

Permanent Contact: nisha.cooch@gmail.com

## NISHA K. COOCH

### EDUCATION

---

2008-2014	<b>UNIVERSITY OF MARYLAND SCHOOL OF MEDICINE</b> Ph.D. Candidate in Neuroscience Program	BALTIMORE, MD
2005-2007	Passed Qualifying Exam, Advanced to Candidacy, 2010 <b>GEORGETOWN UNIVERSITY</b>	WASHINGTON, DC
2001-2005	Post Baccalaureate Pre-Medical Program <b>WASHINGTON AND LEE UNIVERSITY</b> Bachelor of Science Degree in Psychology, <i>Cum Laude</i> Nominated as Peer Writing Tutor, 2004 Kappa Kappa Gamma Sorority Member (2002-2005) and Social Chair (2002-2003)	LEXINGTON, VA

### RESEARCH EXPERIENCE

---

2008-present	<b>UNIVERSITY OF MARYLAND SCHOOL OF MEDICINE</b> <b>NATIONAL INSTITUTE ON DRUG ABUSE</b> <i>Dissertation Student</i> (Department of Anatomy & Neurobiology) – Use <i>in vivo</i> electrophysiological techniques to investigate mechanisms by which the reward system encodes value information to support decision making	BALTIMORE, MD
2006-2007	<b>GEORGETOWN UNIVERSITY</b> <i>Research Assistant</i> (Department of Pharmacology) – Studied the role of hippocampus in valuing outcomes using fMRI	WASHINGTON, DC
2005	<b>WASHINGTON &amp; LEE UNIVERSITY</b> <i>Co-Researcher in Neuropsychology</i> (Department of Neuroscience) – Investigated the effects of alcohol abuse on attention and memory in college students	LEXINGTON, VA
Summer 2004	<b>UNIVERSITY OF VIRGINIA</b> <i>Research Assistant</i> (Department of Neurology) – Explored techniques used to identify mismatched mitochondrial DNA	CHARLOTTESVILLE, VA

### TEACHING EXPERIENCE

---

2013-present	<b>TOP TEST PREP</b> <i>Instructor &amp; Consultant</i> – Prepare students for their Medical College Admissions Tests (MCAT) and Law School Admissions Tests (LSAT) and counsel on application processes	BALTIMORE, MD
2009-present	<b>UNIVERSITY OF MARYLAND SCHOOL OF MEDICINE</b> <i>Volunteer Lab Instructor</i> – Teach 1 <sup>st</sup> year medical students in Gross Anatomy course <i>Tutor</i> – Write lesson plans and tutor students in preparation for their Medical College Admissions Tests (MCAT)	BALTIMORE, MD
2006	<b>GEORGETOWN UNIVERSITY</b> <i>Teacher's Assistant</i> (Department of Mathematics) – Taught students in Probability and Statistics	WASHINGTON, DC

### CLINICAL EXPERIENCE

---

2007-2008	<b>CLINICAL SKIN CENTER</b> <i>Medical Assistant</i> – Prepared and assisted in examinations, biopsies, and surgeries; Managed administrative tasks	FAIRFAX, VIRGINIA
Summer 2005	<b>LINDAMOOD-BELL LEARNING PROCESSES</b> <i>Clinician</i> – Worked one on one with children to promote cognitive functioning in accordance with mental deficits	WASHINGTON, DC

### CURRENT ACTIVITIES

---

2013-present	<b>NIH ENTREPRENEUR &amp; COMMERCIALIZATION CLUB</b> <i>Member</i> - determine the commercial viability of bio-health technologies and development strategies	BETHESDA, MD
2010-present	<b>ALUMNI ASSOCIATIONS PROGRAM</b> <i>Member</i> – Conduct interviews for local prospective students of Washington & Lee University	BALTIMORE, MD
2009-present	<b>PROGRAM IN NEUROSCIENCE TRAINING COMMITTEE</b> <i>President</i> – Present student recommendations on curriculum to faculty; Organize and recruit for community outreach events	BALTIMORE, MD

Dissertation Title: Mechanisms of Valuation: Encoding of Outcome Variables in Orbitofrontal Cortex and Ventral Striatum

Nisha Kaul Cooch, Doctor of Philosophy, 2014

Dissertation directed by Geoffrey Schoenbaum, National Institute on Drug Abuse and Joseph Cheer, University of Maryland School of Medicine

### **Abstract**

Adaptive decision making requires that we consider not only the inherent value of our options but also more specific features of those options, which can be used to compute the current value of each choice in a dynamic environment. General value information is conveyed via ‘model-free’ signaling, whereas details of outcomes that are independent of value are conveyed via ‘model-based’ representations of outcomes. Orbitofrontal cortex (OFC) and ventral striatum (VS), which function as part of the reward system, have been implicated in value-guided behaviors, but their specific contributions to model-free and model-based signaling remains elusive. Several studies suggest that OFC is not critical for distinguishing differentially valued outcomes of a common currency. Instead, encoding of specific features of outcomes in OFC appears to provide the flexibility required for advantageous choice selection in several value-guided behaviors. However, VS has been shown to be essential when either value or specific feature information is necessary for adaptive behavior.

In the following set of experiments, we tested the hypothesis that OFC signals model-based information, which is incorporated with model-free signals in VS. To isolate

model-free and model-based signals, we independently manipulated size (value) and flavor (specific feature) of rewards while recording single units in the rat OFC (Experiment 1) or VS (Experiment 2). Our data provide evidence for model-based signals in OFC and evidence for a hybrid of model-based and model-free signals in VS. Whereas OFC lesions disrupted model-based representations of flavor in VS, they did not eliminate all model-based signaling. We therefore conclude that OFC provides some, but not all, of the model-based information in VS.

**Mechanisms of Valuation:  
Encoding of Outcome Variables in Orbitofrontal Cortex and Ventral Striatum**

Author: Nisha Kaul Cooch

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, Baltimore in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2014

## Acknowledgments

### **Schoenbaum Lab and Collaborators**

Tom Stalnaker, PhD

Tzu-Lan Liu

Heather Wied

Federica Luncantonio, PhD

Guillermo Esber, PhD

Mihaela Iordanova, PhD

Mike McDannald, PhD

Joshua Jones, PhD

Sheena Chaudary

Alex Hernandez

Yuji Takahash, PhD

Teghi Singh, PhD

Donna Calu, PhD

### **Thesis Committee**

Geoffrey Schoenbaum MD, PhD

Joseph Cheer, PhD Patricio

O'Donnell MD, PhD Greg

Elmer, PhD

Todd Gould MD

### **GPILS at University of Maryland School of Medicine**

Jennifer Aumiller

Rachael Holmes

### **National Institute on Drug Abuse**

## **Table of Contents**

Chapter 1: Model-Based and Model-Free Components of Decision Making	1
Chapter 2: Model-Based Signaling of Outcomes in Orbitofrontal Cortex	15
Chapter 3: Model-Based Signaling of Outcomes in Ventral Striatum	38
Chapter 4: General Discussion of Valuation Processes	58
References	66

## List of Figures

Figure 1. Experimental design.....	20
Figure 2. Influence of reward size and flavor on free choice behavior.....	26
Figure 3. Influence of reward size and flavor on forced choice behavior.....	27
Figure 4. Recording sites in rat OFC.....	28
Figure 5. Outcome selectivity in OFC.....	30
Figure 6. Size and flavor indices for each OFC neuron.....	31
Figure 7. Signaling of Reward Configuration.....	33
Figure 8. Change in OFC firing pattern during inference of new outcome.....	35
Figure 9. Relationship between OFC task structure signals and subsequent choice behavior.....	36
Figure 10. Influence of reward size and flavor on free choices.....	46
Figure 11. Influence of reward size and flavor on accuracy in forced choices.....	47
Figure 12. Influence of reward size and flavor on reaction time in forced choices.....	48
Figure 13. Changes in free choice behavior across block transitions.....	48
Figure 14. Flavor and size selectivity in VS of control rats.....	51
Figure 15. Block-specific activity in VS of control rats.....	52
Figure 16. Change in VS firing pattern during inference of new outcome.....	53
Figure 17. Flavor and size selectivity in VS of OFC-lesioned rats.....	54
Figure 18. Block-specific activity in VS of OFC-lesioned rats.....	55
Figure 19. Block-specific activity in VS, according to factor.....	55
Figure 20. Change in VS firing pattern in OFC-lesioned rats during inference.....	56

## **CHAPTER 1: MODEL-BASED AND MODEL-FREE COMPONENTS OF DECISION MAKING**

The goal of any decision, no matter how complex, is to maximize gains and minimize losses. Our ability to make effective choices is critical to our well-being and, whether we realize it or not, we each make thousands of decisions every day. It is thus not surprising that a wealth of industries, from medicine and science, to economics and finance, to law and political science, have a vested interest in decision analysis for the purposes of assessing, predicting, and improving decision making. Strides in the machine learning subfield of computer science have been particularly useful for the study of decision making among neuroscientists, largely due to the apparent neural instantiation of algorithms developed to support intelligent technology. One such algorithm, the temporal difference reinforcement learning (TDRL) algorithm, enables machines that interact with their surroundings to learn the consequences associated with stimuli in their environment and thereby make effective choices. In a seminal study, Wolfram Schultz demonstrated in primates that midbrain dopamine (DA) neurons mimic prediction error signals postulated by TDRL (Schultz, Apicella et al. 1992). As a consequence of Schultz's work and algorithmic processing theories proposed by behavioral economists, neuroscientists became well-positioned to investigate the neural mechanisms of decision making by simultaneously collecting choice data and their neural correlates.

### **Choices Based on Model-Free Algorithms**

'Model-free' implies freedom from information that is independent of inherent value. Thus, the output of model-free algorithms is common currency value (Sutton and Barto 1990). A machine programmed with a model-free TDRL algorithm works in the following way when introduced to a new environment. First, it acts arbitrarily, making choices at random, with no expectations regarding outcomes. When an action results in reward, the outcome, which is better-than-expected, generates a positive reward prediction error (RPE) that updates the machine's expectation about the outcome of the preceding action. Following Thorndike's 'law of effect,' which claims that rewarded actions tend to be repeated (Thorndike 1898), the machine becomes more likely to choose that action in the future. TDRL can also reduce the likelihood of a given choice by way of

negative RPEs, which occur when outcomes are worse than expected. This may occur when a choice associated with no expectation of outcome results in punishment or when a choice that has become associated with reward produces no reward or less reward than expected.

In this way, the machine learns through experience the ‘cached’ value of its options, which allows it to choose those options with the highest likelihood of producing net gains for the system. However, as differences can only be quantified when compared in common units, RPEs signal the difference in expected and experienced outcomes in a common currency. Thus, TDRL can only account for choices between options leading to outcomes that can be directly compared on the same scale. Neural signals that behave like RPEs are therefore likely supplemented by other signals that, on their own, or in collaboration with TDRL signals, generate choice behaviors that model-free signals alone cannot explain. Such behaviors are described below.

### **Limitations of Model-Free Information**

In some cases, TDRL reliably predicts choice behavior. However, as with many economic models, there are other cases where TDRL predicts choice behavior with only moderate accuracy, as well as cases in which TDRL predictions are reliably inconsistent with real life choice behavior. In other words, there are certain situations in which we tend to choose what would be considered in TDRL as the less valuable option, and thus TDRL algorithms cannot explain these choices.

When the winner of the Nobel Prize in Economics, Daniel Kahnman and his colleague, Amos Tversky, described Prospect Theory (Kahnman and Tversky 1979), they alluded to situations in which ‘expected value’ fails to explain our choice behavior. Choice based on expected value, as proposed by Blaise Pascal to be the product of the value of the expected outcome and the probability of receiving that outcome (Giordano, Benedikter et al. 2012), is consistent with choice behavior predicted by TDRL. Indeed, TDRL ensures choice with the highest likelihood of acquiring the greatest gain. However, as noted by Kahnman and Tversky (Kahneman and Tversky 1984), when given a choice between A)

an 85% chance of winning \$1000, and B) a 100% chance of winning \$800, people overwhelmingly choose option B, despite option A having a greater expected value ( $0.85 * \$1000 = \$850$ ) than option B ( $1 * \$800 = \$800$ ).

Game theory also provides several examples of situations in which we violate TDRL predictions. One such case occurs in the Ultimatum Game (Guth, Werner et al. 1982). In this task, two people are randomly paired, and one is assigned as the proposer and the other as the responder. The proposer is given a sum of money, and the pair is told they have a single chance to make a deal. The proposer suggests a proportion of the money that each player will receive. If the responder accepts, each player receives the proposed amount, but if the responder rejects the offer, both players forfeit the full sum. When the proposer offers less than 20% of the sum, the responder rejects the offer more than half the time. This effect has been a difficult one to explain, as economists reason that any money is more valuable than no money, and thus responders should accept any offer. A machine behaving in accordance with a TDRL algorithm would certainly choose an option leading to some positive amount of money over an option leading to no money. However, humans and monkeys reliably violate the TDRL choice model.

In response to the failure of ‘expected value’ to accurately predict behavior, Daniel Bernoulli suggested that choice may instead result from ‘expected utility,’ which represents expected subjective value and is computed as the product of the utility of the expected outcome and the probability of receiving that outcome (Bernoulli 1738). Unlike value, utility is often marginal, meaning it does not increase linearly with each additional unit. Thus, utility takes into account that increasing winnings by \$1000 when original winnings were \$0 is not equivalent to increasing winnings by \$1000 when original winnings were \$5000. Though expected utility theory, unlike expected value theory, can explain certain behaviors, particularly in gambling contexts (Hartley and Farrell 2002), it has been demonstrated by Allais’ Paradox and Ellsberg’s Paradox that expected utility also oversimplifies choice behavior.

The framing effect exemplifies the failure of expected utility (and of TDRL) to predict behavior.

One example of the framing effect, as described by Kahnman and Tversky (Kahnman and Tversky 1979), who studied different manifestations of the framing effect, is as follows: In response to the news that a disease outbreak is expected to kill 600 people, 2 programs are proposed:

- In Program 1, 200 people will be saved.

- In program 2, there is a  $1/3$  chance that 600 people will be saved, and there is a  $2/3$  chance that 0 people will be saved.

Given these options, people tend to choose Program 1.

However, if instead of Programs 1 and 2, people are instead offered Programs A and B, choice behavior changes.

- In Program A, 400 people die.

- In Program B, there is a  $1/3$  chance that nobody will die and a  $2/3$  chance that 600 people will die.

Given these options, people tend to choose Program B. Despite that upon close inspection, Programs 1 and A are equivalent, and Programs 2 and B are equivalent, people reliably choose different options, depending on how the options are framed. Neither expected utility nor TDRL have a mechanism for distinguishing Program 1 from Program A or Program 2 from Program B. Thus, utilizing either of these choice strategies would result in a 50% choice rate for each option, an outcome inconsistent with empirical data for choice trends.

### **Choices Incorporating Model-Based Representations of Outcomes**

With numerous demonstrations of choice behavior reliably deviating from behavior predicted by TDRL, it is clear that we represent potential outcomes in a way that is not fully captured by this model-free algorithm. Our dynamic representations of outcomes are therefore thought to result from a combination of model-free and model-based representations. Goal-directed behavior that is prospective and flexible depends on model-based representations of potential outcomes. These representations take into

account factors such as specific sensory information, context, and motivational state, independent of value. Such encoding facilitates optimal choice behavior in new or changing environments and explains behaviors for which model-free accounts fall short. Model-based algorithms are particularly useful for choices that require higher order cognitive functions such as inferential thinking. For example, model-based algorithms, unlike model-free algorithms, can explain adaptive choices between options that have not been experienced in the past.

There are several behavioral tasks, such as reinforcer devaluation, in which people and animals demonstrate their ability to infer consequences of cues that have not before been paired with those cues (Gallagher and Schoenbaum 1999, Izquierdo and Murray 2000, Pickens, Saddoris et al. 2005, Valentin, Dickinson et al. 2007, West, DesJardin et al. 2011). In reinforcer devaluation, we train animals that a light cue leads to a food reward. Once the association is learned, we devalue the specific food reward in a separate setting, independent of the light, by inducing illness once the food is consumed. When we reintroduce rats to the light, they respond less to that cue, revealing their understanding that the cue leads to an outcome whose value has been degraded. Model-free algorithms could only explain such an altered response to the light cue if illness had previously followed presentation of the light. However, the change in response to the light occurs independent of such an experience and thus cannot be explained by these model-free representations, in which values or policies are pre-computed or cached during prior experience. Instead, appropriate responding to the cue after outcome devaluation requires access to a cognitive map (Tolman 1949) or model of the environment that allows for mental simulation of the outcome.

### **Distinguishing Model-Free and Model-Based Strategies**

Identifying neural mechanisms underlying model-based and model-free information relies on our ability to distinguish the use of each of these strategies behaviorally. Blocking, a behavior demonstrated by Kamin (Kamin 1969), offers this opportunity to separate behaviors that utilize model-based information from those that can rely on model-free information alone. Blocking builds on simple Pavlovian conditioning,

whereby a cue (for example a light) is reliably followed by a specific outcome (for example a pellet of food), and an animal observing this pairing learns to expect a pellet of food whenever the light is presented. In blocking, a second cue is added to the light (for example a tone), and the food pellet follows the compound cue. Blocking refers to the observation that animals do not learn to associate the tone with food, and the explanation is that because the tone offers no new information about the impending outcome, information about that cue is 'blocked' from learning.

According to TDRL, when an animal is first exposed to a neutral cue like a light, the animal will not expect a positive or negative outcome. Its expectation in terms of value is thus zero. However, when it experiences the food pellet, this better-than-expected outcome produces a positive RPE, which, physiologically, is a phasic burst of DA. This RPE updates the animal's expectation regarding the light. Thus, the light comes to elicit the expectation of food, and the experience of food ceases to cause an RPE, as receipt of food provides congruence of experience and reality.

When a second cue, like a tone, is added in a blocking paradigm, there is again no positive or negative expectation associated with the new cue. The expectation with respect to value is thus zero. Therefore, when the tone-light compound cue results in the one food pellet, predicted by the light, there is no prediction error to drive learning about the tone. Indeed, the outcome is consistent with what the animal expects as a result of both the light (one food pellet) and the tone (no expectation). To 'unblock' learning, one simply changes the value of the outcome so that the value of the tone-light compound cannot be fully predicted by knowledge of the outcome of the light itself. For example, after training an animal that the light predicts a food pellet, a tone-light compound cue that leads to two food pellets unblocks learning about the tone, via an RPE, and the animal learns that the tone predicts food.

In classical blocking experiments, changing the value of the outcome unblocks learning about the new cue. In other words, the currency of the outcomes is the same, but changing the amount of reward available to the animal drives learning. To be precise, we

call this classical form of unblocking ‘value unblocking’. Model-free algorithms like TDRL easily explain unblocking value. When the animal experiences the compound cue for the first time, he expects one food pellet based on his expectation about the light (one food pellet) and the tone (no expectation). Two food pellets is greater than expected, which results in a positive RPE that updates the animal’s understanding of the new cue.

Another form of unblocking has been observed that cannot be explained by model-free algorithms. We refer to this type of unblocking as ‘identity unblocking’. In identity unblocking, the value of the outcome is not changed as a result of the compound cue. Instead, only the identity of the outcome is changed. For example, if a light predicted a banana pellet, the light-tone compound cue may predict a grape pellet, assuming animals have no preference between banana and grape flavors. When this is done, animals also learn to associate the added cue with a rewarding outcome. Identity unblocking demonstrates animals’ ability to learn based on distinctions that are independent of value. Such learning cannot be driven by model-free information and instead must reflect model-based representations. Employing behaviors that depend on model-free versus model-based representations allows us to elucidate the circuitry and mechanisms underlying each representation type.

### **Physiological Substrates of Model-Free and Model-Based Information**

Distinct physiological circuits are thought to signal general, model-free and specific, model-based information (Cardinal, Parkinson et al. 2002, Daw, Niv et al. 2005). There are several neural candidates to consider as playing a role in model-free and model-based signaling, as motivated behaviors are executed as a result of the interaction of different brain areas. The basal ganglia are thought to be particularly essential for action selection, and their ventral portion, including the ventral striatum (VS) and ventral tegmental area (VTA) are involved in reward learning (Berridge 2001), an important component of choice behavior. The discovery that DA neurons in VTA of monkeys mimic activity predicted by TDRL led researchers to believe that areas connected to VTA, such as VS, are likely involved in model-free signaling like TDRL.

Adaptive decision making also relies on the prefrontal cortex. The lack of maturity of this portion of the cortex in adolescents is thought to underlie faulty executive functions associated with teenage behavior (Dahl 2001, Pfeifer and Allen 2012). A part of the prefrontal cortex, the orbitofrontal cortex (OFC) projects heavily to VS (Berendse, Galis-de Graaf et al. 1992) and, like VS, is intimately tied with VTA (Takahashi, Roesch et al. 2009). Together, the prefrontal cortex, VS, and VTA comprise a large portion of the putative reward circuit, which facilitates the development and evolution of choice behaviors. Given their roles in value-guided behaviors, both VS and OFC have been proposed to signal value information in a model-free manner. However, it has been hypothesized that distinct sub-circuits of the reward system separately encode information related to the value versus the identity of outcomes (Cardinal, Parkinson et al. 2002), bringing into question the precise relationship between the circuits from which model-free and model-based representations of outcomes arise.

### **Orbitofrontal Cortex**

OFC, which receives processed sensory information and is reciprocally connected to other reward structures of the brain, such as amygdala, hippocampus, VTA, and basal ganglia (Schoenbaum, Setlow et al. 2003), has been proposed as the seat of economic value in the brain. This suggested function for OFC is based primarily on evidence supporting its role in value-guided behaviors. The effect of orbitofrontal injury on decision making was famously described by Dr. Harlow, whose patient, Phineas Gage, suffered an accident in which an explosion drove a metal pole through the socket of his eye and out the back of his skull. Though Phineas survived, his OFC was severely damaged and his personality was markedly altered. Most noticeably, Gage behaved in a socially inappropriate manner, seemingly unable to consider the consequences of his actions (Damasio 1994, Damasio, Grabowski et al. 1994).

Since Phineas Gage, other patients with orbitofrontal damage have contributed to our understanding of OFC. Such damage consistently leads to deficits in reversal learning tasks, wherein the contingencies between cues and outcomes are switched. Perhaps the best known reversal learning test in humans is the Iowa gambling task, where people

choose decks of cards based on rewards associated with different decks. Patients with damage to OFC fail to modify their choices on this task as the value of the decks evolve (Fellows 2007). Leslie Fellows, a neurologist interested in orbitofrontal function, refers to the behavioral deficits in patients with orbitofrontal injury as demonstrating inconsistent preferences (Fellows and Farah 2003, Fellows and Farah 2005). Another neurologist, Antonio Damasio, whose clinical research on emotional processing includes investigation into the mechanisms underlying decision making deficits, describes the behavioral manifestation of orbitofrontal injury as “the [in]ability to select an advantageous response among an array of available options” (Damasio, Grabowski et al. 1994). The inconsistent preferences and failure to adaptively choose between options described by Fellows and Damasio are intuitive manifestations of an inability to consider potential consequences, and these observations build on the clarification of OFC function afforded by Phineas Gage.

What is particularly interesting about patients with orbitofrontal injury is that though their behavioral deficits prevent them from maximizing gains and minimizing losses, these patients’ sensitivity to the experience of rewards and punishments is intact. In other words, they demonstrate normal emotional responses to direct rewards and punishments. They appear, however, unable to compare hypothetical consequences. Thus, whereas these patients experience disappointment (Steiner and Redish 2012), they do not experience regret (Camille, Coricelli et al. 2004). Deficits in counterfactual thinking, or the ability to compare actual consequences with the consequences of unselected choices, likely underlies the lack of regret demonstrated in these patients (Wheeler and Fellows 2008, Camille, Griffiths et al. 2011, Steiner and Redish 2012). This failure to incorporate knowledge of non-experienced outcomes may also contribute to patients’ abnormal reward learning.

Although patients have offered great insights into the role of OFC in human behavior, the lack of control over the extent of orbitofrontal damage and individual differences in injuries and related factors represent obstacles in using patients to elucidate OFC function. As such, experiments using imaging or single-unit electrophysiological

techniques in animals have furthered our understanding of OFC function. Animals demonstrate the same deficits in value-guided behaviors that are the hallmark of orbitofrontal damage in humans. Namely, reversal learning deficits have been demonstrated in monkeys (Clarke, Robbins et al. 2008), rats (Burke, Takahashi et al. 2009), mice (Bissonette, Martins et al. 2008), cats (Teitelbaum 1964), and marmosets (Dias, Robbins et al. 1996) with orbitofrontal damage.

In addition to the behavioral evidence, *in vivo* electrophysiological studies in animals have also suggested OFC is important in value assessment. Specifically, the observation that individual neurons within OFC increase their firing rate more in anticipation of large rewards than in anticipation of small rewards has led researchers to conclude that these cells represent the value of expected outcomes (Roesch and Olson 2004, Padoa-Schioppa and Assad 2006). The suggestion that OFC signals expected outcome information as common currency value signals (Rolls and Grabenhorst 2008) is further supported by the observation that there is greater OFC activation in anticipation of rewards with short delays than in anticipation of those with long delays (Roesch and Schoenbaum 2006), as well as by the finding that OFC encodes both the mean, as well as the variance in the value of outcomes (Mainen and Kepecs 2009). fMRI studies have also produced results consistent with value signals in OFC. BOLD responses in OFC correlate with expectations about reward and punishment (Kahnt, Heinzle et al. 2010). For example, several fMRI studies have found higher activation in OFC when subjects are presented with affectively pleasant sensory stimuli compared to neutral stimuli. (Plassmann, O'Doherty et al. 2007, FitzGerald, Seymour et al. 2009, Elliott, Agnew et al. 2010, Plassman, O'Doherty et al. 2010). OFC appears too to be engaged during economic transactions, and activity in OFC correlates with willingness to pay (Plassmann, O'Doherty et al. 2007).

Despite the evidence supporting the economic view of OFC function, a comprehensive look at studies on OFC function suggests a role for OFC that is perhaps more complicated than encoding common currency value information. Indeed, we would expect damage to an area whose key function is signaling value to affect simple choice

behavior. However, like humans, animals with orbitofrontal damage respond normally to rewards and punishments. Specifically, such damage in humans and animals does not affect choice between big and small rewards (Walton, Behrens et al. 2010), nor does it affect the ability to learn the association between a cue and reward or punishment (Burke, Franz et al. 2008).

OFC signals information about outcomes that may seem related to value but is in fact independent of value. For instance, numerous studies have demonstrated that OFC encodes information about the specific features of outcomes, including identity (Schoenbaum, Roesch et al. 2009), spatial orientation (Feierstein, Quirk et al. 2006, Roesch, Calu et al. 2007), sensory properties (Delamater 2007), salience (Ogawa, van der Meer et al. 2013), and risk (O'Neill and Schultz 2010). Additionally, activity conveying information independent of actual outcomes but relevant to their relative value are also signaled in OFC. Such variables include motivational factors like current hunger or satiety (Mainen and Kepecs 2009), incentive value (Gallagher, McMahan et al. 1999), and representation of reward history (Riceberg and Shapiro in press). Given these functions, it is perhaps not surprising that OFC is needed for reinforcer devaluation in people and animals (Gallagher, McMahan et al. 1999, Pickens, Setlow et al. 2003, Izquierdo and Murray 2004, Valentin, Dickinson et al. 2007, West, DesJardin et al. 2011), as well as for other value-guided behaviors like outcome-specific Pavlovian-to-instrumental transfer (Ostlund and Balleine 2007), outcome-specific conditioned reinforcement (Burke, Franz et al. 2008), and over-expectation (Takahashi, Roesch et al. 2009).

To test the hypothesis that OFC facilitates choice behavior by encoding specific features of outcomes that are independent of value, McDannald, Lucantonio et al. 2011 compared performance on the unblocking tasks described above in rats with and without OFC lesions. OFC was required for identity unblocking but not value unblocking. Thus, OFC seemed essential for recognizing specific features of outcomes and modifying behavior based on violations in expectations of such features. On the other hand, value unblocking, which can be accomplished merely by recognizing changes in value, did not require

OFC. These findings support the theory that OFC contributes to behavior and learning not via model-free value assessment, but by representing value-independent model-based information that defines the task and predicts features of outcomes.

### **Ventral Striatum**

The striatum is the input structure of the basal ganglia, a collection of nuclei that support action selection. Information from the prefrontal cortex projects to the striatum in a topographically organized manner, converging with sensory and motor information (Groenewegen, Berendse et al. 1990, Berendse, Galis-de Graaf et al. 1992). VS is the only part of the striatum to receive information from hypothalamus and amygdala, placing VS in a unique position to incorporate motivational and emotional information into action selection (Groenewegen, Berendse et al. 1990, Haber, Lynd-Balta et al. 1990, Brog, Salyapongse et al. 1993, Heimer, Zahm et al. 1995, Roesch, Singh et al. 2009). It is therefore often considered a motor-limbic interface, providing a conduit from motivation to action (Mogenson, Jones et al. 1980). As action selection is presumably the result of value assessment, VS has been proposed to integrate different factors that influence value so as to compute the overall value of potential options in a common currency. Such integration would allow VS to signal the value of expected outcomes in a model-free way.

The strong reciprocal connection between VS and VTA make VS a likely candidate for both influencing and being influenced by RPEs signaled by midbrain DA neurons. VS does indeed have access to PE signals (van der Meer and Redish 2009). Further, VS fails to increase its firing rate in anticipation of reward when DA input is eliminated. In other words, the reward-anticipatory firing that develops in VS is dependent on midbrain DA activity. It is therefore likely that VS learns the meaning of cues based on RPE input from VTA (Takahashi, Roesch et al. 2009). Additionally, it seems that RPEs directly influence VS, as fast scan cyclic voltammetry studies have demonstrated transient alterations in the concentration of DA in VS, consistent with RPE signaling (Day, Roitman et al. 2007).

Consistent with a role for VS in signaling value information, VS is involved in reward anticipation (Takahashi, Schoenbaum et al. 2008) and is critical in several value-guided behaviors, including second order conditioning (Setlow, Holland et al. 2002), conditioned place preference (Everitt, Morris et al. 1991), general affective conditioned responding (Ito, Robbins et al. 2004), and general affective Pavlovian-to-instrumental transfer (Corbit and Balleine 2011). Imaging studies assessing the role of VS in these and other value-guided behaviors provide evidence for value signals in VS (O'Doherty, Dayan et al. 2004, Hare, O'Doherty et al. 2008).

There are, however, several studies whose outcomes support a role for VS in model-based representations of outcomes. Lesions of VS in rats have abolished the reinforced devaluation effect, suggesting that VS is involved in goal-directed behavior (Singh, McDannald et al. 2010). Such lesions also prevent outcome-specific Pavlovian-to-instrumental transfer (Corbit and Balleine 2011). Thus, in addition to processing general value information, VS appears too to be critical for using information about specific features of outcomes to guide behavior. As such, when assessing the role of VS in value and identity unblocking, McDannald, Luncantonio et al. 2011 found that VS was required for performance in both tasks.

Evidence for model-based representations of outcomes in VS extends beyond behavioral studies to imaging and recording studies as well. Integration of model-free and model-based representations of outcomes have been observed in fMRI BOLD signals (Daw, Gershman et al. 2011). Further, a recent study found that activity of these signals was better characterized by a hybrid of model-free and model-based activity than by either type of activity alone and that the integration of these signals correlated with choice performance (Glascher, Daw et al. 2010). One study even found that the model-based theory alone explained better value-related BOLD signals in striatum than did model-free theory (Simon and Daw 2011). fMRI studies have also demonstrated VS sensitivity to outcome specific devaluation, a phenomenon that cannot be accounted for by traditional TDRL. Single-unit recording studies have demonstrated VS signaling of expectation of specific features of outcomes (van der Meer and Redish 2009, Goldstein, Barnett et al.

2012) and provide support for VS integration of information at decision points to facilitate behavior (van der Meer and Redish 2009). Such findings further support a role for VS in signaling model-based information.

### **Hypothesis**

In the following chapters, we will describe experiments we undertook to test the hypothesis that OFC signals model-based information about outcomes, which is then incorporated in VS signaling of outcomes. OFC influences VS activity in primates (Simmons, Ravel et al. 2007). Further, there is a remarkable parallel in the circuitry connecting OFC and VS across species (Goldman-Rakic, Lidow et al. 1992), suggesting a similarity in the functional relationship between these structures. Thus, the results of our single-unit recording studies in rats presumably elucidate the mechanisms underlying decision making in higher order animals and humans.

## **CHAPTER 2: MODEL-BASED SIGNALING OF OUTCOMES IN ORBITOFRONTAL CORTEX**

### **INTRODUCTION**

The orbitofrontal cortex (OFC), known to encode information about expected outcomes (Schoenbaum, Chiba et al. 1998, Tremblay and Schultz 1999, O'Doherty, Kringelback et al. 2001), is an area of focus for work aimed at understanding choice behavior. One popular notion is that OFC is the seat of economic value in the brain and thus signals value information in a common currency. Support for the economic interpretation of OFC function comes from observations of increased activation in OFC in response to cues that predict reward. This pattern of activity in OFC is observed in BOLD signal in fMRI studies (Thorpe, Rolls et al. 1983, O'Doherty, Rolls et al. 2000, Gottfried, O'Doherty et al. 2003). In addition, single-unit studies identify neurons in OFC that increase their firing rate more in anticipation of larger, more valuable rewards than in anticipation of less valuable outcomes (Roesch and Olson 2005, Padoa-Schioppa and Assad 2006).

The notion that OFC produces pure value signals is simplest to conceptualize if representations of outcomes in OFC are model-free. Model-free representations are fast and computationally inexpensive, but they do not take into account specific features of outcomes. Further, they lack flexibility; therefore optimal behavior requires that model-free representations of outcomes be supplemented by model-based representations. Model-based representations provide specific information about the features of outcomes, as well as the context in which those outcomes are experienced, thus facilitating the adaptation of behavior to the relative value of such outcomes.

Interestingly, OFC is not generally needed for behaviors that could be supported by value information, independent of specific information about outcome features or current value (McDannald, Lucantonio et al. 2011), nor is OFC needed for simple acquisition (Gallagher, McMahan et al. 1999, Izquierdo and Murray 2004, Machado and Bachevalier 2007), simple extinction (Takahashi, Roesch et al. 2009), or even for choosing between small and large rewards (Fellows 2011). On the other hand, OFC plays a critical role in

behavior that cannot rely on model-free value information about expected outcomes but instead requires the use of specific information and thus a model-based strategy (Pears, Parkinson et al. 2003, McDannald, Saddoris et al. 2005, Parkinson, Roberts et al. 2005, Balleine, Daw et al. 2008, Burke, Franz et al. 2008, McDannald, Lucantonio et al. 2011).

The idea that OFC produces a general or common currency value signal is also problematic when one considers the variety of outcome variables to which OFC neurons respond, including outcome sensory properties (Delamater 2007), identity (Schoenbaum, Roesch et al. 2009), salience (Ogawa, van der Meer et al. 2013), location (Feierstein, Quirk et al. 2006, Roesch, Calu et al. 2007), probability (O'Neill and Schultz 2010), and time of delivery (Roesch, Taylor et al. 2006). Thus the firing of many OFC neurons appears to be heterogeneous rather than integrated into a common currency signal and may therefore underlie richer model-based representations of outcomes. Additionally, cells within OFC that fire preferentially in anticipation of more valuable rewards are among another population of cells that fire preferentially in anticipation of less valuable outcomes. A reinterpretation of the role of OFC in value-guided behaviors may therefore render a different result; namely that OFC is critical for value-guided behaviors insofar as OFC signals specific information about expected outcomes.

Model-free and model-based accounts of OFC activity lead to distinct predictions for how OFC encodes information about expected outcomes. To determine if OFC activity is more consistent with model-free or model-based representations of outcomes, we assessed three such pairs of predictions by coupling single-unit recordings with a behavioral task that allowed for independent manipulation of value (size) and a specific feature of outcomes (flavor). Specifically, a milk reward offered on each trial varied in size (big or small) and in flavor (chocolate or vanilla). We have shown previously that rats show no preference for chocolate versus vanilla rewards, which we have interpreted to mean that rats value chocolate and vanilla flavors similarly (McDannald, Lucantonio et al. 2011). Thus, changing the size of an outcome in this task affected the value of that outcome, but changing the flavor did not.

The first pair of predictions we addressed was model-free versus model-based predictions for encoding of outcome flavor. If OFC signals are model-based, OFC should encode flavor information, as flavor represents specific information about potential outcomes. Specifically, OFC should encode both chocolate and vanilla information in such a way that each flavor is distinguishable within OFC. However, because each flavor provides the same amount and kind of information about the impending outcomes, and because each flavor is experienced a similar number of times, model-based accounts predict no bias in representation of one flavor over the other. Similarly, if OFC signals are model-free, there should be no bias in representation of chocolate versus vanilla rewards because these flavors are indistinguishable in terms of value. However, for precisely this reason, unlike the model-based view of OFC, a model-free view predicts little or no encoding of flavor information, as flavor alone is irrelevant in the computation of value signals. Thus, if OFC signals are model-free, chocolate-selective and vanilla-selective populations should not be observed.

The following predictions regarding encoding of outcome value were the basis of our next analysis. If OFC signals are model-based, OFC should represent big and small rewards as specific size features of outcomes, much as it should represent chocolate and vanilla rewards. Thus, big and small rewards should be distinguishable within OFC, but because model-based representations are independent of value, there should be no bias in representation of one reward size over the other. On the other hand, if OFC signals are model-free, and therefore driven by value, signals anticipating big versus small reward should differ significantly so as to be distinguishable downstream of OFC. Simply put, a model-based interpretation of OFC activity predicts significant but similar representation of big and small rewards in OFC, as big and small represent specific size information, separable from value. A model-free interpretation, however, predicts differential representation or signaling of big and small rewards, consistent with a value signaling function.

A third distinction in model-based versus model-free predictions for activity in OFC is representation of the task itself. The general predictions are: if OFC activity is model-

based, it should represent task structure, and if OFC activity is model-free, it should not. Though specific predictions for this activity are less clear, we looked for two features of activity that, if present, would support the model-based view of OFC activity. First, we analyzed cell activity to determine if any cells fired preferentially for a specific block of our task, rather than for a specific outcome. Such activity would reflect a model-based representation of outcomes, as it would represent the current configuration of available rewards. Second, we tested for anticipatory activity in OFC that could not be explained by experience alone but instead required inferential thinking. Block switches in our task are accompanied by a change in reward at the left and right wells. If after experiencing the new reward at one well, OFC activity in anticipation of reward at the other well reflects an accurate expectation of the new reward, OFC activity cannot be based solely on experience. Thus, signaling of such inferential information in OFC would provide strong support for a model-based representation of outcomes in OFC, and further, a correlation between inference signals and subsequent behavior would support the functional role of model-based information in OFC.

Our analysis of neural activity was consistent with each of our predictions for a model-based representation of outcomes in OFC. Specifically, this data set demonstrated that activity in OFC encodes specific features of outcomes that are independent of value, as well as task structure variables that are either independent of or uninfluenced by experience.

## **MATERIALS & METHODS**

### **Subjects**

6 male Long-Evans rats, weighing 275-300 g were acquired from Charles River Laboratories. They were housed individually in a 12 hour light/dark schedule environment and given *ad libitum* access to food prior to testing. Testing occurred during the rats' light cycle, and during this time, rats were food deprived to 85% of their baseline weight. All testing was performed in accordance with the guidelines set forth by the University of Maryland School of Medicine Animal Care and Use Committee and the National Institutes of Health.

## **Surgical Procedures**

Aseptic, stereotaxic surgical techniques were used to implant a drivable bundle of wires in each hemisphere of the OFC in each rat. 10 FeNiCr wires (Stablohm 675, California Fine Wire, Grover Beach, CA), each with a diameter of 25  $\mu\text{m}$ , were cut with surgical scissors to spread  $\sim 1\text{mm}$  beyond the cannula and the electroplated with platinum ( $\text{H}_2\text{PtCl}_6$ , Aldrich, Milwaukee, WI) to an impedance of  $\sim 300$ . The drivable bundle of wires were then immediately chronically implanted dorsal to OFC 3.0 mm anterior to bregma, 3.2 mm laterally, and 4.0 mm ventral to the surface of the brain in each hemisphere of each rat. Following surgery, cephalexin (15 mg/kg p.o.) was administered twice a day for two weeks to prevent infection. At the conclusion of the experiment, a 15  $\mu\text{A}$  current was passed through each electrode to mark the final position of the electrode. Subsequently, rats were perfused, and their brains were subjected to standard histological techniques (Schoenbaum, Chiba et al. 1999)

## **Apparatus**

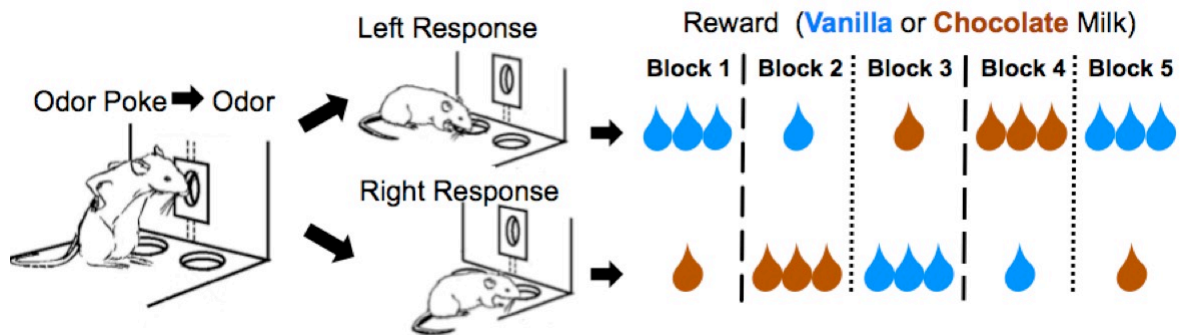
Behavioral training and recording were conducted in 18" x 18" aluminum chambers with sloping walls that narrowed to an area of 12" by 12" at the bottom. On the right wall of each chamber, there were two fluid wells separated by an odor port. Olfactory cues were delivered to the odor port through an air flow dilution olfactometer. These cues were provided by International Flavors and Fragrances (New York, NY). Photobeam interference indicated rat entry into the odor port or fluid wells. All parameters of the task were controlled via computer.

## **EXPERIMENTAL PROCEDURES**

### **Choice Task**

Our choice task is illustrated in Figure 1. Trials in our choice task began with the illumination of a light. Once a rat's nose was in the opening of the odor port for 0.5 seconds, an odor was delivered to a small hemicylinder located behind the odor port opening. The odor was presented for 0.5 seconds. Following odor presentation, there was a 0.5 seconds delay, after which a reward was delivered to the appropriate well, provided

the rat entered the well within 3 seconds of the odor offset.



**Figure 1. Experimental design.** Choice task in which we independently manipulated value and identity of reward outcomes. Across each block shift, either size or flavor of the outcomes differs from outcomes in the previous block. See Experimental Procedure for a full description of the choice task.

Rats learned to go to the well to the right of the odor port or to the well to the left of the odor port, depending on which one of two forced choice odors were presented in the odor port. Alternatively, if a third free choice odor was presented, rats could choose either well, and reward was delivered only at the first well entered. These three odors were constant throughout the experiment. Odors were presented in pseudorandom order such that 7/20 trials were ‘free choice,’ and an equal number of each forced choice odor was presented across the session. No odor was presented on 3 consecutive trials. Reward in this task was chocolate or vanilla flavored milk, diluted to 50% concentration with deionized water. These flavors were chosen because we have previously shown that rats can distinguish chocolate and vanilla flavored milk but do not have a preference for one flavor over the other (McDannald et al., 2012). A ‘small’ reward was a 0.05 mL bolus, and a ‘big’ reward was two such boli delivered 0.5 seconds apart.

After rats were trained to perform this simple task, we introduced blocks where the size and flavor of the reward at each well were independently manipulated. Once rats maintained accurate responding in this more complicated version of the task, we commenced recording sessions.

### Single-Unit Recording

Single-unit recording procedures were the same as previously described (Roesch et al., 2006). Sessions began after active wires were selected, and the electrode was advanced

40 or 80  $\mu\text{m}$  at the end of each session. If activity was not detected at the start of a session, the rat was removed from the chamber, and the electrode was advanced.

Neural activity was recorded using two identical Plexon Multichannel Acquisition Processor systems (Dallas, TX), interfaced with odor discrimination training chambers described above. Signals from the electrode wires were amplified 20x by an op-amp headstage (Plexon Inc HST/8050-G20-GR), located on the electrode array. Outside the training box, the signals were passed through a differential preamplifier (PBX2/16sp-r-G50/16fp-G50; Plexon), in which the single-unit signals were amplified 50x and filtered at 150–9000 Hz. The single-unit signals were then sent to the Multichannel Acquisition Processor box, in which they were further filtered at 250–8000 Hz, digitized at 40 kHz, and amplified at 1–32x. Waveforms ( $>2.5:1$  signal-to-noise) were extracted from active channels and recorded to disk by an associated workstation with event timestamps from the behavior computer. Waveforms were not inverted before data analysis.

Offline Sorter software from Plexon Inc (Dallas, TX) was used to sort units with a template matching algorithm. Files were then converted to Neuroexplorer files, where event markers and unit timestamps were extracted. These data were analyzed in Matlab (Natick, MA).

## **Statistical Analysis**

### *Behavioral Analysis*

For our behavioral analysis, we measured overall performance on the task by calculating the percent of trials where rats correctly chose the well associated with the forced choice odor that was presented, and we measured preference by calculating the percent of free choice trials where the rat chose each size and flavor outcome. To determine if there were significant differences in preference for big versus small rewards or chocolate versus vanilla rewards, we performed t-tests on choice rate for big and small rewards and also for chocolate and vanilla rewards. In order to test whether forced choice trials also indicated size and flavor preference, we performed two separate 2-way within-subjects ANOVA with size and flavor as factors. The first ANOVA compared the percent of

correctly performed forced choice trials, and the second compared reaction time on each forced choice trial. Finally, to determine what factors influenced the change in behavior that occurred across block transitions, we performed a mixed design ANOVA on the difference scores of choice rate in the first 25 trials of a new block compared to the last 25 trials of the previous block, with transition type (size or flavor) and initial flavor (chocolate or vanilla) as factors.

### *Neural Analysis*

We found that neurons displayed two kinds of reward-related activity. First, many neurons showed reward-anticipatory activity, reflecting knowledge of the impending reward on each trial. Second, neurons also showed activity that varied from block to block. Because blocks differed only in the rewards that were available at each reward well, we hypothesized that this activity reflected the available rewards. We performed separate analyses to examine the characteristics of each of the two kinds of reward-related activity, as described below.

To test for reward-anticipatory activity, we analyzed correct forced choice trials because trials with each reward feature were equally represented within a session for forced choice trials, whereas there was a bias toward the larger reward for free choice trials. We first classified the selectivity of neurons according to a 3-way ANOVA on firing rate, with reward size (big or small), flavor (chocolate or vanilla), and direction (left well or right well) as factors. Baseline firing rates were subtracted on each trial because we found that baseline rates carried information about available rewards (see Results). To allow averaging of neurons with different firing rates, we also peak-normalized firing rates by dividing all firing rates for each neuron by firing rate in the 100 ms bin with the highest mean firing rate across the first ten or last ten trials of each condition. Neurons were classified as reward-anticipatory when they showed a main effect of reward size or reward flavor, or an interaction of these factors with each other or of either of these factors with direction. We found that interaction effects in this ANOVA could reflect available-reward, or block-specific activity due to the manner in which available rewards were paired in each block. Thus, so as not to conflate block-specific activity with reward-

anticipatory activity, we included an additional stipulation that neurons with a significant interaction had to show a main effect of direction for at least one block to be classified as reward-anticipatory.

Once neurons were categorized as reward-anticipatory, we performed chi square tests to determine if differences existed in the size of the populations of big versus small-selective cells or chocolate versus vanilla-selective cells. Further, to identify any bias in the overall selectivity of the population of OFC cells, we calculated size and flavor indices for each cell and performed t-tests on these indices. We also ran a regression on the cells' size and flavor selectivity to determine if there was any correlation in the encoding of size and flavor.

To identify cells with block-specific, rather than reward-anticipatory signaling, we ran a different ANOVA on each neuron's firing rate across trials. This ANOVA had factors of available sizes, available flavors, and direction. Specifically, the available size factor had two levels: 1) big on left – small on right, or 2) small on left – big on right. Each block could be described by one of these sets of available sizes. Available flavors were similarly assigned to each block, except that the levels were 1) chocolate on left – vanilla on right, or 2) vanilla on left – chocolate on right. Neurons showing an effect of available sizes, available flavors, or an interaction between the two, without that effect interacting with direction, were classified as block-specific cells. We added the latter stipulation because available rewards did not differ between directions within a block, and we sought to isolate the population that signaled available rewards without signaling anticipated reward.

To discover the timing of reward-anticipatory activity and block-specific activity, we performed a sliding ANOVA analysis on 500 ms epochs, with the first epoch ending at the beginning of the trial, and each subsequent epoch sliding by 50 ms forward. Sliding epochs were aligned to various events in the trial. At each epoch, cells were classified according to each of the two ANOVA analyses, and the proportion showing significance for each analysis was calculated. As shown in Figure 5, each kind of reward encoding

incorporated both reward size and reward flavor information, but the two kinds showed different profiles across the time of the trial. Encoding of available rewards was present even before the trial began and peaked early in the trial. Encoding of anticipated reward for a particular trial peaked at the end of the trial, just before reward was delivered. For subsequent analyses of reward-anticipatory activity, we focused on the 500 ms epoch immediately before reward delivery, when anticipatory activity was represented in the highest proportion of neurons.

We tested whether neurons signaled inference information by examining activity on the first forced-choice trial in whichever direction occurred *second* at the start of each block. This activity was compared to the last two trials in the same direction in the previous block. ANOVAs were performed on firing rates across neurons, with preference direction (left of right), time (before or after the block switch), group (control or lesion), and reward feature (size or flavor) as factors. We included only neurons with a main effect of size or flavor, because predictions about their activity were most straightforward. Significant interactions between preference direction and time for the cells selective for the feature that had been altered (size or flavor) would indicate inference signaling. As a control, we did the same comparison across block switches for the feature that was not altered and thus provided no opportunity for inferring new outcomes. For this condition, activity related to the unchanged feature should be constant across the block switch.

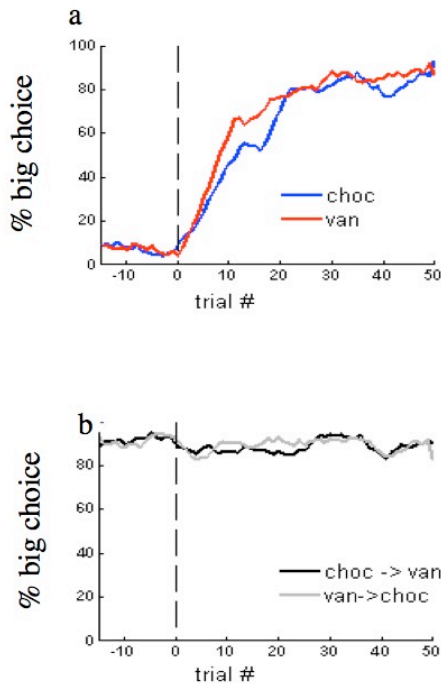
We were also interested in determining if there was a relationship between block-selectivity or inference signals and subsequent behavior. We first calculated performance as the percent of free choice trials for which the rat chose the well with the big reward. We then plotted the difference in activity of block-selective cells during the first 10 trials of a new block minus the activity in the previous block against rats' performance on free choice trials within that block. Next, we calculated an inference score for each cell on each trial, which was a measure of the accuracy of the cell's anticipatory information (i.e. a high score indicated that the cell was signaling the expectation of the correctly inferred new outcome, and a low score indicated that the cell was signaling the expectation of the

reward that had been delivered in the previous block). A plot of these scores, against rats' performance on free choice trials within that same block., was then constructed.

## **RESULTS**

### **Behavior Reveals Size Preference but not Flavor Preference**

Because rats, like all animals, are motivated by the value of expected outcomes, we expected rats to prefer big rewards to small rewards and to demonstrate this through their choice behavior once the task was learned. We also expected, as previously shown, that rats would demonstrate no preference between chocolate and vanilla rewards. We tested these predictions in four ways. First, we analyzed choice rate for each reward type in the free choice trials and expected rats to choose big rewards more often than small rewards but to show no significant difference in choice for chocolate versus vanilla rewards. Second, we compared the accuracy for the different rewards on the forced choice trials and expected that when presented with the forced choice odor directing them to the well with the big reward, rats would be more likely to accurately choose the correct well than if directed to the well with the small reward. Third, we compared reaction times for obtaining reward on forced choice trials. Similar to our second prediction, we expected rats to be faster to respond when directed to a big reward than when directed to a small one. Finally, because each block begins with a change in odor-outcome contingencies, and because we expected rats to be influenced by reward size but not reward flavor, we expected to see behavior change across block transitions according to the location of the big reward but without regard to the location of each flavor.



**Figure 2.** Influence of reward size and flavor on free choice behavior.

- a) Before a size shift, indicated by the dotted vertical line, rats rarely chose the well that delivered the small reward on free choice trials. Once the size changed, the rats' responses changed accordingly, and they consistently chose the well that delivered the big reward, regardless of the flavor of reward. Overall choice for the big reward was significantly greater than for that of the small reward ( $t_{93} = 54.9$ ,  $p < 0.0001$ ).
- b) Before a flavor shift, indicated by the dotted vertical line, rats consistently chose the well with big reward on free choice trials. When the flavors changed, the rats continued to choose the well with the big reward, regardless of the direction of the flavor switch (chocolate to vanilla or vanilla to chocolate). Overall choice for chocolate versus vanilla rewards did not differ significantly ( $t_{93} = 1.6$ ,  $p = 0.11$ ).

Rats performed the task accurately, choosing the correct well on  $87.6 \pm 0.90\%$  of forced choice trials across all recording sessions. As expected, rats preferred big rewards to small

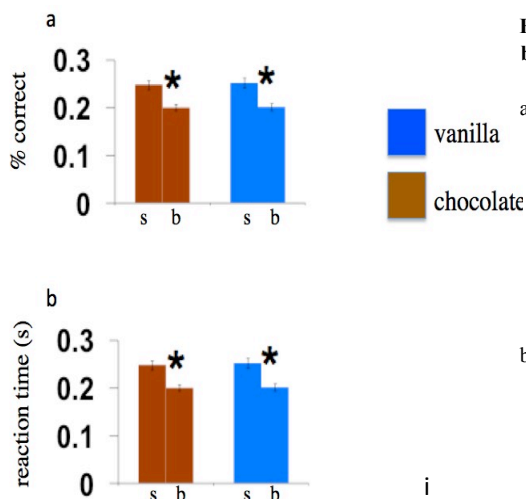
reward (Figure 2a), choosing the well delivering the big reward more often ( $86.7 \pm 0.67\%$ ) than the well delivering the small reward ( $13.3 \pm 0.67\%$ ) on free choice trials. Additionally, rats demonstrated no preference between chocolate and vanilla flavors, choosing chocolate rewards ( $47.9 \pm 1.3\%$ ) and vanilla rewards ( $52.1 \pm 1.3\%$ ) at a similar rate (Figure 2b). T-tests across sessions revealed that choice for big and small reward differed significantly from a 50% choice rate for each outcome ( $t_{93} = 54.9$ ,  $p < 0.0001$ ), whereas choice for chocolate and vanilla rewards did not differ significantly from a 50% choice rate ( $t_{93} = 1.6$ ,  $p = 0.11$ ). Thus, while rats displayed differential preference for big versus small rewards, they preferred chocolate and vanilla rewards equally.

Also as expected, rats demonstrated their preference for big over small rewards in forced choice trials, for which we examined both accuracy and reaction time during the last 25 trials of each block, when odor-outcome expectations were well established. In these cases, given the forced choice odor directing them to the well with the big reward, rats performed with higher accuracy (Figure 3a) and greater speed (Figure 3b) than when given the forced choice odor directing them to the well with the small reward. Such behavior is likely a manifestation of rats' preference of big over small rewards. Two

separate 2 within-subjects ANOVAs with factors size (big or small) and flavor (chocolate or vanilla) were performed on these data.

The first ANOVA, used to assess accuracy on forced choice trials, was performed using percent correct scores (Figure 3a). This analysis revealed a main effect of size ( $F_{1,92} = 182.3, p < 0.001$ ) but also an effect of flavor ( $F_{1,92} = 5.3, p < 0.05$ ) and an interaction between the two ( $F_{1,92} = 5.1, p < 0.05$ ). Post-hoc tests showed that the effect of flavor and the interaction were driven by a slightly higher accuracy on trials with the small vanilla reward ( $85.0 \pm 1.3\%$  correct) compared to that on trials with small chocolate reward ( $81.4 \pm 1.6\%$  correct;  $p < 0.01$ ), with no difference between percent correct on trials with the big vanilla reward (96.8% correct) and the big chocolate reward (96.8% correct;  $p = 0.97$ ).

The second ANOVA, used to assess speed on forced choice trials, was performed using average reaction times, defined as the time from odor cessation to odor port exit, are illustrated in Fig 3B. Here we found a main effect of size ( $F_{1,92} = 62.2, p < 0.001$ ) but not flavor ( $F_{1,92} = 0.3, p = 0.57$ ), and no interaction ( $F_{1,92} = 0.1, p = 0.73$ ). These results indicate rats were faster to respond when big chocolate or big vanilla rewards were available ( $201 \pm 7.2$  ms, and



**Figure 3. Influence of reward size and flavor on forced choice behavior.**

a) On forced choice trials in the last 25 trials of blocks, rats' accuracy was higher when the reward was big versus small. There was an effect of size, ( $F_{1,92}=182.3, p < 0.001$ ), an effect of flavor ( $F_{1,92}=5.3, p < 0.05$ ), and an interaction ( $F_{1,92}=5.1, p < 0.05$ ). Post-hoc tests revealed the flavor effect was driven by slightly higher accuracy for the small vanilla reward ( $85.0 \pm 1.3\%$  correct) than for the small chocolate reward ( $81.4 \pm 1.6\%$  correct;  $p < 0.01$ ) but no difference in accuracy for the big rewards (96.8% correct each;  $p = 0.97$ ).

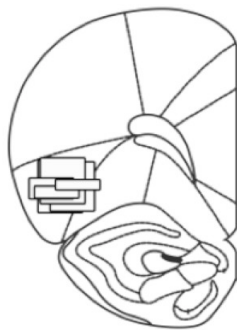
b) On forced choice trials in the last 25 trials of blocks, rats' reaction time was faster when the reward was big versus small, regardless of its flavor. There was a main effect of size ( $F_{1,92}=62.2, p < 0.001$ ) but not flavor ( $F_{1,92}=0.3, p = 0.57$ ), and no interaction ( $F_{1,92}=0.1, p = 0.73$ ).

$201 \pm 8.1$  ms, respectively) and slower when small chocolate or small vanilla rewards were available ( $248 \pm 9.2$  ms, and  $252 \pm 9.8$  ms, respectively). Thus, across these various measures of the relative value attributed to rewards in this task (choice rate,

accuracy, and reaction time), we found that the size of the reward always had a large effect, whereas flavor had a small, usually non-significant effect.

Finally, as expected, we found that changes in choice behavior across block transitions, measured as the difference between choice rate in the first 25 trials of a new block and the last 25 trials of the previous block, were influenced by changes in reward size at each well but not by changes in reward flavor. The results of a mixed design ANOVA performed on these difference scores, with transition type (size or flavor) and initial flavor (chocolate or vanilla) as factors are illustrated in Figure 2C. This analysis revealed a main effect of transition type ( $F_{1,92} = 195.7, p < 0.001$ ), driven by significant different scores across size transitions (planned contrast:  $F_{1,92} = 445.9, p < 0.0001$ ), with insignificant difference scores observed across flavor transitions (planned contrast:  $F_{1,92} = 1.3, p = 0.27$ ). There was no effect of initial flavor ( $F_{1,92} = 0.0, p = 0.93$ ). Planned contrasts also revealed no differences between chocolate and vanilla for size transitions ( $F_{1,92} = 1.6, p = 0.21$ ), and no difference between chocolate to vanilla and vanilla to chocolate for flavor transitions ( $F_{1,92} = 2.3, p = 0.13$ ). These results indicate that choice behavior was sensitive to size changes without regard to flavor.

### OFC Encoding of Outcome Flavor is Consistent with Model-Based Signaling



**Figure 4. Recording sites in rat OFC.** The black boxes indicate the location of electrodes throughout the experiment. The width represents the width of the electrode (~1 mm), and the height represents the approximate extent of recording across all sessions.

Bregma 3.72 mm + 0.28 to +0.96

entered the reward well, and stayed in the reward well for at least 500 ms. Because OFC encodes information about expected outcomes, we expected OFC to signal reward-relevant information once outcomes were predictable (after odor presentation) but before reward was delivered. We therefore confined our neural analysis to the 500 ms period during which the rat had entered the reward well but had not yet received reward. We recorded 831 neurons in OFC over the course of 94 sessions in 6 rats (Figure 4).

Though both model-based and model-free accounts of OFC activity predict that OFC encodes information about expected outcomes, these accounts offer distinct predictions in the nature of such encoding. First, with regard to flavor information, a model-based view predicts both chocolate and vanilla rewards to be significantly encoded by OFC because each provides information about the specific outcome. Specifically, if OFC signals model-based information, we would expect to observe a population of cells within OFC that fired preferentially in anticipation of chocolate rewards, as well as a population that fired preferentially in anticipation of vanilla rewards. Further, if OFC represents information about outcomes in a model-based way, it should represent chocolate and vanilla rewards similarly. Thus, populations of chocolate-selective and vanilla-selective cells should be approximately the same size, and the overall population should demonstrate no bias in selectivity toward chocolate or vanilla outcomes.

If, on the other hand, OFC representation of outcomes is model-free, OFC would not represent chocolate or vanilla outcomes independent of their value. We would therefore not expect to find chocolate-selective and vanilla-selective populations in OFC if OFC signals model-free information. Further, if individual neurons in OFC do not encode flavor independent of value, it follows that there should be no statistical difference in the number of chocolate-selective and vanilla-selective cells in OFC and no population bias toward selectivity of either flavor. However, unlike what we would expect to see if OFC were model-based, we would not expect to find distinct significant populations of chocolate-selective and vanilla-selective cells in OFC.

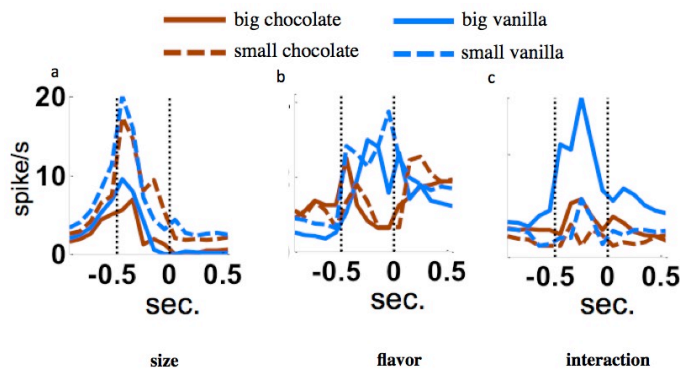
Neural encoding of outcome flavor was consistent with model-based representations of outcomes in OFC. Both chocolate-selective and vanilla-selective populations were identified (Figure 5), the populations did not differ significantly in size, and there was no bias in selectivity toward chocolate or vanilla rewards when the entire population was considered (Figure 6). Our classification of selectivity of OFC neurons allowed us to categorize flavor-selective cells as either chocolate-selective or vanilla-selective. According to the ANOVA performed on each cell, 135 of 831 (16%) neurons were flavor-selective, with 70 cells firing more in anticipation of chocolate and 65 cells firing

more in anticipation of vanilla ( $p=0.67$  by  $\chi^2$ ). To determine whether the overall selectivity of the population of OFC neurons was biased toward chocolate or vanilla rewards, we calculated a flavor index for each neuron and plotted the distribution for the population (Figure 6). The flavor index was calculated as the difference in peak normalized firing in anticipation of chocolate and vanilla reward.

$$\text{Flavor Index} = (\text{Chocolate-Vanilla})$$

Positive size indices corresponded to chocolate-selective cells, and negative size indices corresponded to vanilla-selective cells. The overall average of these indices was  $0.002 \pm 0.003$ , which was not significantly different than zero ( $t_{830} = 0.57$ ,  $p = 0.57$ ).

### OFC Encoding of Outcome Size is Consistent with Model-Based Signaling



**Figure 5. Outcome selectivity in OFC.** Distinguishable populations of cells within OFC fire in anticipation of big, small, chocolate, and vanilla rewards, with some cells displaying a size x flavor interaction. Firing in outcome-selective neurons begins increasing during the odor epoch, when rats are presented with information that allows for predictions about the upcoming reward, and grows as the rat moves to the well (movement epoch) and waits for the impending reward (expectancy epoch).

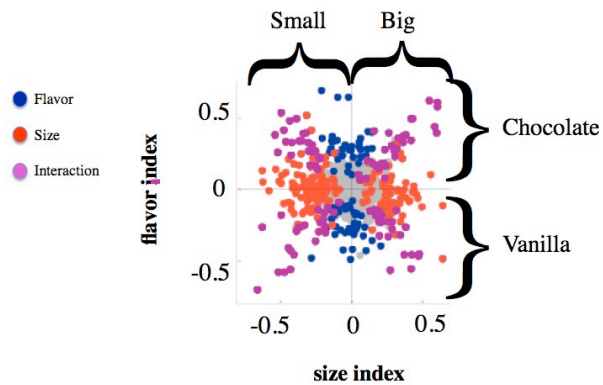
Both model-based and model-free views of OFC signaling predict that OFC neurons will respond based on the size of expected outcomes. Again, however, each view makes distinct predictions for the content of the information encoded. According to the model-based view, OFC encodes value information not only because of its economic function, but because value is derived from

size information, and size is a specific feature of an outcome. Thus, in accordance with model-free accounts of OFC activity, big and small rewards should be represented in OFC in much the same way as chocolate and vanilla rewards. That is, there should be separable populations of big-selective and small-selective cells, the populations should be comparable in size, and the overall population should not be biased in selectivity toward big or small rewards.

The model-free view of OFC offers a different set of predictions for the encoding of size information, consistent with a role of OFC in signaling value. Because reward size is directly related to its value, a model-free OFC would differentially represent big and small rewards. There are several ways distinct representation of big and small rewards could be accomplished within OFC, three of which we explored. For example, only one reward size may be encoded. Alternatively, if both reward sizes are encoded, the population of big-selective cells may be significantly

bigger or smaller than the population of small-selective cells, or, if the sizes of the populations are comparable, there may exist a selectivity-bias toward one size.

Consistent with model-based representations of outcomes in OFC, significant populations of both big-selective and small-selective cells were identified (Figure 5), the sizes of the populations were not significantly different, and there was not a bias in selectivity toward big or small outcomes in the overall population of OFC cells (Figure 6). As with flavor, our classification of selectivity of OFC neurons allowed us to categorize size-selective cells as big-selective or small-selective. According to this analysis, 193 of 831 (23%) of all neurons showed size-selectivity, with 87 firing more for big and 106 firing more for small rewards ( $p = 0.17$  by  $\chi^2$ ). To determine whether the overall selectivity of the population of OFC neurons was biased toward big or small rewards, we calculated a size index for each neuron and plotted the distribution for the population (Figure 6). The size index was calculated as the difference in peak normalized firing in anticipation of big and small reward.



**Figure 6. Size and flavor indices for each OFC neuron.** In the population of OFC neurons, the number of big-selective cells did not differ significantly from the number of small-selective cells ( $p = 0.17$  by  $\chi^2$ ), nor did the number of chocolate-selective cells differ significantly from the number of vanilla-selective cells ( $p = 0.67$  by  $\chi^2$ ). There was also no skew in selectivity for either size ( $t_{830} = -0.96$ ,  $p = 0.34$ ) or flavor ( $t_{830} = 0.57$ ,  $p = 0.57$ ). Additionally, an OFC neuron's size index (dark blue) does not provide predictive information about that cell's flavor index (light blue) and vice versa ( $r = 0.04$ ,  $p = 0.26$ ). Specifically, both small-selective (left) and big-selective cells (right) have a full range of flavor-selectivity, and both chocolate-selective (top) and vanilla-selective cells (bottom) have a full range of size-selectivity.

$$\text{Size Index} = (\text{Big} - \text{Small})$$

Positive size indices corresponded to big-selective cells, and negative size indices corresponded to small-selective cells. The overall average of indices (across all 831 neurons) was  $-0.003 \pm 0.004$ , which was not significantly different from zero by t-test ( $t_{830} = -0.96$ ,  $p = 0.34$ ).

### **OFC Independently Encodes Flavor and Size Information about Outcomes**

If OFC integrates information into a common value signal, we would expect size and flavor to be encoded in conjunction with one another. In other words, we would expect the information OFC signals about chocolate and vanilla rewards to be related specifically to the value, rather than to other features, of those rewards. To illustrate the relationship between size and flavor selectivity, we plotted each neuron's size index against its flavor index (Figure 6). This plot shows similar proportions of neurons in each of the four quadrants, with the distribution of the overall population symmetric about the origin. A small number of neurons encoded both size and identity, including 18 (2%) that showed effects of both, and 44 (5%) that showed an interaction between the two (while meeting the additional statistical test of differentiating the two rewards presented in any one block). Across the entire population, there was no evidence of a correlation between size and flavor indices ( $r = 0.04$ ,  $p = 0.26$ ).

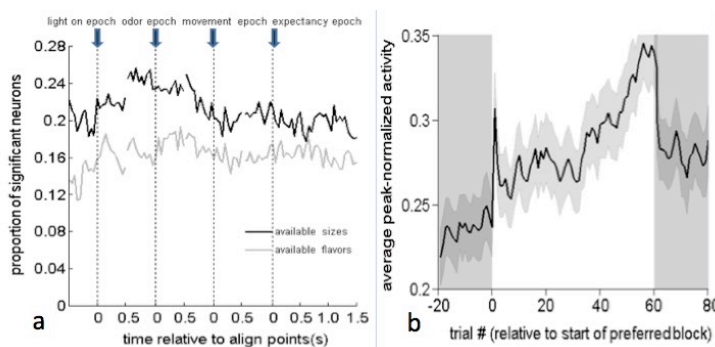
### **OFC Encoding of Task Structure is Consistent with Model-Based Signaling**

While a model-based view of encoding of outcome information in OFC leads to some simple predictions about the signaling of expected outcome information in OFC, it is also important to consider the function of model-based versus model-free representations. An advantage of model-based encoding is flexibility and the ability to incorporate contextual information or rules of a task to optimize behavior. Our task consisted of 5 consecutive blocks of approximately 60 trials each. The rewards presented at each well changed only 4 times during each session, at the start of each new block. Thus, each time reward outcomes change, they remain in their new configuration for about 60 trials. Understanding this simple rule would likely lead to different behavior than would be predicted if behavior were based on experience alone.

Imagine you began our task according to the configuration in Figure 1. Each time you go left, you receive a big reward, and each time you go right, you receive a small reward. Suddenly, after going left, you receive a small reward. What do you do next? Knowing the structure of the task allows you to infer that the big reward is now on the right. You would likely therefore choose to go right, even though you had not yet experienced a big reward on the right. Such a choice – and any accompanying change in neural activity – would be a product of a model-based representation of the task and associated rewards.

Now imagine you are starting a new task, for which you do not know the rules. You again receive big rewards at the left well and small rewards at the right well. After 85 trials, you receive a small reward at the left well. What do you do next? With no knowledge of the rules of the new task, it is impossible to predict with confidence what will happen on the next trial. However, if you rely on your experience thus far in the task, it is 85x more likely to receive the big reward at the left well. Thus, a model-free representation of the task, based on experience, would likely lead you back to the left well. Whether such a choice proved optimal would depend on the rule, but with no representation of the rule, we, like rats and other animals, would likely default to what our experience tells us is the best choice.

We tested OFC for two types of activity that could result from an understanding of task structure and which would enrich a model-based representation of outcomes in OFC. The



**Figure 7. Signaling of Reward Configuration.**

- a) OFC neurons demonstrate block-specific activity in which they signal the current location of each size and flavor of reward.
- b) The firing rate of block-specific cells increases over the course of the block.

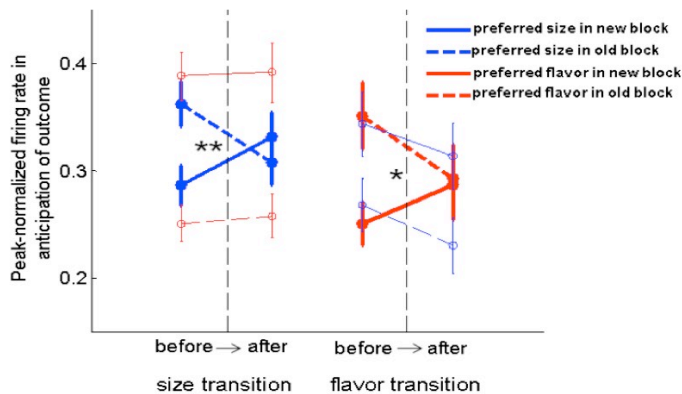
first type of activity for which we tested neurons was signaling of the configuration of available rewards. In every block, there was a chocolate reward at one well and a vanilla reward at the other well. Additionally, there was

a big reward at one well and a small reward at the other well. However, the combinations of reward (i.e. big chocolate, big vanilla, small chocolate, small vanilla) and locations of variations in size and flavor changed from block to block. Our 3-way ANOVA on firing rate, with reward size, flavor and location (left or right well) as factors, identified block-selective cells that did not encode the reward delivered but instead encoded reward configuration, tracking information about both the size and flavor of rewards, as illustrated in Figure 7a. Further, signaling about reward configuration increased across trials within blocks (Figure 7b).

The second type of activity we tested for in OFC was activity consistent with inferential reasoning. In our task, specific rewards available at each well change simultaneously. If, in accordance with Figure 1, a rat unexpectedly receives a small reward at the left well at the start of the second block, understanding this rule, or the task structure, would lead to the inference that the right well now delivers a big reward. Importantly, ‘inference’ implies knowledge *before* experience. Though behavior is believed to be a consequence of both model-based and model-free representations of outcomes, the activity of an area of the brain that signals model-based information should reflect inferential knowledge. On the other hand, model-free areas would be unable to signal inferred information because they are limited to experience.

Model-based and model-free accounts of OFC activity lead to distinct predictions for activity in OFC at the start of a new block, after an unexpected reward has been experienced at one well. If OFC is model-based, OFC signals in anticipation of reward at the second well should reflect knowledge that the reward has changed there as well. However, if OFC is model-free, such signals should reflect anticipation of the reward the rat received throughout the previous block. To illustrate this distinction with one specific case, consider the second block in Figure 1. If on the first trial of block 2, the rat gets a small reward at the left well, each account of OFC activity makes a distinct prediction for OFC activity that occurs on the subsequent trial when the rat anticipates reward at the right well but has not yet received a reward at the right well within the new block. Model-based activity would reflect the accurate anticipation of a big reward at the right well,

with big-selective cells firing more and small-selective cells firing less than at the end of the previous block. Model-free activity, however, would reflect anticipation of a small reward at the right well, consistent with what the rat has been experiencing at this well. In this case, there would be little difference in the firing rates of these cell types at this point, compared to firing rates at the end of the previous block.



**Figure 8. Change in OFC firing pattern during inference of new outcome.** On the first trial after transitions (dotted vertical lines), neurons that are selective for the outcome of the 2<sup>nd</sup> forced choice odor in the current block “new” (solid lines), becoming more active in anticipation of the corresponding reward, while neurons selective for the outcome on the previous block “old” (dashed lines), becoming less active. The interaction between “before/after” and “new/old” (\*  $p < 0.05$ , \*\*  $p < 0.01$ ) indicate that activity in OFC reflects inference of the outcome of the 2<sup>nd</sup> forced choice odor.

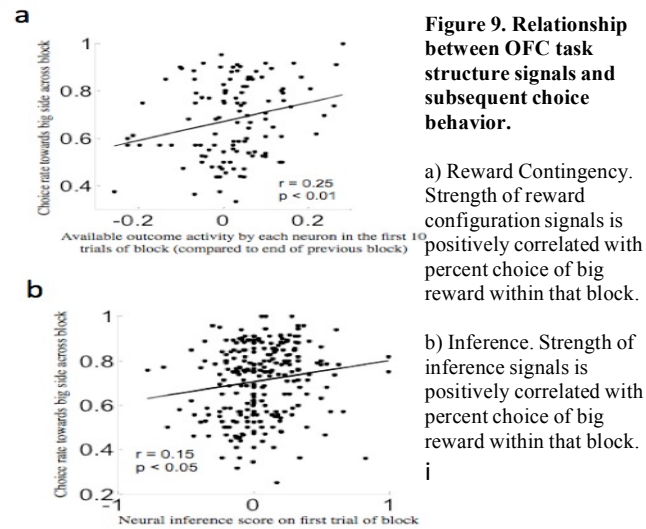
In support of model-based activity in OFC, we found that signaling in OFC reflected inferred outcomes before they were experienced (Figure 8). There was a significant interaction in our 2 ANOVA on firing rates between time (before and after the block switch) and preference point (old block or new block) (\*  $p < 0.05$ , \*\*  $p < 0.01$ ). This was true for both changes in reward size and in

reward flavor (Figure 7). Corresponding interactions in control conditions were non-significant,  $p$ 's  $> 0.80$ .

### Model-Based Encoding of Task Structure in OFC Correlates with Behavior

In general, behavior appears to be the consequence of both model-based and model-free signaling. We were interested in the extent to which model-based signals in OFC influenced behavior in this task. We found that the degree to which OFC signaled either reward configuration or inferred reward was correlated with subsequent behavior. Specifically, the more OFC signaled task specific information, the more likely rats were to maximize their reward acquisition within that block, as both the strength of reward configuration signals (Figure 9a) and the strength of inference signals (Figure 9b) correlated with percent choice of big reward within blocks (reward configuration:  $r =$

0.25,  $p < 0.01$ , inference:  $r = 0.15$ ,  $p < 0.05$ ). These correlations between model-based signals in OFC and subsequent behavior provide evidence for the functional role of OFC model-based representations of outcomes and suggest that the OFC assists in adapting to changing contingencies. Further, the importance of these signals for adaptive behavior is supported by the findings that optimal choice behavior resulted more frequently when model-based signals in OFC were more pronounced.



## SUMMARY

Here we provide support for the contribution of model-based representations in OFC to decision making. We observed OFC signaling of information about outcomes that reflected a rich, flexible understanding of task structure and potential outcomes. Such flexibility is inconsistent with model-free accounts of OFC function, including the notion that OFC signals value in a common currency. Indeed, if OFC signaled information in a common currency, we would expect distinct encoding of outcomes with distinguishable values. However, we found big and small rewards were equally represented in OFC. Further, equally valued flavors were also differentially represented within OFC, suggesting that the function of OFC is not limited to value assessment. Finally, if OFC represented information in a model-free way, expectancy signals would reflect the most recently experienced or most abundantly experienced outcomes at each well. However, OFC signaled information consistent with inferred outcomes that had not yet been paired with the stimulus predicting them. That these signals were correlated with subsequent behavior suggests that these OFC signals are important for the flexible behavior that is a hallmark of decision making facilitated by model-based signals. This model-based account of OFC activity is consistent with findings that OFC encodes a variety of

variables associated with outcomes and that OFC activity is modulated by contextual information and task structure. Further, these findings point to a potential mechanism underlying the failure of these factors to adequately influence behavior when OFC is offline.

## **CHAPTER 3: MODEL-BASED SIGNALING OF OUTCOMES IN VENTRAL STRIATUM**

### **INTRODUCTION**

Though the precise role ventral striatum (VS) plays in decision making is unclear, several studies support the view that VS signals value information in a common currency (Li, McClure et al. 2006, Daw, Gershman et al. 2011, Garrison, Erdeniz et al. 2013), making it part of the model-free reinforcement learning system. Indeed it has been argued that VS serves as the critic in a model-free actor-critic system in which a critic supplies expected value information that is used by an actor to select the most advantageous action (O'Doherty, Dayan et al. 2004, Li, McClure et al. 2006). However, VS receives heavy input from orbitofrontal cortex (OFC), an area implicated in representing higher order, model-based information (Jones, Esber et al. 2012), and we have recently shown that both VS and OFC are necessary for behaviors that require model-based representations of outcomes (McDannald, Lucantonio et al. 2011). These data suggest that both model-free and model-based information may be encoded within VS.

We tested this hypothesis by coupling single-unit recording in VS with a behavioral task that allowed independent manipulation of value (size) and a specific feature (flavor) of reward. In our task, rats learned to go to two separate locations for a milk reward, depending on which one of two odors was presented. Alternatively, if a third odor, the free choice odor, was presented, rats could choose the location at which to receive reward. Rewards came in two sizes (big and small) and two flavors (chocolate and vanilla). Thus, we distinguished VS activity related to expectations of value versus specific features of outcomes by independently manipulating size (value) and flavor (specific feature) while recording neural activity. This procedure allowed us to assess the presence of model-based signals in VS, as model-free signals are affected only by the reward size information in this task. Specifically, VS encoding of specific features of outcomes (flavor) would support our hypothesis that VS incorporates model-based

information. Further, to explore the importance of input from OFC, a putative model-based region, we also assessed activity in VS in rats with unilateral OFC lesions.

The results show that some VS neurons discriminate between cues predictive of big versus small rewards, whereas other neurons discriminate impending reward flavors, independent of value. These results suggest that VS contains both general, model-free and also specific, model-based representations of outcomes. Further, consistent with a role for OFC in contributing specific information to VS, we found that OFC lesions disrupted encoding of the flavor features but left relatively intact encoding based on reward value.

## **MATERIALS & METHODS**

### **Subjects**

Nine male Long-Evans rats, weighing 275-300 g were acquired from Charles River Laboratories. They were housed individually in a 12 hour light/dark schedule environment and given *ad libitum* access to food prior to testing. Testing occurred during the rats' light cycle, and during this time, rats were food deprived to 85% of their baseline weight. All testing was performed in accordance with the guidelines set forth by the University of Maryland School of Medicine Animal Care and Use Committee and the National Institutes of Health.

### **Surgical Procedures**

Aseptic, stereotaxic surgical techniques were used to implant a drivable bundle of wires in each hemisphere of the VS in each rat. 10 FeNiCr wires (Stablohm 675, California Fine Wire, Grover Beach, CA), each with a diameter of 25  $\mu\text{m}$ , were cut with surgical scissors to spread  $\sim 1\text{mm}$  beyond the cannula and the electroplated with platinum ( $\text{H}_2\text{PtCl}_6$ , Aldrich, Milwaukee, WI) to an impedance of  $\sim 300$ . The drivable bundle of wires were then immediately chronically implanted dorsal to VS 1.6 mm anterior to bregma, -1.5 mm laterally, and 4.5 mm ventral to the brain surface in each hemisphere of each rat. Following surgery, cephalexin (15 mg/kg p.o.) was administered twice a day for two weeks to prevent infection.

A glass micropipette attached to a picospritzer via plastic tubing was used to create unilateral OFC lesions with infusions of 0.1  $\mu$ L of NMDA (20  $\mu$ g/ $\mu$ L) at two separate sites: 1) 4.0 mm anterior to bregma, and at 2.2 and 3.7 mm lateral to the midline, at a depth of 4.2 mm ventral to the skull surface, 2) 3.0 mm anterior to bregma, 3.2 and 4.2 mm lateral to the midline, and 5.2 mm ventral to the skull surface. X rats received NMDA infusions. At the conclusion of the experiment, a 15  $\mu$ A current was passed through each electrode to mark the final position of the electrode. Subsequently, rats were perfused, and their brains were subjected to standard histological techniques (Schoenbaum, Chiba et al. 1999).

### **Apparatus**

Behavioral training and recording were conducted in 18" x 18" aluminum chambers with sloping walls that narrowed to an area of 12" by 12" at the bottom. On the right wall of each chamber, there were two fluid wells separated by an odor port. Olfactory cues were delivered to the odor port through an air flow dilution olfactometer. These cues were provided by International Flavors and Fragrances (New York, NY). Photobeam interference indicated rat entry into the odor port or fluid wells. All parameters of the task were controlled via computer.

## **EXPERIMENTAL PROCEDURES**

### **Choice Task**

Our choice task is illustrated in Figure 1. Trials in our choice task began with the illumination of a light. Once a rat's nose was in the opening of the odor port for 0.5 seconds, an odor was delivered to a small hemicylinder located behind the odor port opening. The odor was presented for 0.5 seconds. Following odor presentation, there was a 0.5 seconds delay, after which a reward was delivered to the appropriate well, provided the rat entered the well within 3 seconds of the odor offset.

Rats learned to go to the well to the right of the odor port or to the well to the left of the odor port, depending on which one of two forced choice odors were presented in the odor

port. Alternatively, if a third free choice odor was presented, rats could choose either well, and reward was delivered only at the first well entered. These three odors were constant throughout the experiment. Odors were presented in pseudorandom order such that 7/20 trials were ‘free choice,’ and an equal number of each forced choice odor was presented across the session. No odor was presented on 3 consecutive trials. Reward in this task was chocolate or vanilla flavored milk, diluted to 50% concentration with deionized water. These flavors were chosen because we have previously shown that rats can distinguish chocolate and vanilla flavored milk but do not have a preference for one flavor over the other (McDannald, Takahashi et al. 2012). A ‘small’ reward was a 0.05 mL bolus, and a ‘big’ reward was two such boli delivered 0.5 seconds apart.

After rats were trained to perform this simple task, we introduced blocks where the size and flavor of the reward at each well were independently manipulated. Once rats maintained accurate responding in this more complicated version of the task, we commenced recording sessions.

### **Single-Unit Recording**

Single-unit recording procedures were the same as previously described (Roesch et al., 2006). Sessions began after active wires were selected, and the electrode was advanced 40 or 80  $\mu\text{m}$  at the end of each session. If activity was not detected at the start of a session, the rat was removed from the chamber, and the electrode was advanced.

Neural activity was recorded using two identical Plexon Multichannel Acquisition Processor systems (Dallas, TX), interfaced with odor discrimination training chambers described above. Signals from the electrode wires were amplified 20x by an op-amp headstage (Plexon Inc HST/8050-G20-GR), located on the electrode array. Outside the training box, the signals were passed through a differential preamplifier (PBX2/16sp-r-G50/16fp-G50; Plexon), in which the single-unit signals were amplified 50x and filtered at 150–9000 Hz. The single-unit signals were then sent to the Multichannel Acquisition Processor box, in which they were further filtered at 250–8000 Hz, digitized at 40 kHz, and amplified at 1–32x. Waveforms ( $>2.5:1$  signal-to-noise) were extracted from active

channels and recorded to disk by an associated workstation with event timestamps from the behavior computer. Waveforms were not inverted before data analysis.

Offline Sorter software from Plexon Inc (Dallas, TX) was used to sort units with a template matching algorithm. Files were then converted to Neuroexplorer files, where event markers and unit timestamps were extracted. These data were analyzed in Matlab (Natick, MA).

## **Statistical Analysis**

### *Behavioral Analysis*

For our behavioral analysis, we measured overall performance on the task by calculating the percent of trials where rats correctly chose the well associated with the forced choice odor that was presented, and we measured preference by calculating the percent of free choice trials where the rat chose each size and flavor outcome. To determine if there were significant differences in preference for big versus small rewards or chocolate versus vanilla rewards in control rats, we performed t-tests on choice rate on free choice trials for big and small rewards and also for chocolate and vanilla rewards. To compare free choice behavior in controls and lesioned rats, we performed a mixed ANOVA on the difference between choice rate and 50%. Our factors were group (sham or lesion) and reward feature (size or flavor).

In order to test whether accuracy on forced choice trials indicated a preference between reward sizes or between reward flavors, we performed a 3-way-mixed design ANOVA on percent of correct forced choice trials on the last 25 trials of each block, with size (big or small) and flavor (chocolate or vanilla) as within-subject factors and group (control or lesion) as the between-subject factor. We repeated this analysis, replacing percent correct scores with reaction time to test if response latency on forced choice trials further demonstrated preference for big rewards without regard to flavor, and if there were differences in this measure across groups.

Finally, to determine what factors influenced the change in behavior that occurred across block transitions, we performed a mixed design ANOVA on the difference scores of choice rate in the first 25 trials of a new block compared to the last 25 trials of the previous block, with transition type (size or flavor) and initial flavor (chocolate or vanilla) as within-subject factors and group (control or lesion) as between-subject factors.

### *Neural Analysis*

To determine if cells were selective for big, small, chocolate, or vanilla rewards, we analyzed reward-anticipatory activity on correct forced choice trials because trials with each reward feature were equally represented within a session, whereas the numbers of free choice trials were biased towards the larger reward. We first performed a sliding ANOVA analysis on 500 ms epochs sliding forward by 50 ms, beginning right before odor delivery and ending during the reward. We performed a more extensive analysis of the 500 ms epoch beginning when the rat reached the reward well and ending upon receipt of reward. These were 3-way ANOVAs on firing rate, with reward size, flavor and direction (left or right well) as factors. Baseline firing rates were subtracted on each trial because we found that baseline rates carried information about reward features that were available in a particular block (see Results). To allow averaging of neurons with different firing rates, we also peak-normalized firing rates by dividing all firing rates for each neuron by the 100 ms bin with the highest mean firing rate across the first ten or last ten trials of each condition.

Once neurons were categorized according to selectivity, we performed chi square tests to determine if differences existed in the size of the populations of big versus small-selective cells or chocolate versus vanilla-selective cells. Further, to identify any bias in the overall selectivity of the population of VS cells, we calculated size and flavor indices for each cell and performed t-tests on these indices. We also used chi square tests to compare the proportion of cells encoding big and small and chocolate and vanilla in OFC-lesioned hemispheres versus those in sham-lesioned rats. And finally, we compared the magnitude of size and flavor indices in neurons recorded from OFC-lesioned hemispheres versus that in neurons recorded from sham-lesioned rats.

The ability of VS neurons to incorporate information about task structure was assessed through analysis of block-specific activity and inference signaling in VS. Block-specific activity, which we hypothesized to encode the response-reward contingencies during each block, was identified using a 3-way ANOVA, with response-size contingencies (big on left and small on right or small on left and big on right), response-flavor contingencies (chocolate on left and vanilla on right or vanilla on left and chocolate on right), and response direction (left or right) as factors, on each cell's firing rate across trials. We considered that a cell exhibited significant block-selectivity when it had a main effect of available size or flavor, or an interaction between them, without that effect also interacting with direction (the reason for the last condition was that, if an effect interacted with response direction, that would mean that that cell was signaling not the block *per se*, but something about the reward available on a particular trial within a block). We ran this ANOVA as a sliding analysis, with 500 ms epochs sliding forward by 50 ms, beginning before the odor was delivered and ending when the reward was delivered. We examined block-selective activity in more detail during the epoch immediately before odor delivery, because this was found to be the epoch with highest proportion of block-selective activity in the sliding analysis. To assess differences in the amount of block-selectivity between control and lesioned rats, we used a 3-way ANOVA with factors group (sham or lesion), feature (size or flavor), and time within block (early, middle, or late) on the proportion of cells showing a significant block-selective effect.

To analyze inference signaling, we examined activity among size-selective or flavor-selective neurons on the first forced-choice trial in whichever direction occurred *second* at the start of blocks. The rationale for calling activity on this trial an *inference signal* is that rats would not yet have experienced that side's outcome on that specific block. Nevertheless, they could infer that a block switch had occurred and thus which outcome they would receive on that trial, based on the new outcome they had received on the other side. We compared activity on this first trial in the second direction with average activity on the last two trials in the same direction in the previous block. We reasoned that an inference could manifest as a spontaneous increase in activity when a neuron's preferred

outcome would be available on that side in that block, or a spontaneous decrease in activity when a neuron's preferred outcome would no longer be available on that side in that block. We called this factor preference point (with levels old block or new block).

We tested for the presence of inference signals by performing an ANOVA on firing rates across neurons, with factors: preference direction (left or right), time (before or after the block switch), group (control or lesion), and reward feature (size or flavor). Included in the analysis were all size-selective and flavor-selective neurons. Inference signals would be reflected in a significant interaction between preference direction and time for size-selective neurons across size switches and for flavor-selective neurons across flavor switches. We also included a control condition, which was to look at the same comparison across block switches in which the relevant feature, for a particular neuron, was not switched. In other words, there should be no interaction between preference direction and time for size-selective neurons across flavor switches, and for flavor-selective neurons across size switches.

To determine if there was a relationship between the so-called inference signal and subsequent behavior, we calculated an inference score for each cell for relevant block switches, which was a measure of the accuracy of the cell's anticipatory information (i.e. a high score indicated that the cell was signaling the expectation of the correctly inferred new outcome, and a low score indicated that the cell was signaling the expectation of the reward that had been delivered in the previous block). These scores were then plotted against the rats' performance on free choice trials within that same block. Performance was calculated as the percent of free choice trials for which the rat chose the well with the big reward.

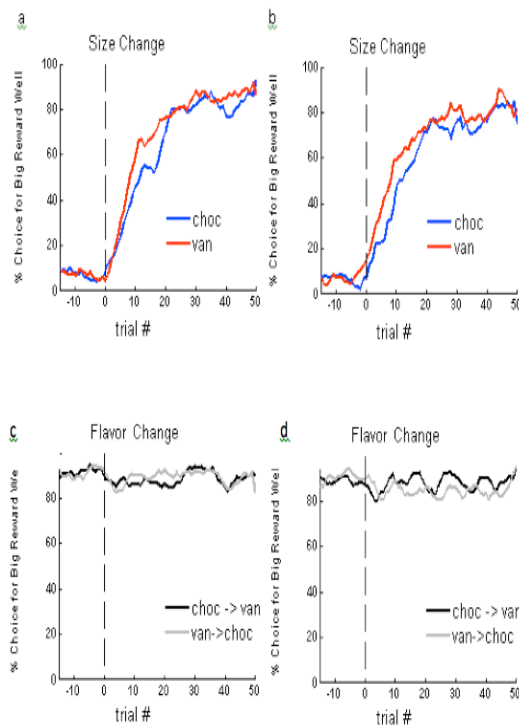
## **RESULTS**

### **Behavior Reveals Size Preference but not Flavor Preference**

Consistent with previous data, we expected rats to demonstrate a preference for big over small rewards in a choice task, while showing no preference between chocolate and vanilla flavors. We expected to observe this pattern of choice behavior not only in

normal, control rats, but also in rats with unilateral OFC lesions, as we have consistently shown that such lesions do not generally alter behavior. To test these predictions, we compared each group's performance on free choice and forced choice trials.

We analyzed free choice behavior by assessing the choice rate for each type of reward. We expected both intact and lesioned rats to choose big rewards more often than small reward but to demonstrate no difference in choice rate for chocolate versus vanilla rewards. Consistent with this expectation,

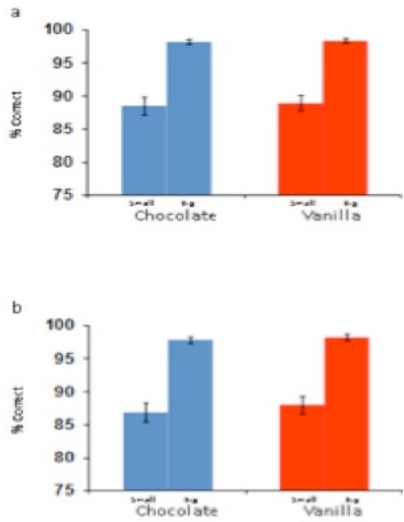


**Figure 10. Influence of reward size and flavor on free choices.**

- a) **Size Shift in Control Rats.** Before a size shift, indicated by the dotted vertical line, rats rarely chose the well that delivered the small reward on free choice trials. Once the size changed, the rats' responses changed accordingly, and they consistently chose the well that delivered the big reward, regardless of the flavor of reward.
- b) **Flavor Shift in Control Rats.** Before a flavor shift, indicated by the dotted vertical line, rats consistently chose the well with big reward on free choice trials. When the flavors changed, the rats continued to choose the well with the big reward, regardless of the direction of the flavor switch (chocolate to vanilla or vanilla to chocolate).
- c) **Size Shift in Lesion Rats.** Lesion rats responded to size shifts similarly to control rats. Both groups chose the big reward significantly more than the small reward ( $F_{1,148} = 9263, p = ?$ ), but there was no interaction between size and group ( $F_{1,148} = 0.67, p = 0.41$ ).
- d) **Flavor Shift in Lesion Rats.** Lesioned rats responded to flavor shifts similarly to control rats. Neither group demonstrated a preference for one flavor over the other ( $F_{1,148} = 1.3, p = 0.25$ ), and there was no interaction between flavor and group ( $F_{1,148} = 1.1, p = 0.29$ ).

both groups performed similarly on free choice trials, choosing big rewards (sham:  $91.0 \pm 0.59\%$ ; lesion:  $90.3 \pm 0.59\%$ ) more often than small rewards (sham:  $9.0 \pm 0.59\%$ ; lesion:  $9.7 \pm 0.59\%$ ) but choosing chocolate (sham:  $49.9 \pm 1.2\%$ ; lesion:  $48.0 \pm 1.4\%$ ) and vanilla (sham:  $50.1 \pm 1.2\%$ ; lesion:  $52.0 \pm 1.5\%$ ) rewards to a similar extent (Figure 10). A mixed ANOVA on difference between choice percentage from 50%, with group (sham or lesion) and reward feature (size or flavor) as factors, demonstrated no effect of group ( $F_{1,148} = 1.7, p = 0.20$ ), but a significant effect of reward feature ( $F_{1,148} = 1703.8, p < 0.000001$ ), with no interaction between group and reward feature ( $F_{1,148} = 0.4, p = 0.54$ ). Planned comparisons revealed that choice for the big reward was significantly greater than 50%

( $F_{1,148} = 9263$ ,  $p < 0.00001$ ). This was true of both groups (Figure 10), as there was no interaction between size (big or small) and group ( $F_{1,148} = 0.67$ ,  $p = 0.41$ ). Choice for chocolate (and thus vanilla as well) was not significantly different from 50% ( $F_{1,148} = 1.3$ ,  $p = 0.25$ ). This was also true of both groups, with no interaction between flavor (chocolate or vanilla) and group ( $F_{1,148} = 1.1$ ,  $p = 0.29$ ).



**Figure 11. Influence of reward size and flavor on accuracy in forced choices.**

**a) Control Rats.** Control rats are more likely to respond accurately on forced choice trials when directed to the well with the large reward, regardless of flavor.

**b) Lesion Rats.** Rats with unilateral OFC lesions performed with similar accuracy on forced choice trials as control rats. There was a main effect of size ( $F_{1,148} = 237.5$ ,  $p < 0.000001$ ) no effect of flavor ( $F_{1,148} = 0.78$ ,  $p = 0.38$ ), or group ( $F_{1,148} = 0.96$ ,  $p = 0.33$ ), and no interactions of group by size ( $F_{1,148} = 0.63$ ,  $p = 0.69$ ), or group by flavor ( $F_{1,148} = 0.15$ ,  $p = 0.69$ ).

We analyzed forced choice behavior using accuracy and response latency. We expected both intact and lesioned rats to perform more accurately and respond with shorter latencies when the big reward was at stake but to show

no effect of whether it was chocolate or vanilla. As expected, during the last 25 trials of each block, the percent of correct forced choice trials was greater for trials where the odor indicated that the rat would receive a big reward compared to those for which the odor predicted a small reward (Figure 11). Additionally, there was no difference in accuracy between trials predicting chocolate and vanilla rewards. Further, this pattern of behavior on forced choice trials was consistent for control and lesion rats. A 3-way mixed ANOVA on percent correct scores on forced choice trials, with group (sham or lesion), size (big or small) and flavor (chocolate or vanilla) as factors, revealed a main effect of size ( $F_{1,148} = 237.5$ ,  $p < 0.000001$ ) and no effect of flavor ( $F_{1,148} = 0.78$ ,  $p = 0.38$ ) or group ( $F_{1,148} = 0.96$ ,  $p = 0.33$ ). There were no interactions of group by size ( $F_{1,148} = 0.63$ ,  $p = 0.69$ ) or flavor ( $F_{1,148} = 0.15$ ,  $p = 0.69$ ).

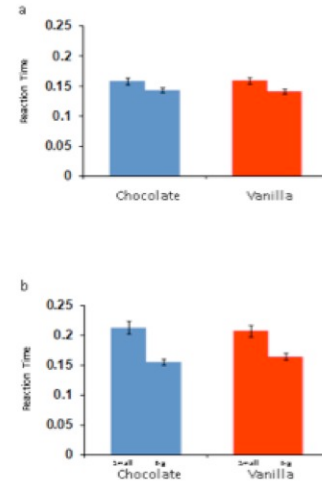
Rats were also faster to respond for big versus small rewards with no effect of flavor (Figure 12). A 3-way mixed ANOVA, performed using average reaction times on the last 25 trials of each block demonstrated a main effect of size ( $F_{1,148} = 55.3$ ,  $p < 0.00001$ ) but

not flavor ( $F_{1,148} = 0.049$ ,  $p = 0.82$ ). Reaction time was specifically defined as the time from odor cessation to odor port exit. Notably, there was an effect of group ( $F_{1,148} = 27.2$ ,  $p = 0.000001$ ), indicating that rats with lesions were significantly slower than control rats in responding, regardless of outcome. This finding is consistent with our previous findings that OFC-lesioned rats fail to modulate their reaction times based on the value of expected outcomes (Schoenbaum, Setlow et al. 2003).

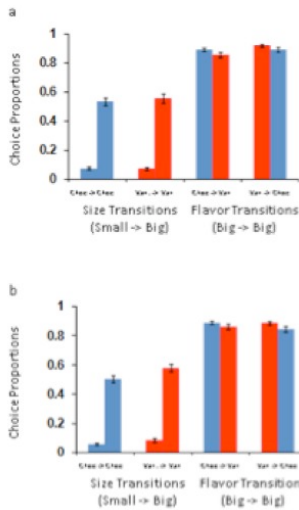
**Figure 12. Influence of reward size and flavor on reaction time in forced choices.**

**a) Control Rats.** Control rats are more likely to respond quicker on forced choice trials when directed to the well with the large reward, regardless of flavor.

**b) Lesion Rats.** Rats with unilateral OFC lesions performed slower than control rats on forced choice trials ( $F_{1,148} = 237.527.2$ ,  $p = 0.000001$ ).



Finally, because each block begins with a change in response-outcome contingencies, and because we expected rats' responding to be influenced by reward size but not reward flavor, we expected to see behavior in both groups change across block transitions



**Figure 13. Changes in free choice behavior across block transitions.**

**a) Control Rats** Free choice behavior in control rats is influenced by reward size but not reward flavor.

**b) Lesion Rats** Free choice behavior in response to size and flavor transition in lesion rats is similar to behavior in control rats.

according to the location of the big reward but without regard to the location of each flavor. Accordingly, both groups showed changes in choice behavior across block transitions in response to shifts in reward size but not flavor (Figure 13). Block transitions were measured as the difference between choice rate in the first 25 trials of a new block and the last 25 trials of the

previous block. A mixed ANOVA with between-subjects factors: group (sham or lesion) and initial flavor (chocolate or vanilla) and a within-subjects factor: altered reward feature (size or flavor), showed a main effect of altered reward feature ( $F_{1,146} = 789$ ,  $p < 0.000001$ ) but no effect of initial flavor ( $F_{1,146} = 0.80$ ,  $p = 0.37$ ) or group ( $F_{1,146} = 0.008$ ,  $p = 0.93$ ) and no interactions ( $F_s < 1.2$ ,  $p$ 's  $> 0.27$ ). Planned comparisons demonstrated that

size transition scores were significantly greater than zero ( $F_{1,146} = 949$ ,  $p < 0.000001$ ), indicating that changing the location of big and small rewards significantly influenced the location rats chose. This was true of both groups, as there was no interaction between size transition and group ( $F_{1,146} = 0.006$ ,  $p = 0.94$ ). For flavor transitions, scores were significantly less than zero ( $F_{1,146} = 10.1$ ,  $p = 0.0018$ ), with no interaction of flavor transition and group ( $F_{1,146} = 0.003$ ,  $p = 0.95$ ). Furthermore, neither group had a preference between chocolate and vanilla rewards, as there was no difference in transition scores for chocolate to vanilla and vanilla to chocolate transitions ( $F_{1,146} = 0.014$ ,  $p = 0.91$ ) and no interaction with group ( $F_{1,146} = 0.40$ ,  $p = 0.53$ ).

### **Anticipatory Activity in VS Reflects Both Size and Flavor Information**

After experiencing the odor, rats entered a reward well, and 500 ms later, reward was delivered. We analyzed activity during this 500 ms delay when rats had the information to predict the impending reward but had not yet experienced that reward; thus, signaling during this period reflected reward-anticipatory activity. We recorded 399 neurons in VS over 83 sessions in four control rats. ANOVAs performed on each cell led us to identify size-selective and flavor-selective activity. Our analysis further allowed us to subcategorize size-selective cells as big-selective or small-selective and flavor-selective cells as chocolate-selective or vanilla-selective. In addition, we calculated a size selectivity index and a flavor selectivity index for each neuron, defined as the difference in average peak-normalized firing rate between the two conditions:

$$\text{Size Index} = (\text{Big} - \text{Small})$$

$$\text{Flavor Index} = (\text{Chocolate} - \text{Vanilla})$$

Thus, big-selective cells were represented by positive size indices, whereas small-selective cells were represented by negative size indices; and chocolate-selective cells were represented by positive flavor indices, whereas vanilla-selective cells were represented by negative flavor indices.

### **VS Encoding of Outcome Size in Control Rats is Partially Consistent with Model-Based Representations of Outcomes**

If encoding of size information is model-free, size signals should reflect the value inherent to reward size. Thus, the discrepancy in value of big and small rewards should translate to differential encoding of reward size. Specifically, if model-free signals are directly proportional to value, a model-free VS should demonstrate a selectivity bias toward big rewards (or small rewards if coding value inversely). If instead, encoding of size information in VS is model-based, then size signals should reflect the specific feature aspects of size – e.g. the number of drops or volume of reward – without regard to value. Therefore, representation of big and small rewards would be comparable in VS. As we believe VS to be a hybrid of model-free and model-based information, we expected a mix of these signal types.

Consistent with model-based signaling, there was not a significant difference in the proportion of big-selective and small-selective cells in VS ( $c^2 = 0.87$ ,  $p = 0.35$ ) (Figure 14c). Further, among size-selective cells, the average size-selectivity index was not significantly different from zero ( $t_{71} = -0.66$ ,  $p = 0.51$ ), suggesting that size-selective cells signal model-based rather than model-free information. However, when considering our entire population of VS cells, size selectivity was significantly skewed toward small rewards ( $t_{398} = -2.3$ ,  $p = 0.020$ ). Though differential encoding of size information may support model-free signaling in VS, it is perhaps surprising that rather than greater representation of the more valuable, big reward, VS shows greater representation of the less valuable, small reward.

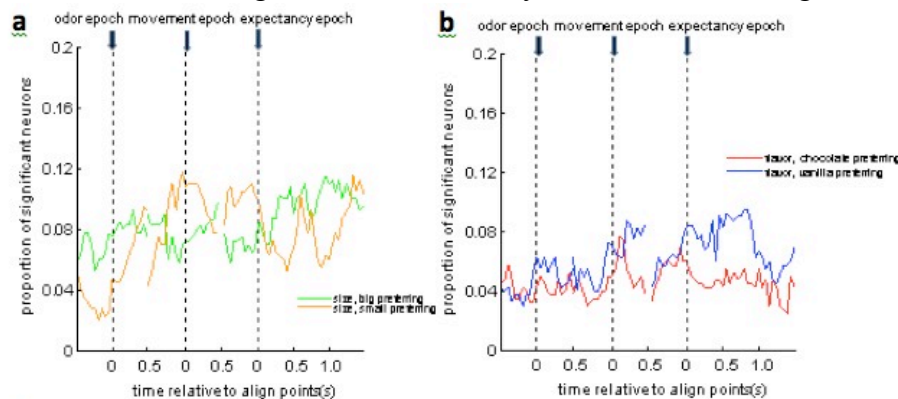
### **VS Encoding of Flavor in Control Rats is Consistent with Model-Based Representations of Outcomes**

If VS signals are purely model-free, chocolate and vanilla flavors should not be reflected in VS activity because each flavor is equivalent with respect to value. However, model-based signals would reflect the specific features of reward, even when the value of reward is equivalent. Thus, in accordance with our hypothesis that VS encodes model-based information, we expected VS to signal distinct expectations of each flavor. Consistent with our predictions, these populations appeared to contribute equally to population activity (Figure 14c), as there was no significant difference in the numbers of

chocolate and vanilla cells ( $c^2 = 1.4$ ,  $p = 0.23$ ). Additionally, there was no overall skew in selectivity toward either flavor among either the flavor-selective population ( $t_{56} = -0.20$ ,  $p = 0.85$ ) or the overall VS population ( $t_{398} = -0.61$ ,  $p = 0.54$ )

## VS Encoding of Task Structure is Consistent with Model-Based Representations of Outcomes

While encoding of size and flavor information provides a descriptive analysis of potential model-free and model-based signals, we were also interested in the functional application of model-based signals. The flexibility in decision making afforded by model-based



**Figure 14. Flavor and size selectivity in VS of control rats.**

**a) Proportions of size-selective cells.** There are both big-selective and small-selective cells in VS of control rats. These populations are of comparable size.

**b) Proportions of flavor-selective cells.** There are both chocolate-selective and vanilla-selective cells in VS of control rats. These populations are of comparable size.

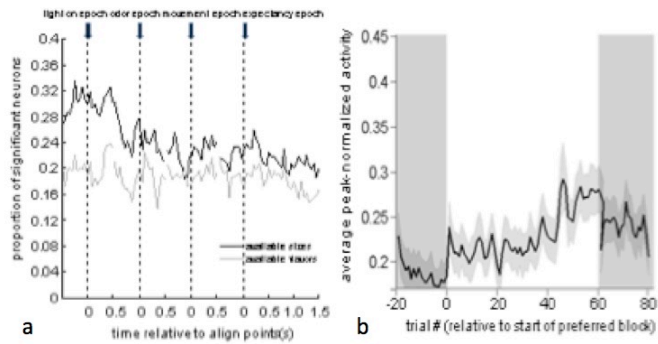
**c) Size and flavor indices.** Big and small populations are of similar size, as are chocolate and vanilla populations. Within the size and flavor-selective cells, there is no bias toward one size or flavor. However, across the population, there is a bias toward small rewards.

influence that factors such as

context have over behavior is thought to result from the use of such models to infer future consequences or outcomes. This is most apparent in our task on free choice trials, in which understanding the rules of the task can help maximize reward. With this idea in mind, we exploited our behavioral design to test for evidence of model-based processing in VS in two ways.

representations of specific features of outcomes stems from animals' ability to incorporate information that is otherwise independent of value to assess the value of current opportunities. Indeed, the rules and

First, we tested VS for block-specific activity that encoded the current configuration of rewards within each block. In each block, chocolate and vanilla rewards were available, big and small rewards were available, and reward was available at both the left and right well. However, the specific combination of size, flavor, and location of the rewards changed from block to block. We believed that VS would contain neurons that kept track of both size and flavor contingencies during each block. Such information would represent the structure, or model, of the task. Consistent with model-based encoding of outcomes in VS, we identified a considerable proportion of block-specific cells that signaled reward configuration (Figure 15). Specifically, these cells signaled the combination of reward features (big, small, chocolate, vanilla) available on the left and on the right throughout individual trials (Figure 15a) and developing across the duration of blocks (Figure 15b).



**Figure 15. Block-specific activity in VS of control rats.**

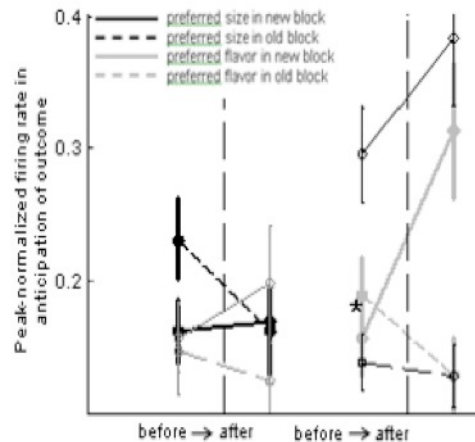
- a) Reward contingency activity across trials.** VS signals information about the current location of reward based on size (black) and flavor (gray).
- b) Reward contingency activity, collapsed across factor, across blocks.** VS maintains or increases its signaling of reward contingencies across the 60 trials within each block.

Next, we tested VS for signaling of inferential information. Based on the task structure, a change in reward at one well indicated a block switch had taken place, meaning that the reward available at both wells was different from those available in the previous block. Therefore, experiencing a reward change at one well could allow rats to infer the size and flavor of the new reward at the other well. Inference signals would indicate knowledge of the task structure, or task model, and could assist in reward maximization when the locations of big and small rewards were switched. If the rat unexpectedly received a small reward on the first trial of a new block, an ability to infer the new location of the big reward would allow the rat to choose accordingly from the first free choice trial, even if the rat had not yet experienced that specific reward-location contingency. Such forward-thinking or mental simulation would require model-based representations of

outcomes. If, on the other hand, the rat's behavior were dominated by model-free signals that are based on experience, the rat would likely only have knowledge of where he last experienced the big and small rewards. Thus neural signals, at the start of a new block, associated with model-free learning, would continue to anticipate the reward configuration from the previous block, whereas

model-based signals would anticipate the reward configuration of the new block without having to experience new reward at both locations. Consistent with model-based signaling, VS neurons demonstrated significant inference signaling (Figure 16). These inference signals appeared to be

stronger for flavor inferences, but the interaction between preference direction and time was significant across both kinds of transitions ( $F_{1,242} = 16.2$   $p < 0.0001$ ).



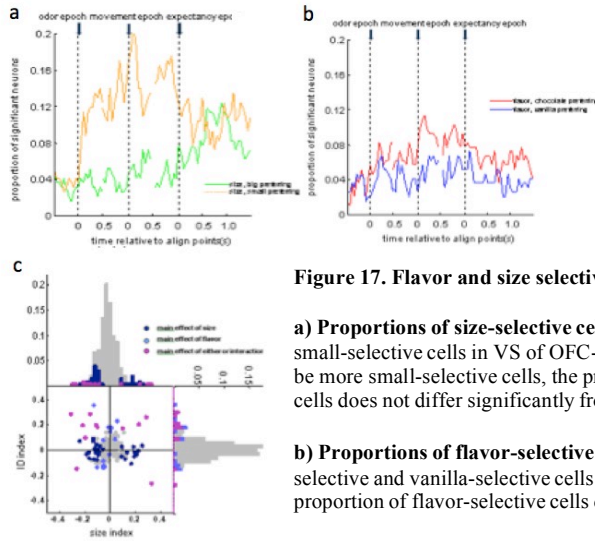
**Figure 16. Change in VS firing pattern during inference of new outcome.** On the first trial after transitions (dotted vertical lines), neurons that are selective for the outcome of the 2<sup>nd</sup> forced choice odor in the current block “new” (solid lines), becoming more active in anticipation of the corresponding reward, while neurons selective for the outcome on the previous block “old” (dashed lines), becoming less active. VS demonstrated significant inferences with respect to flavor but not size.

### Model-Based Size Signals Are Not Significantly Disrupted by OFC Lesions

In accordance with our hypothesis that model-based information encoded by VS neurons derives from OFC input, we expected the model-based component of size in VS to be diminished or eliminated when OFC was lesioned. If this were the case, size selectivity would be more skewed in lesioned rats than in controls. However, there was not a significant difference in the proportion of big or small-selective cells in lesion versus control rats ( $\chi^2 = 0.75$ ,  $p = 0.38$ ). Though Figure 17a appears to demonstrate that activity in VS of lesioned rats was significantly skewed toward selectivity for small rewards, the distribution of size selectivity indices in neurons in these rats did not differ significantly from that of controls ( $t_{590} = 0.45$ ,  $p = 0.65$ ). Additionally, there was not a significant difference in selectivity indices among size-selective cells ( $t_{110} = 0.21$ ,  $p = 0.83$ ) (Figure 17c). Thus, inconsistent with our predictions, we did not observe differences in encoding of size information in control and lesioned animals.

## Model-Based Flavor Signals in VS Are Disrupted by OFC Lesions

Because we hypothesized that model-based representations of outcomes in VS were



**Figure 17. Flavor and size selectivity in VS of OFC-lesioned rats.**

**a) Proportions of size-selective cells.** There are both big-selective and small-selective cells in VS of OFC-lesioned rats. Though there appear to be more small-selective cells, the proportion of big and small-selective cells does not differ significantly from the proportions in control rats.

**b) Proportions of flavor-selective cells.** There are both chocolate-selective and vanilla-selective cells in VS of OFC-lesioned rats. The proportion of flavor-selective cells did not differ from control rats.

and vanilla-selective cells in response to OFC lesions. Surprisingly, the proportion of flavor-selective cells in the population of VS neurons did not differ significantly in lesion and control rats ( $c^2 = 0.27$ ,  $p = 0.60$ ), as illustrated in Figure 17b. However, consistent with our predictions, eliminating OFC input to VS did disrupt normal encoding of flavor information. Rats with OFC lesions demonstrated asymmetrical flavor selectivity such that these rats had a higher proportion of chocolate-selective cells than control rats ( $c^2 = 6.8$ ,  $p = 0.009$ ). Flavor selectivity in VS of lesioned rats was also biased toward chocolate rewards in both flavor-selective cells ( $t_{80} = -2.3$ ,  $p = 0.025$ ) (Figure 17c) and in the population as a whole ( $t_{590} = -2.4$ ,  $p = 0.017$ ).

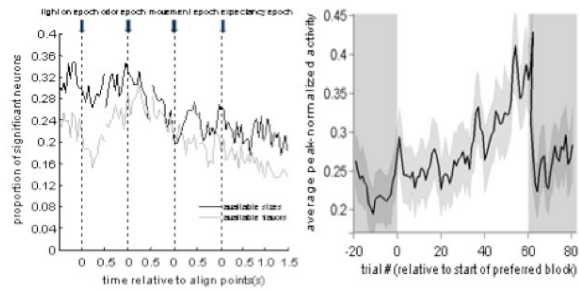
## OFC Lesions Disrupt Flavor-Specific Task Structure Encoding in VS

We expected OFC lesions to disrupt model-based task structure encoding in VS. Thus, we expected VS reward contingency signals and inference signals to be significantly diminished in OFC-lesioned rats compared to controls. In contrast to our predictions, reward configuration was encoded during the epoch immediately before odor delivery (the epoch with the largest proportion in controls) in a similar proportion of neurons in lesioned and control rats (Figure 18; size: in controls 121/399, in lesions 51/193,  $c^2 = 1.38$ ,  $p = 0.24$ ; flavor: in controls 84/399, in lesions 45/193,  $c^2 = 0.58$ ,  $p = 0.44$ ).

dependent on input from OFC, we expected flavor encoding, which is inherently model-based, to be diminished when OFC input was disrupted. Specifically, we expected

an overall reduction in the number of chocolate-selective

However, when we examined the magnitude of the size-configuration and flavor-



**Figure 18. Block-specific activity in VS of OFC-lesioned rats.**

**a) Reward contingency activity across trials.** Like in controls, VS of OFC-lesioned rats signals information about the current location of reward based on size (black) and flavor (gray).

**b) Reward contingency activity, collapsed across factor, across blocks.** Like in controls, VS of OFC-lesioned rats maintains and increases its signaling of reward contingencies across the 60 trials within each block.

configuration signals, we found some evidence that encoding of flavor configuration was slower to develop in lesioned rats. We performed a 3-factor ANOVA on average peak-normalized

firing rates among block-selective neurons during the same epoch as above, with group (sham or lesion), feature (size or flavor), and time within block (early, middle, or late) as factors. This analysis demonstrated a trend toward a 3-way interaction (Figure 19;  $F_{2,342} = 2.3$   $p =$

0.10). This interaction was driven by a significant interaction between group and time

within the block for flavor-configuration

encoding neurons

(planned comparison,

$F_{2,154} = 3.3$   $p < 0.05$ ),

indicating that lesions

caused flavor

encoding to develop

more slowly

compared to controls

but to reach a higher final level. One interpretation of this finding is that, early in the

block, encoding of available flavors (or response-flavor contingencies) was disrupted by

OFC lesions, but later in the block it was unaffected or even accentuated. The parallel

planned comparison for size-configuration encoding neurons showed no significant

interaction ( $F_{2,188} = 1.02$ ,  $p = 0.36$ ), suggesting that, with respect to reward contingencies,

OFC lesions specifically affected flavor encoding.

**Figure 19. Block-specific activity in VS, according to factor.**

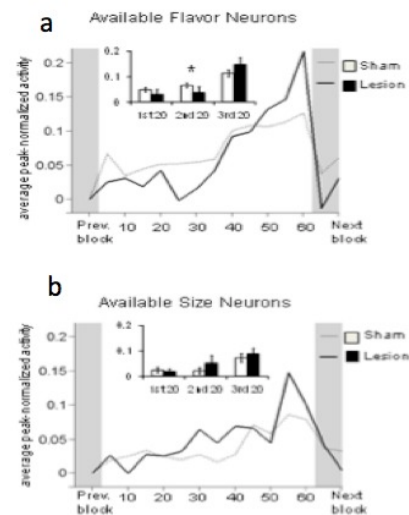
**a) Encoding of flavor contingencies across blocks in VS of control and OFC-lesioned rats.**

Flavor-selective cells showed a significant interaction of group and part of block (early, middle, or late ( $F_{2,154} = 3.3$ ,  $p < 0.05$ ).

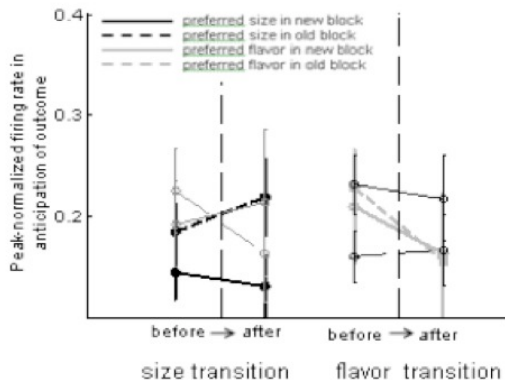
**b) Encoding of size contingencies across blocks in VS of control and OFC-lesioned rats.**

Size-selective cells showed no interaction between group and part of block ( $F_{2,188} = 1.02$ ,  $p = 0.36$ ).

i



Also consistent with our predictions, there was a significant difference in signaling of inference information in VS of control rats and rats with OFC lesions, as demonstrated by a 3-way interaction, with group (sham or lesion), preference point (old block or new block) and time (before switch or after switch) as factors ( $F_{1,242} = 6.8$   $p < 0.01$ ).



**Figure 20. Change in VS firing pattern OFC-lesioned rats during inference.** There was a 3-way interaction between group (control or lesioned), preference point (old block or new block), and time (before switch or after switch) ( $F_{1,242} = 6.8$   $p < 0.01$ ). Specifically, when OFC was lesioned, rats did not show significant inference signals in VS ( $F_{1,242} = 0.05$   $p = 0.82$ ).

Specifically, neurons in lesioned rats did not show any evidence of inference (Figure 20; planned comparison, preference direction by time interaction:  $F_{1,242} = 0.05$   $p = 0.82$ ), whereas, as previously mentioned, those in control rats did ( $F_{1,242} = 16.2$   $p < 0.0001$ ). Thus, OFC lesions seemed to abolish inference signals normally observed in VS. Control and lesioned rats did not differ in terms of inference signaling in VS in control conditions (defined as size transitions for flavor-selective neurons, and flavor transitions for size-

selective neurons) (planned comparison, preference direction by time interaction:  $F_{1,242} = 0.6$   $p = 0.43$ ), suggesting that the effect of lesions was specific for inferences of relevant information, rather than more generic changes in firing rates across any block transitions.

## SUMMARY

These experiments provide evidence for model-based and model-free signal integration in VS, as well as for a role of OFC in providing model-based information to VS. If VS signals were purely model-free, there would be no encoding of value-independent information like flavor or task structure. However, consistent with model-based signaling in VS, we identified chocolate-selective and vanilla-selective populations in VS, as well as populations signaling task structure properties (i.e. reward configuration and inferential information). If, on the other hand, VS signals were purely model-based, we would not observe differential encoding of value, or size of reward. Consistent with model-free signaling in VS, we observed differential reward size. Thus, VS activity appears to include both model-free and model-based signals.

The encoding of specific feature information and task structure elements that was identified in VS was disrupted when OFC input to VS was eliminated. However, model-free encoding of value information in VS remained intact in lesioned animals. Thus, OFC appears to contribute specifically to the model-based signals in VS. However, OFC-lesioned rats retained some VS encoding of flavor and configuration of rewards, suggesting that perhaps not all model-based information in VS is dependent on OFC input.

## CHAPTER 4: GENERAL DISCUSSION OF VALUATION PROCESSES

It is impossible to consider the complexity of decision making without first considering the complexity of value, the entity on which decision making is based. While relative value is simpler to deduce when options are presented in a common currency – for example, \$1 versus \$100 - there are an abundance of choices for which common currency comparison is not possible. Even when common currency comparisons can be made between options, there are often other factors influencing our overall assessment of those options. Such factors may be as abstract as justice, for which outcomes are not intuitively quantifiable. Indeed, fairness and reciprocity are pervasive influences of consumer behavior (Fehr and Gächter 2000), and, as noted earlier, the Ultimatum Game provides an example from Game Theory in which people tend to choose outcomes that do not provide the greatest dollar payout. Specifically, if the monetary offer made by the proposer is deemed unfair by the responder, the responder tends to reject the offer, subjecting both players to a fate of \$0 in winnings. Thus, the assessment of the value of the responders' options (i.e. accept or reject the offer) clearly incorporates more than the dollar amount of payout. One simple explanation for this so-called 'altruistic punishment' (Crockett, Clark et al. 2010) in the Ultimatum Game is that the injustice of the proposed bargain deducts a value amount greater than the value of the potential dollars earned on the deal, making rejection the most valuable option.

Because numerous factors influence the value of our options, our assessment of the current value of options often changes over time, as we become aware of additional relevant factors, or as the environment in which we make our choices changes. It is thus advantageous to have two systems supporting decision making: one that tracks the general value of our options and another that tracks specific information that can be used to calculate the present value of those options. Neural instantiation of such systems has been identified; it is generally believed that the striatum supports the faster, habitual behaviors of the model-free system, whereas the prefrontal cortex underlies the slower, goal-directed behaviors afforded by the model-based system (Daw, Niv et al. 2005).

Though it has been argued that model-free and model-based systems compete, at least at the level of ventral striatum (VS) (Yin, Mulcare et al. 2009), the nature of the interaction of these systems is not well understood.

In the preceding chapters, we have attempted to elucidate the physiological basis of these two systems and their potential integration. Consistent with prefrontal control of model-based signaling, we demonstrated that the orbitofrontal cortex (OFC) signals value-independent information about outcomes. Thus, its essential role in value-guided behaviors is likely due to the flexibility it contributes to value assessment. Further, we found VS signals information consistent with model-free signals. However, we also observed model-based signaling in VS, suggesting VS is a likely candidate for the integration of general value and specific feature information. VS signaling of model-based information, such as that related to abstract concepts like justice, is supported by evidence that activity in both OFC and VS is predictive of social choices that are independent of quantifiable gains and losses (Seo and Lee 2012). Though we found that OFC lesions disrupted model-based signaling in VS, lesions did not altogether eliminate such signaling. Thus, it is possible that VS incorporates model-based signals from multiple areas. Indeed, other areas, such as the hippocampus (van der Meer, Johnson et al. 2010), have demonstrated signals resembling model-based representations of outcomes.

Though we have these two systems to deal with the complex choices we encounter, our limited computational capacity renders us essentially unequipped to respond advantageously to all choice situations. The Monty Hall Problem provides an excellent example of how we are particularly inept in choices involving risk. In 1990, Marilyn vos Savant, a magazine columnist, was approached with a question about the best strategy to use on the game show *Let's Make A Deal* (vos Savant 1990). On the show, a player is told that there is a large prize behind one of three doors, and if he correctly chooses the door, he wins the prize. Once the player has chosen a door, the host opens one of the two remaining doors, revealing no prize behind that door. The host then gives the player the option to change his guess as to which door leads to the prize. When asked whether one

should at this point switch doors, Marilyn responded that it would be to the player's advantage to switch. Her response was met with harsh criticism from hundreds of mathematicians, until eventually, they realized Marilyn was right.

The explanation goes like this: when the player originally chooses a door, there is a  $1/3$  probability that he has chosen the door that will win him the prize and a  $2/3$  probability that he has not. The key to the problem is that the host will always open a door that does not lead to the prize. Thus, regardless of where the prize actually is, the host, by eliminating one of the two doors that does not lead to the prize, gives the player the opportunity to switch their bet from a  $1/3$  probability of winning the prize to a  $2/3$  probability of winning. Convincing ourselves of this counterintuitive concept requires slow, deliberate thinking. Though switching doors in this case is advantageous, our brains are not built to readily identify that advantage. Before the problem was solved, players reliably defaulted to their original choice – an example of what is known as the 'status quo bias.'

One assumption of several theories of model-free and model-based signal integration is that there is a neural trade-off in efficiency and accuracy (Daw, Niv et al. 2005). Consistent with this notion, one likely reason for our lack of aptitude in assessing probability is that we have evolved to prioritize efficiency over precision in our choice analysis. After all, throughout human history, we have dealt with the inundation of stimuli, the meaning of which we have learned through experience. In contrast, gambling-type choices with known risk probabilities represent modern choices without obvious prehistoric analogues.

It has been suggested that model-based signals may influence model-free signaling such that model-free signals evolve to reflect higher order contingency information (Doll, Jacobs et al. 2009). A mechanism of this type would be particularly efficient, as model-free valuation supporting habitual behavior does not seem to significantly involve cortical processing. This concept is supported by the observation that firing of midbrain dopamine (DA) neurons in ventral tegmental area (VTA) is time-locked to stimuli

predicting reward and to unexpected reward (Schultz 2002). Thus, if mechanisms underlying our choice behavior are built on principles of efficiency, maximizing habitual behavior is advantageous, as response time and expended energy are relatively low for habitual responding. Animal studies confirming that behavior is more likely to be dominated by the model-free habitual system after significant behavioral training (Adams and Dickinson 1981) indicate that behaviors that once required model-based signaling can eventually be maintained without depleting cortical resources. Based on this logic, and the importance of efficiency, it is possible that prefrontal areas like the OFC are initially engaged in new types of choice situations, but as rules governing outcomes can be deduced, behavioral control is surrendered to subcortical areas like striatum. Such rules, or heuristics, could explain behaviors that have been deemed by many as ‘irrational,’ such as choices that violate expected value and those demonstrating the framing effect.

Let’s re-consider our example of where expected value fails to predict behavior and imagine that for the sake of efficiency, we use a heuristic when faced with unfamiliar choices that involve potential gains and losses. Roughly speaking, the heuristic could be: guarantee gains and avoid guaranteed losses. People tend to choose a 100% chance of winning \$800 over an 85% chance of winning \$1000, even though the former has an expected value of \$800, while the latter has an expected value of \$850. Statistically speaking, choosing the \$800 is illogical if we are able to make this choice an infinite number of times. Indeed, in the long-run, choosing the 85% chance of winning \$1000 would make us richer. However, given just one shot, perhaps the rational choice is that which guarantees a gain. In the real world, choices often present themselves only once, and an efficient system likely employs general rules to maximize gains while expending minimal resources to navigate those choices. Consideration of the precise magnitude of gains or losses may reasonably occur after heuristic criteria are met, allowing us to reliably choose a 100% chance of \$850 over a 100% chance of \$800.

The framing effect demonstrated by Kahnman and Tversky too is far less surprising if we assume efficiency is a priority for our computationally challenged decision making

engine. Recall that given the choice between saving 200 people in Program 1, or in Program 2, having a 1/3 chance of saving 600 people and a 2/3 chance of saving no one, the people tend to choose Program 1. If, for the sake of efficiency, we ignore specific probabilities and simply evaluate these programs with the goal of guaranteeing gains (i.e. definitely saving lives), Program 1 is superior. If instead of a gain frame, we consider a loss frame, we choose between Program A, in which 400 people die and Program B, in which we have a 1/3 chance of no one dying and a 2/3 chance of 600 people dying. In this case, our goal may be to avoid guaranteed losses (i.e. definite deaths), which can only be accomplished through Program B. Accordingly, given these options, people tend to choose Program B. Thus, though upon calculation, Program 1 is equivalent to Program A, and Program 2 is equivalent to Program B, our application of a choice heuristic (i.e. choose definite gains and avoid definite losses), significantly enhances our decision making efficiency. Such heuristics, likely developed for efficiency, can explain why we appear to be ‘risk seeking with respect to losses’ and ‘risk averse with respect to gains’ (Kahnman and Tversky 1979).

Rather than changing our general attitude toward risk depending on our situation, we may simply use an efficient rule that leads us to choose guaranteed gains and choose against guaranteed losses. Signals directing choice behaviors consistent with such heuristics may be deployed by the model-free system, but in this proposed system, model-based signaling would likely contribute to the development of the rules. Thus, our ‘cognitive biases’ may be symptoms of an efficient system. Indeed, evidence of these biases has also been observed in animals. In one such demonstration, birds were given a choice between a tray with a fixed number of seeds and a tray with an amount that varied around the same mean. In warm weather, when the fixed amount was sufficient for maintaining a ‘positive energy budget,’ birds chose the tray with a fixed number of seeds. However, in cold weather, the fixed amount was not sufficient for a positive energy budget, and birds chose the risky tray (Caraco 1981). If we apply our proposed heuristics to these specific observations, we conclude that in warm weather, birds chose the definite gain, and in cold weather, they chose to avoid the definite loss. ‘Risk aversion’ has also been demonstrated in fish, birds, and bumblebees (Stephens and Krebs 1986, Kacelnik and

Bateson 1996), suggesting that, rather than an error in human logic, the effect may instead be a byproduct of a system attempting to reduce computational overhead. In other words, the effect may arise not because we come to irrational conclusions when thinking deeply about choices, but because we tend not to think deeply about all choices.

Interestingly, humans' tendency towards these short-cuts depends on the domain within which we are making our choices (Platt and Heutzel 2008). Because it is rare for people to demonstrate these 'risk seeking' and 'risk averse' effects in all domains of their lives, it is likely that degree of familiarity with the context of one's choices influences the likelihood that heuristics are utilized or if additional, model-based resources are recruited for the choice. Chip Heath, a psychologist, engineer, and business strategist, who teaches behavior and strategy courses at Stanford University, was recently quoted in the *Washington Post*, claiming that in his experience, people's instincts have led to good choices if those instincts were based on at least ten years of experience (Cunningham 2013). This reflection lends empirical credence to the idea that our habitual systems can learn to support quick behaviors that seem to reflect knowledge beyond the scope of traditional model-free systems. Experimental evidence too provides support for this theory.

Choosing a default option, or the status quo, as often observed in the Monty Hall problem, has been associated with increased activity in the VS (Yu, Mobbs et al. 2010). This finding was interpreted to mean that choosing the status quo is itself rewarding. However, it is perhaps more likely that VS activation during status quo choices reflects the habitual nature of those choices. It has further been shown that the frontal cortex is engaged when choices are made against the status quo and that, in this context, the frontal cortex specifically modulates activity in the basal ganglia (Fleming, Thomas et al. 2010). Thus, though the status quo may represent a choice not traditionally thought to be supported by model-free signaling, model-free signals may actually support such choices after significant experience and yet remain susceptible to the modulatory influence of the model-based system.

A recent theory regarding the interaction of model-free and model-based signals provides an explanation for why these systems may communicate, rather than act as separate, parallel systems (Doll, Simon et al. 2012). These researchers reason that model-based capacities likely evolved more recently to complement model-free signaling in choice execution. While activity in prefrontal cortex, striatum, and VTA are all associated with choice behavior (Pessiglione, Seymour et al. 2006, Jocham, Klein et al. 2011, Steinberg, Keiflin et al. 2013), we have previously provided evidence that the influence model-free signaling has over behavior is not altogether independent of model-based signaling (Takahashi, Roesch et al. 2009). Specifically, we have demonstrated that when OFC signals reward predictions, DA neurons in VTA do not signal the reward prediction errors (RPEs) that are sufficient for driving reward-seeking behavior (Steinberg, Keiflin et al. 2013). Thus, expectations signaled by the model-based system are incorporated in the value assessment performed by the model-free system.

Though important components of decision making have been elucidated, several lines of research regarding the physiological basis of valuation remain largely untapped. Particularly fruitful will likely be further investigation into the mechanisms by which DA may modulate model-free and model-based signals. Both the VS, which appears to support both model-free and model-based systems, and the OFC, which supports the model-based system, are dopaminergically innervated (Alexander, DeLong et al. 1986), and DA neurons of VTA that signal RPEs communicate other information that is not normally available to the model-free system (Bromberg-Martin, Matsumoto et al. 2010). It is therefore possible that there are no pure model-free signals in the brain and that any physiological representation of ‘value’ is thus inherently integrative. Experience-driven, habitual choices may therefore result from a model-free system influenced by model-based signals and hence reflect model-based representations of outcomes more than originally supposed. Additionally, DA cells outside of VTA may make distinct contributions to decision making. For example, DA cells of substantia nigra pars compacta do not signal RPEs and thus may support another mechanism of valuation, such as model-based prediction errors (Daw, Gershman et al. 2011). Prediction errors of this sort could potentially provide a mechanism for model-based supervision of model-free

control. According to this theory, model-free signals would efficiently direct the bulk of our behavior, but the slower, model-based system would maintain the ability to take over behavioral control.

- Adams, C. and A. Dickinson (1981). "Instrumental responding following reinforcer devaluation." Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology **33(B)**: 109-121.
- Alexander, G. E., et al. (1986). "Parallel organization of functional segregated circuits linking basal ganglia and cortex." Annual Review of Neuroscience **9**: 357-381.
- Balleine, B. W., et al. (2008). Multiple forms of value learning and the function of dopamine. Neuroeconomics: Decision Making and the Brain. P. W. Glimcher, C. F. Camerer, E. Fehr and R. A. Poldrack. Amsterdam, Elsevier: 367-388.
- Berendse, H. W., et al. (1992). "Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat." Journal of Comparative Neurology **316**: 314-347.
- Bernoulli, D. (1738). "Exposition of a new theory on the measurement of risk." Econometrica **22(1)**: 22-36.
- Berridge, K. C. (2001). Reward learning: reinforcement, incentives, and expectations. The Psychology of Learning and Motivation. D. L. Medin. New York, Academic Press.
- Bissonette, G. B., et al. (2008). "Double dissociation of the effects of medial and orbital prefrontal cortical lesions on attentional and affective shifts in mice." Journal of Neuroscience **28**: 11124-11130.
- Brog, J. S., et al. (1993). "The patterns of afferent innervation of the core and shell in the "accumbens" part of the rat ventral striatum: immunohistochemical detection of retrogradely transported fluoro-gold." Journal of Comparative Neurology **338**: 255-278.
- Bromberg-Martin, E. S., et al. (2010). "Dopamine in motivational control: rewarding, aversive and alerting." Neuron **68**: 815-834.
- Burke, K. A., et al. (2008). "The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards." Nature **454**: 340-344.
- Burke, K. A., et al. (2009). "Orbitofrontal inactivation impairs reversal of Pavlovian learning by interfering with 'disinhibition' of responding for previously unrewarded cues." European Journal of Neuroscience **30**: 1941-1946.

Camille, N., et al. (2004). "The involvement of the orbitofrontal cortex in the experience of regret." Science **304**: 1168-1170.

Camille, N., et al. (2011). "Ventromedial frontal lobe damage disrupts value maximization in humans." Journal of Neuroscience **31**: 7527-7532.

Caraco, T. (1981). "Energy budgets, risk and foraging preferences in dark-eyed juncos." Behavioral Ecology & Sociobiology **8**: 213-217.

Cardinal, R. N., et al. (2002). "Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex." Neuroscience and Biobehavioral Reviews **26**: 321-352.

Clarke, H. F., et al. (2008). "Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex." Journal of Neuroscience **28**: 10972-10982.

Corbit, L. and B. Balleine (2011). "The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell." Journal of Neuroscience **31**(33): 11786-11794.

Crockett, M., et al. (2010). "Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion." Emotion **10**(6): 855-862.

Cunningham, L. (2013). "Decision making for the indecisive." The Washington Post On Leadership.

Dahl, R. (2001). "Affect regulation, brain development, and behavioral/emotional health in adolescence." CNS Spectrums **6**(1): 60-72.

Damasio, A. R. (1994). Descartes Error. New York, Putnam.

Damasio, A. R., et al. (1994). "The return of Phineas Gage: clues about the brain from the skull of a famous patient." Science **264**: 1102-1105.

Daw, N., et al. (2005). "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control." Nature Neuroscience **8**(12): 1704-1711.

- Daw, N. D., et al. (2011). "Model-based influences on humans' choices and striatal prediction errors." Neuron **69**: 1204-1215.
- Day, J. J., et al. (2007). "Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens." Nature Neuroscience **10**: 1020-1028.
- Delamater, A. R. (2007). "The role of the orbitofrontal cortex in sensory-specific encoding of associations in pavlovian and instrumental conditioning." Annals of the New York Academy of Science **1121**: 152-173.
- Dias, R., et al. (1996). "Dissociation in prefrontal cortex of affective and attentional shifts." Nature **380**: 69-72.
- Doll, B., et al. (2009). "Instructional control of reinforcement learning: A behavioral and neurocomputational investigation." Brain Research **1299**: 74-94.
- Doll, B., et al. (2012). "The ubiquity of model-based reinforcement learning." Current Opinion in Neurobiology **22**: 1075-1081.
- Elliott, R., et al. (2010). "Hedonic and informational functions of the human orbitofrontal cortex." Cerebral Cortex **20**(1): 198-204.
- Everitt, B. J., et al. (1991). "The basolateral amygdala-ventral striatal system and conditioned place preference: further evidence of limbic-striatal interactions underlying reward-related processes." Neuroscience **42**: 1-18.
- Fehr, E. and G. Gächter (2000). "Fairness and retaliation: The economics of reciprocity." Journal of Economic Perspectives **14**(3): 159-181.
- Feierstein, C. E., et al. (2006). "Representation of spatial goals in rat orbitofrontal cortex." Neuron **51**: 495-507.
- Fellows, L. (2007). "The role of orbitofrontal cortex in decision making: a component process account." Annals of the New York Academy of Sciences **1121**: 421-430.

Fellows, L. (2011). "Orbitofrontal contributions to value-based decision making: Evidence from humans with frontal lobe damage." Annals of the New York Academy of Sciences **1239**: 51-58.

Fellows, L. K. and M. J. Farah (2003). "Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm." Brain **126**: 1830-1837.

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

FitzGerald, T., et al. (2009). "The role of human orbitofrontal cortex in value comparison for incommensurable objects." Journal of Neuroscience **29**(26): 8388-8395.

Fleming, S., et al. (2010). "Overcoming status quo bias in the human brain." Proceedings of the National Academy of Sciences **107**(13): 6005-6009.

Gallagher, M., et al. (1999). "Orbitofrontal cortex and representation of incentive value in associative learning." Journal of Neuroscience **19**: 6610-6614.

Gallagher, M. and G. Schoenbaum (1999). "Functions of the amygdala and related forebrain areas in attention and cognition." Annals of the New York Academy of Sciences **877**: 397-411.

Garrison, J., et al. (2013). "Prediction error in reinforcement learning: A meta-analysis of neuroimaging." Neuroscience & Biobehavioral Reviews **37**(7): 1297-1310.

Giordano, J., et al. (2012). "Neuroeconomics. An emerging field of theory and practice." The European Business Review.

Glascher, J., et al. (2010). "States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning." Neuron **66**: 585-595.

Goldman-Rakic, P. S., et al. (1992). "The anatomy of dopamine in monkey and human prefrontal cortex." Journal of Neural Transmission Supplements**36**: 163-177.

- Goldstein, B., et al. (2012). "Ventral striatum encodes past and predicted value independent of motor contingencies." Journal of Neuroscience **32**(6): 2027-2036.
- Gottfried, J. A., et al. (2003). "Encoding predictive reward value in human amygdala and orbitofrontal cortex." Science **301**: 1104-1107.
- Groenewegen, H. J., et al. (1990). "The anatomical relationship of the prefrontal cortex with the striatopallidal system, the thalamus and the amygdala: evidence for a parallel organization." Progress in Brain Research **85**: 95-118.
- Guth, et al. (1982). "An experimental analysis of ultimatum bargaining." Journal of Economic Behavior and Organization **3**(4): 367-388.
- Haber, S. N., et al. (1990). "Topographic organization of the ventral striatal efferent projections in the rhesus monkey: an anterograde tracing study." Journal of Comparative Neurology **293**: 282-298.
- Hare, T. A., et al. (2008). "Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors." Journal of Neuroscience **28**: 5623-5630.
- Hartley, R. and L. Farrell (2002). "Can expected utility theory explain gambling?" The American Economic Review **92**(3): 613-624.
- Heimer, L., et al. (1995). Basal Ganglia. The Rat Nervous System. G. Paxinos. San Diego, Academic Press: 579-628.
- Ito, R., et al. (2004). "Differential control over cocaine-seeking behavior by nucleus accumbens core and shell." Nature Neuroscience **7**: 389-397.
- Izquierdo, A. D. and E. A. Murray (2000). "Bilateral orbital prefrontal cortex lesions disrupt reinforcer devaluation effects in rhesus monkeys." Society for Neuroscience Abstracts **26**: 978.
- Izquierdo, A. D. and E. A. Murray (2004). "Combined unilateral lesions of the amygdala and orbital prefrontal cortex impair affective processing in rhesus monkeys." Journal of Neurophysiology **91**: 2033-2039.

Jocham, G., et al. (2011). "Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices." Journal of Neuroscience **31**(5): 1606-1613.

Jones, J. L., et al. (2012). "Orbitofrontal cortex supports behavior and learning using inferred but not cached values." Science **338**: 953-956.

Kacelnik, A. and M. Bateson (1996). "Risky theories - the effects of variance on foraging decisions." American Zoologist **36**: 402-436.

Kahneman, D. and A. Tversky (1984). "Choices, values, and frames." American Psychologist **39**: 341-350.

Kahnman, D. and A. Tversky (1979). "Prospect theory: an analysis of decision under risk." Econometrica **47**(2): 263-292.

Kahnt, T., et al. (2010). "The neural code of reward anticipation in human orbitofrontal cortex." Proceedings of the National Academy of Science **107**: 6010-6005.

Kamin, L. J. (1969). Selective association and conditioning. Fundamental issues in associative learning. N. J. Mackintosh and W. K. Honig. Halifax, Dalhousie University Press: 42-64.

Li, J., et al. (2006). "Policy adjustment in dynamic economic game." PLoS **1**(1): e103.

Machado, C. J. and J. Bachevalier (2007). "The effects of selective amygdala, orbital frontal cortex or hippocampal formation lesions on reward assessment in nonhuman primates." European Journal of Neuroscience **25**: 2885-2904.

Mainen, Z. F. and A. Kepecs (2009). "Neural representation of behavioral outcomes in the orbitofrontal cortex." Current Opinion in Neurobiology **19**: 84-91.

McDannald, M., et al. (2012). "Model-based learning and the contribution of the orbitofrontal cortex to the mode-free world." European Journal of Neuroscience **35**(7): 991-996.

- McDannald, M. A., et al. (2011). "Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning." Journal of Neuroscience **31**: 2700-2705.
- McDannald, M. A., et al. (2005). "Lesions of orbitofrontal cortex impair rats' differential outcome expectancy learning but not conditioned stimulus-potentiated feeding." Journal of Neuroscience **25**: 4626-4632.
- Mogenson, G. J., et al. (1980). "From motivation to action: functional interface between the limbic system and the motor system." Progress in Neurobiology **14**: 69-97.
- O'Doherty, J., et al. (2004). "Dissociable roles of ventral and dorsal striatum in instrumental conditioning." Science **304**: 452-454.
- O'Doherty, J., et al. (2001). "Abstract reward and punishment representations in the human orbitofrontal cortex." Nature Neuroscience **4**: 95-102.
- O'Doherty, J., et al. (2000). "Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex." Neuroreport **11**: 893-897.
- O'Neill, M. and W. Schultz (2010). "Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value." Neuron **68**: 789-800.
- Ogawa, M., et al. (2013). "Risk-responsive orbitofrontal neurons track acquired salience." Neuron **77**: 251-258.
- Ostlund, S. B. and B. W. Balleine (2007). "Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental learning." Journal of Neuroscience **27**: 4819-4825.
- Padoa-Schioppa, C. and J. A. Assad (2006). "Neurons in orbitofrontal cortex encode economic value." Nature **441**: 223-226.
- Parkinson, J. A., et al. (2005). "Acquisition of instrumental conditioned reinforcement is resistant to the devaluation of the unconditioned stimulus." Quarterly Journal of Experimental Psychology **58**: 19-30.

Pears, A., et al. (2003). "Lesions of the orbitofrontal but not medial prefrontal cortex disrupt conditioned reinforcement in primates." Journal of Neuroscience **23**: 11189-11201.

Pessiglione, M., et al. (2006). "Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans." Nature **442**: 1042-1045.

Pfeifer, J. and N. Allen (2012). "Arrested development? Reconsidering dual-systems models of brain function in adolescence and disorders." Trends in Cognitive Sciences **16**(6): 322-329.

Pickens, C. L., et al. (2005). "Orbitofrontal lesions impair use of cue-outcome associations in a devaluation task." Behavioral Neuroscience **119**: 317-322.

Pickens, C. L., et al. (2003). "Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task." Journal of Neuroscience **23**: 11078-11084.

Plassman, H., et al. (2010). "Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making." Journal of Neuroscience **30**: 10799-10808.

Plassmann, H., et al. (2007). "Orbitofrontal cortex encodes willingness to pay in everyday economic transactions." Journal of Neuroscience **27**: 9984-9988.

Platt, M. and S. Heutzel (2008). "Risky business: The neuroeconomics of decision making under uncertainty." Nature Neuroscience **11**(4): 398-403.

Riceberg, J. S. and M. L. Shapiro (in press). "Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior." Journal of Neuroscience.

Roesch, M. R., et al. (2007). "Should I stay or should I go? Transformation of time-discounted rewards in orbitofrontal cortex and associated brain circuits." Annals of the New York Academy of Sciences **1104**: 21-34.

Roesch, M. R. and C. R. Olson (2004). "Neuronal activity related to reward value and motivation in primate frontal cortex." Science **304**: 307-310.

Roesch, M. R. and C. R. Olson (2005). "Neuronal activity in primate orbitofrontal cortex reflects the value of time." Journal of Neurophysiology **94**: 2457-2471.

Roesch, M. R. and G. Schoenbaum (2006). From associations to expectancies: orbitofrontal cortex as gateway between the limbic system and representational memory. The Orbitofrontal Cortex. D. H. Zald and S. L. Rausch. London, Oxford University Press.

Roesch, M. R., et al. (2009). "Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards." Journal of Neuroscience **29**: 13365-13376.

Roesch, M. R., et al. (2006). "Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation." Neuron **51**: 509-520.

Rolls, E. T. and F. Grabenhorst (2008). "The orbitofrontal cortex and beyond: From affect to decision-making." Progress in Neurobiology **86**: 216-244.

Schoenbaum, G., et al. (1998). "Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning." Nature Neuroscience **1**: 155-159.

Schoenbaum, G., et al. (1999). "Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning." Journal of Neuroscience **19**: 1876-1884.

Schoenbaum, G., et al. (2009). "A new perspective on the role of the orbitofrontal cortex in adaptive behaviour." Nature Reviews Neuroscience **10**: 885-892.

Schoenbaum, G., et al. (2003). "Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals." Learning and Memory **10**: 129-140.

Schoenbaum, G., et al. (2003). "A systems approach to orbitofrontal cortex function: recordings in rat orbitofrontal cortex reveal interactions with different learning systems." Behavioral Brain Research **146**: 19-29.

Schultz, W. (2002). "Getting formal with dopamine." Neuron **36**(2): 241-263.

Schultz, W., et al. (1992). "Neuronal activity in monkey ventral striatum related to the expectation of reward." Journal of Neuroscience **12**: 4595-4610.

Seo, H. and D. Lee (2012). "Neural basis of learning and preference during social decision-making." Current Opinion in Neurobiology **22**: 1-6.

Setlow, B., et al. (2002). "Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive Pavlovian second-order conditioned responses." Behavioral Neuroscience **116**: 267-275.

Simmons, J. M., et al. (2007). "A comparison of reward-contingent neuronal activity in monkey orbitofrontal cortex and ventral striatum: guiding actions toward rewards." Annals of the New York Academy of Science **1121**: 376-394.

Simon, D. A. and N. D. Daw (2011). "Neural correlates of forward planning in a spatial decision task in humans." Journal of Neuroscience **31**: 5526-5539.

Singh, T., et al. (2010). "An essential role for the nucleus accumbens core in behaviors guided by outcome expectancies." Society for Neuroscience Abstracts.

Steinberg, E. E., et al. (2013). "A causal link between prediction errors, dopamine neurons and learning." Nature Neuroscience **16**: 966-973

Steiner, A. P. and A. D. Redish (2012). "The road not taken: neural correlates of decision making in orbitofrontal cortex." Frontiers in Neuroscience **6**: 131.

Stephens, D. and J. Krebs (1986). "Foraging theory."

Sutton, R. S. and A. G. Barto (1990). Time-derivative models of Pavlovian reinforcement. Learning and Computational Neuroscience: Foundations of Adaptive Networks. M. Gabriel and J. Moore. Boston, MIT Press: 497-537.

Takahashi, Y., et al. (2009). "The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes." Neuron **62**: 269-280.

Takahashi, Y., et al. (2008). "Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model." Frontiers in Neuroscience **2**: 86-99.

Teitelbaum, H. (1964). "A comparison of effects of orbitofrontal and hippocampal lesions upon discrimination learning and reversal in the cat." Experimental Neurology **9**: 452-462.

Thorndike, E. L. (1898). "Animal intelligence: an experimental study of the associative processes in animals." Psychological Review **2**: 1-107.

Thorpe, S. J., et al. (1983). "The orbitofrontal cortex: neuronal activity in the behaving monkey." Experimental Brain Research **49**: 93-115.

Tolman, E. C. (1949). "There is more than one kind of learning." Psychological Review **56**: 144-155.

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

Valentin, V. V., et al. (2007). "Determining the neural substrates of goal-directed learning in the human brain." Journal of Neuroscience **27**: 4019-4026.

van der Meer, M., et al. (2010). "Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task." Neuron **67**: 25-32.

van der Meer, M. and A. Redish (2009). "Covert expectation-of-reward in rat ventral striatum at decision points." Frontiers in Integrative Neuroscience **3**(1): 1-15.

van der Meer, M. A. and A. D. Redish (2009). "Covert expectation of reward in rat ventral striatum at decision points." Frontiers in Integrative Neuroscience **3**: Epub 2009 Feb 2005.

vos Savant, M. (1990). "Ask Marilyn." Parade Magazine **16**.

Walton, M. E., et al. (2010). "Separable learning systems in the macaque brain and the role of the orbitofrontal cortex in contingent learning." Neuron **65**: 927-939.

West, E. A., et al. (2011). "Transient inactivation of orbitofrontal cortex blocks reinforcer devaluation in macaques." Journal of Neuroscience **31**: 15128-15135.

Wheeler, E. Z. and L. K. Fellows (2008). "The human ventromedial frontal lobe is critical for learning from negative feedback." Brain **131**: 1323-1331.

Yin, H. H., et al. (2009). "Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill." Nature Neuroscience **12**: 333-341.

Yu, R., et al. (2010). "Insula and striatum mediate the default bias." Journal of Neuroscience **30**(44): 14702-14707.