stimuli and the monitoring of behavioral events all take place with a 1 ms resolution. The MAP system takes the 8-bit word and records it along with the neurophysiological data onto a single data file.

Reward delivery was controlled via the card's digital output port by sending a 5-volt TTL pulse to an ISMATEC-IPC8 peristaltic pump (ISMATEC SA, Glattbrug

Switzerland). The pump was configured to deliver a reward for a duration directly proportional to the duration of the TTL pulse at a constant flow rate of 0.15 ml/sec. The gustatory working memory experiment required subjects to hold one of three different juices in working memory for a period of 3 seconds with an intervening water drop midway through the delay. To achieve this, we built a custom circuit, designed to open a three-way solenoid valve in response to a TTL pulse from the DAQ card. The peristaltic pump has eight independent lines, and by using four of these valves, we delivered four different kinds of liquids to the subjects via a custom-made mouth-piece with four independent cannulae, which diverted the relevant juice into the mouthpiece while routing the others back to juice container.

Subjects' response to events in the task were registered using a 4-TPS-E1 Touch Sensor Module (Crist Instrument, Damascus, MD) connected to the digital input port of the DAQ card. The touch sensor was a contact sensitive device designed to send a 5-volt TTL pulse when a grounded subject touched it.
Eye position was recorded using an infrared eye monitoring system (ISCAN, Burlington, MA). This system has an infrared camera aimed at the subject's eye and uses image tracking software to find the x and y positions of center of the subject's pupil as well as the pupil's diameter. These three signals are fed separately to the DAQ analog input channels and recorded for the duration of the session.

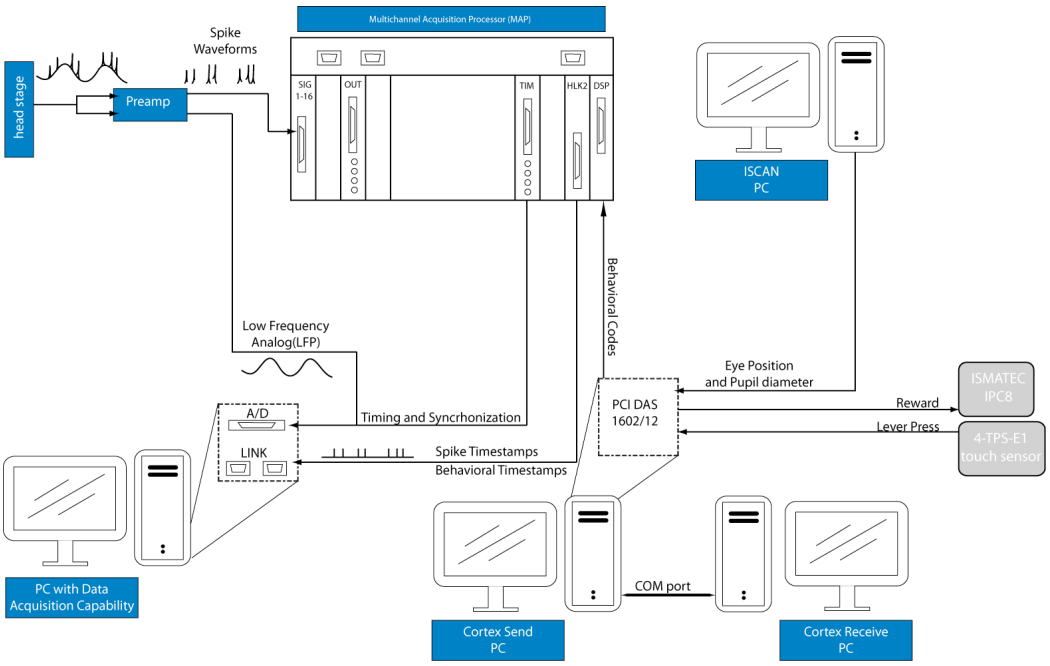


Figure 2.1. Schematic diagram of the experimental setup used to control behavioral events and record neurophysiological data.

2.2 Neurophysiological Techniques

We obtained images of the subjects' brain using a 1.5 T magnetic resonance imaging (MRI) scanner at the U.C. Davis Center for Imaging Sciences. Using these images, we calculated the stereotactic coordinates for chamber implantation using commercial graphics software (Adobe Illustrator CS3, San Jose, CA). We confirmed the correspondence between the MRI sections and our recording chambers by mapping the position of sulci and grey and white matter boundaries using neurophysiological recordings.

Subject C was implanted with two recording chambers. One chamber was placed in the right hemisphere centered over the principal sulcus approximately 31 mm anterior of the interaural line parallel to the anterior-posterior (AP) axis (i.e. AP 31) angled at 27º from the vertical (Figure 2.2 column 4). This placement allowed us access to the prefrontal and orbital cortices. The second chamber was placed in the left hemisphere at approximately AP 27 parallel to the interaural line (Figure 2.2 column 1); this permitted access to the primary gustatory cortex.

Subject G was also implanted with two recording chambers; one centered over the principal sulcus at AP 30 at an angle of 26º in the left hemisphere (Figure 2.2 column 3), and one centered over primary gustatory cortex at AP 28 in the right hemisphere (Figure 2.2 column 2). For the color change detection experiment, subject G's right hemisphere chamber was repositioned so that it lay over the principal sulcus (AP 31) and the left hemisphere chamber was removed (Figure 2.2 column 5).

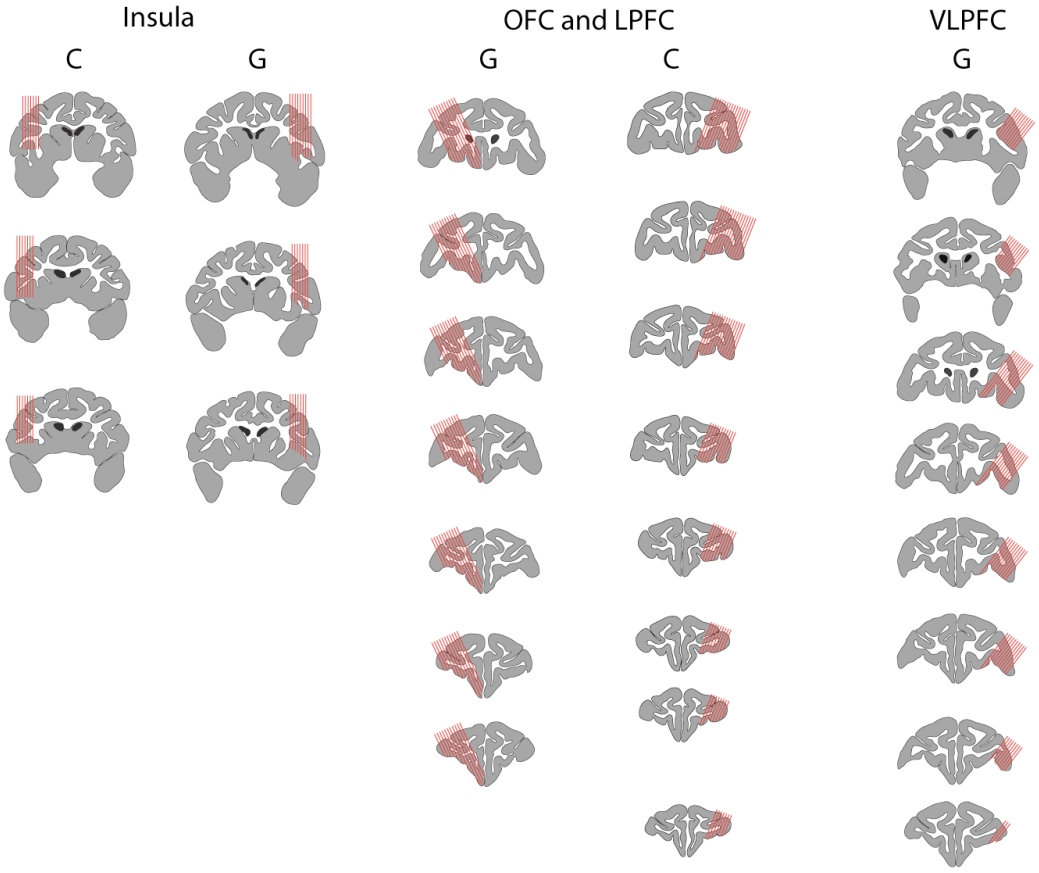


Figure 2.2 Representative coronal sections showing the location of the recording chambers in each subject. Columns 1-4 show the placement of the chambers for the gustatory working memory experiments. Columns 1 and 2 show the placement of the chamber over the insula, while columns 3 and 4 show the placement of the chamber over OFC and LPFC of subjects C and G respectively. Column 5 shows the placement of the chamber of the VLPFC of subject G for the multi-item color change detection experiment. Red lines represent the electrodes.

2.2.1 Neurophysiological Recordings

We recorded simultaneously from the different brain areas using arrays of 8-16 dura-piercing tungsten microelectrodes (FHC Instruments, Bowdoin, ME) assembled using custom-made plastic grids that contain an array of 22-gauge holes spaced 1 mm apart. Stainless steel 22-gauge hypodermic needles glued to the bottom of the plastic grid guided the electrodes to the desired position above the dura. We determined the approximate distance to lower the electrodes from the MRI images and advanced the

electrodes using custom-built, manual microdrives until they were located just above the cell layer. We then slowly lowered the electrodes into the cell layer until we obtained a neuronal waveform. After we successfully isolated single neurons on most or all the channels, the electrodes were allowed to rest undisturbed for 1 to 1.5 hours in order for the brain to return to its resting position after being displaced by the force of the penetrating electrode. Typically, we were able to hold cells for the entire duration of the resting period and the subsequent experiment with only a minimal amount of drift in the neuronal signal. During recording, no attempt was made to change the electrode position in order to maintain recording stability, since doing so would have disturbed the animal from the performance of his task. Therefore, we discarded channels that had unstable neuronal waveforms (less than 10% of the channels) during our offline analysis (see below).

2.2.2 Recording Apparatus

Waveforms were recorded and digitized using the Plexon Multichannel Acquisition Processor (MAP) system (Plexon Inc, Dallas TX). This system recorded the voltage measured at the tip of each of the electrodes with respect to a reference electrode, typically taken as one of the titanium screws holding the head positioner in the skull or the head positioner itself. The signal was amplified by a series of two operational amplifiers (Figure 2.1, top left). The first stage consists of an op-amp based headstage amplifier placed close to the electrodes that provides up to 20x signal gain. The second stage consists of a pre-amplifier that further increases signal gain up to 1000x and uses hardware filters to separate the signal in to a low frequency analog signal (band pass filtered between 3 and 150Hz), which reflects the local field potential (LFP), and a high frequency signal (150 Hz to 8 KHz), which reflects neuronal spiking (Figure 2.1).

Both signals were digitized at 40 KHz at 12-bit resolution per channel. The low frequency signal was stored in a hard drive for later analysis. The high frequency signal was further passed trough a simple threshold discriminator using the Real-Time Acquisition System Programs for Unit Timing in Neuroscience software suite. When the signal crossed the threshold the program recorded the timestamp of the threshold crossing as well as the actual signal in a 1400μs window around the time of crossing. Signals that did not cross the threshold were discarded. The recorded waveforms were then analyzed offline and sorted in to single units using the Offline Sorter software package. This software takes each waveform and calculates 12 different features (e.g. 3 principal component projections, peak-valley ratio, waveform energy, etc.). Combinations of these features are then plotted against each other in two or three dimensions. Clusters of waveforms are grouped together manually and assigned to single units (Figure 2.3).

We randomly sampled neurons; no attempt was made to select neurons based on responsiveness. This procedure ensured an unbiased estimate of neuronal activity thereby allowing a fair comparison of neuronal properties between the different brain regions. Our recordings from DLPFC were dorsal to the principal sulcus consisting of areas 9 and

9/46. Our recordings from VLPFC were ventral to the principal sulcus consisting of areas 47/12 and 45. Our recordings from OFC were located between the lateral and medial orbital sulci consisting of areas 11 and 13. Recordings from FEF were located in the convex area between the superior and inferior arcuate sulci.

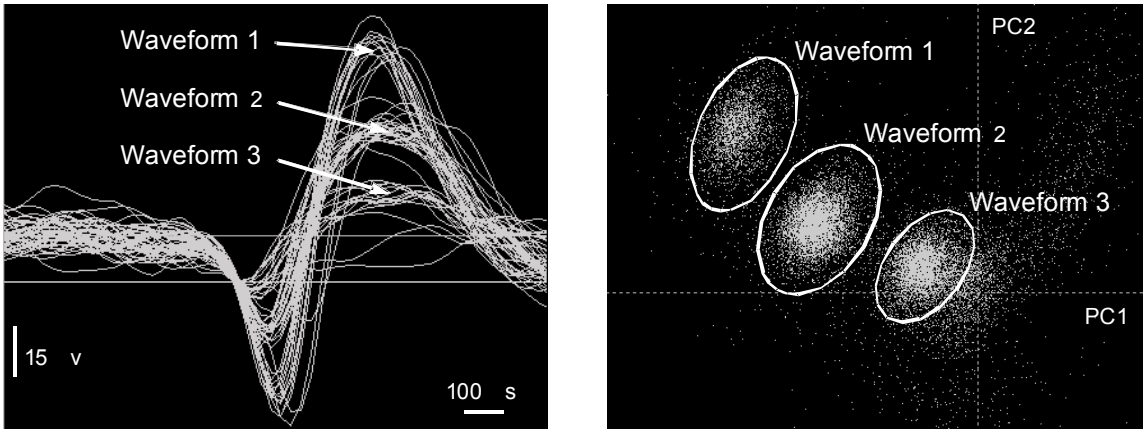


Figure 2.3: a) Trace illustrating one second of recording from a single channel with three neuronal waveforms. b) Subsequent offline cluster analysis of this same channel showing the plot of principal component projection PC1 against PC2 for this channel. The three clusters sorted into separate units by drawing a circle around them manually.

2.3 Statistical Analysis

All analyses were made using MATLAB 2007b (Mathworks, Natick, MA). Neuronal responses were analyzed by splitting the trial in to several epochs (typically sample and delay epochs) and calculating the neuron's average spike-count firing rate in each epoch. We examined the manner in which stimuli are encoded, both at the time of presentation, and subsequently in working memory by looking at neuronal activity during both sample and delay epochs. Baseline activity was assessed using the average activity during the initial 500 ms of the fixation epoch at the beginning of the trial. By having the baseline data we can compute the neuron's firing rate during the sample and delay epochs either in absolute terms, or expressed as a percentage of the baseline firing rate.
Trials were grouped into different conditions depending on the task and selectivity was assed using a general linear model using the mean firing rate as the dependent variable. From the general linear model we were able to calculate the strength of selectivity by determining the percentage of variance in the neuron's firing rate that can be explained by the manipulation in question.

Chapter 3

Representation of gustatory stimuli in working memory by orbitofrontal neurons

3.1 Introduction

There are two prominent models regarding the functional organization of working memory in prefrontal cortex (PFC). The domain-specific model states that PFC areas receiving direct projections from a specific sensory modality are responsible for maintaining and manipulating information about that modality in working memory. The operational model postulates that two PFC regions, dorsolateral (DLPFC) and ventrolateral (VLPFC), manipulate and maintain respectively all types of sensory information within working memory. Previous studies focused on visual, auditory and somatosensory modalities, which project to VLPFC making it difficult to distinguish the two models. In contrast, gustatory information enters PFC via orbitofrontal cortex (OFC). Thus, the domain-specific model would argue that OFC should be responsible for maintaining gustatory information, while the operational model would argue that VLPFC should maintain gustatory information. To distinguish these models, we recorded the activity of single neurons throughout PFC and gustatory cortex (GUS) from two subjects while they performed a gustatory delayed-match-two-sample task with intervening gustatory distraction. Neurons that encoded the identity of the gustatory stimulus across the delay, consistent with a role in gustatory working memory, were most prevalent in OFC and GUS. Gustatory information in OFC was more resilient to intervening distraction, paralleling previous findings regarding visual working memory processes in PFC and posterior sensory cortex. Our findings provide support for the domain-specific model of working memory organization. Maintaining gustatory information may be one aspect of a wider function for OFC in reward working memory that could contribute to its role in decision-making.

3.2 Materials and Methods

3.2.1 The task

In order to distinguish between the process model and the content model of working memory organization in PFC, we trained two rhesus monkeys (*Macacca mulatta)* to perform a gustatory delayed match to sample (DMS) task. To start the task the subject held a lever for 1000-ms. A picture appeared on the screen for 650-ms followed by 500-ms of juice delivery to the subject's mouth (Fig. 3.1). The picture helped warn the subject

of the juice's delivery and enabled us to examine neuronal selectivity relating to the expectancy of receiving a specific juice.

However, the subject could not use it to solve the task (see below). The picture remained on the screen until the end of the juice delivery. We used three different juice stimuli: each of the juices was predicted by one of two pictures (requiring six pictures total) thereby enabling us to distinguish neuronal responses related to the expectancy of a specific juice from neuronal responses related to encoding the visual properties of the picture. The juices we used were orange juice (Berkeley Farms Inc., Berkeley, CA), vegetable juice (V8 CSC Brands, Camden, NJ) and 0.1 M sucrose solution mixed with guava flavoring (LorAnn Oils, Lansing, MI). The juice entered the subject's mouth through a bundle of stainless steel cannulae (13 gauge, Small Parts, Miami Lakes, FL) to prevent cross-contamination of juice flavors. After a 1250-ms delay, we delivered 500-ms of water. This served as a distractor and helped ensure that the subject's mouth was rinsed of any remnants of the first juice. After a second 1250-ms delay, a second picture appeared for 650-ms. It was followed by 500-ms of juice delivery. On half the trials, the second juice was the same as the first juice (match), while on the other half of the trials it was one of the two other juices (non-match). After a delay of 800-ms, the subject indicated whether the two juices were the same or different by releasing (match) or holding (non-match) the lever. Thus, in order to solve the task the subject must hold the first juice stimulus in working memory for the 3000-ms delay between juices. We rewarded correct responses with 2000-ms of water. Incorrect responses resulted in the screen turning red and 10-s and the subject received no reward. There was a 3000-ms inter-trial interval (ITI).

To determine the neuronal response purely to juice stimuli, on 30% of the trials the subject received a 500-ms juice stimulus outside of the context of the task. We term these "free-juice" trials. These trials were randomly interspersed with task trials throughout the session. All three juice stimuli were given and the subjects could not predict when these trials would occur.

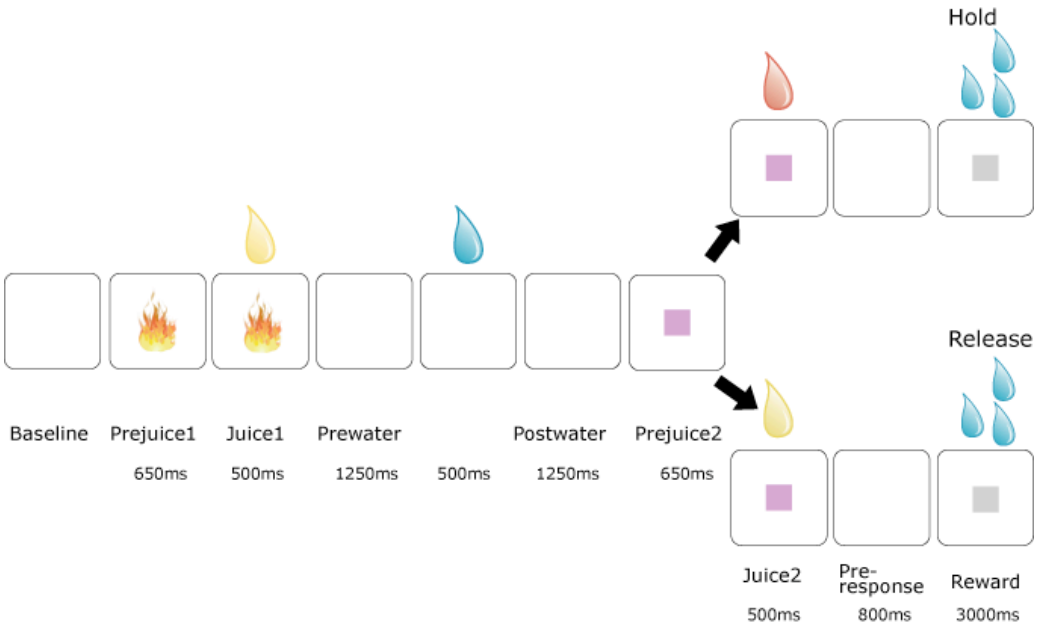


Figure 3.1. Illustration of the behavioral task. The subject grasped a lever to initiate the trial. We presented two juices sequentially separated by a brief delay. The subject had to release a lever if the two juices were the same (match) or continue holding the lever if the two juices were different (non-match). We presented a water drop half way through the delay which served as a distractor and ensured that remnants of the first juice were rinsed from the subject's mouth.

3.2.2 Neurophysiological recordings

Our recordings from DLPFC were dorsal to the principal sulcus consisting of areas 9 and 9/46. Our recordings from VLPFC were ventral to the principal sulcus consisting of areas 47/12 and 45. Our recordings from OFC were located between the lateral and medial orbital sulci consisting of areas 11 and 13. We recorded simultaneously from the different brain areas using arrays of 8-16 electrodes. For a full description of our neurophysiological techniques see Chapter 2.

3.2.3 Reconstruction of recording locations

We reconstructed our recording locations by measuring the position of the recording chambers using stereotactic methods. We plotted the positions onto the MRI sections using commercial graphics software (Adobe Illustrator CS3). We traced and measured the distance of each recording location along the cortical surface from the lip of the ventral bank of the principal sulcus. We also measured the positions of the other sulci,

relative to the principal sulcus, allowing the construction of the unfolded cortical maps shown in Figure 3.2.

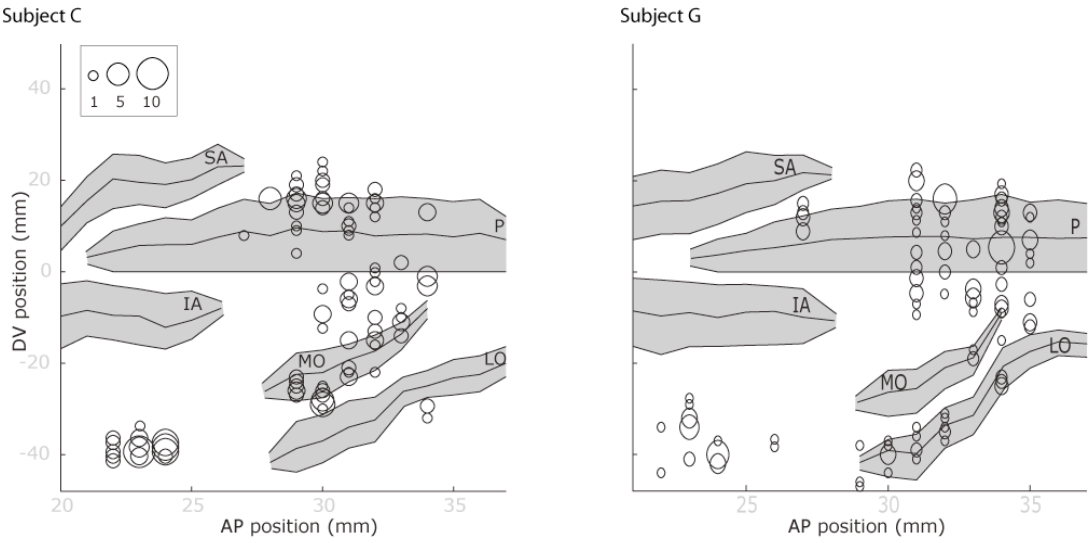


Figure 3.2. Flattened reconstructions of the cortex indicating the locations of recorded neurons in DLPFC and VLPFC. The size of the circles indicates the number of neurons recorded at that location. We measured the anterior-posterior position from the interaural line (x-axis), and the dorsoventral position relative to the lip of the ventral bank of the principal sulcus (0 point on y-axis). Gray shading indicates unfolded sulci. SA = superior arcuate sulcus, IA = inferior arcuate sulcus, P = principal sulcus, LO = lateral orbital sulcus, MO = medial orbital sulcus.

3.2.4 Statistical methods

We only analyzed correct trials. There were too few error trials to permit meaningful analysis of neuronal activity on those trials. We constructed spike density histograms by averaging activity across the appropriate conditions using a sliding window of 100-ms.

We quantified neuronal selectivity during the gustatory DMS using several defined time epochs. To analyze selectivity relating to the expectancy of receiving a specific juice reward, we defined the *pre-juice1* epoch beginning 100-ms after the first picture appears on the screen and lasting 550-ms until the delivery of the first juice (we discarded the first 100-ms to account for the latency of neuronal responses in PFC). To analyze selectivity

relating to the initial identification of the juice, we defined the *juice1* epoch as beginning 250-ms after onset of the first juice and lasting 500-ms until 250-ms after the first juice offset. The 250-ms offset allowed for the latency of gustatory responses as determined from the free juice trials (see results). To analyze selectivity relating to gustatory working memory, we defined two epochs either side of the distracting water drop. The *pre-water* epoch starts 250-ms after the offset of the first juice and lasts 1000-ms until the delivery of the water drop. The *post-water* epoch starts 250-ms after the offset of the water drop and lasts 1000-ms until the appearance of the second picture on the screen (Figure 3.1).

For each neuron and each epoch in turn, we calculated the neuron's mean firing rate. We performed a 2-way ANOVA on the firing rate, using factors of Juice (the identity of the first juice) and Set (the picture set from which we drew the reward-predictive cue), evaluating significance using an alpha level of $p<0.01$. We then used χ^2 tests to determine whether there were significant differences in the prevalence of neurons encoding the different experimental factors in different brain areas. We also used the 2-way ANOVA to calculate the strength of selectivity. We did this by calculating the magnitude of our statistical effects using eta-squared. This is equivalent to the percentage of explained variance attributable to the juice manipulation (PEVjuice).

To analyze the free-juice trials, we defined two time epochs. The *juice* epoch consisted of the 500-ms of juice delivery. The *post-juice* epoch started 250-ms after the offset of the juice and lasted 1000-ms. This epoch was designed to be comparable to the *pre-water* epoch in the gustatory DMS task. For each neuron and each epoch, we calculated the neuron's mean firing rate. To determine the prevalence of neurons encoding the juice, we performed a 1-way ANOVA on the firing rate, using the factor of the juice's identity. We also used the free-juice trials to determine the latency at which neurons encoded the identity of the juice. We took a 200-ms window of time, beginning 500-ms before the onset of juice, and performed a 1-way ANOVA on the firing rate, using the factor of the juice's identity. We then moved the window forward by 10-ms, and repeated the analysis. We continued in this fashion until we had analyzed up to 500-ms after the offset of the juice. We defined the latency of selectivity as the time when the *p*-value fell below 0.005 for three consecutive time bins. We chose this criterion by comparing it with the selectivity evident in the spike density histograms. However, to verify that it was indeed a sensible criterion, we examined how many neurons would have reached the criterion in the 1000-ms before the delivery of the juice (i.e. when it would have been impossible for the neurons to encode the juice's identity). Just 19/392 (4.8%) neurons reached criterion in this time period, indicating that our choice of criterion yielded a reasonable type I error.

To analyze the processes that underlie the subject's decision as to whether or not the two presented juices match, we focused on the period of the trial that followed the presentation of the second juice. We defined two time epochs. The *juice2* epoch began 250-ms after the onset of the second juice delivery and lasted 500-ms until 250-ms after the offset of the juice. The *pre-response* epoch began 250-ms after the offset of the juice

and lasted 650-ms until 100-ms after the onset of the go cue. For each neuron and each epoch, we calculated the neuron's mean firing rate. We performed a 2-way ANOVA on the firing rate, using factors of Juice (the identity of the second juice) and Response (whether the subject would indicate match or non-match). We calculated the magnitude of our statistical effects using eta-squared.

To calculate the latency at which the neurons determined the behavioral response, we performed a 2-way ANOVA using factors of Juice (the identity of the second juice) and Response (whether the subject would indicate match or non-match) as the experimental factors. We took a 200-ms window of time, beginning 500-ms before the onset of the second juice and performed the 2-way ANOVA on the firing rate. We then move the window forward by 10-ms, and repeated the analysis, continuing in this fashion until we had analyzed up to the presentation of the go cue. We defined the latency of selectivity as the time when the *p*-value for the main effect of Response fell below 0.005 for three consecutive time bins. We chose this criterion by comparing it with the selectivity evident in the spike density histograms. However, to verify that it was indeed a sensible criterion, we examined how many neurons would have reached the criterion in the 1000-ms before the delivery of the second juice (i.e. when it would have been impossible for the subject to know the correct behavioral response). Just 19/392 (4.8%) neurons reached criterion in this time period, indicating that our choice of criterion yielded a reasonable type I error.

3.3 Results

3.3.1 Behavior

Both monkeys performed the task well above the chance level of 50% (C 82 ± 5.7% over 40 recording sessions, G 84 ± 5.1% correct over 18 recording sessions). Subject C completed a mean of 377 correct trials (± 91 trials) per session and subject G completed 419 correct trials (± 41 trials) per session. Subject C showed significantly better performance when he had to remember guava (86 ± 7.7%) or tomato (85 ± 8.2%) compared to orange (76 ± 9.4%, 1-way ANOVA, $F_{2,117} = 18$, $p<5x10^{-7}$). Subject G showed no differences in performance between the three juices (orange: 81 ± 10%, guava: 84 ± 7.4%, tomato: 87 ± 5.0%; 1-way ANOVA, $F_{2,51} = 2.1$, $p>0.1$). Both subjects performed better on non-match conditions compared to match (C 91 ± 4.5% non-match, 74 ± 12% match, 1-way ANOVA, $F_{1,78} = 68$, $p<5x10^{-12}$; G 89 ± 4.2% non-match 79 ± 11% match, 1-way ANOVA, $F_{1,34} = 13$, $p<0.005$). Subject C had a median reaction time of 336 ± 66-ms, while G had a median reaction time of 251 ± 27-ms.

We also conducted a behavioral probe test in order to confirm that the subjects were solving the task in the manner we anticipated. We ensured that the subjects were not solving the task by remembering the identity of the reward-predictive pictures rather than the identity of the juice. Such an explanation would be necessarily convoluted, since

although they could remember the picture, they would still need to recall by the end of the delay what juice was paired with that picture in order to determine whether or not the second juice matched the first. Nevertheless, we tested whether the subjects had adopted this strategy by replacing the reward-predictive pictures with gray squares for one session. This behavioral manipulation had little effect on our subject's performance. Subject C performed at 84% (compared to 82% with the pictures), while subject G performed at 76% (compared to 84% with the pictures). Thus, the subjects did not appear to be using the reward-predictive cues to solve task.

3.3.2 Neurophysiological results

We recorded neuronal activity simultaneously from four cortical areas, DLPFC, VLPFC, OFC and GUS. We recorded from 392 neurons. There were 141 cells in DLPFC (C 58, G 83), 89 cells in VLPFC (C 40, G 49), 76 cells in OFC (C 46, G 30) and 86 cells in GUS (C 54, G 32).

3.3.3 Neuronal responses during the gustatory DMS task

For each neuron and each of the first four epochs in turn, we used a 2-way ANOVA to determine whether the neuron encoded the identity of the juice or the visual properties of the reward-predictive cue (see Methods). Very few neurons showed evidence of encoding the visual properties of the reward-predictive cue. Only 21/392 (5%) of the neurons showed a main effect of picture set in at least one of the four epochs. This did not significantly differ from what would have been expected by chance (0.01 alpha level x 4 epochs = 0.04, binomial, $p>0.1$). A similar number of neurons (16/392 or 4%) showed a significant interaction between juice and picture set in at least one of the four epochs. This proportion also did not significantly differ from what would have been expected by chance (binomial, $p>0.1$). Consequently, the remainder of our report will focus on those neurons that showed a main effect of juice, but with no significant main effect or interaction with picture set.

3.3.4 Reward expectancy

We found neurons that showed a selective modulation of their firing rate during the pre-juice1 epoch depending on which juice was being held in working memory. Figure 3.3A shows a neuron that has a significantly higher firing rate prior to the onset of the first juice only when the juice to be delivered was orange juice (2-way ANOVA, $F_{2,193} = 15.2$, $p<1x10^{-6}$). Such a neuron is consistent with previous reports detailing the encoding of expected rewards in PFC (Watanabe 1996; Tremblay and Schultz 1999; Hikosaka and Watanabe 2000). This type of encoding was a lot more prevalent in PFC than GUS (Table 3.1). A statistical comparison of the proportion of selective neurons in the different brain areas confirmed that all the comparisons between PFC and GUS were significant ($\chi^2>5$, $p<0.05$) while none of the comparisons between the different PFC areas reached

significance (χ^2<1, p>0.1). Indeed, the proportion of selective neurons in GUS did not exceed that expected by chance (binomial test, p>0.1).

	BEHAVIORAL TASK				FREE JUICE	
	Pre-juice	Juice1	Pre-water	Post-water	Juice	Post-juice
DLPFC	10	10	11	6	3	10
VLPFC	12	17	11	9	6	13
OFC	13	16	21	18	11	23
GUS	*1*	15	31	17	6	36

Table 3.1. Percentage of juice-selective neurons during the first four epochs of the behavioral task, and the two epochs of the free juice trials. Numbers in italics indicate that the proportion of selective neurons did not exceed that expected by chance (binomial test, p>0.1).

3.3.5 Juice identification

Many neurons encoded the identity of the first juice. The neuron in Figure 3.3B shows a significantly higher firing rate during the juice1 epoch for guava and tomato juice compared to orange juice (2-way ANOVA, $F_{2, 270} = 14.1$, $p<1x10^{-5}$). Neurons of this type were present in all four cortical areas from which we recorded (Table 3.1) with no significant differences among areas (all comparisons χ^2 <1.8, p>0.05).

3.3.6 Gustatory working memory

Many neurons showed a selective modulation of firing rate during the delay dependent on the identity of the first juice, consistent with a role in gustatory working memory. Figure 3.3C shows such a neuron. It shows a significantly higher firing rate when the subject was holding orange or tomato juice in working memory compared to guava juice ($F_{2,222}$=10, p<0.0001). After the water drop, the selectivity disappeared ($F_{2,222}$<1, p>0.1). However, not all neurons lost their selectivity after the water drop. The neuron shown in Figure 3.3D had a significantly higher firing rate when the subject was holding orange or tomato juice in working memory compared to guava juice ($F_{2,280}$=150,

$p<1x10^{-16}$). The difference in firing rate is still evident after the intervening water drop ($F_{2,280}=85$, $p<1x10^{-16}$).

We compared the proportion of juice-selective neurons between the different areas during both delay epochs. Such neurons were common in GUS and OFC, but were relatively infrequent in DLPFC and VLPFC. We used chi-squared tests to determine whether the differences in the prevalence of these neurons were significant. During the pre-water epoch there were significantly more selective neurons in GUS compared to DLPFC and VLPFC (both comparisons, $\chi^2 >9$, $p<0.01$). However, there was no significant difference between GUS and OFC ($\chi^2 =1.7$, $p>0.1$). Finally, there were no significant differences between DLPFC, VLPFC and OFC (all comparisons, $\chi^2 <3.6$, $p>0.05$). During the post-water epoch, there was a significant difference between DLPFC and both OFC and GUS ($\chi^2 >4$, $p<0.05$). All other comparisons were not significant ($\chi^2 <2.4$, $p>0.05$).

Next, for each area in turn, we examined whether there was a significant difference in the proportion of selective neurons during the pre- and post-water epochs. For all three PFC areas, there was no significant difference in the prevalence of the juice-selective neurons during the pre- and post-water epochs (all comparisons $\chi^2<1.1$, $p>0.05$). However, in GUS there were significantly fewer selective neurons in the post-water epoch compared to the pre-water epoch ($\chi^2=3.94$, $p<0.05$). Thus, PFC neurons were better able to maintain information about the juice across the intervening water drop than GUS neurons.

To determine the strength of neuronal selectivity (as opposed to its prevalence), we calculated PEVjuice for all neurons during the pre- and post-water epochs (see Methods). Figure 3.4 shows the mean value of this measure for each of the four brain areas and two time epochs. Neurons in OFC and GUS encoded the juice during the delay more strongly compared to DLPFC and VLPFC. In addition, OFC neurons encoded the juice in WM with approximately the same strength in the pre- and post-water epochs while GUS neurons showed weaker selectivity during the post-water epoch compared to the pre-water epoch. To analyze these effects statistically, we performed a 2-way ANOVA with Area (the four different brain areas) and Epoch (pre- and post-water) as factors. There was a significant Area x Epoch interaction ($F_{3,776}=3.3$, $p<0.05$) which we characterized by analyzing the simple effects and using post-hoc Tukey-Kramer tests (evaluated at $p<0.05$). There was a significant simple effect of Area on both the pre-water ($F_{3,776}=14$, $p<5x10^{-9}$) and post-water epochs ($F_{3,776}=2.9$, $p<0.05$). Post-hoc tests revealed that during the pre-water epoch, mean PEVjuice in GUS was significantly higher than DLPFC and VLPFC and significantly higher in OFC than DLPFC. For the post-water epoch, mean PEVjuice in OFC was significantly higher compared to both DLPFC and VLPFC but not GUS. The simple effect of Epoch on each of the four Areas was only significant for GUS ($F_{1,776}=15$, $p<0.0005$, all other areas, $p>0.1$). GUS neurons had significantly weaker selectivity during the post-water epoch compared to the pre-water epoch, while in the other areas the strength of selectivity remained unchanged after the intervening water distracter.

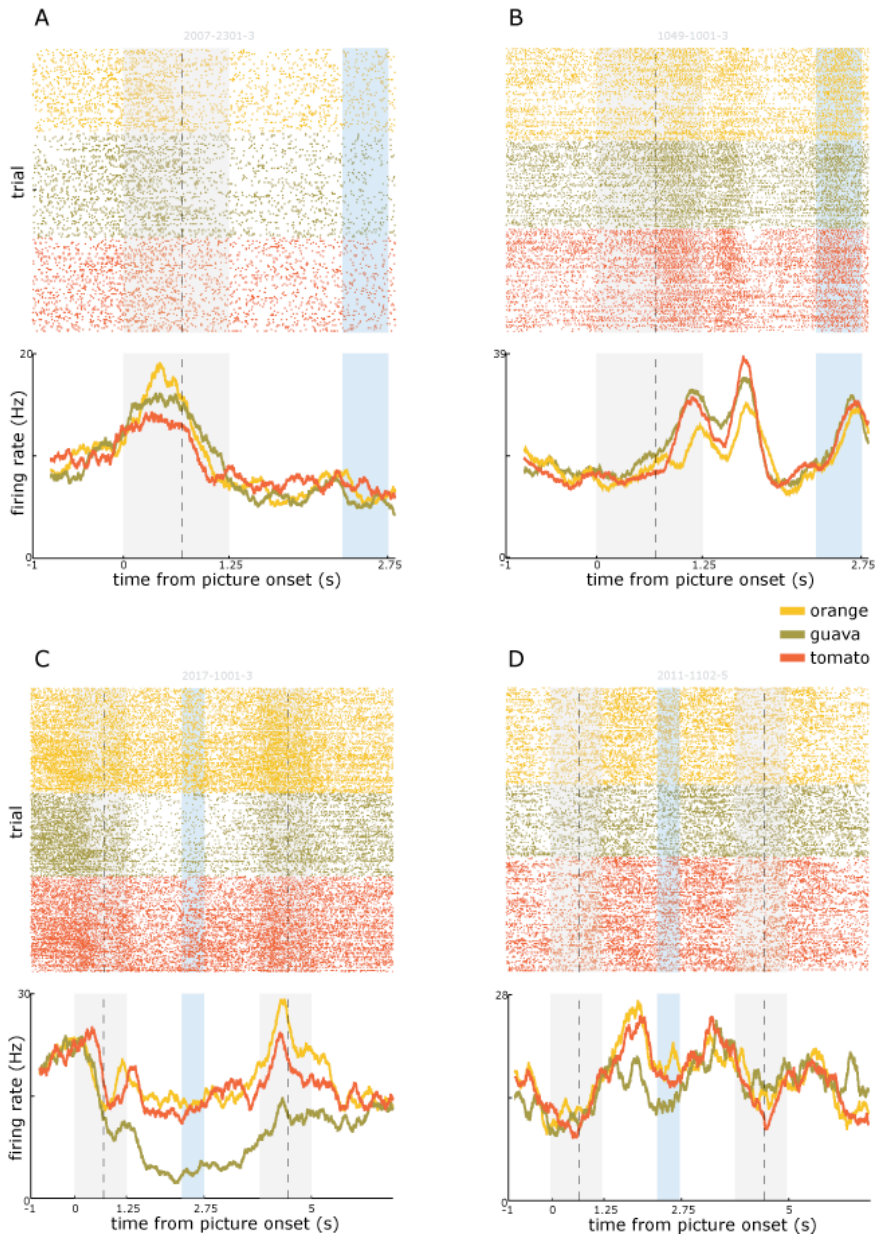


Figure 3.3. A) Spike density histogram illustrating an OFC neuron that had a significantly higher firing rate when the subject expected to receive orange juice. The gray shading illustrates the pre-juice and juice epochs with the vertical dotted line indicating the onset of juice delivery. The blue shading illustrates the delivery of the water drop. B) An OFC neuron that had a significantly higher firing rate to the delivery of guava and tomato juice relative to orange juice. C) A GUS neuron that had a significantly higher firing rate to the delivery of orange juice and tomato juice relative to guava juice that lasted through the *pre-water* epoch. The encoding disappeared following the delivery of the water drop. D) An OFC neuron

that showed a significantly higher firing rate to the delivery of orange juice and tomato juice relative to guava juice. The encoding lasted throughout the *pre-water* epoch and survived the delivery of the water drop.

3.3.7 Characterization of the neuronal encoding of the juices

There appeared to be two distinct ways of encoding the juices prominent in our neuronal populations. Some neurons, such as the neuron in Figure FFA, showed an increase in firing rate for one of the juices relative to the other two juices. We refer to such encoding as "standard selectivity" since this is the pattern one typically envisages when discussing neuronal selectivity. In contrast, other neurons, such as those in figures FFB-D, showed a lower firing rate to one of the juices relative to the other two juices. We refer to such encoding as "inverse selectivity". We determined the relative prevalence of these two types of selectivity in each brain area. For each neuron that showed selectivity in any of the four epochs (as defined by the 2-way ANOVA) we determined the epoch in which it showed maximal selectivity (calculated via PEVjuice). We then calculated its mean firing rate for each of the three juices in turn as well as its overall mean firing rate during this epoch. We determined the number of neurons for which the mean firing rate of just one of the juices exceeded the overall mean firing rate (consistent with standard selectivity), and the number of neurons for which the mean firing rate of two of the juices exceeded the overall mean firing rate (consistent with inverse selectivity). In DLPFC and OFC, approximately half the neurons showed the standard pattern of selectivity while the remainder showed the inverse pattern (DLPFC: 18/36 or 50%, OFC: 15/29 or 52%, binomial test, $p>0.1$ in both cases). In VLPFC and GUS, the standard pattern of selectivity dominated (VLPFC: 23/31 or 74%, GUS: 25/36 or 69%, binomial test, $p<0.05$ in both cases).

We also examined whether there was any consistency in the rank order of mean firing rates elicited by the three different juices. We did this for each neuron that showed selectivity in any of the four epochs, once again focusing on the epoch that showed maximal selectivity. In subject G, all six possible rankings of the three juices were equally prevalent and none occurred any more frequently than would be expected by chance (expected frequency=0.17, binomial test, $p>0.1$ in all cases). In subject C, there was a slight tendency for neurons to show their lowest firing rate to orange and their highest firing rate to guava (18/69 or 26%, binomial test, $p<0.05$). To examine whether this related to the subject's preferences among the juices we performed a discrimination task in which the subject selected pictures associated with the different juices using a protocol that we previously designed to examine choice behavior (Kennerley, Dahmubed et al. 2009). This revealed that the subject preferred orange juice over guava and guava over tomato. In summary, there was little evidence to suggest that neurons favored a particular encoding of the juices, or that subjective preferences mapped onto neuronal preferences in this particular task. Although we did not break the analysis down by brain area (since

there would have been too few neurons encoding any specific rank order) there did not appear to be any encoding of rank order in any of the areas.

3.3.8 Neuronal responses to juice stimuli outside of the behavioral task

Neuronal responses to the juices delivered on free juice trials enabled us to examine how quickly the different brain areas identified the juices, since on these trials there was no visual cue to alert the subject as to which juice would occur. During the juice epoch, only a small number of neurons encoded the identity of the juice (Table 3.1) and there was no significant difference between the areas in terms of the proportions of juice-selective neurons (all comparisons, $\chi^2<3.1$, $p>0.05$). During the post-juice epoch, there were significantly more juice-selective neurons in GUS compared to both DLPFC and VLPFC (both comparisons, $\chi^2>10$, $p<0.005$), and significantly more in OFC compared to DLPFC ($\chi^2=5.3$, $p<0.05$). The remaining comparisons among areas were not significant (all comparisons, $\chi^2<2.2$, $p>0.05$). We compared the proportion of selective neurons during the post-juice epoch of the free reward trials with the pre-water epoch in the behavioral task. We found that there was no significant difference between the number of neurons that showed juice selectivity in either epoch in any of the areas (all comparisons $\chi^2<1$, $p>0.1$).

We determined the latency at which neurons encoded the juices' identity using a sliding ANOVA analysis. We defined the latency of selectivity as the time when the p-value fell below 0.005 for three consecutive time bins (see Methods). The proportion of neurons that reached this criterion in DLPFC (5/141 or 3.3%) did not exceed that expected by chance (binomial test, $p>0.1$) and so we excluded them from the remainder of the analysis. In the remaining areas, 9/89 (10%) of VLPFC neurons reached criterion, 12/76 (16%) of OFC neurons and 16/86 (19%) of GUS neurons. There was no significant difference between the areas with regard to the time at which they first encoded the juice (VLPFC: 461 ± 81-ms, OFC: 511 ± 60-ms, GUS: 516 ± 70-ms, 1-way ANOVA, $F_{2,30}<1$, $p>0.1$).

3.3.9 Match/Non-match selectivity

We used a 2-way ANOVA with factors of Juice (the identity of the second juice) and Response (whether the subject indicated that the juice was a match or non-match) to characterize neuronal selectivity during the juice2 and pre-response epochs. During both the juice2 and the pre-response epochs, only 9/392 (2%) of the neurons showed a significant interaction between the two factors. Given the small proportion of such neurons, we did not analyze them further. Instead, we focused on the majority of the selective neurons that showed a main effect of Juice and/or Response.

During the juice2 epoch, the majority of the selective neurons showed a significant main effect of Juice (Table 3.2). However, there were no significant differences in the proportion of selective neurons between the areas (all comparisons, $\chi^2<2.7$, $p>0.05$). The

number of neurons that showed a main effect of Response during the juice2 epoch did not exceed that expected by chance in any area (binomial, p>0.1). Thus, neuronal selectivity during the juice2 epoch appeared to relate primarily to the identification of the presented juice.

During the pre-response epoch, neurons showing a main effect of Juice were only present in OFC and GUS. The proportion in DLPFC or VLPFC did not exceed that expected by chance (binomial, p>0.1). There were more juice-selective neurons in GUS than OFC and the difference in the proportions approached significance (χ^2=3.4, p=0.06). Many neurons also showed a significant main effect of Response. However, the only significant difference in the proportion of response-selective neurons was between OFC and DLPFC (χ^2=4.6, p<0.05); all other comparisons between areas were not significant (χ^2<1.7, p>0.05). There was a similar number of response-selective neurons that showed a higher firing rate when the subject had to release the lever (59%) as when the subject held the lever (41%). These proportions did not significantly differ from an even split (binomial test, p>0.1).

	Response		Juice	
	Juice2	Pre-response	Juice2	Pre-response
DLPFC	*1*	16	*1*	*1*
VLPFC	*3*	11	6	*3*
OFC	*0*	5	8	10
GUS	*2*	9	14	20

Table 3.2. Percentage of neurons that encoded the identity of the second juice or the upcoming behavioral response during the *juice2* and *pre-response* epochs. Numbers in italics indicate that the proportion of selective neurons did not exceed that expected by chance (binomial test, *p*>0.1).

As well as determining the proportion of selective neurons, we also quantified the strength of encoding by calculating the PEV due to both experimental factors (PEVjuice and PEVresponse) during the juice2 and pre-response epochs (see Methods). Figure FFB shows the strength of encoding of the juice (PEVjuice) for each of the four brain areas and two time epochs. It was stronger during the pre-response epoch than the juice2 epoch and strongest in GUS than the other brain areas. To analyze these effects statistically, we

performed a 2-way ANOVA on the PEVjuice values of the neurons with Area (the four different brain areas) and Epoch (juice2 and pre-response) as factors. There was a significant Area x Epoch interaction ($F_{3,775}$=3.2, $p<0.05$) which we characterized by analyzing the simple effects and using post-hoc ANOVA tests (evaluated at $p<0.05$). There was a significant simple effect of Area during the pre-response epoch ($F_{3,776}$=16, $p<1x10^{-9}$) but not the juice2 epoch ($F_{3,776}$=2.44, $p>0.05$). The effect during the pre-response epoch was due to a significantly smaller PEVjuice in DLPFC and VLPFC compared to GUS. No other comparisons between areas were significantly different. There was also a significant simple effect of Epoch in GUS ($F_{1,776}$=14, $p<0.005$), but not the other three areas (all tests, $F_{1,776}<2$, $p>0.1$).

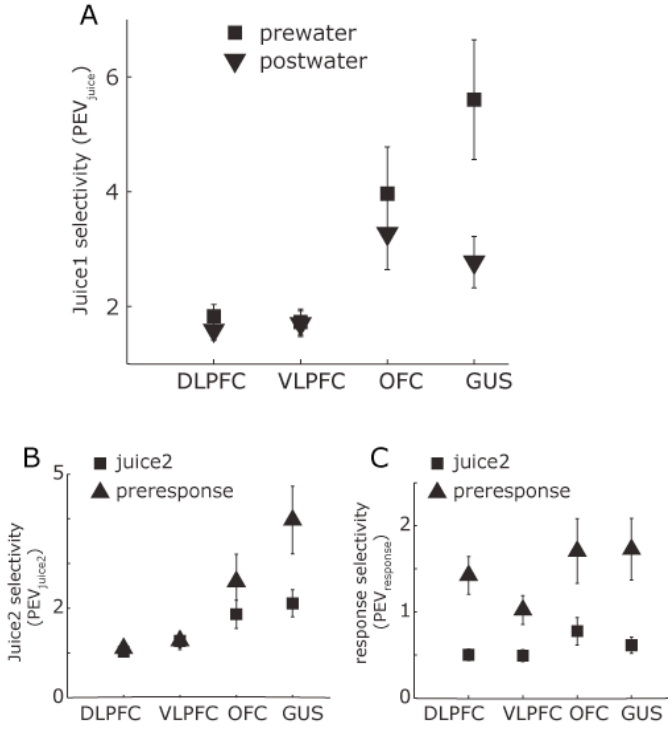


Figure 3.4. A) The mean strength of neuronal selectivity for the encoding of the gustatory stimulus (PEV_{juice}) during the *pre*-and *post-water* epochs. There was stronger encoding in OFC and GUS relative to DLPFC and VLPFC. In addition, OFC neurons encoded the juice in WM with approximately the same strength in the *pre-* and *post-water* epochs while GUS neurons showed weaker selectivity during the *post-water* epoch compared to the *pre-water* epoch. B) The mean strength of neuronal selectivity for the encoding of the gustatory stimulus (PEV_{juice}) during the *juice2* and *pre-response* epochs. Selectivity was stronger during the *pre-response* epochs than the *juice2* epochs and strongest in GUS than the other brain areas. C) The mean strength of neuronal selectivity for the

encoding of the behavioral response ($PEV_{response}$) during the *juice2* and *pre-response* epochs. Response selectivity was significantly higher during the *pre-response* epoch compared to the *juice2* epoch.

Figure FFC shows that encoding of the behavioral response (PEVresponse) was significantly stronger during the pre-response epoch compared to the juice2 epoch. To analyze these effects statistically, we performed a 2-way ANOVA on the PEVresponse values of the neurons with Area (the four different brain areas) and Epoch (juice2 and pre-response) as factors. There was a significant main effect of Epoch ($F_{1,775}$=34, $p<1x10^{-8}$), but no significant main effect of Area ($F_{3,775}$=1.9, $p>0.1$) nor a significant Area x Epoch interaction ($F_{3,775}<1$, $p>0.1$).

We determined the latency at which neurons encoded the behavioral response using a sliding ANOVA analysis. We defined the latency of selectivity as the time when the p-value fell below 0.005 for three consecutive time bins (see Methods). A similar proportion of neurons that reached this criterion in all four new areas (DLPFC: 37/141 or 26%, VLPFC: 23/89 or 26%, OFC: 24/76 or 32% and GUS: 30/86 or 35%). There was no significant difference between the areas with regard to the time at which they first encoded the behavioral response (DLPFC: 686 ± 38-ms, VLPFC: 665 ± 65-ms, OFC: 624 ± 62-ms, GUS: 616 ± 52-ms, 1-way ANOVA, $F_{3,110}<1$, $p>0.1$).

3.4 Discussion

To summarize, we found that many neurons in PFC showed differential firing rates during the delay epoch of a gustatory matching-to-sample task dependent on the identity of the first juice stimulus, consistent with a role for maintaining gustatory information in working memory. Furthermore, our findings are consistent with the idea that PFC is to some extent organized according to the content of the information that must be maintained across a delay, since there was a strong bias for neurons involved in gustatory working memory to be located in OFC rather than VLPFC or DLPFC. This shows that OFC, an area that is not typically associated with working memory, can in fact show stronger involvement in working memory than LPFC when the type of information that must be maintained relates to the type of information for which OFC is specialized to process. Indeed, this is the first time that neurons in an area outside of LPFC have shown stronger working memory selectivity than LPFC neurons themselves.

Demonstrating analogous effects in humans may prove difficult, since it may be impossible to prevent human subjects from encoding a gustatory stimulus with a verbal label. However, a recent neuroimaging study did conclude that OFC was particularly important for maintaining emotional information in working memory. Stern and colleagues used a matching-to-sample task with face stimuli (LoPresti, Schon et al. 2008). Subjects were required to make match judgments regarding either the identity or the emotional expression of the face. Within posterior sensory cortex, separate areas encoded

these two aspects of the facial stimulus. In contrast, OFC exhibited sustained delay activity for both emotion and identity trials. This suggests an important role for OFC in holding socially and emotionally relevant information in working memory, consistent with its stronger limbic connections relative to other PFC areas.

However, we must be careful in proposing too strong a content model of PFC functional organization. OFC and LPFC differ not just in the regions with which they connect, but also in terms of their intrinsic anatomical organization (Zald 2007). For example, LPFC has a well differentiated layer IV, which is absent in OFC (Barbas and Pandya 1989; Carmichael and Price 1994; Petrides and Pandya 1994). Such anatomical differences highlight the fact that it is unlikely that all PFC areas are performing the same process, merely differing in the type of information processed depending on their anatomical connections (Zald 2007). Instead, such differences suggest that different PFC areas perform fundamentally different computations.

Our results help to constrain how those computations might differ, since it is clear that OFC neurons are capable of maintaining information across a delay consistent with a role in working memory. However, the exact functional organization of PFC remains unclear. Perhaps the most striking functional difference between OFC and LPFC, from both neuropsychological and neurophysiological studies, relates to their involvement in decision-making and goal-directed behavior (Fellows and Farah 2005; Wallis 2007). For example, in a task that involves a simple choice between two stimuli associated with different behavioral outcomes, OFC shows robust encoding of multiple variables necessary to make a decision, which is noticeably absent in LPFC (Kennerley, Dahmubed et al. 2009).

3.4.1 Delay activity in sensory cortex

In addition to OFC, we also saw neuronal activity in GUS related to encoding of the identity of the juice stimulus across the delay. However, this activity differed from that in OFC in a number of key respects. First, it was noticeably weaker following the intervening distracting water drop. Miller and colleagues observed an analogous effect in the visual system. Neuronal activity related to the encoding of a visual stimulus across a delay occurs in both PFC and inferior temporal cortex. However, unlike PFC, intervening, distracting visual stimuli disrupted the delay activity in inferior temporal cortex (Miller, Erickson et al. 1996). From these results, Miller concluded that the ability to maintain information across distraction was an important property of PFC that differentiated this region from posterior sensory areas (Miller and Cohen 2001). Our results support this assertion, by showing that it applies to other sensory systems in addition to the visual modality.

Second, OFC delay activity differs from GUS delay activity in that its prevalence depends on the stage of the task. There was little difference between OFC and GUS in the prevalence and strength of encoding of the identity of the first juice during the *pre-* and

post-water epochs. In contrast, OFC encoding of the second juice during the *pre-response* epoch was noticeably weaker than in GUS. Thus, the encoding of gustatory information in OFC appears to depend on its relevance to the successful performance of the task. During the first delay, when the subject must remember the identity of the first juice, we see robust encoding of gustatory information in OFC. However, during the second delay the subject has had the opportunity to compare the first juice with the second and decide whether they match. Thus, he has simply to remember his intended behavioral response and so at this stage of the task gustatory information is irrelevant. The tendency to encode information only when it is behaviorally relevant is a hallmark feature of working memory that differentiates it from delay activity in posterior sensory cortex (Rao, Rainer et al. 1997; Rainer, Asaad et al. 1998; Duncan 2001; Everling, Tinsley et al. 2002).

3.4.2 Interpretational issues

Although we refer to our task as a gustatory delayed match-to-sample task, there is almost certainly a contribution of olfactory information. We chose our juice stimuli so that they were discriminable and palatable to our subjects rather than to dissociate the olfactory and gustatory components of the task. However, the underlying rationale of our experiment does not depend on whether our subjects solve the task by remembering the gustatory component of the juice, its olfactory component, or the integration of both gustatory and olfactory components ('flavor'). This is because OFC is the major PFC recipient of both gustatory and olfactory information (Morecraft, Geula et al. 1992; Barbas 1993; Carmichael and Price 1995; Cavada, Company et al. 2000) and it is the first site of convergence of these inputs that enables a representation of flavor (Rolls and Baylis 1994; Rolls 1996). Thus, irrespective of precisely what information about the juice the subjects are maintaining in order to solve the task, we would still expect to see a dissociation between LPFC and OFC depending on whether the process model or content model of working memory holds true.

There did not appear to be any consistent pattern in the order in which the neurons encoded the three different juices. This might appear to contradict recent studies suggesting that OFC plays an important role in encoding subjective preferences (Padoa-Schioppa and Assad 2006; Fellows and Farah 2007; Padoa-Schioppa and Assad 2008). However, our task did not require the subject to make judgments or choices on the basis of their preferences between the juices, but rather attend to the sensory characteristics and identity of the juices. This suggests that OFC neurons, like neurons in other PFC areas (Rao, Rainer et al. 1997; Duncan 2001; Miller and Cohen 2001), may be able to alter the aspect of the stimulus that they encode in response to the overarching task demands. Indeed, recent neuroimaging findings in humans show that activity in OFC is modulated by selectively attending to either the pleasantness or intensity of a gustatory stimulus (Grabenhorst and Rolls 2008).

All of the areas from which we recorded showed as strong encoding of the juice in the free-juice trials as they did in the behavioral task. At first, this might seem to suggest that encoding of the juice during the delay reflects a passive response to the gustatory stimulus, rather than the active maintenance of information in working memory. However, such a conclusion may be unwarranted. Simply because we do not require our subjects to hold the juice in working memory, does not mean that they are not doing so. It may be easier for our subjects to hold the juice in working memory on all trials, rather than trying to discriminate which trials will require working memory from those trials that will not. The work described in this chapter has been published in the Journal of Neuroscience (Lara, Kennerley et al. 2009).

Chapter 4

Capacity Limits of Visual Working Memory: A Behavioral Study

4.1 Introduction

The ability to hold multiple pieces of information in WM is paramount to complex cognitive abilities such as learning, problem solving and language comprehension (Cowan 2001; Cowan, Elliott et al. 2005) and it is highly correlated with general intelligence (Conway, Kane et al. 2003). This ability, however, is extremely limited in capacity (Miller 1956; Cowan 2001). Recent estimates place the limit on the number of items that can be simultaneously maintained at an average of 4 items with some subjects having a capacity as low as 2 items (Luck and Vogel 1997; Vogel, McCollough et al. 2005; Bays and Husain 2008; Zhang and Luck 2008). The precise nature of the neuronal mechanisms underlying this limit, however, is still a subject of great debate.

Several theoretical models have been proposed that attempt to explain how this limit might arise (Wilken and Ma 2004; Bays and Husain 2008; Rouder, Morey et al. 2008; Zhang and Luck 2008). From this work, two competing models with diverging predictions for mechanism of the limited capacity have been proposed. One model, the shared resources model, states that there is a limited resource pool that is shared amongst the items in working memory. When there is a single item in memory, all the resources will go towards representing it, which will in turn produce a very high precision memory trace. When the number of items increases the resources are split between the items at the expense of representational precision (Bays and Husain 2008). This model states that there is no capacity limit *per se*; rather, when the number of items in memory is large the representations become so imprecise that information about the items in memory is lost. An alternative model, the fixed-precision "slots" model, states that subjects have a limited number of memory "slots" that hold a fixed-precision memory trace. Items are placed in these slots until there are no more slots to form a memory trace. This gives rise to a definite upper limit on the number of items that can be placed in working memory (Zhang and Luck 2008).

Despite decades long interest in the topic of a limited capacity memory store and the recent advances in cognitive modeling, only recently has the storage of multiple items begun to be investigated at the neuronal level (Warden and Miller 2007; Siegel, Warden et al. 2009). However, a systematic investigation of the neuronal underpinning of the capacity limit and the exact nature of the precision of the memory representations is still

lacking. In this chapter we describe experiments we performed to estimate the capacity limit in visual working memory in the macaque monkey. In addition, we aimed to determine whether this limit arises from a loss of precision in the memory traces as predicted by Bays and Husain (2008) or if precision remains fixed as predicted by Zhang and Luck (2008) in which case we would not see adverse effects of increasing the memory load until the upper limit in the number of "slots" is reached. In Chapter 5 we will look at the neuronal underpinnings of the capacity limit by recoding from prefrontal cortex neurons and focusing on the changes that occur in neuronal encoding when a subject is remembering one item in working memory compared to two items.

4.2 Methods

In this section we will provide details of our experiments and analysis. These methods are relevant to this chapter and the next, both of which use the same behavioral task, with the only difference being the number of memoranda and the colors of the stimuli.

4.2.1 The task

We trained two subjects on a color change detection task adapted from the human literature (Luck and Vogel 1997), illustrated in Figure 4.1. Subjects sat in a primate chair in front of an LCD screen. At the start of the trial a fixation square (0.5 x 0.5º of visual angle) appeared in the center of the screen. Subjects had to maintain their gaze within 1.5º of the center of the fixation spot for 800 ms. After subjects achieved fixation, a sample array of 1 to 4 different colored squares appeared on the screen for 500 ms (Figure 4.1). At the end of the 500 ms sample period the array disappeared from the screen for 1000 ms. During this time, subjects had to keep the color of all the squares in working memory. At the end of the delay period one of the squares in the array was presented again and the subject had to indicate if the color of that square had changed or remained the same. The subjects indicated whether this was the case by moving a lever. Subjects were free to indicate their response as soon as the test square appeared on the screen. Correct responses were rewarded with 0.5 ml of orange juice delivered directly to the subjects' mouth. If the subject made an incorrect response the screen turned gray for 4 seconds to indicate an incorrect response and no reward was given. If at anytime during the sample or the delay periods subjects broke fixation, the entire screen turned red for 10 seconds, after which a new trial started. There was a 3000 ms inter trial interval between all trials.

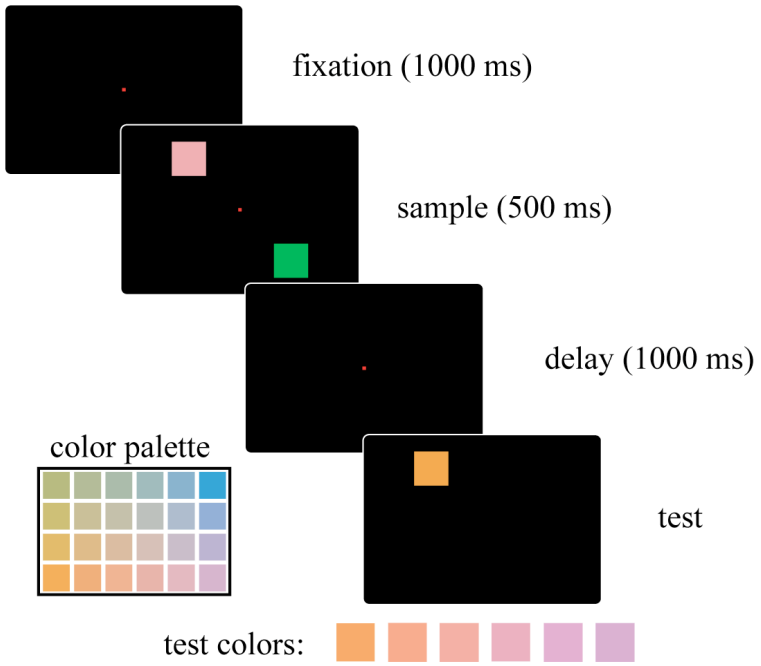


Figure 4.1. Visual change detection task. Trials start with 1 sec of fixation. An array of 1-4 colored squares appears for 500ms. A test color appears after a 1000ms delay. The distance between sample colors and the test color is varied parametrically. In this example, the test color can be chosen from any of the colors shown at the bottom of the figure. The inset shows the color palette used in the experiments discussed in this chapter.

4.2.2Stimuli

Stimuli were squares 3.5º x 3.5º of visual angle. They were presented in any of four fixed locations 5º away from the fixation spot. Colors were chosen from the 1976 *CIE L a* b** color space. This color space was designed by the Commission internationale de l'éclairage (International Commission on Illumination) to be perceptually uniform. In other words, moving away from a reference color in one direction is perceived as the same amount of change in color for any other direction as long as the distance from the reference color remains the same (Shevell 2003). Colors are specified by three parameters; the *L* parameter corresponds to luminance while the *a** and *b** parameters correspond to the color hue. In order to make all stimuli equiluminant we fixed the luminance at $L = 70$ and varied the *a** and *b** parameters to produce the colored squares. The *L a* b** color space has additional characteristic that makes it very well suited for our purposes, namely, distances between colors in this space correspond to perceived distances between colors by an observer. For example, consider any three colors *color1, color2* and *color3* on a plane with fixed luminance (Figure 4.1 inset, L=70); if the distance (ΔE) between *color2* and *color1* is the same as the distance

between *color2* and *color3*, an observer will judge *color2* to be equally different from *color1* and *color3*. Moreover, if, as shown in the figure, *color1* and *color3* are separated by the same distance *d*, the observer will judge *color1* and *color3* equally different from each other, as they are different from *color2*.

This property of the *L a* b** color space is particularly useful as it allows us to parametrically vary the distance between the sample and test objects in a uniform fashion for all colors avoiding any perceptual differences between colors that may arise. The distance between colors in this space, ΔE, is given by the geometrical distance between the two colors. In other words, if we have two colors, $color1 = \left(L_1, a_1^*, b_1^*\right)$ and $color2 = \left(L_2, a_2^*, b_2^*\right)$, then, the distance ΔE, between them is given by $\Delta E = \sqrt{\left(L_2 - L_1\right)^2 + \left(a_2^* - a_1^*\right)^2 + \left(b_2^* - b_1^*\right)^2}$. Using this, we constructed a set of stimuli to be used both as sample and as test colors and we restricted the distances between sample and test colors to the set $\Delta E \in$ (0, 40, 50, 60, 70, 80, 90, 100).

4.2.3 Analytical methods

To estimate the precision of internal memory representations we calculated the probability of subjects detecting a change in color as a function of distance between sample and test colors. We fitted a Gaussian function of the form $p = ae^{-\frac{\Delta E^2}{\sigma^2}}$, where *p* is the probability of detecting a change; ΔE is the distance between colors, σ is the standard deviation, and *a* is a scaling factor. We used nonlinear regression to compute the parameters using a robust fitting algorithm that iteratively reweighs the parameter values and re-computes a least-squares fit. All analyses were done in MATLAB (Mathworks Inc., release 2008b).

4.3 Results

We trained subjects on the visual change detection task shown in Figure 4.1. Both subjects learned the task well and could discriminate all colors in the palette. To determine if subjects had learned the task and could discriminate all colors, we calculated the subjects' performance for each color using trials in which only one item appeared on the screen and trials in which the difference between sample and test color was large ($\Delta E \geq 100$). Figure 4.2 shows the subjects' performance for trials in which each color was used as the sample. Both subjects were able to discriminate all the colors at a high level of performance. The mean performance for subject G was 91%; and 89.4 % for subject I, with performance on all colors > 82% and >70% respectively.

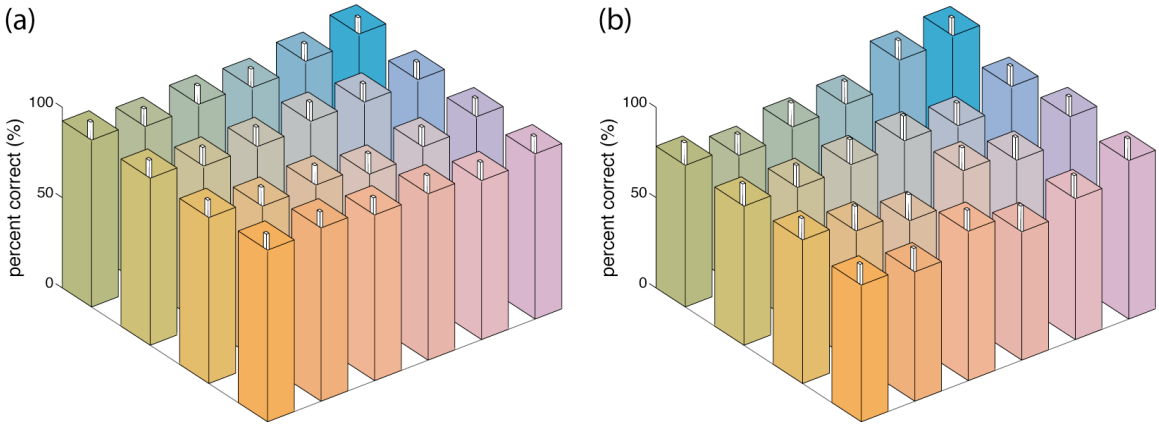


Figure 4.2. Subjects performance for each color. The height of the bar shows the percent correct while the color of the bar represents the color of the sample stimulus. The small line in the center of the bars indicate the standard error of the mean. Performance on all colors > 82% for subject G (a) and >70% for subject I (b).

4.3.1 Estimation of visual working memory capacity

To determine the capacity of visual working memory we calculated subjects' performance as a function of the number of items in memory (set size). Figure 4.3 shows that for both subjects as the set size increases, performance decreases significantly (ANOVA; subject G: $F_{3,13556}= 280.13$, $p < 2\text{x}10^{-16}$, subject I: $F_{3,1408}= 38.43$, $p < 2\text{x}10^{-16}$). For one-item trials, performance is close to 90% in both cases. However, when a second item is added to the display, performance falls significantly (*post-hoc* Tukey-Kramer test; both subjects $p<1\text{x}10^{-4}$). For both subjects, there is also a significant drop in performance between set size two and three (*post-hoc* Tukey-Kramer, $p<0.001$). Although adding a fourth color to the memory set, adversely affected performance in both subjects as well, both subjects' performance for set-size three was not significantly different from set-size four (*post-hoc* Tukey-Kramer test, $p>0.05$).

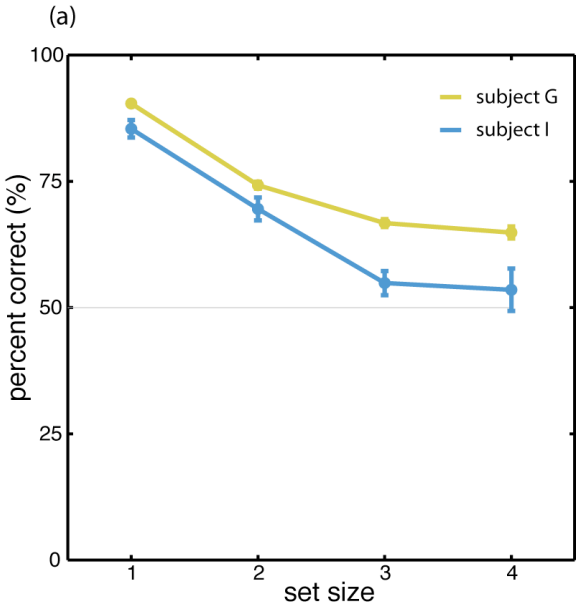


Figure 4.3. Performance as a function of set size for each subject. For both subjects, as the set size increased from one to four, performance significantly decreased. The gray line shows chance level.

4.3.2 Reaction time as a function of set size

After the memory period, a test color was presented on the screen and subjects were able to indicate their response as soon as the test appeared. We calculated the subjects' reaction time for each set size for all trials and it is shown in Figure 4.4. As the number of items in memory increased, the subjects took significantly longer to respond (ANOVA, subject G: $F_{3,14971} = 499.5$, $p<1\text{x}10^{-16}$ subject I: $F_{3,\ 7815} = 20.4$, $p<3\text{x}10^{-13}$). Both subjects' reaction time for *set-size one* was significantly faster compared to all other set sizes. For trials with set-size larger than one, adding one more item significantly slowed subjects' reaction time (subject G: $p<1\text{x}10^{-5}$; subject I: $p<0.05$). The reaction time difference between set size 3 and 4, however, was not significantly different.

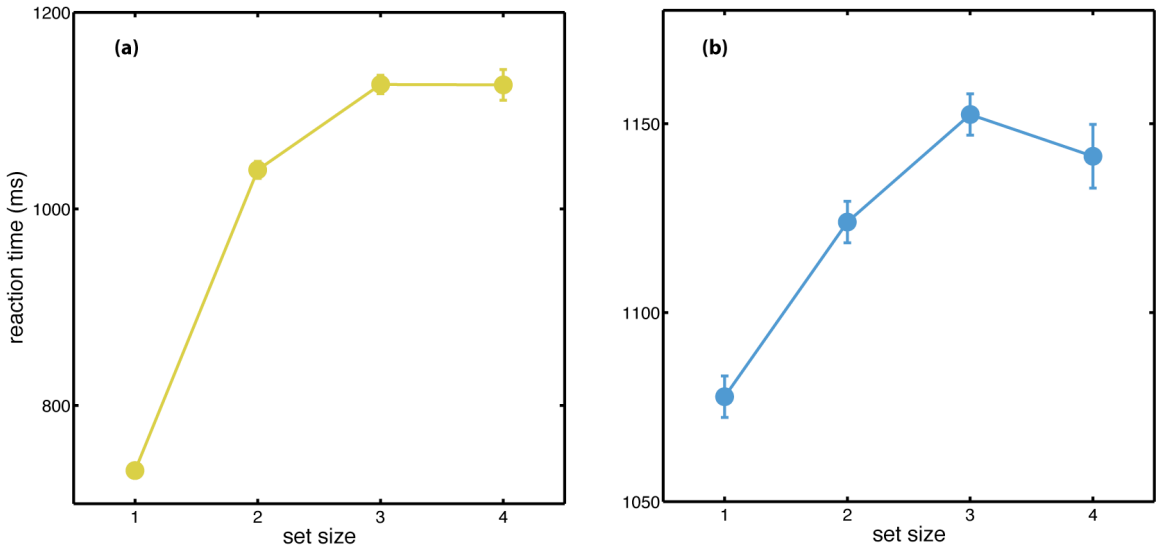


Figure 4.4. Reaction time as a function of set size. Both subjects were significantly faster for one-item trials compared to higher set sizes. **(a)** Subject G, **(b)** Subject I.

4.3.3 Performance at different ΔE

The results from the preceding sections were calculated using all trials. They included trials where the difference between the sample and the test colors ranged from very small (ΔE=40) to very high (ΔE=100). Performance on trials in which ΔE was small might be worse than on trials with large ΔE because a small ΔE might make it difficult for subjects to distinguish between the sample and test colors. Consequently a performance decrease might not reflect the subject's working memory capacity limit, but rather a difficulty in discriminating the colors. Thus, we investigated the effect of ΔE on performance by looking at subjects' performance as a function of ΔE separated by set-size (Figure 4.5).

For both subjects, and at all set-sizes, performance is not significantly better than chance at the lowest value of ΔE. Additionally, as set-size increases performance drops significantly below chance. This might indicate that subjects were unable to discriminate between colors separated by a distance of 40 units or less and were judging both colors to be the same (thus making more incorrect responses than pure guessing). Performance for the one-item trials, however, significantly improved for all other values of ΔE and eventually plateaus at around 93%.

For set-size of two, performance is significantly above chance only for trials with the highest ΔE (Binomial test $p<1.4\text{x}10^{-15}$ for ΔE=100). The same was true for three-

item trials with the exception of subject's performance at ΔE=100 which was marginally significantly above chance (binomial p<0.04). Indeed, both subjects performed above chance only when we used highly discriminable colors in trials with set-size of two and three. However, at set size of four, both subjects' performance was at or below chance level.

In summary, our subjects performed the task more accurately and more quickly as the difference between the sample and test colors increased. Furthermore, these effects became less pronounces as the set size of to-be-remembered items increased. Taken together these results suggest that as the set size increases, the precision with which items are stored decreases, a result that is more compatible with the resources model than the slots model. In the next section, we will quantify how precision changes as a consequence of set size.

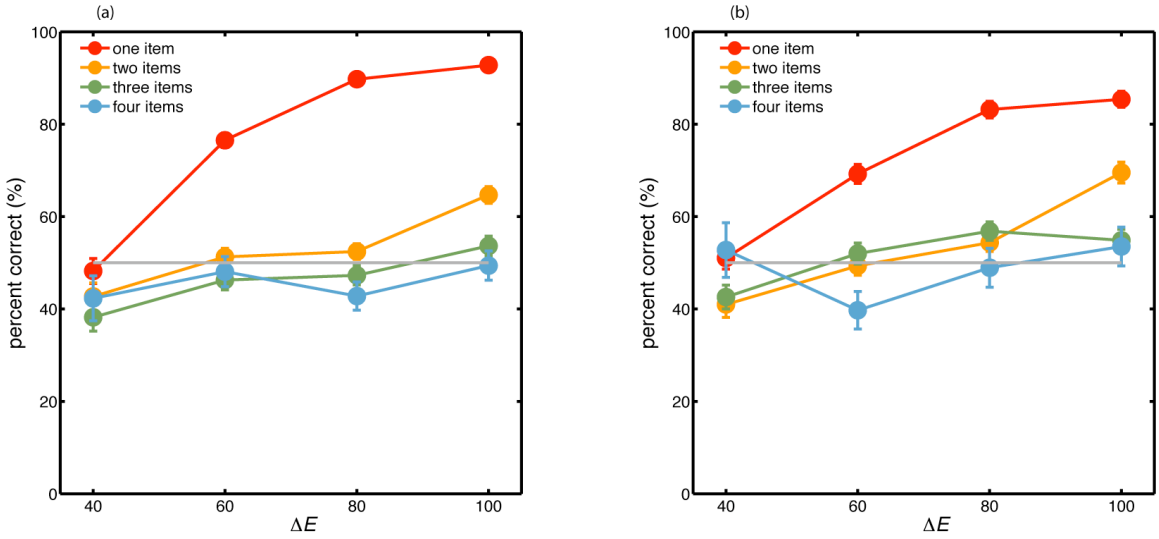


Figure 4.5. Mean performance as a function of Δ*E* separated by set-size. For trials with set-size of one performance is at chance (grey line) only for the lowest Δ*E*; for higher values of Δ*E* performance is significantly above chance. For other set sizes performance is significantly above chance only for *set-size two* and only at Δ*E* = 100. Error bars indicate the standard error of the mean; absence of error bars indicates that error bars are smaller than the marker.

4.3.2 Precision of internal representations

The poor performance and slow reaction times for small Δ*E* might indicate that subjects had difficulty distinguishing the sample and test colors and were reporting that there was no change in color from when in fact there was a change. When we plotted the

subjects' probability of responding 'no change' as function of ΔE for one-item trials (figure 4.6 (a) and (e) subject G and I respectively) we saw that as the sample and test colors became more similar to each other, subjects were more likely to report 'no-change'. By fitting a Gaussian function to these probabilities we obtained an estimate of the precision of subjects' internal color representations (given by the reciprocal of the standard deviation, σ, of the fitted Gaussian).

We repeated this procedure for all set-sizes and we found that as the number of items in working memory increased, the subjects' internal representations lost precision. Figure 4.7 shows the estimated precision as a function of set-size. Both subjects showed a significant drop in precision for trials with more than one item, indicating that as soon as a second item was added to memory, the internal representation of both objects was significantly degraded. Subject G had a further drop in precision for three item trials compared to the two item trials, while adding a fourth item did not seem to decrease precision further. Subject I, however, did not show an effect of adding a third item to the set, but did show a dramatic drop in precision when a fourth item was added into memory.

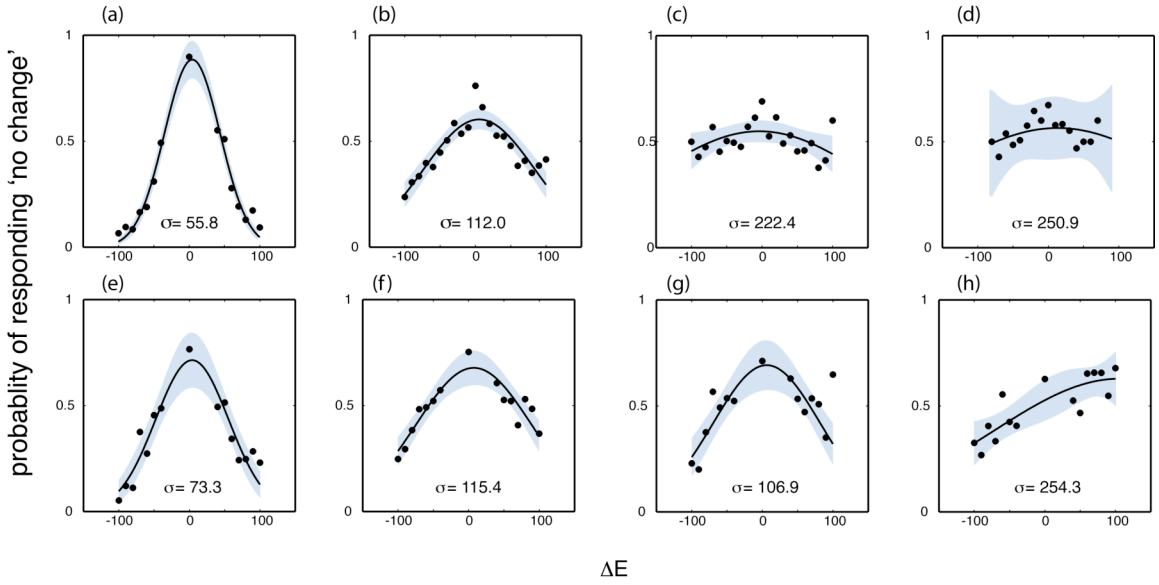


Figure 4.6. Subjects probability of responding 'no-change' as a function of ΔE. Both subjects were more likely to respond 'no change' for small ΔE and as the set-size increased the probability tended towards chance. Solid black lines are the least-squares fitted Gaussian functions. The blue shaded regions indicate the 95% confidence interval of the fit.

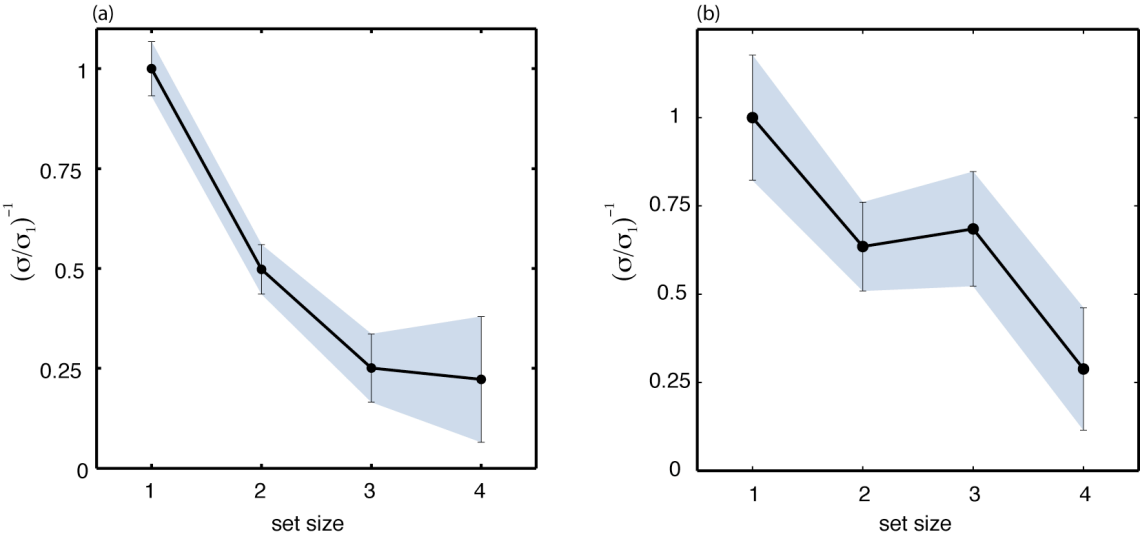


Figure 4.7. Normalized precision estimate as a function of set size. Both subjects showed a dramatic drop in precision in two-item trials compared to one-item trials. Subject G had a further drop in precision when a third item was added with no further drop following the addition of a fourth item. Subject I in contrast showed a drop from three to four items rather than two to three. **(b)**. The shaded area represents the 95% confidence interval for the parameter estimates. Precision estimates are normalized to maximum precision given by σ at set-size=1.

4.4 Discussion

In this chapter we showed that macaque monkeys were able to hold at most two items in visual working memory. Moreover, our data suggests that the effect of adding an extra item to memory is to reduce the overall precision of the memory representations. This result is similar to results from human psychophysical experiments that have shown that visual working memory can be regarded as a limited resource that can be flexibly allocated to encode multiple items in memory. According to this view, there is a limited pool of resources that is shared out between items such that adding more items to memory decreases the precision with which they are stored (Bays and Husain 2008; Bays, Catalao et al. 2009). In their experiments, Bays and colleagues found that precision falls to about 60% of the maximum when a second object is added to memory, and adding a third caused a further decrease in precision to about 40% of the maximum. In fact, they found that the relationship between the number of items and the precision could be described by a simple power law of the form $\sigma \propto N^k$, where N is the number of items in the display. Our results showed a similar pattern. Adding a second item to the display caused the precision to drop to almost half of the maximum, while a third and fourth caused the precision to drop further. We pooled the data from both subjects and found

that our data could indeed be described by a power law relationship between the precision and the number of items in memory, with k=0.96 ± 0.16. Thus, it appears that in macaque monkeys, visual working memory can also be thought of as a limited resource that is shared between items to produce variable-precision memory representations.

An alternative prominent model in the human psychophysical literature argues that visual working memory has a limited number of slots that represent memories with a fixed precision (Zhang and Luck 2008). According to this model, subjects are able to maintain multiple representations with the same precision up to the limit of the number of slots available. In this model, the various behavioral measures (like percent corrects and reaction times) should stay relatively unchanged as a function of set size until the limit is reached. When there are no more slots available, behavioral parameters would show a sudden and sharp decrease. Our results do not seem to agree with this view, as we saw a systematic drop in performance as soon as a second item was added to the display. Subjects performed well above chance for highly discriminable colors at set-size of two but their performance was significantly worse than one-item trials. This suggests that subjects were able to maintain at least two items in memory, and possibly three given that their *overall* performance with highly discriminable colors was for set-size three was still above chance.

Our findings confirm recent reports that macaque monkeys can hold more than one item in visual working memory (Warden and Miller 2007; Siegel, Warden et al. 2009). We have extended these finding to show that monkeys can hold at most two colors in memory. Moreover, we also demonstrated that the precision of stored representations mirrors the findings in humans and we showed that monkey visual working memory can also be thought of as a limited resource that is shared between items in memory. In the next two chapters we will attempt to elucidate the neuronal basis of this limited resource by analyzing the activity patterns of prefrontal cortex neurons while subjects hold multiple colors in working memory.

Chapter 5

Neuronal mechanisms of multi-item working memory

5.1 Introduction

In the previous chapter, we found that internal memory representations lose precision as the memory load increases. In this chapter, we begin to investigate the neuronal basis of this loss of precision. We will focus our efforts on understanding how neurons in the VLPFC encode more than one item in working memory, and how their activity changes when there are two items in working memory compared to a single item.

The neuronal mechanisms or multi-item working memory remain largely unexplored. To date, only two studies have been done where animals were required to maintain more than one item in working memory (Warden and Miller 2007; Siegel, Warden et al. 2009). In their experiments, Miller colleagues showed that when animals are required to remember two images presented sequentially, neurons in the PFC encoded information about both pictures during the delay period. However, the delay activity for the two images did not simply relate in a linear, additive fashion to the activity for single objects. Instead, PFC neurons appeared to response to specific sequences of two objects. One reason for this might be that animals were trained to remember the sequence of the objects as well as the objects themselves, i.e. animals were trained to treat the sequence A-B differently from sequence B-A, even though in both cases the animals were remembering objects A and B. This introduces the potential confound that the complexity of the delay period activity they saw was related to the order in which the objects were presented in addition to the objects themselves.

A further issue is that, to date, no attempt has been made to measure precisely how PFC neuronal tuning changes when more than one object is being held in working memory. To enable us to do this, we used a color change detection task that has been widely used in the human psychophysics literature (Luck and Vogel 1997) and adapted it for non-human primates. We recorded the activity of single neurons in the VLPFC which receives inputs from posterior color areas (Webster, Bachevalier et al. 1994), specifically from an area in temporal cortex that has recently been shown to contain a very high number of color tuned neurons (Conway and Tsao 2009). We found that neurons in the VLPFC showed color tuning during the sample and working memory periods of the task. By parametrically varying the distance between the sample and test colors, we were able to estimate the sharpness of neuronal tuning. When two items were being held in memory,

tuning became less sharp indicating that neurons were encoding less information as more objects were loaded into memory.

5.2 Materials and methods

5.2.1 The task

In this experiment, we used the same color change detection task described in the previous chapter. However, trying to understand the encoding of three or four items in working memory was prohibitively difficult, and so we focused our efforts on first understanding the simplest case of multiple encoding, the encoding of two items in working memory. Thus, we eliminated the three and four item trials from the task. In addition, we modified our stimulus palette by taking a different sample of colors from the L a* b* color-space. For this experiment we used the colors shown in Figure 5.1, the colors from a ring in the $L = 60$ plane centered at $a^* = 17$ and $b^* = 6$. Using this arrangement, colors can be specified by two parameters, the radius from the center, r, and the color angle, θ, measured counter-clockwise from the horizontal. This change in stimuli did not require any additional training of the subjects since the basic structure of the task does not depend on the specific colors we use. Both subjects performed the task just as well (performance >87% for all colors) with the new stimulus colors.

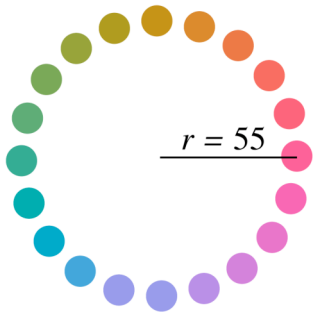


Figure 5.1. Color stimuli used in the neurophysiological experiments. The colors form a ring in the $L = 60$ plane centered in $a^* = 17$ and $b^* = 6$ with radius $r = 55$.

5.2.2 Reconstruction of recording locations

We reconstructed our recording locations by measuring the position of the recording chambers using stereotactic methods. We plotted the positions onto the MRI

sections using commercial graphics software (Adobe Illustrator CS3). We traced and measured the distance of each recording location along the cortical surface from the lip of the ventral bank of the principal sulcus. We measured the anterior-posterior position from the interaural line (x-axis), and the dorsoventral position relative to inter hemispheric fissure (0 point on y-axis). We also measured the positions of the other sulci, relative to the principal sulcus, allowing the construction of the unfolded cortical maps shown in Figure 5.2.

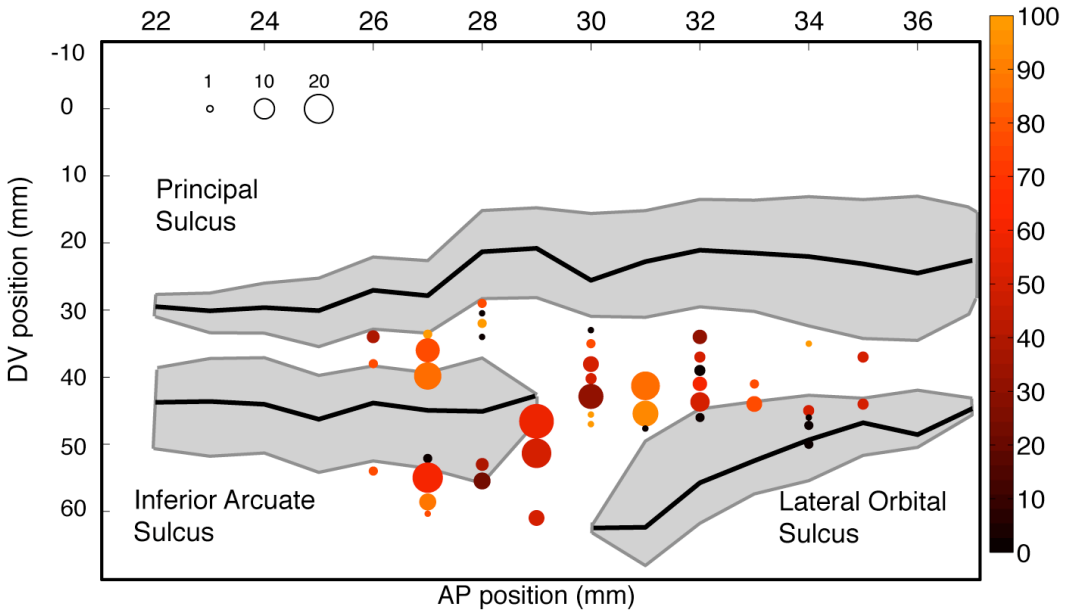


Figure 5.2. Flattened reconstructions of the cortex indicating the locations of recorded neurons in the VLPFC. The size of the circles indicates the number of neurons recorded at that location. The color of the circle indicates the proportion of selective neurons recorded in that location. Gray shading indicates unfolded sulci.

5.2.3 Statistical analysis

We quantified neuronal selectivity during the color change detection task using two defined time epochs. To analyze selectivity relating to presentation of colored stimuli, we defined the *sample* epoch beginning 100-ms after the sample array appears on the screen and lasting 500-ms (we discarded the first 100-ms to account for the latency of neuronal responses in PFC). To analyze selectivity relating to relating encoding colors in working memory, we defined the *delay* epoch beginning 100 ms after the stimulus array disappears from the screen and lasting 1000 ms until the appearance of the second picture on the screen.

In order to assess the effect of attention in the neuronal response, we measured the subject's micro-saccades during the sample period. When the stimulus array appeared on the screen, subjects tended to make micro-saccades in the direction of the colored squares. In the two-item trials, subjects tended to make micro-saccades to one or both of the stimuli. By examining micro-saccades we were able to infer which of the two colors in the array subjects were attending to and loading into working memory. We defined the start of a micro-saccade as the point in time when the eye moved more than three standard deviations away from the mean position of the fixation period (before the stimuli appeared on the screen). We defined a time window starting from 50-ms after the start of the micro-saccade and lasting 250-ms and termed it the *attention* epoch. On some proportion of trials, subjects tended to make a second micro-saccade, however, the proportion of trials that subjects made two micro-saccades was not high enough to permit us to define a second *attention* epoch.

For each neuron and each epoch in turn, we calculated the neuron's trial average spike count firing rate. To analyze neurons' spatial tuning we performed an ANOVA on the firing rate, using the location of the cue on one-item trials, evaluating significance using an alpha level of $p<0.01$. To quantify the neurons' color selectivity we performed the Rayleigh test on the neurons average firing rate for each color stimulus. The Rayleigh test measures the degree to which a distribution of responses around a circle is non-uniform. It is particularly well suited for detecting a unimodal deviation from uniformity (Fisher 1995). A small *p-value* indicates a significant departure from uniformity and indicates that the null hypothesis, that the data are uniformly distributed around the circle, should be rejected. In order to analyze neuronal responses using circular statistical methods we treat the firing rate as if it were binned to the to the color angle, θ, of the presented stimulus with the number of spikes representing the number of samples falling into that bin.

We constructed neuronal tuning curves by we plotting the neuron's average response to each color in polar coordinates (r, θ). The radial distance from the center, r, is given by the neuron's average firing rate, while the angle θ is given by the color-angle of the stimulus (see Figure 5.4). We calculate the mean resultant vector, shown in red, using vector addition. We define the strength of tuning, s, as the length of the mean resultant vector. To measure the sharpness of the neuronal tuning, we calculated the circular variance. This parameter is analogous to the linear variance and it is closely related to the mean resultant vector. The circular variance measures the spread in neurons response to colors around the circle. If the neuron responds only to a few colors in a particular direction of the circle, then the resulting vector will be large and the corresponding variance will be small. If the neuron responds equally to all colors, the resultant vector will be zero and the circular standard deviation will be correspondingly large.

For some neurons we were able to construct a tuning curve for trials where there were two colors present. To do this, for all color combinations, we assigned the two-item response to both colors in the array and divided the final result by the number of times each color appeared in a two-item array. This has the effect of normalizing the responses and producing a well-defined two-item tuning curve, since all colors appeared an equal number of times in combination with all other colors. For example, consider a neuron that shows a firing rate of 10 Hz when colors *A* and *B* are on the screen. Using the procedure outlined above we would assign 10 Hz to both color *A* and color *B*. Now consider all other trials in which color *A* appears with any of the remaining colors, all of those colors will also be assigned a response of 10 Hz, and color A will have been assigned 10 Hz a total of *N* times, where N is the number of times *A* appeared with another color. Now we repeat the same procedure with color *B*. If we assume that the neuron is perfectly tuned for color *A* in the two-item conditions (i.e. it only responds when *A* is on the screen, and it does not respond to any other color) then the responses to all other color combinations will be 0. When we divide all the responses by *N* all colors will have a final response of 10 Hz/*N*, with the exception of color *A*, which will have a final response of 10Hz. Thus, the tuning curve will correctly reflect the neuron's tuning to color *A*.

To characterize the nature of the encoding of two items by PFC neurons, we performed a multiple linear regression analysis of the form: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ on the neuron's response to the two colors in the array (y) using the neuron's response to the constituent colors when presented in isolation (x_1 and x_2) as regressors. All analyses were done in MATLAB (R2008b). All circular statistical analyses were done using the *CircStat* toolbox for MATLAB (Berens 2009).

5.3 Results

5.3.1 Location selectivity

We recorded from 263 neurons in the VLPFC from one subject. In order to characterize the neuron's receptive field (RF), we first analyzed trials in which only one color was presented. Stimuli were presented in four fixed spatial locations, and while the subjects were not required to remember the location of stimuli, we found that 152/263 neurons showed a significant effect of sample location during the sample epoch and 153/263 neurons during the *delay* period (ANOVA, $p<0.01$). Neurons typically had stronger responses to stimuli presented in one visual hemi-field compared to the opposite hemi-field. The neuron shown in Figure 5.3 had a significantly stronger response to colors presented in the left hemi-field compared to the right (ANOVA, $F_{3,385}=105.4$, $p<1\text{x}10^{-16}$). The responses to items within one hemi-field were not significantly different from each other (*post-hoc* Tukey-Kramer test $p>0.05$). Thus, we defined a neuron's RF as the hemi-field that elicited the strongest response.

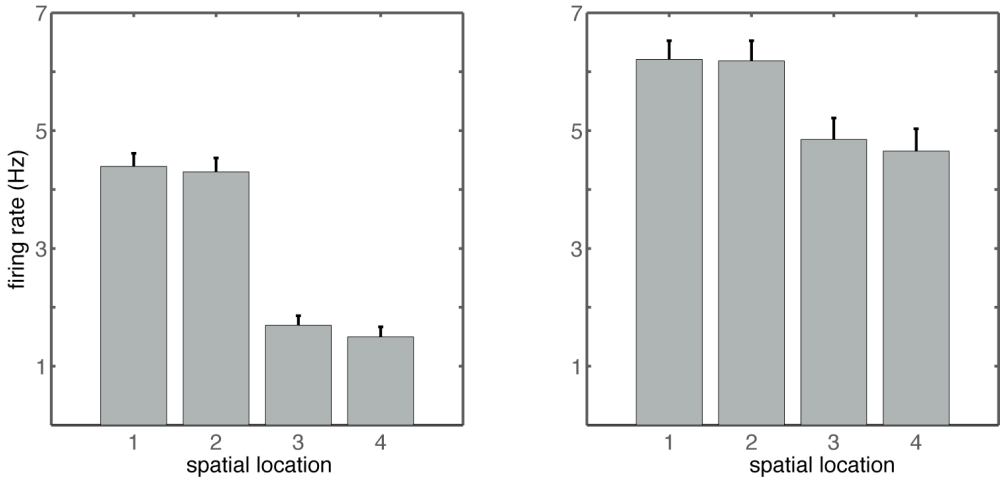


Figure 5.3. Example of the average response of a single neuron to items presented at each spatial location during the *sample* epoch (a) and during the *delay* epoch (b).

5.3.2 Color tuning in the VLPFC

Neurons in the VLPFC showed selective tuning for color during both the *sample* and the *delay* epochs. To determine if neurons showed color tuning, we performed the Rayleigh test on their response to the colors for one-item trials (see methods). In order to obtain a more accurate characterization of neuronal tuning, we restricted our analysis to stimuli that appeared in the neurons RF. Figure 5.4 shows the activity of two typical color selective neurons during the s*ample* epoch.

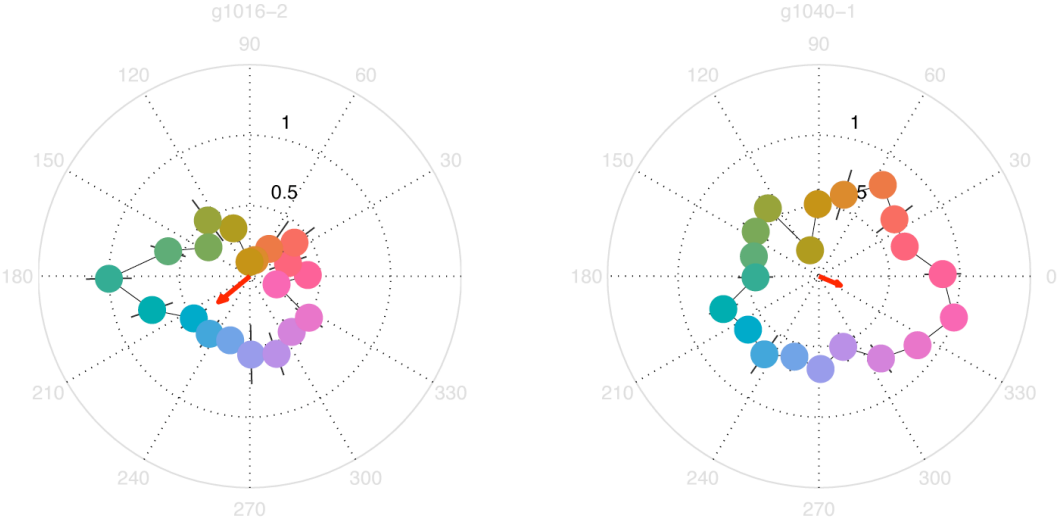


Figure 5.4. Color tuning curve for two neurons recorded from the VLFC computed using activity during the *sample* epoch. The distance from the center corresponds to the neuron's firing rate normalized by the maximum firing rate. The color of the marker represents the color on the screen. The

red arrow indicates the resultant vector and it is proportional to the strength of tuning.

The neuron in the left panel showed a significantly higher response to colors in the lower left quadrant of the circle compared to the other colors (*Rayleigh* test, $p<0.002$). The neuron in the right panel was tuned to colors in the lower right quadrant (*Rayleigh* test, $p<0.005$). To quantify the strength of tuning, s, we examined the length of the mean resultant vector. This quantity ranges from 0 (no color tuning) to 1 (maximal color tuning: the neuron responds exclusively to one color). For the neuron in the left panel of Figure 5.4, $s = 0.3$, while for the neuron in the right panel $s = 0.17$. To quantify the sharpness of neuronal tuning we calculated the *circular variance*, which measures the spread in the neuron's color response. The neuron in the left panel showed an increased firing rate to more colors in the circle panel and thus was more broadly tuned compared to the neuron in the left (*circular variance* = 0.84 vs. 0.7).

Of the 263 neurons we recorded, 121 (46%) showed significant color tuning ($p<0.01$) during the *sample* epoch. In order to determine if there was a section of the color circle that was represented significantly more than others, we performed the Rayleigh test on the distribution of mean response direction for all color-tuned neurons. On the population level, we found neurons tuned for all directions around the color circle with no particular direction being over represented (*Rayleigh* test, $p>0.15$). For all color-tuned neurons, the strength of tuning ranged from weakly tuned neurons ($s = 0.02$) to more strongly tuned neurons ($s = 0.4$) (Figure 5.5a).

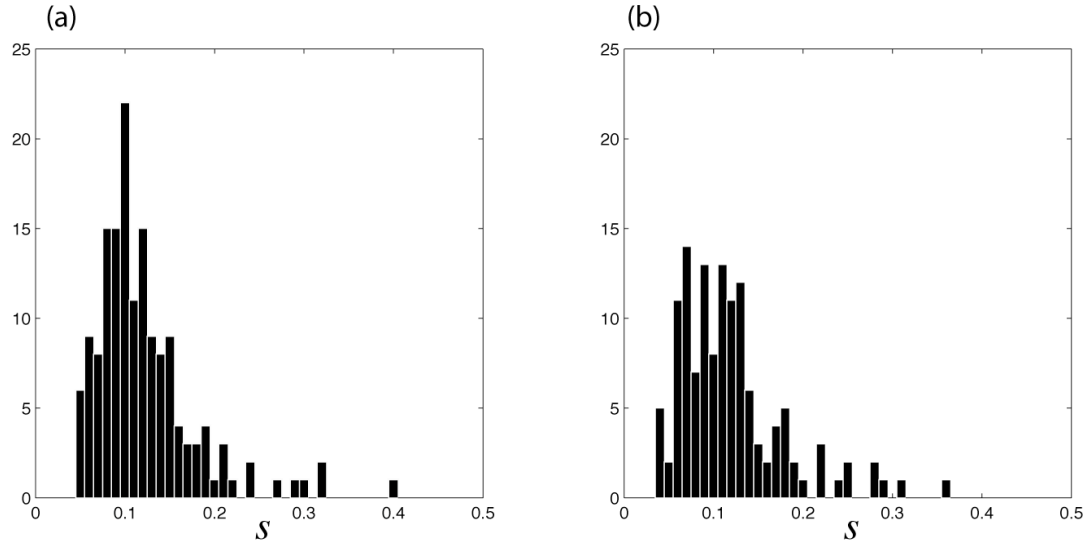


Figure 5.5 Distribution of tuning strength for all color-tuned neurons in the *sample* epoch (a) and in the *delay* epoch (b)

5.3.3 Color working memory for one item

During the *delay* epoch, 52% of neurons (138/263) showed significant color tuning in the one-item condition. Figure 5.6 shows an example of two color working memory neurons. The neuron on the left panel showed significant color-working memory tuning for colors in the upper right quadrant of the circle (*Rayleigh* test, $p<0.001$), while the neuron on the right panel had significant tuning for colors on the lower right quadrant (*Rayleigh* test, $p<0.01$).

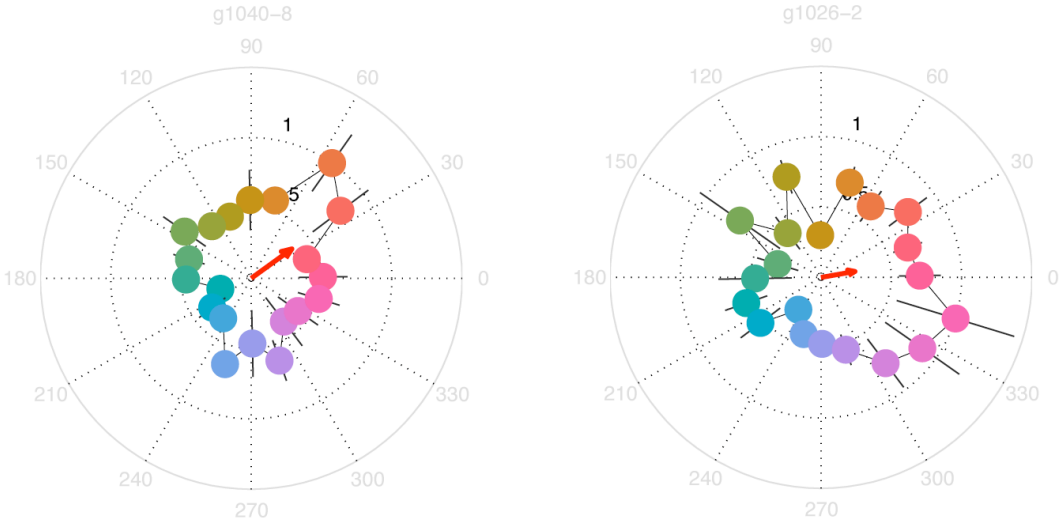


Figure 5.6. Color tuning curve for two neurons during the *delay* epoch. The distance from the center corresponds to the neuron's firing rate normalized by the maximum firing rate. The color of the marker represents the color on the screen. The red arrow indicates the resultant vector and it is proportional to the strength of tuning.

During the *delay* epoch, we found neurons tuned to all directions of the color circle with no significant direction being overrepresented (*Rayleigh* test, $p>0.5$). We found that color tuned neurons in the *delay* epoch had a similar range of tuning strength compared to the *sample* epoch (Figure 5.5b). The strength of tuning was not significantly different between the *sample* and *delay* epochs (Mann-Whitney test, $p>0.6$).

5.3.4 Color working memory for two items

Our first analysis examined whether we could predict the firing rate of the neuron to two items, based on which item fell in the neuron's RF and which item was outside of the neuron's RF. We performed a multiple linear regression of the form

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, on the neuron's response to the two-color arrays (y). The predictor variables (x_1 and x_2) consisted of the neuron's response to the color that was either inside (x_1) or outside (x_2) the neuron's RF as determined on the one-item trials. The model did a poor job in predicting neuronal firing rates on two-item trials. Only 4 % of neurons showed a significant fit, which was not significantly different from the proportion expected by chance.

A second possibility is that the neuron's firing rate is driven by whichever color it prefers irrespective of where that color falls with respect to the RF. To test this, we rank-ordered the neuron's response to individual colors on the one-item trials. For each two-item trial, we then determined the neuron's rank preference of the colors in the display. The predictor variable x_1 was the neuron's response to the neuron's more preferred color on the one-item trials. The predictor variable x_2 was the neuron's response to the neuron's less preferred color on the one-item trials. A significant effect in this regression would be consistent with a coding scheme where the neurons two-item response is a linear combination of the constituent items depending on the neuron's color preference. If the neuron's response on two-item trials was driven solely by its preferred color we would expect $\beta_1 = 1$ and $\beta_2 = 0$ in the regression equation. Unfortunately, this model also did a poor job of predicting neuronal firing rates on two-item trials. Only 4 % of neurons showed a significant fit, which was again not significantly different from the proportion expected by chance.

A third analysis examined whether the two-item response was driven by the stimulus that was in the subject's focus of attention. In order to determine if this was the case, we used the subject's micro-saccades as a proxy for the focus of covert attention. We monitored our subject's eye position throughout the entire period of time that he was fixating the center of the screen. When the array of colors appeared on the screen, he tended to make micro-saccades in the direction of one or both of the stimuli. We designated the color that was in the direction of the micro-saccade as being in the focus of attention. On the majority of two-item trials, he made only one micro-saccade to one stimulus location and not the other. This prevented us from confidently designating the second color in the array as being the focus of attention. We therefore performed a regression analysis on the neuron's two-item responses using the neuron's one-item response to the first attended color as the only regressor. We found that for 17% of color-tuned neurons (21/121) in the *sample* epoch, the response to the two-item trials was linearly related to the response to the attended color. Figure 5.7 shows one such neuron's response to the two-item trials plotted as a function of the response to the attended color. For this neuron, there was a clear linear relationship between the two-item responses and the attended color response ($R^2 = 0.46$, $p<0.001$).

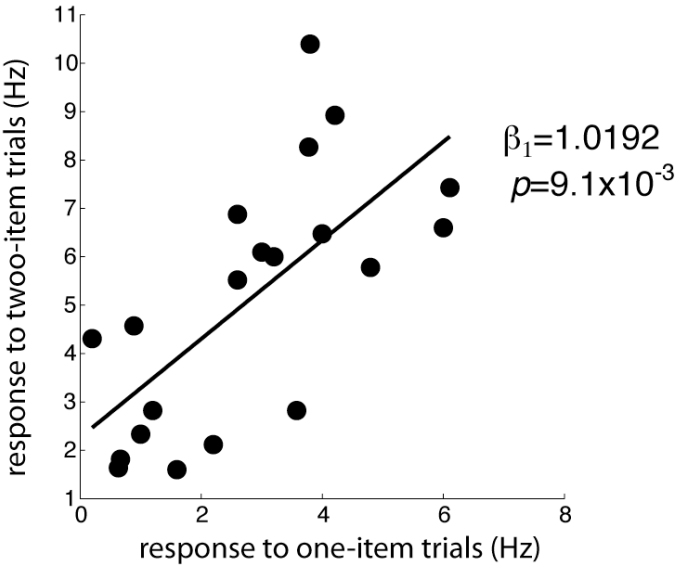


Figure 5.7 Single unit response to the two-item trials as a function of the one-item responses of the attended color. Solid line indicated linear regression fit.

For neurons that show this linear relationship, we can construct a two-item tuning curve by assigning the two-item response to both constituent colors and normalizing by the number of times the color appeared in the two-item arrays (see methods). This method of constructing a two-item tuning curve assumes a linear relationship between the two-item trials responses and the one-item responses, and thus is only applicable to the 17% of neurons in our sample that showed such a relationship.

Figure 5.8b shows the one and two-item tuning curve for one of these neurons. The figure shows that color tuning for the two-item conditions was significantly weaker and less sharp compared to the one-item conditions shown in Figure 5.8a. Tuning strength for the one-item condition: $s_{one\text{-}item}$=0.37, while for the two-item condition it was significantly smaller $s_{two\text{-}item}$=0.06 (parametric bootstrap, $p<0.01$) . This was the case for all neurons for which we could obtain a two-item tuning curve and it indicates that when the subject is maintaining two colors in working memory, color tuning becomes smaller and it less sharp. In fact, for the neuron in this example, the tuning in the two-item trials completely disappears (Rayleigh test, $p>0.2$).

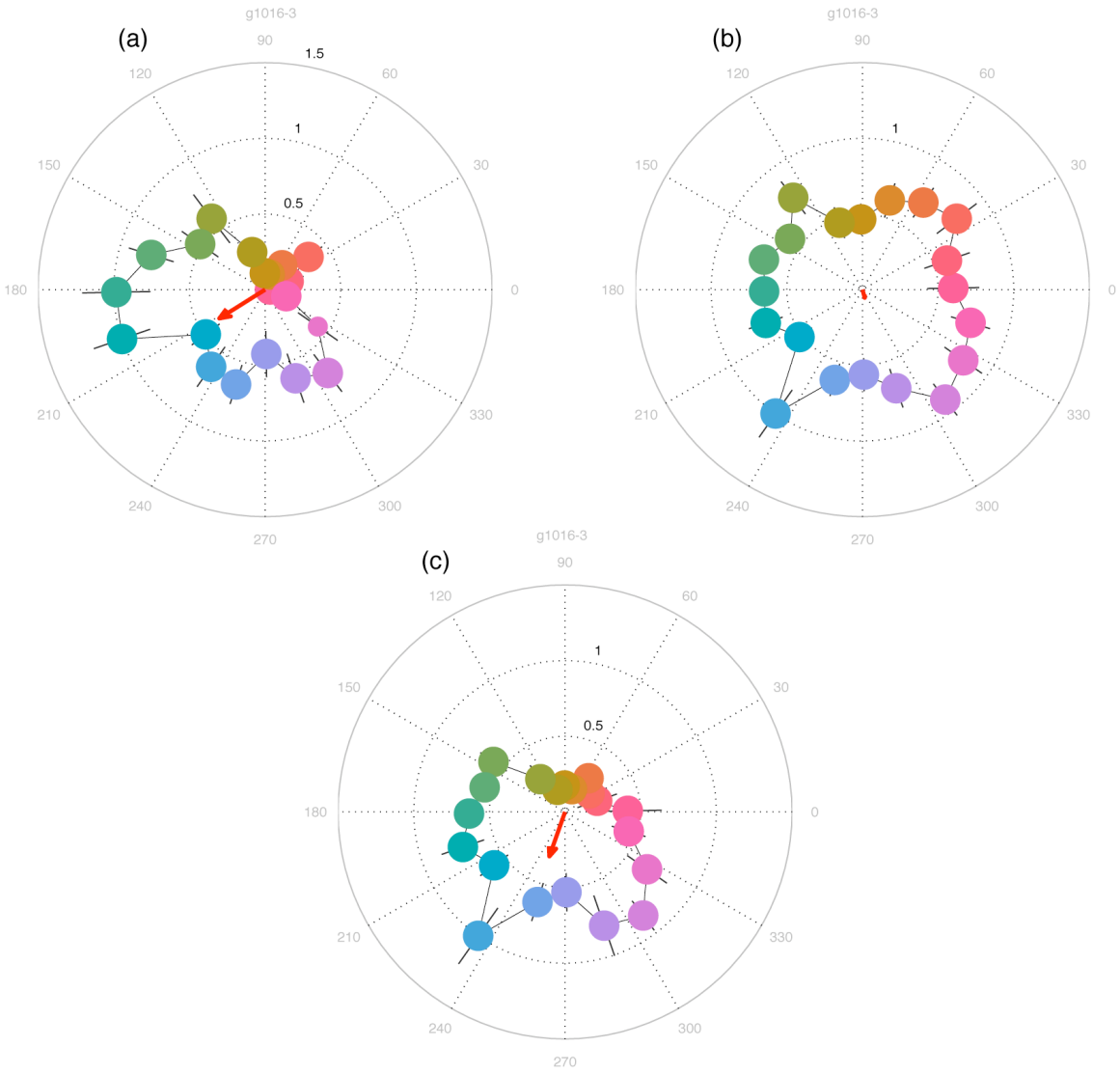


Figure 5.8 Tuning curves for a neuron with a linear relationship between the attended color and the two-item response. (a) Tuning curve computed using the one-item trials, the neuron was strongly tuned for colors in the lower left quadrant. (b) Two-item tuning curve computed using all the colors in the two item arrays. The neuron did not have a significant tuning. (c) Modified two-item tuning curve computed using the attended colors. This tuning curve shows that the neuron is strongly tuned for colors in the lower left quadrant when attention is taken into account.

This method of constructing a two-item tuning curve does not take in to account which color was being attended. Thus, we examined the effect of attention on color

tuning in the two-item conditions. We did this by computing a modified two-item tuning curve by assigning the two-item response to the attended color. By doing this, we were able to recover the color tuning seen in the one-item trials (Figure 5.8c). Although the one-item and modified two-item tuning curves are not identical, when attention is taken into account, the strength of tuning for the two-item trials ($s_{two\text{-}item\text{-}attention}$= 0.31) was not significantly different than the one-item trials ($s_{one\text{-}item}$ = 0.36; Parametric bootstrap test, p>0.05). This was the case for 80% of the neurons that showed a linear relationship between the attended color and the two-item response.

5.4 Discussion

In this set of experiments, we found that neurons in the VLPFC showed significant color tuning during both the *sample* phase and the *delay* phase of a color change detection task. Additionally, we found that when two-colors had to be maintained in working memory, for a subset of neurons the response to the two-color array was only well described as a linear function of the constituent colors when that color was in the focus of attention.

Understanding the neuronal coding of more than a single stimulus, at its core is a difficult theoretical problem. In the visual system, this has been addressed in area V4 (Gawne and Martin 2002) and in IT cortex (Zoccolan, Cox et al. 2005). Gawne and Martin found that when two stimuli are presented inside the RF of a V4 neuron, the response could be described as the maximum firing rate of each stimulus presented separately (MAX model). Our results show that this model cannot be applied to neurons in the VLPFC. If this were the case we would expect to see a β_1 = 1 and β_2 = 0 in the rank-preference regression.

Zoccolan *et al* found that for neurons in the IT cortex, responses to pair of objects were best explained by the average of the responses to the constituent objects. This corresponds to $\beta_1 = \beta_2 = 0.5$ in the regression equation. This model also failed to describe our data since we did not see a significant number of neurons whose responses could be modeled by this regression model.

Only when we take into account where subjects are covertly attending, can we describe the responses using a linear model. One reason for this might be that in our experiments subjects were not just passively viewing the stimuli, but were actively using information about the stimuli in an ongoing task. Crucial to solving the task, was the ability to encode information about the two colors in working memory. From the pattern of their micro-saccades we were able to determine that subjects were adopting a strategy of sequentially attending to the colors in the display. Thus, the neuronal responses reflected the items in the display only when they were in the focus of attention.

For those neurons for which we were able to compute a well-defined two-item tuning curve, we found that sharpness of tuning decreased when subjects were required to maintain two colors in working memory. This decrease in sharpness mirrors the decrease in precision of internal memory representations seen at the behavioral level. It is therefore possible that the underlying reason for the loss of precision in the internal memory representations as the memory load increases is due to the decrease in the neurons' ability to encode more information (as evidenced in the widening of tuning curves). This loss of precision, however, was not seen when we took into account the color subjects were attending to, similar to human psychophysics findings. In their experiments, Bays and Husain (2008) found that when covert attention was directed towards a particular item, the proportion of resources allocated to representing that item was increased, which lead to the item being retained with greater precision than other objects on the scene.

Our results also agree with one prominent model of attention which states that the effect of attention of neuronal coding is to increase neuronal sensitivity to the feature being attended, which is akin to a sharpening of a neuron's tuning curve (Desimone and Duncan 1995; Reynolds and Chelazzi 2004). Our results did not show a sharpening of tuning, *per se,* however we observed wider tuning curves when attention was not taken into account. Unfortunately, our current method of analyzing the pattern of microsaccades as a proxy for attention does not allow us to determine precisely which color is being attended on every single trial. In order to fully determine the effect of covert attention on the encoding of multiple items in working memory would need to have a more direct measure of attention, by looking at responses from the frontal eye fields, for example (Buschman and Miller 2009) or by cueing subjects to covertly attend to a certain color during the task.

Chapter 6

Conclusions and future work

In this dissertation we have described our efforts to provide answers to two long-standing questions in cognitive neuroscience: the organization of the working memory within the prefrontal cortex and the mechanisms underlying the limited capacity of working memory. In this chapter we will summarize our findings and turn to remaining questions and future work that might be done to address them.

6.1 Gustatory working memory

Our results demonstrated that OFC neurons could show greater encoding of information in working memory than LPFC neurons when the information is from an informational domain for which OFC is specialized. This supports the content model regarding the functional organization of PFC, but it is unlikely to be the whole story. For example, it has previously been found that OFC neurons can maintain abstract information in working memory (Wallis, Anderson et al. 2001) and PFC neurons show the capacity to alter the nature of the information they are maintaining in working memory depending on the demands of the task (Rao, Rainer et al. 1997). Further, our results do not necessarily contradict the process model, since DLPFC may still be an area specialized for the monitoring and manipulation of the contents of working memory. Indeed, a hybrid version of the two models may be the most accurate, with DLPFC manipulating and monitoring information, while VLPFC and OFC share the load of maintaining information dependent on informational content. Olfactory, gustatory, emotional, social and reward information may be maintained in working memory by OFC neurons, while somatosensory, auditory and visual information may be maintained by VLPFC neurons.

6.2 Working memory capacity limit

The work presented in Chapters 4 and 5 described our efforts to understand the nature of the limited capacity of working memory. We started by showing that a popular behavioral paradigm, the multiple-item change detection task, widely used in the human psychophysics literature could be adapted to primates. Using this task we showed that macaque monkeys can hold more than a single item of information in working memory. Moreover, our results showed that as the number of items in working memory increases, the internal memory representations loose precision. This fits well with a model of working memory in which there is a limited amount of resources that must be shared out

to represent all items in memory. Although the controversy over the 'resources' or 'slots' models is likely to continue, our results lend support to the resources model by showing that the resources model better predicts the behavioral data from non-human primates.

Our experiments required that subjects maintain information in working memory for a period of one second. It has been suggested that information activated in working memory is subject to decay unless it is reactivated by additional stimulus presentations or through processes (Baddeley 1986; Cowan 1995). It would be interesting to see if increasing the length of time that items have to be actively maintained has an effect on the fidelity of the memory representations. If indeed memories systematically loose precision the longer they are maintained, by looking at the one-item case, it might be possible to determine the minimum amount of precision required to successfully represent information in working memory.

In Chapter 5 we saw that neurons in the VLPFC are involved in maintaining colors in working memory. We found that over 50% of neurons encoded color information in working memory when in trials where one item had to be remembered. Additionally, we found that for a large number of these neurons, we could not predict their responses during the two-item trials based on the one-item responses. However, a subset of these color-tuned neurons did have a well defined tuning curve.in the two-item. We found that there was a decrease in sharpness when two items were being held in working memory. This reduction in fidelity mirrored what we saw in the behavior of the subjects. As the number of objects increases, neuronal tuning becomes broader thus encoding less information. This is analogous loss in precision of the internal memory representations.

However, this may not be the complete story. Although we did not require that subjects preferentially attend to a color, nonetheless we found that on some trials subjects were covertly attending to one color and not the other. When we incorporated this into our analysis we found that neurons were just as sharply color-tuned when compared to trials in which there was only one item in memory. In other words, it was as if the neuron only saw one color in the display. Our measure of attention was indirect and therefore our results should be interpreted with caution. It would be interesting to have a more direct measure of attention by recording from FEF neurons and decoding where attention is being focused, as was done in a study recently by Buschman and Miller (2009) where they were able to decode the focus of attention even in the absence of eye movements. By using this method to decode what color subjects were attending to, we could increase the power of our analysis since we would not have to limit our analysis to the subset of trials in which the subjects makes microsaccades towards an item in the array.

Additionally, it has been shown that attention increases neuronal sensitivity to objects that are behaviorally relevant (Desimone and Duncan 1995; Treue and Maunsell 1996; Desimone 1998; Reynolds, Pasternak et al. 2000; Hayden and Gallant 2005; Saalmann, Pigarev et al. 2007). In our task, both objects were behaviorally relevant, since

both colors could change with equal probability. It would be interesting to conduct an experiment where one of the items had a higher probability of changing and subjects were cued to direct attention to that item. We might expect to see that neurons in the PFC also show increase in neuronal sensitivity for the colors that are in the focus of attention analogous to what has been observed in lower visual areas.

Another mechanism that might account for the loss in memory precision could be operating at the population level. It has been shown that for a large population of neurons, correlations in the noise of neuronal responses can cause the information encoded by the population to saturate (Zohary, Shadle et al. 1994; Abbot and Dayan 1999). These correlations can place a fundamental constraint on the precision with which the brain can represent information (Averbeck, Latham et al. 2006). In order to test this model we would need to record simultaneously the activity of hundreds of neurons. In our experiments our recordings were limited to a maximum of 16 simultaneous electrodes per session. With an average yield of 1.5 neurons per electrode we could at best record from around 24 single neurons per session. This number of neurons per session, although not small by conventional standards, does not lend it self for population coding analysis. One possible solution would be to construct pseudo-populations as in Meyers *et al.* (Meyers, Freedman et al. 2008). Although this would affect the estimates of the absolute amount of information that could be decoded (thus limiting our ability to estimate the absolute limit of information that can be held in working memory), we could still use this method to obtain relative estimates of information being encoded in one-item compared to two-item conditions. Alternately, we could implant a number of 64 electrode arrays across the PFC, which could potentially increase more than ten-fold the number of neurons recorded per session. Even though our laboratory has no experience with such arrays, we have close collaborations with the laboratories of Dr. Jose Carmena and Dr. Robert Knight which do have extensive experience with this technique.

Finally, it would be interesting to examine if working memory activity is related to the local field potential (LFP) recorded while subjects were maintaining two items in working memory. Lisman and Idiart (1995) proposed a mechanism by which different memory traces are represented by neuronal firing that is synchronized to different phases of the local field potential. Recently Siegel *et al.* (2009) showed that neurons involved in representing two sequentially presented items in working memory, fire at different phases of the LFP. These accounts assume that items are loaded sequentially into working memory, and although in our task stimuli were presented simultaneously, we saw that on some trials subjects made sequential eye movements towards each stimulus location. Thus it would be interesting to see if there is evidence of this kind of frequency-locked pattern of activity in our data.

6.3 Conclusion

The nature of how the brain maintains active representations of things it can no longer sense for very short period of time has been studied for more than a century. And while

great advances have been made throughout the years, with countless doctoral dissertations being written along the way, this field is still full of interesting and important questions still to be answered. It has certainly proved interesting and puzzling enough to keep a young fellow originally trained in physics occupied for over six years. And it does not end here; hopefully the work in this dissertation can provide a spark, however small, to inspire many more young curious minds.

Literature Cited

Abbot, L. and P. Dayan (1999). "The Effect of Correlated Variability on the Accuracy of a Population Code." Neural Computation **11**(1): 91-101.

Alvarez, G. and P. Cavanagh (2004). "The capacity of visual short-term memory is set both by visual information load and by number of objects." Psychological Science **15**(2): 106-111.

Averbeck, B. B., P. E. Latham, et al. (2006). "Neural correlations, population coding and computation." Nature Reviews Neuroscience **7**(5): 358-366.

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

Baddeley, A. D. (1992). "Working Memory." Science **255**(5044): 556-559.

Barbas, H. (1993). "Organization of cortical afferent input to orbitofrontal areas in the rhesus monkey." Neuroscience **56**(4): 841-864.

Barbas, H. and M. M. Mesulam (1985). "Cortical afferent input to the principalis region of the rhesus monkey." Neuroscience **15**(3): 619-637.

Barbas, H. and D. Pandya (1991). Patterns of connections of the prefrontal cortex in the rhesus monkey associated with cortical architecture. Frontal Lobe Function and Dysfunction. H. S. Levin, H. M. Eisenberg and A. L. Benton. New York, Oxford Univeristy Press**:** 35 - 58.

Barbas, H. and D. N. Pandya (1989). "Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey." J Comp Neurol **286**(3): 353-375.

Bays, P. M., F. Catalao, et al. (2009). "The precision of visual working memory is set by allocation of a shared resource." Journal of Vision **9**(10): 1-11.

Bays, P. M. and M. Husain (2008). "Dynamic Shifts of Limited Working Memory Resources in Human Vision." Science **321**(5890): 851-854.

Berens, P. (2009). "CircStat: a MATLAB toolbox for circular statistics." Journal of Statistical Software **31**(10).

Bodner, M., J. Kroger, et al. (1996). "Auditory memory cells in dorsolateral prefrontal cortex." Neuroreport **7**: 1905-1908.

Buschman, T. J. and E. K. Miller (2009). "Serial, Covert Shifts of Attention during Visual Search Are Reflected by the Frontal Eye Fields and Correlated with Population Oscillations." Neuron **63**(3): 386-396.

Carmichael, S. T. and J. L. Price (1994). "Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey." Journal of Comparative Neurology **346**(3): 366-402.

Carmichael, S. T. and J. L. Price (1995). "Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys." Journal of Comparative Neurology **363**(4): 615-641.

Carmichael, S. T. and J. L. Price (1995). "Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys." Journal of Comparative Neurology **363**: 642 - 664.

Carmichael, S. T. and J. L. Price (1996). "Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys." J Comp Neurol **371**(2): 179-207.

Cavada, C., T. Company, et al. (2000). "The anatomical connections of the macaque monkey orbitofrontal cortex. A review." Cereb Cortex **10**(3): 220-242.

Cavada, C. and P. S. Goldman-Rakic (1989). "Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe." J Comp Neurol **287**(4): 422-445.

Conway, A. R., M. J. Kane, et al. (2003). "Working memory capacity and its relation to general intelligence." Trends in Cognitive Sciences **7**(12): 547-552.

Conway, B. R. and D. Y. Tsao (2009). "Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex." Proceedings of the National Academy of Sciences **106**(42): 18034-18039.

Cowan, N. (1995). Attention and memory: An integrated framework., Oxford University Press.

Cowan, N. (2001). "The magical number 4 in short-term memory: A reconsideration of mental storage capacity." Behavioral and Brain Sciences **24**: 87-185.

Cowan, N., E. M. Elliott, et al. (2005). "On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes." Cognitive Psychology **51**(1): 42-100.

Daneman, M. and P. Carpenter (1980). "Individual differences in working memory and reading." Journal of Verbal Learning and Verbal Behavior **19**: 450-466.

Desimone, R. (1998). "Visual attention mediated by biased competition in extrastriate visual cortex." Philosophical Transactions of the Royal Society B **353**: 1245-1255.

Desimone, R. and J. Duncan (1995). "Neural mechanisms of selective visual attention." Annual Reviews of Neuroscience **18**: 193-222.

Duncan, J. (2001). "An adaptive coding model of neural function in prefrontal cortex." Nat Rev Neurosci **2**(11): 820-829.

Everling, S., C. J. Tinsley, et al. (2002). "Filtering of neural signals by focused attention in the monkey prefrontal cortex." Nat Neurosci **5**(7): 671-676.

Fellows, L. K. and M. J. Farah (2005). "Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans." Cereb Cortex **15**(1): 58-63.

Fellows, L. K. and M. J. Farah (2007). "The role of ventromedial prefrontal cortex in decision making: Judgment under uncertainty or judgment per se?" Cereb Cortex **17**(11): 2669-2674.

Fisher, N. I. (1995). Statistical Analysis of Circular Data. Cambridge, Cambridge University Press.

Freedman, D. and J. Assad (2009). "Distinct Encoding of Spatial and Nonspatial Visual Information in Parietal Cortex." Journal of Neuroscience **29**(17): 5671-5680.

Freedman, D., M. Riesenhuber, et al. (2001). "Categorical representation of visual stimuli in the primate prefrontal cortex." Science **291**: 312-316.
Funahashi, S., C. J. Bruce, et al. (1989). "Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex." J Neurophysiol **61**(2): 331-349.
Funahashi, S., M. Chafee, et al. (1993). "Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task." Nature **365**: 753-756.
Fuster, J. M. (1973). "Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory." J Neurophysiol **36**(1): 61-78.
Fuster, J. M. (2008). The prefrontal cortex. San Diego, Academic Press.
Fuster, J. M. and G. E. Alexander (1971). "Neuron activity related to short-term memory." Science **173**(997): 652-654.
Fuster, J. M., R. H. Bauer, et al. (1982). "Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks." Exp Neurol **77**(3): 679-694.
Gawne, T. and J. Martin (2002). "Responses of Primate Visual Cortical V4 Neurons to Simultaneously Presented Stimuli." Journal of Neurophysiolgy **88**: 1128-1135.
Goldman-Rakic, P. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. Handbook of physiology. V. Mountcastle, F. Plum and S. Geiger. Bethesda, MD, American Physiology Society. **5**.
Goldman-Rakic, P. S. (1996). "The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive." Philosophical Transactions of the Royal Society B **351**(1346): 1445-1453.
Goodale, M. A. and A. D. Milner (1992). "Separate visual pathways for perception and action." Trends in Neuroscience **15**(1): 20-25.
Grabenhorst, F. and E. T. Rolls (2008). "Selective attention to affective value alters how the brain processes taste stimuli." Eur J Neurosci **27**(3): 723-729.
Hayden, B. Y. and J. L. Gallant (2005). "Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4." Neuron **47**(5): 637-643.
Henderson, L. (1972). "Spatial and verbal codes and the capacity of STM." The Quarterly Journal of Experimental Psychology **24**(4): 485-495.
Hikosaka, K. and M. Watanabe (2000). "Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards." Cerebral Cortex **10**(3): 263-271.
Jacobsen, C. F. (1935). "Functions of the frontal association area in primates." Archives of Neurology and Psychiatry **33**: 558 - 569.
Kennerley, S. W., A. F. Dahmubed, et al. (2009). "Neurons in the frontal lobe encode the value of multiple decision variables." Journal of Cognitive Neuroscience **21**(6): 1162-1178.
Kubota, K. and H. Niki (1971). "Prefrontal cortical unit activity and delayed alternation performance in monkeys." J Neurophysiol **34**(3): 337-347.
Lara, A. H., S. W. Kennerley, et al. (2009). "Encoding of gustatory working memory by orbitofrontal neurons." Journal of Neuroscience **29**(3): 765-774.

Levy, R. and P. S. Goldman-Rakic (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale. **133:** 23-32.

Lisman, J. E. and M. A. Idiart (1995). "Storage of 7 +/- 2 short-term memories in oscillatory subcycles." Science **267**(5203): 1512-1515.

LoPresti, M. L., K. Schon, et al. (2008). "Working memory for social cues recruits orbitofrontal cortex and amygdala: a functional magnetic resonance imaging study of delayed matching to sample for emotional expressions." J Neurosci **28**(14): 3718-3728.

Luck, S. J. and E. K. Vogel (1997). "The capacity of visual working memory for features and conjunctions." Nature **390**: 279-281.

Meyer, D. E. and D. E. Kieras (1997). "A computational theory of executive cognitive processes and multiple-task performance: I. Basic mechanisms." Psychological Review **104**(1): 3-65.

Meyers, E., D. Freedman, et al. (2008). "Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex." Journal of Neurophysiolgy **100**(3): 1407.

Miller, E., C. Erickson, et al. (1996). "Neural Mechanisms of Visual Working Memory in Prefrontal Cortex of the Macaque." Journal of Neuroscience **16**(16): 5154.

Miller, E. K. and J. D. Cohen (2001). "An integrative theory of prefrontal cortex function." Annual Review of Neuroscience **24**: 167-202.

Miller, E. K. and J. D. Cohen (2001). "An integrative theory of prefrontal cortex function." Annu Rev Neurosci **24**: 167-202.

Miller, E. K. and R. Desimone (1994). "Parallel neuronal mechanisms for short-term memory." Science **263**(5146): 520-522.

Miller, E. K., C. A. Erickson, et al. (1996). "Neural mechanisms of visual working memory in prefrontal cortex of the macaque." J Neurosci **16**(16): 5154-5167.

Miller, G. A. (1956). "The magical number seven, plus or minus two: some limits on our capacity for processing information." The Psychological Review **63**: 81 - 97.

Mishkin, M. and F. J. Manning (1978). "Non-spatial memory after selective prefrontal lesions in monkeys." Brain Res **143**(2): 313-323.

Morecraft, R. J., C. Geula, et al. (1992). "Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey." Journal of Comparative Neurology **323**(3): 341-358.

Niki, H. (1974). "Differential activity of prefrontal units during right and left delayed response trials." Brain Research **70**(2): 346-349.

Niki, H. (1974). "Prefrontal unit activity during delayed alternation in the monkey. I. Relation to direction of response." Brain Research **68**(2): 185-196.

Niki, H. and M. Watanabe (1979). "Prefrontal and cingulate unit activity during timing behavior in the monkey." Brain Res **171**(2): 213-224.

Owen, A., A. Evans, et al. (1996). "Evidence for a Two-Stage Model of Spatial Working Memory Processing within the Lateral Frontal Cortex: A Postitron Emission Tomography Study." Cerebral Cortex **6**: 31-38.

Owen, A., C. Stern, et al. (1998). "Functional organization of spatial and nonspatial working memory processing within the human lateral frontal cortex." Proceedings of the National Academy of Sciences **95**(13): 7721-7726.

Padoa-Schioppa, C. and J. A. Assad (2006). "Neurons in the orbitofrontal cortex encode economic value." Nature **441**(7090): 223-226.

Padoa-Schioppa, C. and J. A. Assad (2008). "The representation of economic value in the orbitofrontal cortex is invariant for changes of menu." Nat Neurosci **11**(1): 95-102.

Pandya, D. N. and C. L. Barnes (1987). Architecture and connections of the frontal lobe. The Frontal Lobes Revisited. E. Perecman. New York, The IRBN Press**:** 41-72.

Pandya, D. N. and E. H. Yeterian (1985). Architecture and connections of cortical association areas. Association and auditory cortices. A. Peters and E. G. Jones. New York, Plenum Publishing Corporation. **4:** 3 - 61.

Pashler, H. (1988). "Familiarity and visual change detection." Perception & Psychophysics **44**(4): 369-378.

Petrides, M. (1996). "Specialized Systems for the Processing of Mnemonic Information within the Primate Frontal Cortex." Philosophical Transactions of the Royal Society B **351**: 1455-1462.

Petrides, M. and D. N. Pandya (1994). Comparative architectonic analysis of the human and macaque frontal cortex. Handbook of Neuropsychology. F. Boller and J. Grafman. New York, Elsevier. **9:** 17 - 57.

Rainer, G., W. F. Asaad, et al. (1998). "Selective representation of relevant information by neurons in the primate prefrontal cortex." Nature **393**(6685): 577-579.

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

Reynolds, J. and L. Chelazzi (2004). "Attentional modulation of visual processing." Neuroscience **27**(1): 611.

Reynolds, J., T. Pasternak, et al. (2000). "Attention Increases Sensitivity of V4 Neurons." Neuron **26**: 703-714.

Rolls, E. T. (1996). "The orbitofrontal cortex." Philos Trans R Soc Lond B Biol Sci **351**(1346): 1433-1443; discussion 1443-1434.

Rolls, E. T. and L. L. Baylis (1994). "Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex." J Neurosci **14**(9): 5437-5452.

Romanski, L. and P. Goldman-Rakic (2002). "Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex." Nature Neuroscience **2**(12): 1131-1136.

Romanski, L. M., J. F. Bates, et al. (1999). "Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey." J Comp Neurol **403**(2): 141-157.

Romo, R., C. D. Brody, et al. (1999). "Neuronal correlates of parametric working memory in the prefrontal cortex." Nature **399**: 470-473.

Rouder, J., R. Morey, et al. (2008). "An assessment of fixed-capacity models of visual working memory." Proceedings of the National Academy of Sciences **105**(16): 5975.

Saalmann, Y. B., I. N. Pigarev, et al. (2007). "Neural mechanisms of visual attention: how top-down feedback highlights relevant locations." Science **316**(5831): 1612-1615.

Sanders, A. F. (1968). "Short term memory for spatial positions." Psychologie **23**: 1-15.

Shevell, S. K. (2003). The science of color. Oxford, Elsevier.

Siegel, M., M. Warden, et al. (2009). "Phase-dependent neuronal coding of objects in short-term memory." Proceedings of the National Academy of Sciences **106**(50): 21341.

Sperling, G. (1960). "The information available in brief visual presentations." Psychological Monographs: General and Applied **74**(11): 1-29.

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

Toth, L. J. and J. A. Assad (2002). "Dynamic coding of behaviourally relevant stimuli in parietal cortex." Nature **415**(6868): 165-168.

Tremblay, L. and W. Schultz (1999). Relative reward preference in primate orbitofrontal cortex. Nature. **398:** 704-708.

Treue, S. and J. Maunsell (1996). "Attentional modulation of visual motion processing in cortical areas MT and MST." Nature **382**: 539-541.

Vogel, E., A. McCollough, et al. (2005). "Neural measures reveal individual differences in controlling access to working memory." Nature **438**: 500-503.

Vogel, E., G. Woodman, et al. (2001). "Storage of features, conjunctions, and objects in visual working memory." Journal of Experimental Psychology **27**(1): 92-114.

Vogel, E. K. and M. G. Machizawa (2004). Neural activity predicts individual differences in visual working memory capacity. Nature. **428:** 748-751.

Voytek, B., M. Davis, et al. (2010). "Dynamic neuroplasticity after human prefrontal cortex damage." Neuron.

Walker, E. (1940). "A cytoarchitectural study of the prefrontal area of the macaque monkey." Journal of Comparative Neurology **98**: 59 - 96.

Wallis, J. D. (2007). "Orbitofrontal cortex and its contribution to decision-making." Annu Rev Neurosci **30**: 31-56.

Wallis, J. D., K. C. Anderson, et al. (2001). "Single neurons in prefrontal cortex encode abstract rules." Nature **411**(6840): 953-956.

Warden, M. and E. Miller (2007). The Representation of Multiple Objects in Prefrontal Neuronal Delay Activity. Cerebral Cortex.

Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. Nature. **382:** 629-632.

Webster, M. J., J. Bachevalier, et al. (1994). "Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys." Cereb Cortex **4**(5): 470-483.

Wilken, P. and W. Ma (2004). A detection theory account of change detection. Journal of Vision.

Wilson, F., S. Scalaidhe, et al. (1993). Dissociation of Object and Spatial Processing Domains in Primate Prefrontal Cortex. ScienceNew Series. **260:** 1955-1958.

Zald, D. H. (2007). "Orbital versus dorsolateral prefrontal cortex: anatomical insights into content versus process differentiation models of the prefrontal cortex." Ann N Y Acad Sci **1121**: 395-406.

Zhang, W. and S. J. Luck (2008). "Discrete fixed-resolution representations in visual working memory." Nature **453**(7192): 233-235.

Zoccolan, D., D. Cox, et al. (2005). "Multiple Object Response Normalization in Monkey Inferotemporal Cortex." Journal of Neuroscience **25**(36): 8150-8164.

Zohary, E., M. Shadle, et al. (1994). "Correlated neuronal discharge rate and its implications for psychophysical performance." Nature **370**: 140-143.