# Neuronal correlates of metacognition in primate frontal cortex

by

Paul Middlebrooks

B.S. Molecular Biology, University of Texas, 2001

Submitted to the Graduate Faculty of

University of Pittsburgh in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

This dissertation was presented

by

Paul Middlebrooks

It was defended on

May 18th, 2011

and approved by

Aaron Batista, Ph.D., Dept. of Bioengineering

Carol Colby, Ph.D., Dept. of Neuroscience

Stefan Everling, Ph.D., University of Western Ontario

Julie Fiez, Ph.D., Dept. of Psychology

Beatriz Luna, Ph.D., Dept. of Psychology

Dissertation Advisor: Marc Sommer, Ph.D., Dept. of Neuroscience

**Neuronal correlates of metacognition in primate frontal cortex**

Paul Middlebrooks, Ph.D.

University of Pittsburgh, 2011

We spend a large portion of life as the object of our own thoughts. Daily we reflect on all sorts of recent and not so recent decisions, and the products of those reflective thoughts serve to guide future goals, actions, and thoughts. The process of “thinking about thinking”, or metacognition, has garnered scrutiny in psychology studies for decades and recently in some imaging and neurological studies, but its neuronal basis remains unknown. Moreover, metacognition is largely thought a uniquely human ability, and only very recently has some evidence suggested other species may harbor metacognitive skills. To begin investigating neuronal mechanisms underlying metacognition, we performed two experiments.

First, we tested whether rhesus macaques exhibited evidence for metacognition. We trained monkeys to perform a visual oculomotor metacognition task. In each trial, monkeys made a decision then made a bet. To earn maximum reward, monkeys had to monitor their decision and then make a bet to indicate whether the decision was correct or incorrect. We found the monkeys’ behavior was best explained by a metacognitive strategy, and we ruled out possible alternative strategies to perform the task such as reliance on visual stimuli or saccadic reaction times.

Second, we tested whether neurons exhibited activity correlated with metacognition. While monkeys performed the task we recorded from single neurons in three frontal cortical areas known to play roles in higher cognitive functions: the frontal eye field, lateral prefrontal cortex, and the supplementary eye field. Our predictions were that frontal eye field neuronal activity would correlate with making the decisions but not the bets, and that lateral prefrontal cortex and supplementary eye field neuronal activity would correlate with linking the decisions to the bets – the putative metacognitive signals. We found signals in all three brain areas correlated with making decisions and correlated with making bets. The supplementary eye field was the only area of the three that exhibited strong signals correlated with metacognitive monitoring, and these signals appeared early and were sustained throughout the task. Our results identify the supplementary eye field as a likely contributor to metacognitive monitoring.

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# PREFACE

Chapter 2 has been published, and the work in Chapter 3 is in preparation.

Middlebrooks, PG, Sommer, MA (2011) Metacognition in monkeys during an oculomotor task. Journal of Experimental Psychology: Learning, Memory, and Cognition 37(2):325-337.

Middlebrooks, PG, Sommer, MA (*in preparation)* Neuronal correlates of metacognition in monkey frontal cortex.

*Acknowledgments*

My advisor Marc Sommer provided the adventurous spirit and graciousness that allowed and encouraged this dissertation. I am especially indebted to him for sharing his expertise in communicating science, and maybe most of all his willingness to embrace the adage that if you try something risky and it works, it was brave; if it doesn't work, it was stupid. Hopefully we landed closer to the former. Thanks to the rest of my thesis committee, Aaron Batista, Carol Colby, Julie Fiez, and Bea Luna. Their insights enlightened me and vastly improved my research, despite my seemingly best efforts to confuse them. Special thanks to my outside examiner Stefan Everling for hopping the national fence to come down here and spend precious time contributing expertise to this endeavor. I have deep gratitude to Carl Olson and Carol Colby, who taught me new ways to think critically, always had an open door, and cut the ham

and served the fixings every Christmas party. That gratitude extends as well to the CNUP and CNBC communities as a whole.

Both the CNUP and CNBC staff made my ride smoother than I could ever reasonably ask. First, thanks to Joan Blaney, juggler extraordinaire, who kept an eye on me the entire time while juggling the rest of the CNUP as well- she never drops a ball. Thanks to those on the CNUP side, including Marlene Nieri, Kathy Short, Joni Mainier, Debbie Glumac, and Mary Spanoudakis, and the CNBC side, including Barb Dorney, Melissa Stupka, Rebecca Clark, Anna Hegedus, and Lisa Bopp. I also owe a great deal to Karen McCracken who comes in every day and provides excellent care for the entirety of animals.

I would have been lost without the generosity and friendship of many. Among the myriad, thank you to Patrick Mayo, who was always eager to help in any way and who by comparison invariably made me feel like an optimist, and to Robin Ashmore- without his programming skills, I would still be analyzing data.

Mom and dad, thanks for instilling in me the vital mix of drive, curiosity, and freedom of thought that was the foundation for this work. I was fortunate that along the path here Carol Otey and Ben Philpot both took chances on me and taught me how to do experimental science. I would be amiss without a special thank you to Nelson Totah, who hosted the bash that led me to meet my wife. Lastly, thank you to said wife, Catherine, who agreed to marry me even though I do academic research. Her patience and encouragement are without limit, and she provided continual rejuvenation to me.

# 1.0 GENERAL INTRODUCTION

Imagine you are standing in the kitchen preparing a meal. Out of the corner of your eye you see a slight movement, but glancing in that direction, you see nothing. You think, was it an insect? With your eyes still fixed where the movement occurred, you evaluate the situation. What kind of movement was it? Have there been many bugs in here before? Is the kitchen really messy, or pretty clean? You make a decision that prompts some behavior: you may search for the bug, or make a mental note to buy bug spray, or just continue cooking with heightened arousal. This process of monitoring one's cognition and using the information to guide subsequent courses of action and thought is called metacognition.

## 1.1 METACOGNITION

Psychologists have contemplated and studied metacognition for decades. The earliest scientific research on the topic occurred before the term metacognition was invented. In what may have been the first experimental paradigms, subjects were asked to assess their own "feeling of knowing" whether an item was in their memory even though they could not presently recall it (Hart, 1965). Later John Flavell coined the term *metamemory* (Flavell, 1971), referring to thinking about one's memories, and then he broadened the concept to include thinking about ones ongoing cognitive processes, *metacognition* (Flavell, 1976). Flavell proposed some of the first systematic descriptions of various types of metacognitive processes, often in the context of child development (Flavell, 1979).

More recently, Nelson and Narens (1990) established what has since become a well-established psychological framework for understanding and studying metacognition (for a review of the scope of the framework's influence, see Van Overschelde, 2008). In the framework, any cognitive process can be split into at least two separate processing levels (Figure 1): a primary *object-level* and a secondary *meta-level* that contains a dynamic model of the object-level. The object-level and meta-level interact and are related by information exchanged between them. Specifically, when the object-level sends information to the meta-level, it allows the meta-level to *monitor* the object-level. When the meta-level sends information to the object level, it allows the meta-level to *control* the object level. A hierarchical construct like the object-level/meta-level psychological framework is attractive in its simplicity and lends itself to theoretical consideration, but from a neuroscience perspective it has little explanatory power. It is unclear how these processes would be physiologically encoded by neurons in the brain. The overall goal of my dissertation experiments was to advance the field from abstract concepts to concrete facts about the activity of neurons during a metacognitive task. My intended contribution is to help establish a physiological foundation for understanding how metacognitive processes are implemented in the brain, and further inform our general understanding of what is meant by metacognition.

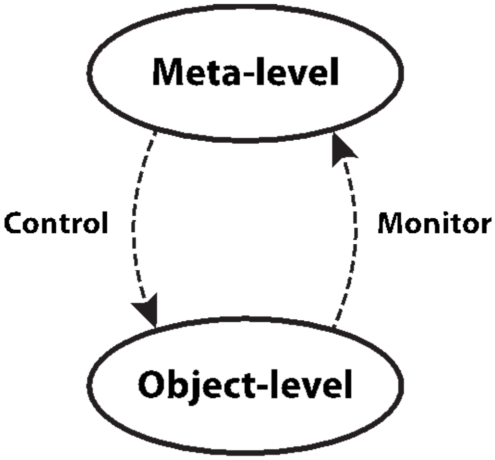


**Figure 1: A framework for metacognition**

Any cognitive process can be split into at least two levels, an object-level and a meta-level. The type of metacognitive process, monitoring or control, depends on the direction of information flow (dashed arrows) between the two levels. Adapted from Nelson and Narens 1990.

## 1.2 METACOGNITIVE BEHAVIOR IN NONHUMAN ANIMALS

Before one can hope to study how a metacognitive process may be implemented at the level of single neurons, it is essential to determine the extent to which laboratory animals can engage in metacognition. Although humans are often assumed to be the only metacognitively skilled creatures on the planet, a growing body of literature suggests at least some nonhuman animals (hereafter, animals) harbor metacognitive skills as well.

Animal metacognition studies to date can be grouped into a few classes, two of which I review below. One line of experiments employ an “opt-out” task designed to test an animal’s ability to monitor its own uncertainty. Opt-out tasks provided the first attempts to test metacognition in animals, and have been the most widely used since then. In each trial of an opt-out task, an animal is presented with a test of varying difficulty and then elects to take or decline the test. If the animal takes the test, it reaps a large reward for a correct response or no reward for an incorrect response, and if the animal declines, it earns a small but ensured reward. An animal monitoring its own uncertainty should opt-out more often on difficult trials than on easy trials, and should perform the task better when given the chance to opt-out than when forced to take the test. Some species that have met the criteria are rats (Foote and Crystal, 2007; Kepecs et al., 2008), dolphins (Smith et al., 1995), rhesus macaques (Shields et al., 1997; Smith et al., 1998; Hampton, 2001) and orangutans (Suda-King, 2008).

Another line of animal metacognition research utilizes betting tasks, designed to test whether an animal can monitor the accuracy of its own task performance. Each trial of a betting task involves two steps: first an animal takes a test of varying difficulty, and then places a wager to report whether the test choice was correct. Reward is earned based on whether the wager accurately reflects the test choice: maximum reward is earned by making high bets after correct test choices and low bets after incorrect test choices. Therefore, an animal able to monitor its own decisions should make high bets after correct test choices and low bets after incorrect test choices. Pioneering studies demonstrated that rhesus monkeys can perform betting tasks as if relying on metacognition (Shields et al., 2005; Kornell et al., 2007). As I review in Chapter 2, however, the tasks were insufficiently controlled to rule out alternative explanations for the betting. When I began my project, therefore, I considered evidence for metacognition in rhesus monkeys to be compelling but not incontrovertible.

While opt-out and betting tasks both provide avenues to explore metacognition in animals, betting tasks offer some important advantages. Betting tasks hew closer to the conceptual framework of Figure 1, in that they involve both an explicit test choice decision (an object-level operation) and an explicit wager (a meta-level operation). Opt-out tasks do not require a test choice decision. The design of opt-out tasks for use in studying animals, furthermore, leads to significant problems in interpreting the results. The opt-out choice must be distinguished from other aborts, and this is often accomplished by requiring the animal to choose a specific stimulus to indicate that it is declining to proceed with the trial. The decline response stimulus may be interpreted simply as a third test choice instead of an escape option. In many opt-out tasks, the decline response stimulus is presented concurrently with the test choice options, allowing the animal to continue gathering evidence from visible stimuli instead of relying solely on internal cognitive operations (see Hampton 2001 for an exception). Finally, opt-out task behavior can be explained by reinforcement learning models that do not invoke uncertainty monitoring (Smith et al., 2008). Considering these factors, we decided to use a betting task to explore the neuronal correlates of metacognition in monkeys. We designed a task in which monkeys made a decision and then had to wager on whether the decision had been correct.  All trials required the monkeys to complete the same series of events; the only major differences from trial to trial were the animals' decisions and bets (task is described in detail in Chapter 2). Our task design improved on previous betting paradigms for monkeys by eliminating extraneous cues that could be used to help betting, thus requiring monkeys to rely fully on the internal representation of their decisions.

## 1.3 LOCALIZING METACOGNITION IN THE BRAIN

Studies in humans that search for brain regions involved in metacognition, like the behavioral animal metacognition studies, have mostly employed paradigms that test metacognitive monitoring. Lesion studies have linked monitoring deficits to damaged medial and/or lateral prefrontal cortex (Schnyer et al., 2004; Pannu et al., 2005b). Transcranial magnetic stimulation

(TMS) delivered over dorsolateral prefrontal cortex impairs subjects' monitoring ability (Rounis et al., 2010). Several functional magnetic resonance imaging (fMRI) experiments have measured blood oxygen level-dependent (BOLD) responses while testing a range of subjects' metacognitive monitoring (Henson et al., 2000; Dobbins et al., 2002; Kikyo et al., 2002; Maril et al., 2003; Schnyer et al., 2005; Chua et al., 2006; Modirrousta and Fellows, 2008; Chua et al., 2009; Kim and Cabeza, 2009). Taken together, these fMRI studies implicate a broad range of brain regions as participating in metacognition, but the more commonly involved regions include medial prefrontal cortex, lateral prefrontal cortex, and parietal cortex. A related body of imaging experiments sought neural correlates for self-referential thoughts. Like the metacognition studies, many self-referential processing studies report BOLD responses in medial prefrontal regions (Gusnard et al., 2001; Kelley et al., 2002; Macrae et al., 2004). An extensive review concluded that a group of midline cortical regions were the main contributors to self-referential processing (Northoff et al., 2006), regions that greatly overlap with those making up the "default mode" system thought to mediate internally directed cognition (Raichle et al., 2001).

## 1.4 SINGLE NEURON STUDIES

Two studies have recorded single neuron activity related to metacognition. One used an opt-out task (Kiani and Shadlen, 2009) and recorded single neurons in lateral intraparietal (LIP) cortex, an area implicated in visuo-spatial cognition (Colby et al., 1996), attention (Gottlieb et al., 1998) and decision-making (Shadlen and Newsome, 2001). Monkeys were trained to discriminate the motion direction of a visual display of randomly moving dots that had overall coherence in one direction. Trial difficulty was controlled by varying the motion coherence strength in one direction and by varying the duration that the moving-dot stimulus appeared. On half of the trials, the monkeys were required to choose between two direction responses by making an eye movement to one of two targets in the two directions. Reward was delivered after correct decisions but not incorrect decisions. On the other half of trials, the monkeys were offered an opt-out response that always delivered a small reward. When offered the opt-out response, monkeys chose the opt-out target more often on difficult trials than easy trials, suggesting they were less confident on the difficult trials. LIP neuronal activity correlated with the decision to choose either the opt-out target or one of the direction targets, thus LIP was suggested as a correlate of the monkeys' confidence (Kiani and Shadlen, 2009).

In the other study (Kepecs et al., 2008), rats were trained to perform a modified version of an opt-out task while recording single neurons in orbitofrontal cortex (OFC), an area associated with reward, risk, and uncertainty (Hsu et al., 2005; Tobler et al., 2007; O'Neill and Schultz, 2010). The rats were trained to discriminate the majority component odor within a mixture of two odors. They reported a decision by poking their nose into one of two ports (one port for each odor) then received reward if correct or no reward if incorrect. As expected, rats made more correct decisions on easy trials (when the proportion of one odor dominated the other). OFC neurons varied as a function of trial difficulty and whether a correct choice was made. The authors then added a key manipulation to the experiment to assess the rats' confidence in their decisions. Once a rat poked its nose into a port, a random delay was imposed before reward delivery after a correct decision (as before no reward was delivered after an incorrect decision). The rats could endure the wait and earn reward (or risk no reward), or they could abort the trial and immediately start the next trial. The rats’ behavior was consistent with confidence. They were more willing to wait for a reward after an easy correct trial than a difficult correct trial, and conversely after errors they aborted more often when the error was made on an easy trial than on a difficult trial. Unfortunately the authors did not record OFC neurons during the delayed-reward trials (Kepecs et al., 2008), but instead offered a model that correctly predicted the animals’ behavior during the delayed-reward trials and matched the pattern of some OFC neurons’ activity during the initial task.

## 1.5 FRONTAL EYE FIELD, LATERAL PREFRONTAL CORTEX, AND SUPPLEMENTARY EYE FIELD

We studied neuronal correlates of metacognition by harnessing one of the most thoroughly characterized networks of the primate brain, the primate visual oculomotor system. A significant achievement in systems neurosciences over many decades has been the detailed description of neurophysiology underlying visual stimulus processing and eye movement generation. Even more impressive, the visual oculomotor system has been used as a key to unlock what behavioral psychologists had considered a black box (Skinner, 1974; Schall, 2004). Using the preparation developed by Wurtz in the late 1960s (Wurtz, 1968), in which eye movements and neuronal activity are recorded from the awake, behaving monkey, researchers have succeeded in elucidating many of the neuronal mechanisms underlying higher-level cognitive processes like working memory, attention, and decision-making (Goldman-Rakic, 1995; Basso, 1998). Based on that rich history, and guided by human lesion and imaging studies mentioned above, we elected to study three frontal cortical regions: frontal eye field (FEF), lateral prefrontal cortex (PFC), and supplementary eye field (SEF).

The FEF, located in the anterior bank of the arcuate sulcus, is integral to the generation and control of saccades (Bruce et al., 1985; Sommer and Tehovnik, 1997; Dias and Segraves, 1999; Brown et al., 2008). FEF neurons typically have receptive fields in contralateral space (Bruce and Goldberg, 1985). Many FEF neurons respond to visual stimuli (Mohler et al., 1973), discharge before saccades (Bruce and Goldberg, 1985), and/or maintain information during a delay (Funahashi et al., 1989; Sommer and Wurtz, 2001). It is increasingly apparent that FEF plays a role in numerous higher cognitive functions, such as target selection (Schall et al., 1995), attention (Moore and Fallah, 2001; Thompson and Bichot, 2005), decision-making (Kim and Shadlen, 1999; Thompson and Schall, 1999), reward (Roesch and Olson, 2003; Ding and Hikosaka, 2006), and conscious perception (Thompson and Schall, 2000; O'Shea and Walsh, 2004; Libedinsky and Livingstone, 2011). In addition, monkey FEF is part of a circuit that internally monitors eye movements. It receives a copy of upcoming eye movement information ("corollary discharge") from the superior colliculus (Sommer and Wurtz, 2002, 2006) and may relay it to other parts of the brain (Sommer and Wurtz, 2008). Corollary discharge signals could promote a variety of functions, such as visual stability across saccades (Crapse and Sommer, 2008b; Sommer and Wurtz, 2008), shifting the focus of attention (Armstrong and Moore, 2007), and perhaps many of the higher cognitive functions mentioned above. Our hypothesis was that FEF neurons would have activity correlated with making decisions in our task, but not with linking the decisions to the bets. We speculate that the decision-related information encoded in FEF would be sent to other brain regions responsible for linking the decisions to the bets.

The PFC, located anteriorly to the FEF and encompassing the principal sulcus area, has been implicated in a number of high cognitive processes such as executive function (Miller and Cohen, 2001), working memory (Fuster, 1973; Funahashi et al., 1989), reward (Watanabe, 1996;

Roesch and Olson, 2003) target selection and attention (Hasegawa et al., 2000; Iba and Sawaguchi, 2003), abstract rule encoding (Wallis and Miller, 2003), behavioral context (Johnston and Everling, 2006), and decision-making (Kim and Shadlen, 1999). Receptive fields of PFC neurons are broader than those in FEF and harder to delimit (Boch and Goldberg, 1989; Funahashi et al., 1990, 1991). It has been suggested that the various and diverse types of information processed in PFC may be commonly used to guide behavior for a desired outcome (for a review, see Tanji and Hoshi, 2008). We hypothesized that neurons in PFC would help link selected decisions to appropriate bets.

The SEF is located in the rostral region of supplementary motor area in dorsomedial frontal cortex. Like FEF and PFC, many SEF neurons respond to visual stimuli and/or are active before, during, and after saccades in the monkey (Schall, 1991; Olson and Gettner, 1995; Russo and Bruce, 1996). Although saccades can be elicited by microstimulation of SEF (Schlag and Schlag-Rey, 1987), SEF seems more regulatory or executive as opposed to being directly involved in initiating saccades (Stuphorn et al., 2010). SEF has also been implicated in performance monitoring by signaling error, conflict, and reward in the context of a visual saccade task (Stuphorn et al., 2000; Nakamura et al., 2010). We hypothesized SEF neuron activity would correlate with monkeys’ ability to monitor their decisions and thus guide an appropriate bet.

## 1.6 EXPERIMENTAL AIMS

The central goal of this dissertation project was to record single neuron activity in frontal cortex while monkeys performed a metacognitive task. The two experimental aims were as follows:

**Aim 1: Assess metacognition in monkeys**

Previous work on metacognition in monkeys was encouraging, yet difficult to interpret and inconclusive. To assess the metacognitive abilities of our monkeys, we designed a novel, streamlined, visual oculomotor betting task and trained monkeys to perform the task as described in Chapter 2. By using multiple analytical approaches and ruling out possible non-metacognitive strategies to perform the task, we established that our monkeys used metacognition to an extent that made neurophysiological studies of this higher-level function feasible.

**Aim 2: Investigate single neuron activity related to metacognition**

To assess how metacognitive processes are encoded at the level of single neurons, we recorded from neurons in FEF, PFC, and SEF while monkeys performed our task. In Chapter 3 we compare neuronal activity across different task trial outcomes while holding visual and/or eye movement parameters constant. If the neurons encode metacognitive processes, we expected to see a difference in activity among trials in which the monkey maintained differing internal records, as inferred from differing bets, for the same decisions. For example we compared neuronal activity during correct decisions followed by high bets (evidence that the monkey monitored its correct decision), versus neuronal activity during correct decisions followed by low bets (evidence that the monkey failed to monitor its correct decision). We interpreted our results in the context of the hypothesized metacognitive roles of each cortical region.

# 2.0 METACOGNITION IN MONKEYS DURING AN OCULOMOTOR TASK

## 2.1 ABSTRACT

This study investigated whether rhesus monkeys show evidence of metacognition in a reduced, visual oculomotor task that is particularly suitable for use in fMRI and electrophysiology. The 2-stage task involved punctate visual stimulation and saccadic eye movement responses. In each trial, monkeys made a decision and then made a bet. To earn maximum reward, they had to monitor their decision and use that information to bet advantageously. Two monkeys learned to base their bets on their decisions within a few weeks. We implemented an operational definition of metacognitive behavior that relied on trial-by-trial analyses and signal detection theory. Both monkeys exhibited metacognition according to these quantitative criteria. Neither external visual cues nor potential reaction time cues explained the betting behavior; the animals seemed to rely exclusively on internal traces of their decisions. We documented the learning process of one monkey. During a 10-session transition phase, betting switched from random to a decision-based strategy. The results reinforce previous findings of metacognitive ability in monkeys and may facilitate the neurophysiological investigation of metacognitive functions.

## 2.2 INTRODUCTION

Humans possess the ability to monitor and control cognition, a function known as metacognition. We monitor our cognition, for example, when contemplating whether we understood a poem. Likewise, we control our cognition when planning to reread the poem to better grasp its meaning.

Behavioral studies have concluded that rhesus macaques can engage in metacognitive processes. Some studies, using opt-out paradigms, tested monkeys' abilities to monitor their own uncertainty (Smith et al., 1998). Monkeys were offered a choice to take or opt out of a test of varied difficulty. They tended to opt out more often when offered a difficult test, suggesting they can monitor their own uncertainty. Other studies, using betting paradigms, provided more direct evidence (Shields et al., 2005; Kornell et al., 2007). Such studies test monkeys' abilities to monitor their previous decisions, referred to as retrospective monitoring (Nelson and Narens, 1990). A betting paradigm involves two steps: The monkey takes a test of varying difficulty, then places a bet to indicate whether its test choice had been correct. Reward is based on how appropriate the bet is relative to test choice; that is, a correct test choice should be followed by a high bet and an incorrect test choice should be followed by a low bet to earn maximum reward. The general conclusion has been that monkeys can advantageously wager on their own recent decisions, indicating that they are able to monitor those decisions.

Betting paradigms have advantages over opt-out paradigms when assessing metacognition (for a discussion, see, e.g., Metcalfe, 2008). For example, in opt-out paradigms, the opt-out stimulus is presented concurrently with the test choice stimuli, so no memory of a decision is required to make an advantageous response, and the monkey may even interpret the opt-out stimulus as simply a third test choice option (for exceptions, see Hampton, 2001; Foote and Crystal, 2007).

Results from betting paradigms should also be interpreted carefully. Subjects may place bets that are influenced by factors extraneous to the internal trace of the decision. Humans, when asked to assess the accuracy of a memory, can rely on a host of factors beyond directly monitoring the memory (for a summary, see (Benjamin and Bjork, 1996). A few such factors include degree of familiarity with the task subject matter, perceptual stimuli related to the task, latency with which a response is made, and the amount and type of information associated with the memory. Reliance on factors such as these has been shown to influence judgments that supposedly rely on memory (Chandler, 1994; Benjamin et al., 1998; Brewer and Sampaio, 2006). In the pioneering studies with monkeys, metacognitive tasks have involved complex visual inputs at the decision stage, for example, clouds of dots in a numerosity judgment task (Shields et al., 2005) or elaborate reward procedures, such as attainment of a threshold number of virtual tokens that are gained or lost across trials (Kornell et al., 2007). Although the prior tasks involved both elements of retrospective monitoring — decision and bet — their additional features may have recruited cognitive processes irrelevant to metacognition (e.g., assessment of gradual accumulation of tokens). Moreover, in all of the prior tasks, decisions were reported with limb or wrist movements; hence, the animals could keep track of their decisions not only internally but also externally, by watching their effectors or sensing postural adjustments. Historically, researchers faced these types of potential confounds in early studies of working memory as well (Jacobsen, 1936; Fuster and Alexander, 1971; Kubota and Niki, 1971; Wang, 2005). Goldman-Rakic and colleagues (Funahashi et al., 1989; Goldman-Rakic, 1995) offered an improvement by designing a highly controlled and reduced paradigm, the oculomotor delayed-response task. Their simplified task has become central to the modern study of working memory at both the neurophysiological and psychological levels.

We have developed a simplified betting paradigm that eliminates as many as possible of the confounding issues discussed above. We apply it here to achieve what we consider the most demanding test yet of the hypothesis that monkeys are capable of metacognition. The task requires a decision and then a bet but little else. Extraneous cognitive demands and spurious cues that might aid performance are minimized. Additionally, the task is immediately suitable for studying metacognition in monkeys at the single neuron level, and this is a future goal of our work. Following the example set in the working memory field, our paradigm is set in a visual oculomotor context that offers many advantages (Goldman-Rakic, 1995). First, it permits little variability in how the task can be performed. The monkeys sit still in a primate chair and report choices using only eye movements, which are stereotyped from trial to trial. No task-related skeletal movements are involved. Second, the task minimizes variability in the perceptual and cognitive state of the animal. The visual environment is tightly controlled, and each trial represents a self-contained unit of decision, bet, and outcome (with no dependence on trial history as in Kornell et al., 2007). Third, we can interpret our findings in the context of decades of research on the primate oculomotor system. Fourth, with only trivial alterations, the task could be used in functional imaging experiments (on monkeys or humans) by any investigators who can display visual stimuli and track eye movements in their scanner. A host of regions have been implicated in metacognition with fMRI, such as prefrontal cortical regions and posterior parietal cortex (Kikyo et al., 2002; Maril et al., 2003; Chua et al., 2006).

We trained two monkeys on the oculomotor betting task and found that the hypothesis was supported: The animals exhibited metacognitive behaviors within a few weeks and maintained near optimal betting behavior thereafter. Our findings confirm that monkeys can monitor and use information about their decisions in a task that is a simple extension of standard oculomotor paradigms and of immediate use to neuroscientists.

## 2.3 METHODS

### 2.3.1 Monkeys

Two rhesus monkeys (Macaca mulatta) served as the subjects. Monkey N had substantial previous experience performing oculomotor tasks. Monkey S was initially naïve to all laboratory tasks. Each monkey was implanted with scleral search coils to monitor eye position with high spatial (0.1-degree) and temporal (1-ms) precision (Judge et al., 1980). The temporal resolution was particularly important for us, because we wanted to study saccadic reaction times and, in future work with these animals, correlate the eye movements with 1-ms resolution neuronal data. The use of scleral search coils is not critical, however, and standard video eye-tracking methods could be used for implementing the task in purely psychophysical or imaging studies. A plug for connecting to the eye coil leads and a plastic post for keeping the head still during experiments were bound with acrylic and affixed to the skull with bone screws using aseptic techniques while animals were anesthetized (Sommer and Wurtz, 2000). Procedures were approved by and conducted under the Institutional Animal Care and Use Committee of the University of Pittsburgh.

### 2.3.2 Task

A monkey sat in a primate chair facing a tangent display screen in a dimly lit room. Visual stimuli from a 60-Hz LCD projector were back-projected onto the screen. The REX real-time system (Hays et al., 1982) was used to control behavioral paradigms and collect eye position data at 1 kHz. Every trial of the task included a decision stage followed by a bet stage (Figure 2a, spatial layout; Figure 2b, timing).

*Decision Stage*

The animal's goal in the decision stage (Figure 2a, top; Figure 2b, left) was to detect and report the location of a peripheral target (Thompson and Schall, 1999). The monkey fixated a center spot for 500–800ms (randomized by trial). Then a dim target appeared in one of four possible locations (also randomized). The locations were constant throughout a session but often varied between sessions; they were always mirror symmetric about the vertical axis, with one location in each quadrant, but their eccentricities could range from 5 to 25 degrees, and their directions, relative to the horizontal meridian, could range from 0 to 60 degrees in angle. Varying the location is not a necessary feature of the task but was meant to accustom the animals to stimuli presented over a wide range of space, because in subsequent neurophysiological recordings, we expected that the centers of neuronal response fields would vary considerably. After the target appeared, mask stimuli appeared at all four locations. The interval between target appearance and appearance of the masks, the stimulus onset asynchrony (SOA) was randomized by trial: 16.7, 33, 50, or 66.7ms. The monkey had to decide, "Where was the target?" Shorter SOAs made the target more difficult to detect, and longer SOAs made it easier. After the masks appeared, a randomized 500 – 1,000ms delay ensued during which the monkey continued maintaining fixation while the masks remained visible. The fixation spot was then extinguished, cueing the monkey to report its decision by making a saccade to the perceived target location within 1,000ms. The monkey received no performance feedback until after the bet stage of the task, but the computer tracked whether the decision was correct (saccade landed in an electronic window around the target location) or incorrect (saccade landed anywhere else). The size of the electronic window was adjusted to contain typical scatter of saccadic endpoints for particular target amplitudes. If at any time during the decision stage, the monkey broke fixation, made a saccade before cued to go, or failed to make a saccade, the trial was aborted, (and repeated again later) and the next trial began immediately. Such an early abort might occur for many reasons— for example, if the monkey failed to see the target at all or simply wanted to rest—so we neither penalized nor quantified such trials. In all trials, the saccade ended the decision stage and started the bet stage.

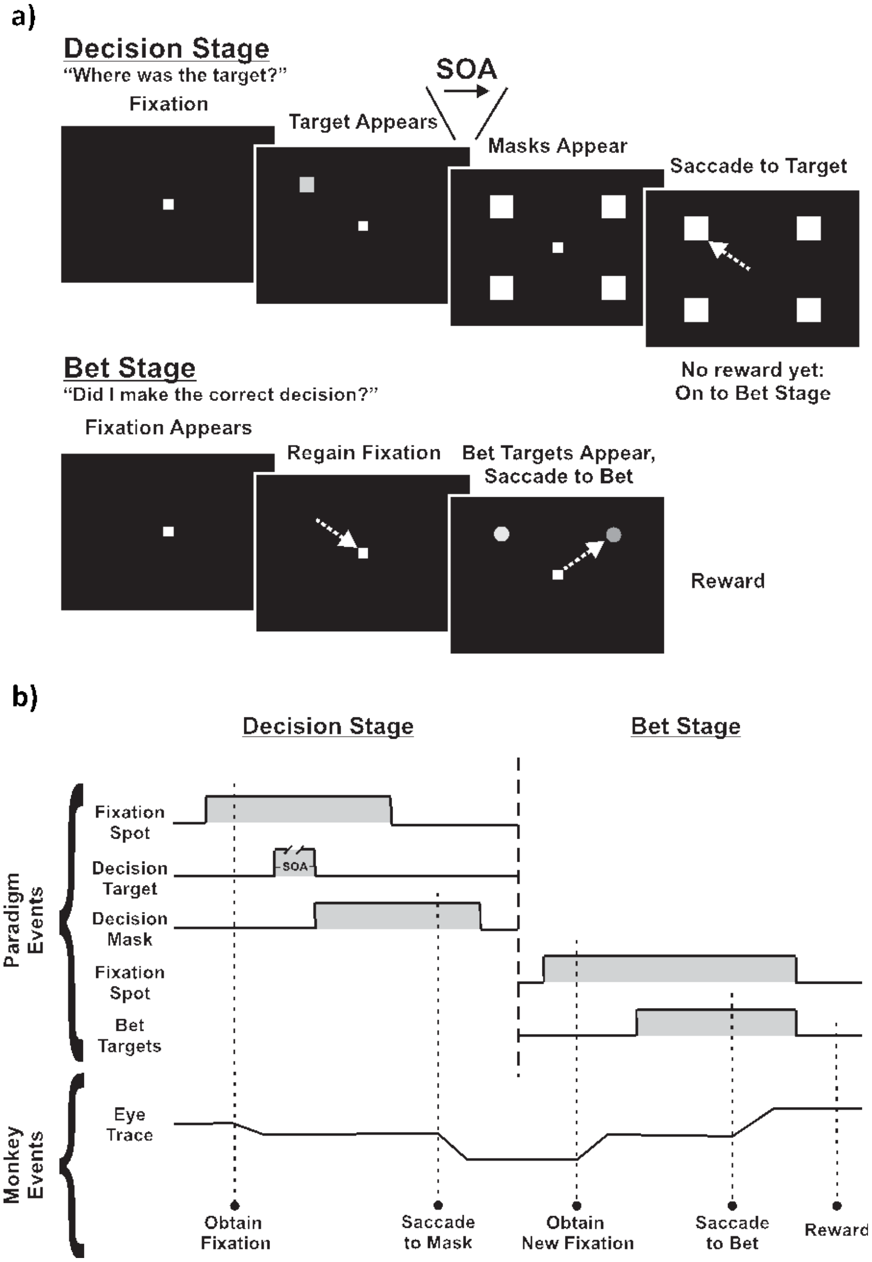


**Figure 2: Spatial aspects and timing of task.**

(a) Spatial layout: Each trial consisted of two stages—a decision stage (top) and a bet stage (bottom). In the decision stage, monkeys foveated a fixation spot. A target appeared at one of four locations in the periphery, and after a variable time (stimulus onset asynchrony [SOA]), masks appeared at the four locations. A correct decision was made (shown) if a saccade went to the location of the target. A saccade to any other location was an incorrect decision. In the bet stage, monkeys foveated a new fixation spot, then the two bet targets appeared in the periphery. A bet was made when a saccade went to one of the two targets, completing a single trial. (b) Timing: Top bracket shows when stimuli were visible (gray shading) and invisible (flat line) during each trial. Bottom bracket shows a hypothetical eye trace and actions performed by monkeys during the sequence of events.

### *Bet Stage*

The goal in the bet stage of each trial (Figure 2a, bottom; Figure 2b, right) was to wager on the decision just made. A new fixation spot appeared, and the monkey foveated it. After a brief delay (500 – 800ms), two bet targets appeared: a red high-bet target and a green low-bet target (for Monkey N; this color assignment was reversed for Monkey S). A monkey reported its bet by making a saccade to one of the targets, then received a reward or a time-out as described below, and the trial ended. A monkey would optimize its reward if it bet high after a correct decision and low after an incorrect decision.

In a given session, the bet targets always appeared in the same two locations. Within those two locations, the appearance of either bet target was randomized by trial. If, during the bet stage, the monkey broke fixation or made a saccade to a non–bet-target location, the trial was aborted, and a brief time-out ensued before a new trial began.

### *Reward*

The amount of reward delivered after each trial was based on how appropriate the bets were relative to the decisions. If the monkey made a correct decision and bet high, it earned the maximum reward: five drops of juice. If the monkey made an incorrect decision and bet high, it received no reward and a 5-s time-out. Betting low earned a sure but minimal reward: three drops after a correct decision and two drops after an incorrect decision. The reward schedule was based on previous studies (Kornell et al., 2007; Persaud et al., 2007) and fine-tuned to elicit the best performance in our monkeys.

### 2.3.3 Training

Retrospective monitoring tasks are challenging for monkeys to learn (Son and Kornell, 2005), so we first familiarized the animals with each stage separately. We ran blocks of trials in which monkeys performed only the decision stage of the task, receiving immediate reward for correct trials. We also ran blocks of trials in which they performed only the bet stage. For those we assigned fictitious, random correct and incorrect variables to control reward that would be delivered after a bet target choice. The bet stage-only sessions exposed the animals to the two bet targets and the potential reward-related outcomes associated with each.

After the monkeys became familiar with the decision and bet stages, we started training them on the full task: the decision stage followed by the bet stage. The animals learned by trial and error. Training of the first animal (Monkey N) proceeded somewhat sporadically, because we tried various stimulus parameters and adjusted the computer programs. We did not document that animal's learning curve. Training of Monkey S was accomplished thereafter, involved no changes to the task, and provided detailed data on the learning progression (see Results).

## 2.4 RESULTS

### 2.4.1 Average Performance

*Decision stage*

As expected, and in agreement with previous work (Thompson and Schall, 1999), both monkeys' proportion of correct decisions increased with stimulus onset asynchrony (SOA). Figure 3a shows Monkey N’s average proportion of correct decisions for each SOA (49 experimental sessions; 8,735 trials). Figure 3c shows the same data for Monkey S (55 sessions; 11,201 trials). All data are from steady-state performance after training. Chance performance is probability .25. At the shortest SOA (16.7ms), both monkeys performed only slightly above chance (~.3; p < 01, one-sample t test). At the longest SOA (66.7ms), performance greatly exceeded chance; the probability of making a correct decision was ~.7. One-way analysis of variance (ANOVA) confirmed a significant effect of SOA on correct decisions (p < .001 for each monkey). Hence, the range of SOAs resulted in a decision task that varied smoothly in difficulty from hard (but not impossible) at the shortest SOA to easy (but not trivial) at the longest SOA.

*Bet stage*

We predicted that if the monkeys were using metacognition, that is, basing their bets on their decisions, then, on average, their proportion of high bets would increase with SOA in parallel with their decisions. Figures 3b and 3d show each monkey’s average rate of high bets for each SOA for the same sessions as Figures 3a and 3c. An ANOVA showed that high bets occurred with probability ~.35 at the 16.7-ms SOA and rose steadily to probability ~.6 at the 66.7-ms SOA (p < .001 for each monkey). Hence, on average, the monkeys’ bets tracked their decisions.

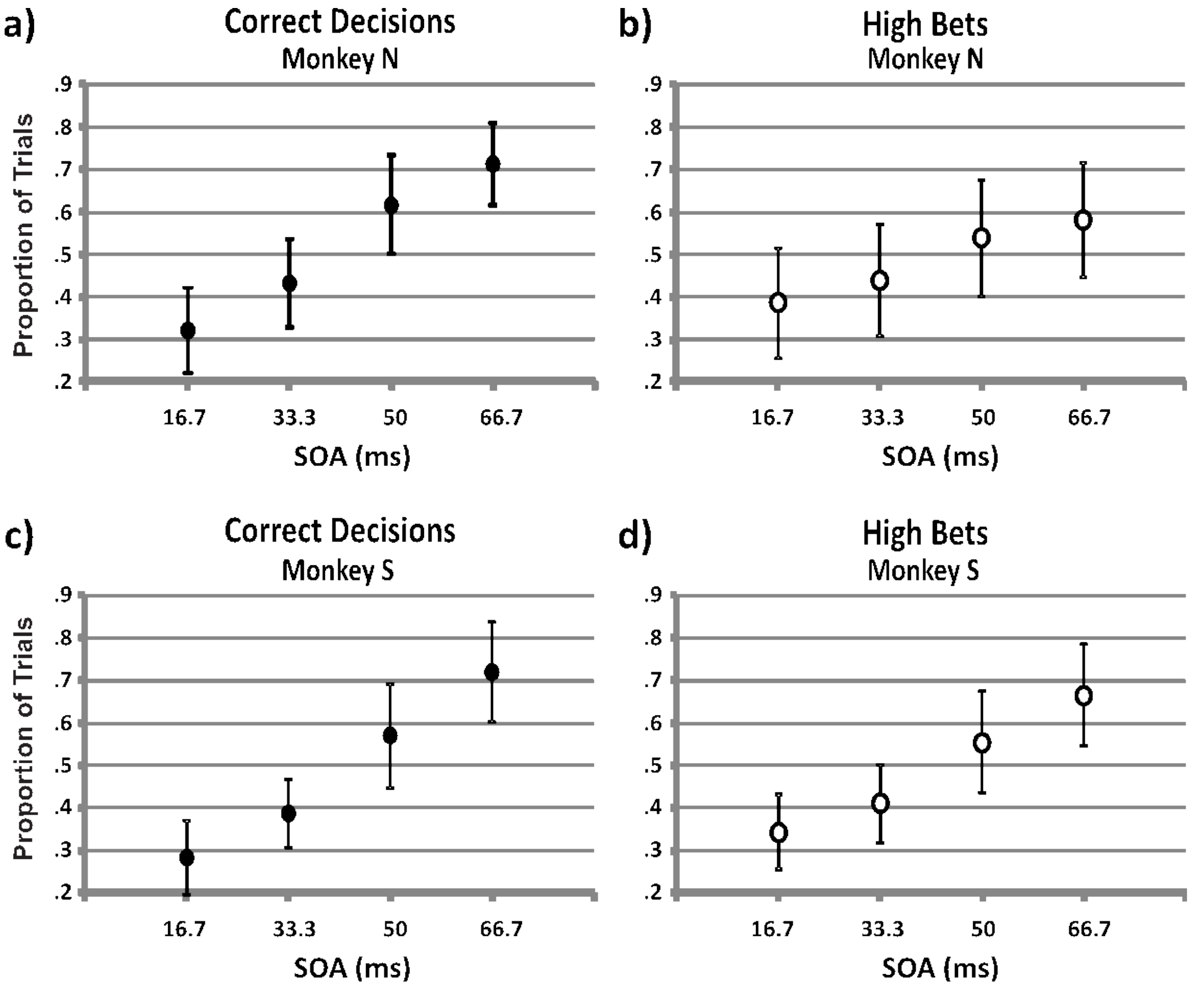


**Figure 3: Average decisions and bets**

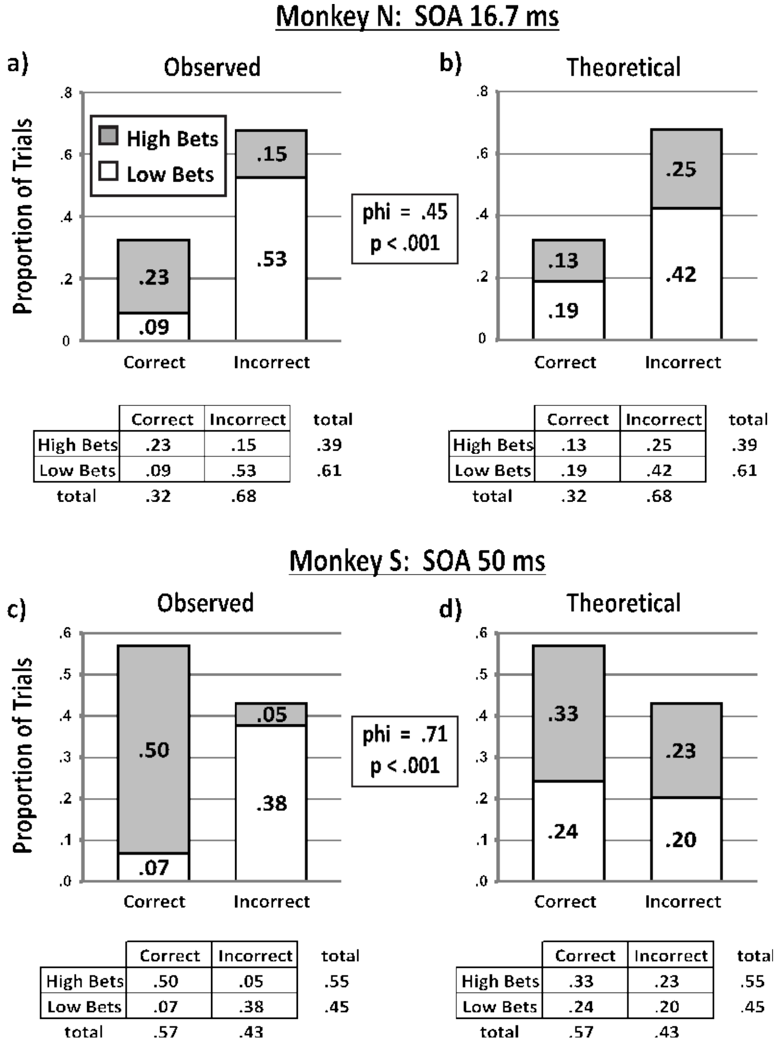
(a) Overall proportion of correct decisions made by Monkey N plotted as a function of each of the four stimulus onset asynchronies (SOAs). Total trials = 8,735 over 49 sessions. (b) Overall proportion of high bets made by Monkey N plotted as a function of each of the four SOAs during the same sessions as in Panel a. (c) Overall proportion of correct decisions made by Monkey S plotted as a function of each of the four SOAs. Total trials = 11,201 over 55 sessions. (d) Overall proportion of high bets made by Monkey N plotted as a function of each of the four SOAs during the same sessions as in Panel c. Error bars represent standard deviations.

### 2.4.2 Trial-by-Trial Performance

Our operational definition of metacognition is that a monkey displays metacognition if its bets track its decisions. As reported above, this was true on average, but, in principle, the same results could occur without metacognition. A monkey could use an external cue such as the sequence of visual stimulation during the decision stage, betting low after short SOAs and high after long SOAs, regardless of its decisions. To test this possibility, we performed a trial-by-trial analysis.

In the trial-by-trial analysis, we examined whether a monkey's bets were correlated with its decisions independent of the sensory stimulation that it experienced. If true, then for identical visual input (constant SOA), a monkey should make more low bets after incorrect decisions and more high bets after correct decisions. Figure 4 illustrates an example of the analysis for each monkey. Figure 4a shows the distribution of Monkey N's decisions and bets for all 16.7-ms SOA trials (n = 2,207) over 49 sessions. After correct decisions, the animal was much more likely to make a high bet (probability .23 over all four possible trial outcomes, gray shading) than a low bet (probability .09, white shading). Conversely, after incorrect decisions, the animal was less likely to make a high bet (probability .15) and more likely to make a low bet (probability .53). Figure 4b shows the theoretical response rates for the case in which the animal made bets irrespective of its decisions, that is, if it made a constant proportion of high to low bets whether its decision was correct or incorrect. Considering just the subset of correct decision trials, we note that high bets would occur with overall probability .13 and low bets would occur with probability .19. In the incorrect decision trials, high bets would occur with overall probability .25 and low bets, with probability .42 (these four probabilities add to .99, rather than 1, because of rounding error). The table below each graph summarizes the proportions of bets (rows) and decisions (columns), illustrating that the marginal distributions—the average probabilities of making the particular bets and decisions—were identical in the observed and simulated data (the latter having been calculated from the former). Consistent with previous analyses (Kornell et al., 2007), we compared these two distributions and found them to be significantly different, $\chi 2$(1, N

= 2,207) = 439, p < .001. Results are also shown for Monkey S's 50-ms SOA trials (Figures 4c and Figure 4d; N = 2,786 from 55 sessions), and the conclusions were the same, $\chi$2 (1, N = 2,786) = 1,417, p < .001. To quantify the magnitude of the difference between observed betting and random (theoretical response rate) betting, we calculated the phi correlation (the chi-square statistic normalized for the total number of trials). A phi correlation of 0 indicates that two distributions are identical, and a phi correlation of 1 indicates that they are unrelated. As shown in Figure 4, the phi correlation was .45 for Monkey N's 16.7-ms SOA trials and .71 for Monkey S's 50-ms SOA trials. We repeated the analysis for every SOA and pooled the data to provide an overall trial-by-trial result (Figure 5). The overall phi correlation was .48 for Monkey N and .74 for Monkey S. All the phi correlations were greater than zero (p < .001) and represented nonrandom betting behavior at a level of skill similar to that reported in previous studies (Kornell et al., 2007). Table 1 shows the phi correlations for each SOA; all were significant (p < .001) and varied little across SOA (the standard deviations were <□5% of the means for both monkeys, and relationships with SOA were insignificant, p < .05 for both monkeys, Pearson's test). In sum, both animals' bets tracked their decisions above chance on a trial-by-trial basis, satisfying our criterion for metacognitive behavior.



|  | Correct | Incorrect | total |
|---|---|---|---|
| High Bets | .23 | .15 | .39 |
| Low Bets | .09 | .53 | .61 |
| total | .32 | .68 |  |

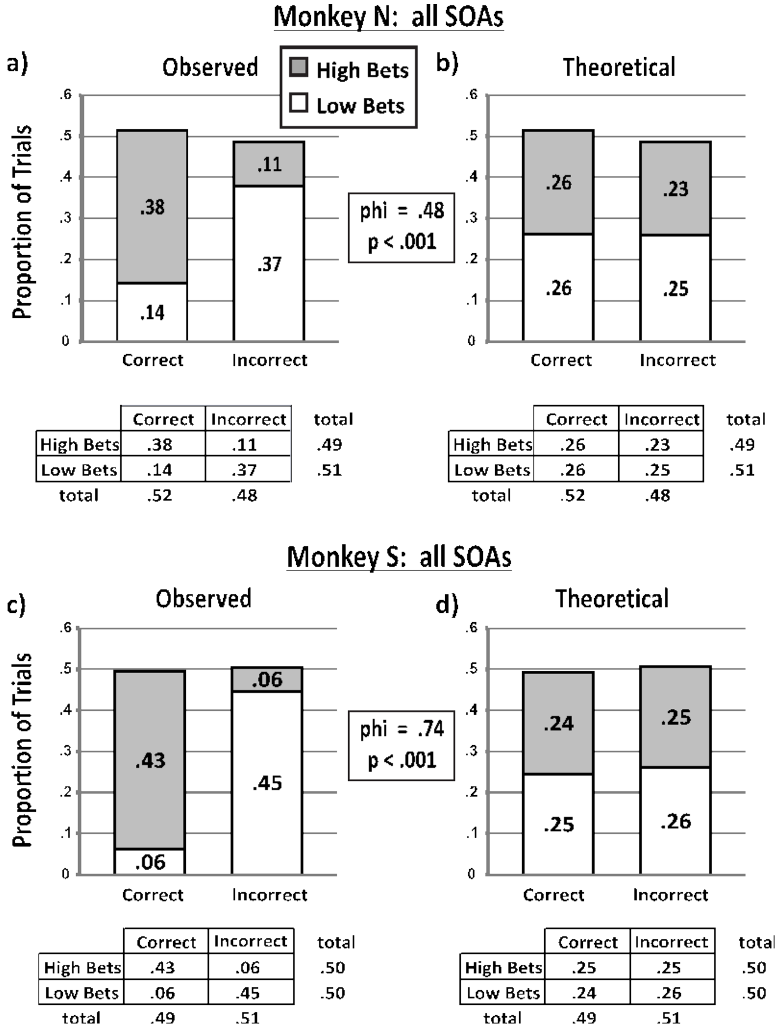
|  | Correct | Incorrect | total |
|---|---|---|---|
| High Bets | .13 | .25 | .39 |
| Low Bets | .19 | .42 | .61 |
| total | .32 | .68 |  |

|  | Correct | Incorrect | total |
|---|---|---|---|
| High Bets | .50 | .05 | .55 |
| Low Bets | .07 | .38 | .45 |
| total | .57 | .43 |  |

|  | Correct | Incorrect | total |
|---|---|---|---|
| High Bets | .33 | .23 | .55 |
| Low Bets | .24 | .20 | .45 |
| total | .57 | .43 |  |

**Figure 4: Trial-by-trial analysis for example stimulus onset asynchronies (SOAs)**

(a) The distribution of trial outcomes for Monkey N's 16.7-ms SOA trials over 49 sessions, 2,207 trials. The bar graph shows the observed proportions of high bets (gray shading) and low bets (white shading) among correct and incorrect decisions. Below the bar graph, the table shows these same outcome proportions listed with the marginal totals. (b) Theoretical response rates if the monkey placed bets on the basis of an external variable (the SOA) rather than tracking its decisions. Note that the total proportion of high bets, low bets, correct decisions, and incorrect decisions are the same in Panels a and b. Chi-square between the distributions of Panels a and b indicate that they are significantly different (p < .001), with a phi correlation of .45. (c) The distribution of trial outcomes for Monkey S's 50-ms SOA trials over 55 sessions, 2,786 trials. As above, the bar graph shows the observed proportions of high bets (gray shading) and low bets (white shading) among correct and incorrect decisions. Below the bar graph, the table shows these same outcome proportions listed with the marginal totals. (d) Theoretical response rates if the monkey placed bets on the basis of an external variable (the SOA) rather than tracking its decisions. Note that the total proportion of high bets, low bets, correct decisions, and incorrect decisions are the same in Panels c and d. Chi-square between the distributions of Panels c and d indicate that they are significantly different (p < .001), with a phi correlation of .71. In the tables, some rows and columns do not add precisely to marginal distributions because of round-off error.



|  | Correct | Incorrect | total |
|---|---|---|---|
| High Bets | .38 | .11 | .49 |
| Low Bets | .14 | .37 | .51 |
| total | .52 | .48 |  |

|  | Correct | Incorrect | total |
|---|---|---|---|
| High Bets | .26 | .23 | .49 |
| Low Bets | .26 | .25 | .51 |
| total | .52 | .48 |  |

|  | Correct | Incorrect | total |
|---|---|---|---|
| High Bets | .43 | .06 | .50 |
| Low Bets | .06 | .45 | .50 |
| total | .49 | .51 |  |

|  | Correct | Incorrect | total |
|---|---|---|---|
| High Bets | .25 | .25 | .50 |
| Low Bets | .24 | .26 | .50 |
| total | .49 | .51 |  |

**Figure 5: Trial-by-trial analysis for all stimulus asynchronies (SOAs) combined.**

(a) The distribution of trial outcomes for Monkey N's trials combined. The bar graph shows the observed proportions of high bets (gray shading) and low bets (white shading) among correct and incorrect decisions. Below the bar graph, the table shows these same outcome proportions listed with the marginal totals. (b) Theoretical response rates if the monkey placed bets on the basis of an external variable (the SOA) rather than tracking its decisions. Note that the total proportion of high bets, low bets, correct decisions, and incorrect decisions are the same in Panels a and b. Chi-square between the distributions of Panels a and b indicate that they are significantly different (p < .001), with a phi correlation of .48. (c) The distribution of trial outcomes for all of Monkey S's SOA trials combined. As above, the bar graph shows the observed proportions of high bets (gray shading) and low bets (white shading) among correct and incorrect decisions. Below the bar graph, the table shows these same outcome proportions listed with the marginal totals. (d) Theoretical response rates if the monkey placed bets on the basis of an external variable (the SOA) rather than tracking its decisions. Note that the total proportion of high bets, low bets, correct decisions, and incorrect decisions are the same in Panels c and d. Chi-square between the distributions of Panels c and d indicate that they are significantly different (p < .001), with a phi correlation of .74. In the tables, some rows and columns do not add precisely to marginal distributions because of round-off error.

**Table 1: Phi correlations**

Phi correlations for each monkey (rows) calculated for each stimulus onset asynchrony and averaged over all stimulus onset asynchronies (columns). All values were significantly greater than zero (p < .001).

| SOA | 16.7ms | 33.3ms | 50ms | 66.7ms | Mean (SD) |
|---|---|---|---|---|---|
| Monkey N | .45 | .46 | .46 | .42 | .45 (.02) |
| Monkey S | .68 | .74 | .71 | .67 | .70 (.03) |

### 2.4.3 Saccade Latencies

Our trial-by-trial analyses indicate that the monkeys were not using external cues to make bets. Another possibility is that the monkeys could detect their saccade latencies during the decision stage and use this information to help place their bets. In metacognition paradigms, humans tend to make shorter latency responses when confident than when not confident (Costermans et al., 1992). Likewise, Kornell et al. (2007) found a negative correlation between monkeys' response latencies (of an arm movement to touch a screen) and high bets but found that the correlation could not account for all of the betting data.

We tested whether our monkeys might have based their bets on their decision stage latencies. First, we examined whether there was a significant difference in saccade latencies between correct and incorrect decisions (using all trials). As expected, both monkeys made correct decisions faster than incorrect decisions (for Monkey N, 218.8 ± 18.6ms vs. 236.7 ± 22.6ms; for Monkey S, 175.3 ± 15.9 vs. 186.1 ± 19.4ms, both ps < .01 with Mann–Whitney rank sum test). Next, we compared latencies across the four outcomes. If the monkeys used saccade latencies as criteria for betting, we would predict differences between correct-high and correct-low outcomes and between incorrect-high and incorrect-low outcomes. All results were negative, however (p > .05 for all, Holm–Sidak multiple comparison tests). These data are summarized in Table 2. In short, we found no evidence that the monkeys were using latencies to make bets.

**Table 2: Saccade Latencies**

Latencies (presented in milliseconds) to saccade onset during the decision stage for each monkey across each trial outcome.

| | Correct-High | Correct-Low | Incorrect-High | Incorrect-Low |
|---|---|---|---|---|
| Monkey N | 217.5 | 221.8 | 280.4 | 235.5 |
| Monkey S | 174.2 | 178.9 | 186.2 | 186.6 |

### 2.4.4 Implicit Opt-Out Behavior

The task offered an implicit opt-out option: A monkey could complete the decision stage but then fail to look at a bet target. Such aborts incurred a longer penalty (10-s time-out) than the worst outcome of betting, so as one might expect, they were rare (Figure 6). Nonetheless, they were more frequent for the difficult, short SOA trials than for the easier, long SOA trials (p < .05, ANOVA). This pattern is reminiscent of data from classic opt-out experiments (Smith et al., 1998; Hampton, 2001) and provides further evidence of metacognition in our animals.

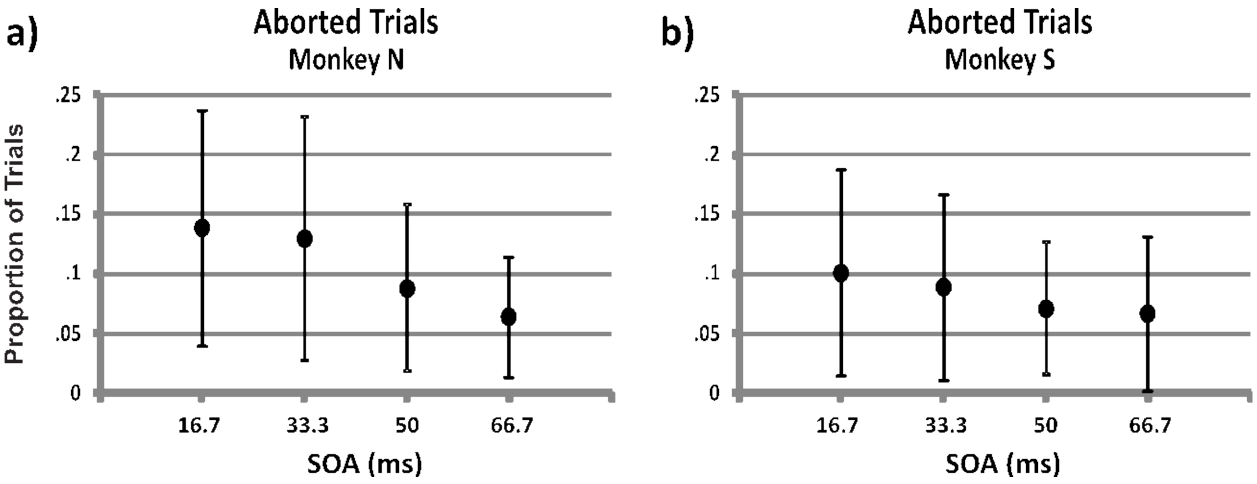


**Figure 6: Implicit opt-out behavior.**

Average proportion of trials aborted during the bet stage as a function of stimulus onset asynchrony (SOA) for (a) Monkey N and (b) Monkey S. In an analysis of variance, $p < .05$ for both panels. Error bars represent standard deviations.

### 2.4.5 Learning

We documented the learning curve of Monkey S during its first 33 days of training on the full task (Figure 7a). Its phi correlations were near zero until approximately Day 13 and then rose steadily to asymptote at around Day 23. We fit the learning curve with a sigmoid function (Figure 7a, black line, R2 = .805) and found that the inflection point occurred between Days 18 and 19. For quantification, we defined a “during” learning period that spanned 10 sessions around the inflection point, Sessions 14–23. In an ANOVA, we compared these data with the 10-session periods that immediately preceded learning (“before” phase, Sessions 4–13) and followed learning (“after” phase, Sessions 24–33). We found that, as expected, in the “before” learning period, the animal was not skilled at betting appropriately (Figure 7b). On average, the animal made high bets with probability .55 regardless of its decisions or SOA ($p = .98$). In the during learning period (Figure 7c), bets started to parallel decisions. Although the animal maintained a high-bet probability of around .60 for the more difficult (shorter SOA) trials, it increased its proportion of high bets on the easier (longer SOA) trials (p < .001). In the after learning phase (Figure 7d), the monkey attained stable, accurate performance; high bets were rare on more difficult trials but frequent on easier trials (p < .001) so that bets closely paralleled decisions.

Figures 7e– 7g summarize the trial-by-trial analysis during each learning period. Before learning, betting was not significantly different from chance (Figure 7e, phi = .027, p < .05). During learning, the monkey's bets became nonrandom, with appropriate tracking of decisions, and thus began to meet our criterion for metacognition (Figure 7f, phi = .28, p < .05). The animal's bets showed that it clearly tracked its decisions in the 10 sessions after learning (Figure 7g, phi = .68, p < .05).

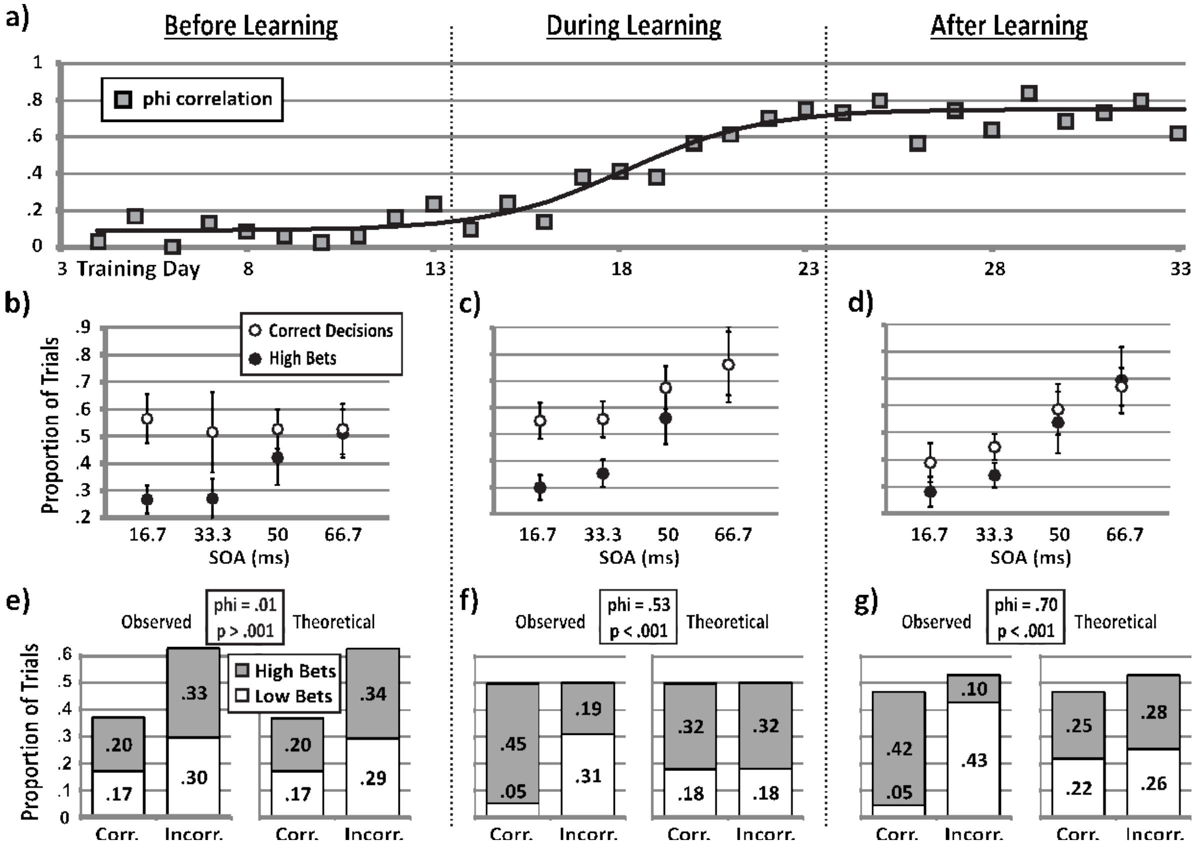


**Figure 7: Learning curves for Monkey S, divided into three stages**

(a) Phi correlation (gray squares) measures how well the bets tracked the decisions on a trial-by-trial basis on Days 4–33 of training (x-axis). A sigmoid function fit the data (black line, $R^2 = .805$, inflection point = 18.1 days). Panels b through d presents the average proportion of correct decisions (solid circles) and high bets (open circles) as a function of stimulus onset asynchrony (SOA) (b) before learning, (c) during learning, and (d) after learning to make bets based on decisions. Error bars represent standard deviations. Panels e through g present trial-by-trial analyses during each phase of learning: (e) The distribution of trial outcomes for Monkey S before learning for all trials combined. The bar graph shows the observed proportions of high bets (gray shading) and low bets (white shading) among correct (Corr.) and incorrect (Incorr.) decisions, as well as the theoretical response rates if the monkey placed bets on the basis of an external variable (the SOA) rather than tracking its decisions. Chi-square between the distributions of observed and theoretical proportions indicates that they are not significantly different ($p < .001$). The phi correlation is only .01. (f) The distribution of trial outcomes for Monkey S during learning for all trials combined. Chi-square between the distributions of observed and theoretical proportions indicates that they are significantly different ($p < .001$), with a phi correlation of .53. (g) The distribution of trial outcomes for Monkey S after learning for all trials combined. Chi-square between the distributions of observed and theoretical proportions indicates that they are significantly different ($p < .001$), with a phi correlation of .70.

### 2.4.6 Signal Detection Theory

There has been recent interest in applying signal detection theory to metacognitive task data, (Higham, 2007; Masson and Rotello, 2009). Signal detection theory (Green and Swets, 1966) assumes that an individual sets a criterion by which to detect a signal (which might be present or absent) in order to make a response (a report of presence or absence). In a retrospective monitoring task such as ours, the signal is the decision (correct or incorrect), and the response is the bet (high or low). In the nomenclature of signal detection theory, a high bet after a correct decision would be a hit, a low bet after a correct decision would be a miss, a high bet after an incorrect decision would be a false alarm, and a low bet after an incorrect decision would be a correct rejection. Applying these signal detection theory principles, we calculated d for the bet patterns of our monkeys (Clifford et al., 2008).

If the monkeys based their bets on their decisions and not on visual inputs, then d′ should remain constant across SOA. Conversely, if visual information contributed to the bets, then d′ values should increase with SOA, because additional visual information (integrated over time as SOA increases) would serve as additional signal on which to bet. Table 3 shows that d′ was, in fact, steady across SOA ($p < .05$, Pearson's correlation).

We found a difference in d′ between monkeys (~1.35 for Monkey N and ~2.35 for Monkey S), and this agrees with our previous analyses (Figures 4 and 5) that the task seemed more challenging for Monkey N. We examined how d′ changed throughout Monkey S's training (Figure 8a), plotting d′ as a function of SOA for the three learning phases described above. As would be expected, there was an overall increase in d′ with learning. For comparison, we also plotted phi correlations for each SOA during training (Figure 8b) and found the same pattern. We tested whether there was a correlation between d′ and SOA in each learning phase, which would suggest an influence of sensory input on bets. There was a significant effect of SOA on d′ before learning (p < .05, Pearson's), but this became insignificant during and after learning (p < .05).

**Table 3: d′ values**

d′ for each monkey (rows) was calculated for each stimulus onset asynchrony (columns).

| | 16.7ms | 33.3ms | 50ms | 66.7ms |
|---|---|---|---|---|
| Monkey N | 1.38 | 1.33 | 1.4 | 1.33 |
| Monkey S | 2.3 | 2.4 | 2.35 | 2.4 |

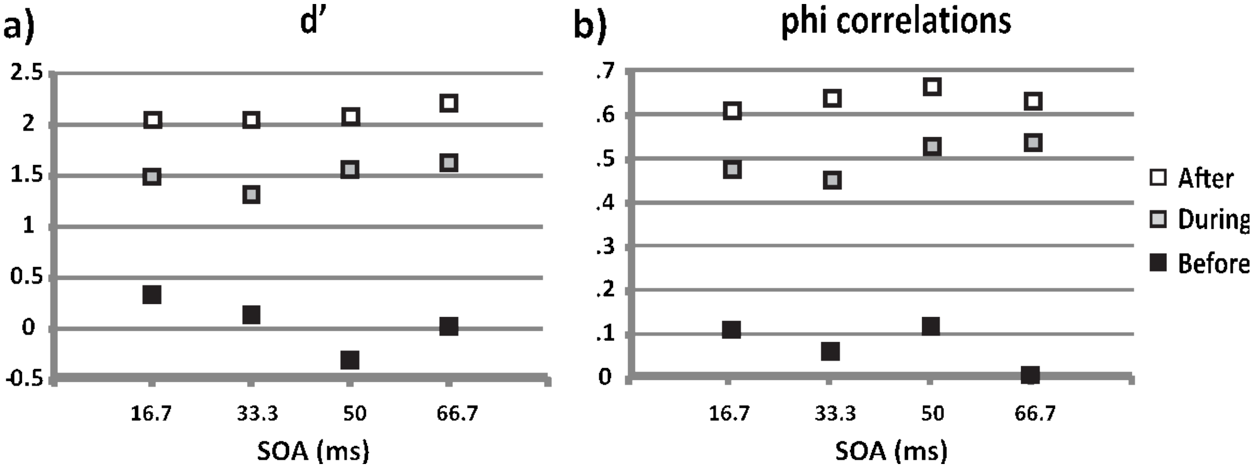


**Figure 8: Signal detection theory analysis of the data**

(a) d′ as a function of stimulus onset asynchrony (SOA) through the three learning phases for Monkey S and (b) phi correlations for the same data, for comparison.

### 2.4.7 Strategies for Performing the Task

We analyzed how well the monkeys' strategy for performing the task earned reward for them, in comparison with idealized strategies. For each SOA, we calculated the monkeys' observed rate of reward collection in units of juice drops per trial as well as the reward rates that would have resulted from three modeled strategies (Figures 9a and 9b). We used the observed proportions of correct decisions and high bets to assign bets to decisions in three different theoretical ways. First, the gray triangles show reward rates that result from random assignment, that is, if the monkeys had placed bets regardless of their decisions. Black triangles show the rates that result from optimal assignment, that is, high bets after all correct decisions and low bets after all incorrect decisions (constrained by the observed, average proportions of bets and decisions). White triangles show the worst possible reward rates, that is, the opposite betting assignment. Finally, we performed a fourth analysis that used only part of the observed data, the decision data alone, to compute the result expected for absolutely perfect betting (white squares). This analysis assigned high bets to each correct decision and low bets to each incorrect decision. We found that both monkeys adopted strategies (black circles) that were near optimal and that for Monkey S, betting was nearly perfect.

We performed the same analysis for Monkey S's learning phases (Figures 9c–9e). Before learning the task, the monkey's bets were random (Figure 9c). During the learning phase, its strategy improved (Figure 9d) and reached near optimal levels as soon as the learning phase concluded (Figure 9e). There was little change in strategy from this immediate post-learning phase to the longer, steady state period (cf. Figure 9b).

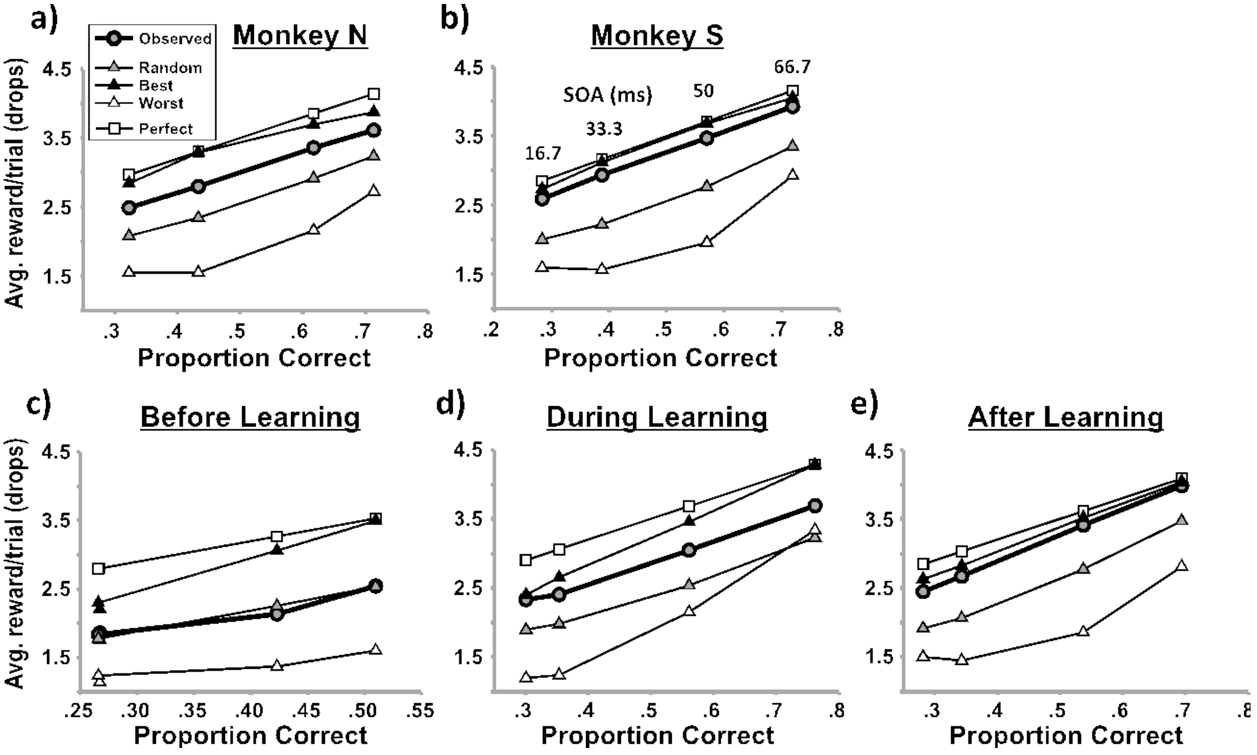


**Figure 9: Strategies for performing the task**

(a) Monkey N and (b) Monkey S and for the learning phases of Monkey S: (c) before learning, (d) during learning, and (e) after learning. In each panel, reward rate (drops of juice per trial) is plotted for each stimulus onset asynchrony (SOA) as a function of proportion of correct decisions. The observed reward rate (gray circles and bold lines) is compared with other hypothetical reward rates from various strategies, given the observed distribution of correct decisions and high bets. The alternative strategies were best possible betting (black triangles), random betting (gray triangles), and worst possible betting (white triangles). For comparison, white squares show reward rates that would have occurred for perfect betting (given the monkeys' observed distribution of correct decisions but not the observed high bets).

## 2.5 DISCUSSION

We designed a visual oculomotor task that required monkeys to retrospectively monitor their own decisions. Two monkeys learned to perform the task advantageously and then maintained that level of performance. Behavior was best explained by metacognitive processes with little or no reliance on sensory (SOA) or motor (reaction time) cues. The results demonstrate that monkeys can monitor their own decisions in a highly reduced task that is suitable for imaging studies and electrophysiology.

Our experiment used a betting paradigm, but this is not the only possible approach. A recent single-neuron study (Kiani and Shadlen, 2009) utilized an opt-out paradigm. The betting approach has some advantages over the opt-out approach, primarily the constant pairing of an explicit decision followed by an explicit bet on the accuracy of that decision. This permits a direct comparison of behavioral (and eventually neuronal) data from both stages of the task: decisions and bets. In the terminology of Nelson and Narens (1990), we may compare an object-level cognitive process (the decision) with its associated meta-level cognitive process (the bet, which results from monitoring the decision). The animals complete the same sequence of events in every trial, from initial fixation through final bet selection, facilitating analysis of neuronal activity related to decisions and bets, for example, allowing for direct identification of firing rates related to metacognitive signals. The neuronal correlates of such signals should predict trial-by-trial bets after identical target–mask stimuli and identical decisions. And if one is interested in opt-out behavior, the betting task is still useful, because it contains an implicit opt-out component. This is not to say that betting tasks are always best for studying metacognition. Hampton (2001), for example, used an opt-out task elegantly to show that monkeys can monitor the integrity of their visual object memory. Betting and opt-out tasks are different tools, and either might be best for answering a particular experimental question.

Our task design ensured that monkeys earned maximal reward only by monitoring decisions to make bets. It was theoretically possible, however, for monkeys to earn rewards with other strategies. Before learning, we did see evidence of nonmetacognitive strategies (Figures 8a, 8b, and 9c), but they were superseded by decision-based betting after successful training, as indicated by multiple analytical methods. During steady-state, trained behavior it was unlikely that the monkeys based their bets on differences in the visual stimuli across trials, because the visual stimuli were the same for each trial; they differed only in location of the masked target and SOA. Keeping track of location would not help, because it contained no information (target location varied randomly by trial). Visual information provided by the SOA could help, as correct trials occurred more often with long SOAs. However, our trial-by-trial analysis indicated that monkeys did not exploit the information. A final alternative explanation might be that the monkeys failed to see the target on shorter SOA trials and used this to their advantage. If a monkey failed to see the target, it could abort the trial to start a new one immediately, or it could continue the trial and make a randomly directed saccade. In the latter scenario, the monkey might know that the saccade was random because it knew that it did not see the target. The best utility of this knowledge would be to switch to a new strategy of betting only low (which can be shown with simple calculations to yield optimal reward for random saccades). Such a scenario is unlikely for two reasons. First, it is not parsimonious. Presuming that the monkeys adopted a new betting strategy when they failed to see a target is more complicated than simply presuming that they aborted such trials. Second, the scenario has no empirical support. Using a special strategy to obtain more reward after failing to see a target (more likely at shorter SOAs) should lead to an anomalous drop in high bets and rise in reward rates at shorter SOAs. We found no hint of either deviation (Figures 3b, 3d, 9a, and 9b). Taken together, all this evidence supports our original hypothesis that the task encouraged covert monitoring of decisions and betting on the basis of those decisions on individual trials.

We proposed an operational definition of metacognition because monkeys cannot use language to relate their cognitive experience. Principled, accepted criteria are needed for inferring their cognition through their behavior. But the definition that we established is not entirely new. Son and Kornell (2005) calculated phi correlations for two monkeys performing a betting paradigm directly analogous to the task we used. The authors did not state that phi correlation values significantly above zero counted as metacognitive behavior, but they indicated as much by stating that the monkeys were “able to make accurate confidence judgments” (p. 311). Our main contribution was to explicitly define the criteria for metacognitive behavior.

What exactly was the cognitive operation that linked the animals’ decisions to their subsequent wagers? Heightened arousal or vigilance seems unlikely to play a role, because monkeys had to fully complete every trial. Placing a low bet required as much effort as placing a high bet. A more convincing explanation is to invoke levels of confidence, which may be considered a weak form of metacognition. High confidence would encode that the decision was correct without necessarily conveying details about the decision. A stronger metacognitive signal would encode decision-related details such as target location. Many studies have examined how metacognition and confidence are related (Koriat et al., 1980; Brewer and Sampaio, 2006), but few experiments have studied how confidence is distinct from metacognition. One report compared human participants’ standardized measures of self-confidence with their standardized measures of metacognitive awareness (Kleitman and Stankov, 2007). The authors of that study concluded that confidence and metacognition are separate processes but found a correlation between the two.

For the sake of simplicity, we have been anthropomorphizing monkey behavior. Doing so facilitates discussion but should not be taken literally. Although monkeys can be trained to monitor their decisions and perform subsequent behaviors based on them, it is unknown to what extent they do this in everyday behavior. Their metacognition may be entirely implicit, below any level of awareness, such as the blindsight process by which monkeys correctly identify stimuli in a blind hemifield (Cowey and Stoerig, 1995). It may well be that any animal capable of making decisions also maintains an internal record of those decisions for future use. Tests of metacognition in animals phylogenetically distinct from primates or dolphins (Smith et al., 1995) have yielded mixed results thus far. Pigeons seem unable to utilize an opt-out response in a manner reflecting metacognition (Sutton and Shettleworth, 2008; Roberts et al., 2009), but such data should be interpreted cautiously (as discussed by Shields et al. 2005, e.g.). Rodents have utilized an opt-out response effectively (Foote and Crystal, 2007), and they may monitor their confidence levels in a task where a choice is made but a reward is delayed- they tend to abort trials before the potential for reward after making incorrect decisions (Kepecs et al., 2008). We suggest a point of comparison in the realm of motor behavior. As far as is known, every animal that moves also issues internal signals about its movements, called corollary discharge, for use in multiple functions such as analysis of sensory input (Crapse and Sommer, 2008a). One could think of metacognition as involving a corollary discharge; when brain networks accomplish a decision, they may issue an internal record of that decision to the rest of the brain for multiple potential purposes.

In the specific context of monkeys performing our task, the frontal eye field seems to contribute to the target-selection decision stage (Thompson and Schall, 1999). It may send a corollary discharge of the decision-related signals to higher areas such as dorsolateral prefrontal cortex, which use the information to wager optimally in the bet stage of the task. Such a hypothesis is testable using the present task and neurophysiological methods of recording, stimulation, and inactivation, or using imaging methods combined with analyses to search for correlated activations in separate, defined areas (Friston et al., 2003; Goebel et al., 2003). We hope future studies will cast light on the neurobiological bases of metacognitive functions in a similar way that studies beginning 40 years ago (Fuster and Alexander, 1971; Kubota and Niki, 1971), and refined 20 years ago with streamlined tasks (Funahashi et al., 1989), began to elucidate the neuronal bases of working memory.

# 3.0 NEURONAL CORRELATES OF METACOGNITION IN PRIMATE FRONTAL CORTEX

## 3.1 ABSTRACT

Humans exhibit metacognition, the ability to monitor and control their cognitive operations. Recent work has provided evidence that non-human primates are capable of metacognition as well. Our hypothesis was that neuronal correlates of metacognition would be found in monkey frontal cortex, within or near regions known to contribute to primary cognitive operations such as decision-making. We recorded single neuron activity in three frontal cortical areas: frontal eye field (FEF), lateral prefrontal cortex (PFC), and supplementary eye field (SEF). Monkeys performed a metacognitive visual oculomotor task (Middlebrooks and Sommer, 2011) in which they made a decision and reported it with a specific vector of saccade.  Unlike conventional tasks, however, they received no immediate reward or other feedback. Instead, they had to internally monitor their decision and place a bet on whether it was correct or not. Their reward took the form of payoff for the bet. We found signals that correlated with the decision (regardless of bet) and the bet (regardless of decision) in all three brain areas. In contrast, sustained activity that linked decisions to appropriate bets -- putative metacognitive signals -- were found almost exclusively in the SEF. Such signals often appeared at the earliest possible time that the monkey could have made its decision, and before the decision was reported with a saccade. These results

offer a survey of neuronal correlates of metacognition in frontal cortex and demonstrate the relative importance of the SEF in linking sequential cognitive functions over short periods of time.

## 3.2 INTRODUCTION

Not only do we perform cognitive functions, but also we monitor and control them. For example, after creating a presentation, we may think about the way we organized its content. Likewise, if the presentation is not ready yet, we can estimate how much more effort will be required to finish it. The process of evaluating cognitive processes and preparing for the future is called metacognition.

Metacognition has been studied for decades in the psychological literature. To our knowledge the term “metacognition” was coined in the late 1970s (Flavell, 1976), and psychological frameworks have been developed to study and understand metacognitive processes (Flavell, 1979; Nelson and Narens, 1990). The early explanatory frameworks were primarily to understand human metacognition, and most neuroscience research into metacognition has focused on humans. Patients with lesions of medial and lateral frontal cortex show impaired metacognitive skills (Schnyer et al., 2004; Pannu et al., 2005a), and transcranial magnetic stimulation (TMS) over dorsolateral prefrontal cortex impairs subjects’ judgments about the visibility of visual stimuli without affecting their ability to correctly discriminate the stimuli (Rounis et al., 2010). A host of functional magnetic resonance imaging (fMRI) studies have measured blood-oxygen level dependent (BOLD) responses while subjects performed metacognitive tasks. Common brain regions shown to have differential BOLD responses include dorsolateral prefrontal cortex (Henson et al., 2000; Kikyo et al., 2002; Chua et al., 2004), medial prefrontal cortex (Schnyer et al., 2005; Chua et al., 2006; Chua et al., 2009; Kim and Cabeza, 2009), and cingulate cortices (Kikyo et al., 2002; Chua et al., 2006; Kim and Cabeza, 2009).

Little is known about how the brain encodes metacognitive processes at the single neuron level. An animal model would facilitate such research, but until recently the extent to which animals engage in metacognition has been unknown. A growing number of behavioral studies have provided evidence for some degree of metacognition in rats (Foote and Crystal, 2007), dolphins (Smith et al., 1995), rhesus monkeys (Smith et al., 1998; Hampton, 2001; Beran et al., 2006; Middlebrooks and Sommer, 2011), and orangutans (Suda-King, 2008). Specifically, such animals may be able to monitor their uncertainty. When offered the chance to take a test or decline it, they opt-out more often on relatively difficult trials, ensuring themselves a small reward rather than risking no reward if they take the test and fail it.

Non-human primates have metacognitive abilities that extend beyond the ability to abort a task due to uncertainty. In experiments where one of a few opaque tubes was baited with food either in full view of an animal or occluded from view, gorillas, chimpanzees, bonobos, orangutans (Call, 2010), and rhesus monkeys (Hampton et al., 2004) sought more information on occluded trials by looking more frequently into the tubes before making their choice. The results indicated that factors related to past mental state (e.g. their memory of the baiting or their uncertainty) play a role in later decisions. Rhesus monkeys can be trained to accurately report, via a post-decision bet, whether a past decision was correct or incorrect (Shields et al., 2005; Kornell et al., 2007). We recently designed a streamlined version of such a "retrospective monitoring" task, suitable for neurophysiology, that involved only visual stimuli and saccadic eye movement reports, and we reported evidence that monkeys can monitor their own decisions (Middlebrooks and Sommer, 2011).

To our knowledge only two studies have recorded single neuron activity related to possible metacognitive processing (see Chapter 1 for details). One used an opt-out task as described above, and reported that monkey lateral intraparietal cortex neuron activity correlated with the animals' choices to abort a trial (Kiani and Shadlen, 2009). The other study used a variation of an opt-out task, and found that rats would not wait through a delay for possible reward on trials with a low likelihood that reward would actually be delivered (Kepecs et al., 2008). Orbitofrontal cortex neuron activity correlated with the pattern of rats' "confidence".

The goal of our study was to determine the neuronal correlates of metacognition in primate frontal cortex. Instead of using an opt-out paradigm, we used a task in which a monkey had to make a decision and then place a bet on the correctness of that decision.  Appropriate wagers required retrospective monitoring, a metacognitive process. In opt-out tasks, critical trials are truncated. In our paradigm, every trial contained the same sequence of task events. Hence we could analyze the entirety of every trial for neuronal correlates of decision-making, wagering, and retrospective monitoring. We focused on three frontal areas known to be involved in visual-oculomotor behavior. The rich history of neurophysiological studies on the monkey visual-saccadic system provides a solid foundation for performing experiments on metacognition (for review see Basso 1998).

All three areas that we studied are known to have neuronal activity related to vision, saccades, and cognitive processes. The frontal eye field (FEF), located in the anterior bank of the arcuate sulcus (Bruce et al., 1985), contains neurons that respond to visual stimuli, it is important for generating and controlling saccades, and it is involved in higher cognitive functions (see

Chapter 1 for details). Our specific hypothesis concerning FEF was that its neurons would have activity correlated solely with making the decision in our task, but not with placing the bet or linking the decision to the bet.

The second region we studied is the lateral prefrontal cortex (PFC), in and around the principal sulcus just anterior to the FEF. Similar to FEF, the PFC contains neurons that respond to visual stimuli and have delay- and saccade-related activity, and the PFC has been implicated in a broad range of higher cognitive functions (see Chapter 1 for details). Our specific hypothesis regarding PFC was that its neuronal activity would not necessarily be related to making the visual-oculomotor decision in our task (the putative role of FEF), but would be related to linking the decision to the bet and placing the bet.

The third region we studied was the supplementary eye field (SEF), located in the rostral region of supplementary motor area in dorsomedial frontal cortex. Like the FEF and PFC, the SEF has myriad types of visual and saccade related activity (Olson and Gettner, 1995; Russo and Bruce, 1996). Among the cognitive functions SEF seems to be involved in (see Chapter 1 for details), its activity during performance monitoring tasks (Stuphorn et al., 2000) suggests it may have a metacognitive role as well. We therefore hypothesized that SEF neurons would have activity correlated with performance in the Decision Stage of the task that could help to link the performance to the eventual bet.

We analyzed neuronal activity from FEF, PFC, and SEF with respect to three main functions of the task: 1) activity related to the decision (regardless of bet), 2) activity related to the bet (regardless of decision), and 3) the putative metacognitive activity that would link the decision to the appropriate bet. We found that all three areas encoded decisions, and likewise all three areas encoded bets. Activity in the SEF, however, provided the strongest link between decision and bet, suggesting that of the three areas, it is the most involved in metacognition.

.

## 3.3 METHODS

### 3.3.1 Surgery

Two male rhesus monkeys (Maccaca mulatta; monkeys N, 6.6 lbs., and S, 6.0 lbs.) were anesthetized and surgically prepared for neuronal recordings and eye position measurements. Using aseptic procedures, ceramic screws and an acrylic implant were affixed to the skull. Recording chambers and a head-restraint socket (Crist Instruments, Hagerstown, MD) were embedded in the implant. Chambers were positioned over FEF/PFC (1 chamber with access to both regions) and SEF using stereotaxic coordinates (FEF/PFC: A25, L20; SEF: A25, midline). In the same surgery, we implanted scleral search coils (Judge et al., 1980). Animals recovered for 1–2 wk before training resumed. Procedures were approved by and conducted under the auspices of the University of Pittsburgh Institutional Animal Care and Use Committee and were in compliance with the guidelines set forth in the United States Public Health Service Guide for the Care and Use of Laboratory Animals.

### 3.3.2 Tasks

*RF mapping tasks*

To determine appropriate target locations to use in the metacognition task (described below), we initially characterized the receptive field of each isolated neuron by using simple visual oculomotor tasks while observing the real-time spike rasters and spike density functions. First, we had the monkey make visually guided saccades to targets in eight different directions (the cardinal directions and diagonals). After the neuron's preferred direction was established, we had the monkey perform visually guided saccades of varying amplitudes in that direction. If necessary, the two tasks (direction task and amplitude task) were repeated to refine the spatial location of a neuron's receptive field (for details, see Sommer & Wurtz 2004). Once the receptive field center was located, we typically had the monkey make memory guided saccades to that location, to distinguish visual-, delay-, and saccade-related activity (Mays and Sparks, 1980; Hikosaka and Wurtz, 1983). The memory guided saccade task involved central fixation, the appearance of a 50ms duration target followed by a 500-1000ms randomized delay, and the requirement to make a saccade to the remembered target location for reward.

*Metacognition task*

The task was described previously in detail (Middlebrooks and Sommer, 2011). Each trial consisted of two stages: a *Decision Stage* and a *Bet Stage* (Figure 10). In the Decision Stage the animal was required to detect and report the location of a peripheral visual target (Thompson and Schall, 1999), and in the Bet Stage to report, via a wager, whether a correct or incorrect decision was made in the Decision Stage (Shields et al., 2005). Appropriate betting, and thus optimal reward delivery, required the animal to maintain an internal representation of its decision. It is the maintenance of that decision signal, and its use for betting, that we refer to as metacognition. To obtain reward on any trial, completion of both the Decision and Bet Stages was required.

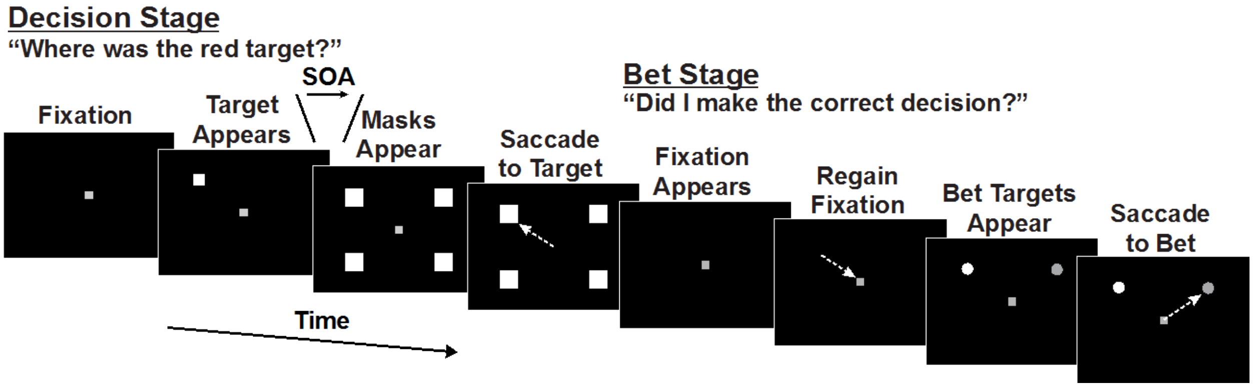


**Figure 10: Metacognition task**

Each trial consisted of two stages, a Decision Stage and a Bet Stage. In the Decision Stage, monkeys foveated a fixation spot. A target appeared at one of four locations in the periphery, and after a variable time (SOA), masks appeared at the four locations. A correct decision was made (shown) if a saccade went to the location of the target. A saccade to any other location was an incorrect decision. In the Bet Stage, monkeys foveated a new fixation spot, then the two bet targets appeared in the periphery. A bet was made when a saccade went to one of the two targets, completing a single trial.

Decision Stage The monkey fixated a center spot for 500-800ms (randomized by trial; Figure 10, left side). Then a dim target appeared in one of four possible locations (also randomized). The locations were constant throughout a session, but often varied between sessions; their eccentricities could range from 5-25 deg. and their directions, relative to the horizontal meridian, could range from 0-60 deg. in angle. For each neuron, these parameters were chosen so that, when possible, at least one target location would be in the hotspot of the receptive field as determined by our mapping procedure. The locations were mirror symmetric across the vertical meridian. After the target appeared, identical mask stimuli (white squares) appeared at all four locations. The interval between target appearance and masks appearance, the *stimulus onset asynchrony* (SOA), was randomized by trial - 16.7, 33, 50, or 66.7ms – thus varying the difficulty of the trial. After the masks appeared, a delay followed (random 500-1000ms) during which the monkey maintained fixation while the masks remained visible. Then, the fixation spot was extinguished, cueing the monkey to report its decision by making a saccade to the perceived target location within 1000ms. The monkey received no performance feedback until after the Bet Stage of the task, but the computer tracked whether the decision was correct (saccade landed in an electronic window around the target location) or incorrect (saccade landed anywhere else). If at any time during the Decision Stage the monkey broke fixation, made a saccade before cued to go, or failed to make a saccade, the trial was aborted (and repeated again later) and the next trial immediately began. In all trials, the saccade ended the Decision Stage and started the Bet Stage.

Bet Stage 350ms after the decision saccade that concluded the Decision Stage, a new fixation spot appeared and the monkey foveated it for 500-800ms (Figure 10, right side). Then two bet targets appeared: a red “high-bet” target and a green “low-bet” target (for monkey N; this color assignment was reversed for monkey S). Within a session the two locations were constant, but the appearance of high-bet or low-bet target varied randomly between the two locations. One location was in the hotspot of the receptive field and the other was at the mirror symmetric location in the other visual hemifield. A monkey reported its bet by making a saccade to one of the targets, then received reward or timeout as described below, and the trial ended. A monkey would optimize its reward if it bet high after a correct decision and low after an incorrect decision. If, during the Bet Stage, the monkey broke fixation or made a saccade to a non-bet-target location, the trial was aborted and a brief timeout ensued before a new trial began.

Reward The amount of reward delivered after each trial was based on how appropriate the bets were relative to the decisions. If the monkey made a correct decision and bet high, it earned the maximum reward: 5 drops of water. If the monkey made an incorrect decision and bet high, it received no reward and a 5-second timeout. Betting low earned a sure but minimal reward: 3 drops after a correct decision and 2 after an incorrect decision. The reward schedule was based on previous studies (Kornell et al., 2007; Persaud et al., 2007) and fine-tuned to elicit the best performance in our monkeys.

### 3.3.3 Neuronal recordings

At the start of a recording session, a single tungsten electrode (300 kΩ to 1MΩ impedance @ 1 kHz; FHC, Bowdoinham, ME) was lowered through a guide tube (23 gauge) using a custom microdrive system (LSR 2008; details at ftp://lsr-ftp.nei.nih.gov/lsr/StepperDrive/). A plastic grid with 1 X 1 mm hole spacing (Crist et al. 1988; Crist Instruments, Hagerstown, MD) was attached inside the recording chamber. The FEF was confirmed with microstimulation (Bruce and Goldberg, 1985) by evoking saccades at low current threshold (< 50 μA). The PFC was recorded from the same chamber as FEF. PFC recordings included locations a few mm anterior to identified FEF, in areas ventral, dorsal, and within the principal sulcus (identified by the isolation of neurons at lower depths than locations dorsal or ventral). The SEF was identified by moderate-current microstimulation (typically 50-100μA) that evoked or delayed saccades (Schlag and Schlag-Rey, 1987; Russo and Bruce, 1996). Standard extracellular recording techniques were used to isolate action potentials from single neurons (FEF; Sommer and Wurtz 2000). Electrode stability was implemented via continuous on-line waveform inspection. All neuronal recording data were collected using the REX real-time system (Hays et al., 1982) and analyzed using MATLAB (R20010a, The MathWorks, Inc.).

### 3.3.4 Analysis

We defined multiple epochs throughout the task, and measured and analyzed the average firing rates within these epochs. *Baseline* was the 300ms epoch before Decision Stage target onset. During the Decision Stage, we analyzed a *visual-1* epoch 100-300ms after target onset, a *delay* epoch 200ms before fixation offset, a *presaccadic-1* epoch 50ms before saccade onset, and a *postsaccadic* epoch 100 – 300ms after saccade onset. After the Decision Stage and before the Bet Stage, we defined an *interstage* epoch as 400ms surrounding the time the animal regained fixation to initiate the Bet Stage, from 200ms before until 200ms after that point. During the Bet Stage, we analyzed a *visual-2* epoch 50-150ms after target onset, a *presaccadic-2* epoch 50ms before saccade onset, a *reward anticipation* epoch 250ms before reward delivery, and a *reward* epoch 50-250ms after reward delivery.

We performed two types of population analyses. In one analysis we included the entire population of recorded neurons. In a second analysis, we focused on only the neurons that were significantly modulated within particular epochs. A neuron was deemed significantly active within a given epoch if its average firing rate within the epoch on all correct trials (collapsed across high and low bets) was above its baseline firing rate as determined by paired t-tests ($p < .05$ criterion). Modulations below baseline were exceedingly rare and such neurons were excluded from analysis.

To analyze decision-related activity, the average firing rate within each epoch was compared between all correct trials and all incorrect trials (regardless of bets). For single neuron analysis, comparisons were made using two-sample t-tests and were considered significant at levels of $p < .05$. For population analyses, comparisons were made using paired t-tests and were considered significant at levels of $p < .05$. Analysis of Bet-related activity was analogous, except we compared average firing rates in epochs between all high-bet trials and all low-bet trials (regardless of decisions).

To analyze metacognitive-related activity, the aim was to compare trials in which decisions were identical, but subsequent bets (our observables of the monkey's internal state) were different. We compared the average firing rates in each epoch between Correct-High trials (correct decisions followed by high bets) and Correct-Low trials (correct decisions – low bets), or between Incorrect-High trials (incorrect decisions – high bets) and Incorrect-Low trials (incorrect decisions – low bets). For single neuron analysis, one-way ANOVAs were first calculated between all four trial conditions. If significant at $p < .05$, multiple comparisons (Tukey-Kramer tests) were calculated between individual conditions and considered significant at $p < .05$. For population analysis, paired t-tests were calculated between trial outcomes at $p < .05$.

## 3.4 RESULTS

### 3.4.1 Behavior

We previously reported a detailed analysis on the monkeys' behaviors during sessions prior to neuronal recordings (Middlebrooks and Sommer, 2011). Here we report behavioral results from the neuronal recording sessions included in this report (Monkey N: total recording sessions = 150; Monkey S total recording sessions = 182). As expected, each monkey performed better in the Decision Stage, and placed more high bets in the Bet Stage, as a function of longer SOA (Figure 11, one-way ANOVAs, each $p < .001$).

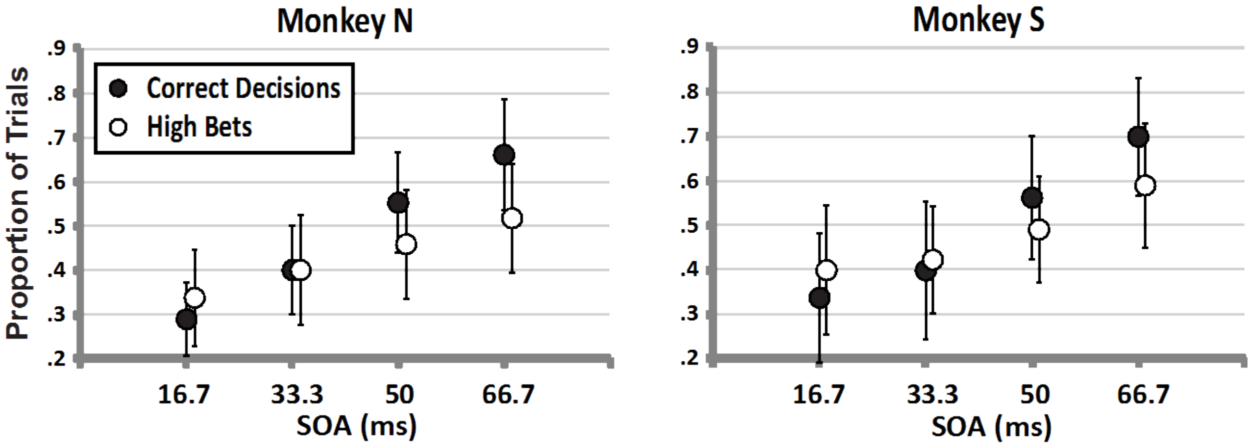


**Figure 11: Average decisions and bets**

(a) Overall proportion of correct decisions (black circles) and high bets (white circles) made by Monkey N plotted as a function of each of the four stimulus onset asynchronies (SOAs) over 150 sessions. (b) Overall proportion of correct decisions (black circles) and high bets (white circles) made by Monkey S plotted as a function of each of the four stimulus onset asynchronies (SOAs) over 182 sessions. Error bars represent standard deviations.

Average data, however (as in Figure 11), do not demonstrate a significant link between individual decisions and bets. Our operational definition of metacognition was that the monkeys' bets had to significantly track their decisions on a trial-by-trial basis, regardless of the SOA. This was critical because the SOA alone provides information that could guide betting; in principle, monkeys could ignore their decisions altogether and just bet high more often if they sensed that the masks appeared later or the task seemed easier. In practice, however, the monkeys used a metacognitive strategy. We found that, for each SOA, both monkeys made bets that correlated appropriately with their decisions ($\chi^2$ test between observed vs. randomly shuffled decision-bet distributions, $p < .001$ for each SOA and each monkey; see Middlebrooks & Sommer 2011 for details). To quantify how effectively each monkey monitored its decisions at each SOA, we calculated phi correlations (Kornell et al., 2007), the $\chi^2$ statistics normalized for number of trials. A phi correlation of 0 indicates random association between decisions and bets, while a phi correlation of 1 indicates perfect association between decisions and bets. Both monkeys' phi correlations were significantly above zero for each SOA (Table 4) and did not vary with SOA (*p* > .05 for both monkeys, one-way ANOVA).

**Table 4: Phi correlations**

Phi correlations for each monkey (rows) calculated for each SOA and averaged over all SOAs (columns). All values were significantly greater than zero (p < .001).

| SOA | 16.7ms | 33.3ms | 50ms | 66.7ms | Mean (SD) |
|---|---|---|---|---|---|
| Monkey N | .45 | .45 | .47 | .42 | .45 (.02) |
| Monkey S | .57 | .58 | .56 | .56 | .57 (.01) |

We also tested whether motor-related cues played a role in the betting. It is possible that monkeys could detect their saccade latencies during the Decision Stage and use this information to help place their bets. For saccade latency detection to be useful, monkeys would need to discriminate between Correct-High vs. Correct-Low trial latencies, and between Incorrect-High and Incorrect-Low trial latencies. Neither pair of latency distributions differed significantly, however (Table 5, *p* > .05 for all, Holm-Sidak multiple comparison tests), similar to our previous results (Table 2 in Chapter 2). It is therefore highly unlikely that the monkeys could have made the necessary discriminations.

**Table 5: Saccade latencies**

Latencies to saccade onset during the Decision Stage for each monkey, across each trial outcome

| | Correct-High | Correct-Low | Incorrect-High | Incorrect-Low |
|---|---|---|---|---|
| Monkey N | 222.3 | 221.7 | 237.8 | 231.9 |
| Monkey S | 167.4 | 163.6 | 174.2 | 177.5 |

In summary, our behavioral analyses indicate that the monkeys performed the task during our neuronal recordings by maintaining information about each decision on a trial-by-trial basis and using the information to place their bet, a metacognitive strategy. The animals seemed to rely negligibly (if at all) on more trivial strategies based on visual or motor cues or a sense of task difficulty.

### 3.4.2 Single neuron recordings

We recorded 87 neurons in FEF (35 Monkey N: 35, Monkey S: 52), 112 in PFC (N: 54, S: 58), and 133 in SEF (N: 61, S: 72). As an overview of how neurons responded during the metacognition task, typical single neuron data from each cortical region are shown in Figure 12, aligned at various task-related time points. Neurons in all three areas were highly modulated by the various events in the task.

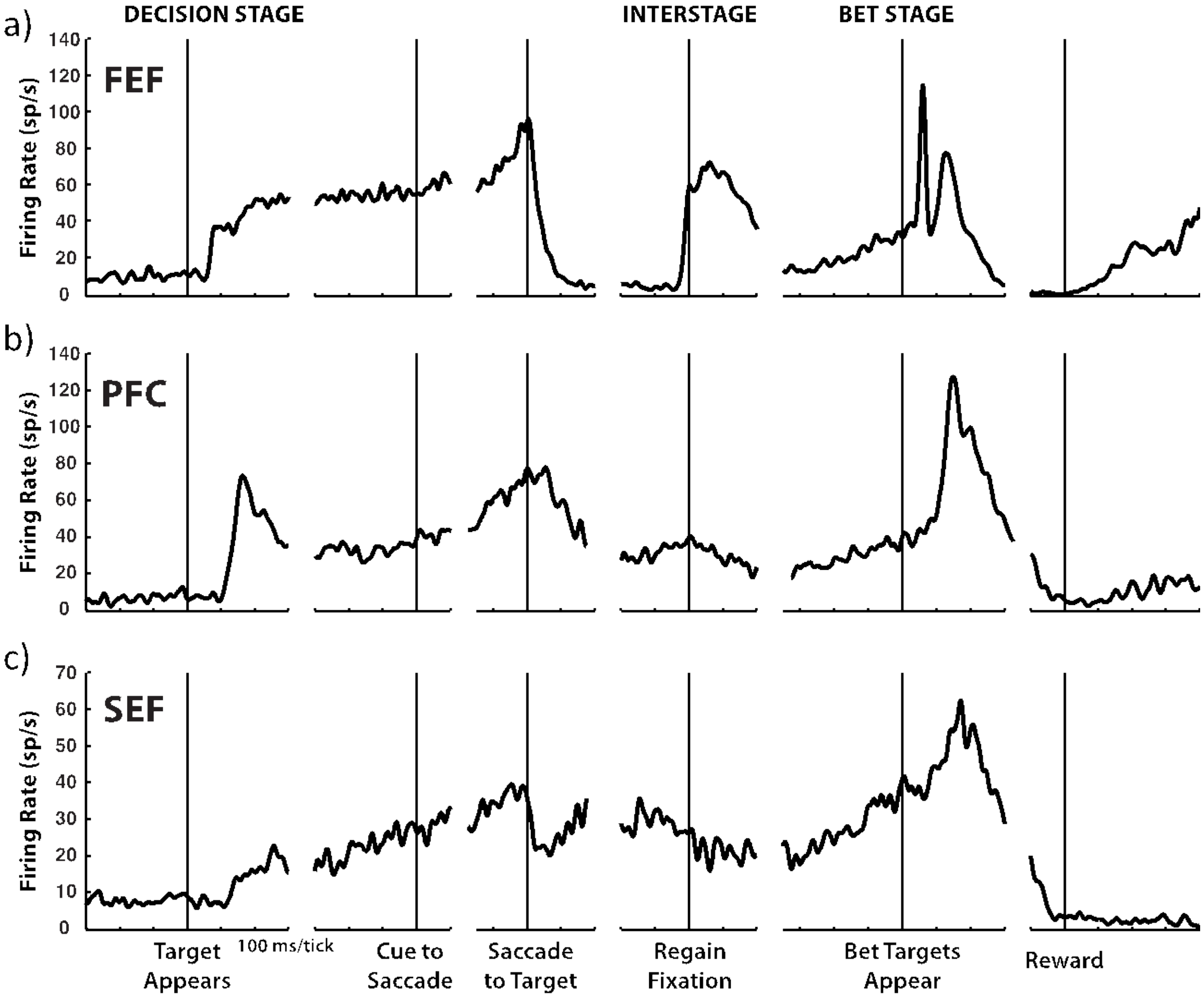


**Figure 12: Single neuron task related activity**

(a) FEF, (b) PFC, and (c) SEF. In each example, neuronal firing rates for all trial outcomes collapsed (solid line) are aligned to various times throughout the whole trial.

### *Decision-related neuronal activity*

To test whether neurons encoded the decision, we compared all correct trials with all incorrect trials, regardless of subsequent bets (i.e. trials from high and low bets pooled).

<u>Sensory-related activity comparison:</u> For the time period prior to the monkeys' decision saccade, we focused on neural activity related to the visual stimuli presented. We analyzed trials in which

the target appeared in the hemifield contralateral to the neuron's location in the brain, because for FEF, visual receptive fields are typically contralateralized (Bruce and Goldberg, 1985). Contralateral biases are common in SEF and PFC, too (Funahashi et al., 1989; Schall, 1991), and we wanted to analyze data from all three areas in the same way to provide a fair comparison. Although targets were always contralateral in these analyses, it should be noted that the subsequent decision saccades could be directed either into the contralateral hemifield (on correct trials) or into the ipsilateral hemifield (on incorrect trials). Firing rates in correct and incorrect trials were compared for the visual-1 and delay epochs (earlier and later, respectively, after target onset, but both before cue to respond; see Methods).

Single neuron examples are shown for FEF (Figure 13a), PFC (Figure 13b), and SEF (Figure 13c). For each neuron the left panel shows correct trials (solid black) and incorrect trials (dashed black) aligned on target onset, and in the right panel the same trials are aligned on fixation spot offset (the cue to make the decision saccade). Each example neuron was active during the visual period and the delay period, and each neuron was more active on correct trials than incorrect trials in both epochs (t-test, $p < .05$). We performed the same analyses on the entire population of neurons recorded in the three cortical regions (Figure 14a-c), and found that each region as a population encoded the upcoming decision during both the visual and delay periods (Table 6). This confirms the FEF results of Thompson and Schall (1999), who used a nearly equivalent target-mask decision paradigm, and extends those results to show that visual and delay activity in the PFC and SEF represent the decision as well.

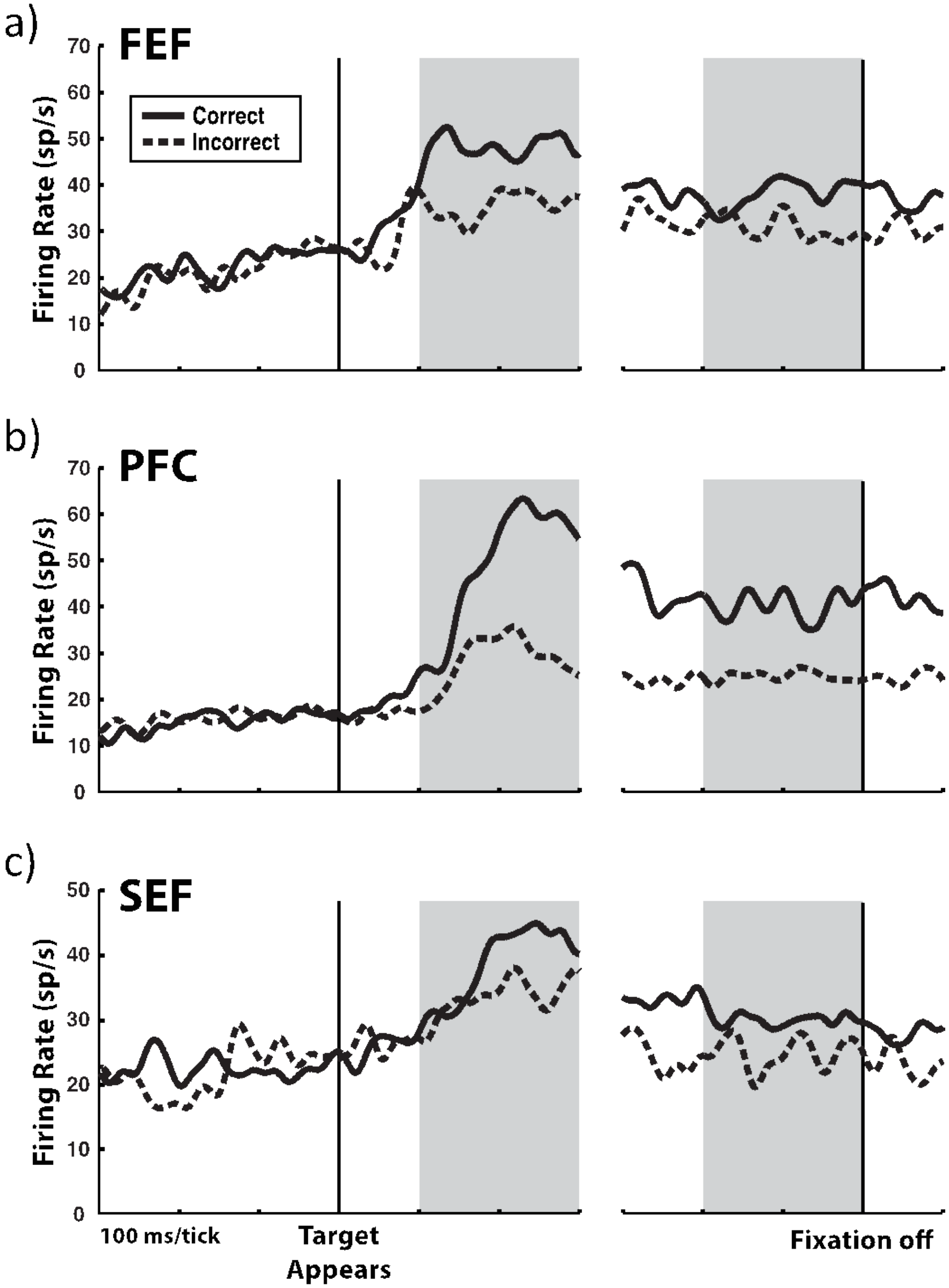


**Figure 13: Single neuron examples of decision-related activity**

(a) FEF, (b) PFC, and (c) SEF. In each example, neuronal firing rates for all correct trials (solid line) and all incorrect trials (dashed line) are aligned to Decision Stage target onset (left panels) and fixation offset (right panels). Grey shading indicates analyzed epochs: visual period (left panel) and delay period (right panel). Each of these examples was more active during correct decisions than incorrect decsisions in both the visual and delay periods.

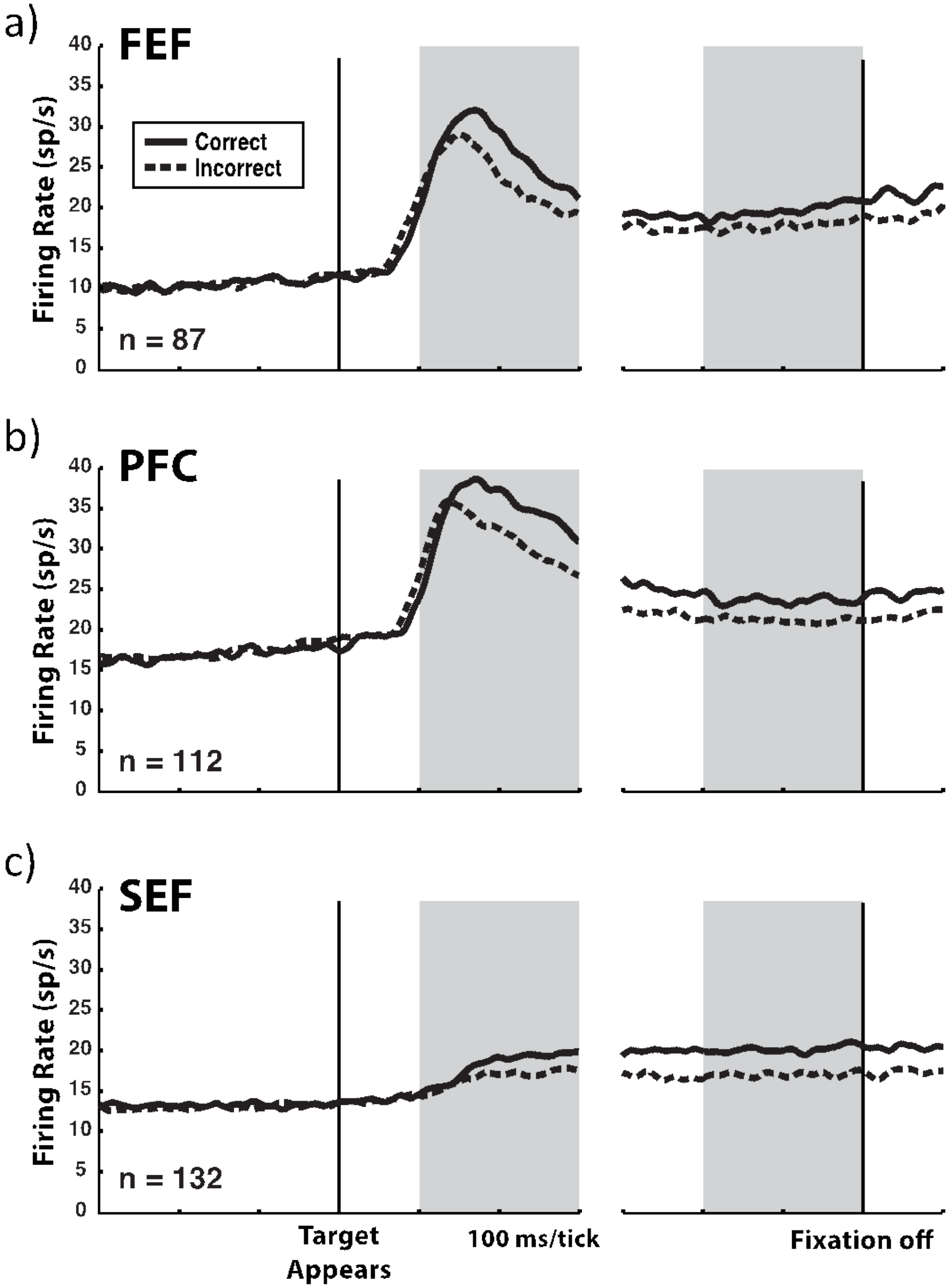


**Figure 14: Population decision-related neuronal activity**

(a) FEF, (b) PFC, and (c) SEF. For each cortical area, averaged neuronal firing rates for all correct trials (solid line) and all incorrect trials (dashed line) are aligned to Decision Stage target onset (left panels) and fixation offset (right panels). Grey shading indicates analyzed epochs: visual period (left panel) and delay period (right panel). Each population was more active during correct decisions than incorrect decisions in both the visual and delay periods.

**Table 6: Decision-related activity: population**

Firing rates during Decision Stage epochs. For each cortical region, all correct vs. all incorrect firing rates (spikes/sec) are shown with standard errors in parentheses, paired t-test p-values underneath. Asterisks represent significant difference between correct and incorrect.

| FEF | Baseline | Visual-1 | Delay | Presaccadic-1 | Postsaccadic |
|---|---|---|---|---|---|
| Correct | 10.4 (.9) | 26.8 (2.0) | 19.5 (1.6) | 29.3 (2.6) | 20.2 (2.4) |
| Incorrect | 10.2 (.9) | 23.8 (1.7) | 17.6 (1.4) | 30.5 (2.7) | 20.4 (2.5) |
| p | .33 | < .001** | .02** | .16 | .75 |
| PFC | | | | | |
| Correct | 16.0 (1.3) | 33.9 (2.6) | 23.5 (1.7) | 25.7 (2.2) | 25.9 (2.4) |
| Incorrect | 16.4 (1.3) | 30.7 (2.4) | 20.7 (1.5) | 25.0 (2.0) | 28.0 (2.3) |
| p | .47 | < .001** | < .001** | .39 | .86 |
| SEF | | | | | |
| Correct | 13.4 (1.1) | 18.5 (1.4) | 20.1 (1.5) | 22.2 (1.5) | 22.7 (1.5) |
| Incorrect | 13.1 (1.1) | 17.0 (1.3) | 17.02 (1.2) | 20.9 (1.4) | 20.5 (1.3) |
| p | .20 | < .001** | < .001** | .02** | < .001** |

The above analysis included the entire population of neurons we recorded. However, we encountered a variety of neuronal response types throughout the recording sessions. For example, some neurons only responded to visual stimuli, others were only active around the time of a saccade, and others had delay-period activity or various combinations of activities, etc. We repeated the analyses above after excluding neurons that were unmodulated in the given epochs. In other words, included in this follow-up analysis were only the subsets of neurons that were significantly active within each epoch (as described in Methods). This analysis yielded the same results; namely FEF, PFC, and SEF visual and delay activities encoded the upcoming decision (Table 7).

**Table 7: Decision-related activity: population subset**

Subset of population with increased activity relative to baseline, decision-related firing rates during Decision Stage epochs. For each cortical region, the number of included neurons in each epoch is listed, all correct vs. all incorrect firing rates (spikes/sec) are shown with standard errors in parentheses, paired t-test p-values underneath. Asterisks represent significant difference between correct and incorrect.

| FEF | Visual-1 | Delay | Presaccadic-1 | Postsaccadic |
|---|---|---|---|---|
| n | 63 | 52 | 41 | 19 |
| Correct | 33.0 (2.2) | 25.9 (2.1) | 44.4 (3.5) | 39.4 (6.4) |
| Incorrect | 28.5 (1.9) | 21.7 (2.0) | 44.7 (4.0) | 37.4 (6.4) |
| p | < .001** | < .001** | .81 | .44 |
| PFC | | | | |
| n | 80 | 54 | 25 | 20 |
| Correct | 42.4 (3.2) | 32.8 (2.6) | 47.7 (6.0) | 40.8 (8.0) |
| Incorrect | 37.4 (3.1) | 26.0 (2.4) | 43.0 (5.2) | 39.9 (8.4) |
| p | < .001** | < .001** | .10 | .64 |
| SEF | | | | |
| n | 68 | 76 | 27 | 46 |
| Correct | 23.5 (2.3) | 26.5 (2.0) | 26.1 (3.0) | 28.8 (2.4) |
| Incorrect | 21.0 (2.1) | 21.2 (1.6) | 25.3 (2.9) | 24.1 (2.2) |
| p | < .001** | < .001** | .28 | < .001** |

Motor-related activity comparison: For the time period during and after the monkeys' decision saccade, we focused on neural activity related to that saccade. Again, because of the tendency for contralateral representation in SEF, PFC, and especially FEF, we analyzed trials in which a saccade was made into the contralateral field. For these saccade-related analyses, the visual target could have appeared in the contralateral hemifield (on correct trials) or the ipsilateral hemifield (on incorrect trials). Correct and incorrect trials were compared for the presaccadic-1 and postsaccadic epochs (just before and after the decision saccade, respectively; see Methods). We found that only SEF as a population had pre- and postsaccadic activity differentiating correct from incorrect decisions (Table 6). Motor-related activity in FEF and PFC was not different between correct and incorrect trials (Table 6).

We repeated this analysis on the subset of neurons active within each epoch (i.e. only for neurons with significant pre- or postsaccadic activity). For those subsets of neurons, SEF was more active during correct than incorrect decisions within the postsaccade epoch (Table 7) but not the presaccade epoch. As in the overall population, neither FEF nor PFC differentially encoded correct versus incorrect saccadic reports in either epoch.

### *Bet-related neuronal activity*

To test whether neurons encoded the bet, we compared all high bet trials with all low bet trials, regardless of the immediately preceding decisions (i.e. trials from correct and incorrect trials pooled). Based on our behavioral results, we predicted that bet-related activity would resemble the results of the decision-related activity, because monkeys showed high trial-by-trial correlations between decision and bet: correct decisions were mostly followed by high bets and incorrect decisions by low bets. As expected, we found a similar pattern of neuronal activity related to bets (for summaries, see Tables 8 and 9). Briefly, across the population of neurons in each cortical region, neuronal activity in both the visual and delay epochs predicted the monkeys’ eventual bets, with the exception of FEF delay period. During motor-related periods, SEF activity predicted the monkeys’ bets as well, though neither FEF nor PFC did. Similarly, across the subset population of neurons with increased activity relative to baseline, neuronal activity in both the visual and delay epochs predicted bets, with the exception of SEF visual period. During motor-related periods, SEF postsaccadic and PFC presaccadic activity predicted bets.

**Table 8: Bet-related activity: population**

Population bet-related firing rates during Decision Stage epochs. For each cortical region, all high vs. all low bet firing rates (spikes/sec) are shown with standard errors in parentheses, paired t-test p-values underneath. Asterisks represent significant difference between correct and incorrect.

| FEF | Baseline | Visual-1 | Delay | Presaccadic-1 | Postsaccadic |
|---|---|---|---|---|---|
| Correct | 10.3 (.9) | 25.9 (1.9) | 18.7 (1.5) | 29.3 (2.6) | 19.6 (2.3) |
| Incorrect | 10.3 (.9) | 24.4 (1.7) | 18.1 (1.5) | 30.5 (2.8) | 20.7 (2.4) |
| p | .83 | < .01** | .31 | .42 | .16 |
| PFC | | | | | |
| Correct | 16.0 (1.3) | 32.7 (2.5) | 22.6 (1.6) | 25.6 (2.2) | 25.8 (2.4) |
| Incorrect | 16.5 (1.3) | 31.4 (2.4) | 21.4 (1.5) | 25.4 (1.9) | 26.0 (2.3) |
| p | .20 | < .01** | .02** | .82 | .62 |
| SEF | | | | | |
| Correct | 13.3 (1.1) | 18.2 (1.4) | 19.9 (1.5) | 22.3 (1.5) | 22.5 (1.5) |
| Incorrect | 13.2 (1.1) | 17.5 (1.3) | 17.5 (1.3) | 20.8 (1.4) | 20.6 (1.3) |
| p | .40 | .04** | < .001** | .01** | < .01** |

**Table 9: Bet-related activity: population subset**

Subset of population with increased activity relative to baseline, bet-related firing rates during Decision Stage epochs. For each cortical region, the number of included neurons in each epoch is listed, all high vs. all low bet firing rates (spikes/sec) are shown with standard errors in parentheses, paired t-test p-values underneath. Asterisks represent significant difference between correct and incorrect.

| FEF | Visual-1 | Delay | Presaccadic-1 | Postsaccadic |
|---|---|---|---|---|
| n | 63 | 52 | 41 | 19 |
| Correct | 31.7 (2.1) | 24.8 (2.0) | 44.4 (3.6) | 38.4 (6.3) |
| Incorrect | 29.3 (2.0) | 22.4 (2.1) | 43.8 (3.8) | 38.3 (6.4) |
| p | < .01** | < .01** | .69 | .97 |
| PFC | | | | |
| n | 80 | 54 | 25 | 20 |
| Correct | 40.6 (3.2) | 30.8 (2.6) | 49.0 (6.3) | 40.9 (8.0) |
| Incorrect | 38.4 (3.1) | 27.3 (2.4) | 42.2 (4.6) | 39.4 (8.1) |
| p | < .001** | < .001** | .02** | .32 |
| SEF | | | | |
| n | 68 | 76 | 27 | 46 |
| Correct | 22.8 (2.2) | 25.8 (2.0) | 26.2 (3.0) | 27.7 (2.4) |
| Incorrect | 21.9 (2.1) | 22.2 (1.8) | 25.2 (3.0) | 24.5 (2.1) |
| p | .11 | < .001** | .26 | .001** |

### *Metacognitive-related neuronal activity*

To test whether neuronal activity correlated with metacognitive monitoring, we compared trial outcomes when the monkey made the *same* decision but *different* bets. Our rationale was that metacognition is the process that links a decision to a bet, allowing bets to be purposeful and not random. Signals related to metacognition should differ between trials when a decision is followed by an appropriate bet and trials when a decision is followed by an inappropriate bet.

We first compared neural activity between Correct-High (CH) trials and Correct-Low (CL) trials. This was a straightforward analysis because visual stimuli and saccade directions were equivalent in CH and CL trials throughout the Decision Stage. All targets were located within, and saccade vectors directed into, the contralateral field. The critical time period was the interstage epoch (the time span after the decision was reported and before the bet targets appeared; see Methods).

In our FEF sample, neuronal activity was no different in CH vs. CL trials during the interstage epoch. A single neuron example from FEF is shown in Figure 15a. Data are aligned on Decision Stage target appearance in the left panel and Bet Stage fixation in the right panel. The neuron was equally active for CH and CL trials, including in the crucial interstage epoch (Figure 15a shaded region:  CH: 11.8 ± 1.1 sp/s, CL: 12.8 ± 1.2 sp/s, $p > .05$, post-ANOVA t-test). The same negative result was seen in the entire FEF sample (Figure 15b). None of the single neurons were differentially active during the interstage for CH versus CL trials (open circles), and as a population, FEF neurons were equally active for CH and CL trials (Figure 15b, CH: 16.4 ± 1.5 sp/s, CL: 16.1 ±1.5 sp/s, $p = .57$ paired t-test). This population activity is reflected in the averaged spike density function (Figure 15c). Throughout the interstage epoch, CH average activity (solid black) overlapped with CL average activity (dashed black).

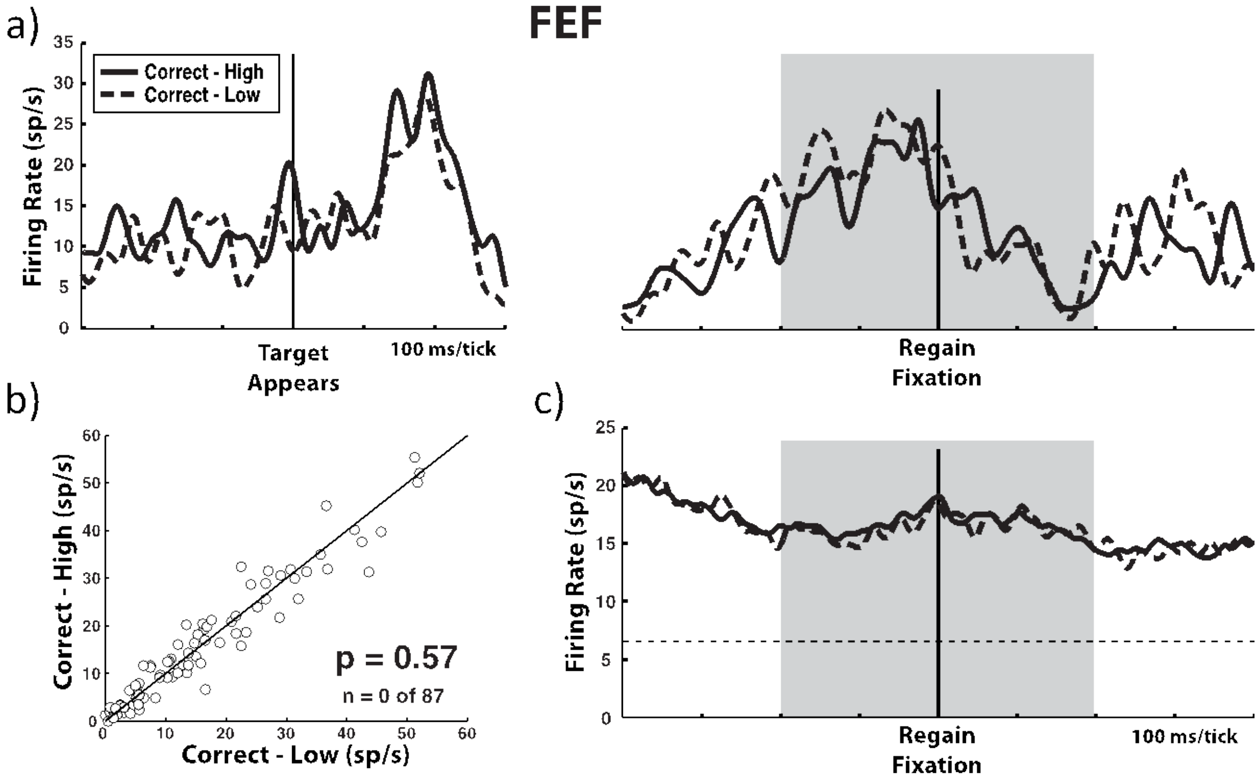


**Figure 15: FEF interstage period activity**

(a) A single FEF neuron during Correct-High trials (solid line) and Correct-Low trials (dashed line) aligned on Decision Stage target appearance (left panel) and on regaining fixation to continue to Bet Stage (right panel), all contralateral trials. Grey shading indicates interstage epoch. This neuron was not significantly different between trial types (CH = 11.8 +/- 1.1 sem; CL = 12.9 +/- 1.3 sem, p > .05). (b) Interstage period comparison of CH versus CL trial activity across the population of FEF neurons recorded. Each circle represents contralateral firing rates of one neuron. No single neurons were significant, and FEF as a population was not significant (paired t-test, p = .57). (c) Average CH and CL spike density functions across FEF population during the interstage epoch.

PFC neuron activity was marginally better at distinguishing CH from CL trials. Figure 16a depicts one of the few PFC single neurons that showed an effect. The neuron was more active for CH trials than CL trials during the interstage epoch (CH: 24.4 ± 1.5 sp/s, CL: 14.7 ± 1.3 sp/s, p < .05 post-ANOVA t-test). Figure 16b shows data from the entire PFC sample. Only 4 of the 112 individual neurons were differentially active for CH versus CL trials (filled circles), and as a population, there was no average activity difference between CH and CL trials (CH:

21.9 ± 2.0 sp/s, CL: 22.0 ± 2.0 sp/s, p = .82 paired t-test). Like for FEF, the averaged spike density functions for PFC overlapped between trial types (Figure 16c).

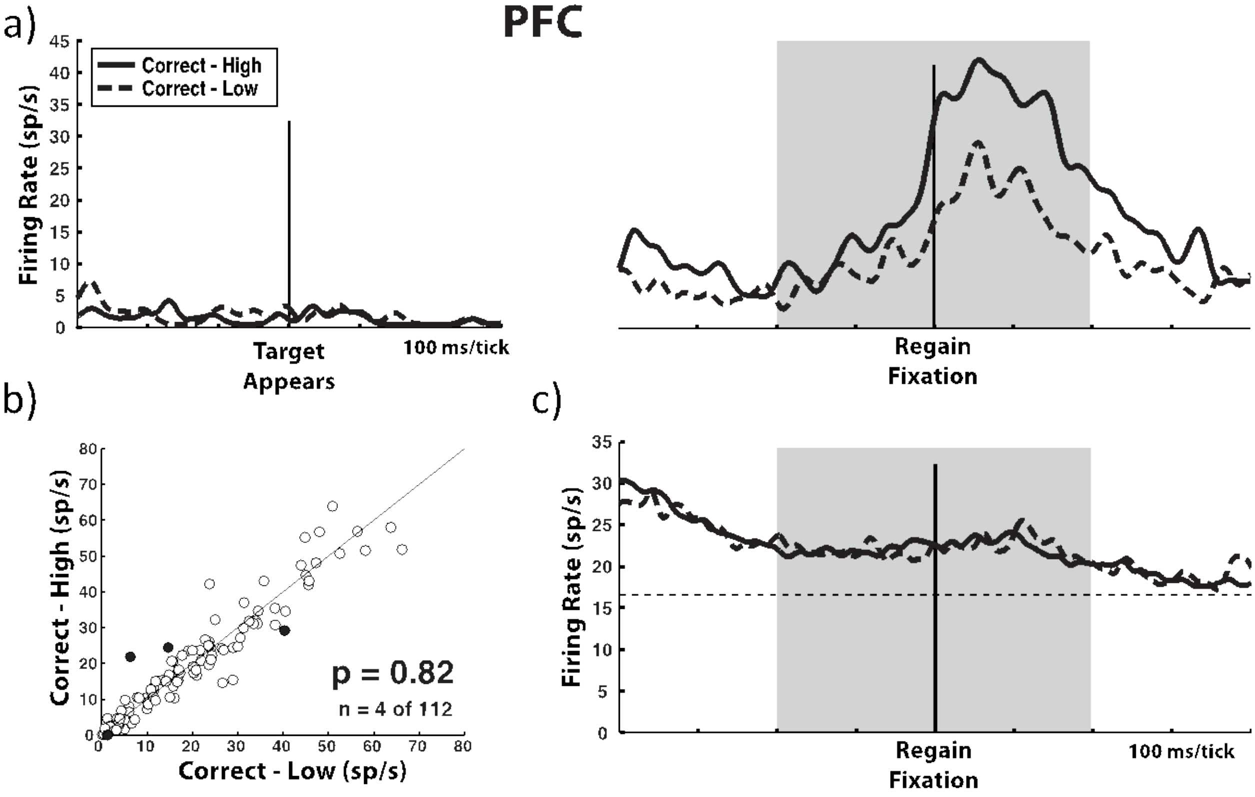


**Figure 16: PFC interstage period activity**

Conventions as in Figure 15. (a) A single PFC neuron with greater activity during CH than CL trials (CH = 24.4 +/- 1.5 sem; CL = 14.7 +/- 1.3 sem, p < .05). (b) Population interstage period CH versus CL. Filled circles represent single neurons significantly different between CH and CL trials. PFC was not significantly different as a population (paired t-test, p = .82). (c) Average CH and CL spike density functions across PFC population during the interstage epoch.

The SEF seemed to be the major player in sustaining a metacognitive signal. For some individual SEF neurons, such as the one shown in Figure 17a, firing rates were profoundly different between CH and CL trials. This example neuron was 2.5 times more active during the interstage epoch for CH trials than CL trials (CH: 63.4 ± 1.8 sp/s, CL: 25.8 ± 3.5 sp/s, p < .05, post-ANOVA t-test). Such strongly modulated neurons were not uncommon; overall, 15% (20/133) of individual SEF neurons in our sample had significantly different activity in CH vs. CL trials (Figure 17b, filled circles). The effect was found in the SEF population as well, which had a higher average firing rate for CH vs. CL trials (CH: 23.6 ± 1.6 sp/s, CL: 21.3 ± 1.5 sp/s, p = .0004 paired t-test). The effect was sustained throughout the interstage epoch, as illustrated by the averaged spike density functions (Figure 17c).

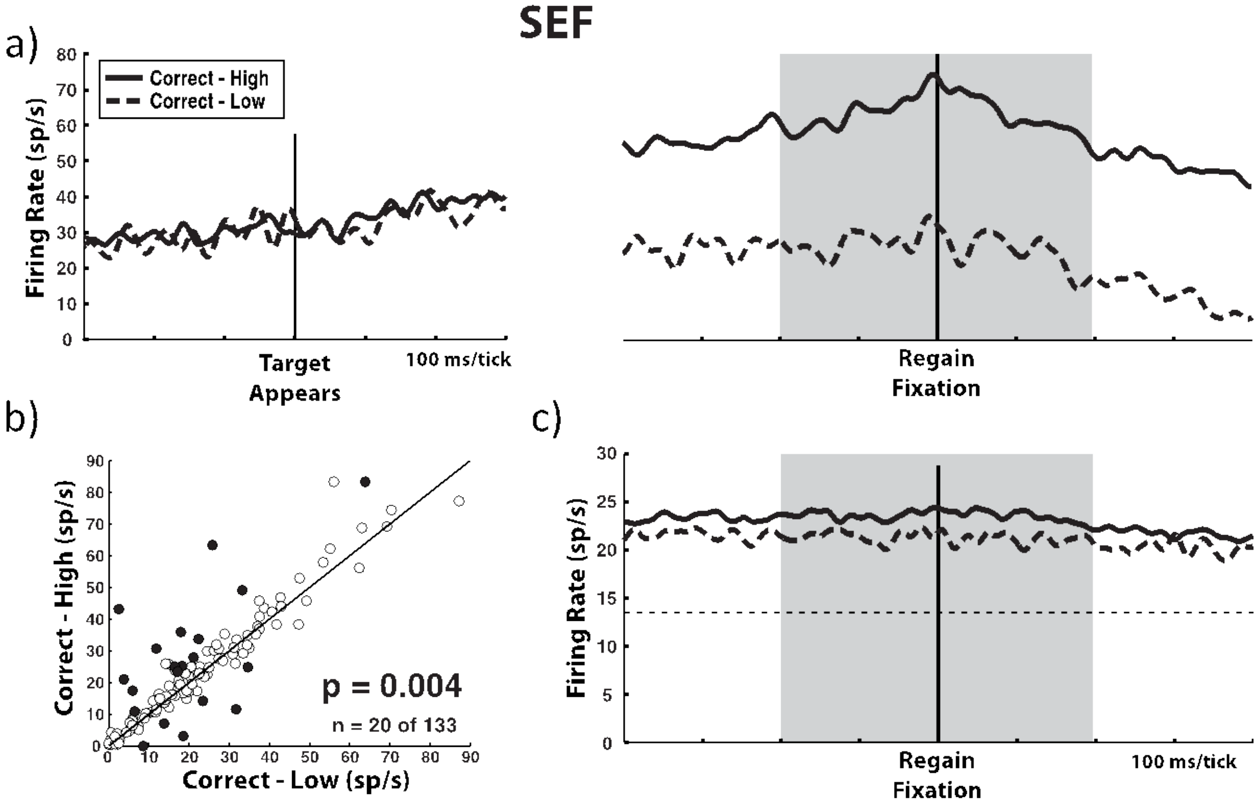


**Figure 17: SEF interstage period activity**

Conventions as in Figure 15. (a) A single SEF neuron with greater activity during CH than CL trials (CH = 63.4 +/- 1.8 sem; CL = 25.8 +/- 3.5 sem, p < .05). (b) Population interstage period CH versus CL. SEF was significantly different as a population (paired t-test, p = .004). (c) Average CH and CL spike density functions across SEF population during the interstage epoch.

We also tested whether time periods earlier than the interstage showed differential CH-CL activity. Figure 18 shows SEF population average neural firing throughout the Decision

Stage of the task, aligned on target appearance (Figure 18a), fixation offset (Figure 18b), saccade onset (Figure 18c), and bet fixation (Figure 18d, same as Figure 17c). SEF neurons as a population began differentiating between CH and CL trials ~200ms after the decision target appeared; that is, at a time when the monkeys may have made their decision (see Discussion), but well before they reported it with a saccade. Firing rates during the individual epochs were as follows: visual-1: CH: 18.5 ±1.4 sp/s, CL: 17.7 ± 1.5 sp/s; delay: CH: 20.1 ± 1.6 sp/s, CL: 18.6 ± 1.4 sp/s; presaccade-1: CH: 22.2 ± 1.5 sp/s, CL: 20.1 ± 1.4 sp/s; postsaccade: CH: 22.9 ± 1.5 sp/s, CL: 21.0 ± 1.4 sp/s. For all of these epochs except the visual, CH and CL firing rates were significantly different ($p < .05$, paired t-tests).

Figure 19 shows two examples of single SEF neuron activity throughout the entire task, including all trial outcomes at various time points. The two neurons differ regarding the pattern of activity among trial outcomes. Whereas the neuron in Figure 19a (same as Figure 17b) had high firing rates during all high bet trials (CH and IH) and low firing rates during all low bet trials (CL and IL) throughout the entire trial, the neuron in Figure 19b had more graded activity (CH > CL≈ IL ≈ IH through the Decision Stage, and CH > IH > IL > CL just before reward). We observed too much variation among activity patterns of single SEF neurons to separate them into subtypes based on the patterns, but we consider it further in the Discussion.

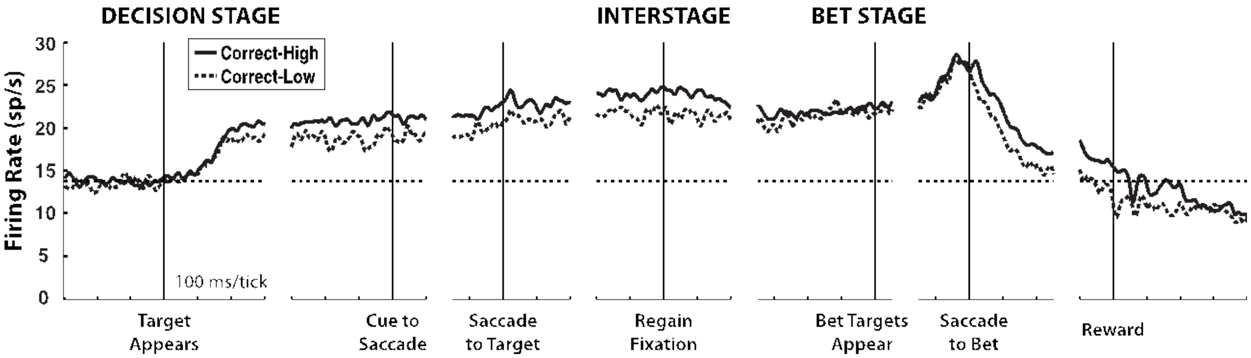


**Figure 18: SEF population CH versus CL throughout the entire trial**

CH trial activity (solid line) begins to separate from CL trial activity (dashed line) ~200ms after Decision Stage target onset, and maintains greater firing rate through the interstage period. Baseline firing rate shown as horizontal dashed line.

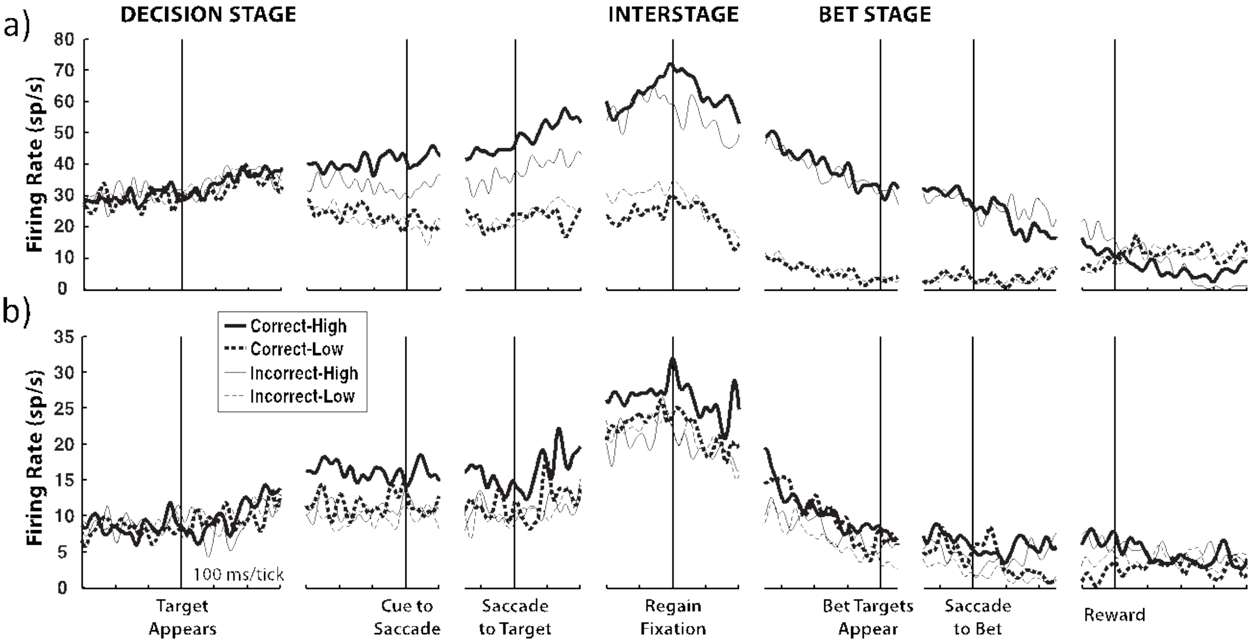


**Figure 19: Two SEF single neurons**

(a) and (b) Neuronal activity of CH (thick solid), CL (thick dashed), IH (thin solid), and IL (thin dashed) trial outcomes at various time points throughout the entire trial.

As in the preceding analyses, we repeated this CH-CL analysis on the subset of neurons significantly active within each epoch. These subsets of neurons resulted in the same pattern of activity within SEF, with the exception that the neurons became predictive of metacognitive performance slightly later, during the delay epoch rather than the visual epoch. In addition, this analysis revealed a potential, though small, contribution of FEF and PFC. The subsets of neurons in FEF and PFC active during the decision presaccadic epoch were differentially active for CH versus CL trials as a population (Figure 20, FEF, n = 41: CH: 41.0 ± 3.2 sp/s, CL: 37.8 ± 3.0 sp/s, p < .05; PFC, n = 25: CH: 43.9 ± 5.3 sp/s, CL: 39.1 ± 4.1 sp/s, p < .05). No single neurons in FEF or PFC showed the effect, however. Neither the FEF nor PFC was differentially active during any other epochs.

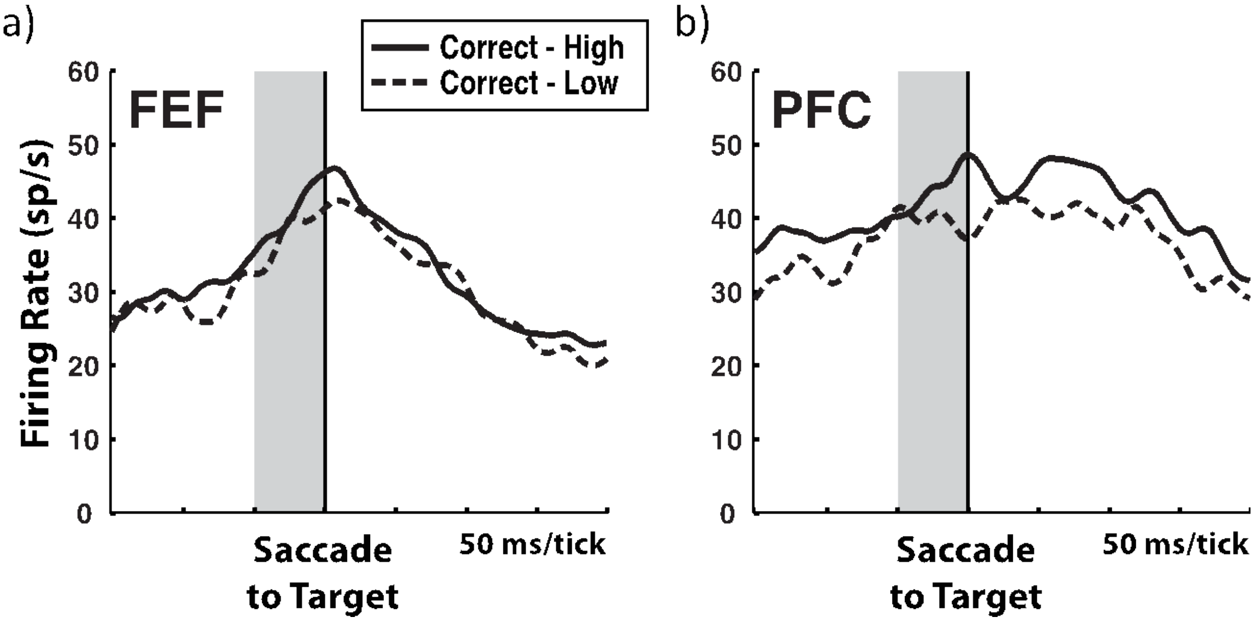


**Figure 20: CH versus CL population presaccadic-1 neuronal activity of (a) FEF and (b) PFC**

Each plot represents the population of neurons with increased activity relative to baseline (FEF: n = 41; PFC: n = 25). Presaccadic-1 epoch shown in shaded region, plots aligned to saccade onset, CH trials: solid lines, CL trials: dashed lines.

The complementary approach to testing whether neuronal activity correlates with metacognitive behavior is to compare Incorrect-High (IH) vs. Incorrect-Low (IL) trials. A complicating factor in analyzing IH vs. IL trials is that the target location is not coincident with the saccade destination. Therefore here, as in our analysis above of decision-related activity, we separated the analysis into a sensory-related activity comparison (visual-1 and delay epochs) and a motor-related activity comparison (presaccadic-1 and postsaccadic epochs). We used the latter comparison method to analyze the interstage epoch, because this epoch was during a time when the monkey's gaze returned to the center of the screen after making its (erroneous) decision saccade.

Our IH vs. IL analysis yielded similar results as our CH vs. IL analysis, except that the effects were weaker (data not shown).  In the population activity, only the SEF exhibited a difference in activity during the interstage epoch for IH vs. IL trials (IH: 21.4 ± 1.6 sp/s, IL: 19.6 ± 1.4 sp/s, p = .005). None of the cortical regions showed differential activity between IH and IL trials during the visual and delay periods. Around the time of the saccade, only the FEF showed an effect; as a population, it was less active for IH trials than IL trials during the postsaccade epoch (IH: 19.0 ± 2.6 sp/s, IL: 21.97 ± 2.9 sp/s, p = .008), but not during the presaccade epoch. Neither SEF nor PFC activity was different in IH vs. IL trials during the pre- or postsaccadic epochs. Finally, we repeated these analyses using only the subsets of neurons with significant activity in the given epochs. Again, SEF was differentially active during the interstage epoch (IH: 23.1 ± 1.0 sp/s, IL: 21.18 ± 1.7 sp/s, p = .02), but none of the other comparisons were significant.

*Bet Stage- and Reward-related activity*

To this point we have reported neuronal activity from the start of a trial through the interstage period that links the Decision Stage to the Bet Stage. Here we consider the neuronal activity through the Bet Stage to end of trial.

As soon as the bet targets appeared, signals putatively related to metacognition seemed to be interrupted. None of the three cortical regions differentiated CH-CL or IH-IL in their visual responses to the bet targets (visual-2 epoch) or in their activities associated with saccades to the bet targets (presaccadic-2 epoch; paired t-tests all $p > .05$). This held true for total population analyses, and for population subsets with increased activity relative to baseline.

The neurons became differentially modulated again around the time of reward delivery. The total SEF population was more active on CH than CL trials during both the reward anticipation (CH: 18.6 ± 1.6 sp/s, CL: 15.8 ± 1.5 sp/s, $p < .001$) and reward periods (CH: 11.5 ± 1.2 sp/s, CL: 10.0 ± 1.1 sp/s, $p = .003$), as was the subset of neurons with increased activity during those periods (reward anticipation, $n = 69$: CH: 25.4 ± 2.4 sp/s, CL: 20.9 ± 2.4 sp/s, $p < .001$; reward, $n = 29$: CH: 21.9 ± 4.1 sp/s, CL: 18.4 ± 3.8 sp/s, $p = .02$). The total population of both FEF and PFC neurons showed the reverse modulation in the reward anticipation period: less activity for CH vs CL trials (FEF: CH: 16.87 ± 2.0 sp/s, CL: 18.86 ± 2.0 sp/s, $p = .005$; PFC: CH: 23.8 ± 2.1 sp/s, CL: 26.1 ± 2.3 sp/s, $p = .04$). Neither the FEF nor PFC was significantly modulated by CH vs. CL during the reward epoch. For IH-IL comparisons, the SEF total population had greater firing rates for IH trials during the reward epoch (IH: 16.8 ± 1.7 sp/s, IL: 11.6 ± 1.4 sp/s, $p < .001$), and the PFC total population had greater firing rates for IH than IL trials during both epochs (reward anticipation: IH: 25.3 ± 2.2 sp/s, IL: 23.0 ± 2.1 sp/s, $p = .04$; reward: IH: 15.91 ± 1.5 sp/s, IL: 11.0 ± 1.0 sp/s, $p < .001$).

*Influence of past trial outcomes*

In strategic decision making tasks, choices can be affected by the outcomes of previous trials. Neurons in PFC and anterior cingulate cortex - an area interconnected with SEF (Huerta and Kaas, 1990; Luppino et al., 2003) - show activity that varies with previous trial outcomes (Barraclough et al., 2004; Seo and Lee, 2007), suggesting the neural activity could be used to guide behavioral choices as predicted by reinforcement learning theory (Sutton and Barto, 1998). Therefore as a final analysis we examined if a monkey's behavior on a current trial was influenced by the outcome of the previous trial.

We calculated the probability a monkey would repeat or switch its current trial bet as a function of past trial outcome. If a monkey's current bet is influenced by the outcome of the previous trial, we predicted that it would switch bets with relatively low likelihood after CH and IL trials (a "win-stay" strategy) but with relatively high likelihood after IH and CL trials (a "lose-switch" strategy). These predictions were not borne out, demonstrated by comparing switch rates after each trial outcome (Figures 21a and 21b). After CH trials the monkeys did not switch to Low bets at rates less than their average Low bet rates, and after IH trials they did not switch to Low bets at rates greater than average Low bet rates (Figures 21a and 21b, left data). Likewise, after CL trials the monkeys did not switch to High bets at rates greater than average High bet rates, and after IL trials they did not switch to High bets at rates less than average High bet rates (Figures 21a and 21b, right data; t-tests, all p > .05).

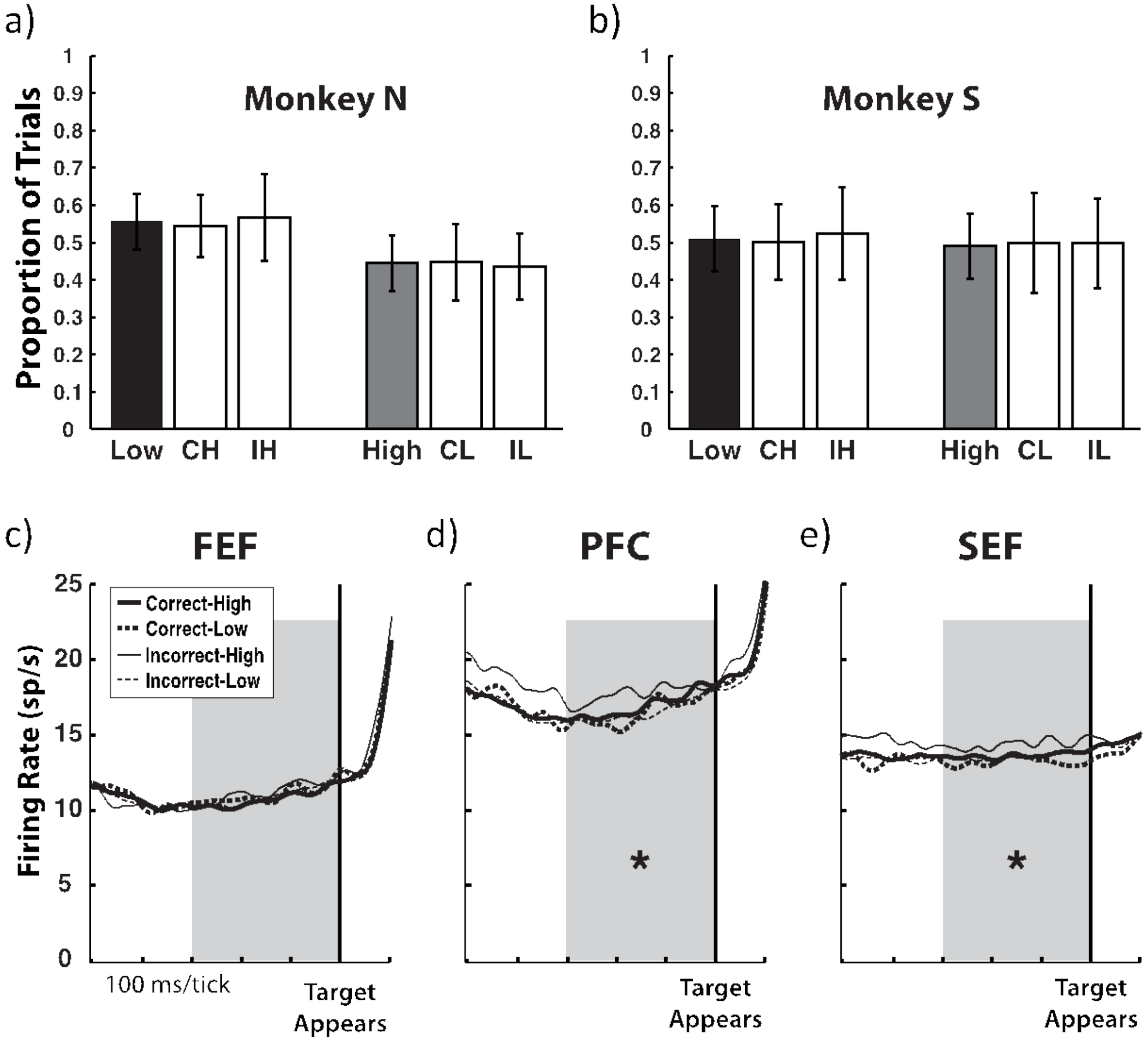


**Figure 21: Influence of previous outcome on current trials**

(a) and (b) Rate of switching bets as a function of previous trial outcome for Monkey N (a) and Monkey S (b). In each bar graph, the average low bet rate (black bar) is plotted next to switch rates for previous CH and IH trial outcomes (white bars), and the average high bet rate (grey bar) is plotted next to switch rates for previous CL and IL outcomes (white bars). Switch rates were not different from respective bet rates for either monkey (paired t-tests, p > .05). Error bars represent standard deviations. (c) – (e) Neuronal activity during baseline period (shaded) as a function of previous trial outcome for population of neurons in FEF (a), PFC (b), and SEF (c). Asterisks indicate IH activity is greater than the three other trial outcomes.

Even though the monkeys' bets were apparently not influenced by previous trial outcomes, we tested whether neurons carried information about the previous outcomes that could in principle be used guide bets. We found that for SEF the baseline epoch within the next trial carried information from the previous trial (Figure 21e). Specifically, IH trials led to significantly higher firing rates than each of the other trial outcomes during the baseline period of the next trial (paired t-tests, p < .05). IH trials yield the worst possible outcome, a timeout and no reward, so this effect is reminiscent of a “prediction error”. However, this differential modulation disappeared as soon as the Decision Stage of the next trial began (target appearance) and did not return throughout the course of the trial. No other epochs in SEF distinguished between previous trial outcomes (paired t-tests, p > .05).

Intriguingly, neurons in both PFC and FEF did carry information from the previous trial into various Decision Stage epochs of the next trial. PFC activity contained a substantial amount of previous-trial information, as was seen previously (Barraclough et al., 2004). As in SEF, in the PFC baseline firing rates were higher after IH trials than after each of the other previous-trial outcomes (Figure 21d, paired t-tests, p < .05). This IH-related signal carried over into Decision Stage epochs, specifically the visual and delay periods (not shown). Also, in PFC previous CH trials led to significantly higher firing rates than previous IL trials during Decision Stage visual and delay periods (paired t-tests, p < .05, not shown). In the FEF, IH trials led to significantly higher firing rates than each of the other three trial outcomes during the postsaccadic period of the next trial (paired t-tests, p < .05, not shown), and IH trials led to significantly higher firing rates than IL trials during the interstage period of the next trial (paired t-test, p < .05, not shown).

In summary, neuronal activity in the PFC and FEF (but not in the SEF) carried information from previous trial outcomes into the next trials, even though the monkeys’ behavior indicates the information was not utilized to make bets. This reinforces the fundamental conclusion of our study, that activity in PFC and FEF is dissociated from metacognitive behavior. The SEF, in contrast, carried information from the previous trial only up to the beginning of the next trial. Activity in the SEF was effectively reset at the start of every trial, consistent with all of the other evidence reported here that the SEF plays a role in metacognitive monitoring on a trial-by-trial basis.

## 3.5 DISCUSSION

We recorded single neuron activity in the FEF, PFC, and SEF while monkeys performed a visual oculomotor task requiring them to monitor their own decisions. As previously reported (Middlebrooks and Sommer, 2011) and replicated here for the recording sessions, monkeys seemed to use a metacognitive strategy to perform the task. Early in the Decision Stage of the task, the FEF, PFC, and SEF each encoded the upcoming decision as seen in many single neurons and at the population levels. Additionally, each area encoded the eventual bet. Activity in the SEF stood out in encoding information that linked the decision to the bet, which we interpret as a correlate of metacognitive monitoring. The SEF initiated its putative metacognitive activity swiftly (~200ms) after the Decision Stage target appeared, and maintained it throughout the Decision Stage and until the Bet Stage was underway. Monkey behavior was independent of previous trial outcome, and so was SEF activity (but not PFC or FEF activity).

### 3.5.1 SEF activity

The SEF results support our hypothesis that it would be involved in metacognitive monitoring. We were surprised, however, that the correlated signal arose so early in the trials. Given that SEF responses to visual stimuli typically occur at latencies of ~80ms or longer (Schall, 1991; Pouget et al., 2005), the ~200ms it took for CH vs. CL trial signals to separate (after target onset) leaves only a little over 100ms for the metacognitive information to emerge. We see this as providing neuronal evidence that the monitoring of a cognitive operation occurs almost immediately after, or possibly concurrent with, the operation itself. This seems analogous to the timecourse of monitoring action plans ("corollary discharge"); when motor areas finalize a movement command, upstream areas monitor it in near simultaneity, within milliseconds (e.g. see Sommer & Wurtz 2008 for review).

The putatively metacognitive SEF activity that we found had not been reported previously in this area. Although to our knowledge no fMRI studies reported human SEF signals during metacognition tasks (note that none were eye-movement tasks or even monitored eye movements), many of the fMRI results have implicated regions known in monkeys to be interconnected with SEF, such as anterior cingulate and various medial prefrontal regions (e.g., Kikyo et al. 2002; Chua et al. 2006; Kim and Cabeza 2009). Previous recording studies in monkeys reported that SEF neurons signal reward, errors, conflict, and/or inhibition of planned saccades, collectively referred to as performance monitoring (Stuphorn et al., 2000; Nakamura et al., 2005). Of these signals previously found in SEF, reward and error signals were the only ones relevant to our task and SEF results.

We found two lines of evidence for reward signals in our SEF data. The first consisted of elevated firing rates during the reward epoch of CH vs. CL trials. Additionally, one could interpret the carryover of information about worst-outcome, IH trials, through the inter-trial period as a "lack of reward" signal. Neither of these reward-related signals can explain our putative metacognitive activity because both of them start after the bet on one trial and end before the decision on the next trial.  Potential reward expectation signals could, however, complicate the interpretation of our metacognitive results; we discuss such signals in a separate section below. With regard to error signals previously found in SEF (Stuphorn et al. 2000), it is not straightforward in our betting task what constitutes an “error”. A simple interpretation is that an error is any trial in which no reward would be earned (i.e. IH trials, which yielded only a timeout). But during the reward anticipation epoch (when one would expect an error signal to be active) we did not observe significantly increased firing rates on IH trials relative to other trial outcomes. A subtler interpretation is that an error is a trial in which reward was earned, but it was less reward than potentially available (CL vs. CH trials respectively). We did not see SEF activity that was greater on CL than CH trials in any epoch. Finally, an error signal might occur after any incorrect decision (e.g. during the postsaccadic and/or interstage epochs), since an incorrect decision was always less advantageous than a correct decision. We did not observe SEF neurons with that signal either. In sum, we saw little or no evidence of error signals in our data.

The signals we observed in SEF as a population were mostly ranked from CH trials (highest firing rates) to IL trials (lowest firing rates). Our interpretation that SEF signals were correlated with metacognition was based on differences between CH and CL trials and differences between IH and IL trials. An alternative expectation about metacognition, however, might predict the highest firing rates for CH and IL trials, when the animals' were accurately monitoring their decisions, and the lowest firing rates for IH and CL trials, when the animals' monitoring was inaccurate. The existence of an abstract signal like that would complement the current findings, and we searched for such a signal, but none of the neurons we recorded had that pattern.

### 3.5.2 FEF and PFC activity

Neither the FEF nor PFC showed much evidence of activity correlated with metacognitive linkage of decision to bet. Instead, activity in both areas was correlated merely with the initial stage of the task: making the decision. This supported our hypothesis about the FEF, which was based on similar results from Thompson and Schall (1999). We did find transient saccade-related FEF activity that correlated with metacognitive behavior, but in retrospect this was not surprising. FEF stimulation in humans (Grosbras and Paus, 2003) and monkeys (Moore and Fallah, 2001) decreases the threshold required to detect and report a visual target's luminance change, suggesting FEF is important for perception of a visual stimulus. One might expect a brain mechanism involved in perceiving and reporting a stimulus to be useful in monitoring the perception as well, and we speculate that FEF saccade-related activity could be sent to other brain areas to support ongoing monitoring.

We expected a more robust metacognitive signal in PFC, given that it has been directly implicated in metacognition in previous work. Human subjects who underwent TMS of dorsolateral PFC (Rounis et al., 2010) reported lower visibility ratings of stimuli even though their ability to detect the stimuli remained unchanged, and PFC lesion patients (Del Cul et al., 2009) that performed a reverse-masking detection task suffered a greater discrepancy than controls between subjective reports of awareness and objective reports of detection. One possibility is that metacognitive processing in PFC (and/or FEF) may be restricted to a smaller subset of neurons- i.e. neurons with particular response patterns, like visuo-movement, visual-only, working memory, etc. We remained agnostic regarding what characteristics of neural activity might be useful for metacognitive processing, so we recorded any neuron that seemed to fire in some way related the task. As a consequence, we did not focus on collecting data from neurons with specific firing rate patterns, and so our population data may have contained relatively few individual neurons in each potential category.

FEF and PFC activity also may be more dependent on the spatial parameters of the task than SEF. We focused our population analyses on contralateral space, but results in SEF were the same when analyzing activity from all directions. FEF neurons often have quite spatially restricted receptive fields (Bruce and Goldberg, 1985). We tried to account for receptive fields by initially observing a neuron's activity during simple visual saccade tasks (see Methods), and often we were able to place one or both Decision Stage contralateral targets within the RF. However, many neurons in all three cortical regions did not have well-structured receptive fields. This makes clear a limitation in recording from three areas and attempting to treat them as equally as possible. It could be that we did not observe metacognitive neural activity in FEF and/or PFC because focusing on the entire contalateral hemifield resulted in lower statistical power of spatially restrictive neural activity.

### 3.5.3 Alternative explanation: reward expectation

It could be argued that the putative metacognitive signals we found could be explained by reward expectation. Many neurons in the SEF (Stuphorn et al., 2000; Roesch and Olson, 2003) and PFC (Leon and Shadlen, 1999) fire in response to, and in preparation for, reward delivery. If the signals we observed were due solely to reward expectation, they arose impressively early in the task (at least in the SEF): just after the Decision target was presented. A reward expectation signal that emerges so early (and continues throughout the trial, as we observed) would be essentially identical to a metacognitive signal, since it provides the same information. However, given the complexity of behavior required to advantageously complete a trial, we doubt that reward expectation explains our metacognitive signals. If we assume that reward expectation neurons encode the relative value of an upcoming reward, we would predict neural activity would be ranked in parallel with reward amount, i.e. CH > CL > IL > IH. We did not observe this pattern of activity in the SEF population throughout the task, however, nor was it common among single neurons (see Figure 19, for example). If we take the term “reward expectation” literally, then we can use the monkeys’ bets as guides to their expectations.  If a monkey places a high bet, we can infer that it expected large reward; if it places a low bet, it expected a small reward. If the neurons encoded this expectation, we should see the pattern of activity (CH = IH) > (CL = IL). We did observe activity approximating this pattern in a few neurons (e.g. Figure 19a), but not as a population.

### 3.5.4 Risk

Another alternative explanation for our data is that activity during each trial outcome varied as a function of the relative riskiness on those trials, as has been reported in single posterior cingulate cortical neurons (McCoy and Platt, 2005). Under that assumption, increased (or decreased) neuronal activity would accompany high (or low) bets regardless of the decision. Although as a population we did not find this pattern of neuronal activity in any of the three cortical regions, some single neurons at first glance seemed to fit this activity profile, such as the SEF example neuron in Figure 19a. However, riskiness itself does not explain the pattern of activity for that neuron. We know from behavioral analyses that IH trials were rare among incorrect trials, so it is highly unlikely the neurons encoded riskiness only during such a small subset of incorrect trials. Likewise, it is unlikely the neurons encoded riskiness most frequently on CH trials among all correct trials. The neural activity, and the betting behavior, were tied to the monkeys' decisions and cannot be explained by levels of riskiness.

### 3.5.5 CH-CL versus IH-IL

Why was there not as robust a difference between IH vs. IL trials as between CH vs. CL? One possibility, mentioned in Results, may involve the slightly different spatial nature of IH and IL trials (discrepant target and saccade directions) compared with CH and CL trials (coincident target and saccade directions). Despite our attempts to perform the analyses as carefully as possible, evaluating CH vs. CL trials is simply a much more straightforward comparison.  It is also feasible that distinct neural circuits encode the monitored perception of correct versus incorrect decisions. Some evidence of this has been shown in schizophrenic patients. In confidence response tasks similar to ours, patients maintain their rate of high confidence responses after correct trials but increased their erroneous high confidence responses after incorrect trials (Moritz and Woodward, 2002; Moritz et al., 2003; Bhatt et al., 2010).

### 3.5.6 An important candidate unstudied here: LIP

Our results complement a recent study of single LIP neurons while monkeys performed an opt-out task (Kiani and Shadlen, 2009). Monkeys viewed a display of randomly moving dots that were overall coherent in one of two directions. The monkeys were allowed to observe the stimulus over time and form a decision regarding the perceived motion of the dots. On half of the trials the monkeys were forced to choose a direction and were rewarded for correct decisions. On the other half a third stimulus appeared that, if chosen, ensured the monkey of a small reward without reporting a decision- otherwise the monkey could report a decision and earn a large reward for a correct decision or no reward for an incorrect decision. LIP firing rates varied with the tendency on relatively difficult trials to choose the response associated with ensured small reward, suggesting LIP neurons correlated with decision confidence. It would be interesting to see how LIP neurons respond in a betting task that separates an explicit decision from a metacognitive report on the decision (the bet), but unfortunately it was impractical to include LIP in our study.

The nature of the visual stimuli used in our task also differed from that of the LIP study. Whereas the moving-dots stimulus required monkeys to accumulate evidence over time to discriminate movement direction, the Decision Stage of our task required detecting the location of a single brief stimulus on each trial. A possible advantage in our task is that such a brief stimulus presentation requires more immediate monitoring of an internal trace to form the decision and eventual metacognitive judgment. It remains to be seen exactly what effect such task differences might have on neuronal activity.

### 3.5.7 Consciousness

Often in the literature, metacognition is associated with conscious awareness (Nelson, 1996; Koriat, 2007). We would not assert that self-monitoring behavior in non-human animals is an indicator of self-awareness, and indeed multiple studies in humans have shown that one does not imply the other (Pierce and Jastrow, 1884; Reder and Schunn, 1996; Kentridge and Heywood, 2000; Kunimoto et al., 2001). Our results are consistent with single neuron studies reporting confidence signals in monkeys (Kiani and Shadlen, 2009) and rats (Kepecs et al., 2008). We suggest that the SEF neuronal activity we observed fits the description of a signal beyond general confidence and more intricately related to monitoring of the monkeys' percept, given how nearly immediately the neuronal signals separate relative to the time of target onset (Figure 18, left panel) and the sustained separation of these signals. As we argued previously (Middlebrooks and Sommer, 2011), metacognition may be to cognition as corollary discharge is to action; both describe the ability of the brain to internally monitor its mental or motor effectors. Just as it appears that all animals that move have internal circuits for monitoring their movements (Crapse and Sommer, 2008a), it may be that all animals with even rudimentary cognitive abilities have the ability to monitor those abilities. This monitoring ability, however, does not necessarily imply states of self-awareness anywhere near the levels experienced by humans and perhaps other anthropoid apes.

# 4.0 GENERAL DISCUSSION

The central pursuit of this dissertation was to relate single neuron activity to a metacognitive process. We trained two monkeys on a visual oculomotor task that afforded maximum reward when performed with a metacognitive strategy (Chapter 2). Using multiple analytical methods including trial-by-trial analysis, signal detection theory, and reward rates, we concluded that metacognition best explained the monkeys' strategy. Their behavior could not be accounted for by relying on their own saccade latencies, visual cues, or trial difficulty to make bets. In Chapter 3 we recorded from single neurons in the FEF, PFC, and SEF while monkeys performed the task. All three regions encoded the decisions, and all three regions encoded the bets. SEF encoded the metacognitive process – the link between decision and bet – more robustly than FEF or PFC. As a population, SEF differentiated CH trials from CL trials beginning ~200ms after Decision Stage target onset and lasting into the Bet Stage, providing a sustained source of metacognitive information. Subsets of FEF and PFC neurons differentiated CH from CL trials transiently around the time of the saccade to report the decision during the Decision Stage. These results provide the first evidence of single neuron activity in frontal cortex that directly correlate with metacognitive behaviors. Below, I discuss the potential roles of frontal cortical regions in metacognition with regard to their known roles in cognition, noting further experiments that would help answer remaining questions. Additionally, I compare our results with the psychological framework of metacognition introduced in Chapter 1 (Figure 1), and finally consider the implications of our research for consciousness and self-awareness.

The sustained SEF signals that correlated with metacognition (Chapter 3) are consistent with known SEF response properties in simpler cognitive tasks. SEF neurons respond to errors, reward, and successes (collectively referred to as “performance monitoring” responses) in a countermanding saccade task that requires inhibition of a planned saccade (Stuphorn et al., 2000; Emeric et al., 2010). In that study it was unclear whether monkeys’ future behavior could be guided by the SEF responses. Our observations offer the first evidence that monitoring-related information in SEF can reliably guide an appropriate action. This was most evident in the comparison of neuronal activity during advantageous (CH) trials relative to disadvantageous (CL) trials (Figure 18), akin to successful versus unsuccessful task performance. With respect to specific types of signals found previously in the SEF, we searched for evidence of error signals (Stuphorn et al., 2000; Emeric et al., 2010). There are two notable points in the task structure where an error might be registered: after an incorrect decision and after completion of an IH trial. The former type of error signal was not found in our data. We saw no elevations in firing rates just after an incorrect decision but before the Bet Stage began (nor did we see elevated firing rates on CL trials, when the monkey may have thought it made an incorrect decision). We did see elevated activity, however after IH trials. This makes sense because from a full-trial perspective, IH trials are the only irretrievable errors. Incorrect decisions can always be redeemed with a low bet (i.e. IL trials), thus yielding some reward. IH trials, however, result in no reward at all. And to make things worse, they are the only trial types that lead to a timeout period. Elevated activity after IH trials could also be a reward signal (more specifically, a lack-of-reward signal). Other hints of reward-related activity were found as well. Regardless, all of the putative error and reward signals that we found started after bets and stopped before the next trial. Other cognitive-related signals previously reported in SEF (e.g. conflict monitoring) were irrelevant to our task design. The bottom line is that none of the well-known signals in SEF were sufficient to explain the metacognitive-related activity that we discovered.

Our FEF results were largely consistent with the hypothesis that FEF activity would encode the decisions and not metacognitive monitoring. Like the SEF, the FEF encoded upcoming decisions during Decision Stage visual and delay epochs. The only apparent metacognitive signal in FEF neurons, however, was conveyed in association with the decision saccade, and was found only for neurons with significant saccade-related activity to begin with and only at the population level of those neurons; none of the individual neurons' firing rates differed between CH and CL trials.

The PFC activity we observed did not support our hypothesis that it would, like SEF, help link the decisions to the bets. Instead, with regard to metacognitive processing, we observed a nearly identical pattern of activity as FEF – only transient signals during the Decision Stage saccade and only among the subpopulation of neurons with increased firing rates relative to baseline. Unlike in FEF, within that PFC subpopulation there were a few individual neurons that differed significantly between CH and CL trials. We expected greater involvement of PFC in metacognitive processing. Even though studies often report similar neuron responses in FEF and PFC (Funahashi et al., 1989), some previous work has suggested a metacognitive role for PFC. Subjects who experienced TMS of human dorsolateral PFC (Rounis et al., 2010) reported lower visibility ratings of stimuli although their ability to detect the stimuli remained unchanged, and PFC lesion patients (Del Cul et al., 2009) performing a reverse-masking detection task suffered greater deficits than controls in their subjective reports of awareness than in their objective reports of detection. In addition, numerous human imaging studies have reported PFC involvement in metacognition tasks (Henson et al., 2000; Kikyo et al., 2002; Chua et al., 2004).

Negative results can be difficult to interpret. The most parsimonious explanation for our negative results concerning FEF and PFC (that they did not show much evidence for metacognitive-related signals) is that neither area plays a major role in metacognition. Another possibility is that metacognitive processing is substantial in the PFC and/or FEF but only in neurons with particular response patterns. We recorded any neuron that seemed to fire in a task-related way, to avoid investigator selection biases and to assemble data that were as representative as possible of the entire neuronal population in an area. As a consequence, we did not focus our efforts on collecting data from particular types of neurons, for instance those that were active specifically during the delay epoch for example. This strategy may have had the effect of diluting otherwise significant metacognitive contributions from certain categories of neurons. Finally, FEF and PFC activity may be more dependent on the spatial parameters of the task than SEF. We focused our population analyses on contralateral space to balance the aim of analyzing all neurons equally with the known contralateral biases of FEF and PFC. Often we were able to place both contralateral Decision Stage targets within the receptive field (for larger receptive fields), but sometimes only one target. It is possible we did not observe metacognitive neuronal activity in FEF or PFC because including the entire contralateral hemifield in analyses, while providing statistical power by increasing the total number of analyzed trials, could have lowered the sensitivity of the analysis by including data from trials outside the exact hotspot of a neuron’s receptive field. Future studies could address these issues by recording only from neurons exhibiting particular activity patterns (e.g. saccade activity only) with well-defined receptive fields. It should be noted, in defense of our SEF conclusions, that they were so robust that even when we pooled all directions together the results pertaining to metacognitive signals were unchanged (this was a supplemental analysis omitted for brevity from this document).

We observed correlations between neuronal activity and metacognitive behavior. To test the causal role of FEF, PFC, or SEF in metacognition, microstimulation and/or reversible inactivation would be needed. Inactivation may not be suitable for FEF in our task, as it seems integrally related to making saccades and inactivation of it may impair the saccadic responses needed for analyzing behavior (Sommer and Tehovnik, 1997; Dias and Segraves, 1999). Inactivation could, in principle be applied to PFC and SEF, without much risk of saccadic deficits (Sommer and Tehovnik, 1999; Sawaguchi and Iba, 2001). Microstimulation could be applied to all three regions with precise timing, at current levels below the threshold required to elicit saccades in FEF and SEF. If any of the cortical regions were causal for metacognition, we would expect microstimulation or inactivation to alter the proportion of high/low bets with respect to correct/incorrect decisions.

## 4.1 WHAT IS METACOGNITION?

Thus far I have simplified metacognition, treating it as if it were a unified computation or function that is implemented in the brain. Under that assumption, it would be tempting to ascribe SEF a core role in metacognitive processing. The variety of human metacognition studies and the frameworks developed to understand metacognition, however, suggest that what is referred to as metacognition is likely a collection of fundamentally distinct, albeit related, processes (Pannu et al., 2005b; Arango-Munoz, 2011). From that perspective (with which I concur), one might expect distinct yet overlapping brain circuits to encode the varieties of metacognitive processing.

Indirect evidence of this can be gleaned by noting the variation in results among fMRI studies that employed different metacognitive tasks. At least one fMRI study has provided direct evidence by comparing BOLD responses between two different metacognition paradigms (Chua et al., 2009)- the authors found both common and separate brain regions were involved. Furthermore, lesions patients (Schnyer et al., 2004; Pannu et al., 2005a) and patients suffering from various neurological syndromes (Pannu and Kaszniak, 2005) often suffer deficits on some metacognitive tasks while performing normally on others.

The implication that metacognition is not simply one process leaves open the possibility that SEF may be important only for the specific type of metacognitive process reported herein, that is, *retrospective monitoring* as assessed by retrospective confidence reports (the bets). It remains possible (and I think likely), however, that SEF contributes to multiple types of metacognitive processes. It is located within the medial frontal cortical region that showed increased BOLD activity across metacogntive tasks (Chua et al., 2009), and this same region is part of the “default mode” network thought to mediate self-reflective processing (Raichle et al., 2001). More experiments, employing different types of metacognitive tasks, are necessary to elucidate the scope SEF plays in metacognition.

One such experiment, requiring only a slight modification to the betting task (Figure 22), could test monkeys’ *prospective monitoring* (Figure 22). Briefly, in the Perception Stage (Figure 22 top row, similar to the original Decision Stage) the goal would be to detect a target in the periphery but maintain fixation (instead of making a saccade as in the original task). In the Bet Stage (Figure 22 middle row), the animal would make a bet to indicate whether the Perception Stage target location was detected and could be identified in the future. After making a bet, a final Decision Stage would ensue (Figure 22 bottom row) in which the goal is to make a saccade to the Perception Stage target location. Reward would be earned just as in the original betting task, based on decision/bet outcomes.

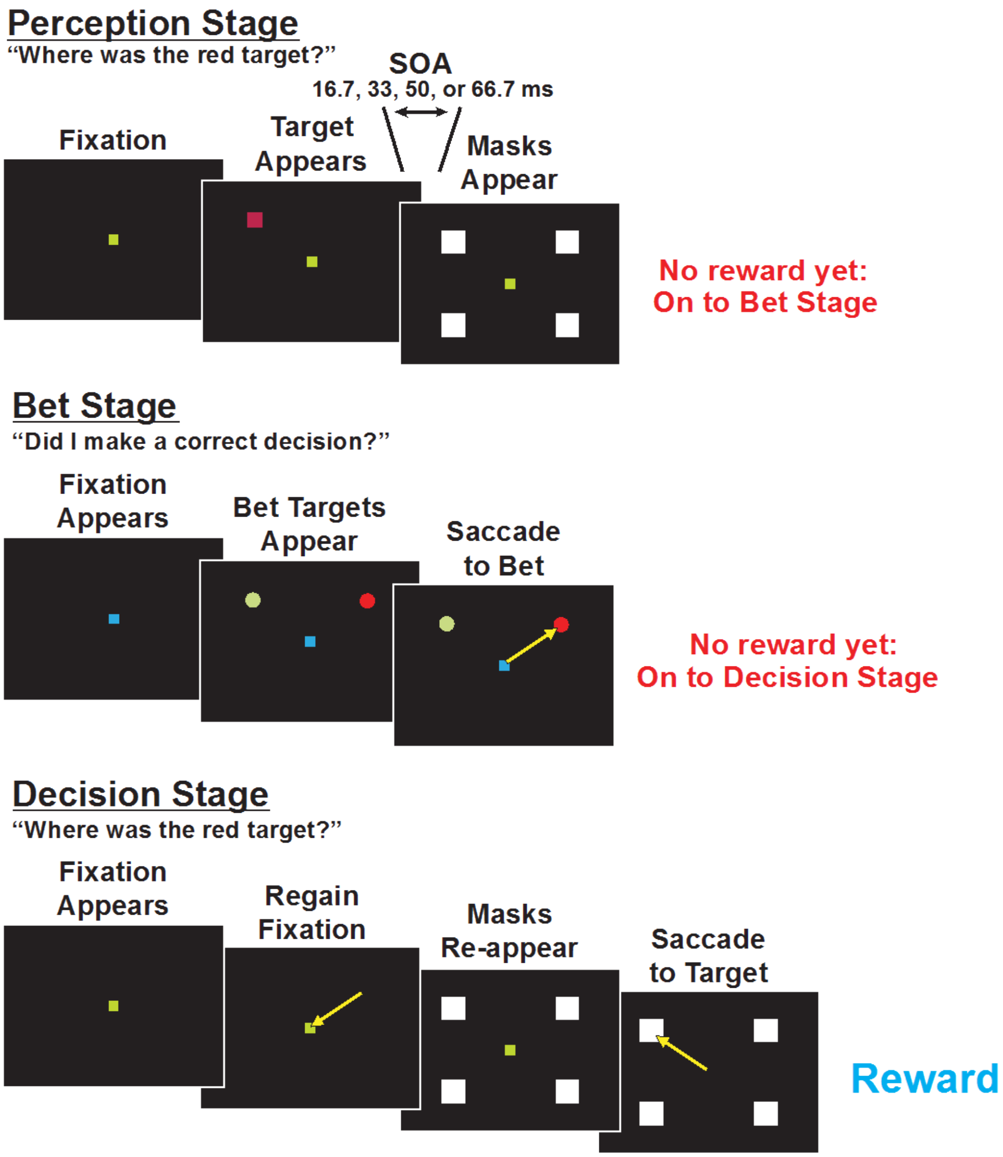


**Figure 22: Prospective monitoring task**

Each trial consists of three stages, a Perception Stage (top row), a Bet Stage (middle row), and a Decision Stage (bottom row). In the Perception Stage, monkeys foveate a fixation spot. A target appears at one of four locations in the periphery, and after a variable time (SOA), masks appear at the four locations. The monkeys remain foveated as the fixation spot changes color to indicate the start of the Bet Stage. In the Bet Stage, the two bet targets appear in the periphery. A bet is made when a saccade goes to one of the two targets. In the Decision Stage, monkeys foveate a new fixatin spot, then the masks re-appear. A correct decision is made (shown) if a saccade goes to the original location of the target. A saccade to any other location is an incorrect decision.

A related question is whether the results described herein are compatible with the psychological framework of metacognition introduced in Chapter 1 (Nelson and Narens, 1990; Figure 1). A simple neurobiological interpretation of the object-level/meta-level construct is that each processing level would consist of distinct brain regions or circuits. It is not clear, given the neuronal activity we observed, where FEF, PFC, and SEF would fit in the framework. We found that each region encoded an upcoming decision, which we take to be an object-level process. SEF also encoded the meta-level monitoring process throughout the task, and FEF and PFC briefly encoded the meta-level as well. We speculate that the same systems or circuitry serving object-level processes also serve meta-level processes, or are at least highly overlapping. That idea is consistent with single neuron recordings in LIP (Kiani and Shadlen, 2009), and may be compatible with recurrent neuronal decision-making models (Wang, 2008). We recorded neurons only in three frontal cortical regions, leaving the possibility that object-level and meta-level processes could be distinct somewhere else in the brain.

## 4.2 RELATION TO CONSCIOUSNESS AND SELF-AWARENESS

It is tempting to suggest that to excel at a metacognition task like the betting task, one must harbor phenomenal self-consciousness/awareness (Nelson, 1996). After all, metacognition is fundamentally a self-referential process, and successful metacognition entails being keenly accurate with respect to self-reference. In studies using tasks similar to the Decision Stage of our task, where a subject’s goal is to detect a target, humans (Weiskrantz, 1986) and monkeys (Cowey and Stoerig, 1995) with visual cortical damage could successfully detect the target without being able to report perception of the target (“blindsight”). In other words, they should be able to perform the Decision Stage of our task without consciousness of the target. In a more recent study with a blindsight patient, a betting task was employed as a metric for consciousness of the target (Persaud et al., 2007). The blindsight patient made a high proportion of low bets after correct target detections, leading the authors to conclude the patient truly was not conscious of the target. However, humans can perform metacognitive tasks without consciousness of the stimuli being presented or the strategies they used to perform the task (Pierce and Jastrow, 1884; Reder and Schunn, 1996; Kentridge and Heywood, 2000; Kunimoto et al., 2001). In addition, a recent fMRI experiment compared BOLD responses between explicit and implicit self-referential tasks (Rameson et al., 2010). Many of the same brain areas were active during both tasks, suggesting self-related processing may engage a common brain mechanism regardless of conscious awareness. We cannot anthropomorphize and conclude that since the monkeys performed the betting task *as if* they were aware of their decisions, they must harbor consciousness of those decisions. Nor can we conclude the neural activity observed is an index of self-awareness. Instead, we consider this research to be a systematic, neurophysiological step toward studying the elusive topic of consciousness, and we intend the work herein to be a contribution to that larger effort.

# BIBLIOGRAPHY

Arango-Munoz S (2011) Two Levels of Metacognition. Philosophia:1‚Äì12.

Armstrong KM, Moore T (2007) Rapid enhancement of visual cortical response discriminability by microstimulation of the frontal eye field. Proceedings of the National Academy of Sciences of the United States of America 104:9499-9504.

Barraclough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. Nature Neuroscience 7:404-410.

Basso MA (1998) Cognitive set and oculomotor control. Neuron 21:665-668.

Benjamin AS, Bjork RA (1996) Retrieval fluency as a metacognitive index. Implicit memory and metacognition:309-338.

Benjamin AS, Bjork RA, Schwartz BL (1998) The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. Journal of Experimental Psychology General 127:55-68.

Beran MJ, Smith JD, Redford JS, Washburn DA (2006) Rhesus macaques (Macaca mulatta) monitor uncertainty during numerosity judgments. Journal of Experimental Psychology Animal Behavior Processes 32:111-119.

Bhatt R, Laws KR, McKenna PJ (2010) False memory in schizophrenia patients with and without delusions. Psychiatry Research 178:260-265.

Boch RA, Goldberg ME (1989) Participation of prefrontal neurons in the preparation of visually guided eye movements in the rhesus monkey. Journal of Neurophysiology 61:1064-1084.

Brewer WF, Sampaio C (2006) Processes leading to confidence and accuracy in sentence recognition: a metamemory approach. Memory (Hove, England) 14:540-552.

Brown J, Hanes D, Schall J, Stuphorn V (2008) Relation of frontal eye field activity to saccade initiation during a countermanding task. Experimental Brain Research 190:135-151.

Bruce CJ, Goldberg ME (1985) Primate frontal eye fields. I. Single neurons discharging before saccades. Journal of Neurophysiology 53:603-635.

Bruce CJ, Goldberg ME, Bushnell MC, Stanton GB (1985) Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. Journal of Neurophysiology 54:714-734.

Call J (2010) Do apes know that they could be wrong? Animal Cognition 13:689-700.

Chandler CC (1994) Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test. Memory & Cognition 22:273-280.

Chua EF, Schacter DL, Sperling RA (2009) Neural correlates of metamemory: a comparison of feeling-of-knowing and retrospective confidence judgments. Journal of Cognitive Neuroscience 21:1751-1765.

Chua EF, Schacter DL, Rand-Giovannetti E, Sperling RA (2006) Understanding metamemory: neural correlates of the cognitive process and subjective level of confidence in recognition memory. NeuroImage 29:1150-1160.

Chua EF, Rand-Giovannetti E, Schacter DL, Albert MS, Sperling RA (2004) Dissociating confidence and accuracy: functional magnetic resonance imaging shows origins of the subjective memory experience. Journal of cognitive neuroscience 16:1131-1142.

Clifford CWG, Arabzadeh E, Harris JA (2008) Getting technical about awareness. Trends in Cognitive Sciences 12:54-58.

Colby CL, Duhamel JR, Goldberg ME (1996) Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. Journal of Neurophysiology 76:2841-2852.

Costermans J, Lories G, Ansay C (1992) Confidence level and feeling of knowing in question answering: The weight of inferential processes. Journal of Experimental Psychology: Learning, Memory, and Cognition 18:142.

Cowey A, Stoerig P (1995) Blindsight in monkeys. Nature 373:247-249.

Crapse TB, Sommer MA (2008a) Corollary discharge across the animal kingdom. Nature Reviews Neuroscience 9:587-600.

Crapse TB, Sommer MA (2008b) Corollary discharge circuits in the primate brain. Current Opinion in Neurobiology 18:552-557.

Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A (2009) Causal role of prefrontal cortex in the threshold for access to consciousness. Brain: A Journal of Neurology 132:2531-2540.

Dias EC, Segraves MA (1999) Muscimol-induced inactivation of monkey frontal eye field: effects on visually and memory-guided saccades. Journal of Neurophysiology 81:2191-2214.

Ding L, Hikosaka O (2006) Comparison of reward modulation in the frontal eye field and caudate of the macaque. The Journal of neuroscience : the official journal of the Society for Neuroscience 26:6695-6703.

Dobbins IG, Foley H, Schacter DL, Wagner AD (2002) Executive control during episodic retrieval: multiple prefrontal processes subserve source memory. Neuron 35:989-996.

Emeric EE, Leslie M, Pouget P, Schall JD (2010) Performance monitoring local field potentials in the medial frontal cortex of primates: supplementary eye field. Journal of Neurophysiology 104:1523-1537.

Flavell JH (1971) First discussant's comments: What is memory development the development of? Human development.

Flavell JH (1976) Metacognitive aspects of problem solving. In: The nature of intelligence (Resnick LB, ed), pp 231-236. Hillsdale, NJ: Erlbaum.

Flavell JH (1979) Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist 34:906-911.

Foote AL, Crystal JD (2007) Metacognition in the rat. Current Biology: CB 17:551-555.

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage 19:1273-1302.

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

Funahashi S, Bruce CJ, Goldman-Rakic PS (1990) Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. Journal of Neurophysiology 63:814-831.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1991) Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex. Journal of Neurophysiology 65:1464-1483.

Fuster JM (1973) Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. Journal of Neurophysiology 36:61-78.

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

Goebel R, Roebroeck A, Kim DS, Formisano E (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn Reson Imaging 21:1251-1261.

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

Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual salience in monkey parietal cortex. Nature 391:481-484.

Green DM, Swets JA (1966) Signal detection theory and psychophysics. 1966: Wiley.

Grosbras MH, Paus T (2003) Transcranial magnetic stimulation of the human frontal eye field facilitates visual awareness. The European Journal of Neuroscience 18:3121-3126.

Gusnard DA, Akbudak E, Shulman GL, Raichle ME (2001) Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. Proceedings of the National Academy of Sciences of the United States of America 98:4259-4264.

Hampton RR (2001) Rhesus monkeys know when they remember. Proceedings of the National Academy of Sciences of the United States of America 98:5359-5362.

Hampton RR, Zivin A, Murray EA (2004) Rhesus monkeys (Macaca mulatta) discriminate between knowing and not knowing and collect information as needed before acting. Animal Cognition 7:239-246.

Hart JT (1965) Memory and the feeling-of-knowing experience. J Educ Psychol 56:208-216.

Hasegawa RP, Matsumoto M, Mikami A (2000) Search Target Selection in Monkey Prefrontal Cortex. J Neurophysiol 84:1692-1696.

Hays AV, Richmond BJ, Optican L (1982) A UNIX-based multiple process system for real-time data acquisition and control. In: WESCON Conf Proc, pp 1-10.

Henson RN, Rugg MD, Shallice T, Dolan RJ (2000) Confidence in recognition memory for words: dissociating right prefrontal roles in episodic retrieval. Journal of cognitive neuroscience 12:913-923.

Higham PA (2007) No special K! A signal detection framework for the strategic regulation of memory accuracy. Journal of Experimental Psychology General 136:1-22.

Hikosaka O, Wurtz RH (1983) Visual and oculomotor functions of monkey substantia nigra pars reticulata. III. Memory-contingent visual and saccade responses. Journal of Neurophysiology 49:1268-1284.

Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science 310:1680-1683.

Huerta MF, Kaas JH (1990) Supplementary eye field as defined by intracortical microstimulation: connections in macaques. The Journal of Comparative Neurology 293:299-330.

Iba M, Sawaguchi T (2003) Involvement of the dorsolateral prefrontal cortex of monkeys in visuospatial target selection. Journal of Neurophysiology 89:587-599.

Jacobsen CF (1936) Studies on cerebral function in primates. Comparative Psychological Monographs 13:1-60.

Johnston K, Everling S (2006) Neural Activity in Monkey Prefrontal Cortex Is Modulated by Task Context and Behavioral Instruction during Delayed-match-to-sample and Conditional Prosaccade---Antisaccade Tasks. J Cognitive Neuroscience 18:749-765.

Judge SJ, Richmond BJ, Chu FC (1980) Implantation of magnetic search coils for measurement of eye position: an improved method. Vision Research 20:535-538.

Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF (2002) Finding the self? An event-related fMRI study. Journal of cognitive neuroscience 14:785-794.

Kentridge RW, Heywood CA (2000) Metacognition and awareness. Consciousness and cognition 9:308-312; discussion 324-326.

Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455:227-231.

Kiani R, Shadlen MN (2009) Representation of confidence associated with a decision by neurons in the parietal cortex. Science (New York, NY) 324:759-764.

Kikyo H, Ohki K, Miyashita Y (2002) Neural correlates for feeling-of-knowing: an fMRI parametric analysis. Neuron 36:177-186.

Kim H, Cabeza R (2009) Common and specific brain regions in high- versus low-confidence recognition memory. Brain Research 1282:103-113.

Kim JN, Shadlen MN (1999) Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nature Neuroscience 2:176-185.

Kleitman S, Stankov L (2007) Self-confidence and metacognitive processes. Learning and Individual Differences 17:161-173.

Koriat A (2007) Metacognition and Consciousness. In: The Cambridge Handbook of Consciousness (Zelazo PD, Moscovitch M, Thompson E, eds). New York: Cambridge Uniersity Press.

Koriat A, Lichtenstein S, Fischhoff B (1980) Reasons for confidence. Journal of Experimental Psychology: Human Learning and Memory 6:107.

Kornell N, Son LK, Terrace HS (2007) Transfer of metacognitive skills and hint seeking in monkeys. Psychological Science: A Journal of the American Psychological Society / APS 18:64-71.

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

Kunimoto C, Miller J, Pashler H (2001) Confidence and accuracy of near-threshold discrimination responses. Consciousness and cognition 10:294-340.

Leon MI, Shadlen MN (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. Neuron 24:415-425.

Libedinsky C, Livingstone M (2011) Role of prefrontal cortex in conscious visual perception. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:64-69.

Luppino G, Rozzi S, Calzavara R, Matelli M (2003) Prefrontal and agranular cingulate projections to the dorsal premotor areas F2 and F7 in the macaque monkey. The European Journal of Neuroscience 17:559-578.

Macrae CN, Moran JM, Heatherton TF, Banfield JF, Kelley WM (2004) Medial prefrontal activity predicts memory for self. Cerebral cortex (New York, NY : 1991) 14:647-654.

Maril A, Simons JS, Mitchell JP, Schwartz BL, Schacter DL (2003) Feeling-of-knowing in episodic memory: an event-related fMRI study. NeuroImage 18:827-836.

Masson MEJ, Rotello CM (2009) Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: implications for studies of metacognitive processes. Journal of Experimental Psychology Learning, Memory, and Cognition 35:509-527.

Mays LE, Sparks DL (1980) Saccades are spatially, not retinocentrically, coded. Science (New York, NY) 208:1163-1165.

McCoy AN, Platt ML (2005) Risk-sensitive neurons in macaque posterior cingulate cortex. Nature Neuroscience 8:1220-1227.

Metcalfe J (2008) Evolution of metacognition. In: Handbook of metamemory and memory (Dunlosky J, Bjork RA, eds), pp 29-46. New York: Psychology Press.

Middlebrooks PG, Sommer MA (2011) Metacognition in monkeys during an oculomotor task. Journal of Experimental Psychology Learning, Memory, and Cognition 37:325-337.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annual Review of Neuroscience 24:167-202.

Modirrousta M, Fellows LK (2008) Medial prefrontal cortex plays a critical and selective role in 'feeling of knowing' meta-memory judgments. Neuropsychologia 46:2958-2965.

Mohler, CW, Goldberg, ME, Wurtz, RH (1973) Visual receptive fields of frontal eye field neurons. Brain Research 61:385-389.

Moore T, Fallah M (2001) Control of eye movements and spatial attention. Proceedings of the National Academy of Sciences of the United States of America 98:1273-1276.

Moritz S, Woodward TS (2002) Memory confidence and false memories in schizophrenia. The Journal of Nervous and Mental Disease 190:641-643.

Moritz S, Woodward TS, Ruff CC (2003) Source monitoring and memory confidence in schizophrenia. Psychological Medicine 33:131-139.

Nakamura K, Roesch MR, Olson CR (2005) Neuronal Activity in Macaque SEF and ACC During Performance of Tasks Involving Conflict. J Neurophysiol 93:884-908.

Nelson TO (1996) Consciousness and metacognition. American Psychologist 51:102-116.

Nelson TO, Narens L (1990) Metamemory: A theoretical framework and new findings. The psychology of learning and motivation 26:125-141.

Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J (2006) Self-referential processing in our brain--a meta-analysis of imaging studies on the self. NeuroImage 31:440-457.

O'Neill M, Schultz W (2010) Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. Neuron 68:789-800.

O'Shea J, Walsh V (2004) Visual awareness: the eye fields have it? Curr Biol 14:R279-281.

Olson CR, Gettner SN (1995) Object-centered direction selectivity in the macaque supplementary eye field. Science 269:985-988.

Pannu JK, Kaszniak AW (2005) Metamemory experiments in neurological populations: a review. Neuropsychol Rev 15:105-130.

Pannu JK, Kaszniak AW, Rapcsak SZ (2005a) Metamemory for faces following frontal lobe damage. J Int Neuropsychol Soc 11:668-676.

Pannu JK, Kaszniak AW, Rapcsak SZ (2005b) Metamemory for faces following frontal lobe damage. Journal of the International Neuropsychological Society: JINS 11:668-676.

Persaud N, McLeod P, Cowey A (2007) Post-decision wagering objectively measures awareness. Nat Neurosci 10:257-261.

Pierce CS, Jastrow J (1884) On small differences in perception. Memoirs of the National Academy of Sciences 3:75-83.

Pouget P, Emeric EE, Stuphorn V, Reis K, Schall JD (2005) Chronometry of visual responses in frontal eye field, supplementary eye field, and anterior cingulate cortex. Journal of Neurophysiology 94:2086-2092.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) A default mode of brain function. Proceedings of the National Academy of Sciences of the United States of America 98:676-682.

Rameson LT, Satpute AB, Lieberman MD (2010) The neural correlates of implicit and explicit self-relevant processing. NeuroImage 50:701-708.

Reder LM, Schunn CD (1996) Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. Implicit memory and metacognition 45-77.

Roberts WA, Feeney MC, McMillan N, Macpherson K, Musolino E, Petter M (2009) Do pigeons (Columba livia) study for a test? Journal of Experimental Psychology Animal Behavior Processes 35:129-142.

Roesch MR, Olson CR (2003) Impact of Expected Reward on Neuronal Activity in Prefrontal Cortex, Frontal and Supplementary Eye Fields and Premotor Cortex. J Neurophysiol 90:1766-1789.

Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau H (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cognitive Neuroscience 1:165-175.

Russo GS, Bruce CJ (1996) Neurons in the supplementary eye field of rhesus monkeys code visual targets and saccadic eye movements in an oculocentric coordinate system. Journal of Neurophysiology 76:825-848.

Sawaguchi T, Iba M (2001) Prefrontal cortical representation of visuospatial working memory in monkeys examined by local inactivation with muscimol. Journal of Neurophysiology 86:2041-2053.

Schall JD (1991) Neuronal activity related to visually guided saccades in the frontal eye fields of rhesus monkeys: comparison with supplementary eye fields. Journal of Neurophysiology 66:559-579.

Schall JD (2004) On building a bridge between brain and behavior. Annu Rev Psychol 55:23-50.

Schall JD, Hanes DP, Thompson KG, King DJ (1995) Saccade target selection in frontal eye field of macaque. I. Visual and premovement activation. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 15:6905-6918.

Schlag J, Schlag-Rey M (1987) Evidence for a supplementary eye field. Journal of Neurophysiology 57:179-200.

Schnyer DM, Nicholls L, Verfaellie M (2005) The role of VMPC in metamemorial judgments of content retrievability. Journal of cognitive neuroscience 17:832-846.

Schnyer DM, Verfaellie M, Alexander MP, LaFleche G, Nicholls L, Kaszniak AW (2004) A role for right medial prefontal cortex in accurate feeling-of-knowing judgements: evidence from patients with lesions to frontal cortex. Neuropsychologia 42:957-966.

Seo H, Lee D (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 27:8366-8377.

Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. Journal of Neurophysiology 86:1916-1936.

Shields WE, Smith JD, Washburn DA (1997) Uncertain responses by humans and rhesus monkeys (Macaca mulatta) in a psychophysical same-different task. Journal of Experimental Psychology General 126:147-164.

Shields WE, Smith JD, Guttmannova K, Washburn DA (2005) Confidence judgments by humans and rhesus monkeys. The Journal of General Psychology 132:165-186.

Skinner BF (1974) About Behaviorism. New York: Vintage.

Smith JD, Shields WE, Allendoerfer KR, Washburn DA (1998) Memory monitoring by animals and humans. Journal of Experimental Psychology General 127:227-250.

Smith JD, Schull J, Strote J, McGee K, Egnor R, Erb L (1995) The uncertain response in the bottlenosed dolphin (Tursiops truncatus). Journal of Experimental Psychology General 124:391-408.

Sommer MA, Tehovnik EJ (1997) Reversible inactivation of macaque frontal eye field. Experimental Brain Research 116:229-249.

Sommer MA, Tehovnik EJ (1999) Reversible inactivation of macaque dorsomedial frontal cortex: effects on saccades and fixations. Experimental Brain Research 124:429-446.

Sommer MA, Wurtz RH (2000) Composition and topographic organization of signals sent from the frontal eye field to the superior colliculus. Journal of Neurophysiology 83:1979-2001.

Sommer MA, Wurtz RH (2001) Frontal eye field sends delay activity related to movement, memory, and vision to the superior colliculus. Journal of Neurophysiology 85:1673-1685.

Sommer MA, Wurtz RH (2002) A pathway in primate brain for internal monitoring of movements. Science (New York, NY) 296:1480-1482.

Sommer MA, Wurtz RH (2006) Influence of the thalamus on spatial visual processing in frontal cortex. Nature 444:374-377.

Sommer MA, Wurtz RH (2008) Brain circuits for the internal monitoring of movements. Annual Review of Neuroscience 31:317-338.

Son LK, Kornell N (2005) Meta-confidence judgments in rhesus macaques: Explicit versus implicit mechanisms. The missing link in cognition: Origins of self-reflective consciousness:296-320.

Stuphorn V, Taylor TL, Schall JD (2000) Performance monitoring by the supplementary eye field. Nature 408:857-860.

Stuphorn V, Brown JW, Schall JD (2010) Role of supplementary eye field in saccade initiation: executive, not direct, control. Journal of Neurophysiology 103:801-816.

Suda-King C (2008) Do orangutans (Pongo pygmaeus) know when they do not remember? Animal Cognition 11:21-42.

Sutton JE, Shettleworth SJ (2008) Memory without awareness: pigeons do not show metamemory in delayed matching to sample. Journal of Experimental Psychology Animal Behavior Processes 34:266-282.

Sutton RS, Barto AG (1998) Reinforcement learning: MIT Press.

Tanji J, Hoshi E (2008) Role of the lateral prefrontal cortex in executive behavioral control. Physiological Reviews 88:37-57.

Thompson KG, Schall JD (1999) The detection of visual signals by macaque frontal eye field during masking. Nature Neuroscience 2:283-288.

Thompson KG, Schall JD (2000) Antecedents and correlates of visual detection and awareness in macaque prefrontal cortex. Vision Research 40:1523-38.

Thompson KG, Bichot NP (2005) A visual salience map in the primate frontal eye field. Progress in Brain Research 147:251-262.

Tobler PN, O'Doherty JP, Dolan RJ, Schultz W (2007) Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. Journal of Neurophysiology 97:1621-1632.

Van Overschelde, JP (2008) Metacognition: knowing about knowing. In: Handbook of metamemory and memory (Dunlosky J, Bjork RA, eds), pp 29-46. New York: Psychology Press.

Wallis JD, Miller EK (2003) From rule to response: neuronal processes in the premotor and prefrontal cortex. Journal of Neurophysiology. 90:1790-806

Wang X-J (2005) Discovering spatial working memory fields in prefrontal cortex. Journal of Neurophysiology 93:3027-3028.

Wang XJ (2008) Decision making in recurrent neuronal circuits. Neuron 60:215-234.

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

Weiskrantz L (1986) Blindsight: a case study and implications: Oxford University Press.

Wurtz RH (1968) Visual cortex neurons: response to stimuli during rapid eye movements. Science 162:1148-1150.