Neural Representation of Economic Parameters by Dopamine Release: a Quantitative Analysis

Andrew S. Hart

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Paul Phillips, Chair

David Perkel

Gregory Horwitz

Program Authorized to Offer Degree:

Neurobiology and Behavior

University of Washington

Abstract

Neural Representation of Economic Parameters by Dopamine Release:

a Quantitative Analysis

Andrew S. Hart

Chair of the Supervisory Committee:

Associate Professor Paul Phillips

Psychiatry and Behavioral Sciences

The goal of the field of neuroeconomics is to understand the neural representations of variables relevant to theories of decision-making. These variables include the expected value of a prospect and the uncertainty associated with a choice. Recordings from dopamine neurons in awake, behaving monkeys strongly indicate that dopamine plays a role in representing these parameters. A complementary role of dopamine is the encoding of a reward prediction error signal, which can be used to iteratively update the expected value associated with a cue that predicts reward. While these roles of dopamine neurons have been qualitatively described, there are very few quantitative analyses that relate dopamine to these variables, and none have been conducted at the level of dopamine release. In the current thesis, I combined simple Pavlovian and operant behavioral tasks with recordings of dopamine release in the nucleus accumbens core of rats, using the electrochemical method fast-scan cyclic voltammetry. To address the relationship between dopamine release and expected value and uncertainty, I applied regression analyses to

conditioned stimulus-evoked dopamine signals during learning. To address the relationship between dopamine release and reward prediction error signaling, I applied an axiomatic reward prediction error model to reward-evoked dopamine release from rats performing an operant task with probabilistic rewards. I report that dopamine release correlates with expected value, uncertainty, and reward prediction error signals. Expected value and uncertainty signals may be blended at the level of dopamine release, and reward prediction error signals can be represented in a balanced or an imbalanced manner depending on the range of errors studied and the behavioral task.

Table of Contents

List of Figures

List of Tables

Table Page

Acknowledgements

I would like to thank Paul Phillips, Mark Walton, Jeremy Clark, Robb Rutledge, and Paul Glimcher for their help designing experiments and analyses and editing manuscripts.

I would like to thank Scott Ng-Evans and Christina Akers for extensive technical support.

I would like to thank my parents, John and Ann, and my siblings Jon-Paul and Jamie for their patience and support.

I would like to thank all of the friends I have met while in graduate school, especially in the Phillips lab for their tolerance and understanding.

Chapter 1:

Introduction*

*The majority of this chapter was originally published online in the article "Neuroeconomics" by Andrew S Hart, Mark E Walton, and Paul E M Philips, in *Encyclopedia of Life Sciences (ELS)* John Wiley & Sons, LTD: Chichester, UK, 2008. MEW and PEMP edited the manuscript.

Abstract

Guided by the common drive to develop normative and descriptive theories of decision-making behavior, experimental psychology and economics have existed for decades as parallel fields. In spite of their common motivations, the two communities rarely entered into a dialog with each other or exchanged methods and ideas. This barrier was created by conceptual differences between the fields. Economists assumed that they had no access to internal states, so they formulated their theories based on behaviors, either observed or idealized. Psychologists, on the other hand, tended to put a great deal of emphasis on investigating the internal states that might govern the relationships between environmental factors and behaviors. The advent of cognitive neuroscience with its noninvasive techniques in human subjects, particularly functional magnetic resonance imaging (fMRI), which allows the recording of human brain activation with good spatial and temporal resolution, has finally placed researchers on both sides of the divide one step closer to the goal of observing internal states. These new techniques have inspired a dialog among economists, psychologists, and neuroscientists that is set to bring new developments in all three areas. This confluence of ideas across disciplines has been dubbed neuroeconomics.

In this chapter, I describe the field of neuroeconomics and its attempt to understand the neural basis of decision-making. I later emphasize the hypothesized role of dopamine in encoding economically relevant variables, such as expected value, uncertainty, and reward prediction errors, and finally give a general description of in vivo voltammetry, the technique exclusively used in this work to record dopamine release in the brain.

Approaches to Decision-making

Economists historically sought out to describe the actions of *Homo economicus*, an idyllic and completely rational human decision maker, whose goal in every action is to maximize value. Value-based decision-making requires a systematic evaluation of costs and benefits of actions available to us, followed by the selection of an appropriate action to complete the choice (Rangel et al. 2008). Value can be considered an objective quantity in the environment, such as 3 oranges or $2, but economists quickly discovered that economic behaviors rarely follow the heuristic of value maximization, and humans are often irrational decision makers, relying heavily on social factors and affect. For example, someone might prefer 3 oranges to $1, but, paradoxically, would prefer $10 to 30 oranges. Economists accounted for this deviation in behavior by introducing the concept of utility, the subjective experience of value within the environment (von Neumann and Morgenstern 1944). The exact relationship between value and utility is unknown, but most agree that it is nonlinear, and as total value increases, the increase in utility for each additional unit of value diminishes (von Neumann and Morgenstern 1944, Kahneman and Tversky 1979). It should be noted that the behaviors that interest economists are not necessarily financial decisions, and in many cases, they don't involve money or any other form of currency. A decision to spend 2 hours at the pub rather than 2 hours exercising is just as valid and interesting an economic problem as a decision to buy an expensive insurance policy for protection against an unlikely loss.

Using the concept of utility, Kahneman and Tversky (1979) formalized prospect theory to describe how humans respond to risk. Prospect theory states that when faced with several options, a human will pick the one that will yield the maximum positive change in utility (or

minimum loss) after all costs and benefits have been considered. Kahneman and Tversky (1979) observed decision-making behavior in human subjects to generate utility curves. They found that the shapes of the curves showed that their subjects were averse to losses as well as risky gains, often choosing smaller deterministic rewards over larger probabilistic rewards. The use of behavioral experiments as a method in economics was perhaps an early sign of the convergence of economics and psychology, as the behavioral economic practices of offering real incentives to participants and avoiding deception have now been adopted as the standard operating procedure in neuroeconomics (Sugrue et al. 2005).

While using behavioral techniques has led to a great deal of progress in economics, economists often remained skeptical that they could use such behaviors to infer the thoughts and feelings that make up the process of choosing. By contrast, experimental psychology and neuroscience have a long history of developing theories of decision-making that include internal states. Psychological theories also tended to include motivational, social and affective components that could account for deviations from rationality. However, until recently, hypotheses that involved internal states in decision-making could not be tested directly by looking at brain activity but instead had to be assessed indirectly from measures such as reaction times and patterns of performance.

Recent developments have changed the nature of the fields of economics, psychology, and neuroscience. Using model-based approaches, neuroscientists have been able to test computational models of processes such as learning and decision-making by directly relating them to data from brain imaging and electrophysiological experiments. These types of experiments lend themselves to the use of more sophisticated decision-making tasks that approach the complexity of those used in economics. This line of work serves to benefit both

neuroscience and economics by using neural data and complex choice tasks to develop better computational theories of decision-making. These theories often include internal processes such as the evaluation of options available in the environment, the representation of anticipated outcomes of actions, action selection, and the evaluation of outcomes (Rangel et al. 2008).

Neurobiology of Rewards

In addition to economic decisions in humans within financial systems, it appears likely that all animal species engage in very similar decision-making processes as they forage for primary rewards such as food, shelter and sex. Thus, the first clues in neuroscience as to the representation of value in the brain have come from the studies of reward that combined Pavlovian conditioning with neural recordings in animals (Phillips and Olds 1969). Early work suggested that in rodents, neurons in the midbrain increase their rate of firing in response to classically conditioned cues that reliably predict food and water rewards (Phillips and Olds 1969). These responses were found to be independent of the physical properties of the stimuli (such as tone frequency) and modulated by the motivational state of the animal. That is, neurons in hungry rats responded more to tones that predicted food than to tones that predicted water, and vice versa for neurons in thirsty rats (Phillips and Olds 1969).

More than 30 years later, Schultz and colleagues were able to show that dopaminergic neurons in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) of the midbrain show similar cue-evoked changes in activity in primates (Ljungberg et al. 1992). While initially these neurons show brief bursts of activity to unexpected fluid reward in thirsty animals, after prolonged training these responses occur at the presentation of the first cue

predictive of a reward rather than the reward itself (Schultz et al. 1997). The signals recorded in dopamine neurons are akin to error signals present in a popular computational model of learning, the temporal difference model (Montague et al. 1996, Schultz et al., 1997, Waelti et al. 2001). This is one of the first examples of how a normative model could be applied to neural data. Moreover, these changes in firing rate scale with both reward magnitude and reward probability, suggesting that the change in firing rate could represent the expected utility of anticipated rewards (Fiorillo et al. 2003, Tobler et al. 2005).

Subsequently, similar findings have come from fMRI studies in humans, with the dopaminergic midbrain and several of its primary target regions, such the ventral striatum and the frontal lobe being activated first by unexpected juice rewards and then by conditioned cues during classical conditioning tasks similar to those used with Schultz's monkeys (Berns et al. 2001, O'Doherty et al. 2003). A similar network of regions also appears activated by non-primary rewards such as money and, as for the juice, the activity scales with the amount of money that is expected and its likelihood of being received (Tobler, O'Doherty et al. 2007). More recently, it has been shown that the size of the signal in response to money is also influenced by personality factors such as introversion (Fujiwara et al. 2008) and the amount of monetary assets that a person possesses (Tobler et al. 2007). Moreover, in humans, neural responses to pain and to financial loss also seem to involve overlapping circuits (Paulus et al. 2003). These results demonstrate a convergence of findings in human and nonhuman models of reward and suggest that the neural representation of primary rewards might be common across species and these might largely overlap with the regions involved in evaluating monetary value in humans.

Neurobiology of Decision-making

Research on the neural representation of value has been extended from simple learning paradigms into the realm of choices with multiple alternatives in both animals and humans. Bergman and colleagues recorded from midbrain dopamine neurons while monkeys were performing a two-armed bandit task where the animals can choose between two options that differ in the likelihood of a reward being delivered (Morris et al. 2006). These choice trials were intermingled with reference ones in which only one option was presented. The neural activity on these reference trials proved to be a better predictor of the monkeys' decision-making behavior on choice trials than the actual rewards received in these trials, and indeed, the responses on choice trials scaled with the values of the monkeys' ultimate choices. These results suggest that dopamine neurons carry a representation of the value of future actions, a critical component in any model of decision-making (Morris et al. 2006).

Other studies on nonhuman primates have implicated important roles for neurons in a variety of cortical regions in aspects of decision-making. For example, Platt and Glimcher (1999) discovered neurons in the lateral intraparietal area (LIP) of cortex that fire at a higher frequency when the monkeys are about to make eye movements that would earn a larger proportion of the total available reward (Platt and Glimcher 1999). Over a range of decisions, this change in activity is correlated to the ratio of the chosen reward magnitude to the total reward available in each trial.

Most of these experiments in animals have varied either the magnitude or likelihood of a single type of juice or food reward. However, many everyday decisions involve an assessment of disparate and abstract potential benefits. Several lines of evidence have indicated that

prefrontal cortex might be important for assessing the relative value of such outcomes. For instance, neurons in the orbitofrontal cortex (OFC), a subregion of the frontal lobe, have been shown to encode the value of available rewards based on their subjective preference (Padoa-Schioppa and Assad 2006). This representation of value is independent of other sensory, motoric or mnemonic processes, suggesting that the brain contains a representation of the value of what is available separate to the means to obtain these items. Moreover, the valuation as coded by the neurons of OFC is independent of the other available options, i.e., whether a particular item is the best or worst available (Padoa-Schioppa and Assad 2006). This finding could have strong implications for economic theory, as economists have long struggled with the concept of transitive preferences (Gilboa 2009). That is, if someone prefers an orange to 3 bananas, and they prefer 3 bananas to 4 apples, then they should prefer an orange over 4 apples. However, no behaviorally derived theory in economics can prove transitive preferences, as subjects often demonstrate paradoxical behavior, such as choosing 4 apples over an orange in this example. The above findings provide economists with a putative neural representation of economic value that could be used to interpolate between goods in the environment.

If OFC neurons reliably represent the value of economic goods, then the degree OFC activation in response to a novel good could be used to accurately predict preferences in hypothetical decisions. For example, if OFC activation is higher in response to a pineapple than to an orange, then we could say that one prefers a pineapple to an orange, even if he or she has never been presented with the choice. Preferences are often revealed through purchasing behavior, so knowing whether one might prefer a pineapple to an orange could translate to knowing that he or she will pay more for a pineapple than for an orange.

When studied with fMRI, humans show similar contributions of networks of subcortical and cortical regions when making financial decisions between different goods. As with the work in animals, this has demonstrated that there is not a single unitary system for assessing value in the brain, but instead several, possibly competing, systems representing different aspects of value. For instance, Knutson and colleagues have studied a variety of purchasing scenarios and have shown that activity in the ventral striatum during the presentation of products available for purchase is a predictor of future preference (Knutson and Bossaerts 2007) and, in a financial investment task, of making a risky investment (Kuhnen and Knutson 2005, Knutson et al. 2008). Other regions, such as the insula and medial prefrontal cortex, also reflect other aspects of investing, purchasing and selling, such as when prices seem particularly high or low (Knutson et al. 2007) or when subjects are likely to make safe investments (Kuhnen and Knutson 2005). Intriguingly, the activation of these brain regions is often a better predictor of purchases than the subjects' own reported preferences. Together these results suggest that people's preferences and financial deliberations can be tracked in their brains, and that this information may provide important insights into how such purchase decisions are made.

Representing Costs

The study by Knutson and colleagues highlights an important feature of decision-making that is not to be overlooked: the representation of costs. A cost is a mechanism to obtain an outcome that alters the net utility of the overall transaction, including delays in reward, the physical effort or persistence that is required to gain the reward, and of course, paying money. The studies of financial loss and purchasing mentioned above have implicated anterior cingulate

cortex and insula as representing financial losses (Kuhnen and Knutson 2005) and excessive prices, respectively (Knutson et al. 2007).

Studies in which subjects choose between options which differ in the delay to the reward suggest that the evaluation of immediate outcomes versus distant ones may engage partially dissociable neural systems. The limbic system, including dopamine neurons, and their cortical targets are activated more strongly when subjects choose immediately available rewards over delayed rewards and show diminishing activation when delayed rewards are chosen (McClure et al. 2004). The limbic system is considered a primitive part of the brain that is primarily engaged with satisfying basic needs and reproduction. In contrast, both immediate and delayed rewards activate lateral frontal cortical regions (McClure et al. 2004). These regions are generally engaged with higher level cognitive functions, and indeed they are more strongly activated when subjects are faced with more difficult choices between immediate and delayed rewards (McClure et al. 2004). Thus it appears that human decisions involving delayed gratification engage multiple neural systems. One system strongly discounts delayed rewards while the other does not discount delayed rewards. This finding invites the hypothesis that when making decisions, humans are often using multiple, competing systems, and that behaviors might depend on which system wins out over the other. Similar conclusions can be drawn from a subfield of neuroeconomics that deals with social decision-making.

Social Decision-making

With increasingly complex behavioral tasks coming into use in neuroscience, much work has been done in the field of social decision-making. This branch of neuroeconomics uses head-

to-head behavioral tasks in which the outcomes depend not only on the subjects' decisions but also on the decisions of another subject, a confederate, or a computer, and often incorporates game theory from economics for the interpretation of behavior (Lee 2008). The cases where behavior deviates from the predictions of game theory offer a unique opportunity for neuroscience to educate the theories of economics.

A commonly used social decision-making task called the ultimatum game demonstrates the power of research in this area (Sanfey et al. 2003). In the ultimatum game, one player must choose how to divide up a sum of money between himself and a responder. The responder can either accept or reject the proposer's decision. If the responder accepts, then the money is split up according to the proposer. If the responder rejects the offer, then neither player receives any money. Game theory, which assumes that both players will behave in order to maximize their own payoff, predicts that the proposer would offer very little money to the responder and the responder would accept any offer made by the proposer. However, players rarely behave in this way. Rather, the responders tend to reject low offers, even when the games are played in single trials so they stand to gain nothing (Sanfey et al. 2003). While game theory cannot explain this behavior, neuroscience can offer some insight into the internal processes that generate these decisions. During this game, players with higher activation in the insular cortex, a region implicated in emotional responding, tend to reject low offers more frequently (Sanfey et al. 2003). These findings invite the possibility of including emotional states in future theories of social decision-making.

Investigation into social decision-making has been able to demonstrate causal roles of some brain regions for observed behaviors. In the ultimatum game a region of the lateral prefrontal cortex is activated as responders decide whether to accept or reject an offer, and

activation is stronger when offers are unfair (Sanfey et al. 2003). Lateral prefrontal cortex is thought to be involved in executive control and inhibition of urges (Tanji and Hoshi 2008). Using powerful magnetic stimulation at the scalp, researchers have been able to temporarily disrupt function in this region in human subjects, causing them to accept unfair offers from other human players more frequently (van 't Wout et al. 2005). This finding suggests that the lateral prefrontal cortex plays an important role in reducing selfish behavior in social settings.

Future Directions and Challenges

The above social decision study demonstrates how neuroeconomics could in the future be used to aid our understanding of psychiatric disorders such as obsessive compulsive disorder, addiction (Redish 2004) or borderline personality disorder (King-Casas et al. 2008). In these disorders, patients seem to mis-value their available alternatives and respond abnormally in social situations. Computational models are already being applied to the study of addictive drugs and their effects on valuation and decision-making (Redish 2004), with the hope of tracking changes in decision-making through the different stages of addiction. Moreover, with further utilization of the broad range of interference tools available in neuroscience such as genetic knockout animals, behavioral pharmacology and focal brain lesions, a better grasp on causal relationships between brain function and decision-making can be obtained.

While it is a promising young field, neuroeconomics still has some important challenges to overcome. While the techniques and behavioral tasks have reached a great deal of sophistication, there is still some question as to how closely behavior in the lab relates to decision-making in real world situations or financial markets. This is especially important in the

case of social decision-making, in which nonverbal communication and personal space make up an important component of the social experience. Such characteristics of social settings are difficult to simulate inside an fMRI scanner. Also, neuroeconomics experiments tend to use money as a reward, but the relationship between money and primary rewards such as food, water, or sex, is not clear. Money may be so deeply ingrained in human behavior that it is interchangeable with primary reward, or it may represent consumable rewards that can be obtained in the future. In order for the total integration of the field, this relationship needs to be explored fully to temper comparisons between human studies that use monetary reward and human and animal studies that use primary rewards.

Neuroeconomics also faces an important conceptual challenge. The concept of competing neural systems that regulate rational versus emotion based decisions has not been universally accepted, and many researchers in the field take the view that neural systems engaged in decision-making contain both rational and emotional components and contribute to decision-making in a more graded fashion. This debate is still awaiting the critical evidence to rule out either hypothesis. Near the center of this debate is the neurotransmitter dopamine's role in Pavlovian learning along with its role as a carrier of expected value information (Clark et al. 2012). Dopamine's role in decision-making may be limited to a process that has been labeled as affective or autonomic (Clark et al. 2012), while counterpart systems contribute to more cognitive valuation and decision-making processes (van der Meer and Redish 2009). Nevertheless, affective or not, dopamine's contribution to decision-making makes it a rich target of research in neuroeconomics.

Dopamine in Neuroeconomics

As mentioned above, the role of dopamine in representing primary rewards, and their predictors make it a strong candidate in the search for a neural currency. The activity of dopamine neurons in response to rewards and reward predictive stimuli suggest that the activity of dopamine neurons is tightly linked to economic parameters, such as expected value (Fiorillo et al. 2003, Tobler et al. 2005). Dopamine neurons also modulate their firing in response to the uncertainty of rewards (Fiorillo et al. 2003), suggesting they have a role in encoding probability or risk, a key parameter in economic theories, such as prospect theory (Kahneman and Tversky 1979). Furthermore, dopamine is strongly implicated as a reward prediction error signal that can be used to update estimates of expected value (Montague et al. 1996, Schultz et al. 1997, Pan et al. 2005, Bayer and Glimcher 2005) and as a pathway in the process leading to drug addiction. Contemporary theories suggest that the aberrant behaviors of drug addicts are the result of a co-opting of the brain's reward learning and economic decision-making system in a so-called "computational process gone awry" (Redish 2004).

Unfortunately for cognitive neuroscientists, the VTA is home to several different cell types so BOLD activity in the VTA does not necessarily indicate dopamine. In similar fashion, the relationship between dopamine release in a brain region such as the nucleus accumbens and the BOLD signal is not yet tractable. Positron emission tomography (PET), using radiolabeled dopamine receptor ligands is one method that has been used to study the dopamine system in humans (Volkow et al. 2009), however the use of radioactivity makes it much more invasive than a typical fMRI experiment. One suitable option for selectively measuring dopamine in the brain is to conduct measurements in rodent models, in which more invasive techniques can be used, such as lesions, pharmacological interventions, and optogenetics. While the behavioral repertoire of rats and mice can be limited, economic parameters should be as applicable to simple

as to more complex tasks. Roesch et al. (2007) and Gan et al. (2010) have successfully monitored the activity of dopamine neurons and dopamine release, respectively, in rats, during economic decision-making tasks. The finding of Gan et al. (2010) that dopamine release encodes expected rewards but not response costs in an effort-based decision-making task, along with the previous findings that dopamine release in the accumbens promote drug- (Phillips et al 2003) and food-seeking (Roitman et al. 2004), as well as studies of the effects of dopamine depletion in the nucleus accumbens (Aberman and Salamone 1999), suggest that dopamine release sets a buying price for prospects in economic decision-making tasks (Phillips et al. 2007). Measuring dopamine release in the rodent nucleus accumbens is therefore a powerful tool for neuroeconomics.

Measuring dopamine release in the brain

Direct measurement of dopamine release in target brain regions is typically performed in rats using a chemical method, such as microdialysis, in which extracellular fluid is sampled from the brain over the course of usually tens of minutes and the dopamine content is measured offline, using techniques such as electrochemistry or liquid chromatography. Alternatively, dopamine release can be measured directly in the brain of freely moving rats using fast-scan cyclic voltammetry (FSCV) at carbon fiber microelectrodes (Garris et al. 1997). FSCV offers many advantages over microdialysis, including miniaturization and speed. At its cutting edge, FSCV can be used to sample extracellular dopamine at a frequency of 10 Hz. This level of temporal resolution is critical, as the signals relevant to neuroeconomics typically involve phasic bursts and pauses in dopamine neuron firing. At their working end, the electrodes are only 7 μm in

diameter and the carbon fiber can be encased in glass tubing with a biocompatible coating only 90 μm wide. Electrodes can be chronically implanted in rats (Clark et al. 2010) or mice (Parker et al. 2010), allowing for stable dopamine recordings over weeks, or even months, and they can even be used to monitor dopamine release in awake humans undergoing neurosurgery (Kishida et al. 2011).

FSCV Methodology

For the current work, I exclusively used FSCV at chronically implantable electrodes as described in Clark et al. (2010). The waveform was a triangular voltage waveform from -0.4 V to 1.3 V relative to a Ag/AgCl electrode and back. The scan rate was 400 V/s, resulting in a scan duration of 8.5 ms, and the waveform was applied every 100 ms. The waveform resolution was 1000 or 500 points depending on whether data was being collected from single or multiple electrodes simultaneously. During the upward scan, dopamine is oxidized at the electrode to dopamine-ortho-quinone, which is reduced back to dopamine during the downward scan. The resultant current was recorded, and then low-pass filtered electronically at 2 kHz using a fourth order Butterworth filter. Background currents were subtracted, and the resultant background-subtracted CV data were converted to estimates of dopamine release using principal components regression (PCR). I performed PCR by taking the dot products of the CV data with the first 2 principal components from singular value decomposition of a training set of dopamine and pH CV's recorded in the nucleus accumbens of rats that underwent electrical stimulation of the medial forebrain bundle, the fiber tract that includes the axons of dopaminergic neurons. This methodology is described fully, with included Matlab code, by Keithley and colleagues (2009). I

converted weights along the principal components into dopamine oxidation currents based upon the peak currents from the training set. Oxidation current is linearly related to dopamine concentration, but the conversion factor can only be determined by calibrating an electrode with known concentration of dopamine in a flow cell. Calibration is typically performed after recordings, but could not be conducted in the current work because I located recording sites by lesioning rats' brains through the electrodes, which destroyed the electrodes before removal. On average, for the type of electrodes used in this work, the conversion factor is around 30 nM / nA (Clark et al. 2010).

Chapter 2:

Dopamine release in the nucleus accumbens during risky Pavlovian conditioning contains correlates of expected value and variance*.

*This chapter was a originally a manuscript of the same title by Andrew S Hart, Jeremy J Clark, and Paul E M Phillips prepared for submission. I performed all experimental procedures, analyzed all data, and wrote the manuscript. JJC and PEMP helped design the behavioral protocol.

Abstract

Cue- and reward-evoked phasic firing of dopamine neurons during Pavlovian and operant conditioning paradigms are well correlated with expected value and reward prediction errors. These neural correlates are predicted by reinforcement learning models, such as Rescorla-Wagner (RW) and temporal difference (TD). Additionally, in risky tasks, during which conditioned cues probabilistically predict rewards, dopamine neurons show sustained responses for the duration of risk-associated cues that are correlated with the variance of reward and are maximal to cues predicting rewards with a probability of 0.5. The RW and TD models do not predict this so-called uncertainty signal, and attempts to reconcile it with learning models have been unsatisfactory. In the current study, we sought to examine the acquisition and maintenance of these neural correlates, as these details may give clues to the origin or function of the uncertainty signal. We conducted fast-scan cyclic voltammetry in rats chronically implanted with carbon fiber electrodes in the nucleus accumbens core during risky Pavlovian conditioning. We report that dopamine release in the accumbens core contains correlates of both expected value and variance, and that the acquisition of the variance correlate, as well as its response to recent outcomes, suggest that it serves as an anticipatory signal for upcoming positive reward prediction errors.

Introduction

Phasic firing of dopamine neurons during operant and Pavlovian conditioning tasks conforms well to models of reinforcement learning (Fiorillo et al. 2003, Schultz et al. 1997, Waelti et al. 2001). In naïve animals, dopamine neurons fire bursts of action potentials in response to primary rewards (Ljungberg 1992, Schultz 1997). In animals trained through Pavlovian or operant conditioning tasks, dopamine neurons represent the expected value (EV) and probability of rewards predicted by conditioned stimuli (CSs), as well as reward prediction error signals (RPEs) when the magnitude or timing of a reward differs from that predicted by the CS (Schultz et al. 1997, Waelti et al. 2001, Fiorillo et al. 2003, Nakahara et al. 2004, Bayer and Glimcher 2005,Tobler et al. 2005, Pan et al. 2005, Bayer et al. 2007, Fiorillo 2011). These signals are mirrored by neural correlates of expected value and RPEs in the nucleus accumbens in the form of the BOLD signal (Preuschoff et al. 2006, Rutledge et al. 2010) and phasic dopamine release (Gan et al. 2010, Flagel et al. 2011). Moreover, they are well-characterized in reinforcement learning models such as the Rescorla-Wagner (Rescorla and Wagner 1972) and temporal difference (TD) models (Sutton 1988), suggesting that dopamine signaling may play an important role as a teaching signal.

In addition to representing EVs and RPEs dopamine neurons represent the uncertainty of reward associated with a partially reinforced CS. Fiorillo et al. (2003) reported a sustained increase or ramping in the firing rate of monkey dopamine neurons for the duration of CSs that predicted rewards probabilistically (Fiorillo et al. 2003). This unprecedented signal correlates with the variance of reward, with a maximum response to CSs that predict rewards with a probability of 0.5 (Fiorillo et al. 2003). A variance-correlated signal can also be found in

accumbens neural activity (Preuschoff et al. 2006) but it is not part of the standard reinforcement learning models (Sutton 1988).

Because it is not a component of standard RL models and because it has been observed in neural recordings only from well-training monkeys, it is unclear how a representation of reward variance is acquired or how it is used. Attempts to reconcile the signal with the TD model have resulted in controversy (Niv et al. 2005, Fiorillo et al. 2005). Niv et al. (2005) showed that the variance correlate could be replicated by a modified TD model that represented positive and negative RPEs asymmetrically, which may be a consequence of the low tonic firing rate of dopamine neurons (Bayer and Glimcher 2005). The Niv (2005) model predicts that while the variance signal can be observed in the average responses of dopamine neurons to a risky CS, it is actually an artifact that would not be present in the individual responses. However, Schultz and colleagues (Fiorillo et al. 2005) later showed that the variance correlate was present in the responses of dopamine neurons on individual trials, so the mechanism that produces the variance correlate and its role in learning or action selection is still unclear.

Studies using fast scan cyclic voltammetry in the nucleus accumbens core of rats during Pavlovian conditioning or operant tasks in rats have demonstrated that EV and RPE correlates are present in the form of phasic dopamine release (Stuber et al. 2008, Gan et al. 2010, Flagel et al. 2011), but a correlate of variance or uncertainty in the form of phasic dopamine release has only been anecdotally described (Sugam et al. 2012), and its acquisition has not been documented. In the current study, we recorded dopamine release in the nucleus accumbens core of rats undergoing Pavlovian conditioning with partial reinforcement. We report that that correlates of both probability and variance are present in CS-evoked dopamine release, but that the correlates emerge at different times during learning. Our findings also suggest that the two

signals are combined in the peak CS-evoked dopamine response. Furthermore, we show that the variance correlate is sensitive to the outcomes of recent trials in a manner consistent with an anticipatory signal for an upcoming positive RPE, rather than a true variance signal.

Materials and Methods

Animals and Surgery

The University of Washington institutional animal care and use committee approved all animal procedures, which were conducted during the animals' light cycle. We anaesthetized male Sprague-Dawley rats (mean ± σ weight = 389 ± 30 grams) with isoflurane and chronically implanted them either unilaterally or bilaterally with carbon fiber microelectrodes targeted to the nucleus accumbens core (from Bregma: anterior 1.3 mm, medial 1.3 mm, ventral 6.8 - 7.2 mm) and unilateral Ag/AgCl reference electrodes (Clark et al. 2010). We connected electrodes to 6-pin data-mate connectors (Harwin, Portsmouth, UK), which were cemented to the skull. Implants were held in place with stainless steel screws in the skull and dental cement. We singly housed rats after surgery for the duration of the experiment. After recovery from surgery, we food-restricted rats to 85% to 90% of their post-recovery body weight before beginning behavioral training. We kept them on food restriction throughout training.

Pavlovian Training

We conducted Pavlovian training and voltammetry recording in modified MED Associates (St. Albans, VT) behavioral chambers (Gan et al. 2010, Flagel et al. 2011). Each chamber was equipped with a house-light, a fan, two retractable levers below cue lights and a food magazine equipped with a cue light, an IR beam, and a photo-sensor, located between the levers. Behavioral chambers were equipped with infrared sensitive cameras, and DVD recorders so that behavior could be recorded and scored offline. We habituated rats to the behavior chambers before the beginning of training. During habituation, the magazine light, house light, and fan were turned on, and approximately ten 45 mg dustless precision pellets (Bio-Serv, Frenchtown, NJ) were placed in the feeder magazine. The rats remained in behavior chambers until they consumed all of the pellets. After habituation, the rats then underwent three sessions of magazine training. During magazine training, the house-light and fan were on continuously and the unconditioned stimulus (US) was delivered 20 times per session 90 ± 30 s apart. The US was composed of the delivery of a single food pellet coincident with the start of a 3-second illumination of the magazine light.

After magazine training, we conducted Pavlovian delay conditioning. The conditioned stimulus (CS) was composed of a lever extension and illumination of a cue light for 8 s. We assigned the CS to the left or right side of the feeder to each rat in a counterbalanced manner. We assigned rats to 5 probability (0: n = 4, 0.25: n = 4, 0.50: n = 5, 0.75: n = 4, 1.00: n = 5) groups. All rats received 24 rewarded trials per session, except one rat in the 0.5 group who received 25 rewarded trials per session. On rewarded trials the CS was immediately followed by the US. On reward omission trials, the CS was presented but was not followed the US. The number of reward omission trials varied between groups and was determined by the probability of reinforcement (Probability: Reward Omission Trials, 0.25: 72, 0.5: 24 (25 for one rat), 0.75: 8,

1.00: 0). Rats in the 0 probability group received 24 non-rewarded CS presentations per session. They were also randomly presented with the US 24 times throughout each session. During behavioral sessions, we automatically recorded feeder entry and lever press responses. We conducted no more than one behavioral session per rat per day and we ran behavioral sessions once every 2 to 3 days. We continued Pavlovian conditioning up to ten sessions for each rat or until the rat no longer had viable voltammetry electrodes.

Approach Behavior

We scored CS approach trials from DVD recordings of Pavlovian training sessions. We counted a trial as a CS approach trial if, during the CS period, any part of the rat's head or forelimb made contact with the lever or occupied the space immediately above or below the lever. We calculated approach probability as the number of cue approach trials divided by the total number of trials. We used this method to score approaches rather than lever presses because we found that rats often made lever contacts that did not register as lever pressing. Rats would bite, pull, or even push up on the lever with their noses, and in other instances, rats approached and fixated on the cue light immediately above the lever but did not make lever contacts. We calculated a 2-way mix model ANOVA for effects of session (session 1 vs. session 6) and probability of reinforcement on CS approach probability. We also conducted Bonferroni-corrected t-tests to compare each group to the 0 probability group in sessions 1 and 6.

Voltammetry

We recorded phasic dopamine release using fast-scan cyclic voltammetry (FSCV) at the implanted carbon fiber electrodes. We began voltammetry recordings no later than the third magazine training session. We conducted recordings as in Gan et al. (2010) and Clark et al. (2010). We plugged each rat into a head stage amplifier (built in our lab by Scott, Ng-Evans, University of Washington), which was attached to a commutator (Crist Instrument Co., Bethesda, MD or Dragonfly Research and Development, Inc., Ridgely, WV), allowing it to move freely during recording sessions. Commutators were connected to National Instruments cards, and signals were digitized and recorded on a PC running Tar Heel Labview Software. Throughout FSCV recordings, electrodes were held at -0.4 V, and a triangular voltage waveform was applied to the electrode every 100 ms. The peak of the waveform was 1.3 V, and the scan rate was 400 V/s; the total duration of the waveform was 8.5 ms. During the voltage sweep, current data from the electrodes was collected in either single or dual-channel modes. In single-channel mode, 1000 points of data were sampled during each 8.5 ms voltage scan, for a sample rate of 117.6 kHz. In dual-channel mode, 500 points of data from each channel were sampled during each scan, for a sample rate of 58.8 kHz. To facilitate conditioning of the electrode surface before recordings, we applied the waveform at 60Hz for approximately 1 hour, before being switched to the normal 10 Hz collection mode and waiting for another 30 minutes. To facilitate synchronization of voltammetry and behavior, TTLs from the behavioral control computer were recorded concurrently with voltammetry data. We verified electrode function before the start of Pavlovian training by examining the mean US-evoked dopamine release from 20 unpredicted USs in magazine session 3, and re-verified electrodes before each Pavlovian session by presenting an unpredicted food pellet. We recorded cyclic voltammagrams (CVs) during unpredicted pellet presentation and inspected background-subtracted CVs for their similarity to a

dopamine CV. If an unpredicted reward failed to produce a dopamine-like CV, then the animal was rested for 2 days. After 3 consecutive failures, we considered electrodes no longer viable for data collection. We excluded data from electrodes that were not viable for at least 6 Pavlovian sessions with voltammetry recording.

Data Processing and Analysis

We recorded dopamine release in the accumbens core from at least 6 Pavlovian sessions with 24 (0: n = 5, 0.25: n = 4; 0.5 n = 6; 0.75 n = 4; 1: n = 5) electrodes implanted in 22 rats. We processed all FSCV data using custom Matlab (The Mathworks Natick, MA) software written by ASH or modified by ASH from Keithley and Wightman (2010). We low-pass filtered FSCV at 2 kHz. We parsed voltammetry data into trials by cutting the data to the onset of the CSs, as well as the unpredicted rewards in the case of magazine training sessions and Pavlovian sessions for the 0 probability group. In order to remove the background current from the data, we subtracted the average of 10 CVs recorded immediately before trial onset from each trial's CV data. We converted CV data to dopamine and pH current using principal components regression (2 principal components) against a training set of electrically evoked dopamine and pH responses. We conducted residual error analysis on each CV and excluded CV data for which the error was great enough to reject the null hypothesis ($\alpha = 0.05$) that it was produced by random noise (Jackson and Mudholkar 1979). To calculate the change in dopamine due to reward delivery or reward omission in the non-zero probability groups, we subtracted the average dopamine signal from the last 10 scans during the CS. When showing dopamine traces from individual trials, we

smoothed data with a 5-point sliding average, however, we conducted all quantitative analyses on unsmoothed data averaged over trials.

We conducted regression analyses to test the CS- and US-evoked dopamine responses from each session for correlates of EV and variance. For binary rewards, EV is equivalent to probability of reinforcement (P_R), and variance is equivalent to $P_R - P_R^2$. Therefore, we fit the data to first (equation 1) and second order (equation 2) polynomial functions of P_R, where $D(t)$ is the dopamine signal recorded at time t. The negative weight for B_2 reflects the expected negative weight on P_R^2 for a variance correlate.

$$D(t) = B_0(t) + B_1(t) * P_R \quad (1)$$

$$D(t) = B_0(t) + B_1(t) * P_R - B_2(t) * P_R^2 \quad (2)$$

We calculated the mean CS-evoked and US-evoked dopamine response from each session for each electrode (n=24). We then calculated regressions of session mean responses at each time point against first- and second-order polynomial models of dopamine as a function of reward probability. We calculated F statistics and r^2 for both models for (1,22 and 2,21 degrees of freedom) as well as the F statistic (1,21 degrees of freedom) and the marginal increase in r^2 for the addition of the second-order term to the first-order model. As an estimate of the relative performance of the two models in explaining the variance of the data, we compared the marginal increase in r^2 for the second-order model with r^2 for the first-order model. We repeated this procedure for all time points (-1 s to 8 s relative to CS-onset, 0.1 s interval) for six sessions of acquisition. For US-evoked dopamine release, we used a similar procedure but calculated regressions against the first-order model only for time points relative to the US (-1 s to 6 s

relative to US-onset). For follow up analyses, we calculated the mean dopamine oxidation current from the session averaged dopamine responses for each electrode over 4 epochs from session 2 and session 6. Three epochs were defined by the CS: CS-onset (0.4 - 1.4 s), CS-peak (1.5 - 2.5 s), CS-late (6.9 - 7.9s), and one epoch was defined by the reward: US-peak (1.5-2.5s from US onset). We also calculated the mean current from the US-peak epoch for reward omission trials. We performed the same regression procedures as above on the mean dopamine responses from each epoch.

We calculated history-based within-electrode contrasts for CS-evoked responses from trials in three stages of Pavlovian training. For subjects in the 0.25, 0.5, and 0.75 probability groups, we sorted trials by the outcomes of the previous two trials. We then calculated the average CS-evoked dopamine response from trials following two rewarded trials and from trials following two omissions from early training (sessions 1 - 3), late training (sessions 4 - 6), and asymptotic (sessions 7-10) stages. We excluded responses from the first two trials of each session. For asymptotic data, we used the subset of electrodes that reached 10 Pavlovian sessions (0.25: n = 4, 0.5: n = 4, 0.75: n = 4). We then calculated the mean dopamine oxidation current in the CS-peak and CS-late epochs defined above, and for contrasts, we subtracted the responses from trials following two omissions from the responses from trials following two rewards. We conducted six paired t-tests for within-electrodes contrasts for each stage/epoch combination and corrected α-levels from t-tests using the Holm-Bonferroni procedure. We performed all statistics using Matlab, except 2-way ANOVAs, for which we used Prism 4.0 (Graphpad Software, San Diego, CA). $\alpha = 0.05$ for all tests, except where corrected.

Histology

After completing Pavlovian training and FSCV data collection, we anaesthetized rats with ketamine (150 mg / kg) and performed electrolytic lesions through their electrodes. We subsequently perfused the rats through the heart with saline, followed by paraformaldehyde (PFA, 40 g / L in phosphate buffered saline). We then removed brains and and stored them in PFA in PBS at 4°C. We later saturated brains in sucrose (300 g / L in PBS) at 4°C. We then froze them on dry ice and sectioned them on a cryostat at 50 μm. We mounted sections containing the nucleus accumbens core on slides and stained with cresyl violet. We verified electrode placement using an adult rat brain atlas (Paxinos and Watson 2005).

Results

Behavior

We found that Pavlovian conditioning with any non-zero probability of reinforcement produced learned CS approach behavior. Rats in the 0.25, 0.5, 0.75, and 1.0 probability groups increased their likelihood of CS approach between the first and sixth sessions, while rats in the non-paired group showed a downward trend in approach probability (Figure 2.1; 2-Way mixed-model ANOVA Session: $F_{1,17} = 26.90$, $P = 0.00007$, Probability: $F_{4,17} = 8.463$, $P = 0.0006$, Interaction: $F_{4,17} = 5.388$, $P = 0.0055$). Post-hoc analysis revealed that rats in the 0.25, 0.5, 0.75, 1.0 probability groups approached the lever more frequently than rats in the 0 probability group in session 6 ($t > 4.784$, $P < 0.001$ for all comparisons), but not in session 1 ($t < 1.553$, $P > 0.05$ for all comparisons).

Figure 2.1

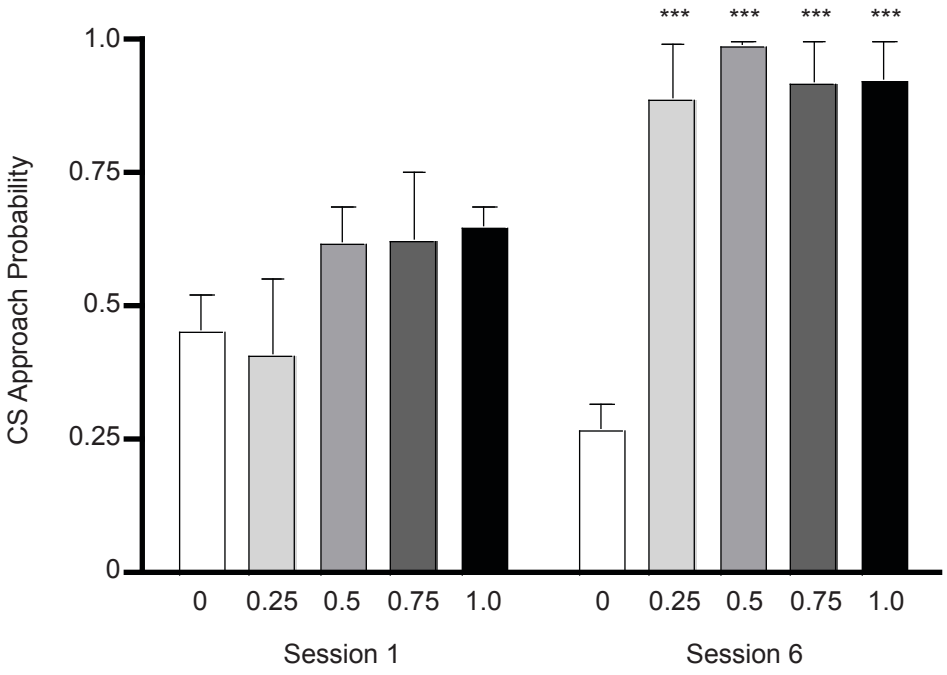

***
***
***
***
1.0
0.75
0.5
0.25
0
CS Approach Probability
0
0.25
0.5
0.75
1.0
0
0.25
0.5
0.75
1.0
Session 1
Session 6

Figure 2.1: Rats in all groups approached the CS during the first session of training, but by session 6, rats in the 0 probability group (n=4) responded significantly less than rats in the 0.25 (n = 4), 0.5 (n=5), 0.75 (n=4), and 1.0 (n=5) groups. Bars show mean plus standard error (***: $P < 0.001$ with respect to non-paired group, Bonferroni corrected t-test).

CS-Evoked Dopamine Correlates of Expected Value and Variance

CS- and US-evoked dopamine responses recorded in the nucleus accumbens core (Figure 2.2) were present in recordings from individual electrodes on individual trials (Figure 2.3a-c) as well as in average responses over sessions (Figure 2.3d-f, Figure 2.4). The CS reliably evoked dopamine release on individual trials over six sessions of training for rats in all non-zero probability groups. Dopamine levels remained elevated during the CS on individual trials for rats in the 0.25, 0.5 (Figure 2.3b), and 0.75 groups, and this sustained response was present in average dopamine traces from later sessions in the 0.25 (Figure 2.4a), 0.5 (Figure 2.3e), and 0.75 (Figure 2.4b) groups. Sustained responses were also present in average dopamine traces from early sessions in the 1.0 group (Figure 2.3f). US-evoked responses were also present, but they attenuated over training in all groups, and were nearly eliminated over training in rats in the 1.0 probability group (Figure 2.3d-f, Figure 2.4).

Session averaged CS-evoked dopamine release contained correlates of both probability and variance over training. Least squares fits of dopamine release at each time-point by first- and second-order polynomial functions of probability of reinforcement (P_R) revealed that the probability correlate emerged first, while the variance correlate appeared later in training. Examination of regression weights revealed that the first order fits were strongest during the early part of the CS (Figure 2.5a), while the second order fits were strongest during the late part of the CS (Figure 2.5b), while there was blending of the two models at intermediate time points later in training. Time points with significant first-order regressions (nonzero slope at $P < 0.05$, F-test) tended to exist in the first 2 seconds of the CS (Figure 2.5c), while time points for which

Figure 2.2

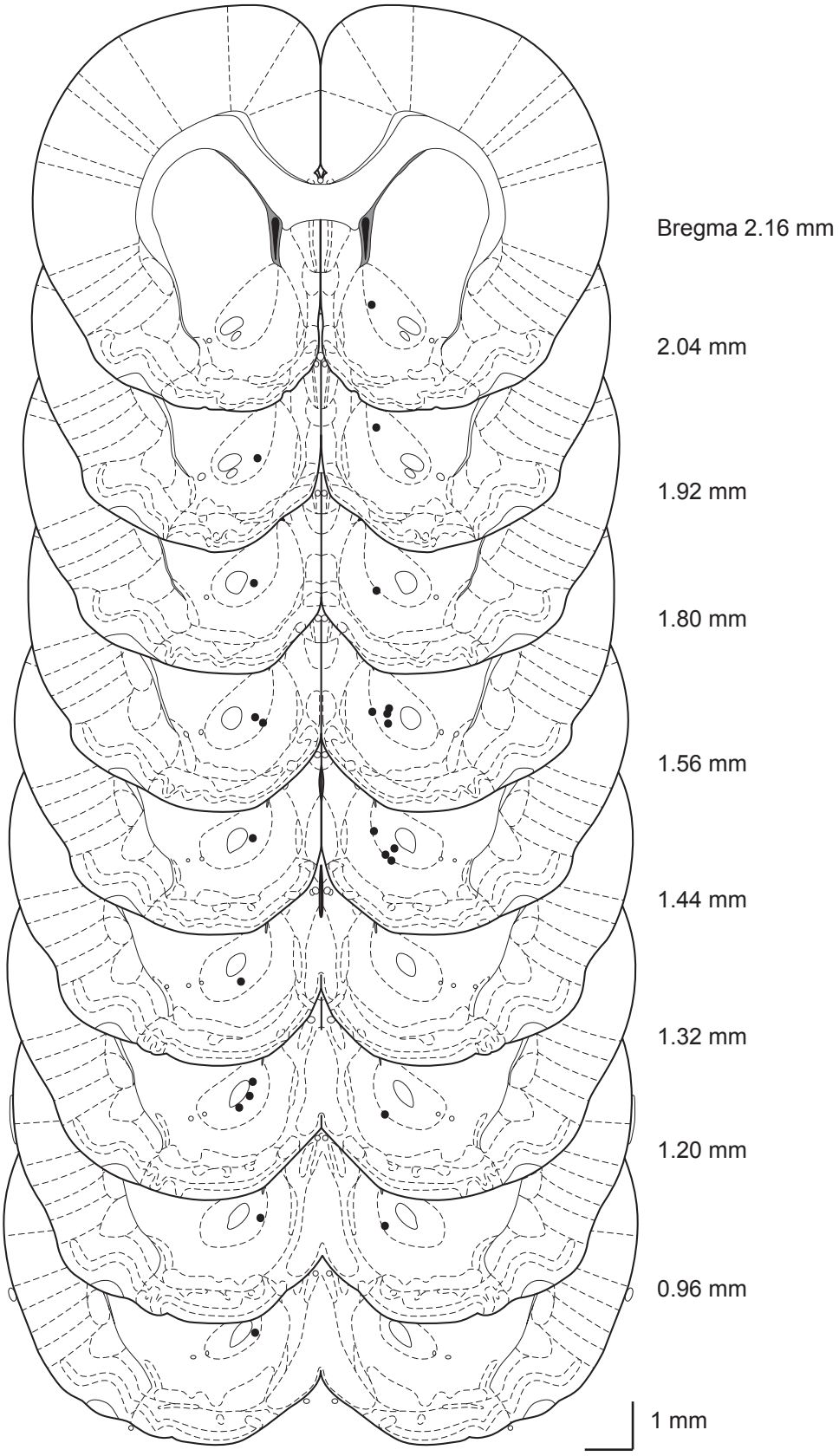

Bregma 2.16 mm
2.04 mm
1.92 mm
1.80 mm
1.56 mm
1.44 mm
1.32 mm
1.20 mm
0.96 mm
1 mm

Figure 2.2: Coronal sections of rat brain show locations for (n = 24) voltammetry electrodes chronically implanted in nucleus accumbens core. Brain atlas sections are from Paxinos and Watson (2005).

Figure 2.3

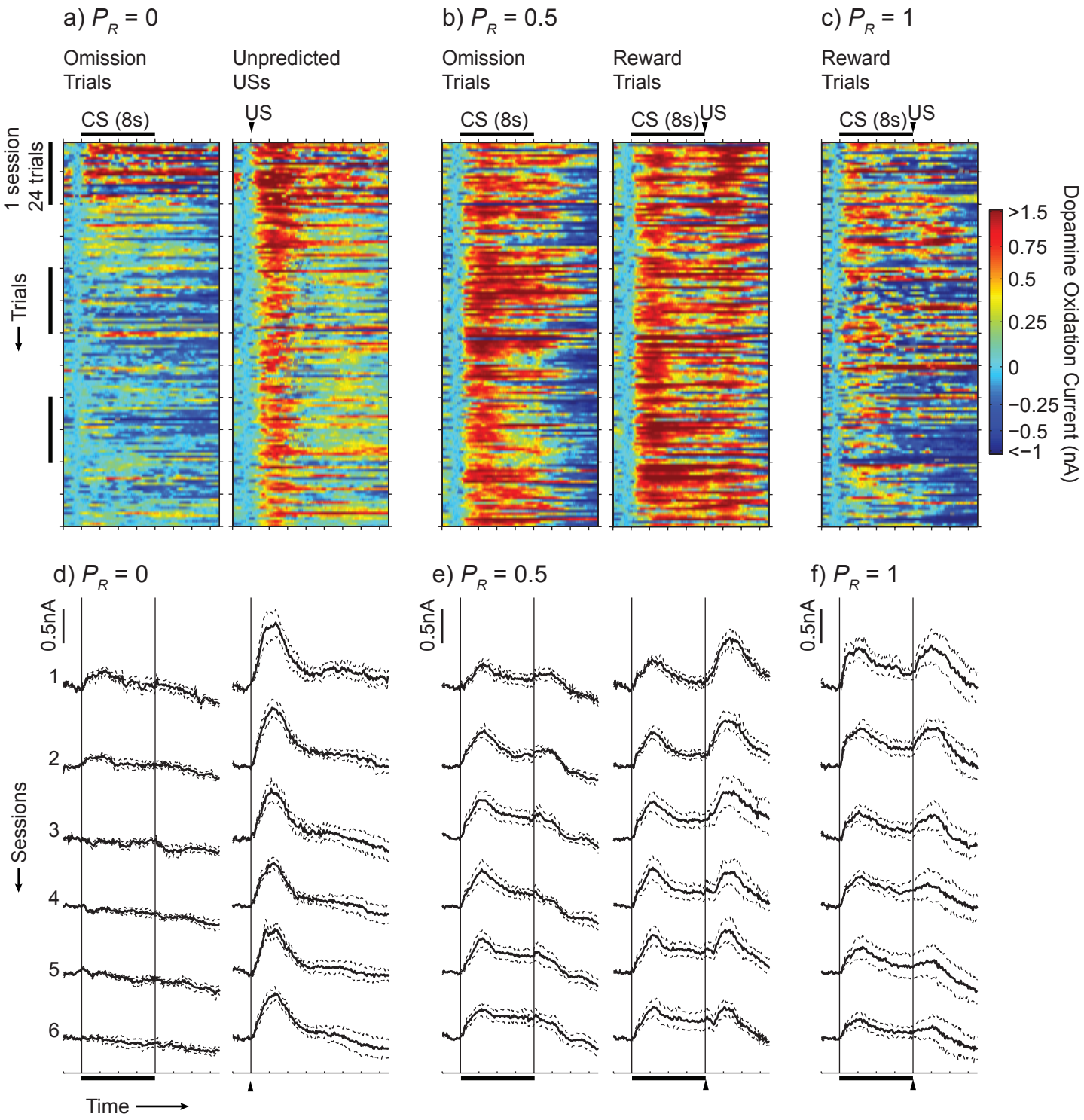

a) P_R = 0
Omission Trials
Unpredicted USs
CS (8s)
US
1 session
24 trials
Trials
b) P_R = 0.5
Omission Trials
Reward Trials
CS (8s)
US
c) P_R = 1
Reward Trials
CS (8s)
US
Dopamine Oxidation Current (nA)
>1.5
0.75
0.5
0.25
0
−0.25
−0.5
<−1
d) P_R = 0
e) P_R = 0.5
f) P_R = 1
0.5nA
Sessions
1
2
3
4
5
6
Time

Figure 2.3: (a-c) Example dopamine traces on individual trials recorded at individual electrodes in the 0, 0.5 and 1.0 groups are shown. For the 0.5 group, reward trials and reward omission trials are shown separately, though they were randomly interleaved during training. Traces were smoothed with a 5-point running average. Gray points indicate data that was excluded because residual error after principal components regression was large enough to reject the null hypothesis ($P < 0.05$) that error was due to random noise. (d-f) Mean ± SEM of session-averaged traces for the 0 ($n = 5$), 0.5 ($n = 6$), and 1.0 ($n = 5$) groups are shown. Traces from reward and reward omission trials for each session were averaged separately to illustrate differential responding at the time of the US.

Figure 2.4

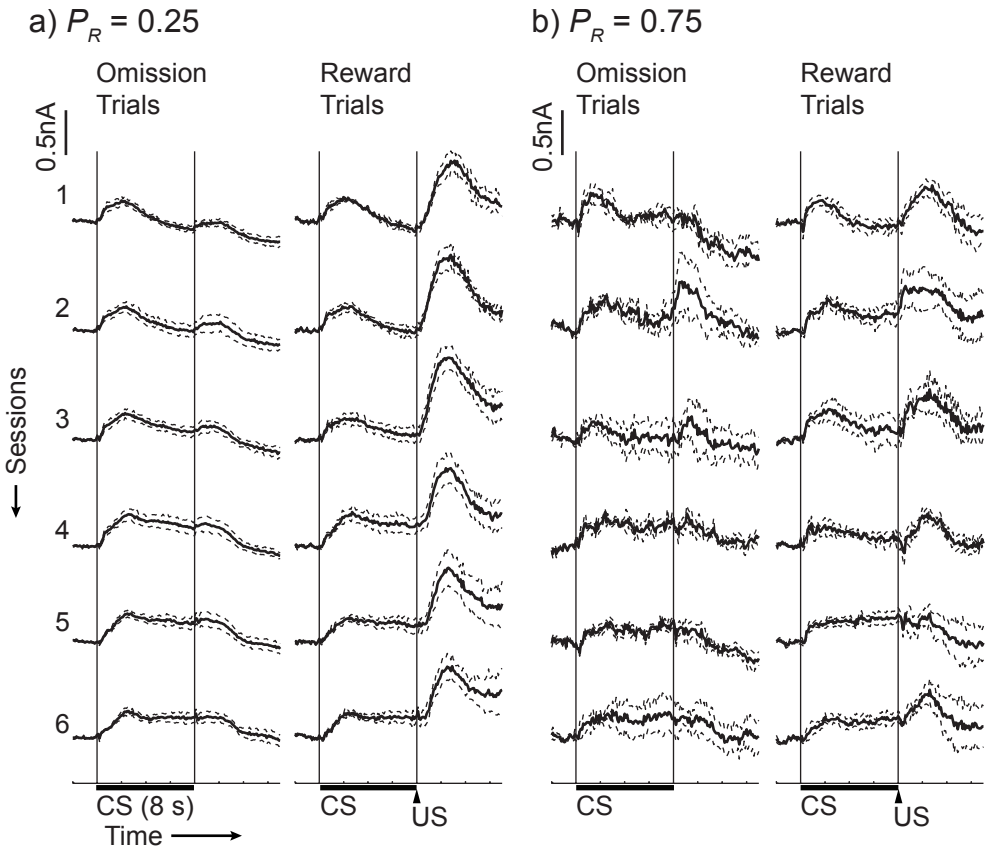

a) P_R = 0.25
b) P_R = 0.75
Omission Trials
Reward Trials
0.5nA
Sessions
1
2
3
4
5
6
CS (8 s)
Time
CS
US

Figure 2.4: Mean ± SEM of session-averaged traces for the 0.25 (n = 4), 0.75 (n = 4) groups are shown. Traces from reward and reward omission trials for each session were averaged separately to illustrate differential responding at the time of the US.

Figure 2.5

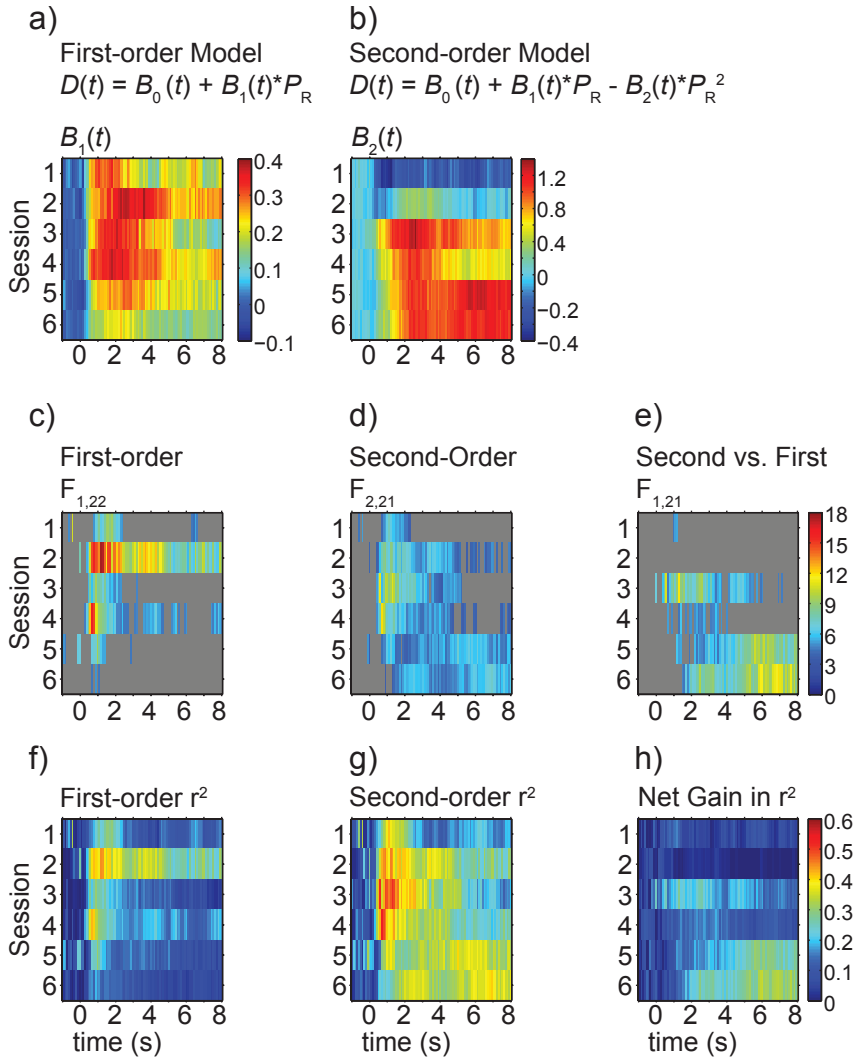

a)
First-order Model
D(t) = B0 (t) + B1(t)*PR
B1(t)
b)
Second-order Model
D(t) = B0 (t) + B1(t)*PR - B2(t)*PR 2
B2(t)
c)
First-order
F1,22
d)
Second-Order
F2,21
e)
Second vs. First
F1,21
f)
First-order r2
g)
Second-order r2
h)
Net Gain in r2
Session
time (s)

Figure 2.5: Regression weights for B_1 (a) from the first-order model and B_2 (b) from the second-order model are shown for each time point (-1s to 8 s from CS-onset, 0.1 s interval) for sessions 1 through 6. (c-e) F-statistics for significant ($P < 0.05$) least squares fits for the first- (c) and second-order (d) models, as well as for comparison between the two models (e) are shown for the regressions in a and b. Gray indicates that the F test was not significant for that time point. (f-h) r^2 values are shown for the first- (f) and second-order (g) models in a and b, as well as the net increase in r^2 (h) for the second-order over the first-order model at each time point.

the second-order model significantly improved the fit ($P < 0.05$, F-test) over the linear model tended to exist during 2 to 8 s after CS onset (Figure 2.5e). Exceptions to these trends existed in session 2, during which the linear model was significant for all time points during the CS, and session 3, during which the quadratic model significantly improved the fit during 1 to 2 s after CS onset. Comparison of marginal increase in explained variance for the second-order term to the model with the explained variance of the first-order model indicated that that linear model tended to perform best during the first second after CS-onset, while the second-order model tended to perform best during the last second of the CS (Figure 2.6). These trends strengthened over training. Follow up analyses on mean CS-evoked dopamine release from early-, peak-, and late-CS epochs from sessions two (Figure 2.7a) and six (Figure 2.7b) highlight the transition from the first-order to the second-order model from early to late learning and within the time series of the CS-evoked response during the late learning phase. For all three epochs tested, the linear model fit the dopamine response in session two, and the quadratic model did not significantly improve the fit. In session six, the quadratic model significantly improved the fit and the second order term accounted for more variance than the first order term for the peak-CS and late-CS epochs but not the early-CS epoch (Figure 2.7, see table 2.1 for statistics).

US-Evoked Dopamine Correlates of Probability

Linear regression analyses of the US-evoked dopamine at each time point from sessions 1 through 6 showed that dopamine release was negatively correlated with P_R throughout learning, with the stronger correlations later in training than early (Figure 2.8). Learning was characterized by a decrease in the intercepts (B_0) rather than a steepening of the slopes (B_1) of the

Figure 2.6

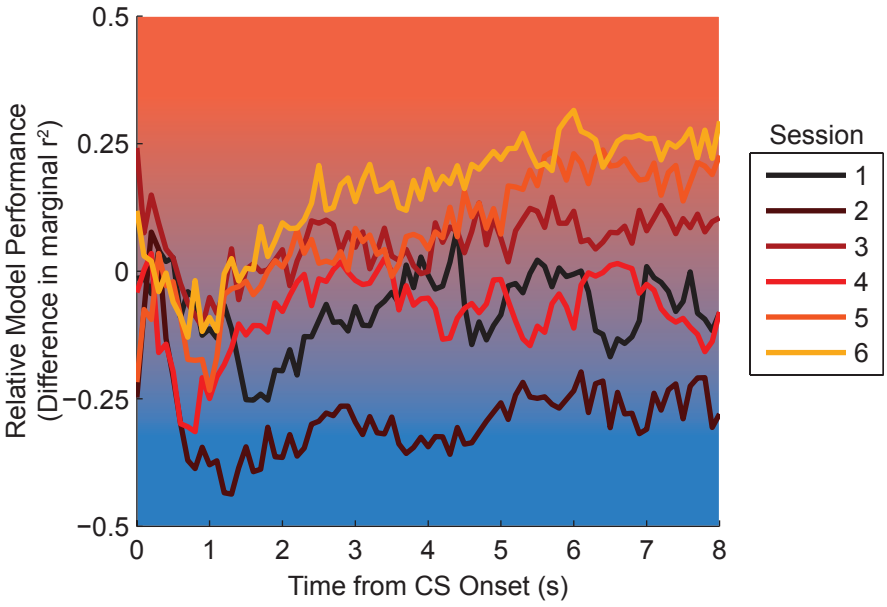

0.5
0.25
0
−0.25
−0.5
Relative Model Performance
(Difference in marginal r²)
0
1
2
3
4
5
6
7
8
Time from CS Onset (s)
Session
1
2
3
4
5
6

Figure 2.6: Plots show the relative model performance as the difference between the net increase in r^2 for adding the second-order term and the first-order $r^2$for time points during the CS in sessions 1 through 6. More positive values indicate that the second-order model outperforms the first-order model in explaining variance within the data. More negative values indicate that the first-order model outperforms the second-order model.

Figure 2.7

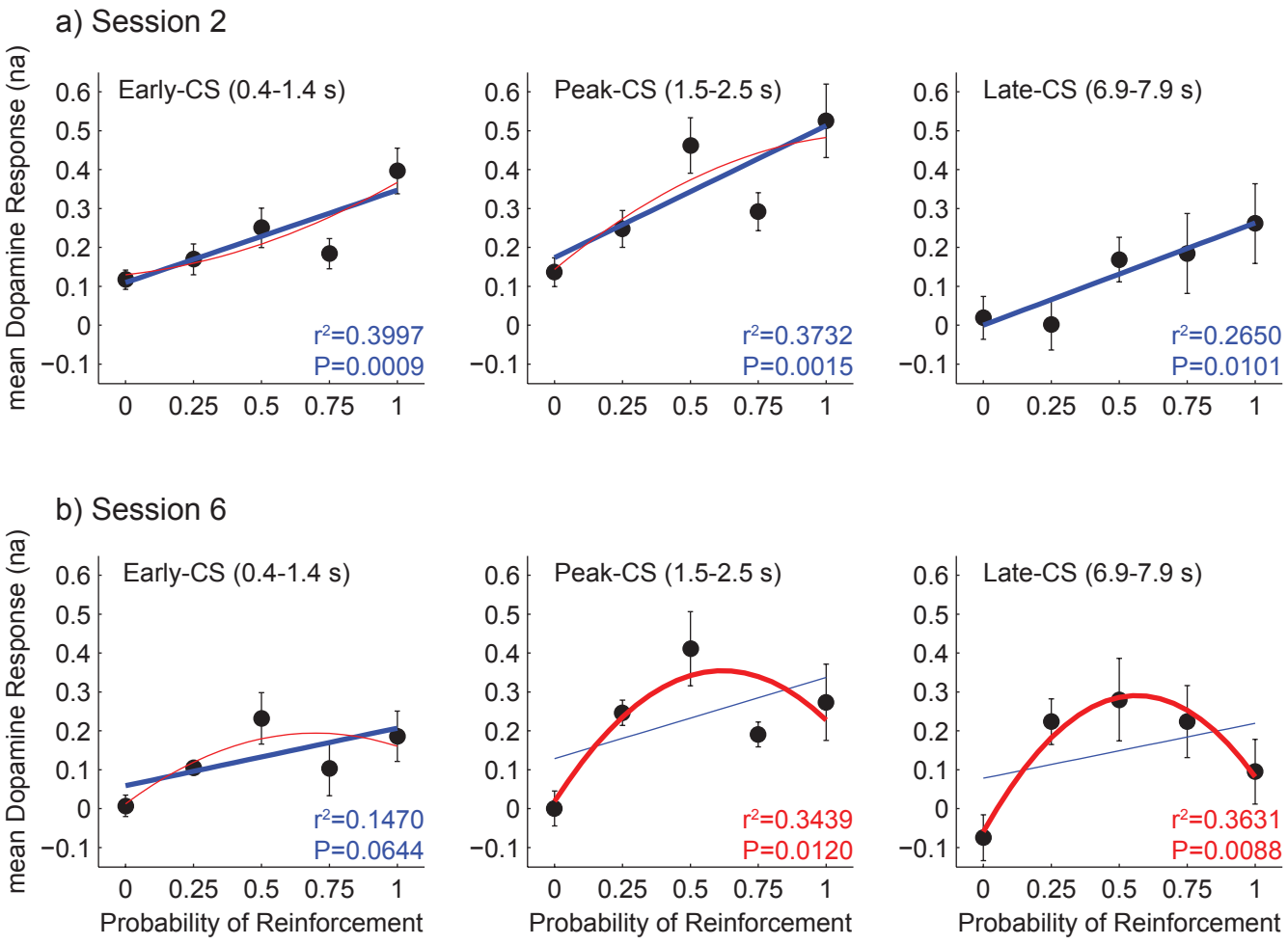

a) Session 2
mean Dopamine Response (na)
Early-CS (0.4-1.4 s)
r²=0.3997
P=0.0009
Peak-CS (1.5-2.5 s)
r²=0.3732
P=0.0015
Late-CS (6.9-7.9 s)
r²=0.2650
P=0.0101
b) Session 6
Early-CS (0.4-1.4 s)
r²=0.1470
P=0.0644
Peak-CS (1.5-2.5 s)
r²=0.3439
P=0.0120
Late-CS (6.9-7.9 s)
r²=0.3631
P=0.0088
Probability of Reinforcement

Figure 2.7: (a-b) Group mean ± SEM for dopamine responses in early-CS (left), peak-CS (middle), and late-CS (right) epochs for sessions 2 (a) and 6 (b) are plotted with respect to reinforcement probability. Curves for first-order (blue) and second-order (red) models fits are shown for each epoch. The heavier curve in each plot is for the model that produces the greater net increase in r^2. All first-order fits in session 2 are significant. Second order fits for peak-CS and late-CS in session 6 are significant. The first order fit for early-CS in session 6 borders on significance (See table 1 for statistics).

Session Epoch	B_1 SEM	B_2 SEM	Lin $F_{1,22}$ P	Quad $F_{2,21}$ P	Q vs. L $F_{1,21}$ P	r^2 for B_1	Net r^2 for B_2
2 Early	0.2376 0.0621	-0.1593 0.2046	**14.6496** **0.0009**	**7.4969** **0.0035**	0.6063 0.4449	0.3997	0.0168
2 Peak	0.3393 0.0938	0.2417 0.3090	**13.0989** **0.0015**	**6.7400** **0.0055**	0.6120 0.4428	0.3732	0.0178
2 Late	0.2629 0.0933	-0.0069 0.3120	**7.9321** **0.0101**	**3.7861** **0.0394**	0.0005 0.9825	0.2650	0.00001
6 Early	0.1481 0.0760	0.3733 0.2408	3.7920 0.0644	3.2188 0.0603	2.4037 0.1360	0.1470	0.0876
6 Peak	0.2086 0.1153	0.8767 0.3347	3.2732 0.0841	**5.5039** **0.0120**	**6.8623** **0.0160**	0.1295	0.2144
6 Late	0.1414 0.1257	1.1003 0.3449	1.2644 0.2729	**5.9849** **0.0088**	**10.1779** **0.0044**	0.0544	0.3087

Table 2.1: Statistics are shown for follow-up regression analysis on mean CS-evoked dopamine release for early (0.4 - 1.4 s), peak (1.5 - 2.5 s), and late (6.9 - 7.9 s) epochs from sessions 2 and 6. B_1 is the slope for the first-order model, and B_2 is the weight for $-P_R^2$ in the second order-model. F and P values that are significant at $P < 0.05$ are shown in bold. Net r^2 is the difference in r^2 between the first- and second-order models. r^2 and net r^2 are colored to signify the term that accounts for the most variance (blue: B_1, red: B_2).

Figure 2.8

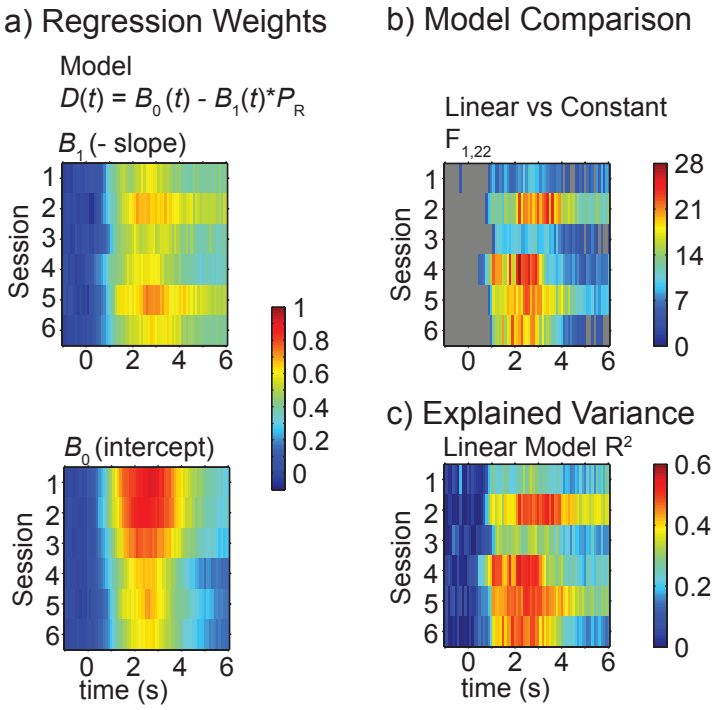

a) Regression Weights
Model
D(t) = B0 (t) - B1(t)*PR
B1 (- slope)
Session
1
2
3
4
5
6
0 2 4 6
1
0.8
0.6
0.4
0.2
0
B0 (intercept)
Session
time (s)
b) Model Comparison
Linear vs Constant
F1,22
28
21
14
7
0
c) Explained Variance
Linear Model R2
0.6
0.4
0.2
0
time (s)

Figure 2.8: (a) Slopes (B_1) and intercepts (B_0) are shown for linear regressions at each time points (-1 s to 6 s relative to US-onset, 0.1 s interval) for rewarded trials in sessions 1 through 6. Significant ($P < 0.05$) F statistics (b) and r^2 (c) values are shown for linear regressions in a.

linear regressions. Follow-up analysis on the mean dopamine release from the peak-US epoch (1.5 - 2.5 s after US onset) from session 2 and session 6 showed consistent results (Figure 2.9). Linear regressions were significant for session 2 (B_0 = 0.8225 ± 0.0910, B_1 = -0.6058 ± 0.1486, r^2 = 0.4303, P = 0.0005) and session 6 (B_0 = 0.5580 ± 0.0716, B_1 = -0.5217 ± 0.1169, r^2 = 0.4753, P = 0.0002), and while the intercept was greater in session 2 than in session 6 (t_{22} = 3.2534, P=0.0036), the slopes did not significantly differ (t_{22} = 0.6378, P=0.5302). Slopes of regression lines of the mean dopamine release during the same epoch on reward omission trials against P_R were not significantly different from zero (Session 2: r^2 = 0.1039, P = 0.1784; Session 6: r^2 = 0.0066, P = 0.7410).

History effect on CS-evoked dopamine responses

Within-electrode contrasts in CS-evoked dopamine responses between trials that followed two rewards and trials that followed two reward omissions showed that history effects reflected slight changes in reward probability rather than reward variance for both the peak-CS and late-CS differed time windows (Figure 2.10). The peak-CS time window had a significant contrast (Holm-Bonferroni corrected paired t-test t_{13} = 3.5414, P = 0.0036) for early learning but not for the late learning (Holm-Bonferroni corrected paired t-test t_{13} = 2.3377, P = 0.036,) or asymptotic stages (Holm-Bonferroni corrected paired t-test t_{11}= 0.6358, P = 0.5379). In contrast, the late-CS time window had significant contrasts for all three learning stages (Holm-Bonferroni corrected paired t-test: early: t_{13} = 3.0188, P = 0.0099, late: t_{13} = 2.7456, P = 0.0167, asymptotic: t_{11} = 3.0810, P = 0.0105). These effects were present even though data were combined between

Figure 2.9

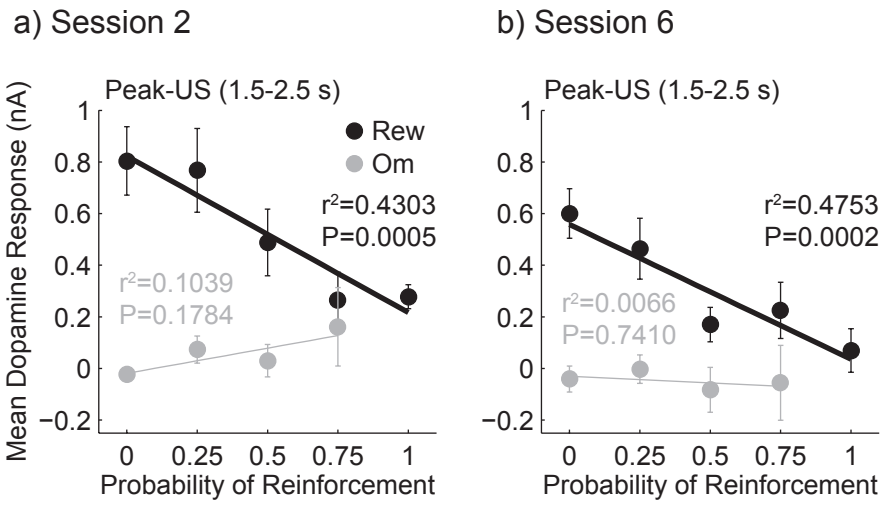

a) Session 2
b) Session 6
Peak-US (1.5-2.5 s)
Mean Dopamine Response (nA)
Rew
Om
r²=0.4303
P=0.0005
r²=0.1039
P=0.1784
r²=0.4753
P=0.0002
r²=0.0066
P=0.7410
1
0.8
0.6
0.4
0.2
0
−0.2
0
0.25
0.5
0.75
1
Probability of Reinforcement

Figure 2.9: Group mean ± SEM for dopamine responses in the peak-US epoch are shown for reward and reward omission trials in sessions 2 (a) and session 6 (b). Linear regressions are significant for responses from reward trials for both sessions, but not for responses from omission trials.

Figure 2.10

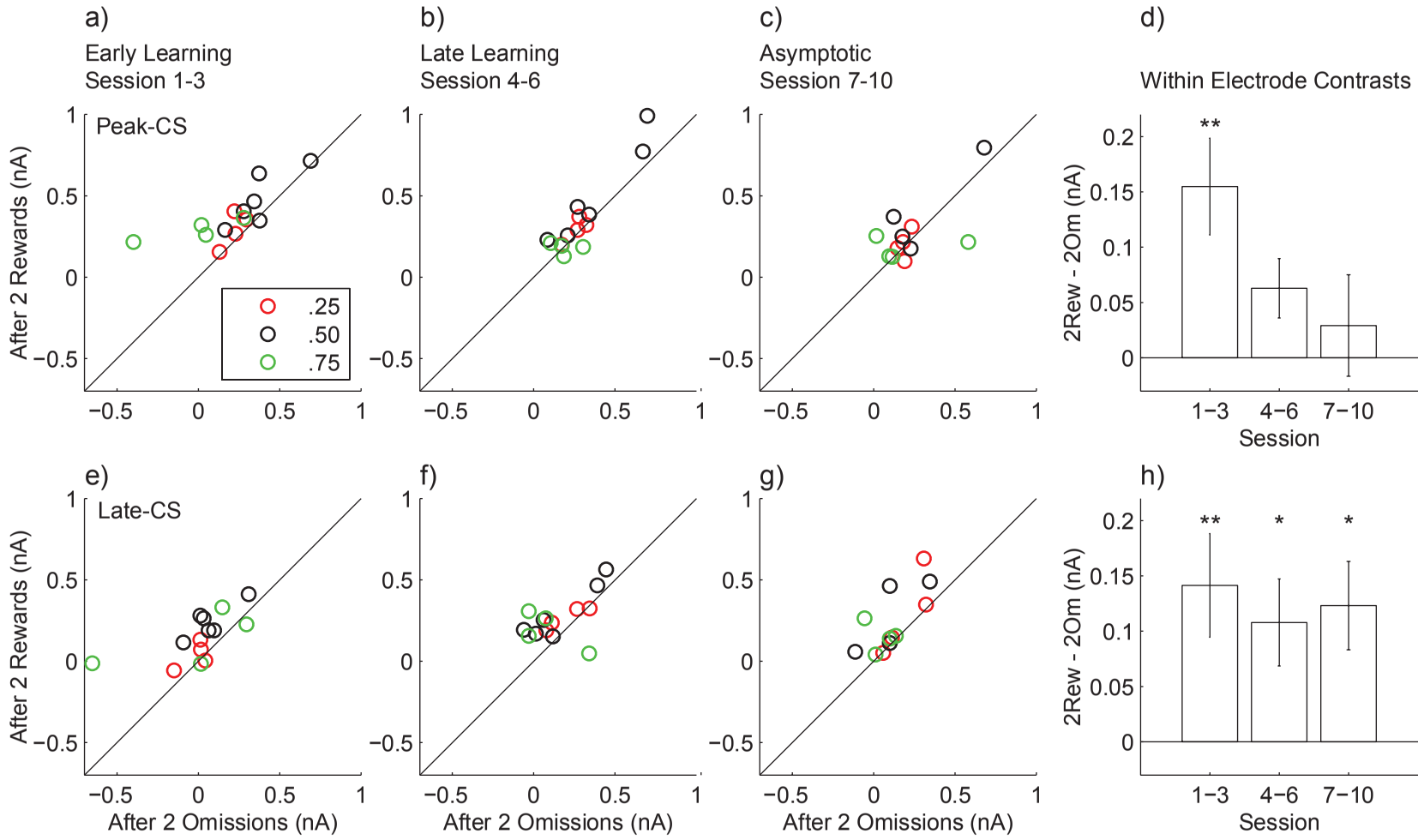

a)
Early Learning
Session 1-3
Peak-CS
b)
Late Learning
Session 4-6
c)
Asymptotic
Session 7-10
d)
Within Electrode Contrasts
.25
.50
.75
After 2 Rewards (nA)
2Rew - 2Om (nA)
1–3
4–6
7–10
Session
e)
Late-CS
f)
g)
h)
After 2 Omissions (nA)

Figure 2.10: (a-c) Scatter plots show the mean dopamine response over the peak-CS epoch from trials following two rewards vs. the mean response from trials following two omissions from early learning (a), late learning (b), and asymptotic (c) stages for electrodes in uncertain probability groups. Points above the line indicate that the signal following two rewards is greater than the signal following two omissions. (d) Bar graph shows Mean ± SE of the difference for within electrode contrasts for the responses in a-c (early: n = 14, late: n = 14, asymptotic: n = 12). (f-g) Scatter plots and bar graph show the same data as a-d but for the late-CS epoch. (**: $P < 0.01$, *: $P < 0.05$ Paired t-test. Holm-Bonferroni correction was applied to α levels)

probability groups, for which runs of rewarded and omission trials would produce distinct and opposite effects on reward variance that would cancel out an overall effect.

Discussion

We used FSCV to record dopamine release in the nucleus accumbens core of rats as they underwent risky Pavlovian conditioning. We report CS- and US-evoked phasic dopamine responses that confirm and expand upon the dopamine neuron firing rates reported by Fiorillo and colleagues (2003). Also in line with their findings, we identified correlates of reward probability and variance in the CS-evoked dopamine signal. Furthermore, our observation of a sustained CS-evoked dopamine response expands the evidence (Fiorillo et al. 2005) that the variance-correlated dopamine signal is not an artifact of averaging (Niv et al. 2005).

We observed that the correlates of EV and variance underwent a dynamic process during learning. The EV correlate dominated early in the CS and early in training, while the variance correlate dominated late in the CS and later in training. Peak CS-evoked dopamine responses to the CS were strongly modulated by both EV and variance, suggesting a blending of the signals. These findings suggest that the discrete coding of EV and variance observed in dopamine neuron firing rate in monkeys (Fiorillo et al. 2003) could give rise to a more blended signal in the form of accumbens dopamine release through the slower time constants for uptake allowing diffusion associated with dopaminergic volume transmission (Garris et al. 1994). Alternatively, this blending could be a characteristic of rat dopamine neurons that is not present in monkeys.

In addition to correlates of EV and variance, we observed a negative correlate of P_R in US-evoked dopamine release. This negative correlation is consistent with dopamine's hypothesized role as a reward prediction error signal and matches observations in monkeys

(Schultz et al. 1997, Fiorillo et al. 2003, Nakahara et al. 2004, Bayer and Glimcher 2005, Bayer et al. 2007). The presence of significant correlations early in learning and a decrease in intercept but not slope between sessions two and six suggest that reward predictions are learned quickly, and that after extensive training on the task, rewards that occur within the task are not as unpredictable as when the animals are new to the task. The context of the behavioral chamber and its set of cues, such as the house light and fan noise, may modestly predict rewards within the behavioral task, so that RPE signals are attenuated across all probabilities. We did not observe significant correlations between P_R and dopamine release on omission trials. This finding is consistent with an imbalance in RPE encoding by dopamine neurons reported by Bayer and Glimcher (2005).

While we did not attempt to determine directly the mechanism that produces the variance-correlated, late-CS dopamine response, we made two observations that give clues about that which it truly represents. First, the late-CS response was best fit by the linear model during session two. This finding suggests that early in training, the sustained response might reflect the rate of rewarded CSs rather than the variance associated with the CSs. Second, the late-CS response was enhanced on trials following two rewards relative to trials following two omissions throughout all stages of the task. Importantly, this enhancement was present among members of all three risky probability groups, suggesting that the variance correlate is not a true variance signal. A true variance signal would show contrasts in opposite directions for the 0.25 and 0.75 groups. This is because when $P_R = 0.25$, a run of two rewards would slightly increase the variance of reward while a run of two omissions would slightly decrease the variance. These changes would result in a net contrast between the two conditions with the signal being slightly larger after two rewards than after two omissions. When $P_R = 0.75$, the runs would have the

opposite effects on variance, resulting in a net contrasts between the two conditions with the signal slightly larger after two omissions than after two rewards. When $P_R = 0.5$, runs of either two rewards or two omissions should both slightly decrease reward variance, resulting in no net contrast between the two conditions. With contrast effects combined across similar numbers of electrodes from the three groups, we would expect to see a mean contrast of 0 between the two conditions because positive contrasts in the 0.25 probability group would be offset by negative contrasts in the 0.75 probability group. The finding of a positive mean contrast with a larger signal after two rewards, therefore suggests that the signal was not a true representation of reward variance.

Rather than a true estimate of reward variance, the late-CS signals we observed may reflect the rate and magnitude of positive reward prediction error signals in the form of P_R*(R-EV), where $R = 1$ for a binary reward, and EV is an estimate of expected value. Early in training, when EV is low, the term R-EV would be close to 1, so the late-CS signal would be correlated to P_R. Later in training, EV should approach P_R, so the late-CS signal would be correlated to P_R*(1-P_R), which is equal to the reward variance.

A CS-evoked dopamine response that reflects the rate and magnitude of positive reward prediction error signals may be indicative of an anticipatory signal that is present in risky situations when a positive reward prediction error signal is likely. Such a signal could be used to focus attention in the period leading up to the receipt of new information and could even be used to adjust the learning rate for the upcoming RPE signal, as may be the case during conditions of high volatility (Behrens et al. 2007, Nasar et al. 2010), or the balance between learning rates for positive and negative errors (Niv et al. 2012). Attention plays an important role in alternative learning models, such as the Pearce-Kay-Hall model and others (Esber and Haselgrove 2011),

and correlates have been observed in the basolateral amygdala (BLA) in conditions of uncertainty in rats performing operant tasks (Roesch et al 2010). The BLA also modulates dopamine release in the accumbens through a presynaptic mechanism (Jones et al. 2010), suggesting a mechanism through which activity in the amygdala could contribute to an uncertainty correlate in the form of accumbens dopamine release. Future work will be needed to determine if indeed, BLA activity is necessary for the uncertainty signal to exist, or if an alternative mechanism is sufficient.

In the current study, we quantitatively demonstrated that CS-evoked dopamine release in the nucleus accumbens contains neural correlates of both EV and variance. We found that the signals are blended to produce the peak CS-evoked response and that the variance correlated, late-CS signal is modulated by recent outcomes in a manner not consistent with a pure variance signal. We hypothesize that the dopaminergic representation of uncertainty reflects anticipation of an upcoming positive RPE and may play a role in directing attention during learning.

Chapter 3:

Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term*

*This chapter was a originally a manuscript of the same title by Andrew S Hart, Robb B Rutledge, Paul W Glimcher, and Paul E M Phillips prepared for submission. I performed all experimental procedures, analyzed all data, and wrote the manuscript. RBR helped design the behavioral protocol and the justification for the experiment. RBR, PWG, PEMP gave advice on data analysis and edited the manuscript.

Abstract

Generating reward-predictive information from environmental stimuli and updating it through feedback is an essential aspect of adaptive behavior. Theorists have argued that dopamine encodes a reward prediction error signal (RPE) that is used in such a process. Recent work with fMRI has demonstrated that the BOLD signal in dopaminergic target areas meets both necessary and sufficient conditions for encoding this hypothesized reward prediction error. However, there has been no direct evidence that dopamine release itself also meets necessary and sufficient criteria for encoding the RPE. Further, the fact that dopamine neurons have low tonic firing rates which yield a limited dynamic range for encoding negative RPE's has led to significant debate about whether the hypothesized RPE signal encodes positive and negative prediction errors on a similar scale. Still, it may be the case that dopamine concentrations, if they encode a RPE, do so symmetrically with regard to positive and negative RPEs. To address both of these issues, we used fast-scan cyclic voltammetry at carbon fiber electrodes chronically implanted in the nucleus accumbens core of rats trained on a probabilistic decision-making task to measure reward-evoked dopamine release. We found evidence that in our task dopamine concentrations encode a RPE. Furthermore, positive and negative RPEs were symmetrically encoded, strengthening the case that dopamine concentration, acting alone, could account for reward prediction error based learning.

Introduction

Hypotheses on the function of the neurotransmitter dopamine provide a link among theories of reinforcement learning, rational choice, and motivated behavior (Schultz et al. 1997, Morris et al. 2006, Phillips et al. 2007, Gan et al. 2010, Glimcher 2011). A large number of studies have now correlated dopaminergic neural activity with the reward-prediction-error (RPE) term in temporal difference (TD) models of learning (Sutton 1988, Montague et al. 1996, Schultz et al. 1997, Waelti et al. 2001). Dopaminergic firing rates have been found to covary with this RPE term both qualitatively (Schultz et al. 1997) and quantitatively (Bayer and Glimcher, 2005.). However, these studies have failed to differentiate putative dopaminergic RPE signals from other related signals, such as salience, or surprise. In response to this uncertainty, Caplin and Dean (2007) developed a mathematically rigorous definition of RPE signals, using three axioms, which are the necessary and sufficient conditions that must be fulfilled by any signal in order for that signal to be equivalent to a RPE.

Rutledge and colleagues (2010) were the first to use the axioms to test a signal in the brain for RPE equivalence. They found that during a task in which humans received probabilistic rewards, the BOLD signal in the nucleus accumbens meet necessary and sufficient criteria for encoding an RPE. Based upon these observations they concluded that, in their task, an RPE-based learning algorithm could be driven by a neural signal in the nucleus accumbens.

While these findings confirmed that the nucleus accumbens carries a RPE-equivalent signal, they left some issues unresolved. The nature of the relationship between the BOLD signal and the extracellular dopamine concentration in a brain region is not fully understood, so it is unclear whether the RPE-equivalent signal in the accumbens is a reflection of dopamine release.

This is a critical shortcoming because dopamine is the centerpiece of the most prevalent RPE-based learning theories, such as temporal difference reinforcement learning (Schultz et al. 1997).

Here we report electrochemical measurements of dopamine release in the nucleus accumbens core of rats performing a simple set of behavioral tasks and demonstrate that these dopamine signals unambiguously encode an RPE representation in our experiments. Our quantitative measurements also indicate that dopamine release, unlike dopamine neuron firing rates, can encode positive and negative reward prediction errors on a symmetric scale.

Materials & Methods

Animals and Surgery

The University of Washington Institutional Animal Care and Use Committee approved all animal procedures used in this experiment. We implanted eight male Sprague Dawley rats with chronic carbon fiber electrodes (Clark et al. 2010) targeted bilaterally to the NAc and with Ag/AgCl reference electrodes placed between the skull and meninges. We secured implants to the skull with screws and dental cement built up into a head-cap. Head-caps also included a 6-pin Datamate connector (Harwin, Portsmouth, UK) and a nylon head post (Clark et al. 2010). After recovery, we food-restricted rats to 85-90% of their post-recovery body weight. Six of the eight rats had electrodes that could detect dopamine, as determined by inspection of cyclic voltammagram (CV) responses to food in a novel environment. One of these six rats was subsequently excluded due to separation of its head-cap

Behavioral Training

Initial training followed the procedures of Gan et al. (2010) with some modifications. We trained rats to lever press for single 45-mg food pellets (Bioserv, Frenchtown, NJ) in a free-operant paradigm and then introduced a trial-based task structure. Rats had 30s to respond after trial start and received a 20mg food pellet as the reward. Rats then trained on a deterministic task in which prizes were guaranteed with four or one pellet from the rich or poor lever, respectively. A session included 20 blocks of trials, each containing four forced-choice trials (Figure 3.1b), during which the rat could only respond on one lever (two trials of each lever per block), and four free-choice trials (Figure 3.1b), during which the rat could respond on either lever. After rats showed preference for the rich lever, we introduced a 5-s delay between response and reward. Reward was signaled by tray light onset, followed by feeder activation. The tray light remained illuminated for 3 s before the inter-trial interval (ITI) began (25 ± 5 s). When rats preferred the rich lever on at least 41 of 80 free-choice trials and successfully reversed their lever preferences in a subsequent session where the assigned contingencies were reversed, we introduced them to the probabilistic task.

In the probabilistic task rats chose between rich and poor lotteries each assigned to one of the levers. The task was designed so that there were four possible lottery-prize combinations per trial. The rich lottery yielded a prize of four pellets with a probability of 0.75 or one pellet with a probability of 0.25 and probabilities were reversed on the poor lottery.. We trained rats until they reached a criterion of 10 rich lever choices out of the last 12 and we performed voltammetry recording on the subsequent session. Recording sessions had an additional criterion that rats choose the rich lever in at least 49 out of 80 free-choice trials. If a rat failed to reach criterion on

a recording session, it was returned to the 10-of-12-criterion stage. After successful recording sessions, we reversed levers and retrained. This cycle continued until each rat had a minimum of six successful recording sessions with counterbalanced lever assignments.

Four rats completed the probabilistic task, and we returned them to the deterministic task. Again, the rich and poor levers always yielded 4 and 1 pellet, respectively. To ensure that rats fully expected deterministic rewards, we introduced an additional training session after the 10-of-12-criterion session. In the extra session, rats needed to choose the rich lever on at least 64 of 80 choice trials. After rats reached this extra criterion, we performed voltammetry on the subsequent session. We used the same 49-of-80-criterion for voltammetry sessions. The cycle continued until each rat had two voltammetry sessions with counterbalanced lever assignments. One rat required an extra cycle of training and recording because its first voltammetry session had noise associated with electrical disconnects. For that rat, we used only the last two recording sessions in the analysis. We counted rats' choices during probabilistic and deterministic sessions used for voltammetry analysis and compared individual rats' choices between the two versions using Chi-squared analysis with Fisher's exact test. We also conducted a paired t-test to compare the group's preference for the rich lever between the two versions of the task. We conducted statistics on behavior in Prism 4.0 (Graphpad Software San Deigo, CA).

Voltammetry Recording and Analysis

We recorded dopamine release in the NAc using fast-scan cyclic voltammetry through chronically implanted carbon fiber electrodes (Clark et al. 2010) at a sample rate of 10 Hz. Before voltammetry sessions, we allowed rats extra time in the chamber to condition the

electrodes. We delivered a food pellet to the feeder before each session and examined the resultant cyclic voltammogram (CV) for similarity to a dopamine CV. Behavioral session onsets also elicited dopamine-like responses. We later verified electrodes by calculating the maximum possible correlation coefficient between the pre-session-pellet or session-onset recordings and a composite of CVs evoked through electrical stimulation of dopaminergic axons. We included electrodes for which both coefficients were above 0.75 and at least one coefficient was above 0.80 for all sessions. Six electrodes implanted in the NAc of four rats that completed all behavioral sessions met this criterion (Figure 3.1a).

We reduced voltammetry data to dopamine oxidation current using background subtraction and principal components regression (PCR) against a training set of electrically evoked dopamine and ΔpH CVs with two principal components (Keithley et al. 2009). Dopamine concentration is proportional to dopamine oxidation current with a factor of approximately 30 nM / nA (Clark et al. 2010). The background for each trial was the average of 10 scans immediately preceding the end of the delay period. We conducted error analysis on all CVs and excluded CVs for which there was a significant ($\alpha = 0.05$) chance of containing a signal other than dopamine and ΔpH (Keithley et al. 2009). We detrended data from each trial by subtracting a line with its slope defined by the mean dopamine oxidation currents between -2 to -1 s and 10 to 11 s.

Dopamine theories of RPE-based reinforcement learning are primarily motivated by dopaminergic responses to unsignaled rewards (Ljungberg et al. 1992), a finding that has been reproduced with electrochemical measurements (Gan et al. 2010, Clark et al. 2010). Therefore, we isolated the best epoch for testing dopamine for RPE-equivalence as the time point likely to contain the maximum response to an unsignaled reward flanked before and after by 0.5 s. We

performed PCR as above on responses pre-session, unsignaled pellets, and identified the peak response to unsignaled rewards as the maximum dopamine signal in the 5 s following reward delivery. We then determined the latency for each (n = 58) maximum pellet response, and counted the number of maximum responses in each 0.1 s bin. We chose 2.0 s, the mode of the latency distribution, as the center of the 1.1 s epoch used in subsequent analyses.

RPE Validation

We tested mean the dopamine signal from the 1.1s window centered at 2.0 s post-reward onset against an axiomatic RPE model (Caplin and Dean 2007, Rutledge et al. 2010). The following axioms are the minimum set of assumptions that define any RPE signal $R(q \in Q, p \in P)$, where Q is the set of prizes, and P is the set of lotteries in which the probabilities of receiving prizes are fully known. When $P_q = 1$, prize q is fully predicted.

Axiom 1: Coherent Prize Dominance: $R(q,p) > R(q',p) \Rightarrow R(q,p') > R(q',p')$

Axiom 2: Coherent Lottery Dominance: $R(q,p) > R(q,p') \Rightarrow R(q',p) > R(q',p')$

Axiom 3: No-surprise Equivalence: $R(q,P_q=1) = R(q',P_{q'}=1)$

Conditional logical statements such as axioms 1 and 2 cannot be proven false if the statement on the left of the conditional arrow is false. Therefore, because it is possible for the axioms to be true for pure noise signal that does not distinguish between any lottery or prize conditions, it is necessary to distinguish between satisfying and strongly satisfying axioms 1 and 2. By strongly satisfy, we mean that the statement on the left of the conditional arrow is true, and the statement

on the right of the conditional arrow is also true. In the current experiment, there were two prizes (4 pellets and 1 pellet) and two lotteries (rich and poor). Axiom 1 would be strongly satisfied in the current experiment if 4 pellets evoked a larger response than 1 pellet on both the rich lottery (left side of Axiom 1), and the poor lottery (right side of Axiom1). Axiom 1 would be proven false in this experiment if, for example, 4 pellets evoked the larger response than 1 pellet as the outcome on one lottery, but 1 pellet evoked a larger response than 4 pellets as the outcome on the other lottery. Likewise, axiom 2 would be strongly satisfied in the current experiment if the poor lottery produced the larger response than the rich lottery for both 4 pellets (left side of Axiom 2) and 1 pellet (right side of Axiom 2). Axiom 2 would be violated if, for example, the poor lottery produced a larger response than the rich lottery to 4 pellets, but the rich lottery produced a larger response than the poor lottery to 1 pellet. The ordering of the prizes (4 > 1) or lotteries (Poor > Rich) is irrelevant to the axiomatic model, as long as ordering is consistent for prizes across lotteries (Axiom 1) and for lotteries across prizes (Axiom 2). Axiom 3 can be falsified if the signal differs between 4 pellets and 1 pellet on the deterministic task. A signal that violates any axiom cannot be a RPE. A signal for which all three axioms are true necessarily encodes a RPE-equivalent representation.

We first tested the axioms by counting the number of observations for which prize and lottery ordering were consistent and thus satisfied the axioms. We then calculated the probability that counts were due to chance using the cumulative binomial distribution for a probability of 0.5, with 12 (6 electrodes by 2 lotteries or 2 prizes) observations. We conducted a similar analysis on axiom 3 violations for which the signal to 4 fully-predicted pellets was greater than the signal to 1 fully-predicted pellet, and calculated the probability that counts were due to chance using the cumulative binomial distribution for a probability of 0.5 with 6 observations.

To more strictly test the signals against the axiomatic model, we conducted four planned paired t-tests on probabilistic dopamine (Rich-4 vs. Rich-1, Poor-4 vs. Poor-1, Rich-4 vs. Poor-4, and Rich-1 vs. Poor-1) . Significant differences in opposite directions for the Rich-4/Rich-1 and Poor-4/Poor-1 comparisons would indicate an axiom 1 violation. Significant differences in opposite directions for the Rich-4/Poor-4 and Rich-1/Poor-1 comparisons would indicate an axiom 2 violation. Significant differences in the same direction for either pair of comparisons would indicate that that axiom is strongly satisfied. We corrected α for multiple comparisons using the Holm-Bonferroni procedure. We ran a paired t-test for the effect of prize on mean deterministic task dopamine from the same epoch. A significant effect of prize in the deterministic task would indicate an axiom 3 violation.

We calculated the model RPE for each lottery-prize combination from both tasks as the difference between the average outcome on the lottery and the actual prize received. We fit the relationship between model RPE and dopamine signal with a line, a modified line, with separate slopes for the positive and negative domains, and a sigmoid function, and we used F-tests to test for significant improvements in fit.

Epoch Analysis

We conducted a set of post-hoc analyses to determine the effect of window size and window center on the RPE-equivalency of the signal, as well as the linear relationship between model RPE and dopamine concentrations. We performed the same counts of strongly satisfying observations and sets of 4 t-tests for axioms 1 and 2, as well as counts of axiom violations and t-tests for axiom 3 and linear regressions on all possible 0.5 s, 1.1 s and 1.9 s time windows (10

samples/s), between 0.1 and 5s after reward onset. For axioms 1 and 2, we calculated the conjunction P-value from the 4 t-tests as largest of the 4 P-values taken to the fourth power. This value provides an estimate of the probability of observing the expected combination of prize and lottery effects given the global null hypothesis that the true signal is not sensitive to either prize or lottery effects (Rutledge et al. 2010). We used Matlab (The Mathworks, Natick, MA) for all data processing and statistics. $\alpha = 0.05$ for all tests except where corrected for multiple comparisons.

Histology

We anaesthetized rats with 150 mg/kg ketamine. We performed electrolytic lesions through the electrodes and perfused the rats through the heart with saline, followed by paraformaldehyde (40 g/l; PFA) in phosphate buffered saline (PBS). We removed brains and stored them for at least 24 hours in 40 g/l PFA in PBS at 4 °C. We saturated brains in 300 g/l sucrose solution at 4 °C, and froze them on dry ice. We cut frozen brains in 50 μm sections on a cryostat and mounted sections on microscope slides. We then Nissl stained mounted sections and located lesion sites using an adult rat brain atlas (Paxinos and Watson 2005).

Results

Behavior

Behavior on choice trials revealed preferences for the lever that returned the higher expected value. During sessions of the probabilistic task used for voltammetry analysis, rats preferred the rich lottery, which yielded four pellets with a probability of 0.75 and one pellet with a probability of 0.25, over the poor lottery for which the probabilities of receiving one or four pellets were reversed. Rats chose the rich lever on 71.93 ± 1.09% (mean ± SD) of choice trials. Later, on the deterministic version of the task, in which the rich and poor lottery deterministically yielded four pellets or one pellet, respectively, preferences were enhanced. Rats chose the rich lever on 85.45 ± 5.93% of choice trials, which was a significant increase over the probabilistic task ($t_3 = 4.112$, P = 0.0260, paired t-test). Rats' individual enhancements of preference ranging from 7.344 to 22.57 percentage points. Individual enhancements were significant in 3 of 4 rats (P<0.001, X^2, Fisher's exact test), while one rat showed the same trend (p=0.0711, X^2, Fisher's exact test).

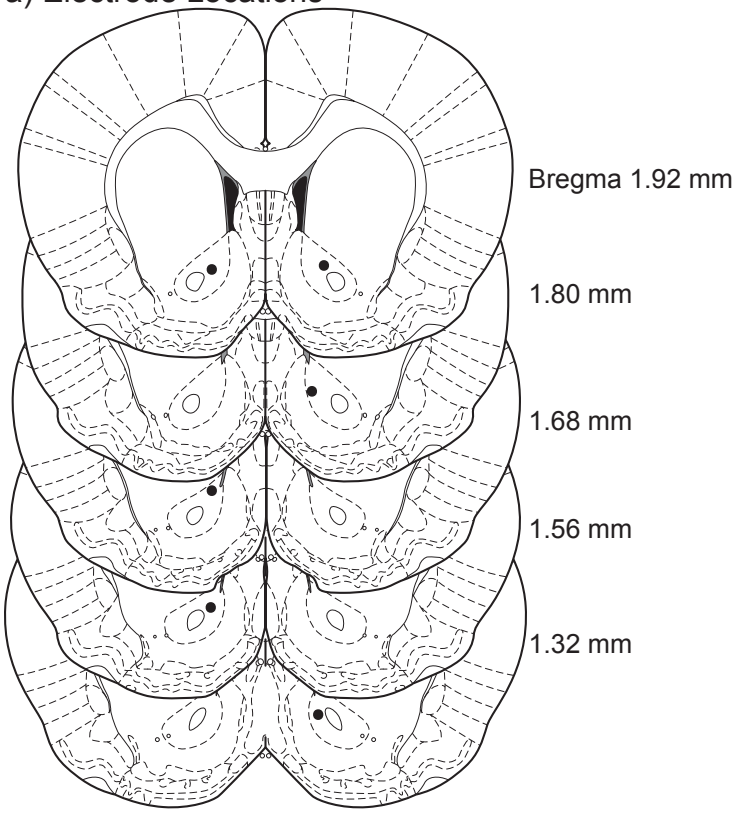
Epoch Selection

Dopamine concentration in the nucleus accumbens core (Figure 3.1a) rapidly rose in response to unsignaled rewards delivered before the beginning of behavioral sessions (Figure 3.2a). The mode of the distribution of latencies to maximum pellet-evoked dopamine was 2.0 s. Therefore, we performed subsequent analyses on mean dopamine responses from 2.0 ± 0.5 s after reward onset.

Test of Axiomatic RPE Model

Figure 3.1

a) Electrode Locations



b) Task Structure

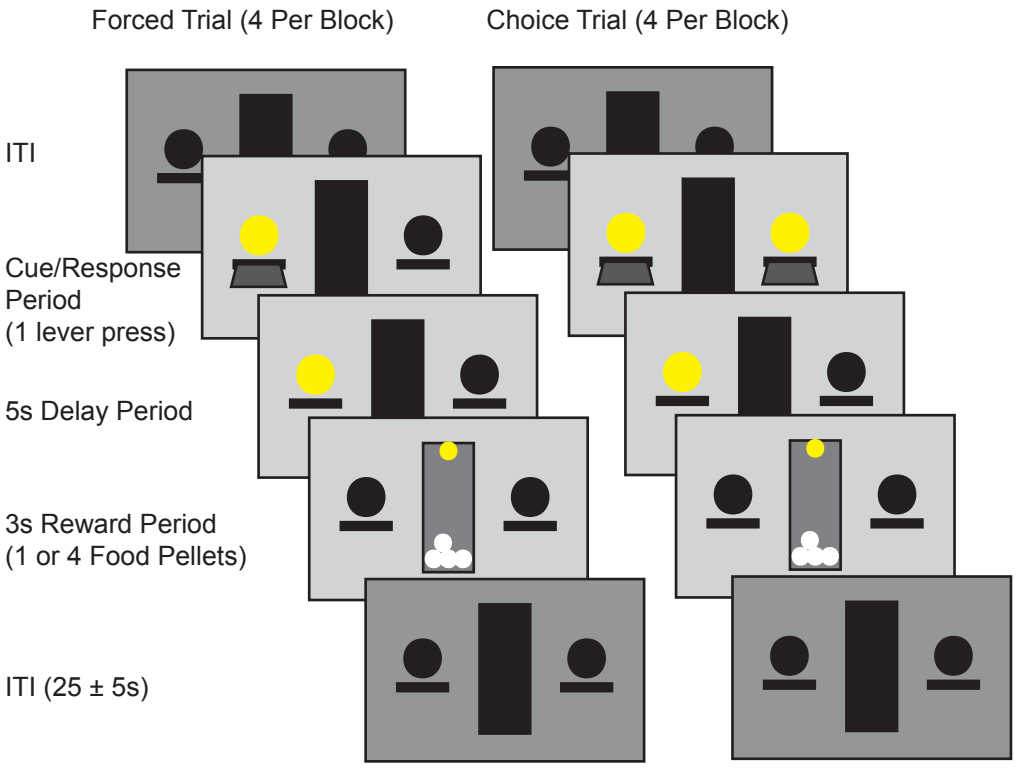


Figure 3.1: (a) Coronal sections show the locations of chronically implanted electrodes. Brain atlas sections are from Paxinos and Watson (2005). (b) The trial structure was the same in both deterministic and probabilistic tasks. A session contained 20 blocks of trials, with four forced-choice (two on each lever) followed by four free-choice trials. In the probabilistic task, responses on the rich lever resulted in four 20mg food pellets on 75% of trials and one pellet on 25% of trials. Probabilities were reversed for the poor lever. In the deterministic task, the rich and poor levers guaranteed four or one pellet, respectively.

Figure 3.2

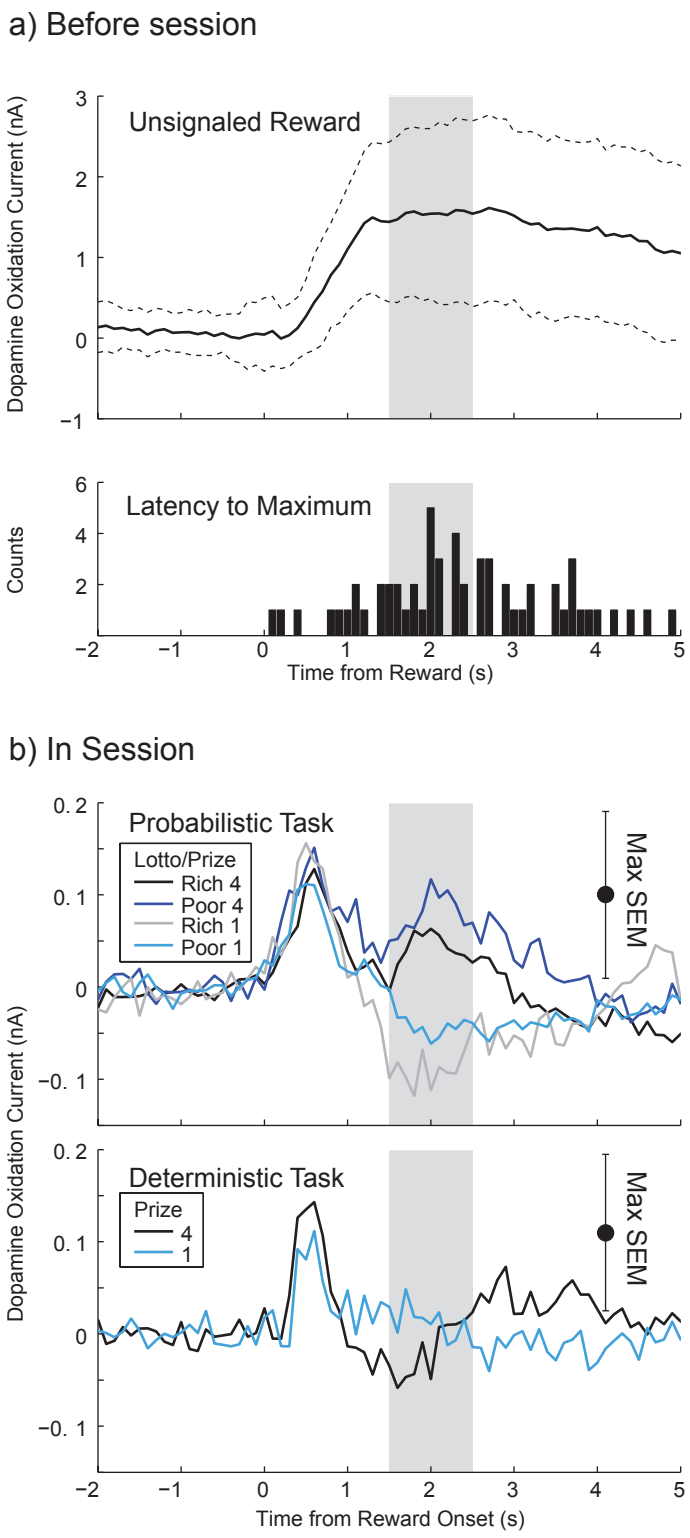

a) Before session
Unsignaled Reward
Dopamine Oxidation Current (nA)
3
2
1
0
−1
Latency to Maximum
Counts
6
4
2
−2
−1
0
1
2
3
4
5
Time from Reward (s)
b) In Session
Probabilistic Task
Lotto/Prize
Rich 4
Poor 4
Rich 1
Poor 1
Max SEM
0.2
0.1
0
−0.1
Deterministic Task
Prize
4
1
Max SEM
Dopamine Oxidation Current (nA)
Time from Reward Onset (s)

Figure 3.2: (a) The upper trace shows the mean ± SD dopamine response to an unsignaled food pellet reward delivered before the beginning of a behavioral session (n = 58, 6 electrodes). The histogram shows the latency to maximum dopamine signal for each unsignaled reward presentation. (b) Traces show the average dopamine current for n = 6 electrodes for each of the lottery-outcome combinations on forced trials during the probabilistic (upper) and deterministic (lower) tasks. Shading indicates the epoch used in subsequent analyses. Lotto: Lottery.

Dopamine release in the nucleus accumbens core met the three necessary and sufficient conditions for RPE-equivalence from the axiomatic RPE model (Caplin and Dean 2007). The average dopamine oxidation current across electrodes (n = 6 electrodes) during the reward-delivery period contained a feature during the tray light period that was characterized by a positive change in dopamine to four pellets and a negative change to one pellet for either lottery. The feature was also ordered with respect to lottery, with poor lottery trials dominating rich lottery trials for both prizes (Figure 3.2b). Average dopamine traces from the deterministic task (Figure 3.2c) showed some modulation with respect to prize, but not as strongly as in the probabilistic task and not in the same time window.

Mean dopamine responses from 2.0 ± 0.5s after reward onset in the probabilistic task were consistent with axiom 1. For axiom 1 to be satisfied, the same prize must evoke the larger dopamine response when it is the outcome of either lottery. For axiom 1 to be violated, prize dominance must differ between lotteries. Figure 3.3a shows that for all 12 observations (6 electrodes x 2 lotteries, $P < 0.001$, binomial test) the four-pellet prize evoked a larger signal than the one-pellet prize, as all points are above the equivalence line. Thus the ordering of dopamine responses from all electrodes was consistent with axiom 1.

Dopamine responses from the probabilistic task were also consistent with axiom 2. For axiom 2 to be satisfied, the same lottery must produce the larger response for both prizes. For axiom 2 to be violated, lottery dominance must differ between prizes. Figure 3.3b shows that for 10 of 12 ($P = 0.0193$, binomial test) observations, the poor lottery produced a larger signal than the rich lottery, while the rich lottery produced a larger signal on only 2 of 12 observations. Thus the ordering of observed dopamine responses was consistent with axiom 2.

Figure 3.3

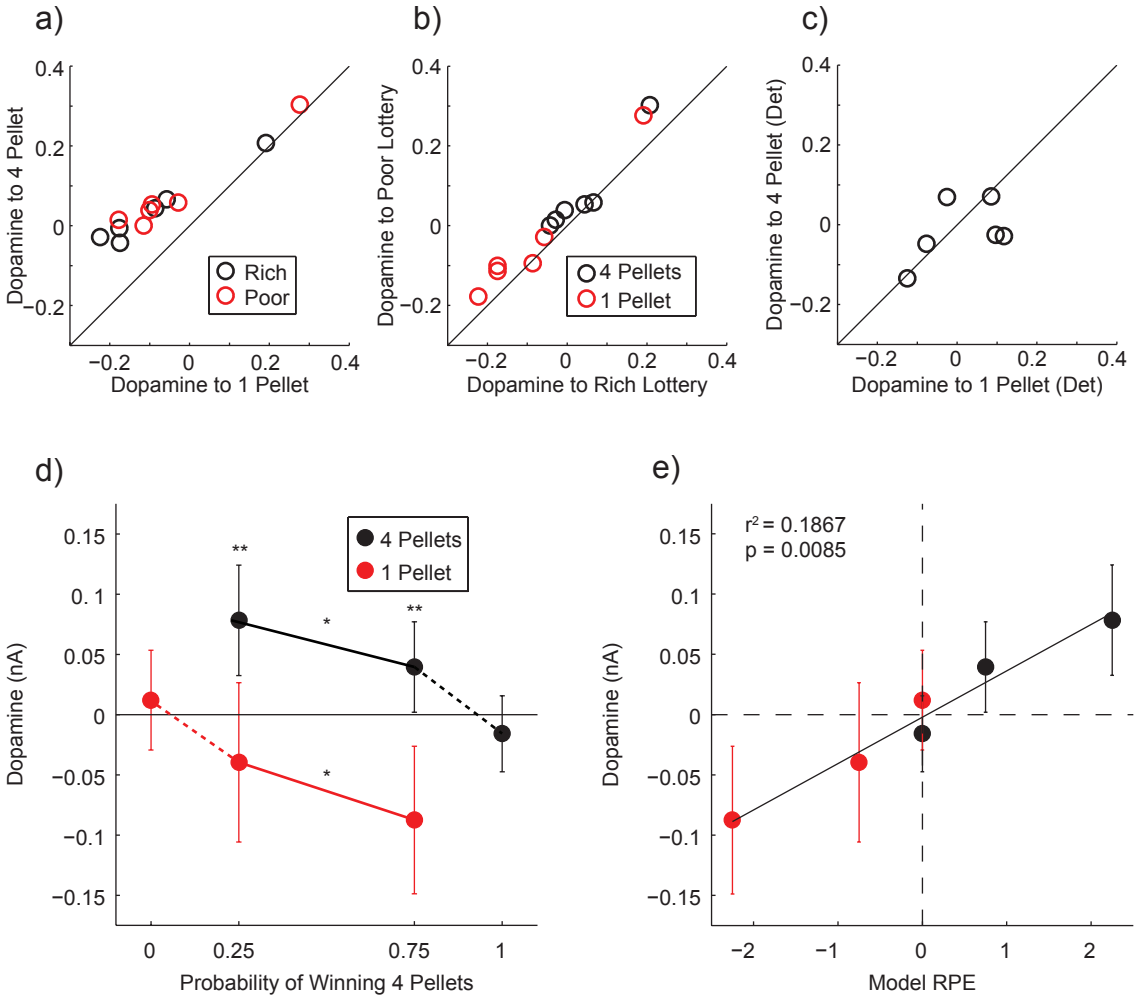

a)
b)
c)
Dopamine to 4 Pellet
Dopamine to 1 Pellet
Rich
Poor
Dopamine to Poor Lottery
Dopamine to Rich Lottery
4 Pellets
1 Pellet
Dopamine to 4 Pellet (Det)
Dopamine to 1 Pellet (Det)
d)
e)
Dopamine (nA)
Probability of Winning 4 Pellets
r² = 0.1867
p = 0.0085
Model RPE

Figure 3.3: (a) Mean dopamine to 4 pellets is plotted against mean dopamine to 1 pellet for both rich and poor lottery trials for each electrode. Points above the line indicate a greater signal to 4 pellets than to 1 pellet. (b) Mean dopamine on poor lottery trials is plotted against mean dopamine on rich lottery trials for both prizes for each electrode. Points above the line indicate a greater signal to the poor lottery than to the rich lottery. (c) Mean dopamine to four pellets is plotted against mean dopamine to one pellet on the deterministic task for each electrode. Responses are heterogeneously distributed around the equivalence line. (d) The mean dopamine signals from a-c for n = 6 electrodes are shown for both the probabilistic and deterministic tasks. Ordering of signals is consistent with the axiomatic RPE model. (e) Mean ± SEM dopamine for each possible lottery/outcome combination across the two tasks is plotted against the model RPE, calculated as the difference between reward and average lottery outcome. The line indicates a significant linear relationship. *: $P < 0.05$ for comparison between lotteries, **$P < 0.01$ for comparison between prizes, paired t-test.

Dopamine responses from the deterministic task showed no systematic deviation from axiom 3. For axiom 3 to be satisfied, responses must not differ between prizes. In Figure 3.3c, responses are heterogeneously distributed around the equivalence line with four values above the line and two values below the line. Thus, although responses from no electrodes were exactly equivalent between prizes, the observed deviations were not different from those expected due to chance (P = 0.3437, binomial test).

A more rigorous set of statistical analyses confirmed that the set of dopamine signals satisfied the axiomatic RPE model (Figure 3.3d). Four planned paired t-tests with alpha levels corrected using the Holm-Bonferroni procedure, revealed results for axioms 1 and 2 consistent with the scatter plots. The dopamine responses were larger for four pellets than for one pellet for both the rich lottery (t_5=5.0532, P = 0.0039) and the poor lottery (t_5=5.0824, P = 0.0038), satisfying axiom 1. The dopamine responses for poor lottery trials were significantly larger than responses for rich lottery trials for both the 4-pellet prize (t_5=2.6796, P = 0.0438) and the 1-pellet prize (t_5=3.4699, P = 0.0179), satisfying axiom 2. Finally, a single paired t-test of responses from the deterministic task revealed that there was not a significant effect of prize when rewards were fully predicted (t_5=0.7576, P = 0.4828), satisfying axiom 3.

Test for Imbalance of RPE Signals

Curve fits of the relationship between model RPEs, calculated as the difference between reward size and average outcome, and measured dopaminergic RPEs revealed no imbalance in encoding of positive and negative RPEs. Least squares regression showed that a line fit the data with a y-intercept of -0.0022 ± 0.0188 and a slope of 0.0384 ± 0.0138 (Figure 3.3e, r^2=0.1867,

$F_{1,34}$=7.8069, P = 0.0085). A modified line with separate slopes for the positive and negative domains did not improve the fit ($F_{1,33}$ = 0.0049, P = 0.9444), and the slopes were nearly identical for the positive and negative domains (RPE < 0: Slope = 0.0399 ± 0.0248, RPE > 0: Slope = 0.0370 ± 0.0248). As with the line the y-intercept was near zero (-0.0008 ± 0.0280). A sigmoid function also failed to improve the fit ($F_{2,32}$=0.0597, P = 0.9951).

Window Analysis

Post hoc analysis of all 118 possible 0.5, 1.1, and 1.9 s windows (10 samples/s) between 0 and 5 s post-reward revealed that there was a narrow range of windows for which the axioms could be strongly satisfied. Windows for all tested durations centered between 1.4 and 2.8 s after reward onset had at least 10 observations for which 4 pellets produced a greater mean dopamine signal than 1 pellet for a given lottery and electrode (Figure 3.4a). Of these windows, a very narrow subset, centered at 2.0 or 2.1 s after reward delivery also had at least ten observations for which the poor lottery produced a larger signal than the rich lottery for a given prize and electrode (Figure 3.4b). Conjunction P-values of t-tests testing axioms 1 and 2 for all of these windows produced conjunction P-values < 0.0001 (Figure 3.4c, 0.0001 < 0.05/472 t-tests). The low conjunction P-values suggest that the consistent ordering of signals by lottery and prize that we observed would be extremely rare if the dopamine signal was truly not modulated by the task. In none of these windows was axiom 3 violated in any systematic way, as the number of electrodes for which four pellets delivered in the deterministic task produced a greater signal than one pellet tended to vary between two and four (Figure 3.4d). P-values for associated t-tests were always greater than 0.05 (Figure 3.4e). There were windows earlier than 1 s after reward

Figure 3.4

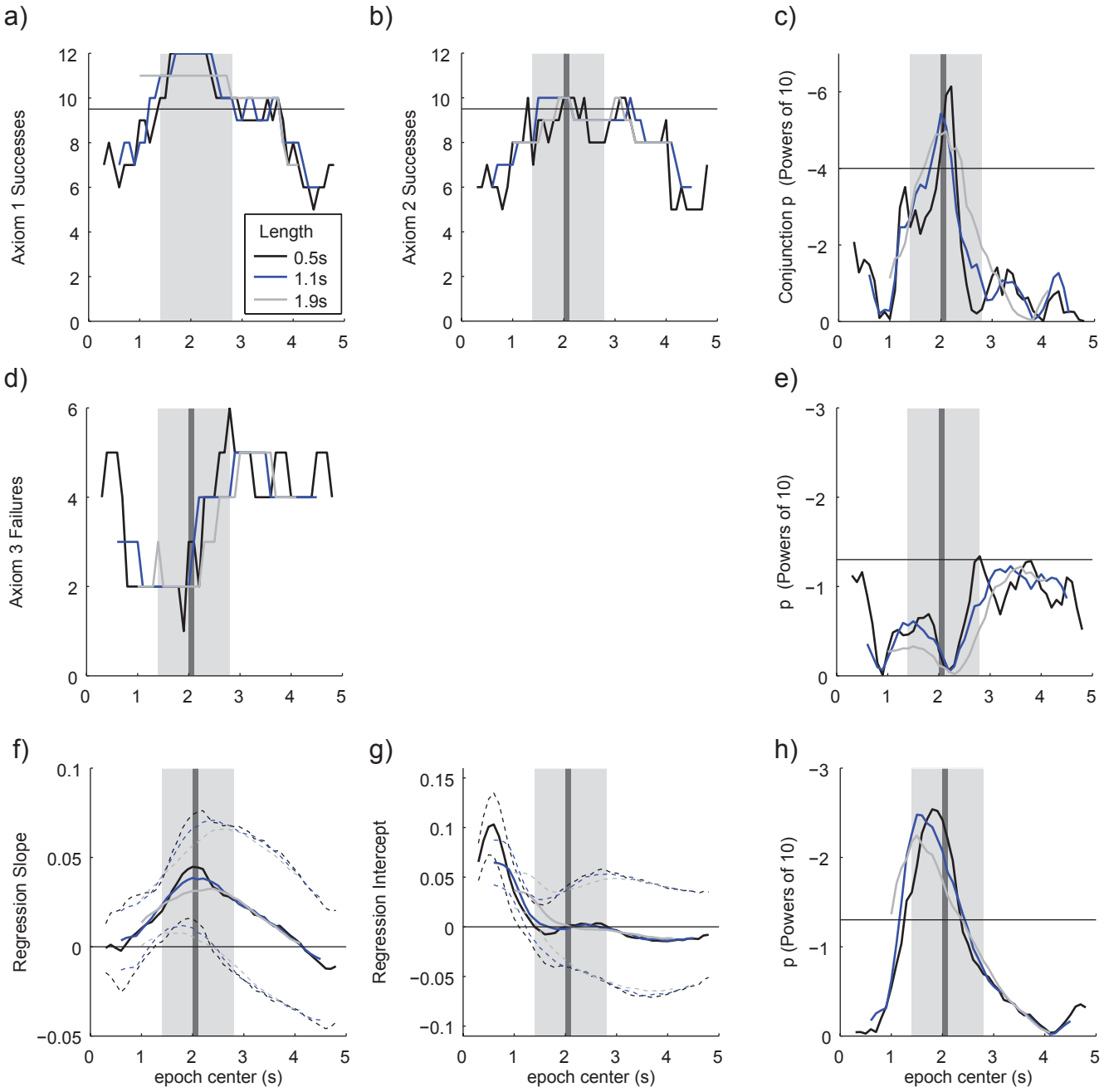

a)
b)
c)
d)
e)
f)
g)
h)
Axiom 1 Successes
Axiom 2 Successes
Conjunction p (Powers of 10)
Axiom 3 Failures
p (Powers of 10)
Regression Slope
Regression Intercept
p (Powers of 10)
epoch center (s)
epoch center (s)
epoch center (s)
Length
0.5s
1.1s
1.9s

Figure 3.4. Epoch analysis for all 118 possible 0.5, 1.1, and 1.9s windows between 0.1 and 5s after reward onset. (a) Counts of observations consistent with axiom 1 for epochs centered at each time. (b) The number of observations consistent with axiom 2 for epochs centered for each time. (c) Conjunction p-values for sets of 4 t-tests for axioms 1 and 2. Values above the line indicate conjunction $p<0.0001$. Shaded areas are from a and b. (d) Counts of observations for which 4 pellets elicits a larger response than 1 pellet in the deterministic task. (e) P values for t-tests for axiom 3. Values above the line indicates $p<0.05$. (f-g) Linear regression slopes (f) and intercepts (g) with 95% confidence intervals and p-values (h) for each epoch. Lightly shaded area indicates the times at which epochs of all 3 lengths have 10 or greater observations consistent with axiom 1. Darkly shaded area indicates times as which all epochs have 10 or greater observations consistent with axiom 2.

onset and later than 2.5 s after reward onset for which the dopamine signal violated axiom 3. These windows did not overlap with any of the windows that could strongly satisfy axioms 1 and 2.

Regressions of the mean dopamine signal from the same range of time windows revealed that the linear relationship between dopamine concentrations and model RPEs was present for a range of windows (Figure 3.4f-h) and was not dependent upon window length. Every window tested centered between 1.3 and 2.3 s after reward delivery had a significant linear regression with a y-intercept near 0. Regression slopes for this set of windows did not vary strongly between window lengths or window centers. As expected this set of windows overlapped with the set of windows that could satisfy axioms 1 and 3, and contained the set of windows that could satisfy axiom 2.

Discussion

In the present study, we used operant decision-making tasks and the axiomatic model of Caplin and Dean (2007) to identify a RPE represented by phasic dopamine release in the NAc. We designed the task so that it offered one of two lotteries which provided either of two rewards probabilistically. This design necessarily avoids the effects of salience associated with uncertainty or reward omissions (Esber and Haselgrove 2011). Both lotteries produced either one or four food pellets with a probability of either 0.75 or 0.25. The reward probability assigned to prizes was reversed between lotteries. Consistent with the axiomatic model, NAc dopamine signals in the probabilistic task were coherently modulated by prize and by lottery

while prizes did not modulate signals in a follow-up deterministic task. By strongly satisfying all 3 components of the axiomatic model (Caplin and Dean 2007) dopamine release in the NAc meets the necessary and sufficient conditions for RPE equivalence. This observation does not necessitate that the RPE signal is the only signal carried by dopamine, and it is important to note that our analysis focused on a particular epoch during which dopamine release is strongly influenced by unsignaled rewards. Post-hoc analysis of the set of 118 possible 0.5, 1.1 and 1.9 epochs centered between 0 and 5 s after reward delivery showed that this RPE equivalent is specific to windows centered at 2.0 or 2.1 s after reward onset, but that our choice of epoch length did not strongly influence the result.

Our data also allowed us to examine a second major issue in the study of dopamine and learning: the nature of the relationship between dopamine release and RPEs across the positive and negative domains. Bayer and Glimcher (2005) reported that the firing rates of putative dopamine neurons in the SNc correlated with positive but not negative RPEs. They hypothesized that a second system might exist which encoded these negative RPEs. Subsequent analyses (Bayer et al., 2007) revealed that negative RPEs were correlated with the duration of the extended inter-spike pause that follows reward delivery. Long pauses provided a signal from which negative, but not positive, RPEs could be extracted. This analysis is of particular relevance because extracellular dopamine concentration, through volume transmission, might reflect temporally integrated spike bursts and pauses due to the interplay among release, diffusion and uptake. This interplay can exist in the case of volume transmission because clearance is slower than in synaptic transmission (Garris et al. 1994). Mathematical models emphasize the non-linearity in the relationship between firing rate of dopamine neurons and extracellular dopamine concentration conferred by uptake (Wightman et al. 1988), diffusion

(Venton et al. 2003), and short-term elastic modulation of release probability (Montague et al, 2004), as well as the impact of dopamine release and uptake on dopamine receptor occupancy (Dreyer et al, 2010). Therefore, we tested the relationship between dopamine signals and model RPEs, calculated as the difference between the prize magnitude and average outcome on the lottery. We found that the relationship between RPEs and extracellular dopamine concentration was best fitted by a straight line with a y-intercept near zero. Neither a dual slope function, nor a sigmoid function significantly improved the fit. These findings show that the dopaminergic RPE signals are linear across the positive and negative domains in our task. Therefore, positive and negative dopaminergic RPEs were represented symmetrically at the level of dopamine concentration without the rectification observed in firing rate of dopamine neurons (Bayer & Glimcher, 2005). While it does not rule out the possibility that opponent neural systems cooperate to compute motivationally relevant variables, the current work demonstrates that dopamine transmission is sufficient to provide a bidirectional teaching signal to the NAc, and that a complimentary opponent system to deliver negative RPEs would not be necessary in the range of RPEs we observed.

The exact mechanism that produces dopaminergic RPE signals is still unknown. Models that take into account network connectivity of the basal ganglia an cortex such as that proposed by Tan and Bullock (2008) are needed to illuminate this area. The current study should serve to inform models of dopaminergic signaling by illustrating that positive and negative RPEs are encoded in a balanced manner on the level of NAc dopamine release.

Chapter 4:

Discussion

In this work, I have demonstrated that dopamine release in the rat nucleus accumbens is correlated to the economic parameters of expected value and uncertainty. I have also demonstrated, using an axiomatic approach, that accumbens dopamine release is a reward prediction error signal. These analyses offer strong evidence that dopamine release quantitatively encodes variables for decision-making and updating expectations, key parameters in the framework of neuroeconomics (Rangel et al. 2008). However, the current work has unresolved issues worthy of discussion. First, there is a difference between the operant and Pavlovian tasks used in terms of the balance between positive and negative RPEs, and second, the role of the uncertainty signal in behavior is still unresolved.

Addressing the Balance Between Positive and Negative RPEs

Responses to rewards and reward omissions in the Pavlovian task in chapter 2 do not entirely share the characteristics of the reward prediction error signals from the operant task in chapter 3. The balance in encoding between positive and negative RPE's in chapter 3 is not apparent in the data from chapter 2. This fact is highlighted when responses to rewards and reward omissions from session 6 of the Pavlovian task (Figure 4.1a) and reward responses from the operant task (Figure 4.1b) are plotted with respect to their model RPE's. I repeated the model comparisons between the linear and dual slope functions from chapter 3 for the reward and omission dopamine responses from chapter 2. I estimated model RPE's as $1\text{-}P_R$ for positive RPEs and $-1*P_R$ for negative RPEs where P_R equals probability of reinforcement. Linear regression of mean dopamine response vs. model RPE for reward responses in the Pavlovian task showed a significant linear relationship (slope = 0.3887 ± 0.0577, intercept = 0.0990 ± 0.0318).

However, the intercept was significantly greater than zero (t_{42} = 3.1132, P = 0.0033), which is a sign of an imbalance between the positive and negative domains. The imbalance was confirmed when a dual slope function with different slopes for the domains of positive and negative reward prediction errors, significantly improved the model fit ($F_{1,41}$ = 5.442, P = 0.0246). The intercept for the dual slope function was 0.0060 ± 0.0504, and the slopes for the positive and negative domains were 0.5621 ± 0.0932 and 0.1148 ± 0.1310, respectively. These slopes were significantly different from each other (t_{41} = 3.9902, P = 0.0003), and the slope for the negative domain was not significantly different from zero (t_{41} = 0.8763, P = 0.3860).

The finding of an imbalance in encoding between positive and negative reward prediction errors between the two tasks suggests that the balance in encoding may depend on the specifics of the task, such as whether the task is operant or Pavlovian, and the length of the delay between cue-onset or response and reward. The model RPEs we estimated were task specific, so it is possible that even though the model RPEs from the operant task span a larger range in pellets, the magnitude of actual RPEs experienced by the subjects may in fact be smaller than that in the Pavlovian task. Dopamine release varied over a much smaller range overall in the operant task than in Pavlovian task, which is consistent with this interpretation. Therefore, it is possible that when RPEs are smaller, they can be encoded in a linear fashion, but there is an upper bound, over which RPEs are imbalanced in the fashion we observed in the Pavlovian task, with a wide dynamic range for positive RPEs and a narrow one for negative RPEs. This upper bound could be determined by tonic dopamine release and uptake.

An alternative explanation for the imbalance in the Pavlovian task but not in the operant task is that negative RPEs were produced by reward omissions in the Pavlovian task, while they were produced by smaller rewards in the operant task. Reward omissions may themselves carry

information that is conveyed by dopamine neurons. Esber and Hasselgrove (2011) have proposed a model of learning in which reward and no-reward conditions both increase the acquired salience of a conditioned cue. Acquired salience here is used to describe the ability of a cue to attract attention beyond its physical properties, such as brightness or loudness. If this model is instantiated by dopamine release, it might be the case that while predictions differed between probability groups, the no-reward condition on omission trials carried the same amount of salience across all probability groups, thus producing the same signal. This signal appears near 0 in Figure 4.1a. However, because the baseline for that signal was taken during the last 1s of the CS, it is possible that the no-reward signal is just an extension of the late-CS dopamine signal, which varied with respect to probability according to the second-order polynomial model, with its highest level at 0.5.

While the specifics of the balance between positive and negative RPEs may vary depending on experimental conditions, and while the effect of reward omissions on dopamine signaling needs to be more directly tested, findings from both studies in this work support the hypothesis that peak reward-evoked dopamine release is a reward prediction error signal. The axiomatic analysis in chapter 3 provided a falsifiable framework for testing the reward prediction error hypothesis on accumbens dopamine release. Over the range of prizes and lotteries tested in chapter 3, dopamine release strongly satisfied the axioms. This finding indicates that dopamine could be used as a teaching signal to update expected values based upon outcomes. However, causality has not been addressed. In order to directly assess the role of reward-evoked dopamine release as a teaching signal, the firing rate of dopamine neurons needs to be silenced or enhanced during reward delivery. Access to optogenetic techniques will allow experimenters to perform this manipulation and assess its influence on behavior. If dopamine reward prediction error

Figure 4.1

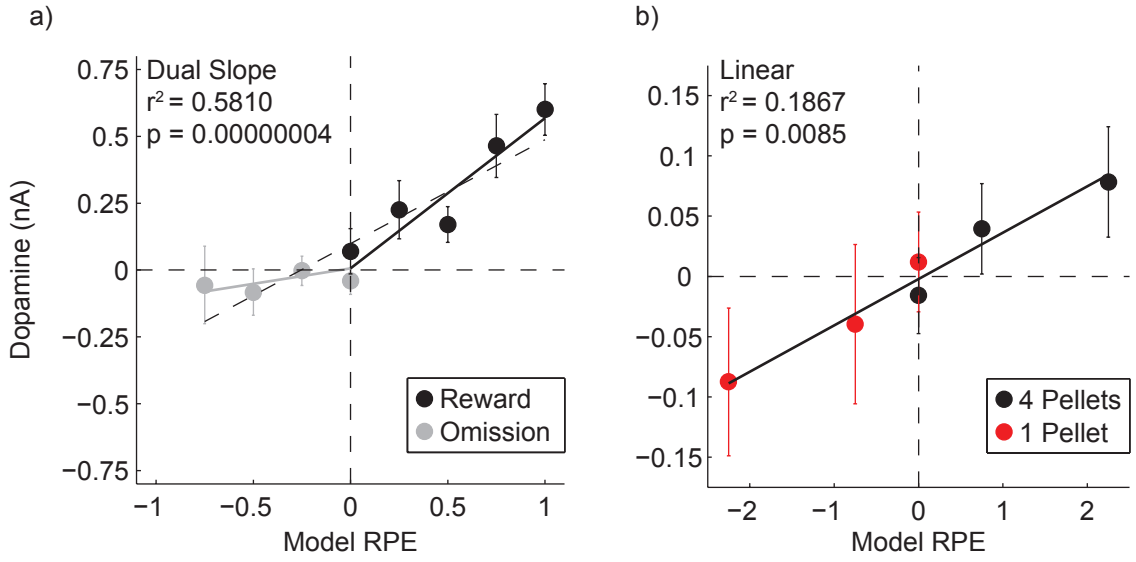

a)
Dual Slope
r² = 0.5810
p = 0.00000004
Reward
Omission
Dopamine (nA)
Model RPE
b)
Linear
r² = 0.1867
p = 0.0085
4 Pellets
1 Pellet
Model RPE

Figure 4.1: (a) Reward-evoked and reward omission dopamine release from session 6 of the Pavlovian task in chapter 2 is plotted with respect to model reward prediction error. A dual-slope function with separate slopes for positive (black) and negative (gray) domains significantly improved the fit over a straight line (dashed). (b) Reward-evoked dopamine release from the operant task in chapter 2. In this case, the straight line fit the data (solid), and the dual slope function did not significantly improve the fit (see chapter 3).

signals are used in learning, silencing them during reward delivery should produce extinction in Pavlovian or operant tasks. Alternatively, increasing their firing rates during delivery of some rewards should produce preferences for rewards paired with optical stimulation in a manner consistent with reinforcement learning models. By clever manipulation, experimenters would be able to artificially produce imbalances between positive and negative reward prediction errors. If dopamine RPEs are truly used as a teaching signal to update an estimate expected value, this manipulation would have predictable effects on decision-making under risk, as shown by Clark, Nasrallah et al. (2012).

Does Uncertainty Make Conditioned Cues More Motivating?

My findings that peak CS-evoked dopamine release during partial reinforcement is correlated to both expected value and uncertainty, and that the peak CS-evoked signal is largest in the 0.50 probability group, suggested that uncertainty may make the CS more motivating. Schultz and colleagues speculated that the dopamine response they recorded under uncertainty in monkeys may promote risk-seeking by enhancing the appeal of uncertain stimuli (Fiorillo et al. 2003). In order to test this hypothesis, I conducted a conditioned reinforcement experiment. Conditioned reinforcement tasks, in which an animal works for a Pavlovian cue that has previously been paired with reward, are thought to assess the motivational salience associated with a conditioned cue (Robinson and Flagel 2009). Therefore, we hypothesized that if the uncertainty-correlated dopamine release confers an uncertainty bonus to cues that make them more motivating, then the conditioned cues that are partially reinforced should elicit more responses in a conditioned reinforcement task.

I ran two cohorts of rats in this experiment, with the cohorts only differing with respect to the length of the CS presentation during the conditioned reinforcement session. For each cohort of 24 male Sprague-Dawley rats, I placed rats into 0.5 (n = 12) and 1.0 (n = 12) probability groups, counterbalanced by weight. I then conducted training as in chapter 2, with three sessions of magazine training, followed by six sessions of Pavlovian training. In this case, I ran sessions on consecutive days, and I did not conduct voltammetry recordings. After Pavlovian training, I conducted a conditioned reinforcement (CR) session the following day. For the CR session, I reconfigured the behavioral chamber, so that a light and lever replaced the feeder in to the center of the wall and were flanked by two nose poke holes. Responses at one nose poke hole, the active hole, would produce a presentation of the light and lever CS. In one cohort, the CS presentation was 2 s in duration, and in the other cohort the CS presentation was 8 s in duration. We used the two durations to control for the possibility that a failure to see an enhancement of nose poke behavior in the 0.5 probability group would be due to the lack of a 2 s CS presentation to elicit the full uncertainty-correlated dopamine response. For each rat, I assigned the active nose poke hole on the opposite side of the chamber as the CS had been during Pavlovian training. Conditioned reinforcement sessions were 40 minutes in duration, and the number of active and inactive nose poke responses were counted. I tested for effects of reinforcer duration (2 s vs 8 s) and Pavlovian reinforcement probability (0.5 or 1.0) and their interaction using a 2 way ANOVA conducted in Prism 4.0 (Graphpad Software, San Diego, CA).

Uncertainty did not enhance the motivational salience of the CS, as rats in the 0.5 probability group did not make more active nose poke responses than rats in the 1.0 probability group in either the 2 s or the 8 s cohort (Figure 4.2a). There were significant effects of P_R ($F_{1,44}$ = 5.989, P = 0.0185) and reinforcer duration ($F_{1,44}$ = 29.41, P = 0.000002), as well as an

Figure 4.2

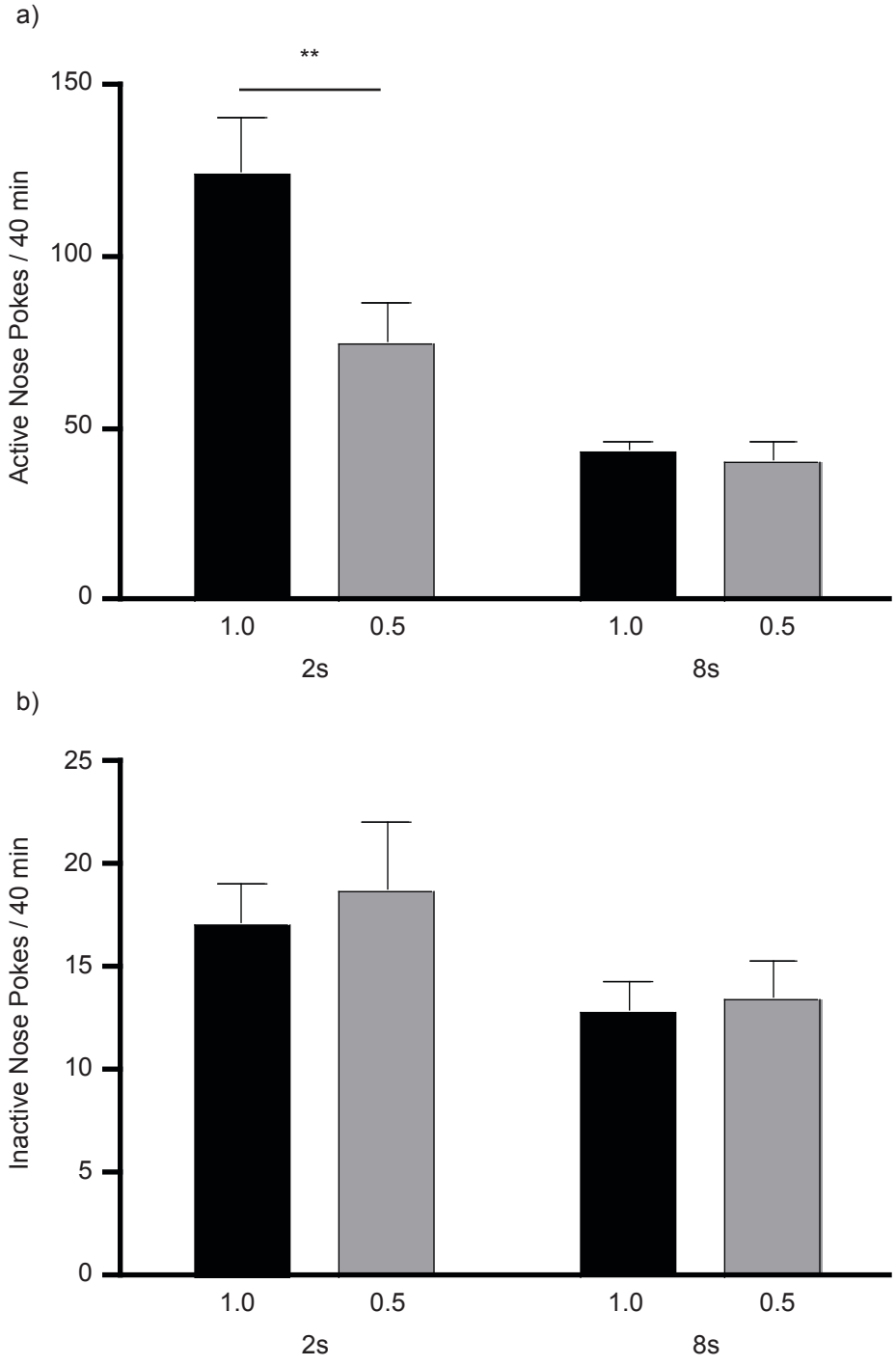

a)
**
150
100
50
0
Active Nose Pokes / 40 min
1.0
0.5
1.0
0.5
2s
8s
b)
25
20
15
10
5
0
Inactive Nose Pokes / 40 min
1.0
0.5
1.0
0.5
2s
8s

Figure 4.2: (a) Active nose pokes during a 40 min conditioned reinforcement session are shown for the 1.0 and 0.5 probability groups for cohorts with 2 s and 8 s conditioned stimulus exposure during conditioned reinforcement session. (b) Inactive nose pokes for the same groups. (**: $P<0.01$, $n = 12$ per group)

interaction effect ($F_{1,44}$ = 4.834, P = 0.0332) on the number of active nose pokes responses, but the direction of the effects was not consistent with an uncertainty bonus. In the 2 s cohort, the 1.0 group made about 1.6 times as many active nose poke responses as the 0.5 group (t = 3.284, P < 0.01, Bonferroni correction), while in the 8 s cohort and there was no difference in active nose poke responses between groups (t = 0.1758, P > 0.05, Bonferroni correction). These results suggest that the lower level of active nose pokes in 8 s cohort produced a floor effect that occluded the difference between probability groups. The 2 s cohort made more active nose poke responses overall than the 8 s cohort, and there was a much more dramatic increase in responding for the 1.0 group than for the 0.5 group, resulting in a larger difference between the groups. Inactive nose pokes did not differ across reinforcer duration (Figure 4.2b, $F_{1,44}$ = 4.022, P = 0.0511) or P_R conditions ($F_{1,44}$ = 0.2092, P = 0.6496), and there was no interaction effect ($F_{1,44}$ = 0.04457, P = 0.8338), suggesting that the difference in responses between the cohorts and the difference between the 1.0 and 0.5 probability groups in the 2 s cohort were related to differences in motivated behavior rather than general activity levels. The lack of an enhancement in conditioned responding for an uncertainty-associated Pavlovian cue suggests that the peak- and late-CS dopamine responses do not play a role in representing motivational salience, and in this case might not serve as an uncertainty bonus. However, an uncertainty bonus might be difficult to identify in rats because they tend to be naturally risk averse, choosing certain options over risky options of the same average outcome (Cardinal and Howes 2005, Nasrallah et al. 2009, Nasrallah et al. 2011), and breakpoint data suggests that their utility function for food rewards is concave down (Clark, Nasrallah et al. 2012), which necessarily produces risk aversion. An alternative interpretation of the lack of an effect of probability on conditioned reinforcement in the 8 s cohort is that the dopamine signal to the full 8 s CS confers

enough of a bonus to make up the disparity in motivation between the 0.5 and 1.0 probability groups. However, this interpretation lacks the parsimony of the floor effect described above.

Conclusion

In the current work, I used fast scan cyclic voltammetry to record dopamine release of rats performing Pavlovian and operant tasks. I then used regression analyses and an axiomatic analysis to quantitatively test the hypotheses that dopamine represents probability, uncertainty, and reward prediction errors. I found that dopaminergic representations of probability and uncertainty vary over training, and that after both probability and uncertainty are learned, the two representations are blended to produce a peak-CS evoked dopamine response. I also found that the positive effect of recent rewards compared to recent reward omissions enhances the uncertainty-correlated signal, across probabilities, in spite of disparate effects on reward variance between probability groups. This finding suggests that the uncertainty representation is not a true representation of variance and is more likely an anticipatory signal for an upcoming positive reward prediction error. However, I failed to demonstrate a behavioral consequence of this signal. Regarding reward prediction error signals, dopamine release to rewards in both the Pavlovian and operant task had the qualities of a RPE. The axiomatic analysis of data from the operant task in chapter 3 showed that this signal meets the necessary and sufficient conditions to function as a RPE. In the case of the operant task, the signals were balanced between positive and negative RPEs, suggesting linear encoding over the range of errors observed. However, a similar analysis of data from the Pavlovian task in chapter 2 revealed that this linear encoding differs depending upon the task used or the range of dopamine signals examined. Future work

using timely manipulations of dopamine signaling, such as those possible with optogenetics, will be necessary to determine the function of uncertainty-correlated CS-evoked dopamine signal, as well as whether dopaminergic RPE signals are truly used to update expected values associated with conditioned stimuli.

References

Aberman JE, Salamone JD. (1999) Nucleus accumbens dopamine depletions make rats more sensitive to high ratio requirements but do not impair primary food reinforcement. Neuroscience. 92:545-552.

Bayer HM, Glimcher PW. (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129-141.

Bayer HM, Lau B, Glimcher PW. (2007) Statistics of midbrain dopamine neuron spike trains in the awake primate. J Neurophysiol 98:1428-1439.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF. (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214-1221.

Berns GS, McClure SM, Pagnoni G, Montague PR. (2001) Predictability modulates human brain response to reward. J Neurosci. 21:2793-2798.

Caplin A, Dean M. (2007) Dopamine, reward prediction error, and economics. Quarterly Journal of Economics 123:663-701.

Cardinal RN, Howes NJ. (2005) Effects of lesions of the nucleus accumbens core on choice between small certain rewards and large uncertain rewards in rats. BMC Neurosci. 6:37.

Clark JJ, Hollon NG, Phillips PE. (2012) Pavlovian valuation systems in learning and decision making. Curr Opin Neurobiol. [Epub ahead of print]

Clark JJ, Nasrallah NA, Hart AS, Collins AL, Bernstein IL, Phillips PE. (2012) Altered risk-based decision making following adolescent alcohol use results from an imbalance in reinforcement learning in rats. PLoS One 7:e37357.

Clark JJ, Sandberg SG, Wanat MJ, Gan JO, Horne EA, Hart AS, Akers CA, Parker JG, Willuhn I, Martinez V, Evans SB, Stella N, Phillips PE. (2010) Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. Nat Methods 7:126-129.

Dreyer JK, Herrik KF, Berg RW, Hounsgaard JD. (2010) Influence of phasic and tonic dopamine release on receptor activation. J Neurosci 30:14273-14283.

Esber GR, Haselgrove M. (2011) Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. Proc Biol Sci. 278:2553-61.

Fiorillo CD. (2011) Transient activation of midbrain dopamine neurons by reward risk. Neuroscience 197**:**162-171.

Fiorillo CD, Tobler PN, Schultz W. (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898-1902.

Fiorillo CD, Tobler PN, Schultz W. (2005) Evidence that the delay-period activity of dopamine neurons corresponds to reward uncertainty rather than backpropagating TD errors. Behav Brain Funct 1:7.

Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PE, Akil H. (2011) A selective role for dopamine in stimulus-reward learning. Nature 469:53-57.

Fujiwara J, Tobler PN, Taira M, Iijima T, Tsutsui K. (2008) Personality-dependent dissociation of absolute and relative loss processing in orbitofrontal cortex. Eur J Neurosci 27:1547-1552.

Gan JO, Walton ME, Phillips PE. (2010) Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci 13:25-27.

Garris PA, Christensen JR, Rebec GV, Wightman RM. (1997) Real-time measurement of electrically evoked extracellular dopamine in the striatum of freely moving rats. J Neurochem. 68:152-161.

Garris PA, Ciolkowski EL, Pastore P, Wightman RM. (1994) Efflux of dopamine from the synaptic cleft in the nucleus accumbens of the rat brain. J Neurosci 14:6084-6093.

Gilboa I. (2009) *Theory of Decision Under Uncertainty*. Cambridge University Press: Cambridge, UK. pp 104-107.

Glimcher PW. (2011) Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc Natl Acad Sci U S A. 108 Suppl 3:15647-54.

Jackson TE, Mudholkar GS. (1979) Control procedures for residuals associated with principal component analysis. Technometrics 21:341-349.

Jones JL, Day JJ, Aragona BJ, Wheeler RA, Wightman RM, Carelli RM. (2010) Basolateral amygdala modulates terminal dopamine release in the nucleus accumbens and conditioned responding. Biol Psychiatry 67:737-44.

Kahneman, D, Tversky A. (1979) Prospect Theory: An Analysis of Decision under Risk. Econometrica. 47:263-291.

Keithley RB, Heien ML, Wightman RM. (2009) Multivariate concentration determination using principal component regression with residual analysis. Trends Analyt Chem 28:1127-1136.

King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR. (2008) The rupture and repair of cooperation in borderline personality disorder. Science. 321:806-810.

Kishida KT, Sandberg SG, Lohrenz T, Comair YG, Sáez I, Phillips PE, Montague PR. (2011) Sub-second dopamine detection in human striatum. PLoS One. 6:e23291.

Knutson B and Bossaerts P. (2007) Neural antecedents of financial decisions. *Journal of Neuroscience*. 27:8174-7.

Knutson B, Wimmer GE, Kuhnen CM, Winkielman P. (2008) Nucleus accumbens activation mediates the influence of reward cues on financial risk taking. Neuroreport. 19:509-513.

Kuhnen CM, Knutson B. (2005) The neural basis of financial risk taking. Neuron. 47:763-770.

Lee D. (2008) Game theory and neural basis of social decision-making. Nature Neuroscience. 11:404-408.

Ljungberg T, Apicella P, Schultz W. (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 67:145-63.

McClure SM, Laibson, DI, Loewenstein G, and Cohen JD. (2004) Separate neural systems value immediate and delayed monetary rewards. *Science*. 306:503-507.

Montague PR, Dayan P, Sejnowski TJ. (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936-1947.

Montague PR, McClure SM, Baldwin PR, Phillips PE, Budygin EA, Stuber GD, Kilpatrick MR, Wightman RM. (2004) Dynamic gain control of dopamine delivery in freely moving animals. J Neurosci 24:1754-9.

Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. (2006) Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 9:1057-1063.

Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O. (2004) Dopamine neurons can represent context-dependent prediction error. Neuron 41:269-280.

Nasrallah NA, Clark JJ, Collins AL, Akers CA, Phillips PE, Bernstein IL. (2011) Risk preference following adolescent alcohol use is associated with corrupted encoding of costs but not rewards by mesolimbic dopamine. Proc Natl Acad Sci U S A. 108:5466-5471.

Nasrallah NA, Yang TW, Bernstein IL. (2009) Long-term risk preference and suboptimal decision making following adolescent alcohol use. Proc Natl Acad Sci U S A. 106:17600-17604.

Nassar MR, Wilson RC, Heasly B, Gold JI. (2010) An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J Neurosci. 30:12366-12378.

Niv Y, Duff MO, Dayan P. (2005) Dopamine, uncertainty, and TD learning. Behav and Brain Funct 1:6.

Niv Y, Edlund JA, Dayan P, O'Doherty JP. (2012) Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J Neurosci 32:551-562.

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. (2003) Temporal difference models and reward-related learning in the human brain. Neuron 38:329-337.

Padoa-Schioppa C, Assad JA. (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223-226.

Pan WX, Schmidt R, Wickens JR, Hyland BI. (2005) Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J Neurosci 25:6235-6242.

Parker JG, Zweifel LS, Clark JJ, Evans SB, Phillips PE, Palmiter RD. (2010) Absence of NMDA receptors in dopamine neurons attenuates dopamine release but not conditioned approach during Pavlovian conditioning. Proc Natl Acad Sci U S A. 107:13491-13496.

Paulus MP, Rogalsky C, Simmons A, Feinstein JS, Stein MB. (2003) Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. Neuroimage. 19:1439-1448.

Paxinos G, Watson C. (2005) The Rat Brain in Stereotaxic Coordinates, Ed 5. London: Elsevier Academic Press.

Phillips MI, Olds J. (1969) Unit activity: motiviation-dependent responses from midbrain neurons. Science 165:1269-1271.

Phillips PE, Stuber GD, Heien ML, Wightman RM, Carelli RM. (2003) Subsecond dopamine release promotes cocaine seeking. Nature. 422:614-618.

Phillips PEM, Walton ME, Jhou TC. (2007) Calculating utility: Preclinical evidence for cost-benefit analysis by mesolimbic dopamine. Pyschopharmacology (Berl) 191:483-495.

Platt ML, Glimcher PW. (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233-238.

Preuschoff K, Bossaerts P, Quartz SR. (2006) Neural differentiation of expected reward and risk in human subcortical structures. Neuron 51:381-390.

Rangel A, Camerer C, Montague PR. (2008) A framework for studying the neurobiology of value-based decision-making. Nature Reviews Neuroscience 9:545-556.

Redish AD. (2004) Addiction as a computational process gone awry. Science. 306:1944-1947.

Rescorla, RA, Wagner, AR. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.H. & Prokasy, W.F. (Eds), *Classical Conditioning II: Current Research and Theory*. Appleton-Century-Crofts, New York, NY, pp. 64–99.

Robinson TE, Flagel SB. (2009) Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. Biol Psychiatry. 65:869-873.

Roesch MR, Calu DJ, Esber GR, Schoenbaum G. (2010) Neural correlates of variations in event processing during learning in basolateral amygdala. J Neurosci 30:2464-71

Roesch MR, Calu DJ, Schoenbaum G. (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci. 10:1615-1624.

Roitman MF, Stuber GD, Phillips PE, Wightman RM, Carelli RM. (2004) Dopamine operates as a subsecond modulator of food seeking. J Neurosci. 24:1265-1271.

Rutledge R, Dean M, Caplin A, Glimcher P. (2010) Testing the reward prediction error hypothesis with an axiomatic model. J Neurosci 30:13525-13536.

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. (2003) The neural basis of economic decision-making in the Ultimatum Game. Science. 300:1755-1758.

Schultz W, Dayan P, Montague PR. (1997) A neural substrate of prediction and reward. Science 275:1593-1599.

Stuber GD, Klanker M, de Ridder B, Bowers MS, Joosten RN, Feenstra MG, Bonci A. (2008) Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. Science 321:1690-1692.

Sugam JA, Day JJ, Wightman RM, Carelli RM. (2012) Phasic nucleus accumbens dopamine encodes risk-based decision-making behavior. Biol Psychiatry. 71:199-205.

Sugrue LP, Corrado GS, Newsome WT. (2005) Choosing the greater of two goods: neural currencies for valuation and decision-making. Nature Reviews Neuroscience. 6:363-375.

Sutton RS. (1988) Learning to predict by the methods of temporal differences. Machine Learning 3:9-44.

Tan CO, Bullock D. (2008) A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. J Neurosci 28:10062-10074.

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

Tobler PN, Fiorillo CD, Schultz W. (2005) Adaptive coding of reward value by dopamine neurons. Science 307:1642-1645.

Tobler PN, Fletcher PC, Bullmore ET, Schultz W. (2007) Learning-related human brain activations reflecting individual finances. Neuron. 54:167-175.

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

van der Meer MA, Redish AD. (2009) Covert Expectation-of-Reward in Rat Ventral Striatum at Decision Points. Front Integr Neurosci. 3:1.

van 't Wout M, Kahn RS, Sanfey AG, Aleman A. (2005) Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. Neuroreport. 16:1849-1852.

Venton BJ, Zhang H, Garris PA, Phillips PE, Sulzer D, Wightman RM. (2003) Real-time decoding of dopamine concentration changes in the caudate-putamen during tonic and phasic firing. J Neurochem 87:1284-95.

Volkow ND, Fowler JS, Wang GJ, Baler R, Telang F. (2008) Imaging dopamine's role in drug abuse and addiction. Neuropharmacology. 56 Suppl 1:3-8.

von Neumann J and Morgenstern O. (1944) *Theory of Games in Economic Behavior.* Princeton University Press: Princeton, NJ.

Waelti P, Dickinson A, Schultz W. (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43-48.

Wightman RM, Amatore C, Engstrom RC, Hale PD, Kristensen EW, Kuhr WG, May LJ. (1988) Real-time characterization of dopamine overflow and uptake in the rat striatum. Neuroscience 1988 25:513-23.