

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

The Prefrontal Cortex and Hierarchical Behavior

### Permalink

<https://escholarship.org/uc/item/70h9d95j>

### Author

Sloan, Jennifer

### Publication Date

2013

Peer reviewed|Thesis/dissertation

The Prefrontal Cortex and Hierarchical Behavior

by

Jennifer Sloan

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jonathan D. Wallis, Chair

Professor Dan Feldman

Professor Tom Griffiths

Professor Pieter Abbeel

Fall 2013

The Prefrontal Cortex and Hierarchical Behavior

Copyright © 2013

by

Jennifer Sloan

## Abstract

### The Prefrontal Cortex and Hierarchical Behavior

by

Jennifer Sloan

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Jonathan D. Wallis, Chair

Every day, often without thinking about it, you create and achieve goals. These goals range in complexity from brushing your teeth in the morning, to avoiding traffic on your way to school, to figuring out how you're ever going to graduate from Berkeley. The intricate workings of your prefrontal cortex enable the planning and execution of such behaviors. For years, scientists have studied the mechanisms of simple choice behavior and learning in the form of stimulus- and action-outcome associations but the question of how increasingly complex and temporally-extended learning arises remains largely unknown. Informed by converging work from the fields of psychology and computer science, we set out to understand the computations performed at the level of single cells that may contribute to the ideation, mental maintenance, and stringing together of actions to perform a hierarchical gambling task. We recorded extracellular activity from two subjects as they performed an  $n$ -armed bandit task and compared and contrasted the information processed simultaneously in three distinct brain areas: the lateral prefrontal cortex, the orbitofrontal cortex, and the anterior cingulate cortex. While our data do not support the preferential processing of superordinate versus subordinate goals in prefrontal cortex, we did find many signals that may underlie hierarchical behavior. These include value and action encoding that depended on the hierarchical level, as well as encoding of past choices that could be used to chunk actions at the same level of the hierarchy. Understanding these mechanisms could help elucidate how the complex behavioral repertoire of the primate is implemented.

*Dedicated to everyone who has believed in me along the way*

## Acknowledgements

Sometimes the difference between doing and *not* doing is a single person letting you know they believe in you.

This is an acknowledgement of many, but certainly not all, of the people who have been *that* person for me.

**Dad**, who taught me my first lesson in behavioral economics, and arguably one of life's most important lessons: the value of short term pain for long term gain. Thank you for introducing me to computers and the internet *way* before it was cool or even *normal*, and for teaching me the mastery of spreadsheets and budgets when I might rather have been "saving the princess".

**Yatish**, who has not only championed me every step of the way but who has made this journey easier, more fun, and undoubtedly more successful than if I were to have traveled it alone. Sometimes doing a little dance makes everything better. Thank you Dr. Sloan \*wink wink\* for the hugs every morning, the kisses every night, and-- of course-- for complementing my lesser strengths. I am the luckiest.

**Mom** who sat with 8<sup>th</sup> grade me at the kitchen table night after night helping me struggle through algebra homework; knowing I was capable of better but had just missed a step along the way.

**Katy**, my "little" big sister, who taught me the value of secondary reinforcement by paying me with coupons I could exchange for escort across a busy street to the candy store as tender for letting her borrow my things

**Kati Markowitz**, for whom my gratitude and admiration is beyond words. You are so much to so many.

**Jon Wallis**, thanks for the second chance. You are a great teacher, an even better advisor, and I am lucky to have learned from you.

**Antonio Homero Lara** and **Steve Kennerley** for paving the Wallis Way before me.

**Alpha, Bravo, Echo, Lima, Papa** for the neurons in your brains and for clarifying that some really strange behaviors are not unique to human primates.

**Chung-Hay Luk**, who has been much more than a labmate but a true friend, cheerleader, and confidant. I am ever grateful for your help in lab; there was a serious void when you left.

**Erin Rich**, friend, labmate, editor, and MonkeyLogic co-commisserator.

**Amy Orsborn**, for being my gChat buddy, the best baker around, and an all-around badass. I think I may have communicated with you more than anyone else in my life up to this point. See you in New York.

**Erin, Amy, Ale, and Chung-Hay**: thanks for the Black Fridays, froyos, and cocktails.

**Jess Brown Redditt**, my cheerleader during the sprint to the finish. Thanks for the inspiring me in the gym and out.

**Meike Visel. Kirstie, Allyson, Chris**, and the matriculating class of 2007. **Heesoo, Liberty, Hania, and Noopur. Matar. Rev Dr Steve Sinha. John Flannery. Iain L. O. Buxton. Kelly Jensen.**

## **1. Introduction**

Much of human behavior is hierarchically structured. Even a simple act, such as making a cup of tea, can be broken down into multiple steps, each of which may be further broken down into requisite operations. Consideration of such hierarchical behaviors has been a cornerstone of many theories regarding the organization of the nervous system. For example, the proposed evolution of higher cognitive processes in humans from simpler instinctual responses in lower organisms informed colleagues Herbert Spencer's and John Hughlings Jackson's theories on the hierarchical organization of our nervous system (Wiest, 2012). Spencer, a biologist and philosopher, postulated that nervous tissue is constructed in layers with evolutionarily newer neural formations added on to more primitive areas (Spencer, 1855). Hughlings Jackson, a British neurologist, described the organization of behaviors to be resultant of this evolution such that each subsequent anatomical layer is responsible for more and more specialized behaviors as well as extended control over the lower centers and more instinctual actions (Jackson, 1884). Karl Lashley argued that, although temporally integrated actions sequences existed in insects and lower organisms, it was only in animals with cerebral cortices that such behaviors were displayed with any degree of complexity (Lashley, 1951).

Spencer also put forth a theory for the mechanism of learning by which these increasingly complex behaviors could be learned. Forming the basis for modern theories of reinforcement learning, Spencer first articulated that pleasurable sensations achieved as the result of motor actions make those actions more likely to be repeated in the future under similar circumstances (Spencer, 1855; Leslie, 2006). This thesis aims to build upon our current understanding of the neural mechanisms supporting reinforcement learning (RL) in hierarchical behavior.

### **1.1 Thesis overview**

The remainder of this chapter will be used to introduce the hierarchical structure of human behavior, what role reinforcement learning plays in that behavior, and how specific areas of the brain are structured to support such complex goal-directed behavior. The following chapter will describe the experimental methods, including the specific task used to study hierarchical behavior, details of training, and the equipment used. Chapter 3 will describe the response properties of neurons in the lateral prefrontal cortex (LPFC), an area that has been implicated in the organization of temporally extended behaviors, during the performance of the task. Chapter 4 will describe neural activity in orbitofrontal cortex (OFC) and anterior cingulate cortex (ACC), two areas that have been implicated in reinforcing behaviors. Finally, Chapter 5 will discuss how the current project has contributed to our understanding of hierarchical behavior and ways in which that understanding remains to be improved.

### **1.2 Hierarchical behavior**

As humans, we perform many complex behaviors every day of our lives. That is, we string together actions with short term objectives into longer sequences of actions with longer-term objectives until we have increasingly complex goals that can take minutes, hours, days or even

years to achieve (Schank and Abelson, 1977; Fuster, 1997). To understand this organization of operations, let us consider the example from Humphreys and Forde (Humphreys and Forde, 1998) of making a cup of tea (Figure 1.1). The superordinate goal of making a cup of tea may be broken down into the steps of putting a teabag into a teapot, pouring hot water into the teapot, pouring tea into a teacup, the addition of milk and tea if preferred, and stirring. Further, pouring tea into a teacup is composed of the subordinate actions of lifting the teapot, moving the teapot to the teacup, and tilting the teapot until tea comes out. One can even imagine that these actions are further broken down into motor programs, or sequential muscle flexions and extensions. In addition, the superordinate goal itself may also belong to higher-order goals (for instance, making breakfast).

The categorization of these actions into so-called chunks, scripts, nodes, or schemas provides a potential mechanism by which a sequence of events can be efficiently stored under a single representation (MacKay, 1987; Cooper and Shallice, 2000; Botvinick et al., 2009). There is some evidence to suggest that the prefrontal cortex (PFC) is critically involved in this process. Indeed, early clinical work described how PFC damage produced disorganized behavior. Wilder Penfield, who operated on his own sister to remove a frontal lobe tumor, described how she was cognitively unimpaired by the operation, with normal intelligence, language, memory and reasoning, and yet she could not perform relatively simple tasks, such as cooking a meal (Penfield and Evans, 1935). Her problem was specifically in organization: she would finish cooking some parts of the meal before others had even begun.

Another component of hierarchical behavior is the ability to keep the final goal in mind. The temporally extended nature of hierarchical behavior means that the final goal can be many steps from the initiation of the behavior. Here too, patients with frontal lobe damage tend to struggle. For example, Duncan and colleagues have reported the phenomenon of “goal neglect”, whereby patients with PFC damage disregard a task requirement, even though it has been understood and remembered (Duncan et al., 1996). The maintenance of the goal likely relies on working memory, a process that is known to be implemented by PFC (Goldman-Rakic, 1987; Miller and Cohen, 2001; Frank and O'Reilly, 2006). Working memory representations are both flexible and dynamic (Frank and O'Reilly, 2006), which would be advantageous for goal representation. Goals frequently need to be adjusted “on the fly” and they only need to be stored temporarily: once the goal is achieved it can typically be forgotten. Indeed, there is strong negative correlation between working memory capacity and the likelihood of experiencing goal neglect (Kane and Engle, 2003)[REF] demonstrating the relationship between the two processes.

Although the above findings are suggestive of a role for PFC in hierarchical organization, the precise contribution remains unclear. In addition, how different PFC subregions interact in order to implement hierarchical behavior is also unknown. The main aim of this thesis is to tease apart the contribution of different PFC areas to hierarchically-structured behavior by recording the electrical activity of single neurons in different PFC regions of awake monkeys while they are engaged in hierarchical behavior.



### 1.3 Reinforcement learning

A key feature of all behavior, whether or not it is hierarchical, is that the organism typically has to choose among multiple alternatives when deciding which course of action to pursue. We may opt to spend our day working to obtain a meal for sustenance or we might choose instead to pursue a mate to maximize our reproductive success. Each of these actions is valuable; one perhaps with more immediate utility than the other, but both serve to perpetuate our genes. While classic theories of behavioral economics (Von Neumann and Morgenstern, 1944; Kahneman and Tversky, 1979) establish heuristics that are useful in determining optimal choices, they don't provide an explanation for how humans and animals learn about and develop preferences for different actions and outcomes to begin with. In this section, we will review the main theories regarding how this is accomplished, and also look at some of the special challenges posed to learning theory by hierarchical behavior.

#### 1.3.1 Historical perspective

With the idea that positive outcomes influence subsequent choice behavior, Edward Thorndike, a contemporary of Herbert Spencer, developed his Law of Effect that spurred a field of research in behavioral conditioning using rewards and punishments to study the process of learning. Thorndike's classic experiments involved putting a cat in a closed box with a lever that, when pressed, opened a gate and enabled escape to food. Early in the experiment, cats tended toward instinctual behaviors like exploring, sniffing, and scratching around the box. Occasionally, by chance or curiosity, they happened upon the lever that released them to a tasty feast. Later in the experiment, after trials were repeated many times, the animals' behavior shifted such that less time was spent exploring before the lever was pressed. Learning could be roughly quantified by the difference in time the animal spent in the box from the beginning to the end of the experiment (Thorndike, 1911). Subsequently, Skinner trained rats to use a response lever to receive rewards (Skinner, 1930). The receipt of the reward "reinforced" the lever press, making it more likely that the animal would press the lever in future.

At around the same time, parallel experiments by the Russian physiologist Ivan Pavlov sought to more carefully elucidate the connection between an external stimulus and the resulting response (Pavlov, 1927). Pavlov's initial experiments examined the salivary reflex of dogs to a food stimulus. Pavlov realized that it was more beneficial for animals to respond to predictive cues of a noxious event (for instance that the certain odor or sound may indicate a predator) than it was to wait until a predator's teeth sank into their flesh. Pavlov characterized how we learn from the temporal pairing of sensory cues. His classic work, called Pavlovian Conditioning, was done by pairing a conditioned stimulus (CS) about which the animal had no prior knowledge or expectation (e.g. a bell), with an unconditioned stimulus (US) which the animal found rewarding (e.g. a piece of meat). Repeatedly pairing the CS and US would eventually elicit the same response to the CS as to the US. In other words, the animal would now salivate to the bell, rather than the piece of meat. Pavlov termed this a "conditioned response". This enabled Pavlov to measure the amount of learning that had taken place about the CS-US association, by measuring the strength of the conditioned response (e.g. amount of salivation to the CS).

Subsequent experiments showed that animals would only learn about a stimulus if it produced a surprising outcome (Kamin, 1969). For example, if the bell in the previous example were now accompanied by a light, the animals would not learn that the light predicted the meat. The bell already provided this information and so the delivery of the meat was not surprising. Learning about the light was “blocked” by the prior learning about the bell. Rescorla and Wagner (1972) put forth a prominent model to account for these findings. They proposed that the change in the strength of the CS-US association was proportional to the difference between what the animal expected to happen (the prior CS-US association) and what actually happened. More formally:

$$V_i^{n+1} = V_i^n + \alpha \left( \lambda_{US} - \sum_i V_i^n \right)$$

where  $V_i$  is the associative strength of conditioned stimulus  $i$  on trial  $n$ ,  $\lambda_{US}$  is the value of the unconditioned stimulus to the animal,  $\alpha$  is a learning rate parameter varying between 0 and 1. Learning is therefore proportional to the discrepancy between the value of the actual outcome ( $\lambda_{US}$ ) and the value that was predicted by summing across all predictive conditioned stimuli. Thus, as the animal learns, and the outcome becomes better predicted, less learning occurs.

As psychologists were working on the problem of understanding the mechanisms of complex learning, drastic technological advancements in computing allowed the creation of the field of artificial intelligence, leading scientists to develop algorithms for making smart machines. In fact, one of the most prominent computational frameworks for machine learning dovetailed nicely with the psychological models. Work done by Sutton aimed to allow naive hedonistic-guided machines to learn and navigate novel environments through trial and error (Sutton and Barto, 1981; Sutton, 1988). The system could make predictions about the best course of action and, based on further information gathered incrementally, update future predictions based on experience. There are many common threads between the problems of human learning and machine learning. Both paradigms (animal and machine learning, respectively) assume that there is a subject or actor that, in a given circumstance or state, performs an action with the intent of achieving a desirable outcome. If the outcome is as predicted, no learning has taken place, and behavior repeats the same in future. However, if the result is better or worse than expected, learning occurs and actions are adjusted on future attempts according to the feedback.

Indeed, Sutton made only two modifications to the Rescorla-Wagner model. First, to move away from the trial structure inherent in the Rescorla-Wagner model, and thereby improve the generalizability of the algorithm, predictions were made on the basis of successive time-points. The algorithm thus became known as the temporal difference (TD) model of RL. One problem with this modification, however, is that (unlike a trial) not every time point results in an outcome. Therefore the algorithm was modified so that it took into account, not just the

immediate reward, but also future potential rewards that arose as a result of the action. More formally:

$$\delta_t = r_t + \gamma V(S_{t+1}) - V(S_t)$$

where  $\delta_t$  is the amount of learning that should take place,  $r_t$  is the amount of reward received at time  $t$ ,  $\gamma$  is a discount factor, ensuring that future rewards are treated as less valuable than immediate rewards,  $V(S_{t+1})$  is the predicted value of future possible states that the actor can access, and  $V(S_t)$  is the value of the current state. In other words, the amount that an action should be reinforced is the difference between what the actor predicted would happen,  $V(S_t)$ , and the value of the thing that actually happened, i.e. the sum of the reward,  $r_t$ , and future potential rewards arising as a result of the action,  $\gamma V(S_{t+1})$ . This calculation is also referred to as a “reward prediction error” (RPE), since it reflects the error in the actor’s prediction as to the reward arising from the action.

In the 1990s there was great excitement in the field of neuroscience, when Wolfram Schultz and colleagues (1997) discovered that dopaminergic neurons in the brain encoded RPEs, providing information about how much better or worse the appetitive outcome was than predicted (Schultz et al., 1997). That is, dopaminergic neurons in the ventral tegmental area (VTA) and the substantia nigra (SNc) become more active (by firing more action potentials than baseline) when an appetitive event was better than the animal predicted, became less active when an event was worse than the animal predicted, and remained at a tonic level of activity when the event occurred as predicted. RL continues to be a prominent model for animal and human learning and RPEs have been increasingly well characterized across species and learning paradigms (Dayan and Niv, 2008; Lee et al., 2012; Roesch et al., 2012). In addition to those in the VTA and SNc, other signals related to RL parameters have also been discovered and characterized in other areas of the brain, most notably the striatum (Valentin and O’Doherty, 2009; Li et al., 2011; Diuk et al., 2013) and anterior cingulate region of the frontal cortex (Behrens et al., 2007; Kennerley et al., 2011).

### *1.3.2 Reinforcement learning and hierarchical behavior*

One difficulty RL faces in describing the learning of natural human behavior is the tenet that an agent learns to behave adaptively by exploring an environment and testing different courses of action in various environmental states or conditions and subsequently experiencing the related outcomes. When we consider that the number of possible environmental states is near infinite, and the number of possible actions in a given state is equally numerous, the combinatorial explosion of possibilities quickly becomes overwhelming. Given that time to learn increases with greater environmental complexity, the system of standard RL in natural settings is eventually infeasible (Botvinick et al., 2008). A possible solution to this problem is abstraction in terms of time or space. Spatial abstraction would mean collapsing across a number of inter-related environmental states (for example, specific locations in a single room) and treating them as equivalent and interchangeable (Botvinick, 2012). This reduces the space over which the agent is required to learn, making the problem more tractable.

The second related approach is the use of temporal abstraction whereby multiple actions are combined into subroutines. Similar to the example given earlier from Forde & Humphries of making a cup of tea is the subroutine of checking email (Botvinick, 2012). The chunking of actions – *open laptop, mouse to browser icon, double-click, enter URL, enter password* etc.— packaged into a single high-level action representation reduces the number of decisions required to solve a problem thereby making it easier to master (Figure 1.2). This hierarchical structure, applied to standard RL paradigms, has yielded the field of Hierarchical Reinforcement Learning (HRL) and is a useful model for how the brain may cope with learning in large spaces with many available options to pursue. The fundamental questions of how these subroutines are established and the neural mechanisms which may support this structured learning remain unanswered. The prefrontal cortex, heavily recruited for executive functions of working memory, the chunking of action sequences, and calculation of reward prediction error learning signals is uniquely poised to facilitate HRL and this thesis aims to elucidate the contributions of three distinct prefrontal areas.

#### **1.4 Prefrontal cortex**

PFC is located at the front of the frontal lobe, rostral to the adjacent premotor cortex as divided approximately by the arcuate sulcus in the monkey (Barbas and Pandya, 1989). Homologies between rodents and primates are difficult and exacerbated by the lack of sulci in rodents. One method for defining PFC across species is to locate the projection zone from the mediodorsal nucleus of the thalamus (Rose and Woolsey, 1948). Using this classification, PFC comprises approximately thirty percent of the human neocortex, a larger proportion of the cortex than in any other species (Fuster, 1997). PFC is the slowest brain region to develop relative to other cortical areas and is only fully mature after a person enters their twenties (Giedd et al., 1999; Fuster). Anatomically positioned for diverse information integration, PFC regions connect with cortical and subcortical motor areas, virtually all sensory areas, and with midbrain and limbic structures associated with the processing of reward, emotion, and memory (Barbas and Pandya; Fuster, 1997; Miller and Cohen, 2001). There are three gross subregions in PFC (Figure 1.3): orbitofrontal cortex (OFC), medial PFC, which includes the anterior cingulate cortex (ACC), and lateral areas (LPFC) and these areas are highly interconnected (Barbas and Pandya, 1989). The following sections will discuss medial, lateral, and orbital PFC subregions in more detail including functionality as well as morphology across species and reciprocal connectivity within the brain.

##### *1.3.2 Lateral prefrontal cortex (LPFC)*

Covering the lateral surface of the anterior end of the brain in monkeys and humans, LPFC has several subregions: areas 9 and 46 dorsally and 44, 45, and 47/12 ventrally. These areas are flanked by the frontopolar cortex (area 10) anteriorly and the frontal eye fields (area 8) posteriorly (Petrides and Pandya, 1999). This thesis will focus on areas 9 and 46, abbreviated LPFC, in the macaque monkey. This region is homologous to the human middle frontal gyrus (Petrides and Pandya, 1999). Because rodent PFC is agranular, it is difficult to identify a homologous region in rodents. However, Uylings and colleagues claim that dorsomedial

shoulder regions of the rat PFC are similar to the dorsolateral portion of PFC in primates (Uylings et al., 2003).

LPFC receives sensory and motor input and is well positioned to integrate complex stimuli information (Petrides and Pandya, 1999). Polysensory information enters from the superior temporal sulcus. Visual signals related to object recognition and location are received from inferotemporal (IT) cortex and V5/MT, respectively. Abstract visual information, including that related to body position in visual space, is projected to LPFC from posterior parietal cortex. Although LPFC is not directly connected to primary motor areas, it is interconnected with premotor cortex and frontal and supplementary eye fields, conveying high-level motor planning and preparatory activity rather than low-level muscle control commands. In contrast to its strong sensorimotor connections, LPFC has little direct limbic connectivity although it can process memory-related information via reciprocal connections with the hippocampus. Limbic information is also received from interconnections with OFC.

Petrides and Pandya (1999, 2002) have shown that there are anatomical subdivisions present in lateral prefrontal cortex such that the ventral and dorsal regions of dorsolateral PFC, as defined by the dorsal and ventral lips of the principal sulcus, connect differentially with distinct regions of parietal and premotor cortices and the cerebellum (Hoshi, 2006). Evidence has been accumulating to support functional divisions within LPFC as well (Hoshi, 2006; Koehlin et al., 2003; Badre et al., 2008; Badre, 2010; Hampshire et al., 2011). Using a simple behavioral task where monkeys were presented with visual cues instructing specific motor responses and then, following a short delay, visual choice cues were shown at various locations and animals could choose an action to achieve a positive outcome, Yamagata et al. (2012) characterized the differential contributions of neurons in dorsal and ventral DLPFC as well as dorsal premotor area. Dorsal DLPFC seemed to encode higher level information about the behavioral goal across the delay whereas neurons in ventral DLPFC encoded stimulus features of the cues as well as spatial information specifying the action to be taken, informing the differentiation of information processing along the perception-action hierarchy underlying goal-directed behavior. Another study examined the activity of single cells in LPFC during a cue-target association task. Here, Sigala et al. (2008) show a hierarchical representation in neural activity whereby responses of single cells in each phase of a task do not predict their response during other phases and instead show orthogonal activity, for example cue information in phase and target information in another. In addition to physiology studies in non-human primates, functional neuroimaging (fMRI) and lesion studies in humans suggest that there may be a hierarchical functional organization within the frontal cortices whereby the post posterior regions control direct, concrete motor responses and information becomes progressively more abstract, goal-oriented, and context dependent anteriorly. Koehlin and colleagues (2003) conducted an fMRI study confirming that the increasing cognitive demands of sensory, contextual, and episodic information engage premotor to caudal to rostral prefrontal regions accordingly. Badre and D'Esposito (2007) subsequently devised an fMRI experiment manipulating competition at four levels of abstraction from simple motor responses to contextual cue-to-dimension mappings, showing an increasing reaction time gradient in

addition to hemodynamic activation along the rostro-caudal axis corresponding with a cognitive representational hierarchy. Furthermore, patients with stroke-related lesions along the rostro-caudal axis show predictable deficits in the same task (Badre and D'Esposito, 2009).

All PFC areas appear to contain neurons that encode reward-related information (Rangel and Hare, 2010; Wallis and Kennerley, 2010). However, in contrast to ACC and OFC, the reward information encoded by LPFC seems to be more highly processed. For example, LPFC neurons are able to predict reward values by distinguishing stimuli on a categorical basis, independent of visual properties (Pan et al., 2008). Reward information in LPFC also interacts with working memory processes, for example, by increasing the precision with which information is stored in working memory (Kennerley and Wallis, 2009). LPFC neurons also show properties that would be useful for allowing reward to control hierarchical behaviors. For example, neurons in LPFC seem to maintain a trace of previous choice activity for several trials in the past in order to maintain an average rate of reward, which would be useful for controlling temporally extended behaviors (Seo et al., 2007). Related, lateral prefrontal activity has been shown to increase with strategic use of working memory. Daniel Bor et al. (2003) has shown using functional neuroimaging that when people employ methods of chunking in a memory task of spatial sequences, behavior improved and LPFC is more active than when no such strategy is used. In addition, LPFC may play a role in allowing higher-level goals to suppress more low-level reward information. For example, Rangel and colleagues showed that LPFC was activated in humans dieting in response to unhealthy food options presented in a functional magnetic resonance imaging (fMRI) study (Hare et al., 2009). The LPFC signal correlated with the strength of the value signal in ventromedial PFC, suggesting that LPFC may have worked as a mechanism to suppress the value information elicited by the unhealthy food.

In summary, LPFC displays many properties that make it ideally suited for the control of hierarchical behavior. Neuronal signals are frequently temporally extended, which may enable the representation of superordinate behaviors. In addition, reward signals in LPFC show the capacity to interact with high-level cognitive representations. However, at this stage, these ideas remain speculative; the current project aims to determine the precise contribution of LPFC.

### *1.3.2 Orbitofrontal cortex (OFC)*

OFC lies on the ventral surface of the PFC right above the eye orbits of the skull. OFC includes Brodmann areas 10, 11, 12, 13, and 14. Along the anterior-posterior axis, the cytoarchitecture of tissue goes from granular (anterior) to dysgranular to agranular (posterior) dependent on the prominence of granular cells in layer IV (Morecraft et al., 1992). Though human OFC shares similar cytoarchitectonic organization (Mackey and Petrides, 2010), rat OFC lacks clear homologous areas and is instead described anatomically as ventrolateral OFC, lateral OFC, and agranular insular cortex (Ongur and Price, 2000).

OFC has few connections with motor areas; indeed, it is the PFC subregion that is most poorly connected to the motor system. However, there are some connections between ventral premotor cortex and areas 12 and 13, and there are interconnections between OFC and LPFC

that may indirectly affect motor behavior (Carmichael and Price, 1995a). While these connections suggest a limited role in motor processing and execution, OFC has many more connections with sensory areas. Area 12 receives visual information from inferotemporal cortex, perirhinal cortex and regions in the superior temporal cortex relay polysensory input (Carmichael and Price, 1995a; Kondo et al., 2005). In addition, areas 12 and 13 receive tactile information about the face, hand, and forelimb from the anterior intraparietal area and secondary somatosensory cortex (Mountcastle et al., 1975; Petrides and Pandya, 1984), taste information from insula and opercular cortex (Carmichael and Price, 1995a), and olfactory information from the pyriform cortex (Carmichael and Price, 1995a). In addition to the vast amount of sensory information it receives, OFC also receives an array of signals related to emotion and reward-related activity. Amygdala, hippocampus, temporal pole, entorhinal, perirhinal, perihippocampal, and cingulate cortices all connect with areas 11, 13 and 14 (Carmichael and Price, 1995b). Overall, though the OFC receives a relative lack of motor information, the area is very rich with limbic and sensory information.

OFC appears to play an important role in value-based decision-making. For example, patients with damage to OFC exhibit choices that are inconsistent with their subjective preferences (Camille et al., 2011b). OFC neurons respond to valuable stimuli in the environment (Rolls, 1996; Schultz et al., 2000; Wallis, 2007, 2012), encode both positive and negative expected outcomes (Morrison and Salzman, 2009) and reflect the value of one reward relative to others (Padoa-Schioppa and Assad, 2006, 2008). Such signals may underlie the role OFC plays in decision-making. Further, the strong limbic input to OFC, as well as its strong connections with all sensory modalities, place it in an ideal location for learning stimulus-outcome associations. OFC lesions impair Pavlovian conditioning in rats but leave instrumental conditioning unaffected (Ostlund and Balleine, 2007). Although intact on a range of neuropsychological exams, patients with OFC damage are unable to cope with stimulus-outcome reversals such that they continue responding in favor of a once-rewarding stimulus even though it may no longer be rewarding (Rolls et al., 1994).

Although much of the recent literature has focused on the role of OFC in encoding reward information, there is also evidence that it plays a role in more cognitive processing. For example, OFC neurons are able to encode high-level, abstract rule information (Wallis et al., 2001). In addition, although working memory processes are more commonly ascribed to LPFC rather than OFC (Bechara et al., 1998), recent evidence has shown that OFC neurons can hold information about rewards in working memory (Lara et al., 2009). In summary, OFC has both the anatomical connections as well as the functional properties to make an important contribution to processing reward information and using reward information to guide behavior, and those contributions could include enabling reward information to interact with high-level cognitive processes.

### *1.3.3 Anterior cingulate cortex (ACC)*

ACC is located in the cingulate sulcus on the medial wall of PFC and consists largely of area 24. In humans, ACC is comparable to monkeys with one difference: human area 24 in the cingulate

sulcus has noticeably larger pyramidal neurons in layer V (Nimchinsky et al., 1996). Medial PFC in rats is more rudimentary than that of monkeys and humans, with simpler cell morphology and fewer divisions (Ongur and Price, 2000). Most primate electrophysiology studies of ACC focus on the region surrounding the portion of the cingulate sulcus anterior to the genu of the corpus callosum (Matsumoto et al., 2003). This is, in part, because this region is most accessible for electrophysiological studies. It lies close to the surface of the brain and is sufficiently lateral to avoid any accidental contact with the central sinus. This is also the region that we will focus on in the current thesis.

ACC has strong connections with limbic and motor-related areas. Of the limbic areas, the amygdala, a key structure for processing affective value and emotion, connects strongly with all part of ACC (Carmichael and Price, 1995b). ACC is also the region of frontal cortex with the heaviest dopaminergic input (Williams and Goldman-Rakic, 1993). As for motor connectivity, ACC strongly connects to the area immediately posterior to it - the cingulate motor area (CMA) - which has direct projections to the spinal cord controlling movements of the arm and leg (Dum and Strick, 1991, 1996). CMA is also connected with the supplementary motor area (SMA), and together their spinal projections make up 40% of all corticospinal projections in the frontal lobe (Dum and Strick, 1996). In contrast, ACC has few connections with sensory areas (Carmichael and Price, 1995a). Recent findings, using diffusion tensor imaging, have shown that ACC has the same pattern of connections in the human as in the monkey (Croxson et al., 2005).

Consistent with its connectivity to limbic areas, ACC neurons encode information about many different aspects of rewards, including their size, probability of delivery and how much work was required to earn the reward (Kennerley et al., 2009), as well as information about negative outcomes (Sallet et al., 2007; Seo and Lee, 2009). Although ACC and OFC share similar limbic connections, and similar responsivity to rewards, their pattern of connections suggests that they are part of two very separate networks performing different functions. Areas in the medial wall tend to connect with one another, but only have weak connections with areas in OFC, while areas in OFC tend to connect with one another, but only have weak connections with areas in the medial wall (Figure 1.4). These anatomical findings have led to the suggestion that there are two distinct limbic networks in frontal cortex (Carmichael and Price, 1996): the medial network and the orbital network.

Given the existence of these two networks, there has been speculation as to what the difference is in their function. One possibility, which would be consistent with the anatomical connections of the two networks, is that OFC is important for associating stimuli with the rewarding outcomes they predict while ACC is important for associating actions with rewarding outcomes (Rushworth et al., 2007). Related ideas have been proposed in the decision-making literature. OFC is argued to be responsible for assigning values to sensory stimuli in the environment, thereby enabling the organism to efficiently make choices between different goods. This decision space is argued to be independent of the action necessary to acquire those goods. In contrast, ACC is argued to calculate the value of actions by integrating information about the action and the object to which the action is directed. There has been considerable debate about whether decision-making occurs solely in the realm of the goods space



(Wunderlich et al., 2010; Padoa-Schioppa, 2011), the action space (Kawagoe et al., 1998; Platt and Glimcher, 1999; Roesch and Olson, 2003) or requires the two systems to operate in parallel (Cisek and Kalaska, 2010; Luk and Wallis, 2013).

Although there is neuropsychological evidence to support these distinctions in humans (Camille et al., 2011a), monkeys (Rudebeck et al., 2008) and rats (Balleine and Dickinson, 1998; Pickens et al., 2003; Ostlund and Balleine, 2007), the evidence at the single-neuron level is more mixed. OFC neurons in monkeys typically encode the value of predicted outcomes rather than the motor response necessary to obtain the outcome (Tremblay and Schultz, 1999; Wallis and Miller, 2003; Padoa-Schioppa and Assad, 2006; Ichihara-Takeda and Funahashi, 2008; Abe and Lee, 2011), but there have been some notable exceptions (Tsujiimoto et al., 2009). Furthermore, robust encoding of actions has been seen in rat OFC (Feierstein et al., 2006; Furuyashiki et al., 2008; Sul et al., 2010; van Wingerden et al., 2010). With regard to ACC, many studies have emphasized the role it plays in predicting the outcome associated with a given action (Ito et al., 2003; Matsumoto et al., 2003; Williams et al., 2004; Luk and Wallis, 2009; Hayden and Platt, 2010), but there have also been studies showing ACC neurons encoding the rewards predicted by stimuli (Seo and Lee, 2007; Kennerley et al., 2009; Cai and Padoa-Schioppa, 2012).

Further yet, there is substantial and growing evidence that implicates ACC in reward prediction error signaling, including data to suggest that separate populations of neurons encode positive and negative errors similarly (Matsumoto et al., 2007; Kennerley et al., 2011; Sallet et al., 2007). In addition to, and in contrast with, standard RPE models where “good” events trigger positive RPEs and “bad” events trigger negative RPEs, it has been suggested that the cellular activity in this region may actually signify the occurrence of an unexpected outcome positively (+ RPE) or a non-occurrence of an expected outcome negatively (RPE -) regardless of whether the outcome is affectively positive or negative (Alexander and Brown, 2011). The reward prediction error activity may also be dependent on task phase or epoch (Kennerley et al., 2011; Sallet et al., 2007) and can occur in response to a stimulus announcing a reward discrepancy before the actual reward has even been dispensed (Sallet et al., 2007). Though results conveying the types of information encoded in ACC are diverse and can be somewhat difficult to interpret in light of each other, that it receives heavy dopaminergic input and makes connections with limbic and motor areas, together with its documented role in action-outcome monitoring, ACC is uniquely poised to convey learning signals in complex behavioral environments such as the present hierarchical experiment where reward is manipulated on multiple levels.

#### **1.4 The role of PFC in hierarchically structured behavior**

Sometimes called the central executive, or association cortex, many prefrontal cortical areas receive information from two or more sensory modalities and project to multiple supplementary and pre-motor regions. This interconnectivity makes PFC suited for integrating and computing information to create abstract, high-level representations, to determine the optimal response to an environmental situation. Such high-level representations would be useful for a diverse array of cognitive processes, including memory and attention, emotional

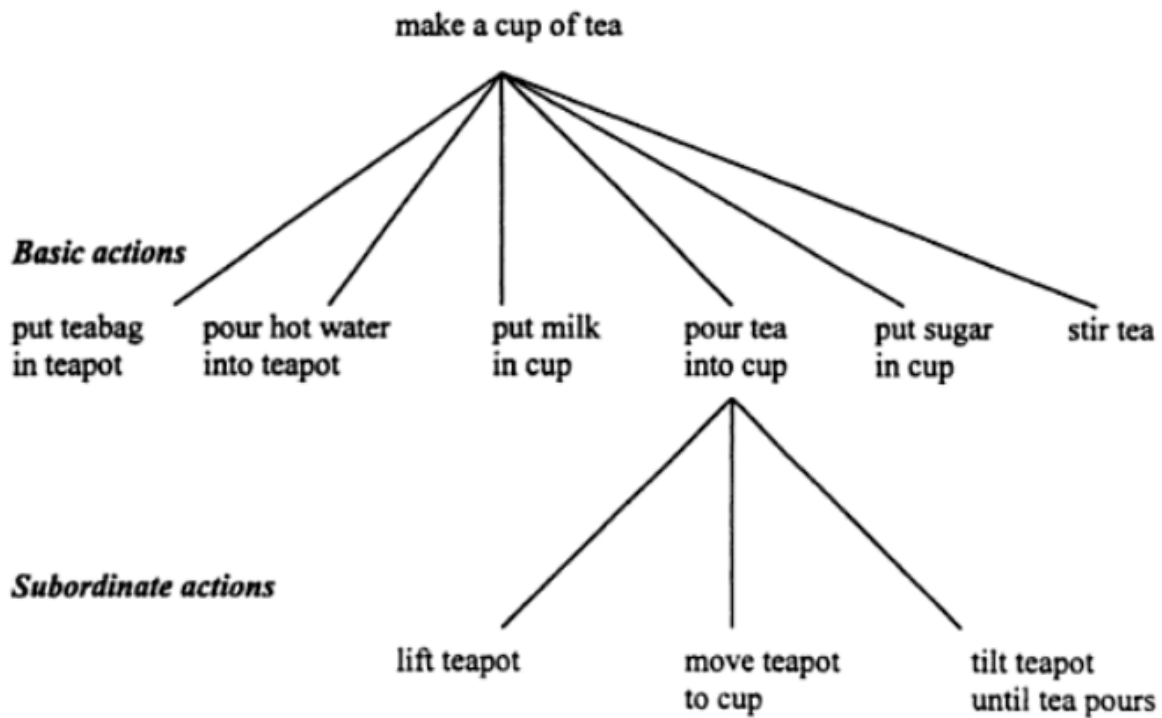
and social processing, as well as planning and goal-directed behavior. Information in PFC may be encoded and maintained over time for working memory and used to select appropriate actions, while inhibiting others, in the pursuit of reward and learning.

As discussed above, there is clear evidence implicating PFC in hierarchical behavior, befitting an area that is at the apex of perception-action cycle. However, as also made clear above, the different PFC areas are unlikely to perform the same function or make the same contribution to hierarchical behavior. The nature of these different contributions are not clear though. The aim of this thesis is to specify these contributions. We designed a primate version of a hierarchical task, and trained two monkeys to perform it. Each trial began with a choice, which we termed the superordinate choice. This choice determined how much juice reward the animal would receive at the end of the trial. However, before the animal received the juice, he had to perform a series of subordinate choices. These choices did not affect the final amount of juice received, but the animal was incentivized to perform that as efficiently as possible, since the more efficiently he performed them, the quicker he would get to the final reward. We then recorded electrical activity from single neurons in LPFC, OFC and ACC while the animal performed the task. Based on the previous evidence summarized above, we proposed different functions of the three different areas.

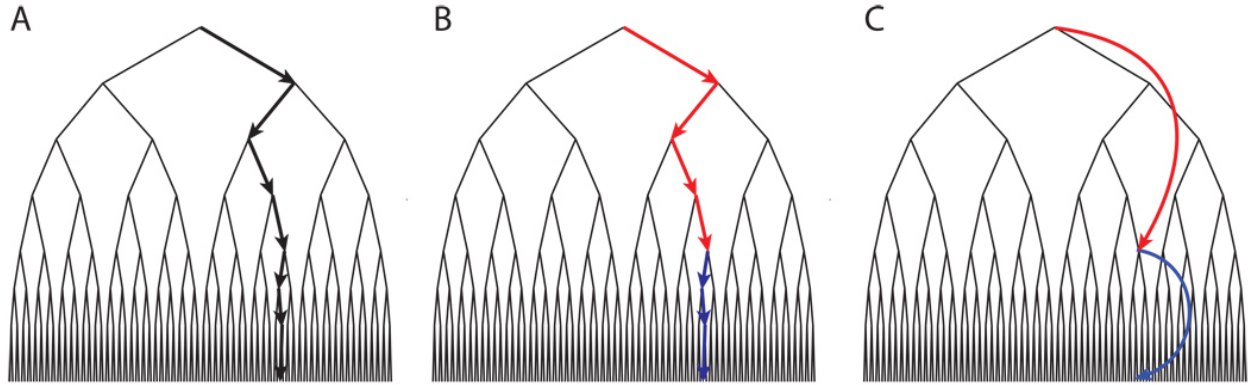
Given the evidence that LPFC is involved in the representation of hierarchical behaviors and has the capacity to maintain information in working memory, we predict that LPFC neurons will encode superordinate motor responses and maintain that information through until the time of reward delivery, even though there will be intervening subordinate motor responses. We know that OFC is important for tracking stimulus values, and so we expect that the activity of OFC neurons will encode value information for low-level subordinate choices. However, given the role of OFC in encoding reward information in working memory, it may also be responsible for maintaining information about the value of the chosen superordinate choice until the time of reward delivery. Such a signal would be important for the animal to learn what the best superordinate choice in terms of yielding the maximum amount of final reward. Finally, given the role of ACC in encoding RPEs, we predict that ACC will encode RPE activity for both subordinate and superordinate choices. Indeed, the final delivery of reward will generate an RPE simultaneously for both levels of choice. To date, RPEs have only been studied using tasks that generate a single RPE, and so how neurons encode multiple RPEs is unknown. One possibility is that ACC will contain different populations of neurons responsible for encoding RPEs from different levels of the hierarchy.

Chapter 3 will focus on LPFC and its role in encoding superordinate and subordinate motor responses. Chapter 4 will focus on OFC and ACC and their role in encoding the value of the choices as well as the final delivery of the reward. Chapter 5 will examine outcome and reward prediction activity in LPFC, OFC, and ACC. Chapter 6 will conclude the thesis with a discussion of the results and suggestions of future directions.

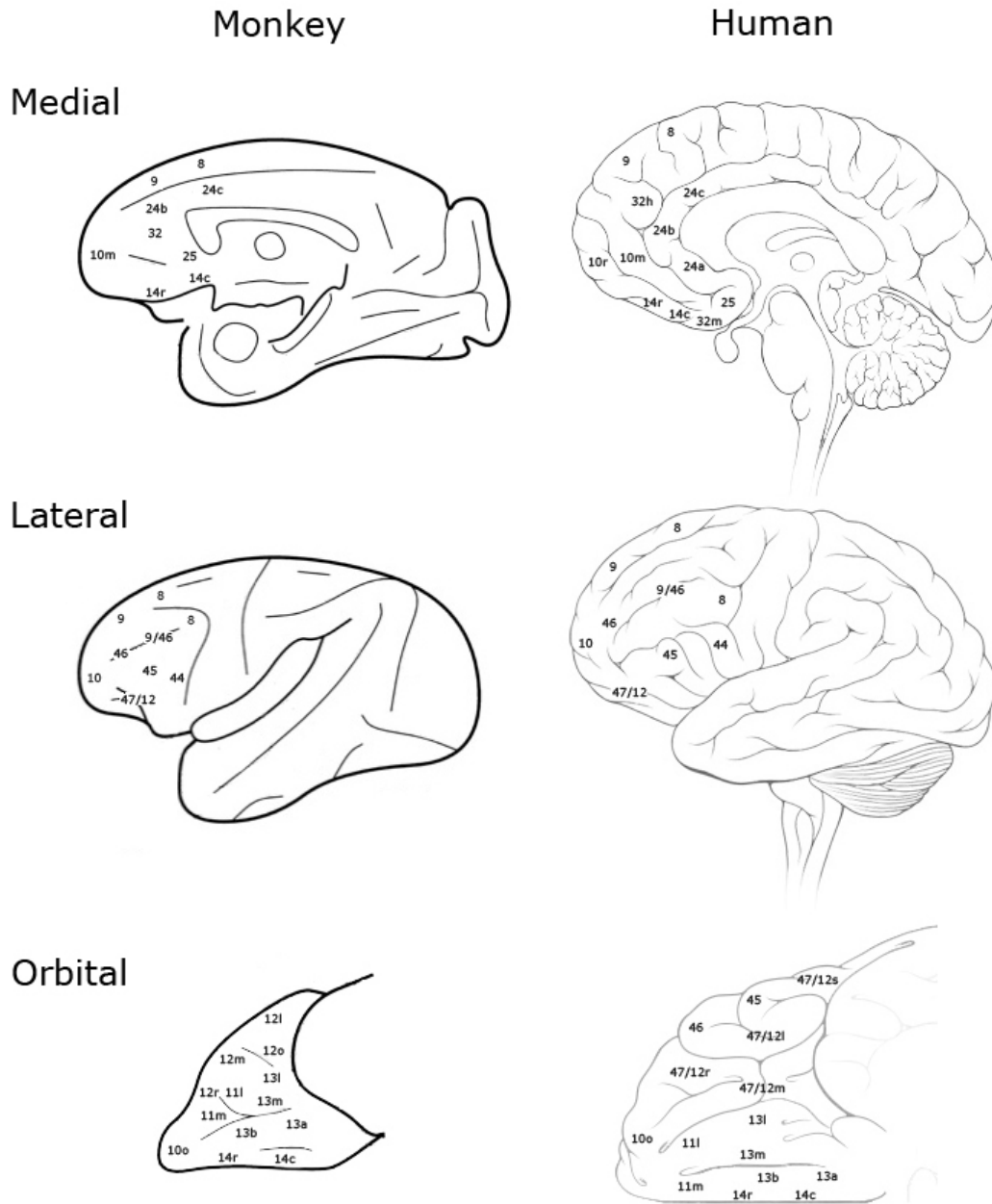
***Superordinate actions***



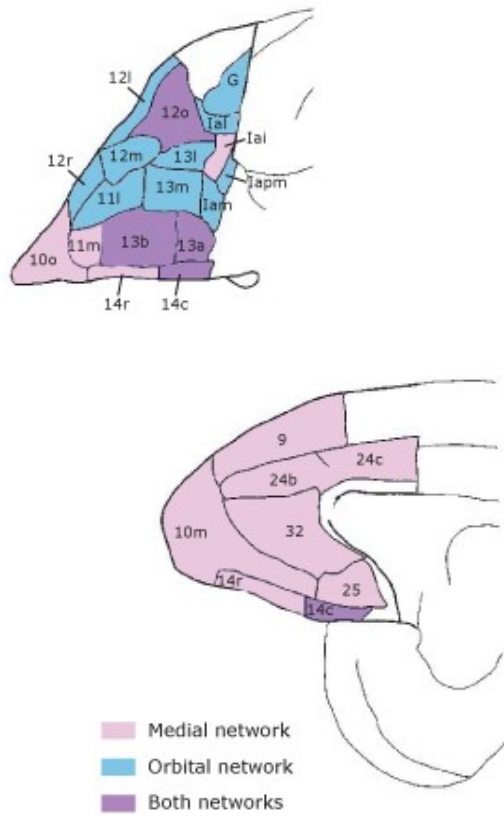
**Figure 1.1** The superordinate task of making a cup of tea may be broken down into subordinate actions. From Humphreys and Forde, 1998.



**Figure 1.2** Panel A shows a standard decision tree where each arrowhead signifies a point at which the agent needs to choose an action from all those available to pursue in order to reach the end goal. Panel B illustrates the benefit of aggregating the first four (red) and second three (blue) actions into single subroutine options. Panel C shows the behavioral trajectory taken from an original 7 steps down to two, an efficient reduction in the decision space. From Botvinick, Niv, and Barto (2009).



**Figure 1.3** The medial, lateral, and orbital surfaces of the prefrontal cortex. Monkey outlines taken from Carmichael and Price (1996) and Petrides and Pandya (1999); human outline taken from Ongur and Price (2000). Adapted and re-printed with permission from Luk (2011).



**Figure 1.4** Medial and orbital networks of the prefrontal cortex taken from Carmichael and Price (1996). Adapted and re-printed with permission from Wallis (2012).

## 2. Methods

### 2.1 Overview

All experimental methods and techniques are described herein.

### 2.2 Behavioral training materials and methods

#### 2.2.1 Subjects

Two rhesus monkeys (*Macaca mulatta*), subjects L and P, were used for this experiment. At the time neurophysiological recordings were taken subject L was 3 years old and weighed 10.1 kg and subject P was 4 years old and weighed 9.5kg. Subjects were housed in pairs when possible as part of a 5 animal colony. Subjects were fed two times per day, provided with behavioral enrichment, and experienced a 12-hour light cycle beginning at 7am daily. Subjects' fluid intake was regulated in order to motivate their participation in the study. All procedures were in accordance with the National Institutes of Health guidelines and the recommendations of the University of California Berkeley Animal Care and Use Committee.

#### 2.2.2 Equipment

Subjects performed tasks seated in a primate chair facing a 19-inch LCD computer screen placed 50cm from the chair. A system of computers controlled the display of behavioral events (Figure 2.1). These computers utilized Monkeylogic (<http://www.monkeylogic.net/>), a toolbox running in conjunction with Matlab (<http://www.mathworks.com/products/matlab/>), for the design and execution of psychophysical tasks. The central Monkeylogic control computer sent commands via a COM port to a receiving computer which then presented the visual stimuli on an LCD monitor. The stimuli were mirrored via a video splitter onto a third monitor in the sound-attenuation box where the monkeys sat. Mirroring the presentation of the task allowed us to monitor exactly what the subjects saw without disturbing them as they worked. The Monkeylogic control computer ran timing routines and interfaced with various external devices via a PCI-6229 data acquisition (DAQ) card (National Instruments, Austin, TX). Each behavioral event in the trial was marked with a code that was sent as an 8-bit number from the DAQ to the multichannel acquisition processor (MAP). The MAP systems read in this number and recorded its value along with a timestamp of when it occurred. Its timestamp was stored along with neurophysiological data in a single '.plx' data file. The Monkeylogic control computer ran with a single interrupt routine that triggered every millisecond and updated both a software clock and sampled all data lines. Thus, the control of the behavioral contingencies, the presentation of visual stimuli, and the monitoring of behavioral event all took place with single millisecond resolution. Visual stimuli in the task were isoluminant as measured by the Spyder luminance meter (Datacolor, Lawrenceville, NJ).

During initial training, when subjects used a manual joystick to respond, choices were registered by using two 4-TPS-E1 Touch Sensor Modules (Crist Instrument, Damascus, MD)

connected to the digital input port of the DAQ card. The touch sensors were contact-sensitive devices designed to send a 5-volt TTL pulse when a grounded subject touched it. Actions were executed using custom-made joysticks that connected to the Touch Sensor Module.

Once subjects learned to register their choices using a joystick, they were then trained to use eye movements (saccades) to select from available stimuli. Eye position was recorded using an infrared eye monitoring system (ISCAN, Burlington, MA). An infrared camera focused on the subject's eye and visualized the results using proprietary image tracking software. The software tracked the center of the subject's pupil as X and Y coordinates as well as the pupil diameter. These three measures were fed separately to three DAQ analog input channels and recorded for the duration of the session.

Juice rewards were delivered by commands from the DAQ analog output ports to the juice pump. The ISMATEC-IPC8 peristaltic juice pump (ISMATEC SA, Glattbrugg, Switzerland) took a 0-5V TTL pulse that delivered a voltage-dependent volume per unit time through polymer tubing which ended in a custom made mouthpiece positioned near the animal's mouth.

### *2.2.3 Behavioral training*

Subjects were trained to perform the behavioral task using positive reinforcement. Sitting in front of a video monitor, subjects used eye movements to choose from available visual stimuli in order to obtain a liquid reward mixture of 50% water and 50% apple juice. Subjects were able to work on the task until receiving as much reward as desired. Once the subjects learned the task, neurophysiological recording sessions began and were typically carried out for five or six days a week.

### *2.2.4 Behavioral task*

A depiction of the behavioral task is shown in Figure 2.5. Animals were required to make two decisions in this task. After completing an initial fixation period of 500 ms, a new trial began with the presentation of the superordinate choice. These two stimuli were chosen from four possible stimuli at random and each indicated a volume of a juice reward the animal would receive upon successful completion of the trial. The red fixation cue turned green and the animal indicated his choice by maintaining visual fixation on the stimulus for 500 ms. Prior to the subordinate choice presentation, the animal is again required to fixate centrally (300 ms). Once the fixation requirement was satisfied, two out of four possible stimuli were presented to the animal and the fixation cue turned from red to green, allowing the animal to proceed with maintaining visual fixation on the stimulus of choice for 500 ms. Each stimulus represented a probabilistic “win” to the animal as indicated by a secondary reinforcement meter presented centrally around the GO cue. The subordinate stimuli had 10%, 35%, 60%, and 85% probabilities of satisfying a “win” requirement and two wins were necessary before the animal was able to receive the liquid juice reward whose volume pertained to the superordinate selection. Juice rewards were delivered over a 1600 ms period and volume was adjusted by the speed of a solenoid pump.



## 2.3 Neurophysiological techniques

### 2.3.1 Isolation of recording sites

To record from the brain areas of interest, we began by placing recording chambers on the overlying skull. Magnetic resonance images (MRIs) of the subjects' brains were taken at the U.C. Davis Center for Imaging Sciences with a 1.5 Tesla scanner prior to the animals' arrival at Berkeley. Those digital images were then imported into commercial graphics software (Adobe Illustrator CS5, San Jose, CA) where stereotactic coordinates for chamber placement were calculated. Correspondence between the MRI scans and electrode placement were verified during recording sessions by mapping the location of sulci and the boundaries of gray and white matter.

Alternate areas were recorded from both brain hemispheres were in each subject. In subject L, ACC and LPFC were recorded from the left hemisphere. This chamber was centered at 25mm anterior of the interaural line (i.e. AP 25) and 6.9mm lateral of the mid-sagittal plane (i.e., LM 6.9) on the skull and angled at 20 degrees outward from vertical. LPFC and OFC were recorded from the right hemisphere in subject L and this chamber was similarly centered at AP 25 and 7.5 LM and angled 12 degrees outward from vertical. In subject P, activity in ACC and LPFC was recorded from the right hemisphere. This chamber was centered at 27mm anterior of the interaural line and 17.7 mm lateral of the mid-sagittal plane on the skull and angled at 20 degrees outward from vertical. LPFC and OFC were recorded from the left hemisphere in subject P and this chamber, angled 10 degrees lateral of vertical, was similarly centered at AP 27 and 17.7 mm LM. (Figures 2.2 to 2.4)

### 2.3.2 Surgery

Subjects underwent an initial surgery to implant a custom-made titanium head-positioning post secured with titanium orthopedic screws. This post kept the subjects head immobile to allow for eye movement tracking as well as head stabilization for electrode recordings. For each animal, after determining the position for chamber placement, two surgical operations were performed: one to implant the chambers and one to make craniotomies through which the recording electrodes enter the brain. Cylindrical titanium recording chambers were secured to the skull using bone cement and titanium screws. Subsequent craniotomies were made within the chambers to allow access to the underlying brain tissue. Chambers were covered with polypropylene CILUX caps from Crist Instrument (Hagerstown, MD) to prevent contamination and infection, as well as to discourage granulation tissue growth.

For surgical procedures, anesthesia was induced with ketamine intramuscularly (10mg/kg IM). Xylocaine or lidocaine spray (14%) was used as a local anesthetic to facilitate intubation. Anesthesia was maintained with isoflurane (2-4%). Depth of anesthesia and vital signs were steadily monitored by the surgical team including heartbeat (90-150 beats per minute), respiration (17-23 breaths/min), body temperature (36-39 degrees celsius) and blood oxygen saturation (>85%). Lactated Ringer's solution was infused intravenously (2-4 ml/kg/hr) to ensure the monkey remained sufficiently hydrated and heating pads and clothing were placed

under and on the body to maintain temperature. Following surgery, gas anesthesia was discontinued, and once an animal showed signs of recovery, the animal was extubated and given buprenorphine at a dose of 0.01-0.03mg/kg subcutaneously (SC) or IM for post-operative pain relief. Animals were checked upon and carefully monitored at least every half hour, the period of which was incrementally lengthened as the animal recovered from anesthesia. After the initial recovery, the animal was checked several times per day at which time appropriate analgesics and antibiotics were administered. Typically, buprenorphine (0.01-0.03mg/kg SC or IM) was administered 2-3 times per day for 2-5 days. All appropriate measures were taken to minimize pain and choice and dose of analgesic were made in consultation with veterinary staff.

### *2.3.3 Recordings*

Neuronal activity was recorded from three brain areas: anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), and lateral prefrontal cortex (LPFC). Tungsten electrodes (FHC, Bowdoin, ME) attached to custom-designed screw microdrives were lowered so as to record from multiple brain areas simultaneously on a given day. The microdrives were mounted to a custom-made plastic grid containing an array of 24-gauge holes spaced 1mm apart, which allowed electrodes to be lowered independently to varying depths. Stainless steel 24-gauge thin-wall hypodermic needles (Terumo, Somerset, NJ) were glued to the bottom of the grid such that the beveled tip of the needles pointed out of the grid. These tips served to puncture the dura mater and guide the electrodes to the desired recording location. Electrodes were lowered manually by handheld screwdrivers to MRI-informed depths. As an electrode approached a cellular layer we slowed the lowering and stopped when we isolated single neurons. After neurons were located on all or most recording channels we waited 1-2 hours for the brain to settle to ensure maximum stability during the recording session which lasted 1.5-2.5 hours. Neuronal drift was seldom a problem. Neurons were sampled randomly to ensure fairer comparison of neuron properties between the different brain regions. During the recording session, no changes to the channels were made and distractions to the subject were minimized.

To minimize the chance of infection chambers were cleaned both before and following all recording sessions. Cleaning began by removing the cap (or grid, for cleanings following recordings) and sterilizing the chamber exterior with alcohol. Next, under sterile conditions, the inside of the chamber was flushed with sterile saline, then with a mixture of povidone (an iodine-based antiseptic) and saline, followed by a final saline flush. Tissue was dabbed dry with a sterile gauze or cotton swab. If cleaning was prior to recording, the plastic grid with microdrives and electrodes was fitted on top of the chamber. The grids used for recording were sterilized in Cidex, a glutaraldehyde solution, overnight. If the cleaning was performed after recording, a new sterilized cap was secured on top of the chamber.

### *2.3.4 Materials and methods for neurophysiology*

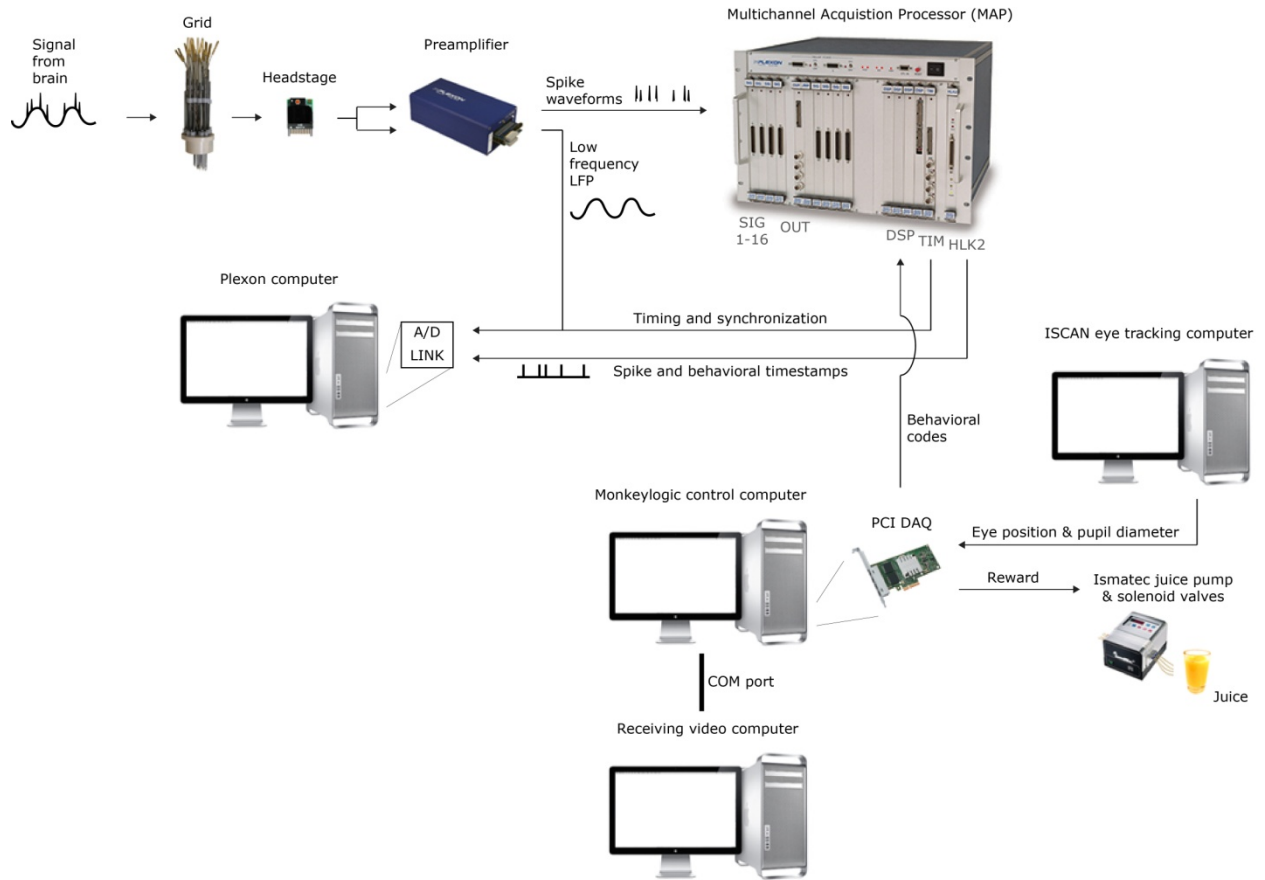
Voltage signals were taken from the tip of the electrodes with respect to the reference, the dead-positioning post affixed to the subject's skull. Those neuronal signals were recorded and

amplified via hardware and software from Plexon, Inc. (Dallas, TX) as shown in Figure 2.1. The signal was amplified 20 times from an op-amp based circuit in the headstage, which connected directly to the electrodes. The signal was then further amplified 100 times through a preamplifier and filtered for spikes in the 100 Hz – 8 kHz frequency band and local field potentials (LFPs) in the 1 Hz – 300 Hz frequency band. Signals were then processed by a Multichannel Acquisition Processor (MAP system) for further amplification and filtering.

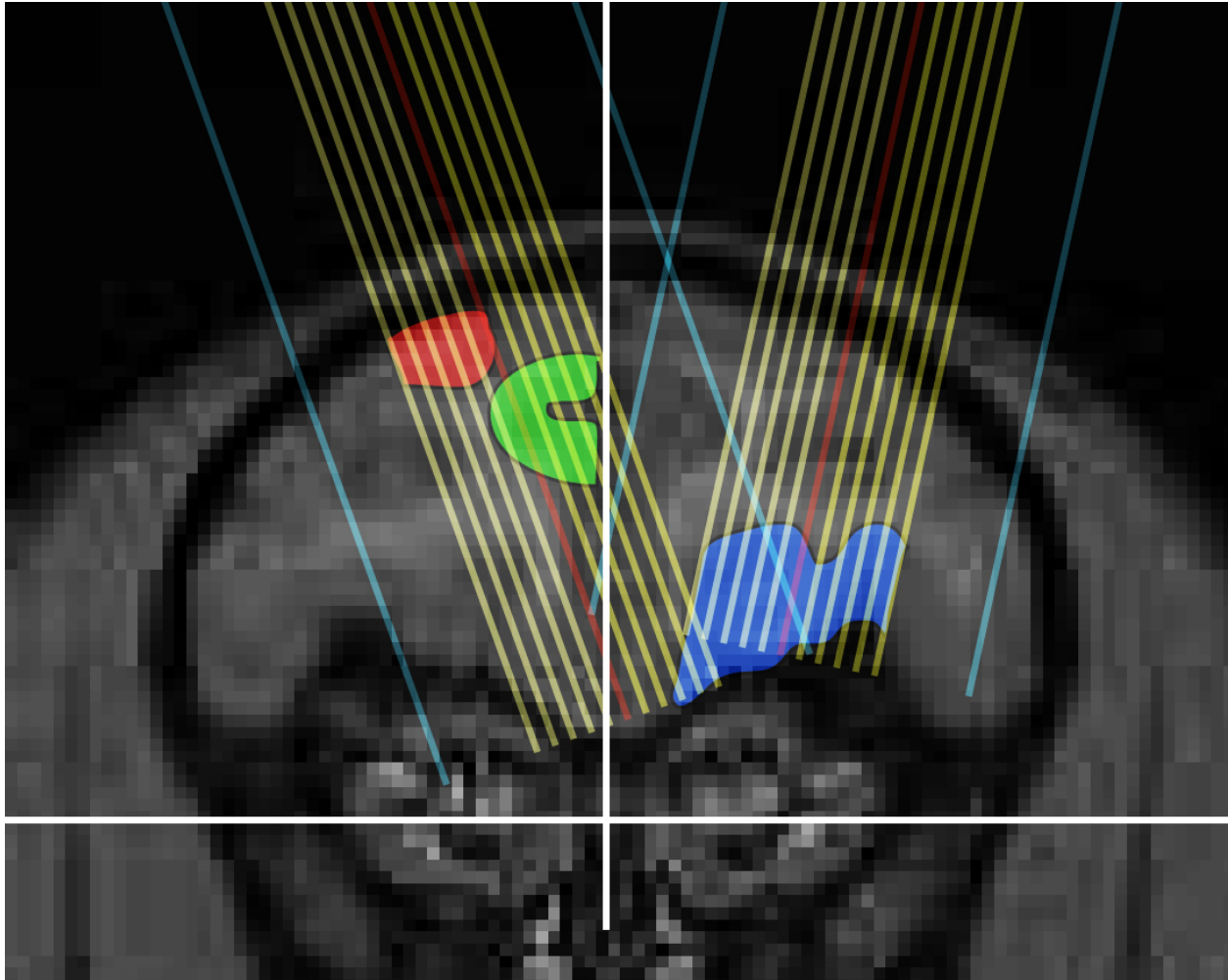
Spikes and LFPs were digitalized at 40 kHz with 12-bit resolution per channel. For the spikes, voltage thresholds were set to ensure that the neuronal signals were a minimum of 4 standard deviations above the background noise (calculated over a 10s period immediately prior to recording.) When the spike waveform crossed the manually set threshold, the program recorded the time stamp of the threshold crossing. Waveforms and voltage fluctuations that did not cross the threshold were discarded. The digitized waveforms were then sorted offline using Offline Sorter software (Plexon, Inc., Dallas, TX). This constructed 2D or 3D plots of a subset of 12 waveform features including the first three projections from principal components analysis, peak-valley differences and widths, and waveform energy. From those 2D or 3D plots, clusters of waveforms were grouped together manually and classified as single units (Figure 2.6). We ensured the separation of neuronal waveforms by rejecting channels where more than 0.1% of the waveforms were separated by intervals of less than 1.5 ms or where neuronal “drift” occurred. Typically, approximately 15% of the channels were discarded.

## **2.4 Statistical analysis**

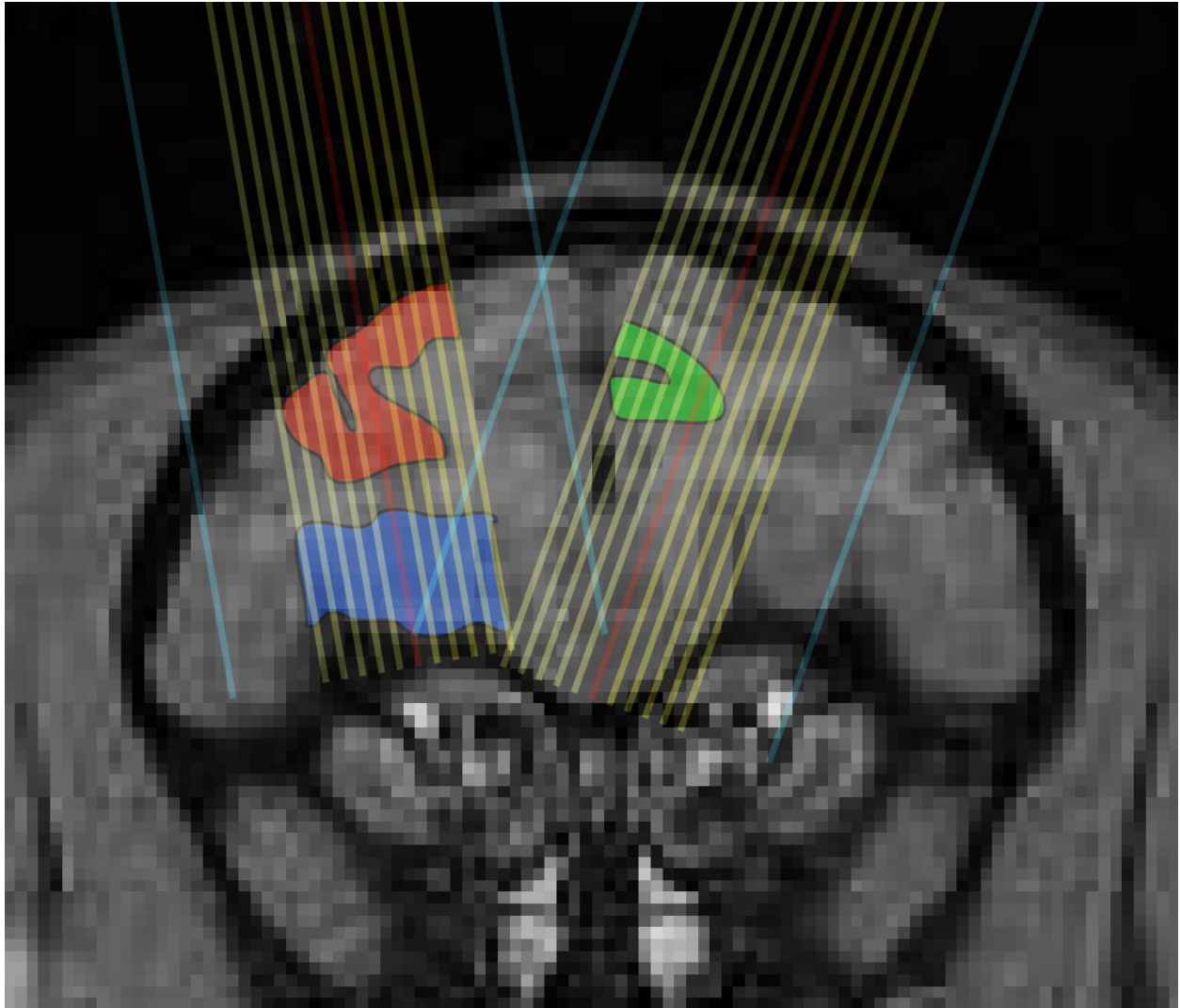
We used MATLAB (MathWorks, Natick, MA) to perform all analyses. Analyses were restricted to successfully completed trials. To characterize selectivity of a neuron, we first calculated its mean firing rate in each trial during a defined time epoch. We compared differences in firing rate between experimental conditions with the null hypothesis that the neuron did not encode a given type of information. We specified the independent variables and statistical test in our description of the experimental results. Once we had classified the neurons according to the type of information they encoded, we assessed differences between brain areas in the prevalence of these different types of neurons using chi-squared tests. We specify the independent variables and particular statistical analyses are described in detail in the results chapter.



**Figure 2.1** Signal acquisition system



**Figure 2.2** Structural magnetic resonance image (MRI) from Animal L showing chamber placement (light blue parallel lines) and possible electrode tracks in yellow. Brain areas recorded from are highlighted in red (LPFC), green (ACC), and blue (OFC).



**Figure 2.3** Structural magnetic resonance image (MRI) from Animal P showing chamber placement (light blue parallel lines) and possible electrode tracks in yellow. Brain areas recorded from are highlighted in red (LPFC), green (ACC), and blue (OFC).

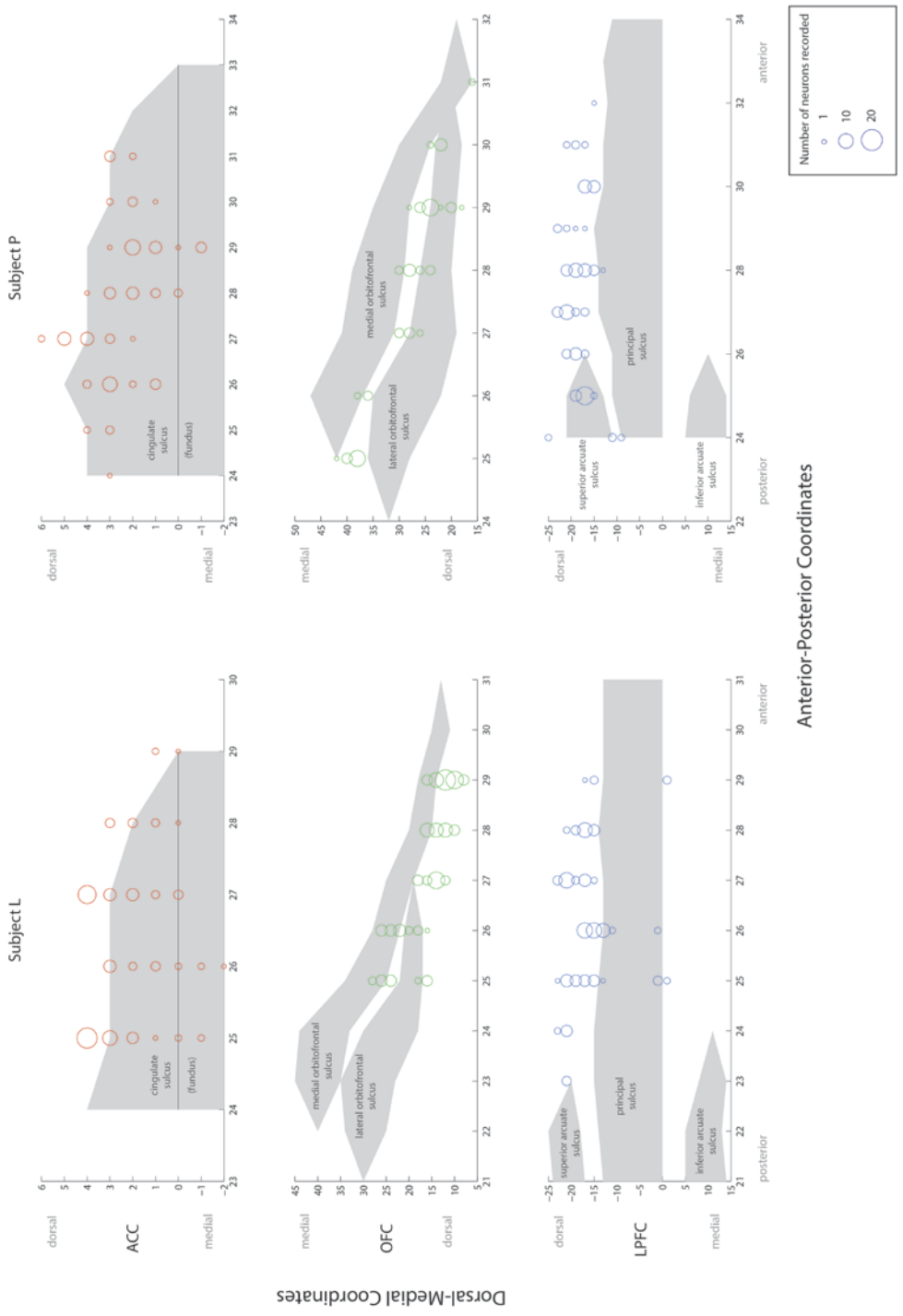
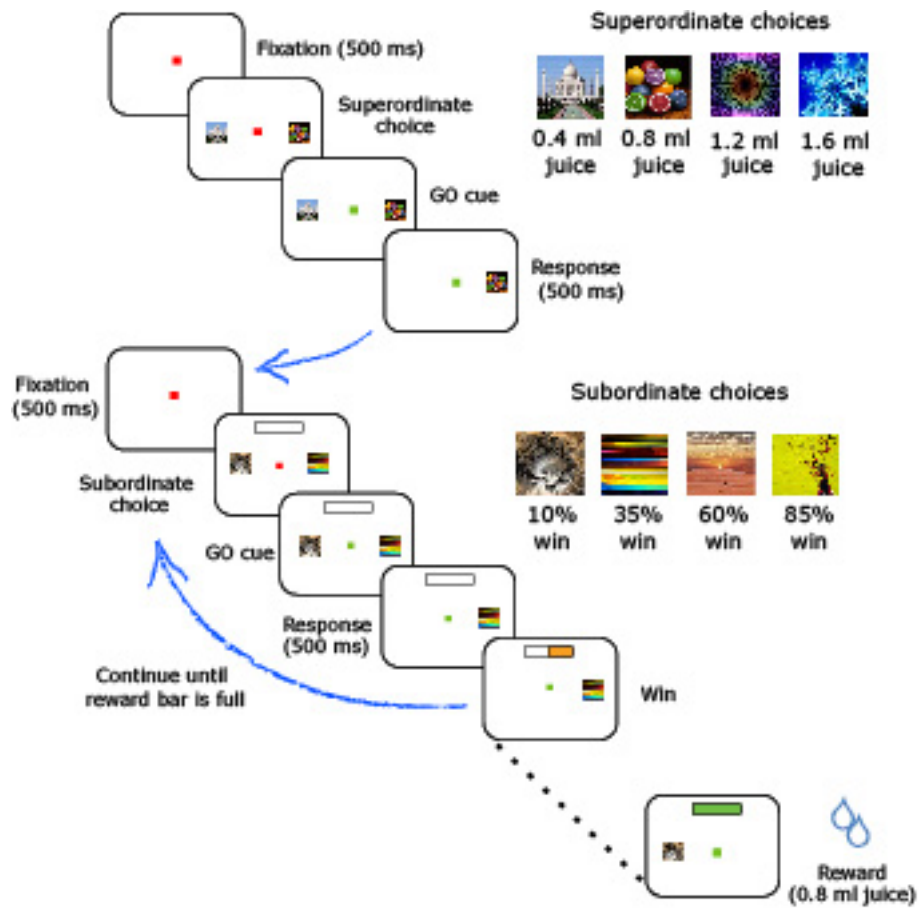
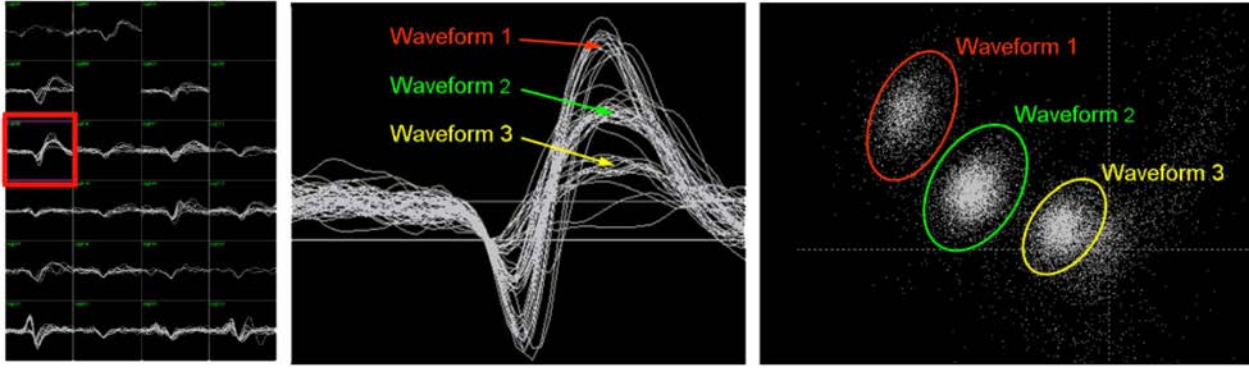


Figure 2.4 Anatomical locations of recording sites in anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), and lateral prefrontal cortex (LPFC) of animals L and P. Shaded areas in grey represent sulci (ACC: cingulate sulcus; OFC: lateral and medial orbitofrontal sulci; LPFC: superior and inferior arcuate sulci and principal sulcus). Dorsal-medial coordinates are with respect to the intrahemispheric line and anterior-posterior coordinates are with respect to the interaural line.



**Figure 2.5** Illustration of the sequence of events in the behavioral task. There were two decision periods: a superordinate choice and a subordinate choice, both presented randomly.





**Figure 2.6** Distinct waveforms cluster on the basis of a principal components analysis using Offline Sorter

## 3. Results

### 3.1 Behavioral analysis

#### 3.1.1 Subject L

Subject L was new to the lab prior to the beginning of this experiment. This monkey was trained for 13 months, through a series of increasingly complex games, before the implantation of recording chambers. During the training period, when given a new stimulus set, subject L consistently learned the slot machine probabilities more quickly than casino values. We set the criterion for learning as choosing the more rewarding stimulus for more than 20 consecutive trials. Upon receiving a new stimulus set each testing day for five days, Subject L reached this criterion on average after 150 slot machine choices and 270 casino choices. Recording chambers were implanted once the subject consistently performed optimally for at least 85% of a given session for two weeks (or ten sessions) of testing on both casino and slot machine choices. Rarely, and typically later in a session when the subject was satiated, the monkey failed to achieve initial fixation or complete whole trials. Data was recorded during 44 sessions. Subject L completed an average of  $538 \pm 100$  trials per session. Subject L chose the highest value slot machine of those presented 99% of the time and the highest value casino offered on 90% of trials. Fewer than 2% of trials were excluded from our analyses because of errors such as failing to initiate fixation.

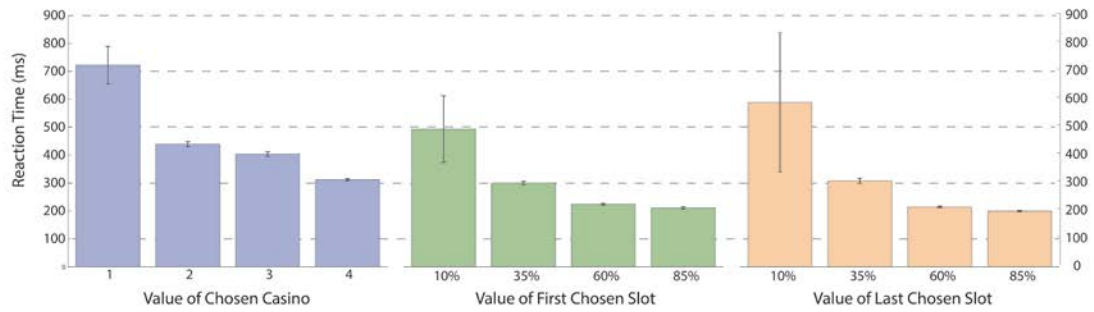
We used a logistic regression to determine which factors were driving the subject's choices. For the casino choice, we examined whether the value of the left casino and the value of the right casino could predict whether the subject would select the casino on the left. As the value of the left casino increased, or the value of the right casino decreased, the subject was significantly more likely to select the left casino ( $p < 1 \times 10^{-15}$  for both predictor variables). Thus, the subject appeared to be using the value of both casinos to determine his choice. For the slot machine choices, we used four predictor variables consisting of the value of the left and right slot machine, as well as value of the left and right casinos. The number of slot machine choices varied from trial to trial, depending on the subjects' choices as well as the payout contingencies of the slot machine. Thus, to analyze choices across trials, we focused on the first and last slot machine choice, as well as the casino choice. The likelihood of choosing the left slot machine increased as the value of the left slot increased or the right slot decreased, for both the first slot and last slot choice ( $p < 1 \times 10^{-15}$  in all cases). There was no effect of the left or right casino values on the slot machine choices ( $p > 0.05$  in all cases).

Further evidence that the subject was comparing both options in order to make a choice was evident in his reaction times (RTs). If the subject was performing such a comparison, we reasoned that his RTs would be slowest when the value difference between the two options was small, and quicker when this value difference was large. To examine whether this was the case, we performed a multiple linear regression to determine whether we could predict the monkey's RT by the values of the chosen stimulus, the unchosen stimulus, and the absolute difference between the chosen and unchosen stimuli. In addition, we included whether the

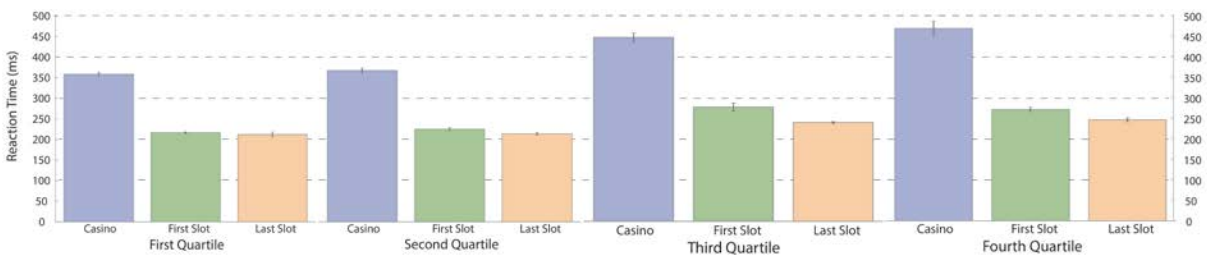
subject was responding to the left or right as a nuisance parameter. Differences here could arise due to a variety of factors such as the calibration of the eye tracker, or due to a subject's attentional bias. Subjects indicated their choices by a steady fixation within 2 degrees of visual angle of the selected stimulus for 500-ms. Because the subjects were able to freely view the stimuli on the screen, they often briefly made saccades to different objects before making their final response. Reaction times were computed as the time between the appearance of stimuli on the screen and the completion of the choice of interest (minus the required choice selection time of 500-ms).

Consistent with the idea that it takes longer to determine the optimal choice when the value of the choices are close together, we found that smaller absolute differences in value significantly predicted longer RTs for the casino choice ( $p < 3.5 \times 10^{-15}$ ) and for the first ( $p < 1 \times 10^{-15}$ ) and last ( $p < 3 \times 10^{-8}$ ) slot machine choice. Not surprisingly, as shown in Figure 3.1, subject L also consistently responded faster to more valuable options for both casino and slot choices ( $p < 1 \times 10^{-15}$  in all three cases). However, we also noticed that casino choices tended to take longer than slot machine choices. At an average of  $382\text{-ms} \pm 5\text{-ms}$  casino decisions took significantly longer than both the first slot machine choice ( $230\text{-ms} \pm 5\text{-ms}$ ) and the last slot machine choice ( $213\text{-ms} \pm 2\text{-ms}$ ; one-way ANOVA,  $F_{2,68522} = 731.83$ ,  $p < 1 \times 10^{-15}$ ).

One possible explanation for the slower responses on casino choices might be that they arise when the subject is losing motivation to perform the task. In this situation it may take longer for the subject to make his first choice (the casino), but once that choice is performed, he makes his subsequent choices (the slots) at normal speed in order to ensure he receives the final reward. To examine whether this was the case, we broke the session into quartiles. Although reaction times did generally increase as the session progressed, casino RTs remained consistently slower than slot machine RTs (Figure 3.2). We performed a two-way ANOVA to test the significance of these effects, with RT as the predictor variable, and independent variables of Choice (casino, first slot, last slot) and Quartile. There was a significant main effect of Quartile ( $F_{3,68511} = 12$ ,  $p < 1 \times 10^{-7}$ ), which arose from RTs getting slower as the session progressed. In addition, there was a significant main effect of Choice ( $F_{2,68511} = 730$ ,  $p < 1 \times 10^{-15}$ ), which arose from casino choices being slower than slot machine choices. Most importantly, a simple effects analysis showed that there was a significant effect of Choice even during the first Quartile ( $F_{2,68511} = 150$ ,  $p < 1 \times 10^{-15}$ ), when the effects of inadequate motivation should have been weakest. Thus, subject L appeared to make casino choices more slowly than slot machines, irrespective of his motivational state, suggesting that casinos required additional cognitive processing compared to slot machines.



**Figure 3.1.** Monkey L: Reaction times for the chosen casino, first slot, and last slot machine values.



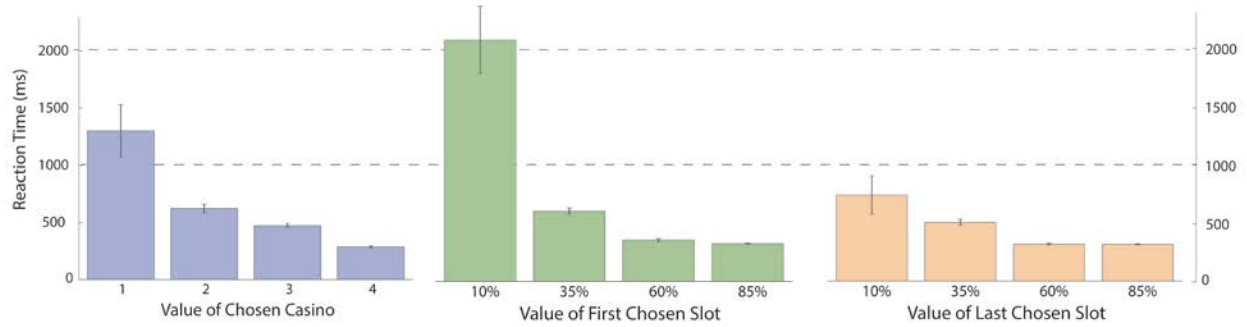
**Figure 3.2.** Monkey L: Mean reaction times (bars are SEM) for chosen values broken down by quartile of behavioral session. Reaction times became slower as a training session progressed. In addition, reaction times for casino decisions were consistently slower than those for slot machine choices.

### 3.1.2 Subject P

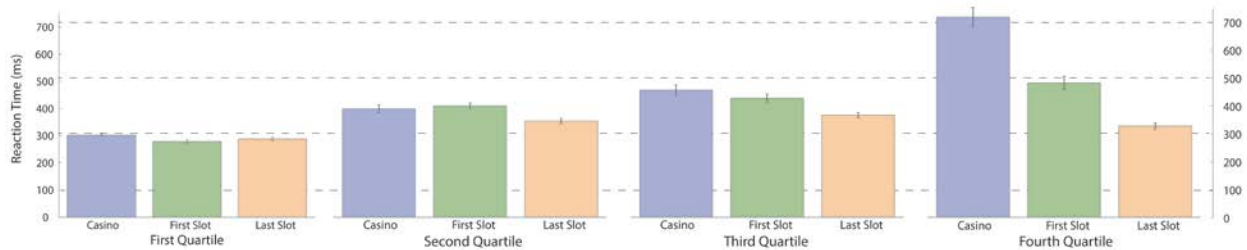
The second subject trained on this task, subject P, took 12 months to reach a consistent level of performance. Similar to subject L, when given a new stimulus set, subject P learned slot machine values more quickly than casino values. After having been given a new stimulus set for each of five days, Subject P took 160 slot machine choices and 400 casino choices to reach a criterion of 20 consecutive trials where he chose the optimal stimuli at both levels of the task. Recording chambers were implanted once the animal performed at or above 85% for every testing session over a two week period. Analyses were computed only for successfully completed trials. During 32 recorded sessions, of  $522 \pm 104$  trials per session, Subject P chose the highest value slot machine of those presented 96% of the time and the highest value casino offered on 91% of trials. Errors, such as fixation failures, occurred on 6% of trials and we excluded these trials from our analyses.

Our analysis of subject P's choices followed the same methods outlined above for subject L. We used a logistic regression to examine how the value of the two options influenced the subject's choice. We determined that subject P's choices were highly driven by the value of the options for both casino and slot machine choices such that as the value of the item on the left increased or the value of the item on the right decreased, the subject was more likely to make a leftward response ( $p < 1 \times 10^{-15}$ ). We also tested whether the casino influenced which slot machines were subsequently chosen and while there was no effect of the casinos on the last slot machine choice ( $p > 1 \times 10^{-15}$  in both cases), it seems that the value of the right casino ( $p = 0.0069$ ) but not the left ( $p = 0.0997$ ) influenced the first slot machine choice. However, note that the magnitude of this effect was considerably weaker than the effect of the value of the slot machines. A one unit change in the value of the slot machines increased the odds of choosing the slot machine by an average of 110%, whereas in contrast, a one unit change in the value of the casinos increased the odds of choosing a particular slot machine by an average of only 8%.

Next, we again considered whether the monkey's RT could be explained by the values of the chosen stimulus, the unchosen stimulus, the absolute difference between the chosen and unchosen stimuli, or the responses of right and left. We concentrated on three choices: the casino choice and the first and last slot machine choice. Similar to subject L, subject P took significantly longer to make choices when they were closer in value ( $p < 0.001$  in all cases). He was also quicker when choices were of higher value (Figure 3.3,  $p < 1 \times 10^{-15}$  in all cases). In addition, casino choices were consistently faster than slot machine choices. At an average of  $455\text{-ms} \pm 10\text{-ms}$  casino decisions took longer than both the first ( $385\text{-ms} \pm 7\text{-ms}$ ) and the last ( $319\text{-ms} \pm 4\text{-ms}$ ) slot machine choices. A 2-way ANOVA of Choice (casino, first slot, last slot) and Quartile on the subjects' RTs showed a significant main effect of Quartile ( $F_{3,45432} = 85$ ,  $p < 1 \times 10^{-15}$ ), indicating that the subjects' choices got slower as the session progressed, and a significant main effect of Choice ( $F_{2,45432} = 69$ ,  $p < 1 \times 10^{-15}$ ), which arose from casino choices being slower than slot machine choices (Figure 3.4). However, a simple effects analysis showed that there was no difference between casino and slot machine choices during the first Quartile ( $F_{2,45432} < 1$ ,  $p > 0.1$ ) suggesting that, in subject P, the casino choices were slower because of motivational issues associated with initiating the sequence of choices that constituted a trial.



**Figure 3.3.** Monkey P: Reaction times for the chosen casino, first slot, and last slot machine values.



**Figure 3.4.** Monkey P: Reaction times for chosen values broken down by quartile of behavioral session. Reaction times slow as the behavioral session progressed, particularly for casino decisions. This could arise because of suggesting a decline in motivation which might differentially affect the tendency to complete choices in the early phase of the trial.

## 3.2 Neurophysiology: LPFC

### 3.2.1 Value encoding

In total, 262 neurons were recorded from LPFC, 136 from subject L and 126 from subject P. Given that value was highly predictive of behavioral parameters, we used a stepwise regression to broadly search for value-related variables that were predictive of each neuron's firing rate during casino and slot machine decision epochs during the behavioral task. We considered seven variables related to the value of the pictures: chosen value; unchosen value; the sum and difference of the chosen and unchosen values; the ipsilateral value, contralateral value, and the difference of the ipsilateral and contralateral values. Note that the sum of the ipsilateral and contralateral values is identical to the sum of the chosen and unchosen values and so was not included as an additional regressor. We also included a parameter that reflected the subject's behavioral response to either the left or right option (section 3.2.2 will describe neural activity related to the behavioral response). We defined the decision period for a given epoch as the 500-ms following the presentation of the stimuli, with a 100-ms lag to account for the time it takes for information to reach LPFC. In other words, the casino decision period is from 100-ms to 600-ms following the appearance of the two casino stimuli on the monitor. Because the number of slot machine choices varies from trial to trial, we restrict most of our analyses to the first and last slot machine choices and epochs. In addition, we examined neural selectivity during the fixation period, the 500-ms immediately preceding the presentation of the casino. During the fixation period, the subject has no information about the value of the upcoming pictures, and so analyzing neural activity during this period provides us with a useful benchmark against which to compare our statistical analyses from the remainder of the trial.

Table 3.1 shows the percentage of neurons encoding each of these parameters for each decision at each stage of the task. During the casino choice period, the most prevalent value encoding (39/262 or 15%) related to the chosen casino. Figure 3.5 shows a neuron that encoded the value of the chosen casino. It responded most strongly when the subject chose the most valuable stimulus, and responded less strongly when the subject had to choose one of the less valuable options. Furthermore, it did not encode the value of the slot machines.

The proportion of neurons that encoded the value of the chosen casinos was significantly higher than the proportion that encoded the value of the slot machines. This was true irrespective of whether we compared the casinos and the first slot (chi-squared test = 7.0,  $p < 0.01$ ) or the casinos and the last slot (chi-squared test = 5.0,  $p < 0.05$ ). When neurons responded selectively to the value of a chosen stimulus, most had a higher firing rate for lower value stimuli (casino: 28/39, binomial test,  $p < 0.005$ ; first slot: 18/19, binomial test,  $p < 5 \times 10^{-6}$ ; last slot: 19/22, binomial test,  $p < 0.0005$ ).

LPFC neurons also encoded stimulus values according to their position on the screen. Figure 3.6 illustrates a neuron located in the right hemisphere of animal P's brain that encoded the value of the contralateral (with respect to the brain hemisphere from which the neuron was recorded) casino but neither contralateral nor ipsilateral first or last slot machine responses. This neuron fired in a graded manner with its highest rate of discharge indicating the highest

value casino on the left side of the screen. This neuron did not respond to slot machines, irrespective of laterality, and did not significantly encode the value of the ipsilateral casino ( $p > 0.05$  in all cases). The proportion of neurons encoding the value of stimuli either ipsilateral or contralateral to the neuron was similar at the time of casino choice (59/262 or 23%), the first slot machine choice (45/262 or 17%) and the last slot machine choice (41/262 or 16%). These proportions did not significantly differ from one another (chi-squared test = 4.0,  $p > 0.1$ ). Neurons were approximately equally likely to have either positive or negative relationships between firing rate and the value of the contralateral or ipsilateral options (binomial test,  $p > 0.1$  in all cases).

We wished to determine whether the neurons encoding values overlapped more than predicted by chance. There were not enough neurons with specific value-related variables to make statistical comparisons across epochs, so we collapsed across the seven different value-related variables. In sum, 143/262 or 55% of neurons encoded value information during the casino, compared to 124/262 or 47% during the first slot machine and 114/262 or 44% during the last slot machine. There were neurons that encoded value information during the casino and the following first slot choices (67/262 or 26%) as well as during the casino and *last* slot choices (66/262 or 25%) but neither of these populations exceeded what would be expected by chance (binomial test,  $p > 0.1$  in both cases). A significant number of neurons did encode value information during both the first and last slot machine choices, however (70/262 or 27%, binomial test,  $p = 0.007$ ). Further, the proportion of neurons that encoded value-related variables during all three choice epochs also exceeded what would be expected by chance alone (42/262, binomial test,  $p = 0.0074$ ).

We next examined whether LPFC neurons maintained information about the value of the casinos while the subject performed the intervening slot machines. To do this, we added additional parameters to the stepwise regression. Specifically, we examined whether the value of the casinos predicted neural activity at the time of the slot machine choices. Thus, we included the seven variables related to value and the behavioral response for both the slot machine as well as the preceding casino (16 parameters total). There was no evidence that LPFC neurons encoded any information about the casino values at the time of slot machine choices. The proportion of neurons encoding each of the parameters relating to the casino values did not exceed chance during either the first slot machine choice or the second slot machine choice (Table 3.2, binomial test,  $p > 0.1$  in all cases).

Finally, we noted an additional difference between the slot machine choices and the casino choices. At the time of the slot machine choice, many LPFC neurons encoded the difference in value between the ipsilateral and contralateral pictures (first slot, 36/262 or 14%; last slot, 27/262 or 10%). In contrast, at the time of the casino choices, few neurons encoded this difference (19/262 or 7%). A statistical comparison of these proportions revealed that significantly more neurons encoded the difference at the time of the first slot machine choice compared to the casino choice (chi-squared test = 5.2,  $p < 0.05$ ). Encoding the difference in value of the two options could be a useful signal to guide action selection so that the most valuable option is consistently chosen.



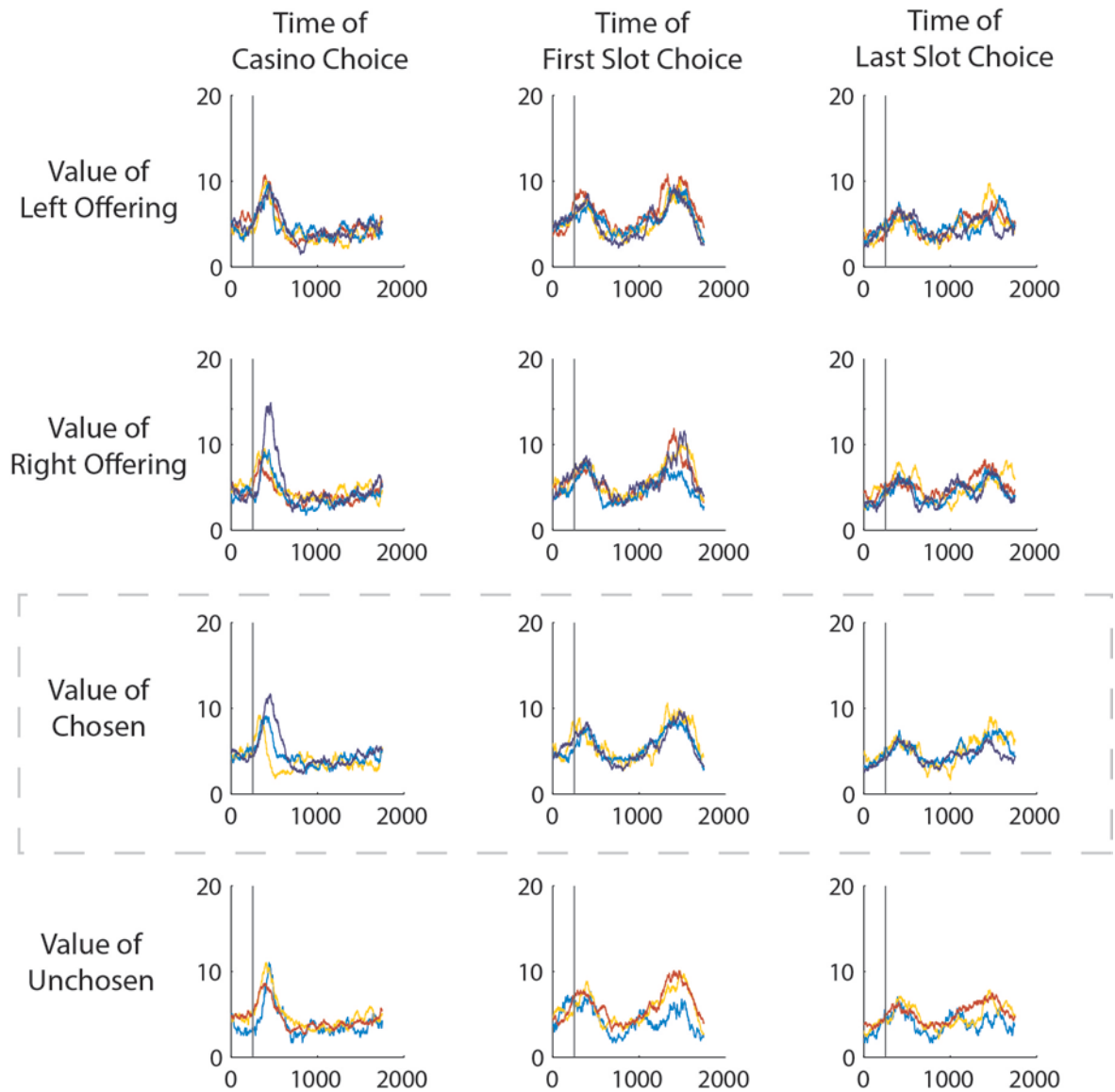
In summary, there was little evidence to support our original hypothesis that LPFC neurons might encode the value of the casinos and maintain that information through until the time of reward. Instead, at each stage of the task, LPFC neurons encoded value information that was pertinent to the choice that the subject was currently facing. However, there was also evidence that suggested that the cognitive processes underlying the casino choices were qualitatively different from those underlying the slot machine choices. Whereas LPFC neurons tended to encode the value of the chosen option at the time of the casino choice, they tended to encode the difference in the value of the options at the time of the slot machine choice. In addition, there was more overlap between neurons encoding value information at the same hierarchical level (first slot vs. last slot) than there was between neurons encoding value information at different hierarchical levels (casino vs. slot).

	Chosen	Unchosen	Chosen + Unchosen	Chosen – Unchosen	Ipsilateral	Contralateral	Ispi- Contra	Response
<b>Fixation</b>	3.44%	1.91%	4.20%	4.96%	2.29%	2.67%	1.91%	3.82%
<b>Casino</b>	14.89%	5.34%	3.82%	7.25%	12.98%	9.54%	7.25%	38.55%
<b>First Slot</b>	7.25%	4.58%	2.67%	9.92%	8.40%	8.40%	13.36%	40.84%
<b>Last Slot</b>	8.40%	2.67%	2.29%	5.34%	7.63%	8.40%	9.54%	48.85%

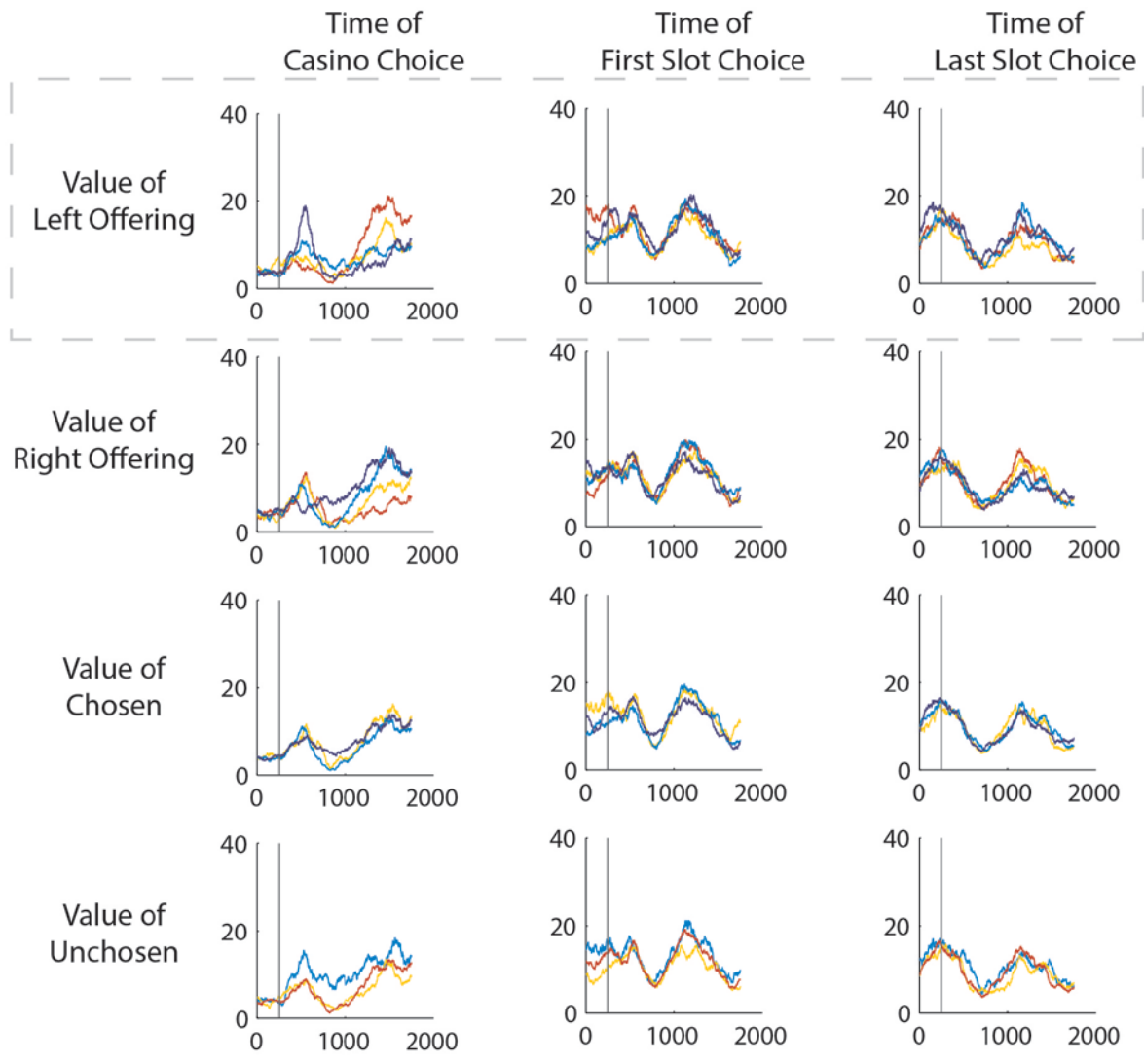
**Table 3.1** The percentage of neurons in LPFC that encode values for the chosen, unchosen, sum of the chosen and unchosen; difference between the chosen and unchosen; ipsilateral stimulus to the hemisphere recorded from; contralateral stimulus to the hemisphere recorded from; the difference between the ipsilateral and contralateral stimuli; and the response (right or left) for the casino, first slot machine, and last slot machine at the time of choice. Percentages of neurons whose activity changes significantly during the fixation period according to different values are shown for reference. Percentages in red are significant, binomial test,  $p < 0.05$ .

	<b>Chosen</b>	<b>Unchosen</b>	<b>Chosen + Unchosen</b>	<b>Chosen – Unchosen</b>	<b>Ipsilateral</b>	<b>Contralateral</b>	<b>Ispi- Contra</b>	<b>Response</b>
<b>First Slot</b>	5.73%	1.15%	3.05%	4.96%	3.82%	3.05%	3.44%	11.45%
<b>Last Slot</b>	1.53%	3.05%	1.91%	4.58%	2.29%	5.34%	3.05%	4.20%

**Table 3.2** The percentage of neurons that encode the *casino* values at the times of first and last slot machine choices. Neurons were classified according to whether they encoded the value of the following options: chosen, unchosen, sum of the chosen and unchosen, difference between the chosen and unchosen, ipsilateral stimulus to the hemisphere recorded from, contralateral stimulus to the hemisphere recorded from, difference between the ipsilateral and contralateral stimuli or the response (right or left). Percentages in red are significant, binomial test,  $p < 0.05$ .



**Figure 3.5** Neuron L129\_18b increased its firing rate during the casino choice as the value of the chosen casino increased. It did not encode the value of the slot machines. Dark blue is the response to the highest value option; light blue is the 2<sup>nd</sup> highest value option; yellow represents the 2<sup>nd</sup> lowest option; orange is the lowest value option. For the chosen value, we combined the lowest value option with the second lowest value option, since there were few trials where the subject chose the lowest value option. Similarly, for the unchosen value, we combined the highest value with the second highest value option, since there were few trials where the subject did not choose the highest value option. Although the neuron did appear to also respond to the highest value option when it was presented on the right (contralateral to the recording location) the effect did not reach significance ( $p > 0.1$ ). Black vertical bar denotes time of stimulus onset. Abscissae correspond to neuronal firing rate in Hz and ordinate axes denote time in milliseconds.



**Figure 3.6** Neuron P106\_25b, located in the right hemisphere of animal P responded selectively to the value of the contralateral casino (but not slot machines) by increasing its rate of discharge in response to higher value choices. Conventions as in Figure 3.5.

### 3.2.2 Response encoding

Many LPFC neurons encoded the behavioral response, i.e. whether the animal selected the option on the left or the right. An example is shown in Figure 3.7. This neuron increased its firing rate whenever the animal selected the option on the right, irrespective of whether it was a casino or slot machine choice. Most neurons recorded from LPFC encoded behavioral responses to at least one of the decisions (casino, first slot or last slot: 174/262 or 66%). During all three decision epochs the cells responding selectively to behavioral responses were equally likely to encode leftward or rightward choices (binomial test,  $p > 0.1$  in all cases). From our initial stepwise regression analysis (Table 3.1), we noted that the proportion of neurons that encoded the behavioral response increased chronologically as the trial progressed from earlier to later choices (casino, 101/262 or 39%; first slot, 107/262 or 41%; last slot, 128/262 or 49%). However, the trend just failed to reach significance (chi-squared test = 5.8,  $p = 0.055$ ).

Across the population, neurons that encoded the behavioral response for one kind of decision often encoded the response for another kind of decision. Thus, there were neurons that encoded the behavioral response to both slot machine decisions (81/262 or 31%), to the casino and first slot machine (66/262 or 25%) or to the casino and last slot machine (67/262 or 26%). There was also a population of cells that, like the neuron in Figure 3.7, encoded the responses to all three choices: casino, first slot, and last slot (52/262 or 20%). We examined whether the overlap in these populations was significantly greater than chance. Specifically, we determined the incidence of encoding of the behavioral response for each choice, and then examined whether the proportion of neurons encoding conjunctions of response encoding exceeded the proportion that one would expect if the neuronal populations were statistically independent of one another. We found that the proportion of neurons encoding the behavioral response for multiple decisions was significantly greater than would be predicted by chance (casino and first slot; casino and last slot; first slot and last slot; and casino, first slot, and last slot; binomial test,  $p < 1 \times 10^{-15}$  in all cases).

However, there was also evidence that some neurons encoded the behavioral response for a specific decision. For example, the neuron in Figure 3.8 encoded the behavioral response for the slot machine choice, but did not encode this information at the time of the casino choice. In sum, 29/262 or 11% of neurons encoded the behavioral response at the time of the slot machine but not at the time of the casino, whereas 20/262 or 8% encoded the behavioral response at the time of the casino but not at the time of the slot machine. These proportions were not significantly different from one another (chi-square test = 1.4,  $p > 0.1$ ).

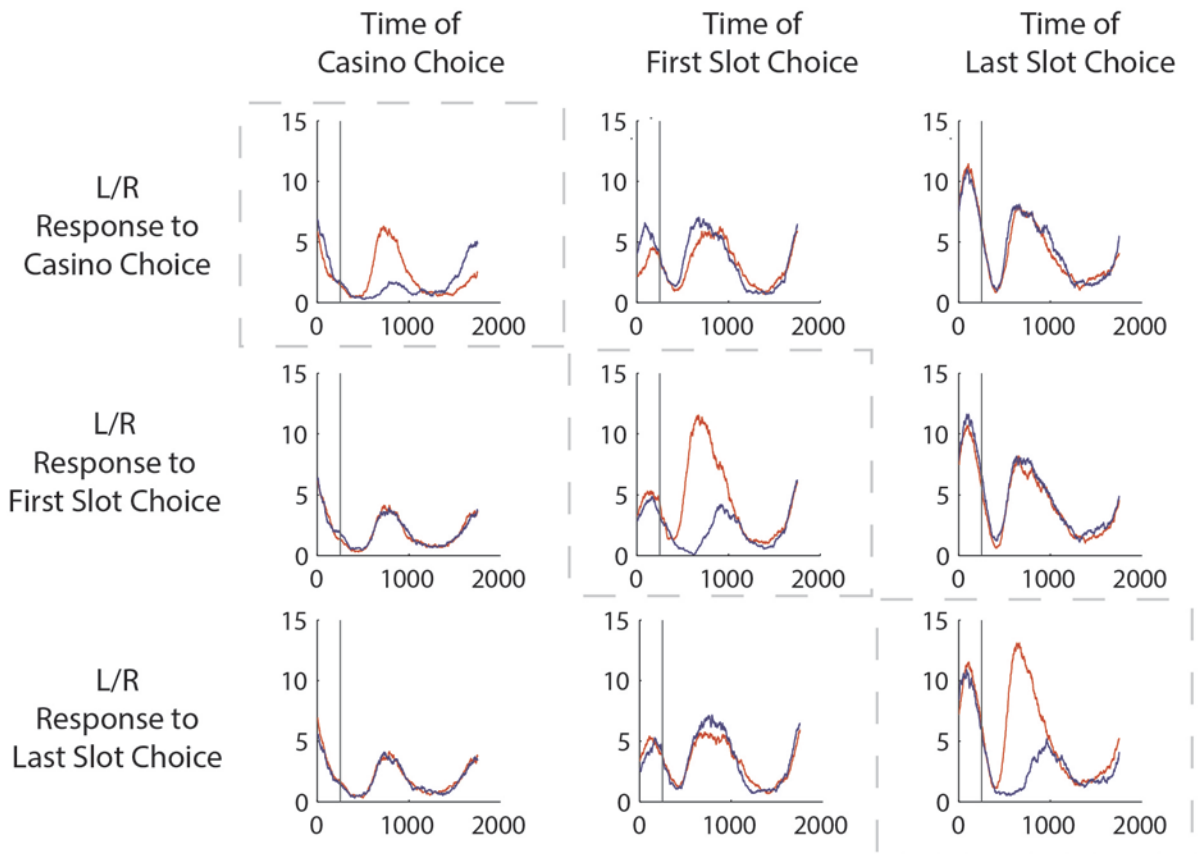
Finally, as mentioned in the previous section on value encoding, we also added a parameter to our stepwise regression model to look at whether neurons in LPFC maintained information about the response to the casino choice at the time of slot machine choices. Interestingly, the response at the time of the casino choice was predictive of the neural activity at the time of the first slot machine choice for some neurons (26/262 or 10%). However, the proportion of neurons encoding the casino response at the time of the second slot machine choice did not exceed chance (11/262 or 4%, binomial test,  $p > 0.1$ ). This result led us to ask whether there

might be encoding of early slot machine responses that are predictive of neuronal activity at subsequent slot responses. In fact, there was a significant number of neurons (38/262, binomial test,  $p < 1 \times 10^{-15}$ ), such as the one shown in Figure 3.9, whose activity at the second slot machine choice was predicted by the response to the first slot machine.

In summary, the encoding of behavioral responses by neurons in LPFC was widespread and complex. Much of the encoding was consistent with a straightforward motor response, with neurons showing similar encoding of the behavioral response to more than one, and sometimes all three, decisions. There were also neurons that showed more complex responses, encoding the behavioral response only at the time of the superordinate or subordinate choice. However, there was no evidence that LPFC neurons were biased towards encoding either the superordinate or subordinate choice. Finally, we also found that encoding of the behavioral response during one choice could predict response encoding at subsequent choices. Such activity could arise if neurons fired when the animal was going to make a specific response, but only if the response followed a specific prior response. For example, a neuron might only respond to rightward slot machine choices when they follow a leftward casino choice. Such an encoding scheme could conceivably represent a mechanism by which individual responses are chunked together to form a specific action sequence.

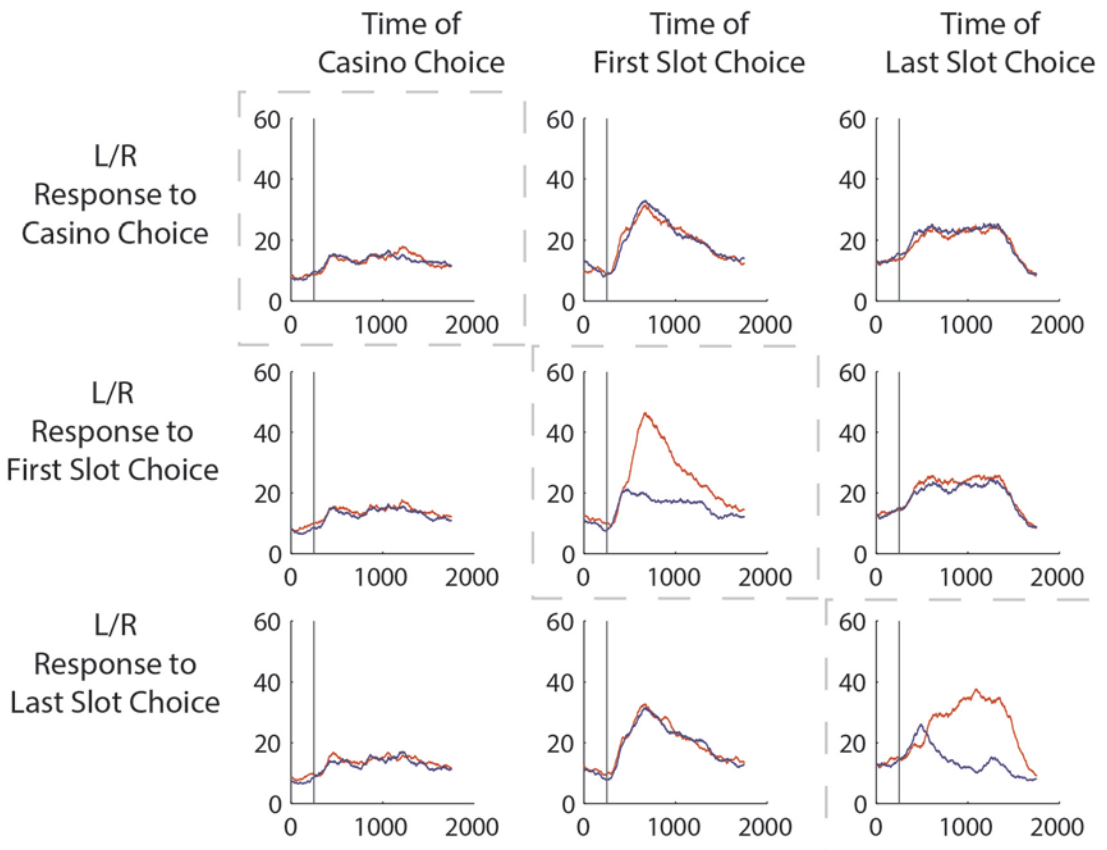
### **3.3 Summary**

LPFC neurons encoded many different variables within the current task, but the encoding of the value of the options was restricted to a minority of the neurons. Instead, the most prevalent encoding was of the behavioral response. However, there was no evidence to support the notion that LPFC was preferentially involved in encoding higher levels of the behavioral hierarchy. Furthermore, there was no evidence that LPFC neurons maintained information about the superordinate choice while the subordinate choices were made.

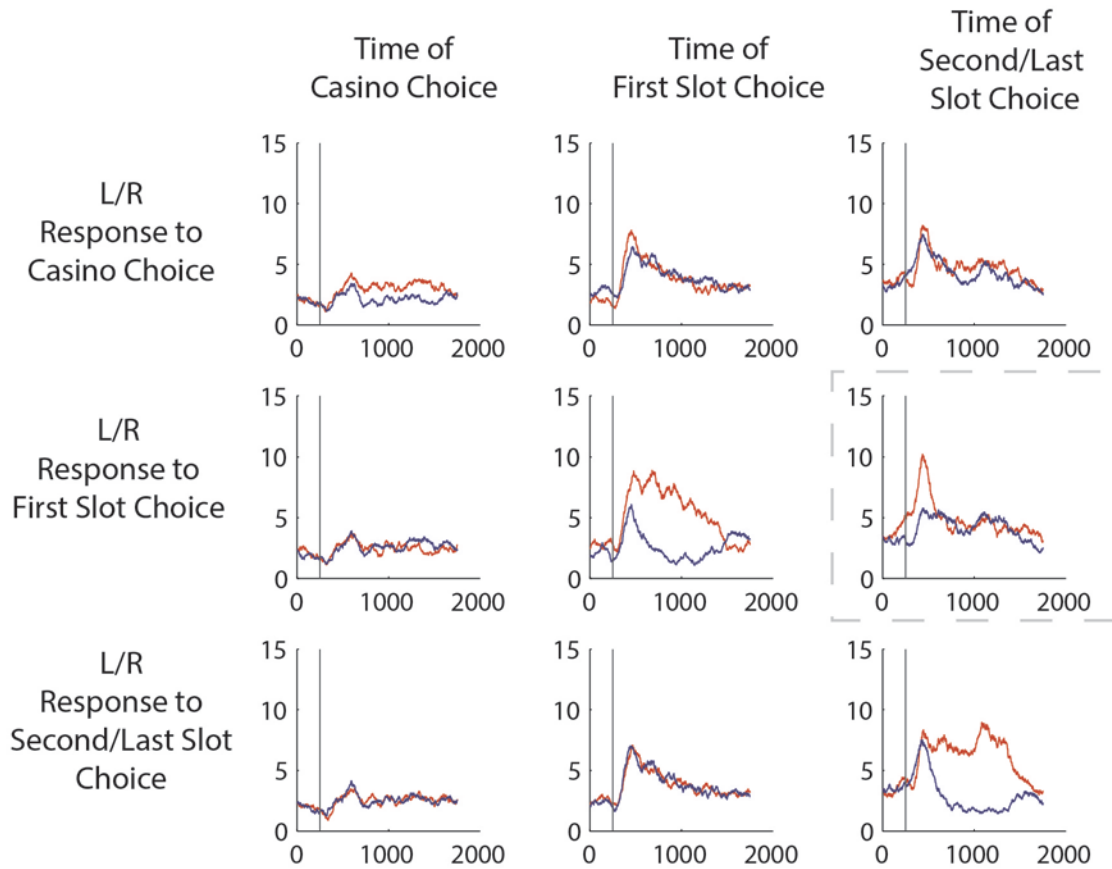


**Figure 3.7** Neuron P109\_18c increased its firing rate whenever the animal selected the right option (red) compared to the left option (blue). This was observed irrespective of whether the subject was making a casino or slot machine choice. Black vertical bar denotes time of stimulus onset. Abscissae correspond to neuronal firing rate in Hz and ordinate axes denote time in milliseconds.





**Figure 3.8** Firing rate of neuron L125\_001b for leftward responses in blue and rightward responses in red. This is an example of a neuron that modulates its firing rate for responses at the subordinate but not superordinate level. Conventions as in Figure 3.7.



**Figure 3.9** Neuron L125\_001a depicts a neuron selective for rightward responses to slot machine choices. At the time of the second/last slot machine choice, the cell encoded the response made at the first slot choice. On the majority of trials there are only two slot machine choices, so the second slot machine is also the last slot machine. However, on trials with less favorable slot options, sometimes the animal is forced to make several choices, the first and second of which are shown above. All other conventions as in Figure 3.7.

## 4. Results

### 4.1 Neurophysiology: OFC

#### 4.1.1 Value encoding in OFC

We recorded the activity of 249 neurons from OFC, 160 from animal L and 89 from animal P. Given the classic role of OFC in encoding value, our first analysis aimed to determine how neurons in OFC compared with those in LPFC in terms of value encoding. We used a stepwise regression taking into account the seven variables related to stimuli value: chosen value, unchosen value, sum of chosen and unchosen values, difference of chosen and unchosen values, the ipsilateral value, contralateral value and the difference between the ipsilateral and contralateral values. Ipsilateral and contralateral were defined with respect to the brain hemisphere in which the neuron was recorded. We also added a parameter for the subject's behavioral response, i.e. whether he chose the left or right option.

The proportions of cells in OFC that alter their activity according to stimulus value at each stage of the task are shown in Table 4.1. During the casino choice period, the most prevalent value encoding (35/249 or 14%) related to the chosen casino. Figure 4.1 shows a neuron that encoded the value of the chosen casino. It responded most strongly when the subject chose the least valuable stimulus, responded less strongly when the subject chose a more valuable option, and did not encode the value of the slot machines. The proportion of cells that encoded the value of the chosen casinos, however, was not significantly different from the proportion that encoded the value of the chosen slot machines irrespective of whether we compared the casinos and the first slot (chi-squared test = 0.28,  $p > 0.1$ ) or the casinos and the last slot (chi-squared test = 0.06,  $p > 0.1$ ). Although there was a bias towards encoding lower value stimuli with a higher firing rate, it was weak and inconsistent (casino: 23/35 or 66%, binomial test,  $p < 0.05$ ; first slot: 18/30 or 60%, binomial test,  $p > 0.1$ ; last slot: 22/32 or 69%, binomial test,  $p = 0.01$ ). Thus, in comparison to LPFC, OFC neurons showed a more even distribution of neurons that either increased their firing rate as value increased or increased their firing rate as value decreased. However, the overall proportion of neurons encoding chosen values was similar in LPFC and OFC at the time of the casino choice and the first and last slot machine choices (chi-squared test,  $p > 0.05$  in all cases).

OFC neurons also encoded stimulus values according to their position on the screen. Figure 4.2 shows a neuron recorded from the right hemisphere of animal L that encoded the value of the contralateral (with respect to the brain hemisphere from which the neuron was recorded) casino. This neuron did not respond to slot machines, irrespective of laterality, nor did it respond significantly to value of the ipsilateral casino ( $p > 0.05$  in all cases). In addition to coding for stimulus values according to their position on the screen, neurons in OFC also encoded the difference in value of the contralateral stimulus from the ipsilateral stimulus. The prevalence of laterality-encoding neurons in OFC did not differ significantly from their prevalence in LPFC (chi-squared tests,  $p > 0.1$  in all cases).

Finally, to examine whether OFC neurons maintained information about the value of the casinos while the subject performed the intervening slot machines we added additional parameters to our stepwise regression model. We included the seven variables related to value, as well as one for behavioral response, for both the slot machine as well as the preceding casino (16 parameters total) to see whether the value of the casinos predicted neural activity at the time of the slot machine choices. As shown in Table 4.2, there were several small but significant populations of neurons during the slot machine epochs that encoded values related to the previous casino choices. This was in marked contrast to what we found in LPFC, where there was no evidence that the value of the casinos were maintained across the slot machine choices. Because the population of selective neurons was small and distributed across several types of encoding scheme, we collated neurons according to whether they encoded any value information about the casinos at the time of the first or last slot choice. Significantly more neurons in OFC (147/249 or 59%) maintained some kind of value information about the casinos during either the first or last slot choice compared to LPFC (121/262 or 46%, chi-squared test = 8.0,  $p < 0.005$ ).

#### *4.1.2 Response encoding in OFC*

Many OFC neurons encoded whether the subject chose the left or right option. An example is shown in Figure 4.3. This neuron increased its firing rate for responses in the right direction and suppressed its firing rate when the subject chose the leftward stimulus. Nearly half of the OFC neurons encoded the response of at least one behavioral choice (122/249 or 49%, binomial test,  $p < 1 \times 10^{-15}$ ). However, the proportion of response encoding neurons was still significantly less than in LPFC for casino, first slot and last slot choices (chi-squared test  $> 7.9$ ,  $p < 0.005$  in all cases). To address whether there were cells dedicated to encoding responses for one type of choice and not another, we looked to see if there were any neurons whose activity significantly correlated with the responses to casinos but not slot machines or to slot machines but not casinos. There was a significant population of neurons that encoded the response associated with the casinos but not the slots (24/249 or 10%, binomial test,  $p > 0.1$ ), while neurons that showed the opposite pattern (encoding the response associated with the slots but not the casinos) were at chance levels.

The final question we addressed with regard to response coding in OFC was whether there were neurons whose activity reflected information about the earlier responses at later points in time. Indeed, there was a small, but significant population of OFC neurons that encoded the casino choice response during the first slot machine choice (24/249 or 10%, binomial test,  $p < 0.001$ ) and the second slot machine choice (16/249 or 6%, binomial test,  $p = 0.05$ ). We similarly wished to know whether responses to the first slot machine choice were predictive of neural activity at the time of subsequent slot machine choices and, as in LPFC, there was some evidence that some neurons (24/249 or 10%, binomial test,  $p < 0.001$ ) do indeed maintain information about the first slot machine response at the time of responding to the second/last slot machine.

### **4.1.3 OFC Summary**

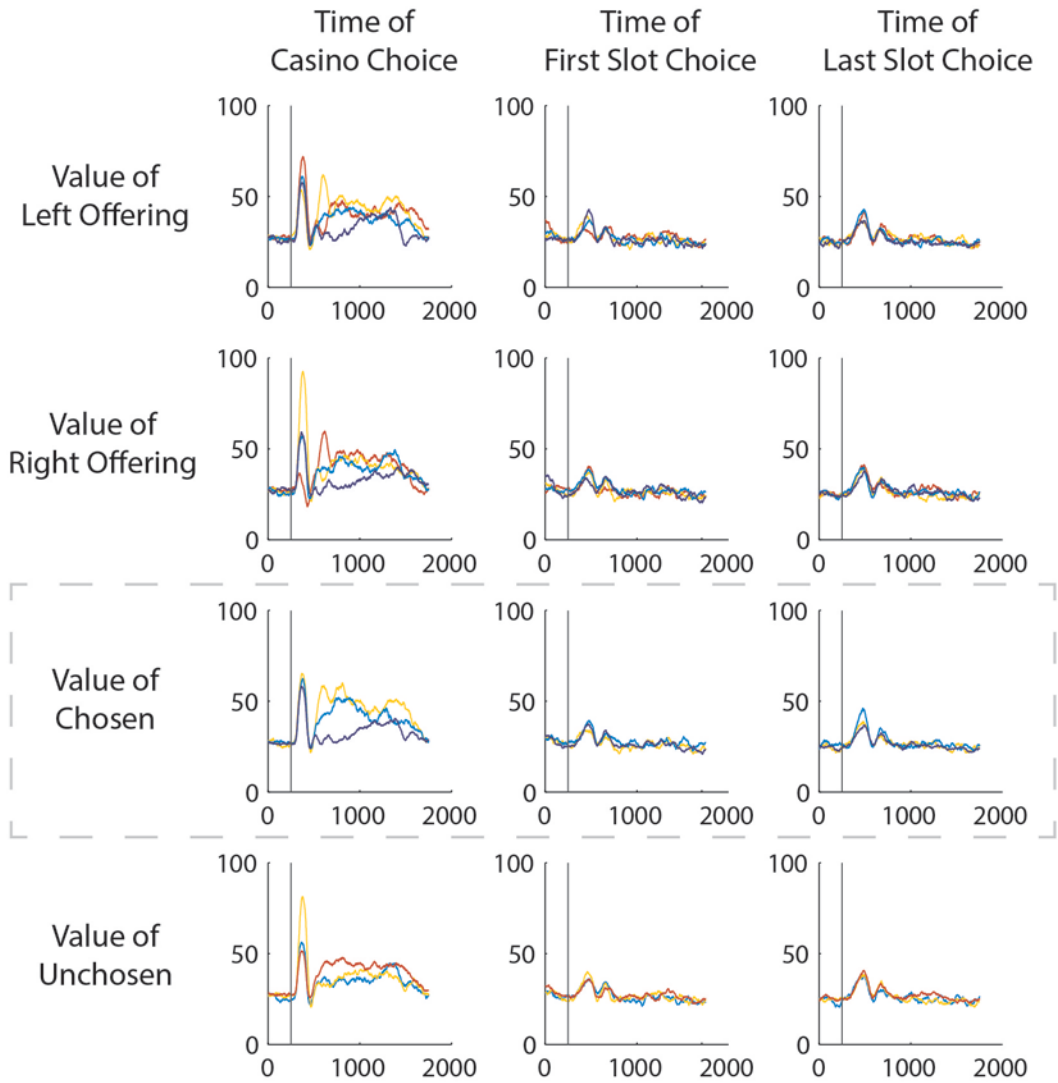
OFC neurons encoded many different task variables relevant to performance of the task. However, two key differences emerged relative to LPFC coding. First, OFC neurons were more likely to maintain value information about the casinos across the intervening slot machine choices. Second, encoding of the response was generally weaker in OFC at all the choices.

	Chosen	Unchosen	Chosen + Unchosen	Chosen – Unchosen	Ipsilateral	Contralateral	Ipsi – Contra	Response
<b>Fixation</b>	2.01%	2.81%	4.02%	4.02%	1.20%	3.21%	2.01%	4.42%
<b>Casino</b>	14.06%	6.02%	4.42%	10.44%	9.64%	8.84%	10.84%	26.51%
<b>First Slot</b>	12.05%	1.61%	4.42%	5.62%	12.05%	10.04%	10.04%	24.10%
<b>Last Slot</b>	12.85%	1.61%	2.01%	8.43%	7.23%	8.03%	6.43%	28.92%

**Table 4.1** The percentage of neurons in OFC that encoded chosen value, unchosen value, sum of the chosen and unchosen values, difference between the chosen and unchosen values, value of the ipsilateral stimulus, value of the contralateral stimulus, the difference between the value of the ipsilateral and contralateral stimuli, and the subject's behavioral response, i.e. whether they chose the left or right option. The percentage of selective neurons is shown separately at the time of the casino choice, first slot machine choice and last slot machine choice. Percentages of neurons whose activity changed significantly during the fixation period according to different values are shown for reference. Percentages in red are significant, binomial test,  $p < 0.001$ .

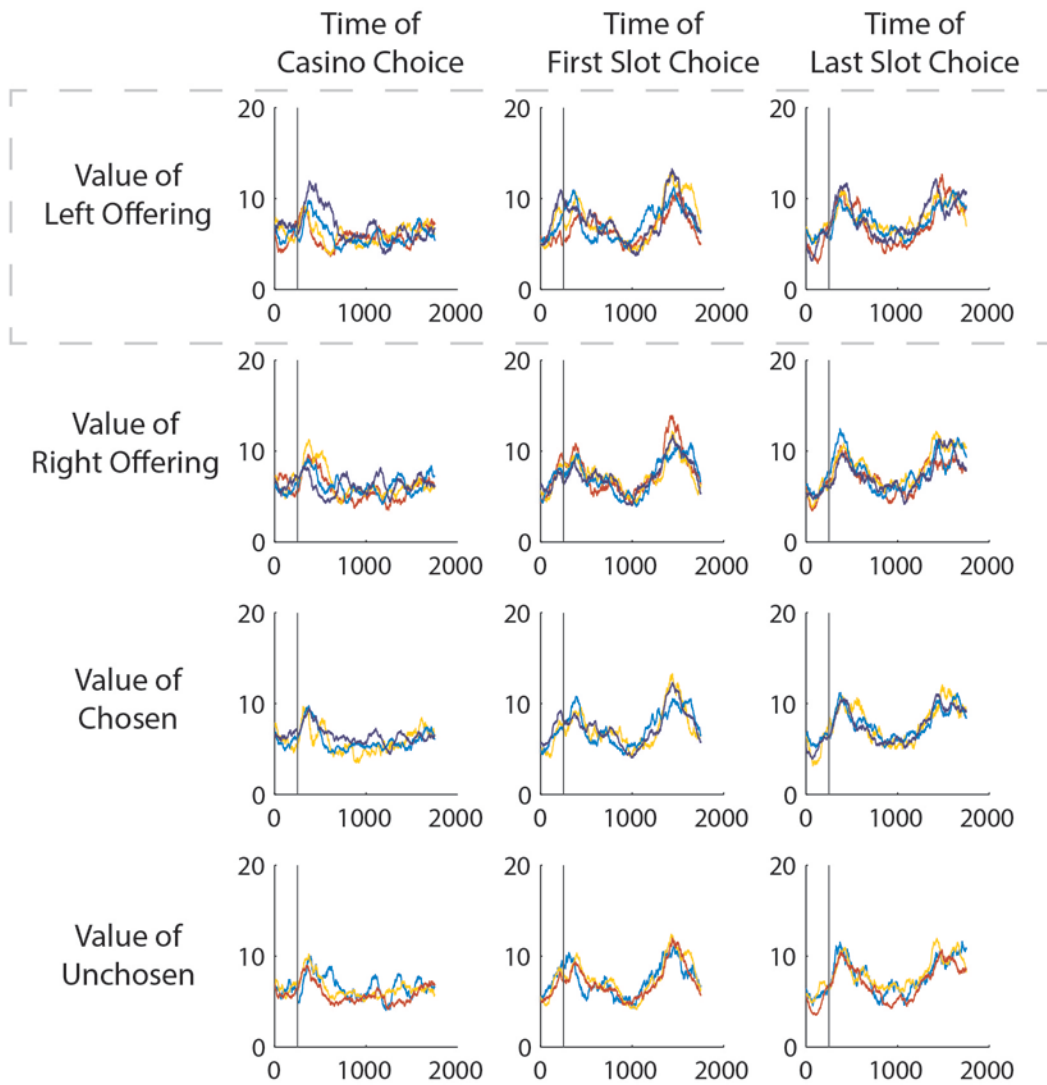
	Chosen	Unchosen	Chosen + Unchosen	Chosen – Unchosen	Ipsilateral	Contralateral	Ipsi – Contra	Response
<b>First Slot</b>	4.82%	2.41%	2.01%	5.22%	6.43%	6.02%	4.02%	9.64%
<b>Last Slot</b>	6.02%	2.81%	4.82%	6.43%	6.02%	2.01%	1.61%	3.61%

**Table 4.2** The percentage of neurons in OFC that encoded the *casino* values for chosen, unchosen, sum of the chosen and unchosen; difference between the chosen and unchosen; ipsilateral stimulus to the hemisphere recorded from; contralateral stimulus to the hemisphere recorded from; the difference between the ipsilateral and contralateral stimuli; and the response (right or left) at the times of first and last slot machine choices. Percentages in red are significant, binomial test,  $p < 0.05$ .

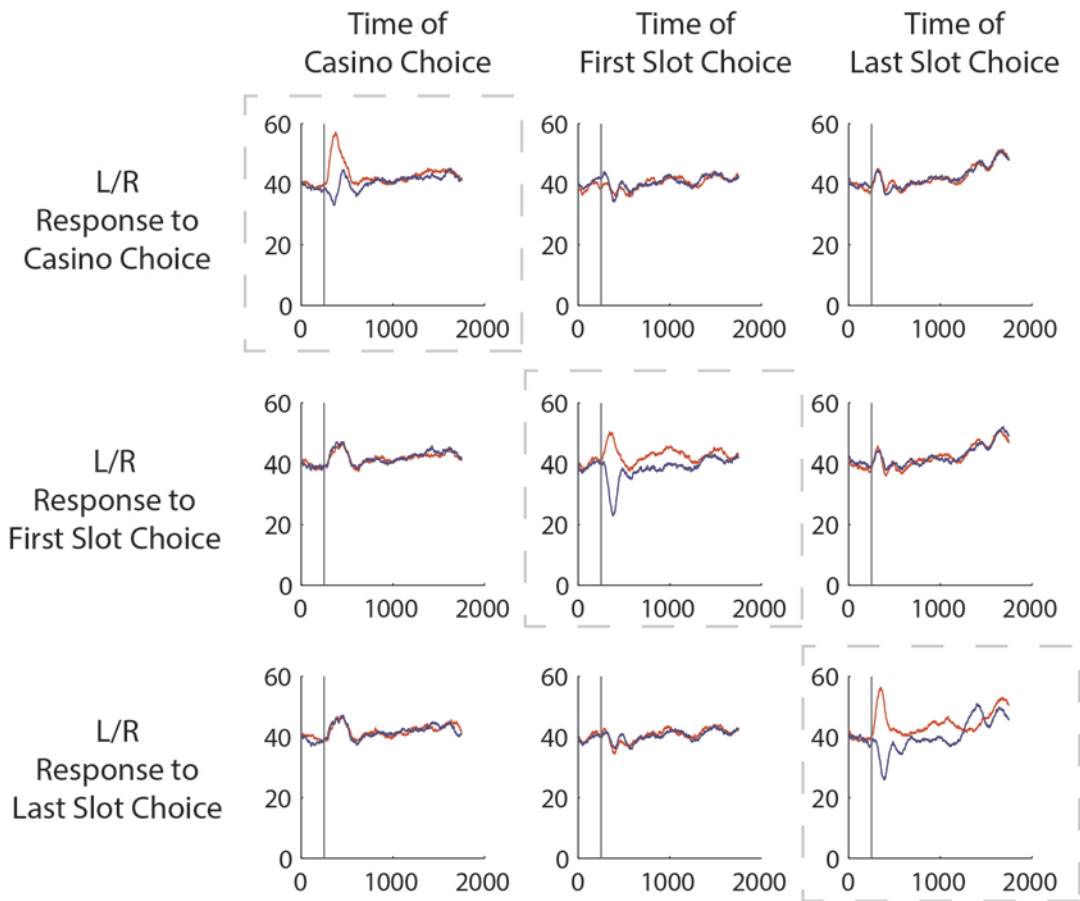


**Figure 4.1** Cell P121\_017a modulated its response based on the value of the chosen casino at the time of the casino choice. However, it did not encode the value of the slot machine at the time of the slot machine choice. Dark blue is the response to the highest value option; light blue is the 2<sup>nd</sup> highest value option; yellow represents the 2<sup>nd</sup> lowest option; orange is the lowest value option. For the chosen value, we combined the lowest value option with the second lowest value option, since there were few trials where the subject chose the lowest value option. Similarly, for the unchosen value, we combined the highest value with the second highest value option, since there were few trials where the subject did not choose the highest value option. Black vertical bar denotes time of stimulus onset. Abscissae correspond to neuronal firing rate in Hz and ordinate axes denote time in milliseconds.

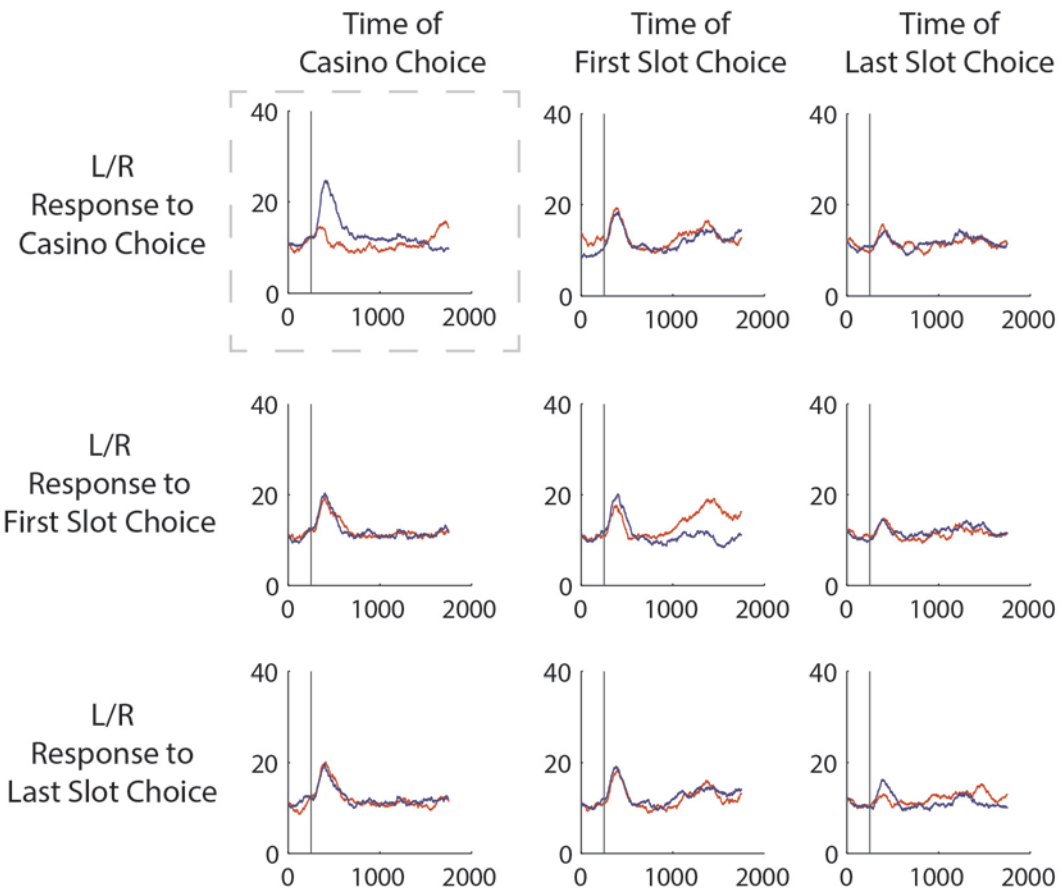




**Figure 4.2** Neuron L147\_25c, located in the right hemisphere of animal L responded selectively to the value of the contralateral casino (but not slot machines) by increasing its rate of discharge in response to higher value choices. Conventions are as in Figure 4.1.



**Figure 4.3** Neuron L147\_28a increased its firing rate whenever the subject chose the option on the right (red) and suppressed its firing rate when the subject chose the option on the left (blue). This pattern occurred at each choice. Conventions are as in Figure 4.1.



**Figure 4.4** Neuron L128\_27b encoded responses at the superordinate but not subordinate level. Black vertical bar denotes time of stimulus onset. Conventions as in Figure 4.1.

## 4.2 Neurophysiology: ACC

### 4.2.1 Value encoding in ACC

In total, 221 neurons were recorded from ACC, 112 from animal L and 109 from animal P. Further toward the goal of understanding how hierarchical information may be coded for in ACC, we again ran a stepwise regression to broadly search for value-related variables that were predictive of each neuron's firing rate at different task levels. Table 4.3 shows the percentage of neurons encoding each of these parameters for all three decision periods in the task. As with LPFC and OFC, the most prevalent value encoding related to the chosen values although there was not a difference in the proportion of neurons that encoded the chosen casino, first slot, or last slot machine values (casino: 43/221 or 19%; first slot: 48/221 or 22%; last slot machine: 39/221 or 18%, chi-squared test = 0.9,  $p > 0.1$ ). A neuron encoding the value of the chosen casino, by increasing its firing rate in response to higher values, is shown in Figure 4.5. There was no difference in the proportion of neurons that encoded the chosen casino in ACC as compared to LPFC or OFC (LPFC: 39/262; OFC: 35/249, chi-squared test = 2.4,  $p > 0.1$ ). However, the proportion of neurons in ACC encoding the value of the chosen first slot machine was significantly greater than that in LPFC (LPFC: 19/262, chi-squared test = 19.8,  $p < 1.0 \times 10^{-5}$ ) as well as OFC (OFC: 30/249, chi-squared test = 7.2,  $p < 0.05$ ). Further, the proportion of neurons in ACC encoding the value of the chosen last slot machine was significantly greater than that in LPFC (LPFC: 22/262, chi-squared test = 8.5,  $p < 0.005$ ) but *not* OFC (OFC: 32/249, chi-squared test = 1.7,  $p < 0.1$ ). Another set of variables encoded by neurons in ACC were related to the position of stimuli on the screen. Shown in Figure 4.6 is a neuron in the right hemisphere of animal P that modulated its firing rate in a graded manner based on the value of the ipsilateral (right) casino, with higher rates of discharge reflecting higher values. None of the proportions of neurons encoding ipsi/contra values at any of the three decision epochs in ACC differed significantly from LPFC or OFC (chi-squared tests,  $p > 0.1$  in all cases).

Next we examined whether variables at the time of the casino choice predicted neural activity at the time of the slot machine choices. To our stepwise regression, we added the seven variables related to value and the behavioral response for both the slot machine as well as the preceding casino for a total of 16 parameters. There were significant proportions of neurons that computed and encoded the difference between the chosen and the unchosen casino values at each subsequent choice epoch (first slot: 10/221; last slot: 10/221, binomial test,  $p < 1.0 \times 10^{-5}$  in both cases), potentially carrying information about the goodness of choice through to the time of reward. In addition to selectivity for the difference of the chosen and unchosen stimuli there were also populations of neurons in ACC that, at the time of the first slot machine choice, encoded the chosen casino (14/221 or 6%, binomial test,  $p < 0.05$ ).

In summary, ACC neurons play a prominent role in encoding the values necessary to perform well on our hierarchically structured probabilistic task. Although many ACC neurons encoded the superordinate choice, they differed from LPFC and OFC in also showing particularly strong encoding of subordinate choices.

#### 4.2.2 Response encoding in ACC

As with LPFC and OFC, the largest proportion of selective neurons we found in ACC encoded whether the subject would choose the left or right option. This was true for the casino choice as well as the slot machine choices, although response encoding appeared to get stronger as the trial progressed. Encoding of the behavioral response was strongest at the time of the last slot machine choice (83/221 or 38%, binomial test,  $p < 1.0 \times 10^{-15}$ ). While this group was not larger than the group of neurons encoding responses to the first slot machine choice (68/221 or 31%, chi-squared test = 1.8,  $p > 0.1$ ), it was significantly larger than the proportion of neurons encoding the response to the casino decision (57/221 or 26%, chi-squared test = 6.5,  $p = 0.01$ ). In comparison to the other areas, response encoding in ACC was consistently weaker than in LPFC (chi-squared test,  $p < 0.05$  in all cases), but did not differ statistically from response encoding in OFC (chi-squared tests,  $p > 0.1$  in all cases.)

Although there was a small population of ACC neurons that encoded superordinate but not subordinate choices (14/221 or 6%, binomial test,  $p < 1.0 \times 10^{-4}$ ) or superordinate but not subordinate choices (19/221 or 9%, binomial test,  $p < 1.0 \times 10^{-11}$ ) the majority of ACC response encoding neurons were selective for multiple choices. Indeed, more than one quarter of ACC neurons encoded the behavioral response at more than one time point. For every conjunction of time points tested, the number of response-encoding neurons was greater than chance (casino + first slot: 35/221 or 15%, binomial test,  $p < 1.0 \times 10^{-15}$ ; casino + last slot: 42/221 or 19%, binomial test,  $p < 1.0 \times 10^{-15}$ ; first slot + last slot: 53/221 or 24%, binomial test,  $p < 1.0 \times 10^{-15}$ ; casino + first slot + last slot: 34/221 or 15%, binomial test,  $p < 1.0 \times 10^{-15}$ ).

Finally, we examined whether there were cells in ACC whose activity reflected information about earlier responses at later points in time. Similar to LPFC and OFC, there was a population of neurons that encoded the casino response at the time of the first slot machine (28/221 or 13%, binomial test,  $p < 1.0 \times 10^{-10}$ ) and a population that maintained the response to the first slot machine choice at the time of the second or last slot machine choice (23/221 or 10%, binomial test,  $p < 1.0 \times 10^{-6}$ ).

#### 4.2.3 ACC Summary

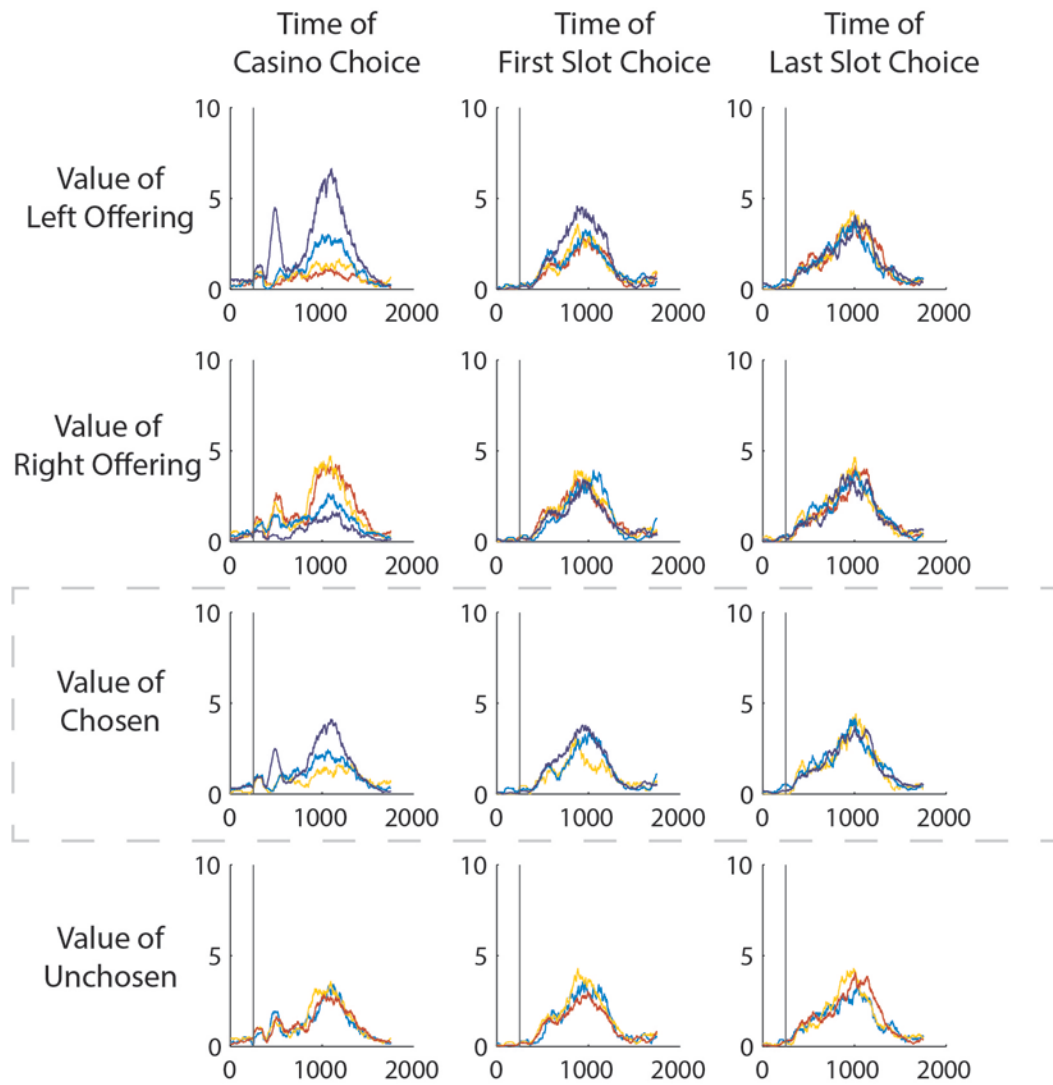
There were a number of features of ACC encoding that differentiated the neurons from the other areas. First, ACC strongly encoded the chosen value, even for the slot machines. Second, it maintained information about the value of the casinos across the slot machines. Finally, there was a progressive increase in response encoding as the trial progressed.

	Chosen	Unchosen	Chosen + Unchosen	Chosen – Unchosen	Ipsilateral	Contralateral	Ipsi – Contra	Response
<b>Fixation</b>	2.01%	2.81%	4.02%	4.02%	1.20%	3.21%	2.01%	4.42%
<b>Casino</b>	19.46%	4.52%	2.26%	8.14%	6.33%	15.84%	7.24%	25.79%
<b>First Slot</b>	21.72%	2.71%	5.43%	5.88%	6.79%	9.05%	8.60%	30.77%
<b>Last Slot</b>	17.65%	2.71%	3.62%	9.95%	5.58%	10.41%	6.33%	37.56%

**Table 4.3** The percentage of neurons in ACC that encoded values for the chosen, unchosen, sum of the chosen and unchosen; difference between the chosen and unchosen; ipsilateral stimulus to the hemisphere recorded from; contralateral stimulus to the hemisphere recorded from; the difference between the ipsilateral and contralateral stimuli; and the response (right or left) for the casino, first slot machine, and last slot machine at the time of choice. Percentages of neurons whose activity changes significantly during the fixation period according to different values are shown for reference. Percentages in red are significant, binomial test,  $p < 0.05$ .

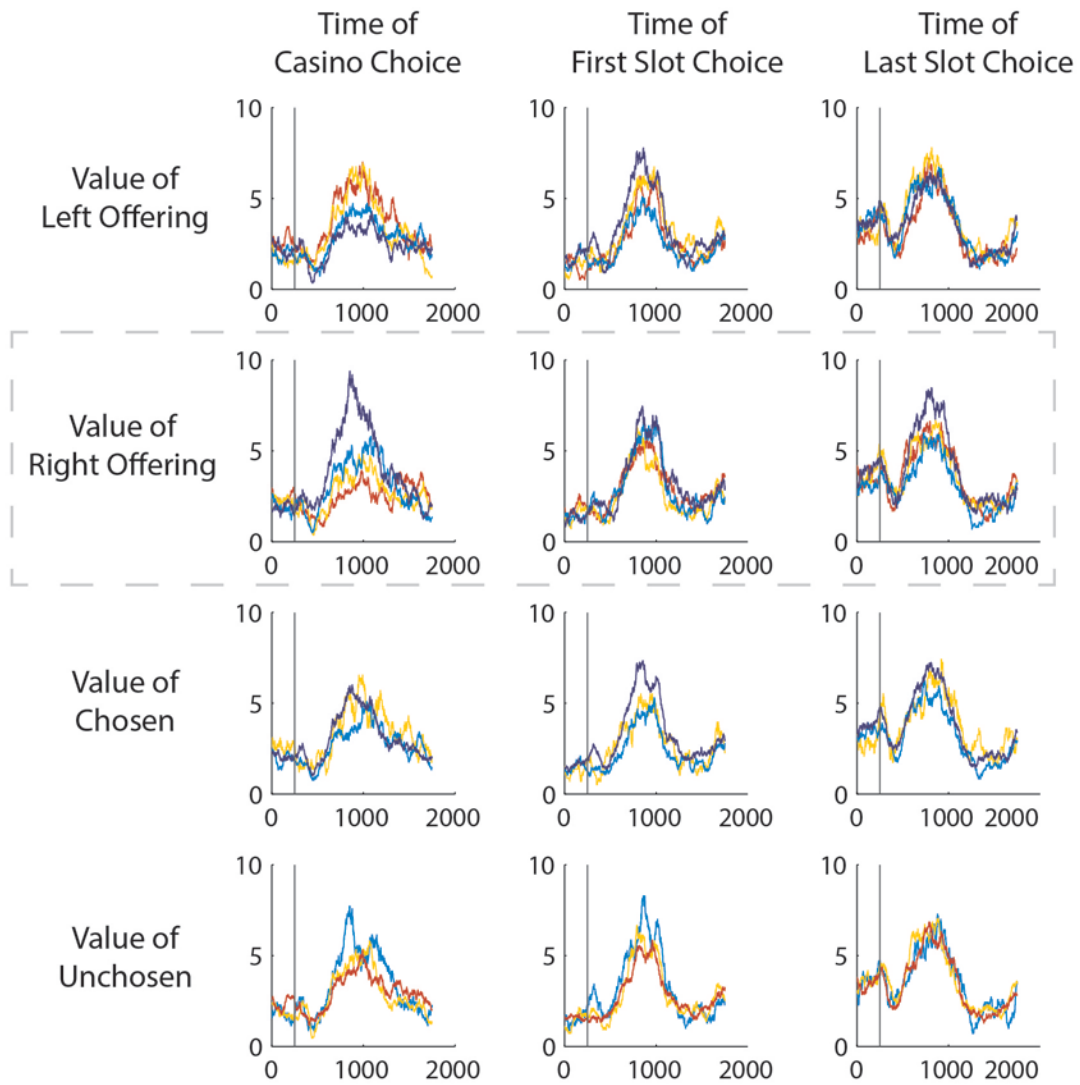
	Chosen	Unchosen	Chosen + Unchosen	Chosen – Unchosen	Ipsilateral	Contralateral	Ipsi- Contra	Response
<b>First Slot</b>	6.33%	2.71%	4.98%	4.52%	2.26%	4.98%	4.07%	12.67%
<b>Last Slot</b>	4.07%	4.52%	3.62%	4.52%	3.17%	3.62%	1.81%	3.17%

**Table 4.4** The percentage of neurons in ACC that encode the *casino* values for chosen, unchosen, sum of the chosen and unchosen; difference between the chosen and unchosen; ipsilateral stimulus to the hemisphere recorded from; contralateral stimulus to the hemisphere recorded from; the difference between the ipsilateral and contralateral stimuli; and the response (right or left) at the times of first and last slot machine choices. Percentages in red are significant, binomial test,  $p < 0.005$ .



**Figure 4.5** Neuron P132\_27a encoded the value of the chosen casino at the time of decision making but not during the subsequent slot machines. Conventions are as in Figure 4.1.





**Figure 4.6** Neuron P111\_25a, located in the right hemisphere of animal P responded selectively to the value of the ipsilateral casino (but not slot machines) by increasing its rate of discharge in response to higher value choices. Conventions are as in Figure 4.1.

## 5. Outcome and Prediction Error Encoding

In multi-step behaviors an animal may make several choices before experiencing the reinforcing outcomes that are inherently good or bad. While it is understood that primary rewards reinforce the immediately preceding actions in simple behaviors, it remains to be understood how temporally extended choices are reinforced such that the animal is able to attribute positive value to a superordinate choice across intervening subordinate actions. To determine how this process takes place, we contrasted neural activity at the time of feedback for the first slot machine choice with neural responses to the final outcome in the trial.

### 5.1 Subordinate choices

A key feature of reinforcement learning is the calculation of the prediction error, the difference between the value of what was expected to happen and the value of what actually happened. We examined whether neurons encoded this information at the time of feedback for the first slot machine (Table 5.1). Specifically, we examined whether they encoded a positive prediction error (the extent to which an outcome was better than expected) or a negative prediction error (the extent to which an outcome was worse than expected). For example, if the animal received a win when he chose the slot machine that was rewarding 35% of the time we would code this as a positive prediction error with a value of 0.65 ( $|1 - 0.35|$ ). If he had received a loss, it would have been coded as a negative prediction error with a value of 0.35 ( $|0 - 0.35|$ ). We also included a parameter that detected whether neurons encoded the outcome in a binary manner, i.e. if the neuron was simply whether or not the choice resulted in a win.

In each brain area, there were many neurons that fired according to whether the subordinate slot machine choice had resulted in a win (LPFC: 41/262 or 16%; OFC: 37/249 or 15%; ACC: 44/221 or 20%) though these populations did not differ significantly from one another (chi-squared test = 2.0,  $p > 0.1$  for all comparisons). There were also neurons that encoded positive prediction errors, particularly in OFC (LPFC: 22/262 or 8%; OFC: 40/262 or 16%; ACC: 22/221 or 10%). While there was no difference between ACC (22/221, binomial test,  $p < 0.05$ ) and either LPFC or OFC, significantly more neurons in OFC (40/249 or 16%) encoded positive prediction errors than those in LPFC (22/262 or 8%, chi-squared test = 6.3,  $p < 0.05$ ). Neurons that encoded negative prediction errors, or outcomes that were worse than expected, were relatively infrequent and only exceeded chance in ACC (13/221 or 6%, binomial test,  $p < 0.05$ ).

### 5.2 Superordinate Choices

To examine neural encoding of the final outcome, we regressed neural activity against the volume of juice delivered. In addition, because the reward magnitudes corresponding to the superordinate casino choices were probabilistic rather than deterministic, 25% of the time the amount of reward the animal received was different than he was expecting. Thus, we also included regressors indicating positive and negative prediction errors. As with the subordinate choices, there were neurons in all brain areas that encoded the outcome to the casino choice, that is the amount of juice delivered. One such neuron from LPFC is shown in Figure 5.1.

However, across the different neuronal populations, there were significantly more neurons in ACC (76/221 or 34%) and OFC (71/249 or 29%) that encoded the outcome of the casino choice than there were in LPFC (54/262 or 21%, chi-squared test = 3.9 and 10.9 respectively,  $p < 0.05$  for both comparisons).

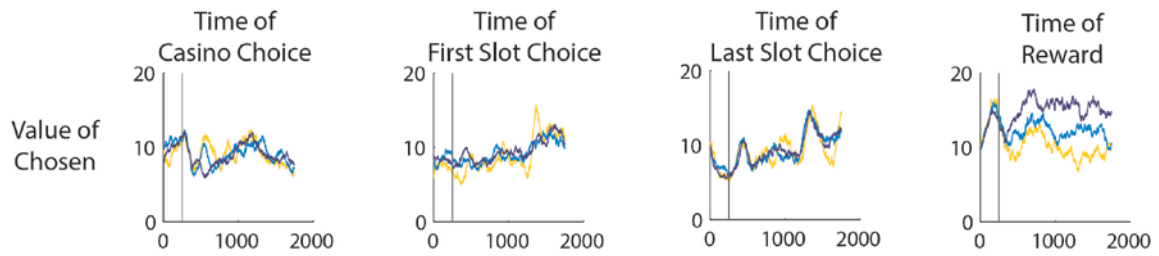
In all brain areas, populations of neurons encoding positive prediction errors did not exceed chance (LPFC: 8/262 or 3%; OFC: 14/249 or 6%; ACC: 10/221 or 5%, binomial test,  $p > 0.05$  in all cases). However, as with the subordinate choice epochs, in ACC there was a small, but statistically significant, population of neurons that encoded negative prediction errors (15/221 or 7%, binomial test,  $p < 0.05$ ). The proportion of such neurons did not exceed chance in OFC or LPFC (OFC: 5/249 or 2%; LPFC: 12/262 or 5%, binomial test,  $p > 0.05$  in both cases).

Finally, we compared the number of neurons encoding the most common signal (the binary outcome) at the time of either slot or casino feedback. There was no difference in the number of LPFC neurons encoding outcomes during slot or casino feedback (slot: 41/262 or 16%; casino: 54/262 or 21%, chi-squared test = 1.9,  $p > 0.1$ ), but there were significantly more neurons that encoded the outcome at the time of casino feedback compared to slot feedback in both OFC (casino: 71/249 or 29%, slot: 37/249 or 15%, chi-squared test = 12.9,  $p < 0.001$ ) and ACC (casino: 76/221 or 34%, slot: 44/221 or 20%, chi-squared test = 11,  $p < 0.001$ ).

<b>SLOT</b>	<b>OUTCOME</b>	<b>+RPE</b>	<b>-RPE</b>
LPFC	15.65%	8.40%	4.58%
OFC	14.86%	16.06%	4.82%
ACC	19.91%	9.95%	5.88%

<b>CASINO</b>	<b>OUTCOME</b>	<b>+RPE</b>	<b>-RPE</b>
LPFC	20.61%	3.05%	4.58%
OFC	28.51%	5.62%	2.01%
ACC	34.39%	4.52%	6.79%

**Table 5.1** Percentages of neurons in each brain area that encoded outcome related variables when feedback was received for the different choices of the task. Neurons were classified according to whether they encoded the value of the outcomes; positive prediction errors when outcomes were better than expected; and negative prediction errors when outcomes were worse than expected. Percentages in red are significant, binomial test,  $p < 0.05$ .



**Figure 5.1** Neuron P101\_024b encoded the value of the outcome when the reward was dispensed by increasing its firing rate in response to higher value outcomes. Dark blue is the response to the highest value option; light blue is the 2<sup>nd</sup> highest value option; yellow represents the lowest option. Although there appears to be a difference in the firing rate of this neuron for the value of the chosen casino, this effect was not significant ( $p > 0.05$ ). Black vertical bar denotes time of stimulus onset for the choice plots (first three plots) and the time of reward delivery for the reward plot (rightmost plot). Abscissae correspond to neuronal firing rate in Hz and ordinate axes denote time in milliseconds.

## 6. Discussion

### 6.1 Summary of findings

We trained two monkeys to perform a hierarchically-structured probabilistic bandit task. They demonstrated, by their optimal choices, that their decisions were driven in a goal-directed manner by the value of future rewards. Most importantly, our subjects were able to link the value of the final reward with the initial casino choice, despite the intervening slot choices. However, contrary to our original hypothesis, we did not find substantial evidence that PFC neurons favored the encoding of either subordinate or superordinate values or actions. On the other hand, we found considerable evidence that PFC neurons process information related to casino and slot machine levels differently. Many neurons encoded value information or the choice response during one epoch of the task and not another. Thus, although different regions of PFC are not biased to encode different levels of the task, the temporal structure of the task is still encoded in PFC activity.

We did notice some differences in neural encoding across the areas. Perhaps most relevant to our understanding of hierarchical behavior is the manner in which the different areas maintained information about the casino choice across the intervening slot machine choices. Most strikingly, OFC was the area that showed the clearest maintenance of value information from the time of the casino choice through to the final delivery of the reward. This encoding was weaker in ACC and virtually absent in LPFC. This information could serve as an eligibility trace, ensuring that the value of the chosen option is updated once the animal receives the final outcome.

There were other, more general trends that were apparent in all three areas that may be relevant to the representation of hierarchical information. Neuronal responses tended to show a greater correlation within hierarchical levels compared to across hierarchical levels. Thus, neurons that encoded either the value or response associated with the first slot machine tended to show similar encoding to the second slot machine, but necessarily to the casino. Such relationships reflect the hierarchical structure of the task in the neural representation. In addition, throughout the course of a trial, we observed a tendency for neurons to encode both the current choice, as well as the choice that was made on the previous choice. Thus, many neurons encoded the previous casino choice at the time of the first slot machine, and many neurons encoded the first slot machine choice at the time of the second slot machine choice. Such activity may conceivably be related to the mechanism by which individual actions are tied together to form options (Botvinick et al., 2009).

The areas also differed in the relative proportion of neurons that were encoding value-related information in comparison to action-related information. LPFC neurons showed stronger encoding of the choice response than ACC or OFC at all stages of the task. In comparison, ACC and OFC tended to encode more reward-related information than LPFC. For example, activity

reflecting winning outcomes to the subordinate choices was greater in OFC than LPFC. Further, the proportions of neurons in OFC that selectively encoded binary win/loss information and magnitude-dependent reward information were greater in both OFC and ACC compared to LPFC.

Finally, ACC showed several unique features that differentiated it from LPFC and OFC. First, it encoded information about chosen stimulus values, but information for subordinate level choices was stronger than in the other brain areas. In addition, response encoding seemed to increase as the trial progressed with more cells encoding each subsequent response, peaking at the final behavioral response preceding reward delivery

## 6.2 Findings in perspective

Though we did not find clear evidence that PFC distinguishes between different levels of hierarchical tasks, we did find that many neurons in different areas treated superordinate and subordinate task-relevant information differently and often specialized in one or the other. Our findings should be considered in the broader context of the empirical literature related to hierarchical behavior, neural organization, and PFC function.

In our task, casino level choices required more temporally extended maintenance of information in a goal-directed manner to compare outcomes with previous choices. Our prediction that LPFC may support the encoding of this information over intervening time was informed by previous work (Koechlin et al., 2003; Badre and D'Esposito, 2007) showing a functional-anatomical hierarchy supporting increasingly abstract information as one moves to progressively anterior regions of LPFC. We found no evidence that LPFC neurons were predisposed to encoding superordinate choices. There are two possible reasons why our data may not converge with previous findings. First, while the casino paradigm we used had multiple layers involving different contexts, it was not *truly* hierarchical in the sense that the subordinate states were not contingent upon the superordinate option in play. For example, previous work has emphasized the dependence of superordinate values on subordinate choices available therein (Glascher et al., 2010). The task we used maintained the two levels as nested but the reward contingencies associated with each level were purposefully independent of one another. Our strategy was to begin with the current task and work towards a truly hierarchical task in later experiments. This more measured approach will enable us to deconstruct the processes underlying hierarchical behavior. In particular, our task required that the subject link the final reward with a choice made at the beginning of a sequence of choices, while simultaneously monitoring the outcomes associated with intervening choices. It is therefore ideally suited for testing whether different PFC regions are responsible for representing choice behavior that evolves over either a long or short behavioral timeframe. Future experiments can now develop tasks in which subordinate choices depend on superordinate choices.

A second reason why we may not have observed hierarchical representations relates to the training that the animals had received in order to perform the task. The stimuli we used were well-learned and likely required working memory faculties to a lesser degree than if the value

contingencies were learned online during recording. These details likely contribute to a comparatively less abstract and less cognitively demanding behavior than if the animals were learning new stimuli with more complex reward contingencies. In addition, recent studies have emphasized how goal-directed behavior, thought to be dependent on PFC, transitions to a habitual system with repeated training, that is thought to depend on the striatum (Daw et al., 2005; Smith and Graybiel, 2013). In future work, we can study the processes underlying this transition by using chronic implants to record PFC neural activity throughout learning.

The ubiquitous encoding of responses in our data is consistent with recent work examining response encoding in PFC that has shown it to be pervasive and complex. That LPFC neurons strongly encode responses supports previous work (Seo et al., 2007; Wallis, 2007; Luk and Wallis, 2009; Tsujimoto et al., 2009). Previous research has been less in accord on findings of response encoding in OFC. Many studies, typically using simple decision-making tasks, have reported that OFC neurons do not encode the behavioral response related to the animal's choice (Tremblay and Schultz, 1999; Wallis and Miller, 2003; Padoa-Schioppa and Assad, 2006). However, there are some notable exceptions to this finding (Tsujimoto et al., 2009; Luk and Wallis, 2013), including the current results. A common feature of all of the tasks in which prominent response coding is observed in OFC is that they require information to be maintained over delays in order to evaluate whether the optimal response was made. Luk and Wallis (2013) used a task in which choice options were presented sequentially requiring the animal to remember, at the time of the outcome, which action was chosen and how that related to the options presented earlier in the trial. Similarly, the task used by Tsujimoto et al. (2009) required that the subject remembered which response had been made on the previous trial in order to determine the optimal response for obtaining reward on the current trial. Thus, OFC may be critically required for associating rewards with earlier actions.

ACC was the only area that appeared to encode subordinate choices more strongly than superordinate choices. This could conceivably relate to the hierarchical task structure. However, it might also relate to the different contingencies that we used for the subordinate and superordinate choices. Although both levels of our task are probabilistic, they are not necessarily probabilistic to the same extent. Recent work has suggested that ACC may be involved in encoding the volatility of probabilistic choices (Behrens et al., 2007). Although the casino choice led to a different magnitude of reward on 25% of the trials, the slot machine choices were potentially more volatile. The probabilistic contingencies were frequently more extreme for slot machine choices, and instead of leading to different sizes of reward, the effect was all-or-none: the animal either won a token or it did not. Thus, the stronger encoding of the subordinate choices in ACC may have reflected the increased volatility of the reward contingencies associated with those choices. We could test this in future experiments by reversing the volatility of the probabilistic contingencies. For example, the casino choice could be associated with the all-or-none delivery of the final reward, whereas the slot machine choices could be associated with the probabilistic delivery of a certain size of reward bar completion.



Also interesting, with respect to the data collected in ACC, is the trend of stronger encoding of responses increasingly closer to the time of reward. Studies of learning in rats (Balleine et al., 1995; Killcross and Coutureau, 2003) have suggested that actions more proximal to reward are more readily sensitive to devaluation, even in over-learned states, suggesting persistent control of behavior by goal-directed systems. Thus, the degree of neural encoding of actions in ACC may reflect the degree to which the action is under the control of the goal-directed system.

### **6.3 Future directions**

While we have contributed to the growing corpus of knowledge on reinforcement learning in PFC, questions remain to be addressed. The task we have presented here was feasible for animals to learn but to better address questions at the core of hierarchical reinforcement learning, it may be necessary use a more *truly hierarchical* adaptation emphasizing the learning process. Determining how neurons in multiple brain areas modulate their activity over the transition from naïve to well-learned environments will better elucidate the dynamic updating of state-dependent action values as well as the establishment of behavioral preferences. Because the superordinate option values in our task were not determined by state or action values at the subordinate level, in the well-learned state it may not have been as necessary to maintain the identity of the chosen option online for the purpose of updating.

Further, though we have observed lower-level learning signals in OFC and ACC, additional work is required to gain understanding about how temporally abstracted actions are initially chunked and organized into sensible subroutines. The subordinate level of our task also did not require unique action sequences contingent upon the higher-level option guiding behavior, unlike in the example from Chapter 1 of putting a teabag in a pot, pouring hot water into a teapot, etc., There are likely differences in the underlying neural mechanisms between behaviors necessary in our task and those more ethologically relevant to human behavior.

### **6.4 Closing remarks**

There is great clinical significance to the understanding of hierarchical behavior. Disorders such as Parkinson's disease, schizophrenia, and addiction involve dysfunction in learning and decision making. This thesis has examined the involvement of PFC in hierarchically structured decision making and behavior. Specifically, we have considered how single neurons in three distinct frontal brain regions—LPFC, ACC, and OFC—encode value, response, and outcome information enabling animals to make successful subordinate-level choices toward the achievement of a superordinate goal. Although we did not find evidence to support our original hypothesis that different PFC areas would be differentially involved in representing the hierarchical task, we nevertheless found many signals that could underlie hierarchical behavior, including value and action encoding that depended on the hierarchical level, as well as encoding of past choices that could be used to chunk actions at the same level of the hierarchy. Understanding these mechanisms could help elucidate how the complex behavioral repertoire of the primate is implemented.

## Bibliography

- Abe H, Lee D (2011) Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70:731-741.
- Alexander WH, Brown JW (2011) Medial prefrontal cortex as an action-outcome predictor. *Nat Neurosci* 14:1338-1344.
- Badre D, D'Esposito M (2007) Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J Cogn Neurosci* 19:2082-2099.
- Badre D, Hoffman J, Cooney JW, D'Esposito M (2009) Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nat Neurosci* 12:515-522.
- Badre D (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci* 12:193-200.
- Badre D, Kayser AS, D'Esposito M (2010) Frontal Cortex and the Discovery of Abstract Action Rules. *Neuron* 66:315-326.
- Balleine BW, Garner G, Gonzales F, Dickinson A (1995) Motivational Control of Heterogeneous Instrumental Chains. *J Exper Psych* 21:203-217.
- Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407-419.
- Barbas H, Pandya DN (1989) Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 286:353-375.
- Bechara A, Damasio H, Tranel D, Anderson SW (1998) Dissociation of working memory from decision making within the human prefrontal cortex. *J Neurosci* 18:428-437.
- Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214-1221.
- Bor D, Duncan J, Wiseman RJ, Owen AM (2003) Encoding Strategies Dissociate Prefrontal Activity from Working Memory Demand. *Neuron* 37:361-367.
- Botvinick MM, Niv Y, Barto AC (2009) Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* 113:262-280.

- Botvinick MM (2012) Hierarchical reinforcement learning and decision making. *Curr Opin Neurobiol* 22:956-962.
- Cai X, Padoa-Schioppa C (2012) Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *J Neurosci* 32:3791-3808.
- Camille N, Tsuchida A, Fellows LK (2011a) Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage. *J Neurosci* 31:15048-15052.
- Camille N, Griffiths CA, Vo K, Fellows LK, Kable JW (2011b) Ventromedial frontal lobe damage disrupts value maximization in humans. *J Neurosci* 31:7527-7532.
- Carmichael ST, Price JL (1995a) Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology* 363:642 - 664.
- Carmichael ST, Price JL (1995b) Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *J Comp Neurol* 363:615-641.
- Carmichael ST, Price JL (1996) Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *J Comp Neurol* 371:179-207.
- Cisek P, Kalaska JF (2010) Neural mechanisms for interacting with a world full of action choices. *Annual review of neuroscience* 33:269-298.
- Cooper R, Shallice T (2000) Contention scheduling and the control of routine activities. *Cogn Neuropsychol* 17:297-338.
- Croxson PL, Johansen-Berg H, Behrens TE, Robson MD, Pinski MA, Gross CG, Richter W, Richter MC, Kastner S, Rushworth MF (2005) Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *J Neurosci* 25:8854-8866.
- Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8:1704-1711.
- Dayan P, Niv Y (2008) Reinforcement learning: the good, the bad and the ugly. *Current opinion in neurobiology* 18:185-196.

- Diuk C, Tsai K, Wallis J, Botvinick M, Niv Y (2013) Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 33:5797-5805.
- Dum RP, Strick PL (1991) The origin of corticospinal projections from the premotor areas in the frontal lobe. *J Neurosci* 11:667-689.
- Dum RP, Strick PL (1996) Spinal cord terminations of the medial wall motor areas in macaque monkeys. *J Neurosci* 16:6513-6525.
- Duncan J, Emslie H, Williams P, Johnson R, Freer C (1996) Intelligence and the frontal lobe: the organization of goal-directed behavior. *Cognit Psychol* 30:257-303.
- Feierstein CE, Quirk MC, Uchida N, Sosulski DL, Mainen ZF (2006) Representation of spatial goals in rat orbitofrontal cortex. *Neuron* 51:495-507.
- Frank MJ, O'Reilly RC (2006) A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. *Behav Neurosci* 120:497-517.
- Furuyashiki T, Holland PC, Gallagher M (2008) Rat orbitofrontal cortex separately encodes response and outcome information during performance of goal-directed behavior. *J Neurosci* 28:5127-5138.
- Fuster JM (1997) *The Prefrontal Cortex: Anatomy, Physiology and Neuropsychology of the Frontal Lobe*. Philadelphia, PA: Lippincott-Raven.
- Fuster JM (2001) The prefrontal cortex--an update: time is of the essence. *Neuron* 30:319-333.
- Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL (1999) Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci* 2:861-863.
- Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585-595.
- Goldman-Rakic PS (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: *Handbook of physiology, the nervous system, higher functions of the brain* (Plum F, ed), pp 373 - 417. Bethesda, MD: American Physiological Society.

- Hampshire A, Thompson R, Duncan J, Owen J (2011) Lateral Prefrontal Cortex Subregions Make Dissociable Contributions during Fluid Reasoning. *Cereb Cortex* 21:1-10.
- Hare TA, Camerer CF, Rangel A (2009) Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324:646-648.
- Hayden BY, Platt ML (2010) Neurons in anterior cingulate cortex multiplex information about reward and action. *J Neurosci* 30:3339-3346.
- Hoshi, E (2006) Functional specialization within the dorsolateral prefrontal cortex: A review of anatomical and physiological studies of non-human primates. *Neurosci Res* 54:73-84.
- Humphreys GW, Forde EME (1998) Disordered action schema and action disorganization syndrome. *Cogn Neuropsychol* 15:771-811.
- Ichihara-Takeda S, Funahashi S (2008) Activity of primate orbitofrontal and dorsolateral prefrontal neurons: effect of reward schedule on task-related activity. *Journal of cognitive neuroscience* 20:563-579.
- Ito S, Stuphorn V, Brown JW, Schall JD (2003) Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* 302:120-122.
- Jackson JH (1884) The Croonian Lectures on Evolution and Dissolution of the Nervous System. *British medical journal* 1:660-663.
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263 - 291.
- Kamin LJ (1969) Predictability, surprise, attention and conditioning. In: Punishment and aversive behavior (Campbell BA, Church RM, eds). New York, NY: Appleton-Century-Crofts.
- Kane MJ, Engle RW (2003) Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *J Exp Psychol Gen* 132:47-70.
- Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1:411-416.
- Kennerley SW, Wallis JD (2009) Reward-dependent modulation of working memory in lateral prefrontal cortex. *J Neurosci* 29:3259-3270.

- Kennerley SW, Behrens TE, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature neuroscience* 14:1581-1589.
- Kennerley SW, Dahmubed AF, Lara AH, Wallis JD (2009) Neurons in the frontal lobe encode the value of multiple decision variables. *J Cogn Neurosci* 21:1162-1178.
- Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex* 13:400-408.
- Koechlin E, Ody C, Kouneiher F (2003) The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science* 302:1181-1185.
- Kondo H, Saleem KS, Price JL (2005) Differential connections of the perirhinal and parahippocampal cortex with the orbital and medial prefrontal networks in macaque monkeys. *J Comp Neurol* 493:479-509.
- Lara AH, Kennerley SW, Wallis JD (2009) Encoding of gustatory working memory by orbitofrontal neurons. *J Neurosci* 29:765-774.
- Lashley KS (1951) The problem of serial order in behavior. In: *Cerebral Mechanisms in Behavior*, pp 112-136. New York, NY: John Wiley & Sons.
- Lee D, Seo H, Jung MW (2012) Neural basis of reinforcement learning and decision making. *Annual review of neuroscience* 35:287-308.
- Leslie JC (2006) Herbert Spencer's contributions to behavior analysis: a retrospective review of principles of psychology. *J Exp Anal Behav* 86:123-129.
- Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND (2011) Differential roles of human striatum and amygdala in associative learning. *Nature neuroscience* 14:1250-1252.
- Luk CH, Wallis JD (2009) Dynamic encoding of responses and outcomes by neurons in medial prefrontal cortex. *J Neurosci* 29:7526-7539.
- Luk CH, Wallis JD (2013) Choice coding in frontal cortex during stimulus-guided or action-guided decision-making. *J Neurosci* 33:1864-1871.
- MacKay DG (1987) *The Organization of Perception and Action: A theory for language and other cognitive skills*. Berlin: Springer-Verlag.

- Mackey S, Petrides M (2010) Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *The European journal of neuroscience* 32:1940-1950.
- Matsumoto K, Suzuki W, Tanaka K (2003) Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301:229 - 232.
- Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. *Nat Neurosci* 10:647-656.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167-202.
- Morecraft RJ, Geula C, Mesulam MM (1992) Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey. *J Comp Neurol* 323:341-358.
- Morrison SE, Salzman CD (2009) The convergence of information about rewarding and aversive stimuli in single neurons. *J Neurosci* 29:11471-11483.
- Mountcastle VB, Lynch JC, Georgopoulos A, Sakata H, Acuna C (1975) Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *Journal of neurophysiology* 38:871-908.
- Nimchinsky EA, Hof PR, Young WG, Morrison JH (1996) Neurochemical, morphologic, and laminar characterization of cortical projection neurons in the cingulate motor areas of the macaque monkey. *The Journal of comparative neurology* 374:136-160.
- Ongur D, Price JL (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex* 10:206-219.
- Ostlund SB, Balleine BW (2007) Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *J Neurosci* 27:4819-4825.
- Padoa-Schioppa C (2011) Neurobiology of economic choice: a good-based model. *Annual review of neuroscience* 34:333-359.
- Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441:223-226.

- Padoa-Schioppa C, Assad JA (2008) The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat Neurosci* 11:95-102.
- Pan X, Sawa K, Tsuda I, Tsukada M, Sakagami M (2008) Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. *Nat Neurosci* 11:703-712.
- Pavlov IP (1927) Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences* 17:136-141.
- Penfield W, Evans J (1935) The frontal lobe in man: A clinical study of maximum removals. *Brain* 68:115 - 133.
- Petrides M, Pandya DN (1984) Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J Comp Neurol* 228:105-116.
- Petrides M, Pandya DN (1999) Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *Eur J Neurosci* 11:1011-1036.
- Petrides M, Pandya DN (2002) Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur J Neurosci* 16:291-310.
- Pickens CL, Saddoris MP, Setlow B, Gallagher M, Holland PC, Schoenbaum G (2003) Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *J Neurosci* 23:11078-11084.
- Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233-238.
- Rangel A, Hare T (2010) Neural computations associated with goal-directed choice. *Current opinion in neurobiology* 20:262-270.
- 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.
- Roesch MR, Bryden DW, Cerri DH, Haney ZR, Schoenbaum G (2012) Willingness to wait and altered encoding of time-discounted reward in the orbitofrontal cortex with normal



- aging. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32:5525-5533.
- Rolls ET (1996) The orbitofrontal cortex. *Philos Trans R Soc Lond B Biol Sci* 351:1433-1443; discussion 1443-1434.
- Rolls ET, Hornak J, Wade D, McGrath J (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J Neurol Neurosurg Psychiatry* 57:1518-1524.
- Rose JE, Woolsey CN (1948) The orbitofrontal cortex and its connections with the mediodorsal nucleus in rabbit, sheep and cat. *Res Publ Assoc Res Nerv Ment Dis* 27 (1 vol.):210-232.
- Rudebeck PH, Behrens TE, Kennerley SW, Baxter MG, Buckley MJ, Walton ME, Rushworth MF (2008) Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J Neurosci* 28:13775-13785.
- Rushworth MF, Behrens TE, Rudebeck PH, Walton ME (2007) Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends Cogn Sci* 11:168-176.
- Sallet J, Quilodran R, Rothe M, Vezoli J, Joseph JP, Procyk E (2007) Expectations, gains, and losses in the anterior cingulate cortex. *Cogn Affect Behav Neurosci* 7:327-336.
- Schank RC, Abelson R (1977) *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593-1599.
- Schultz W, Tremblay L, Hollerman JR (2000) Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereb Cortex* 10:272-284.
- Seo H, Lee D (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci* 27:8366-8377.
- Seo H, Lee D (2009) Behavioral and neural changes after gains and losses of conditioned reinforcers. *J Neurosci* 29:3627-3641.
- Seo H, Barraclough DJ, Lee D (2007) Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cereb Cortex* 17 Suppl 1:i110-117.

- Skinner BF (1930) On the conditions of elicitation of certain eating reflexes. *Proceedings of the National Academy of Sciences of the United States of America* 16:433-438.
- Smith KS, Graybiel AM (2013) A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron* 79:361-374.
- Spencer H (1855) *The Principles of Psychology*. London: Longman, Brown, Green and Longmans.
- Sul JH, Kim H, Huh N, Lee D, Jung MW (2010) Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* 66:449-460.
- Sutton RS (1988) Learning to predict by the methods of temporal differences. *Machine Learning* 3:9-44.
- Sutton RS, Barto AG (1981) Toward a modern theory of adaptive networks: expectation and prediction. *Psychol Rev* 88:135-170.
- Thorndike EL (1911) *Animal Intelligence: Experimental studies*. New York, NY: MacMillan.
- Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. *Nature* 398:704-708.
- Tsujimoto S, Sawaguchi T (2004) Neuronal representation of response-outcome in the primate prefrontal cortex. *Cereb Cortex* 14:47-55.
- Tsujimoto S, Genovesio A, Wise SP (2009) Monkey orbitofrontal cortex encodes response choices near feedback time. *J Neurosci* 29:2569-2574.
- Uylings HB, Groenewegen HJ, Kolb B (2003) Do rats have a prefrontal cortex? *Behavioural brain research* 146:3-17.
- Valentin VV, O'Doherty JP (2009) Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. *J Neurophysiol* 102:3384-3391.
- van Wingerden M, Vinck M, Lankelma JV, Pennartz CM (2010) Learning-associated gamma-band phase-locking of action-outcome selective neurons in orbitofrontal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30:10025-10038.

- Von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton: Princeton University Press.
- Wallis JD (2007) Neuronal Mechanisms in Prefrontal Cortex Underlying Adaptive Choice Behavior. *Ann N. Y. Acad Sci* 1121:447-460.
- Wallis JD (2007) Orbitofrontal cortex and its contribution to decision-making. *Annu Rev Neurosci* 30:31-56.
- Wallis JD (2012) Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature neuroscience* 15:13-19.
- Wallis JD, Miller EK (2003) Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur J Neurosci* 18:2069-2081.
- Wallis JD, Kennerley SW (2010) Heterogeneous reward signals in prefrontal cortex. *Curr Opin Neurobiol* 20:191-198.
- Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953-956.
- Wiest G (2012) Neural and mental hierarchies. *Frontiers in psychology* 3:516.
- Williams SM, Goldman-Rakic PS (1993) Characterization of the dopaminergic innervation of the primate frontal cortex using a dopamine-specific antibody. *Cereb Cortex* 3:199-222.
- Williams ZM, Bush G, Rauch SL, Cosgrove GR, Eskandar EN (2004) Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat Neurosci* 7:1370-1375.
- Wunderlich K, Rangel A, O'Doherty JP (2010) Economic choices can be made using only stimulus values. *Proceedings of the National Academy of Sciences of the United States of America* 107:15005-15010.
- Yamagata T, Nakayama Y, Tanji J, Hoshi E (2012) Distinct Information Representation and Processing for Goal-Directed Behavior in the Dorsolateral and Ventrolateral Prefrontal Cortex and the Dorsal Premotor Cortex. *J Neurosci* 32(37):12934-12949.