Valuation and decision-making in cortical-striatal circuits

by
Jeffrey R. Pettibone

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
2016

Doctoral Committee:
Professor Joshua D. Berke, Chair
Professor Brandon J. Aragona
Professor Kent C. Berridge
Professor Robert T. Kennedy

Dedicated to my mother.

“If it’s anything at all, it’s nothing.”

Acknowledgements

I would first like to thank the members of my committee for their service toward my growth as a scholar. I began this path 13 years ago as an undergraduate in Kent Berridge's Introduction to Biopsychology. His excitement for neuroscience is enduring and contagious, and these pages would not exist without him. Brandon Aragona is my spirit animal, and my dynamic sphere. Bob Kennedy trusted in my work, sample after sample. Josh Berke has tremendously shaped the way I think. He has taught me the value of clarity, and given me the confidence to meet any challenge. Thanks Josh.

During my time at Michigan, I have had the pleasure of working with some exceptional people. Dan Leventhal, Fujun Chen, and Michael Farries deserve my thanks for their mentorship and great conversations. Ali Mohebi has been a true friend and a great person to bounce ideas off of. Nico Mallet and his terrific family kept me sane and positive in my early years—the best people you'd ever want to meet. Jonte Jones is always leading the finest example anywhere. Vaughn Hetrick and Marc Bradshaw taught me how to fish, metaphorically speaking. Jenny-Marie Wong kept the machine running and helped push this work forward.

Finally, some deeply caring people have been with me since childhood. I would like to take some space here and acknowledge their importance in shaping my path. My

parents, Jeff and Patti, did most of the heavy lifting—sometimes by saying No and sometimes by saying Yes. My sister Amanda is tough and smart and a perpetual inspiration. Nick's alright, too. Grandpa Bob and Tom Gabrielson stayed up all night at the fire telling me about the Road and reminding me to never forget about the guy who keeps the power on. Grandma Jean is my second mother, and has always believed I can accomplish anything. Grandpa Steve taught me how to garden, how to box, and how to build. Chris Hayward, my best man, walked the railroad tracks and read the poems and finished all of my drinks when I fell asleep. Sarah Williams provided compelling distractions from this work and for that I am eternally grateful. She also made me a better person. Henry Pettibone, my son, has given me new perspective daily and I thank him for that. I look forward to watching him grow.

TABLE OF CONTENTS

LIST OF FIGURES

Abstract

Adaptive decision-making relies on a distributed network of neural substrates that learn associations between behaviors and outcomes, to ultimately guide future behavior. These substrates are organized in a system of cortical-striatal loops that offer unique contributions to goal-directed behavior and receive prominent inputs from the midbrain dopamine system. However, the consequences of dopamine fluctuations at these targets remain largely unresolved, despite aggressive interrogation. Some experiments have highlighted dopamine's role in learning via reward prediction errors, while others have noted the importance of dopamine in motivated behavior. Here, we explored the precise role of dopamine in shaping decision-making in cortex and striatum. First, we measure dopamine in ventral striatum during a trial-and-error task and show that it uniformly encodes a moment-by-moment estimate of value across multiple timescales. Our optogenetic manipulations demonstrate that changes in this value signal can be used to immediately enhance vigor, consistent with a motivational signal, *and* alter subsequent choice behavior, consistent with a learning signal. Next, I measured dopamine in multiple cortical-striatal loops to examine the uniformity of the value signal. I report that dopamine is non-uniform across circuits, but is consistent within them, implying that dopamine may offer unique contributions to the information processed in each loop. Finally, I performed single-unit recordings in the dorsal striatum, a major recipient of dopamine, to examine whether distinct its subcompartments—the patch and matrix—

carry distinct value signals used in the selection of actions. I report preliminary data and summarize improvements in my electrode localization technique.

Chapter 1

Introduction

1.1 Overview

Adaptive decision-making is essential for survival. If the conditions necessary for comfortable survival were fixed— if there were always money in the bank account or always fruit on the tree— a simple set of fixed rules would be enough to get by. In a changing environment, however, the conditions are flexible and the problem of survival reasserts itself in perpetuity. The animal must adapt to the changes or perish.

Most of the decisions we make on a daily basis are not so dire, yet the neural substrates with which we make even trivial decisions evolved long ago and were likely refined by severe consequences. In this thesis I will investigate some of the underlying neural mechanisms of adaptive decision-making, with a particular emphasis on how decisions are evaluated (what is being learned) and how the evaluation is intertwined with motivation (how the learning is put to use).

1.2 From the Law of Effect to Reinforcement Learning

In 1905, psychologist Edward Thorndike noted that 'responses that produce a satisfying effect in a particular situation become more likely to occur again in that situation, and responses that produce a discomforting effect become less likely to occur again in that situation.' This simple observation, what would later become the Law of

Effect (Thorndike 1911), foreshadowed more than a century of research that continues to grapple with understanding the nature of learning and motivation. Fundamentally, the Law of Effect recognizes the basic principles of instrumental conditioning. The probability of repeating a behavior increases when the behavior produces a satisfying effect. Moreover, this association can be drawn upon to modify future behavior to the potential benefit of the agent. After so many years, these basic features of Law of Effect still provide a useful framework for understanding how we learn. It has been expanded and refined as the field has grown.

The recent introduction of a family of algorithms called Reinforcement Learning (RL) is one such refinement (Sutton and Barto 1998). In a typical RL model, an agent interacts with a set of states, each with an associated value. Each time the agent experiences a state, the value assigned to that state is updated by a prediction error term. The prediction error is the difference between what was expected, based on the previous value, and what was experienced. Despite being designed for machine learning, these models have proven useful for providing a quantitative approach to studying instrumental learning and classical conditioning, based on the key features mentioned above.

1.3 *Does dopamine=RPE?*

One of the most striking examples of an RL concept mapping onto brain activity is the finding that midbrain dopamine cell firing bi-directionally scales with the amount of unexpected reward that an outcome generates, similar to a reward prediction error (RPE) (Montague, Dayan, and Sejnowski 1996; Wolfram Schultz, Dayan, and Montague 1997). That is, in repeated pairings of a cue and a reward, dopamine cells fire more in early trials when a reward is surprising than in later trials when it is entirely expected.

There is also a temporal component to reward expectation. Dopamine cells fire more when a reward is delivered sooner than expected and show pauses in firing when an outcome is expected but not delivered *on time*. The conclusion from these observations is that dopamine conveys a learning signal by comparing an expected amount of reward to a received amount of reward.

The idea that dopamine *is equal to* RPE, and thus provides a role in learning, has gained strong support. More recent electrophysiology studies of dopamine cells have found evidence for this relationship in both probabilistic and deterministic variants of classical conditioning tasks (Bayer, Lau, and Glimcher 2007; Matsumoto and Hikosaka 2009; Eshel et al. 2016). There is also some evidence of this relationship in terminal dopamine release patterns. (Hart et al. 2014; Saddoris et al. 2015)

However, the support for the 'Dopamine=RPE' hypothesis is not universal (John D Salamone and Correa 2012; K. Berridge 2007; Bromberg-Martin, Matsumoto, and Hikosaka 2010b). Recent experiments offer evidence that questions the uniformity of dopamine signaling. In tasks that require instrumental responding, such as maze running for reward (Howe et al. 2013) or lever-pressing for cocaine (Stuber et al. 2005), mesolimbic dopamine concentrations have been reported to gradually 'ramp up' from the onset of behavior until trial outcome, similar to a motivational signal. Additionally, dopamine cell recordings have shown subpopulations that respond in a heterogeneous manner to aversive stimuli (Coizet et al. 2006; Bromberg-Martin, Matsumoto, and Hikosaka 2010a) and some cells that respond to unexpected salient cues, independent of reinforcement value, which are hypothesized to provide an 'alerting' signal (Bromberg-Martin, Matsumoto, and Hikosaka 2010a). Attempts have been made to squeeze these observations into the 'Dopamine=RPE' mold (Niv 2013; Morita and Kato 2014; Lloyd

and Dayan 2015; Gershman 2014), incorporating refinements in modeling and language. Given these non-uniform findings about the activity of putative dopamine cells, is it not surprising that much work has been generated toward resolving this issue.

1.4 A Unifying Account of Dopamine Function

One proposed resolution to the conflict between dopamine providing a learning signal or a motivational signal lies in the distinct modes of dopamine cell firing (A. A. Grace et al. 2007a). It has been suggested that phasic bursting events are responsible for the RPEs that convey a learning signal, while slower, non-synchronous 'tonic' firing alters extracellular dopamine tone and enhances or attenuates aspects of motivation (Niv et al. 2007; Daw, Kakade, and Dayan 2002; Guitart-Masip et al. 2011). In Chapter 2, we test this hypothesis directly by measuring dopamine on sub-second (phasic) and minute ('tonic') timescales while rats perform a stochastically rewarded trial-and-error task. We report that dopamine in the nucleus accumbens core conveys a motivational value signal on both timescales, and that the rapid (sub-second) fluctuations of this value signal are functionally equivalent to the RPEs that inform future choice behavior. This result is concordant with dopamine playing a role in both learning and motivation and thus provides a unifying account of dopamine function.

1.5 Cortical-striatal loops offer unique contributions to decision-making

In the twenty years following the first Schultz RPE paper, much theorizing has predicted specific roles for brain circuitry and neuromodulators in carrying signals related to RL model parameters (Daw and Doya 2006; Daw, Niv, and Dayan 2005). These efforts have produced a modest renaissance in the computational modeling of how brain

circuitry performs adaptive decision-making. An attractive integration of these efforts is the suggestion that distinct cortical-striatal circuits may processes discrete aspects of the decision and work in conjunction to arrive at an appropriate action or set of behaviors (Pennartz et al. 2009; Daw, Niv, and Dayan 2005; Cools 2015).

In a simplified mapping, these cortical-striatal loops can be reduced to a dorsal-ventral neocortical axis projecting to a roughly dorsal-ventral striatal axis (Figure 1.1). The 'executive' loop (anterior cingulate and dorsal prelimbic neocortex to dorsomedial striatum) may carry information necessary for the overt modification of behavior. For example, this loop is involved in reversal learning (Ragozzino 2003) and arbitrating among candidate motor plans based on the accumulation of evidence (Shenhav, Botvinick, and Cohen 2013; Demanuele et al. 2015). Other likely roles for this 'executive' loop include performance and error monitoring, rule learning and rapid strategy switching (Seger 2009; Kehagia, Murray, and Robbins 2010)—all of which rely on a working memory of recent performance (Euston, Gruber, and McNaughton 2012). This type of overt valuation is more computationally demanding and may only be engaged when the cost of losing out on potential reward exceeds the cost of exerting executive control (Cools 2015; Shenhav, Botvinick, and Cohen 2013).

In a complimentary fashion, the ventral 'motivational' loop, also known as the limbic loop (ventral prelimbic and infralimbic cortex to NAc core and shell), integrates information about context, satiety, and reward history (J D Salamone, Cousins, and Snyder 1997)— decision variables that inform a *state value* (Houk, Joel L. Davis, and Beiser 1995). This network is implicit in nature and drives instrumental responding and the learning of simple action-outcome associations (Robbins et al. 1989; Kelley and Delfs

1991). The executive loop and motivational loop work in coordination to generate the range behaviors appropriate to the contingencies of the task at hand (Seger 2009).

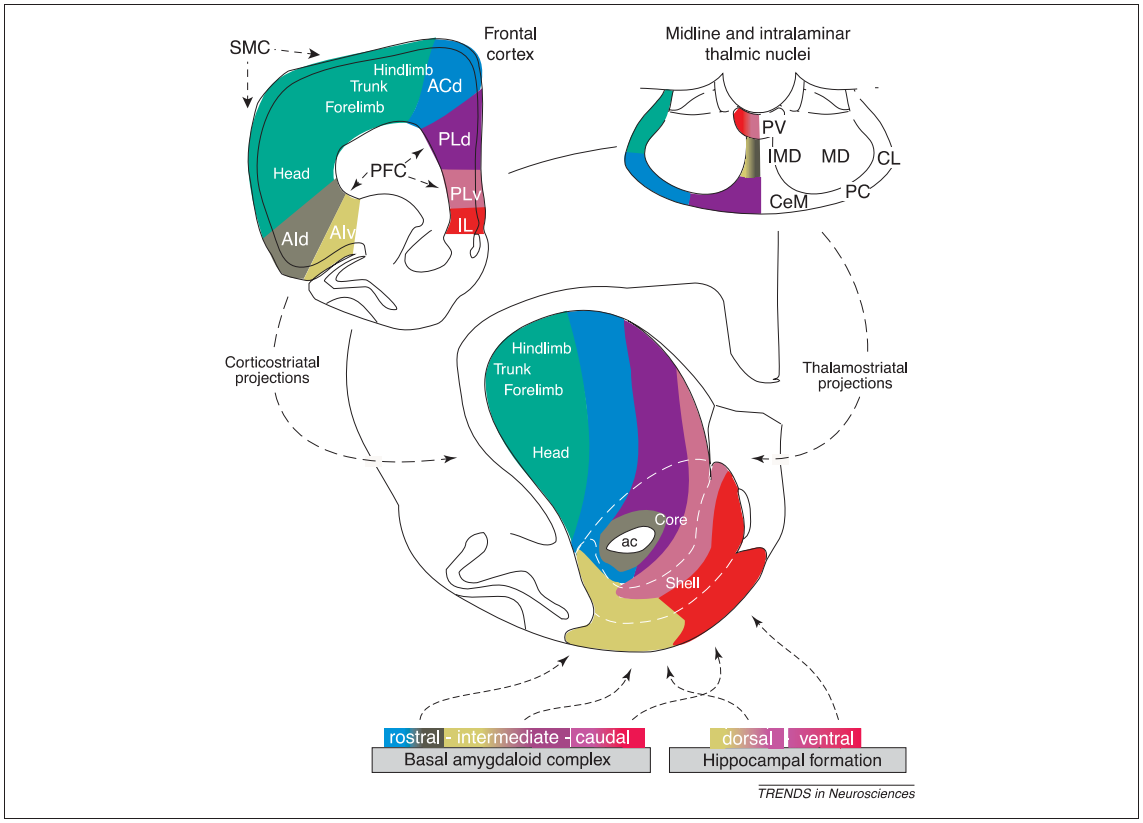


Figure 1.1 Schematic diagram of cortical-striatal connectivity. Adapted from (Voorn et al. 2004).

1.6 A Universal Dopamine Signal?

Dopamine broadcasts a diffuse signal to the cortical-striatal circuitry outlined above, and is thus likely a key modulator in adaptive decision-making. Yet, it remains unclear to what extent this signal is heterogeneous and further, whether the heterogeneity arises in the midbrain via distinct subpopulations of dopamine cells or at the terminal via

local metabolism and circuitry. Midbrain dopamine neurons receive a gradient of reciprocal inputs from their limbic and dorsal striatal targets (Voorn et al. 2004), which provide online feedback and alter the activity of subpopulations. Additionally, they project a distinct density of inputs to neocortex and striatum (Juarez and Han 2016), which originate from spatially organized portions of the ventral tegmental area and substantia nigra pars compacta (Lammel et al. 2008). Terminal dopamine concentration is also subject to differences in local metabolism and circuitry. There is a dorsal-ventral gradient of decreasing dopamine transporter availability in medial prefrontal cortex (Heidbreder and Groenewegen 2003), whereas there is a strong dependence on dopamine transporter in both dorsal and ventral striatum for synaptic dopamine clearance. These metabolic differences can impact how long dopamine stays in the synapse, which in turn affects the dynamics of D1- and D2- receptor occupancy.

Given such complexities of anatomy, it is unlikely that the dopamine signal is universal across the many networks involved in adaptive behavior. In Chapter 3, I address this hypothesis directly by simultaneously measuring extracellular dopamine in cortical and striatal targets while rats perform our adaptive decision-making task. I show that while the dopamine signal does indeed vary across targets, there is a consistent relationship between dopamine and reward rate, an estimate of value, within at least one cortical-striatal loop—the ventral prelimbic neocortex and NAc core.

.

1.7 Striatal Microcircuits and Value

The striatum is the major input nucleus of the basal ganglia and receives massive innervation from dopamine neurons. It is critically involved in action selection and is hypothesized to achieve this through differential activity of the direct and indirect

pathways, which exclusively express D1- and D2- receptors, respectively. The mapping of reinforcement learning onto neural processes has been most fruitful here, with specific evidence of action value (Samejima, Ueda, and Doya 2005) and prediction error (Oyama et al. 2010) signaling among the medium spiny projection neurons which comprise ~95% of its population. In one framework (Maia and Frank 2011), the phasic bursts of dopamine that occur during positive prediction errors engage the normally unoccupied low-affinity D1 receptors, resulting in long-term potentiation at the synapses which were active during the action which produced the good outcome. Conversely, the pauses in dopamine that occur during negative prediction errors decrease D2 receptor occupancy, resulting in a long-term depression at those synapses. In this way, selected actions that resulted with an unexpectedly positive outcome are reinforced and actions that resulted in an unexpectedly negative outcome are suppressed.

An additional layer of complexity emerges, however, when examining the expression patterns of mu-opioid receptors, choline acetyltransferase, enkephalin, calbindin, and substance P (Saka and Graybiel 2003). There are clear subcompartments—the mu-opioid receptor rich *patches* and the mu-opioid receptor deplete *matrix*—which operate in parallel with distinct efferent and afferent connectivity. The matrix compartment comprises the majority of striatum, receives somatosensory input from superficial layers of cortex, and gives rise to the majority of the direct and indirect pathways. Patches receive preferential projections from anterior cingulate and prelimbic neocortex (Crittenden and Graybiel 2011; Eblen and Graybiel 1995; Friedman et al. 2015), specifically the ventral portion of the prelimbic region (Sesack et al. 1989). Patches send downstream projections to the basal ganglia, like matrix, but send a unique

projection to the dopamine neurons of the substantia nigra pars compacta (Fujiyama et al. 2011).

The observation that patches have rather limbic connectivity and a unique influence on dopamine neurons has led to the hypothesis that they carry a state value signal, similar to the critic of an Actor-Critc RL model (Houk, Joel L. Davis, and Beiser 1995; Daw, Niv, and Dayan 2005), while the matrix compartment signals specific action values. I describe my approach to testing this hypothesis in Chapter 4.

1.8 Summary of Chapters

The goal of this thesis was to explore specific questions about the cortical-striatal circuitry that shapes decision-making. In Chapter 2, we tested whether dopamine provides a learning signal or a motivational signal. By optogenetically manipulating DA release at selective time points, we showed that DA can bidirectionally modulate both motivation *and* learning. Moreover, we show that mesolimbic dopamine encodes value on fast and slow timescales. Fast fluctuations of extracellular dopamine concentration correspond with a value function, which estimates a moment-by-moment state value. Slow changes in extracellular dopamine concentration similarly corresponded with reward rate, an alternative estimate for the value of doing work.

In Chapter 3, I extend our previous work to investigate whether dopamine broadcasts the same value signal to multiple cortical-striatal loops, or whether the dopamine signal is distinct. I replicated my previous finding that minute-by-minute dopamine fluctuations in the nucleus accumbens core reflect reward rate. Unexpectedly, I found that dopamine also encodes reward rate in a corresponding neocortical afferent, the prelimbic cortex. I determined that this relationship was limited to the more ventral

portion of the prelimbic cortex, forming a 'hotspot' of value coding in the neocortex. This finding establishes that (i) dopamine in cortex and striatum can carry the same signal and (ii) this signal is confined to regions that interact in the same loop. This finding relates well to human fMRI data showing the same hotspots are engaged across a range of tasks that require estimating subjective value.

In Chapter 4, I explore the hypothesis that anatomically distinct striatal compartments, the patch and the matrix, contribute unique value signals to decision-making. I describe challenges to testing this hypothesis and my evolving approach to recording from and identifying these compartments as well as the data collected thus far.

Chapter 2

Mesolimbic dopamine signals the value of work[*]

Abstract

Dopamine cell firing can encode errors in reward prediction, providing a learning signal to guide future behavior. Yet dopamine is also a key modulator of motivation, invigorating current behavior. Existing theories propose that fast (phasic) dopamine fluctuations support learning, whereas much slower (tonic) dopamine changes are involved in motivation. We examined dopamine release in the nucleus accumbens across multiple time scales, using complementary microdialysis and voltammetric methods during adaptive decision-making. We found that minute-by-minute dopamine levels covaried with reward rate and motivational vigor. Second-by-second dopamine release encoded an estimate of temporally discounted future reward (a value function). Changing dopamine immediately altered willingness to work and reinforced preceding action choices by encoding temporal-difference reward prediction errors. Our results indicate that dopamine conveys a single, rapidly evolving decision variable, the available reward for investment of effort, which is employed for both learning and motivational functions.

[*] Additional Supplementary Figures available in full published online format. (Hamid, Pettibone et al. 2015)

2.1 Introduction

Altered dopamine signaling is involved in many human disorders, from Parkinson's disease to drug addiction. Yet the normal functions of dopamine have long been the subject of debate. There is extensive evidence that dopamine affects learning, especially the reinforcement of actions that produce desirable results (J. N. J. Reynolds, Hyland, and Wickens 2001). Specifically, electrophysiological studies suggest that bursts and pauses of dopamine cell firing encode the reward prediction errors (RPEs) of reinforcement learning (RL) theory (W Schultz, Dayan, and Montague 1997). In this framework, RPE signals are used to update estimated values of states and actions, and these updated values affect subsequent decisions when similar situations are re-encountered. Further support for a link between phasic dopamine and RPE comes from measurements of dopamine release using fast-scan cyclic voltammetry (FSCV) (Day et al. 2007; Hart et al. 2014) and optogenetic manipulations (K. M. Kim et al. 2012; Steinberg et al. 2013).

There is also extensive evidence that dopamine modulates arousal and motivation (K. Berridge 2007; Beierholm et al. 2013). Drugs that produce prolonged increases in dopamine release (for example, amphetamines) can markedly enhance psychomotor activation, whereas drugs or toxins that interfere with dopamine transmission have the opposite effect. Over slow timescales (tens of minutes) microdialysis studies have demonstrated that dopamine release ([DA]) is strongly correlated with behavioral activity, especially in the nucleus accumbens (Freed and Yamamoto 1985) (that is, mesolimbic [DA]). It is widely thought that slow (tonic) [DA] changes are involved in motivation

(Niv, Daw, and Dayan 2006; Cagniard et al. 2006; John D Salamone and Correa 2012). However, faster [DA] changes also appear to have a motivational function (Satoh et al. 2003). Subsecond increases in mesolimbic [DA] accompany motivated approach behaviors (Phillips et al. 2003; Roitman et al. 2004), and dopamine ramps lasting several seconds have been reported as rats approach anticipated rewards (Howe et al. 2013), without any obvious connection to RPE. Overall, the role of dopamine in motivation is still considered to be mysterious (John D Salamone and Correa 2012).

We sought to better understand just how dopamine contributes to motivation and to learning simultaneously. We found that mesolimbic [DA] conveys a motivational signal in the form of state values, which are moment-by-moment estimates of available future reward. These values were used for making decisions about whether to work, that is, to invest time and effort in activities that are not immediately rewarded, to obtain future rewards. When there was an unexpected change in value, the corresponding change in [DA] not only influenced motivation to work, but also served as an RPE learning signal, reinforcing specific choices. Rather than separate functions of phasic and tonic [DA], our data support a unified view in which the same dynamically fluctuating [DA] signal influences both current and future motivated behavior.

2.2 Results

2.2.1 Motivation to work adapts to recent reward history

We used an adaptive decision-making task (Fig. 1a and Online Methods) that is closely related to the reinforcement learning framework (a 'two-armed bandit'). On each trial, a randomly chosen nose poke port lit up (Light-On), indicating that the rat might

profitably approach and place its nose in that port (Center-In). The rat had to wait in this position for a variable delay (0.75–1.25s) until an auditory white noise burst (Go cue) prompted the rat to make a brief leftward or rightward movement to an adjacent side port. Unlike previous behavioral tasks using the same apparatus, the Go cue did not specify which way to move; instead, the rat had to learn through trial- and-error which option was currently more likely to be rewarded. Left and right choices had separate reward probabilities (each was either 10, 50 or 90%), and these probabilities changed periodically without any explicit signal. On rewarded trials only, entry into the side port (Side-In) immediately triggered an audible click (the reward cue) as a food hopper delivered a sugar pellet to a separate food port at the opposite side of the chamber.

Trained rats readily adapted their behavior in at least two respects (Figure 2.1b,c). First, actions followed by rewards were more likely to be subsequently selected (that is, they were reinforced), producing left and right choice probabilities that scaled with actual reward probabilities (Samejima, Ueda, and Doya 2005) (Figure 2.1d).

Second, rats were more motivated to perform the task while it produced a higher rate of reward (Guitart-Masip et al. 2011; Wang, Miura, and Uchida 2013). This was apparent from latency (the time taken from Light-On until the Center-In nose poke), which scaled inversely with reward rate (Figure 2.1e–g). When reward rate was higher, rats were more likely to be already waiting near the center ports at Light-On (engaged trials), producing very short latencies. Higher reward rates also produced shorter latencies even when rats were not already engaged at Light-On, as a result of an elevated moment-by-moment probability (hazard rate) of choosing to begin work (Figure 2.1h,i).

These latency observations are consistent with optimal foraging theories (Stephens 1986), which argue that reward rate is a key decision variable (currency). As

animals perform actions and experience rewards, they construct estimates of reward rate and can use these estimates to help decide whether engaging in an activity is worthwhile. In a stable environment, the best estimate of reward rate is simply the total magnitude of past rewards received over a long time period, divided by the duration of that period. It has been proposed that such a long-term average reward rate is encoded by slow (tonic) changes in [DA] (Niv, Daw, and Dayan 2006). However, under shifting conditions such as our trial-and-error task, the reward rate at a given time is better estimated by more local measures. Reinforcement learning algorithms use past reward experiences to update estimates of future reward from each state: a set of these estimates is called a value function (Sutton and Barto 1998).

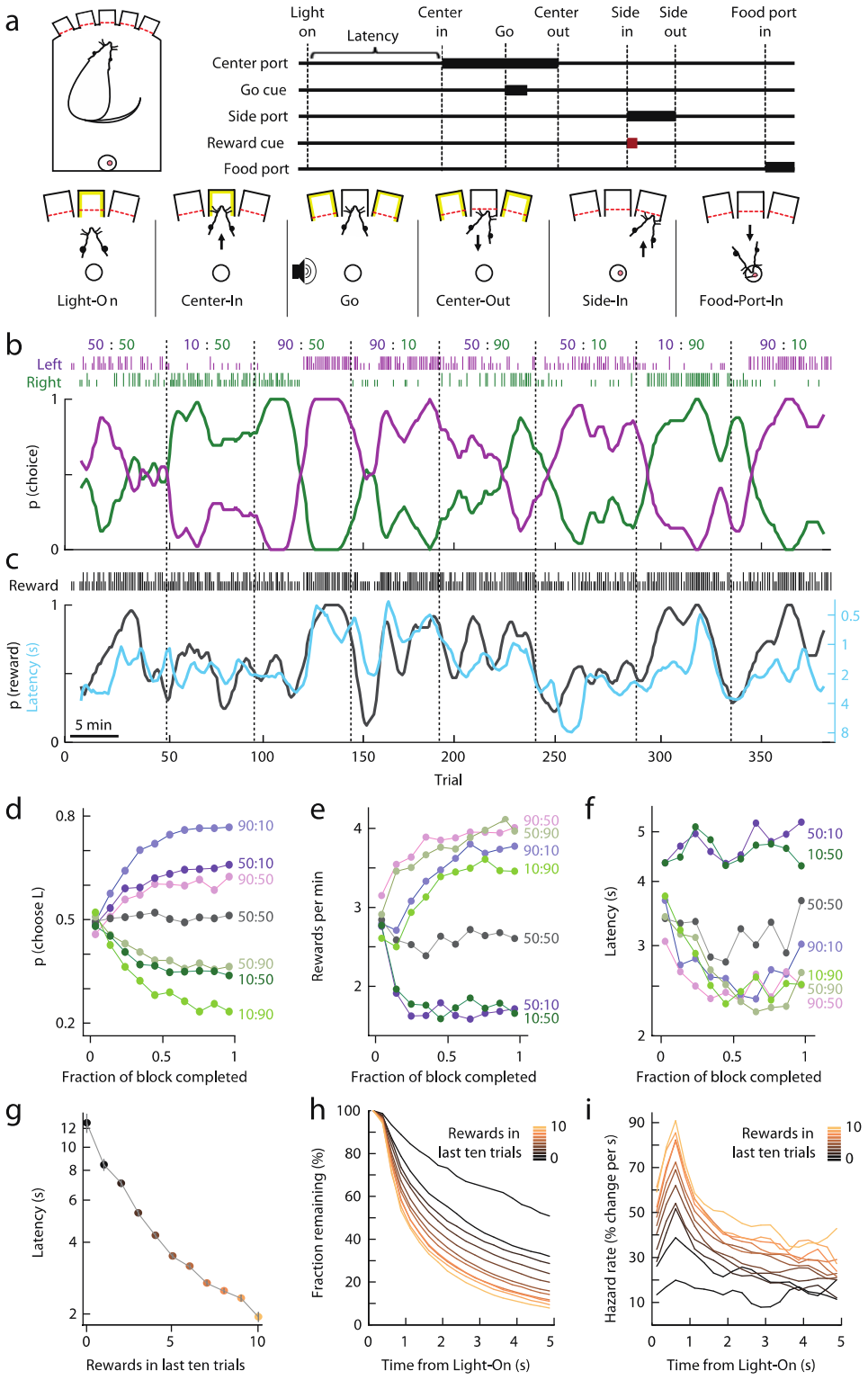


Figure 2.1. Adaptive choice and motivation in the trial-and-error task. (a) Sequence of behavioral events (in rewarded trials). (b) Choice behavior in a representative session. Numbers at top denote nominal block-by-block reward probabilities for left (purple) and right (green) choices. Tick marks indicate actual choices and outcomes on each trial (tall

ticks indicate rewarded trials, short ticks unrewarded). The same choice data is shown below in smoothed form (thick lines; 7-trial smoothing). (c) Relationship between reward rate and latency for the same session. Here tick marks are used to indicate only whether trials were rewarded or not, regardless of choice. Solid black line shows reward rate, and cyan line shows latency (on inverted log scale), both smoothed in the same way as B. (d) Choices progressively adapt towards the block reward probabilities (data set for panels d-i: n = 14 rats, 125 sessions, 2738 +/- 284 trials per rat). (e) Reward rate breakdown by block reward probabilities. (f) Latencies by block reward probabilities. Latencies become rapidly shorter when reward rate is higher. (g) Latencies by proportion of recent trials rewarded. Error bars represent s.e.m. (h) Latency distributions presented as survivor curves (i.e. the average fraction of trials for which the Center-In event has not yet happened, by time elapsed from Light-On) broken down by proportion of recent trials rewarded. (i) Same latency distributions as panel h, but presented as hazard rates (i.e. the instantaneous probability that the Center-In event will happen, if it has not happened yet). The initial bump in the first second after Light-On reflects engaged, after that hazard rates are relatively stable and continue to scale with reward history.

2.2.2 Minute-by-minute dopamine correlates with reward rate

To test whether changes in [DA] accompany reward rate during adaptive decision-making, we first employed microdialysis in the nucleus accumbens combined with liquid chromatography–mass spectrometry. This method allows us to simultaneously assay a wide range of neurochemicals, including all of the well-known low–molecular weight striatal neurotransmitters, neuromodulators and their metabolites (Figure 2.2a), each with 1-min time resolution. We performed regression analyses to assess relationships between these neurochemicals and a range of behavioral factors: reward rate, the number of trials attempted (as an index of a more general form of activation/arousal), the degree of exploitation versus exploration (an important decision parameter that has been suggested to involve [DA]; see Methods) and the cumulative reward obtained (as an index of progressively increasing factors such as satiety).

We found a clear overall relationship between [DA] and ongoing reward rate (R^2 = 0.15, $P < 10^{-16}$). Among the 19 tested analytes, [DA] had by far the strongest relationship to reward rate (Figure 2.2b), and this relationship was significant in six of

seven individual sessions, from six different rats (P = 0.0052 or lower in each case; Figure 2.2c**)**. Modest relationships were also found for the dopamine metabolites DOPAC and 3-MT. We found a weak relationship between [DA] and the number of trials attempted, but this was entirely accounted for by reward rate; that is, if the regression model already included reward rate, adding number of attempts did not improve model fit. We did not find support for alternative proposals that tonic [DA] is related to exploration or exploitation, as higher [DA] was not associated with an altered probability of choosing the better left or right option (Figure 2.2b). [DA] also showed no relationship to the cumulative total rewards earned (though there was a strong relationship between cumulative reward and the dopamine metabolite HVA, among other neurochemicals; Figure 2.2b).

We conclude that higher reward rate is associated specifically with higher average [DA], rather than other striatal neuromodulators, and with increased motivation to work. This finding supports the proposal that [DA] helps to mediate the effects of reward rate on motivation (Niv, Daw, and Dayan 2006). However, rather than signaling an especially long-term rate of reward, [DA] tracked minute-by-minute fluctuations in reward rate. We therefore needed to assess whether this result truly reflects an aspect of [DA] signaling that is inherently slow (tonic) or could instead be explained by rapidly changing [DA] levels, that signal a rapidly changing decision variable.

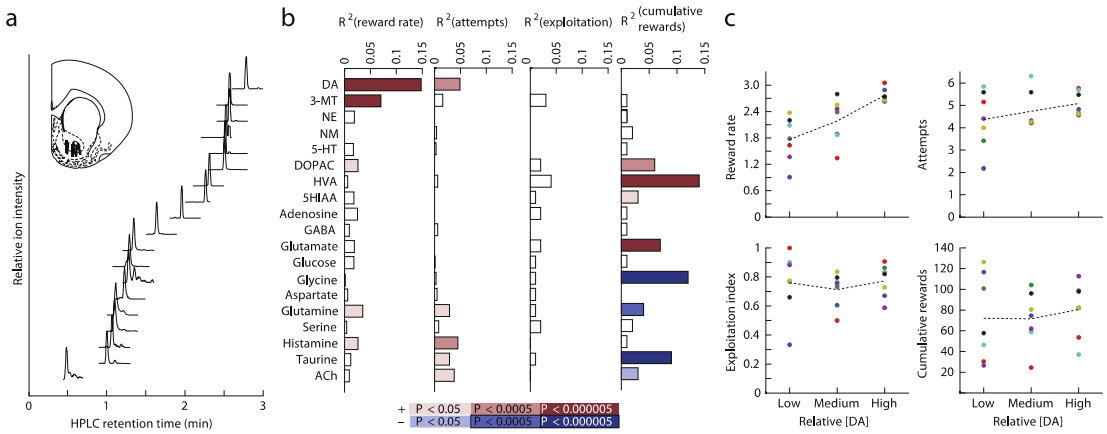


Figure 2.2. Minute-by-minute dopamine levels track reward rate. (a) Total ion chromatogram of a single representative microdialysis sample, illustrating the set of detected analytes in this experiment. X-axis indicates HPLC retention times, y-axis indicates intensity of ion detection for each analyte (normalized to peak values). (Inset) locations of each microdialysis probe within the nucleus accumbens (all data shown in the same Paxinos atlas section; six were on the left side and one on the right). Abbreviations: DA, dopamine; 3-MT, 3-methoxytyramine; NE, norepinephrine; NM, normetanephrine; 5-HT, serotonin; DOPAC, 3,4-dihydroxyphenylacetate acid; HVA, homovanillic acid; 5HIAA, 5-hydroxyindole-3-acetic acid, GABA, ɣ-aminobutyric acid; ACh, acetylcholine. (b) Regression analysis results indicating strength of linear relationships between each analyte and each of four behavioral measures (reward rate; number of attempts; exploitation index; and cumulative rewards). Color scale shows p-values, Bonferroni-corrected for multiple comparisons (4 behavioral measures * 19 analytes), with red bars indicating a positive relationship and blue bars a negative relationship. Since both reward rate and attempts showed significant correlations with [DA], we constructed a regression model that included these predictors and an interaction term. In this model R^2 remained at 0.15 and only reward rate showed a significant partial effect ($p < 2.38 \times 10^{-12}$). (c) An alternative assessment of the relationship between minute-long [DA] samples and behavioral variables. Within each session [DA] levels were divided into three equal-sized bins (LOW, MEDIUM, HIGH); different colors indicate different sessions. For each behavioral variable, means were compared across [DA] levels using one-way ANOVA. There was a significant main effect of reward rate ($F(2,18)=10.02$, $p=0.0012$), but no effect of attempts ($F(2,18)=1.21$, $p=0.32$), exploitation index ($F(2,18)=0.081$, $p=0.92$), or cumulative rewards ($F(2,18)=0.181$, $p=0.84$). Post-hoc comparisons using the Tukey test revealed that the mean reward rates of LOW and HIGH [DA] differed significantly ($p=0.00082$).

2.2.3 Dopamine signals time-discounted available future reward

To help distinguish these possibilities, we used FSCV to assess task-related [DA] changes on fast timescales (from tenths of seconds to tens of seconds; Figure 2.3). In each trial, [DA] rapidly increased as rats poked their nose in the start hole (Figure 2.3c,d),

and for all rats this increase was more closely related to this approach behavior than to the onset of the light cue (for data from each of the single sessions from all six rats). A second abrupt increase in [DA] occurred following presentation of the Go cue (Figure 2.3c,d). If received, the reward cue prompted a third abrupt increase (Figure 2.3c,d). [DA] rose still further as the rat approached the food port (Figure 2.3c,d), then declined once the reward was obtained. The same overall pattern of task-related [DA] change was observed in all rats, albeit with some variation. [DA] increases did not simply accompany movements, given that, on the infrequent trials in which the rat approached the food port without hearing the reward cue, we observed no corresponding increase in [DA] (Figure 2.3c,d).

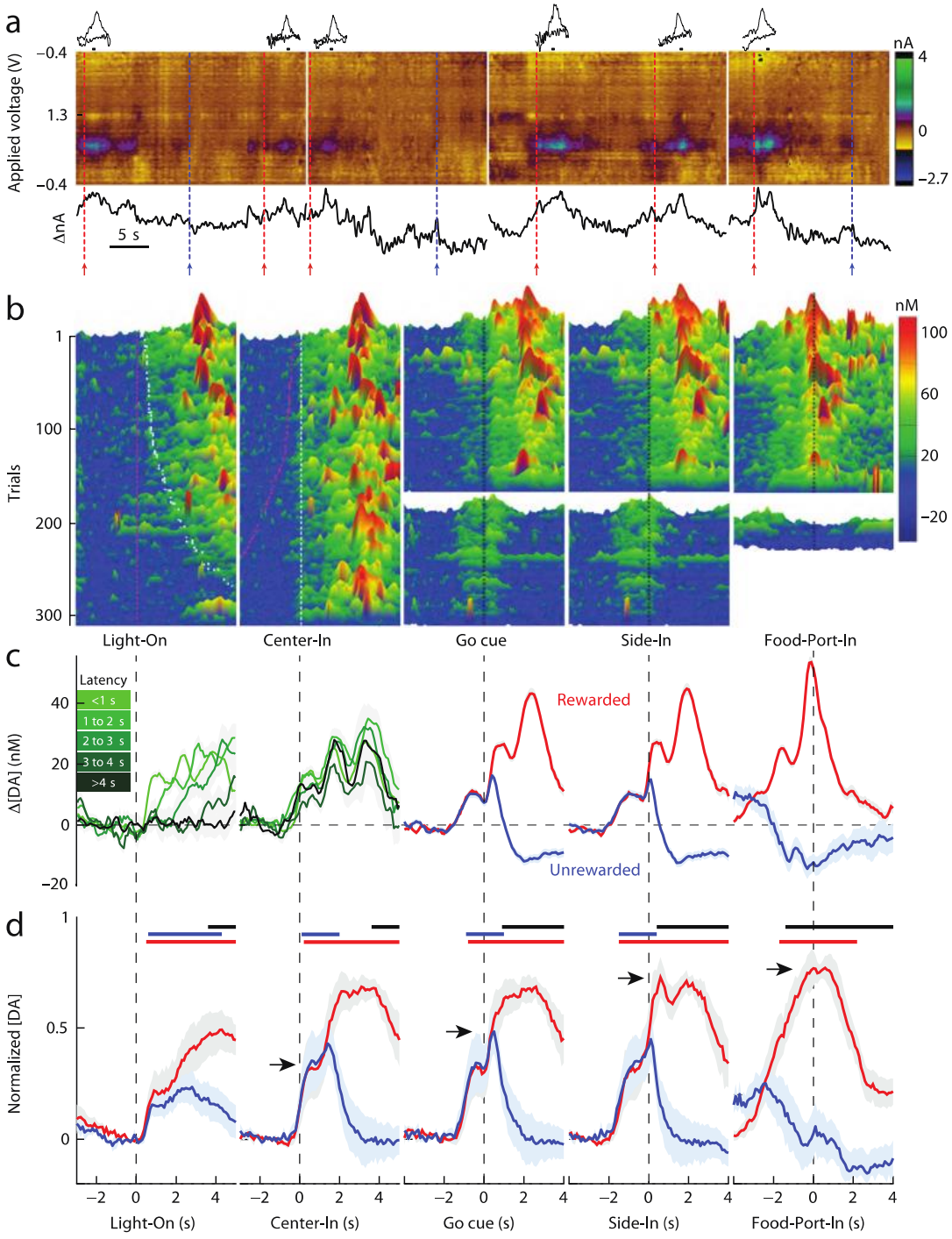


Figure 2.3. A succession of within-trial dopamine increases. (a) Examples of FSCV data from a single session. Color plots display consecutive voltammograms (every 0.1s) as a vertical colored strip; examples of individual voltammograms are shown at top (taken from marked time points). Dashed vertical lines indicate Side-In events for rewarded (red) and unrewarded (blue) trials. Black traces below indicate raw current values, at the applied voltage corresponding to the dopamine peak. (b) [DA] fluctuations for each of the 312 completed trials of the same session, aligned to key behavioral events. For Light-On

and Center-In alignments, trials are sorted by latency (pink dots mark Light-On times; white dots mark Center-In times). For the other alignments rewarded (top) and unrewarded (bottom) trials are shown separately, but otherwise in the order in which they occurred. [DA] changes aligned to Light-On were assessed relative to a 2s baseline period, ending 1s before Light-On. For the other alignments, [DA] is shown relative to a 2s baseline ending 1s before Center-In. (c) Average [DA] changes during a single session (same data as b; shaded area represents s.e.m.). (d) Average event-aligned [DA] change across all six animals, for rewarded and unrewarded trials. Data are normalized by the peak average rewarded [DA] in each session, and are shown relative to the same baseline epochs as in b. Black arrows indicate increasing levels of event-related [DA] during the progression through rewarded trials. Colored bars at top indicate time periods with statistically significant differences (red, rewarded trials greater than baseline, one-tailed t-tests for each 100ms time point individually; blue, same for unrewarded trials; black, rewarded trials different to unrewarded trials, 2-tailed t-tests; all statistical thresholds set to p=0.05, uncorrected).

The overall ramping up of [DA] as rats drew progressively closer to reward suggested some form of reward expectation (Howe et al. 2013). Specifically, we hypothesized that [DA] continuously signals a value function: the temporally discounted reward predicted from the current moment. To make this more clear, consider a hypothetical agent moving through a sequence of distinct, unrewarded states leading up to an expected reward (perhaps a rat running at constant speed along a familiar maze arm; Figure 2.4a). As the reward is more discounted when more distant, the value function will progressively rise until the reward is obtained.

This value function describes the time-varying level of motivation. If a reward is distant (so strongly discounted), animals are less likely to choose to work for it. Once engaged, animals are increasingly motivated, and so less likely to quit, as they detect progress toward the reward (the value function produces a ‘goal gradient’) (Hull 1932). If the reward is smaller or less reliable, the value function will be lower, indicating less incentive to begin work. Moving closer to our real situation, suppose that reward is equally likely to be obtained, or not, on any given trial, but a cue indicates this outcome

halfway through the trial (Figure 2.4a). The increasing value function should initially reflect the overall 0.5 reward probability, but if the reward cue occurs, estimated value should promptly jump to that of the (discounted) full reward.

Such unpredicted sudden transitions to states with a different value produce 'temporal-difference' RPEs (Figure 2.4b). In particular, if the value function is low (for example, the trajectory indicating 0.25 expectation of reward), the reward cue produces a large RPE, as value jumps up to the discounted value of the now-certain reward. If instead reward expectation was higher (for example, 0.75 trajectory), the RPE produced by the reward cue is smaller. Given that temporal difference RPEs reflect sudden shifts in value, under some conditions they can be challenging to dissociate from value itself. However, RPE and value signals are not identical. In particular, as reward gets closer, the state value progressively increases but RPE remains zero unless events occur with unpredicted value or timing.

Our task includes additional features, such as variable timing between events and many trials. We therefore considered what the 'true' value function should look like, on average, based on actual times to future rewards (Figure 2.4c). At the beginning of a trial, reward is at least several seconds away and may not occur at all until a later trial. During correct trial performance each subsequent, variably timed event indicates to the rat that rewards are getting closer and more likely, and thus causes a jump in state value. For example, hearing the Go cue indicates both that reward is closer and that the rat will not lose out by moving too soon (an impulsive procedural error). Hearing the reward cue indicates that reward is now certain and only a couple of seconds away.

To assess how the intertwined decision variables, state value and RPE, are encoded by phasic [DA], we compared our FSCV measurements to the dynamically

varying state value and RPE of a reinforcement learning model (Online Methods). This simplified model consisted of a set of discrete states whose values were updated using temporal-difference RPEs. When the actual sequence of behavioral events experienced by the rat was given as input, the model's value function consisted of a series of increases in each trial (Figure 2.4d,e), resembling the observed time course of [DA] (Figure 2.3c).

Consistent with the idea that state value represents motivation to work, model state value early in each trial correlated with behavioral latencies for all rats (across a wide range of model parameter settings). We identified model parameters (learning rate = 0.4, discount factor = 0.95) that maximized this behavioral correlation across all rats combined and examined the corresponding within- trial correlation between [DA] and model variables. For all of the six FSCV rats, we found a clear and highly significant positive correlation between phasic [DA] and state value V (Figure 2.4f). [DA] and RPE were also positively correlated, as expected given that V and RPE partially covary. However, in every case, [DA] had a significantly stronger relationship to V than to RPE (Figure 2.4f). We emphasize that this result was not dependent on specific model parameters; in fact, even if parameters were chosen to maximize the [DA]:RPE correlation, the [DA]:V correlation was stronger.

Correlations were maximal when V was compared with the [DA] signal measured ~0.4–0.5 s later (Figure 2.4g). This small delay is consistent with the known brief lag associated with the FSCV method using acute electrodes (Venton, Troyer, and Wightman 2002) and prior observations that peak [DA] response occurs ~0.5 s after cue onset with acute FSCV recordings (Day et al. 2007). As an alternative method of incorporating temporal distortion that might be produced by FSCV and/or the finite speeds of DA release and update, we convolved model variables with a kernel consisting of an

exponential rise and fall, and explored the effect of varying kernel time constants. Once again, [DA] always correlated much better with V than with RPE across a wide range of parameter values. We conclude that state value provides a more accurate description of the time course of [DA] fluctuations than RPE alone, even though RPEs can be simultaneously signaled as changes in state value.

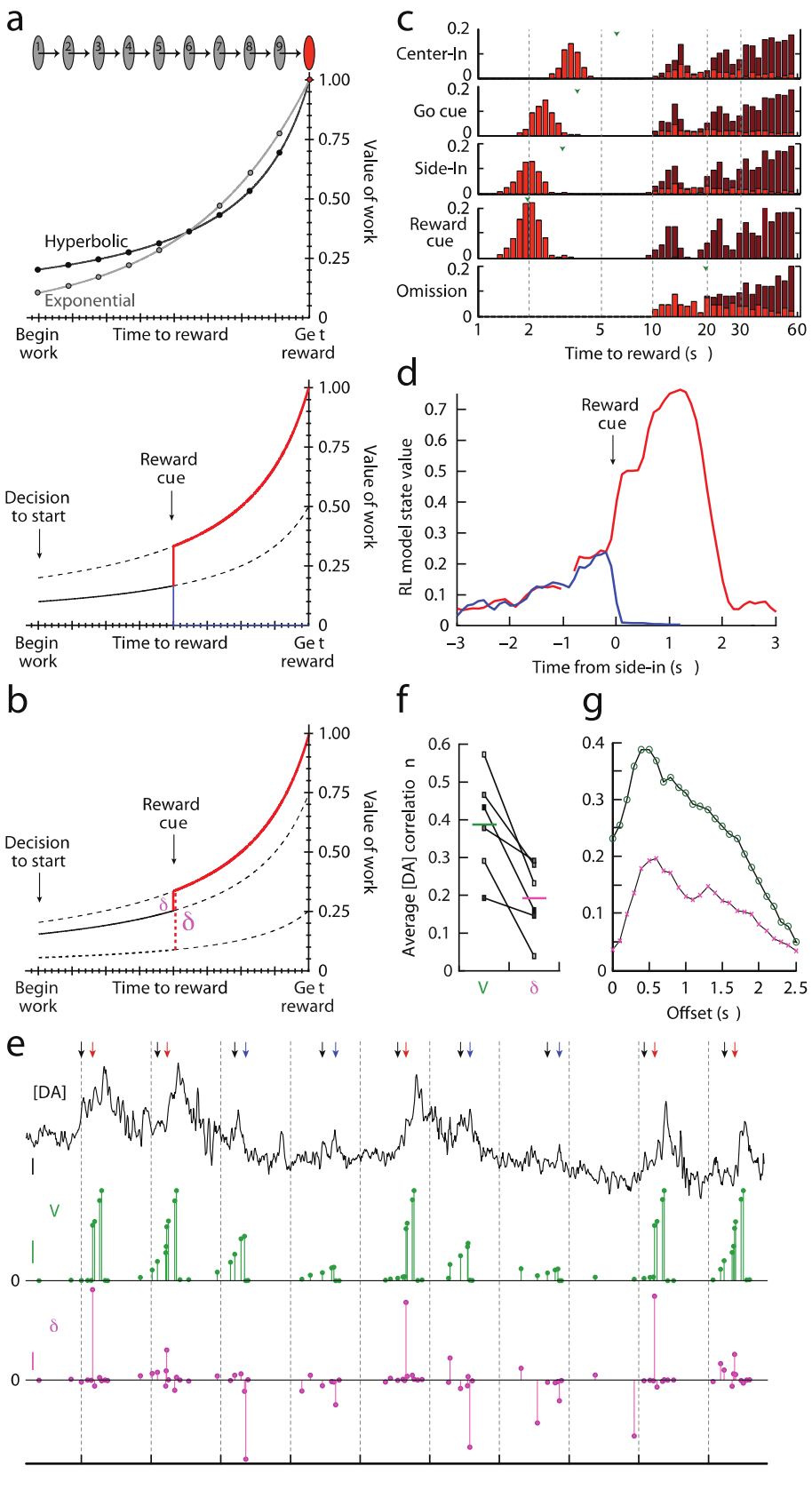

a
1
2
3
4
5
6
7
8
9
1.00
0.75
0.50
0.25
0
Value of work
Hyperbolic
Exponential
Begin work
Time to reward
Get reward
Decision to start
Reward cue
b
δ
c
Center-In
Go cue
Side-In
Reward cue
Omission
0.2
1
2
5
10
20
30
60
Time to reward (s)
d
RL model state value
Time from side-in (s)
f
Average [DA] correlation
V
g
Offset (s)
e
[DA]

Figure 2.4. Within-trial dopamine fluctuations reflect state value dynamics. (a) Top, Temporal discounting: the motivational value of rewards is lower when they are distant in time. With the exponential discounting commonly used in RL models, value is lower by a constant factor ɣ for each time step of separation from reward. People and other animals may actually use hyperbolic discounting which can optimize reward rate (since rewards/time is inherently hyperbolic). Time parameters are here chosen simply to illustrate the distinct curve shapes. Bottom. Effect of reward cue, or omission, on state value. At trial start the discounted value of a future reward will be less if that reward is less likely. Lower value provides less motivational drive to start work - producing e.g. longer latencies. If a cue signals that upcoming reward is certain, the value function jumps up to the (discounted) value of that reward. For simplicity, the value of subsequent rewards is not included. (b) The reward prediction error (delta) reflects abrupt changes in state value. If the discounted value of work reflects an unlikely reward (e.g. probability = 0.25) a reward cue prompts a larger delta than if the reward was likely (e.g. probability = 0.75). Note that in this idealized example, delta would be zero at all other times. (c) Top, Task events signal updated times-to-reward. Data is from the same example session as Fig.3c. Bright red indicates times to the very next reward, dark red indicates subsequent rewards. Green arrowheads indicate average times to next reward (harmonic mean, only including rewards in the next 60s). As the trial progresses, average times-to-reward get shorter. If the reward cue is received, rewards are reliably obtained ~2s later. Task events are considered to prompt transitions between different internal states whose learned values reflect these different experienced times-to-reward. (d) Average state value of the RL model for rewarded (red) and unrewarded (blue) trials, aligned on the Side-In event. The exponentially-discounting model received the same sequence of events as in Fig.3c, and model parameters (α=0.68, ɣ=0.98) were chosen for the strongest correlation to behavior (comparing state values at Center-In to latencies in this session, Spearman r=-0.34). Model values were binned at 100ms, and only bins with at least 3 events (state transitions) were plotted. (e) Example of the [DA] signal during a subset of trials from the same session, compared to model variables. Black arrows indicate Center-In events, red arrows Side-In with Reward Cue, blue arrows Side-In alone (Omission). Scale bars are: [DA], 20nM; V, 0.2; □, 0.2. Dashed grey lines mark the passage of time in 10s intervals. (f) Within-trial [DA] fluctuations are more strongly correlated with model state value (V) than with RPE (delta). For every rat the [DA] : V correlation was significant ($p<10^{-14}$ in each case; Wilcoxon signed-rank test of null hypothesis that median correlation within trials is zero) and significantly greater than the [DA] : delta correlation ($p<10^{-24}$ in each case, Wilcoxon signed-rank test). Groupwise, both [DA] : V and [DA] : delta correlations were significantly non-zero, and the difference between them was also significant (all p=0.031, Wilcoxon signed-rank test). Model parameters (α=0.4, ɣ =0.95) were chosen to maximize the average behavioral correlation across all 6 rats (Spearman r = -0.28), but the stronger [DA] correlation to V than to delta was seen for all parameter combinations. (g) Model variables were maximally correlated with [DA] signals ~0.5s later, consistent with a slight delay caused by the time taken by the brain to process cues, and by the FSCV technique.

2.2.4 Abrupt dopamine changes encode RPEs

FSCV electrode signals tend to drift over a timescale of minutes, so standard practice is to assess [DA] fluctuations relative to a pre-trial 'baseline' of unknown

concentration (as in Figure 2.3). Presented this way, reward cues appeared to evoke a higher absolute [DA] level when rewards were less common (Figure 2.5a,b), consistent with a conventional RPE-based account of phasic [DA]. However, our model implies a different interpretation of this data (Figures 2.4b and 2.5c). Rather than a jump from a fixed to a variable [DA] level (that encodes RPE), we predicted that the reward cue actually causes a [DA] jump from a variable [DA] level (reflecting variable estimates of upcoming reward) to a fixed [DA] level (that encodes the time-discounted value of the now certain reward).

To test these competing accounts, we compared [DA] levels between consecutive pairs of rewarded trials with Side-In events < 30 s apart (that is, well within the accepted stable range of FSCV measurements (Heien et al. 2005); for included pairs of trials, the average time between side-in events was 11.5 s). If the [DA] level evoked by the reward cue reflects RPE, then this level should tend to decline as rats experience consecutive rewards (Figure 2.5d,e). However, if [DA] represents state value, then baseline [DA] should asymptotically increase with repeated rewards while reward cue-evoked [DA] remains more stable (Figure 2.5f,g). The latter proved correct (Figure 2.5h,i). These results provide clear further evidence that [DA] reflects reward expectation (the value function), not just RPE.

Considering the microdialysis and FSCV results together, a parsimonious interpretation is that, across multiple measurement timescales, [DA] simply signals estimated availability of reward. The higher minute-by-minute [DA] levels observed with greater reward rate reflect both the higher values of states distal to rewards (including baseline periods between active trial performance) and the greater proportion of time spent in high-value states proximal to rewards.

By conveying an estimate of available reward, mesolimbic [DA] could be used as a motivational signal, helping to decide whether it is worthwhile to engage in effortful activity. At the same time, abrupt relative changes in [DA] could be detected and used as an RPE signal for learning. But is the brain actually using [DA] to signal motivation or learning, or both, during this task?

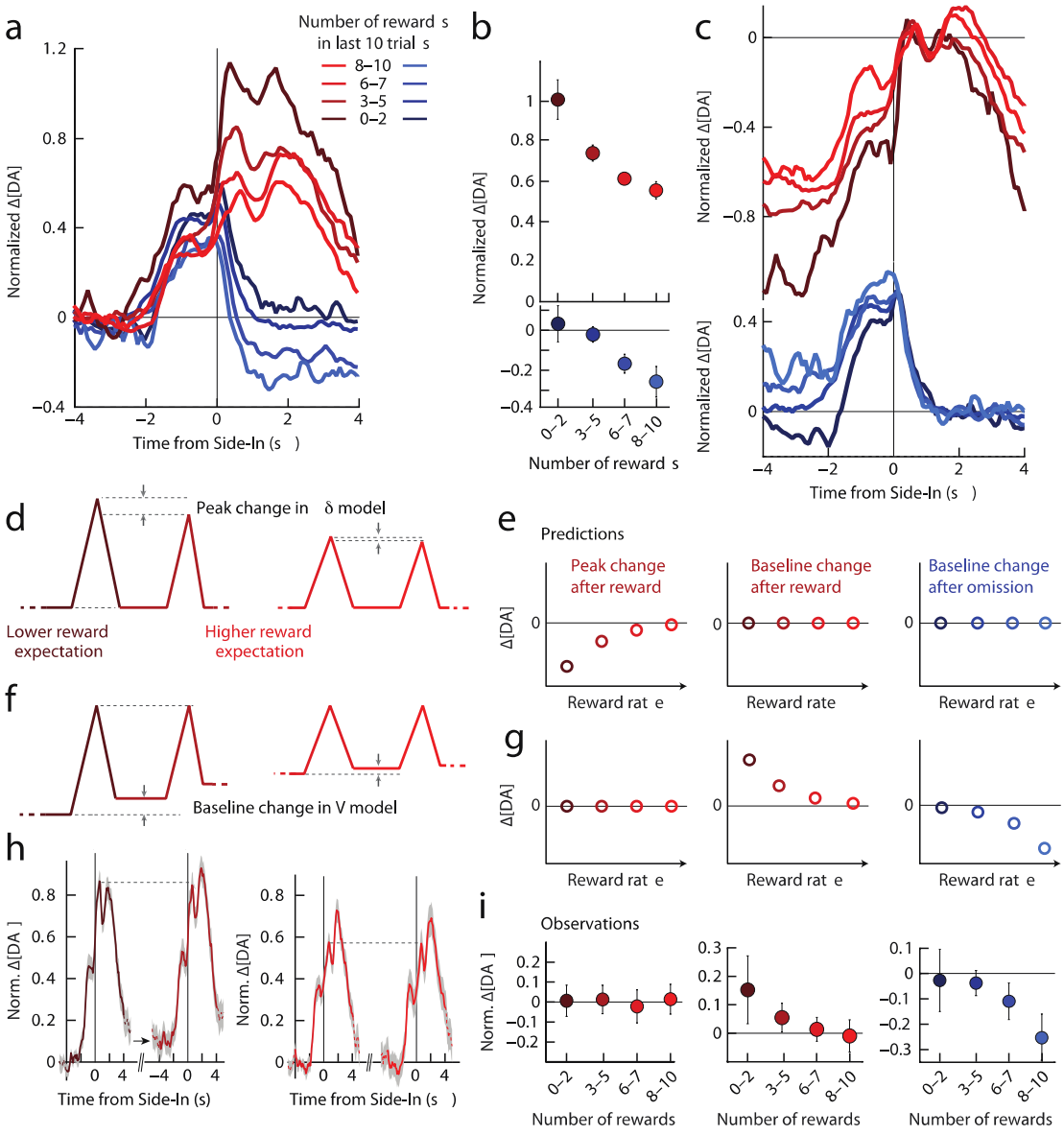


Figure 2.5. Between-trial dopamine shifts reflect updated state values. (a) Less-expected outcomes provoke larger changes in [DA]. [DA] data from all FSCV sessions together (as in Fig.3d), broken down by recent reward history and shown relative to pre-trial "baseline" (-3 to -1s relative to Center-In). Note that the [DA] changes after reward omission last at least several seconds (shift in level), rather than showing a highly transient dip followed by return to baseline as might be expected for encoding RPEs alone. (b) Quantification of [DA] changes, between baseline and reward feedback (0.5-1.0s after Side-In for rewarded trials, 1s-3s after Side-In for unrewarded trials). Error bars show SEM. (c) Same data as (a), but plotted relative to [DA] levels *after* reward feedback. These [DA] observations are consistent with a variable "baseline" whose level depends on recent reward history (as in Fig.4b model). (d) Alternative accounts of [DA] make different predictions for between-trial [DA] changes. When reward expectation is low, rewarded trials provoke large RPEs, but across repeated consecutive rewards RPEs should decline. Therefore if absolute [DA] levels encode RPE, the peak [DA] evoked by the reward-cue should decline between consecutive rewarded trials (and baseline levels should not change). For simplicity this cartoon omits detailed within-trial dynamics. (e)

Predicted pattern of [DA] change under this account, which also does not predict any baseline shift after reward omissions (right). (f) If instead [DA] encodes state values, then peak [DA] should not decline from one reward to the next, but the baseline level should increase (and decrease following unrewarded trials). (g) Predicted pattern of [DA] change for this alternative account. (h) Unexpected rewards cause a shift in baseline, not in peak [DA]. Average FSCV data from consecutive pairs of rewarded trials (all FSCV sessions combined, as in a), shown relative to the pre-trial baseline of the first trial in each pair. Data were grouped into lower reward expectation (left pair of plots, 165 total trials; average time between Side-In events = 11.35s +/- 0.22s SEM) and higher reward expectation (right pair of plots, 152 total trials; time between Side-In events = 11.65s +/- 0.23s) by a median split of each individual session (using # rewards in last 10 trials). Dashed lines indicate that reward cues evoked a similar absolute level of [DA] in the second rewarded trial, compared to the first. Black arrow indicates the elevated pre-trial [DA] level for the second trial in the pair (mean change in baseline [DA] = 0.108, p=0.013, one-tailed Wilcoxon signed rank test). No comparable change was observed if the first reward was more expected (right pair of plots; mean change in baseline [DA] = 0.0013, p=0.108, one-tailed Wilcoxon signed rank test). (i) [DA] changes between consecutive trials follow the pattern expected for value coding, rather than RPE coding alone.

2.2.5 Dopamine both enhances motivation and reinforces choices

To address this question, we turned to precisely timed, bidirectional, optogenetic manipulations of dopamine. Following an approach validated in previous studies (Steinberg et al. 2013), we expressed channelrhodopsin-2 (ChR2) selectively in dopamine neurons by combining *Th-Cre*$^{+}$ rats with DIO-ChR2 virus injections and bilateral optic fibers in the ventral tegmental area. We chose optical stimulation parameters (10-ms pulses of blue light at 30 Hz, 0.5-s total duration; Figure 2.6a,b) that produced phasic [DA] increases of similar duration and magnitude to those naturally observed with unexpected reward delivery. We provided this stimulation at one of two distinct moments during task performance. We hypothesized that enhancing [DA] coincident with Light-On would increase the estimated motivational value of task performance; this would make the rat more likely to initiate an approach, leading to shorter latencies on the same trial. We further hypothesized that enhancing [DA] at the time of the major RPE (Side-

In) would affect learning, as reflected in altered behavior on subsequent trials. In each session, laser stimulation was given at only one of these two times, and on only 30% of trials (randomly selected) to allow within-session comparisons between stimulated and unstimulated trials.

Providing phasic [DA] at Side-In reinforced choice behavior: it increased the chance that the same left or right action was repeated on the next trial, whether or not the food reward was actually received (n = 6 rats, two-way ANOVA yielded significant main effects for laser, $F(1,5) = 224.0$, $P = 2.4 \times 10^{-5}$; for reward, $F(1,5) = 41.0$, $P = 0.0014$; without a significant laser × reward interaction, $P = 0.174$; Figure 2.6c). No reinforcing effect was seen if the same optogenetic stimulation was given in littermate controls (n = 6 *Th-Cre*$^{-}$ rats, laser main effect $F(1,5) = 2.51$, $P = 0.174$; Figure 2.6c). For a further group of *Th-Cre*$^{+}$ animals (n = 5), we instead used the inhibitory opsin Halorhodopsin (eNpHR3.0). Inhibition of dopamine cells at Side-In reduced the probability that the same left or right choice was repeated on the next trial (laser main effect $F(1,4) = 18.7$, $P = 0.012$; without a significant laser × reward interaction, $P = 0.962$). A direct comparison between these three rat groups also demonstrated a group-specific effect of Side-In laser stimulation on choice reinforcement (two-way ANOVA, laser × group interaction $F(2,14) = 69.4$, $P = 5.4 \times 10^{-8}$). These observations support the hypothesis that abrupt [DA] fluctuations serve as an RPE learning signal, consistent with prior optogenetic manipulations (Berridge 2007). However, extra [DA] at Side-In did not affect subsequent trial latency, indicating that our artificial [DA] manipulations reproduced some, but not all, types of behavioral change normally evoked by rewarded trials.

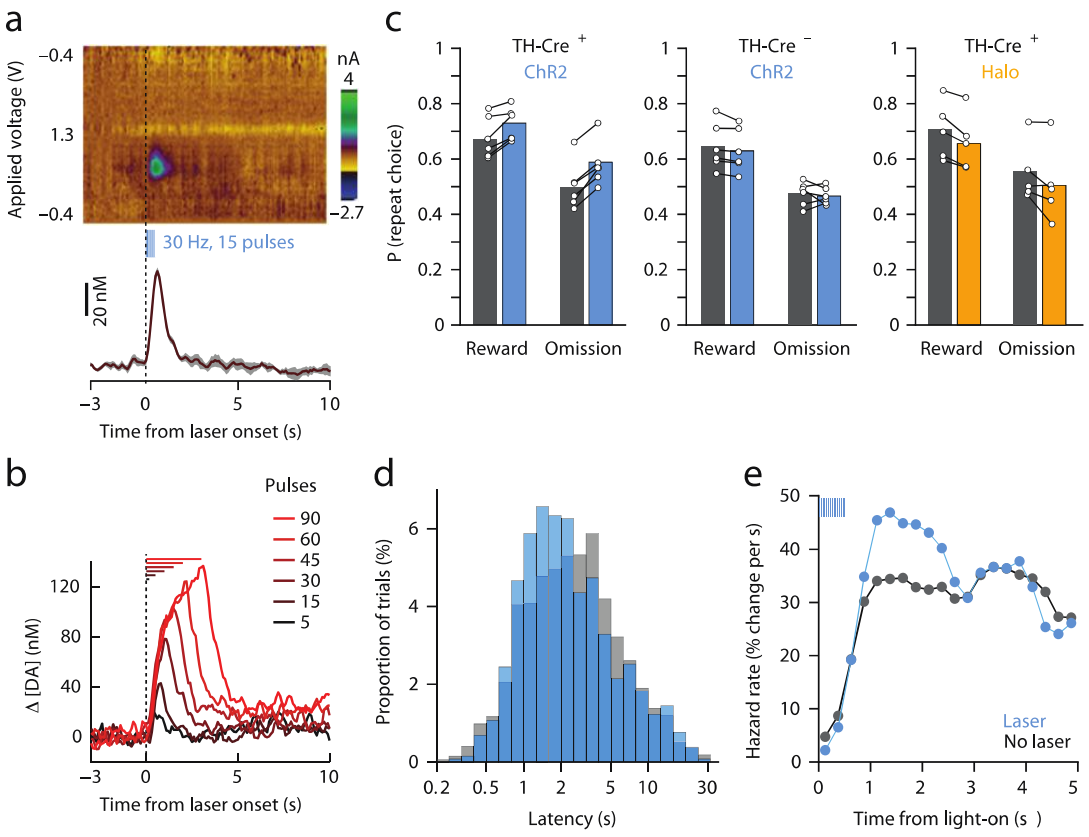


Figure 2.6. Phasic dopamine manipulations affect both learning and motivation. (a) FSCV measurement of optogenetically-evoked [DA] increases. Optic fibers were placed above VTA, and [DA] change examined in nucleus accumbens core. Example shows dopamine release evoked by a 0.5s stimulation train (average of 6 stimulation events, shaded area indicates +/-SEM). (b) Effect of varying the number of laser pulses on evoked dopamine release, for the same 30Hz stimulation frequency. (c) Dopaminergic stimulation at Side-In reinforces the chosen left or right action. *Left,* in *TH-Cre*$^{+}$ rats stimulation of ChR2 increased the probability that the same action would be repeated on the next trial. Circles indicate average data for each of 6 rats (3 sessions each, 384 trials/session ± 9.5 SEM). *Middle,* this effect did not occur in *TH-Cre*$^{-}$ littermate controls (6 rats, 3 sessions each, 342±7 trials/session). *Right,* in *TH-Cre*$^{+}$ rats expressing Halorhodopsin, orange laser stimulation at Side-In reduced the chance that the chosen action was repeated on the next trial (5 rats, 3 sessions each, 336±10 trials/session). (d) Laser stimulation at Light-On causes a shift towards sooner engagement, if the rats were not already engaged. Latency distribution (on log scale, 10 bins per log unit) for non-engaged, completed trials in *TH-Cre*$^{+}$ rats with ChR2 (n=4 rats with video analysis). (e) Same latency data as d, but presented as hazard rates. Laser stimulation (blue ticks at top left) increases the chance that rats will decide to initiate an approach, resulting in more Center-In events 1-2s later (for these n=4 rats, one-way ANOVA on hazard rate $F(1,3) = 18.1$, $p=0.024$).

Optogenetic effects on reinforcement were temporally specific: providing extra [DA] at Light-On (instead of Side-In) on trial *n* did not affect the probability that rats

made the same choice on trial $n + 1$ (laser main effect $F(1,5) = 0.031$, $P = 0.867$) nor did it affect the probability that choice on trial n was the same as trial $n - 1$ (laser main effect $F(1,5) = 0.233$, $P = 0.649$).

By contrast, extra [DA] at Light-On markedly affected latency for that very same trial (Figure 2.6d**).** The effect on latencies depended on what the rat was doing at the time of Light-On (two-way ANOVA yielded a significant laser × engaged interaction, $F(1,3) = 28.1$, $P = 0.013$). If the rat was already engaged in task per- formance, the very short latencies became slightly longer on average (median control latency = 0.45 s, median stimulated latency = 0.61 s; simple main effect of laser, $F(1,3) = 10.4$, $P = 0.048$). This effect apparently resulted from additional laser-evoked orienting movements on a subset of trials. By contrast, for non-engaged trials extra [DA] significantly reduced latencies (median control latency = 2.64 s, median stimulated latency = 2.16 s; simple main effect of laser, $F(1,3) = 32.5$, $P = 0.011$; Figure 2.6d). These optogenetic results are consistent with the idea that mesolimbic [DA] is less important for the initiation of simple, cue-evoked responses when a task is already underway (Nicola 2010), but is critical for motivating 'flexible approach' behaviors (Ikemoto and Panksepp 1999).

The shorter latencies produced by extra [DA] were not the result of rats approaching the start port at faster speeds, as the average approach trajectory was unaffected. Instead, extra [DA] transiently increased the probability that rats initiated the approach behavior. As the approach itself lasted ~1–2 s, the result was an increased rate of Center-In events ~1–2 s after the laser pulse train (Figure 2.6e). This effect of Light-On laser stimulation on hazard rates was dependent on rat group (two-way ANOVA, laser × group interac- tion $F(2,14) = 26.28$, $P = 0.000018$). *Post hoc* pairwise comparison of simple laser effects showed a significant increase in hazard rate for *Th-Cre*$^{+}$ ChR2 rats

($F(1,14) = 62.06$, $P = 1.63 \times 10^{-6}$) and a significant reduction in hazard rate for *Th-Cre*$^{+}$ eNpHR3.0 rats ($F(1,14) = 6.31$, $P = 0.025$), with no significant change in *Th-Cre*$^{-}$ ChR2 rats ($F(1,14) = 2.81$, $P = 0.116$). Overall, we conclude that, beyond just correlating with estimates of reward availability, mesolimbic [DA] helps translate those estimates into decisions to work for reward.

2.3 Discussion

2.3.1 A dopamine value signal used for both motivation and learning

Our results help confirm a range of disparate prior ideas, while placing them within a newly integrated theoretical context. First, phasic [DA] has been previously related to motivated approach (Phillips et al. 2003; Roitman et al. 2004), reward expectation (Howe et al. 2013) and effort-based decision-making (Gan, Walton, and Phillips 2010), but our demonstration that [DA] specifically conveys the temporally discounted value of future rewards grounds this motivational aspect of dopamine fluctuations in the quantitative frameworks of machine learning and optimal foraging theory. This idea is also consistent with findings using other techniques; for example, fMRI signals in ventral striatum (often argued to reflect dopamine signaling) encode reward expectation in the form of temporally-discounted subjective value (Kable and Glimcher 2007).

Second, using the complementary method of microdialysis to assess slower changes, we partly confirmed proposals that reward rate is reflected specifically in increased [DA], which in turn enhances motivational vigor (Niv, Daw, and Dayan 2006). However, our critical argument is that this motivational message of reward availability

can dynamically change from moment to moment, rather than being an inherently slow (tonic) signal. Using optogenetics, we confirmed that phasic changes in [DA] levels immediately affect willingness to engage in work, supporting the idea that subsecond [DA] fluctuations promptly influence motivational decision-making (Satoh et al. 2003; Adamantidis et al. 2011). This dynamic [DA] motivation signal can help to account for detailed patterns of time allocation (Niyogi et al. 2014). For example, animals take time to reengage in task performance after getting a reward (the post-reinforcement pause), and this pause is longer when the next reward is smaller or more distant. This behavioral phenomenon has been a long-standing puzzle (Schlinger, Derenne, and Baron 2008), but fits well with our argument that the time-discounted value of future rewards, conveyed by [DA], influences the moment-by-moment probability (hazard rate) of engaging in work.

Third, we confirmed the vital role of fast [DA] fluctuations, including transient dips, in signaling RPEs to affect learning (Hart et al. 2014; K. M. Kim et al. 2012; Steinberg et al. 2013). However, a notable result from our analyses is that RPEs were conveyed by fast relative changes in the [DA] value signal, rather than by deviations from a steady (tonic) baseline. This interpretation explains for the first time, to the best of our knowledge, how [DA] can simultaneously provide both learning and motivational signals, an important gap in prior theorizing. Our results also highlight the importance of not assuming a consistent baseline [DA] level across trials in voltammetry studies.

One interesting implication is that, among the many postsynaptic mechanisms that are affected by dopamine, some are concerned more with absolute levels and others with fast relative changes. This possibility needs to be investigated further, together with the natural working hypothesis that [DA] effects on neuronal excitability are closely involved in motivational functions (du Hoffmann and Nicola 2014), whereas [DA] effects on

spike-timing-dependent-plasticity are responsible for reinforcement-driven learning (J. N. J. Reynolds, Hyland, and Wickens 2001). It is also intriguing that a pulse of increased [DA] sufficient to immediately affect latency, or to alter left or right choice on subsequent trials, does not appear to be sufficient to alter latency on subsequent trials. This suggests that state values and left and right action values (Samejima, Ueda, and Doya 2005) may be updated via distinct mechanisms or at different times in the trial.

Although dopamine is often labeled as a reward transmitter, [DA] levels dropped during reward consumption, consistent with findings that dopamine is relatively less important for consuming, and apparently enjoying, rewards (K. Berridge 2007; Cannon and Palmiter 2003). Mesolimbic [DA] has also been shown to not be required for performance of simple actions that are immediately followed by reward, such as pressing a lever once to obtain food (Ishiwari et al. 2004). Rather, loss of mesolimbic [DA] reduces motivation to work, in the sense of investing time and effort in activities that are not inherently rewarding or interesting, but may eventually lead to rewards (John D Salamone and Correa 2012). Conversely, increasing [DA] with drugs such as amphetamines increases motivation to engage in prolonged work, in both normal subjects and those with attention-deficit hyperactivity disorder (Rapoport et al. 1980; Wardle et al. 2011).

2.3.2 Dopamine and decision dynamics

Our interpretation of mesolimbic [DA] as signaling the value of work is based on rat decisions to perform our task rather than alternative 'default' behaviors, such as grooming or local exploration. In this view, mesolimbic [DA] helps to determine whether to work, but not which activity is most worthwhile (that is, it is activational more than

directional (John D Salamone and Correa 2012)). It may be best considered as signaling the overall motivational excitement associated with reward expectation or, equivalently, the perceived opportunity cost of sloth (Niv, Daw, and Dayan 2006; Niyogi et al. 2014).

Based on prior results (Gan, Walton, and Phillips 2010), we expect that [DA] signals reward availability without factoring in the costs of effortful work, but we did not parametrically vary such costs here. Other notable limitations of our study are that we only examined [DA] in the nucleus accumbens and we did not selectively manipulate [DA] in various striatal subregions (and other dopamine targets). Our functional account of [DA] effects on behavioral performance is undoubtedly incomplete and it will be important to explore alternative descriptions, especially more generalizable accounts that apply throughout the striatum. In particular, our observation that mesolimbic [DA] affects the hazard rate of decisions to work seems compatible with a broader influence of striatal [DA] over decision-making, such as setting 'thresholds' for decision process completion (Gan, Walton, and Phillips 2010; Nagano-Saito et al. 2012; Leventhal et al. 2014). In sensorimotor striatum, dopamine influences the vigor (and learning) of more elemental actions (Leventhal et al. 2014; Turner and Desmurget 2010), and it has been shown that even saccade speed in humans is best predicted by a discounting model that optimizes the rate of reward (Haith, Reppert, and Shadmehr 2012). In this way, the activational, invigorating role of [DA] on both simple movements and motivation may reflect the same fundamental, computational-level mechanism applied to decision-making processes throughout striatum, affecting behaviors across a range of timescales.

Activation signals are useful, but not sufficient, for adaptive decision-making in general. Choosing between alternative, simultaneously available courses of action requires net value representations for the specific competing options (Gan, Walton, and

Phillips 2010; Morris et al. 2006). Although different subpopulations of dopamine neurons may carry somewhat distinct signals (Matsumoto and Hikosaka 2009), the aggregate [DA] message received by target regions is unlikely to have sufficient spatial resolution to represent multiple competing values simultaneously (Dreyer et al. 2010) or sufficient temporal resolution to present them for rapid serial consideration (McClure, Daw, and Montague 2003). By contrast, distinct ensembles of GABAergic neurons in the basal ganglia can dynamically encode the value of specific options, including through ramps-to-reward (Tachibana and Hikosaka 2012; van der Meer and Redish 2011) that may reflect escalating bids for behavioral control. Such neurons are modulated by dopamine, and in turn provide key feedback inputs to dopamine cells that may contribute to the escalating [DA] patterns observed here.

2.3.3 Relationship between dopamine cell firing and release

Firing rates of presumed dopamine cells have been previously reported to escalate in trials under some conditions (Fiorillo, Tobler, and Schultz 2003), but this has not been typically reported with reward anticipation. Several factors may contribute to this apparent discrepancy with our [DA] measures. The first is the nature of the behavioral task. Many important prior studies of dopamine (W Schultz, Dayan, and Montague 1997; Day et al. 2007), although not all (Morris et al. 2006), used Pavlovian situations, in which outcomes are not determined by the animal's actions. When effortful work is not required to obtain rewards, the learned value of work may be low and corresponding decision variables may be less apparent.

Second, a moving rat receives constantly changing sensory input, and may therefore more easily define and discriminate a set of discrete states leading up to reward

compared with situations in which elapsed time is the sole cue of progress. When such a sequence of states can be more readily recognized, it may be easier to assign a corresponding set of escalating values as reward gets nearer in time. Determining subjects' internal state representations, and their development during training, is an important challenge for future work. It has been argued that ramps in [DA] might actually reflect RPE if space is nonlinearly represented (Gershman 2014) or if learned values rapidly decay in time (Morita and Kato 2014). However, these suggestions do not address the critical relationship between [DA] and motivation that we aim to account for here.

Finally, release from dopamine terminals is strongly influenced by local microcircuit mechanisms in striatum (Threlfell et al. 2012) producing a dissociation between dopamine cell firing and [DA] in target regions. This dissociation is not complete: the ability of unexpected sensory events to drive a rapid, synchronized burst of dopamine cell firing is still likely to be of particular importance for abrupt RPE signaling at state transitions. More detailed models of dopamine release, incorporating dopamine cell firing, local terminal control and uptake dynamics, will certainly be needed to understand how [DA] comes to convey a value signal.

2.4 Methods

2.4.1 Animals and behavioral task.

All animal procedures were approved by the University of Michigan Committee on Use and Care of Animals. Male rats (300–500 g, either wild-type Long-Evans or *Th-Cre*$^{+}$ with a Long-Evans background (Witten et al. 2011) were maintained on a reverse 12:12 light:dark cycle and tested during the dark phase. Rats were mildly food deprived,

receiving 15 g of standard laboratory rat chow daily in addition to food rewards earned during task performance. Training and testing was performed in computer-controlled Med Associates operant chambers (25 cm × 30 cm at widest point) each with a five-hole nose-poke wall, as previously described (Gage et al. 2010; Leventhal et al. 2012; Schmidt et al. 2013). Training to perform the trial-and-error task typically took ~2 months, and included several pretraining stages (2 d to 2 weeks each, advancing when ~85% of trials were performed without procedural errors). First, any one of the five nosepoke holes was illuminated (at random), and poking this hole caused delivery of a 45-mg fruit punch–flavored sucrose pellet into the Food Port (FR1 schedule). Activation of the food hopper to deliver the pellet caused an audible click (the reward cue). In the next stage, the hole illuminated at trial start was always one of the three more-central holes (randomly-selected), and rats learned to poke and maintain hold for a variable interval (750–1,250 ms) until Go cue onset (250-ms duration white noise, together with dimming of the start port). Next, Go cue onset was also paired with illumination of both adjacent side ports. A leftward or rightward poke to one of these ports was required to receive a reward (each at 50% probability), and initiated the inter-trial interval (5–10 s randomly selected from a uniform distribution). If the rat poked an unlit center port (wrong start) or pulled out before the end of the hold period (false start), the house light turned on for the duration of an inter-trial interval. During this stage (only), to discourage development of a side bias, a maximum of three consecutive pokes to the same side were rewarded. Finally, in the complete trial-and-error task left and right choices had independent reward probabilities, each maintained for blocks of 40–60 trials (randomly selected block length and sequence for each session). All combinations of 10, 50 and 90% reward probability were used

except 10:10 and 90:90. There was no event that indicated to the rat that a trial would be unrewarded other than the omission of the Reward cue and the absence of the pellet.

For a subset of ChR2 optogenetic sessions, overhead video was captured at 15 frames per s. The frames immediately preceding the Light-On events were extracted, and the positions of the nose tip and neck were marked (by scorers blind to whether that trial included laser stimulation). These positions were used to determine rat distance and orientation to the center port (the one that will be illuminated on that trial). Each trial was classified as 'engaged' or 'unengaged', using cutoff values of distance (10.6 cm) and orientation (84°) that minimized the overlap between aggregate distributions. To assess how path length was affected by optogenetic stimulation, rat head positions were scored for each video frame between Light-On and Center-Nose-In. Engaged trials were further classified by whether the rat was immediately adjacent to one of the three possible center ports, and if that port was the one that became illuminated at Light-On or not (that is lucky, unlucky guesses).

Smoothing of latency (and other) time series for graphical display (Figure 2.1b,c) was performed using the MATLAB *filtfilt* function with a seven-trial window. To quantify the impact of prior trial rewards on current trial latency, we used a multiple regression model

$$\log_{10}(\text{latency}) \sim \beta_1 r_1 + \beta_2 r_{t-2} + \cdots + \beta_{10} r_{t-10}$$

where $r = 1$ if the corresponding trial was rewarded. All latency analyses excluded trials of zero latency (that is those for which the rat's nose was already inside the randomly-chosen center port at Light-On). For analysis of prior trial outcomes on left/right choice

behavior we used another multiple regression model, just as previously described (Lau and Glimcher 2005).

Latency survivor curves were calculated simply as the proportion of trials for which the Center-In event had not yet occurred, at each 250-ms interval after Light-On (an inverted cumulative latency distribution), smoothed with a three-point moving average ($x_t' = 0.25x_{t-1} + 0.5x_t + 0.25x_{t+1}$). These survivor curves were then used to calculate hazard rates, as the fraction of the remaining latencies that occurred in each 250-ms bin (the number of Center-In events that happened, divided by the number that could have happened).

We defined reward rate as the exponentially weighted moving average of individual rewards (a leaky integrator) (Simen, Cohen, and Holmes 2006; Daw, Kakade, and Dayan 2002; Sugrue, Corrado, and Newsome 2004). For each session the integrator time constant was chosen to maximize the (negative) correlation between reward rate and behavioral latency. If instead we defined reward rate as simply the number of rewards during each minute (ignoring the contributions of trials in previous minutes to current reward rate), the relationship between microdialysis- measured [DA] in that minute and reward rate was lower, although still significant ($R^2 = 0.084$, $P = 5.5 \times 10^{-10}$).

An important parameter in reinforcement learning is the degree to which agents choose the option that is currently estimated to be the best (exploitation) versus trying alternatives to assess whether they are actually better (exploration), and dopamine has been proposed to mediate this trade-off (Humphries, Khamassi, and Gurney 2012; Beeler, Frazier, and Zhuang 2012). To assess this we examined left/right choices in the second half of each block, by which time choices have typically stabilized (Figure 2.1d; this

behavioral pattern was also seen for the microdialysis sessions). We defined an exploitation index as the proportion of trials for which rats choose the better option in these second block halves (so values close to 1 would be fully exploitative, and values close to 0.5 would be random/exploratory). As an alternative metric of exploration/exploitation, we examined the number of times that the rat switched between left and right choices in each minute; this metric also showed no significant relationship to any neurochemical assayed in our microdialysis experiments.

2.4.2 Microdialysis

After 3–6 months of behavioral training rats were implanted with guide cannulae bilaterally above the nucleus accumbens core (NAcc; +1.3–1.9 mm AP, 1.5 mm ML from bregma) and allowed to recover for at least 1 week before retraining. On test days (3–5 weeks after cannula implantation) a single custom-made microdialysis probe (300-µm diameter) with polyacrylonitrile membrane (Hospal; 20-kDa molecular weight cutoff) was inserted into NAcc, extending 1 mm below the guide cannula. Artificial CSF (composition in mM: $CaCl_2$ 1.2; KCl 2.7, NaCl 148, $MgCl_2$ 0.85; ascorbate, 0.25) was perfused continuously at 2 µl min^{-1}. Rats were placed in the operant chamber with the house light on for an initial 90min period of probe equilibration, after which samples were collected once every minute. Following five baseline samples the house light was extinguished to indicate task availability.

For chemical analyses, we employed a modified version of our benzoyl chloride derivatization and HPLC-MS analysis method (Song, Mabrouk, et al. 2012). Immediately after each 2-µl sample collection, we added 1.5 µl of buffer (sodium carbonate monohydrate 100 mM), 1.5 µl of 2% benzoyl chloride in acetonitrile, and 1.5 µl of a

13C- labeled internal standard mixture (total mixture volume 6.5 µl). The mixture was vortexed for 2 s between each reagent addition. Since ACh is a quaternary amine and thus not derivatized by benzoyl chloride, it was directly detected in its native form (transition 146→87). Deuterated ACh (d4-ACh) was also added to the internal standard mixture for improved ACh quantification (Song, Hershey, et al. 2012) 5 µl of the sample mixture was automatically injected by a Thermo Accela HPLC system (Thermo Fisher Scientific) onto a reverse-phase Kinetex biphenyl HPLC column (2.1 mm × 100 mm; 1.7 particle size; Phenomenex). The HPLC system was interfaced to a HESI II ESI probe and Thermo TSQ Quantum Ultra (Thermo Scientific) triple quadrupole mass spectrometer operating in positive mode. Sample run times for all analytes were 3 min. To quantify neurochemicals in dialysate samples, we constructed six-point external calibration curves encompassing known physiological concentrations. Thermo Xcalibur 2.1 software (Thermo Fisher Scientific) automatically detected chromatographic peaks and quantified concentrations. To reduce noise each resulting minute-by-minute time series was smoothed with a three-point moving average (as above), then converted to Z-scores to facilitate comparison between subjects.

Regression analysis of microdialysis data was performed stepwise. We first constructed models with only one behavioral variable as predictor and one outcome (analyte). If two behavioral variables showed a significant relationship to a given analyte, we constructed a model with both behavioral variables and an interaction term, and examined the capacity of each variable to explain analyte variance without substantial multicollinearity.

To determine cross-correlogram statistical thresholds we first shuffled the time series for all sessions 200,000 times, and calculated the average Pearson correlation

coefficients (that is the zero-lag cross-correlation) for each shuffled pair of time series. Thresholds were based on the tails of the resulting distribution: that is for uncorrected two-tailed alpha = 0.05 we would find the levels for which 2.5% of the shuffled values lay outside these thresholds. As we wished to correct for multiple comparisons we divided alpha by the number of tests (276; number of cross-correlograms = 23 timeseries × 22 timeseries divided by two, as the crosscorrelograms are just mirror-reversed when the order is changed, plus 23 autocorrelograms).

2.4.3 Voltammetry

FSCV electrode construction, data acquisition and analysis were performed as described (Aragona et al. 2009). Rats were implanted with a guide cannula above the right NAcc (+1.3–2.0 mm AP, 1.5 mm ML from bregma), a Ag/AgCl reference electrode (in the contralateral hemisphere) and a bipolar stimulation electrode aimed at the VTA (−5.2 mm AP, 0.8 mm ML, 7.5 mm DV). Carbon fiber electrodes were lowered acutely into the NAcc. Dopaminergic current was quantified offline by principal component regression (PCR) (Heien et al. 2005) using training data for dopamine and pH from electrical stimulations. Recording time points that exceeded the PCR residual analysis threshold (Qα) were omitted from further processing or analysis. Current to [DA] conversion was based on *in vitro* calibrations of electrodes constructed in the same manner with the same exposed fiber length. On many days data was not recorded due to electrode breakage or obvious movement-related electrical noise. FSCV recordings were made from 41 sessions (14 rats total). We excluded those sessions for which the rat failed to complete at least three blocks of trials, and those in which electrical artifacts caused >10% of trials to violate the assumptions of PCR residual analysis. The remaining ten

sessions came from six different rats. To avoid aggregate results being overly skewed by a single animal, we only included one session from each of the six rats (the session with the largest reward-evoked [DA] increase). Upon completion of FSCV testing, animals were deeply anesthetized and electrolytic lesions were created (40 μA for 15 s at the same depth as recording site) using stainless steel electrodes with 500 μm of exposed tip (AM Systems). Lesion locations were later reconstructed in Nissl-stained sections.

For between-session comparisons we normalized [DA] to the average [DA] difference between the pre-trial baseline and Food-Port-In aligned peak levels. To visualize the reward-history-dependence of [DA] change between consecutive trials (Figure 2.5h), we first extracted time series of normalized [DA] from consecutive pairs of rewarded trials (Side-In event to subsequent Side-In event separated by less than 30 s). For each session we divided these traces into 'low-reward-rate' and 'high-reward-rate' groups, using the (number of rewarded trials in the last 10) that best approximated a median-split (so low- and high- reward-rate groups had similar trial numbers). We then averaged all low-reward-rate traces, and separately all high-reward-rate traces.

2.4.4 Reinforcement learning model

To estimate the time-varying state value and RPE in each trial, we used a Semi-Markov Decision Process (Daw et al. 2006) with temporal difference learning, implemented in MATLAB. The model consisted of a set of states, with rat behavioral events determining the times of transitions between states. Each state was associated with a stored ('cached') value of entering that state, $V(s)$. At each state transition a reward prediction error □ □was calculated using

$$\delta_t = r_t + V_t(s_t) - \gamma^{-n} V_t(s_{t-n})$$

where n is the number of time steps since the last state transition (a time step of 50ms was used throughout), r is defined as one at reward receipt and zero otherwise, and γ specifies the rate at which future rewards are discounted at each timestep ($\gamma < 1$). The V terms in the equation compare the cached value of the new state to the value predicted, given the prior state value and the elapsed time since the last transition (as illustrated in Figure 2.4c). Each state also had $e(s)$, an eligibility trace that decayed with the same time parameter γ (following the terminology of ref. 21, this is a TD(1) model with replacing traces). RPEs updated the values of the states encountered up to that point, using

$$V'(s) = V(s) + \alpha \delta e_t(s)$$

where α is the learning rate. V and δ were defined only at state transitions, and V was constrained to be non-negative. The model was 'episodic', as all eligibilities were reset to zero at trial outcome (reward receipt, or omission). V is therefore an estimate of the time-discounted value of the next reward, rather than total aggregate future reward; with exponential discounting and best-fit parameters subsequent time-discounted rewards are negligible (but this would not necessarily be the case if hyperbolic discounting was used).

We also examined the effect of calculating prediction errors slightly differently

$$\delta_t = r_t + \gamma^{n} V_t(s_t) - V_t(s_{t-n})$$

This version compares a discounted version of the new state value to the previous state value. As expected, the results were the same. Specifically, overall [DA] correlation to V remained ~0.4, overall δ correlation was ~0.2, and each individual session [DA] was significantly better correlated to V than to δ, across the full parameter space.

We present results using γ in the 0.9 to 1 range, because 0.9 is already a very fast exponential discount rate when using 50-ms time steps. However we also tested smaller γ (0.05–0.9) and confirmed that the [DA]:δ correlation only diminished in this lower range (data not shown).

To compare within-trial [DA] changes to model variables, we identified all epochs of time (3 s before to 3 s after Center-In) with at least six state transitions (this encompasses both rewarded and unrewarded trials). Since the model can change state value instantaneously, but our FSCV signal cannot (Kile et al. 2012), we included an offset lag (so we actually compared V and δ to [DA] a few measurements later). The size of the lag affected the magnitude of the observed correlations (Figure 2.4f), but not the basic result. Results were also unchanged if (instead of a lag) we convolved model variables with a kernel consisting of an exponential rise and fall, demonstrating that our results are not a simple artifact of time delays associated with the FSCV method or sluggish reuptake. Finally, we also tried using the SMDP model with hyperbolic (instead of exponential) discounting (Mazur 1986; Ainslie 2005; Kobayashi and Schultz 2008; Kacelnik 1997), and again found a consistently stronger correlation between [DA] and V than between [DA] and δ (data not shown).

2.4.5 Optogenetics

We used three groups of rats to assess the behavioral effects of VTA DA cell manipulations (first *Th-Cre*$^{+}$ with AAV-EF1□-DIO-ChR2-EYFP virus, then littermate *Th-Cre*$^{-}$ with the same virus, then *Th-Cre*$^{+}$ with AAV- EF1□-DIO-eNpHR3.0-EYFP). All virus was produced at the University of North Carolina vector core. In each case rats received bilateral viral injections (0.5 or 1 μl per hemisphere at 50 nl min^{-1}) into the VTA (same coordinates as above). After 3 weeks, we placed bilateral optic fibers (200-□m diameter) under ketamine/xylazine anesthesia with FSCV guidance, at an angle of 6° from the sagittal plane, stopping at a location that yielded the most laser-evoked [DA] release in NAc. Once cemented in place, we used FSCV to test multiple sets of stimulation parameters from a 445-nm blue laser diode (Casio) with Arroyo Instruments driver under LabView control. The parameters chosen for behavioral experiments (0.5-s train of 10-ms pulses at 30 Hz, 20 mW power at tip) typically produced [DA] increases in *Th-Cre*$^{+}$ / ChR2 rats comparable to those seen with unexpected reward delivery. All rats were allowed to recover from surgery and retrained to pre-surgery performance. Combined behavioral / optogenetic experiments began 5 weeks after virus injection. On alternate days, sessions either included bilateral laser stimulation (on a randomly selected 30% of trials, regardless of block or outcome), or not. In this manner, each rat received three sessions of Light-On stimulations and three sessions of Side-In stimulation, interleaved with control (no laser) sessions, over a 2-week period. Halorhodopsin rats were tested with 1 s of constant 20-mW illumination from a 589-nm (yellow/orange) laser (OEM Systems), starting either at Light-On or Side-In as above. One *Th-Cre*$^{+}$

/ChR2 rat was excluded from analyses due to misplaced virus (no viral expression directly below the optic fiber tips).

For statistical analysis of optogenetic effects on behavior we used repeated measure ANOVA models, in SPSS. For each rat we first averaged data across the three sessions with the same optogenetic conditions. Then, to assess reinforcing effects we examined the two factors of LASER (off versus on) and REWARD (rewarded versus omission), with the dependent measure the probability that the same action was repeated on the next trial. For assessing effects on median latency we examined the two factors of LASER (off versus on) and ENGAGED (yes versus no). For assessing group-dependent effects on hazard rate we examined the factors of LASER (off versus on) and GROUP (*Th-Cre*$^{+}$ /ChR2; *Th-Cre*$^{-}$ /ChR2; *Th-Cre*$^{+}$ /eNpHR3.0), with the dependent measure the average hazard rate during the epoch 1–2.5 s after Light-On. This epoch was chosen since it is 1–2 s after the laser stimulation period (0–0.5 s) and approach behaviors have a consistent duration of ~1–2 s. *Post hoc* tests were Bonferroni-corrected for multiple comparisons.

Supplementary Figures

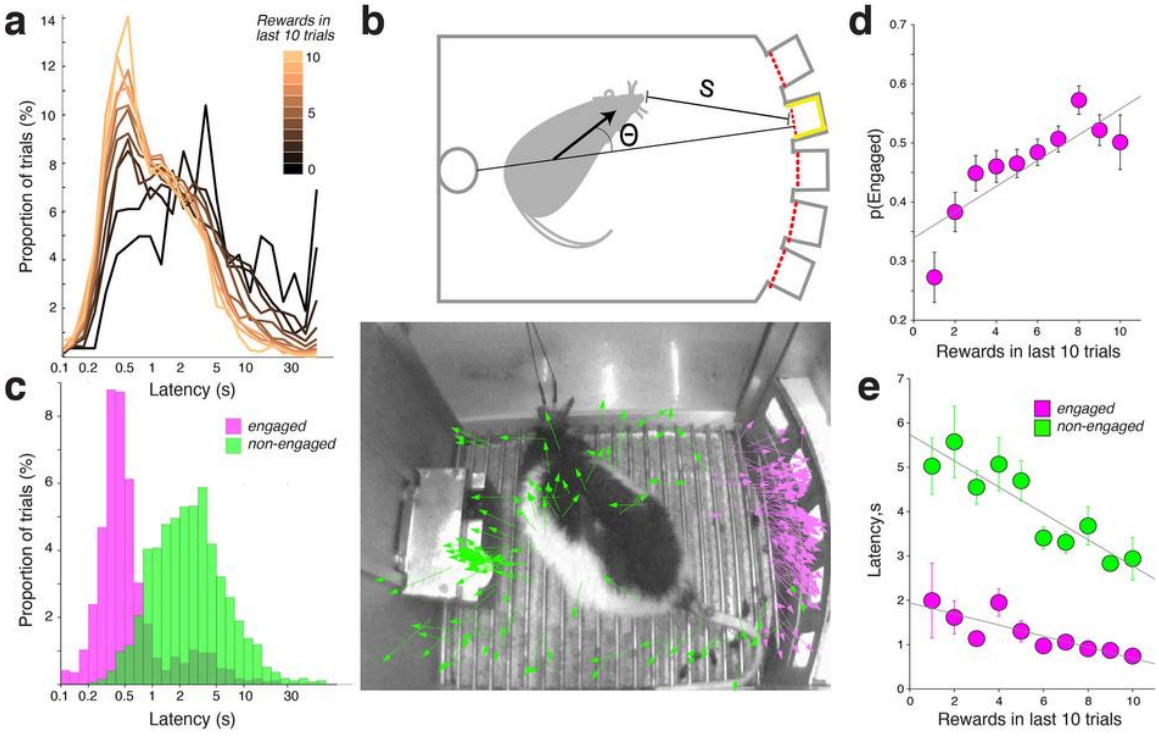


Figure 2.7. Reward rate affects the decision to begin work. Latency distributions are bimodal, and depend on reward rate. Very short latencies (early peak) preferentially occur when a greater proportion of recent trials have been rewarded (same data set as Fig 1d–i). (b) (top) Schematic of video analysis. Each trial was categorized as "engaged" (already waiting for Light-On) or non-engaged based upon distance (s) and orientation (θ) immediately before Light-On (see Methods). (bottom) Arrows indicate rat head position and orientation for engaged (pink) and non-engaged (green) trials (one example session shown). (c) Categorization into engaged, non-engaged trials accounts for bimodal latency distribution (data shown are all non-laser trials across 12 ChR2 sessions in TH-Cre+ rats). (d) Proportion of engaged trials increases when more recent trials have been rewarded (3336 trials from 4 rats, r=0.82, p=0.003). (e) Especially for non-engaged trials, latencies are lower when reward rate is higher (r=−0.11,p=0.004 for 1570 engaged trials, r=−0.18, p=5.2×10−19 for 1766 non-engaged trials).

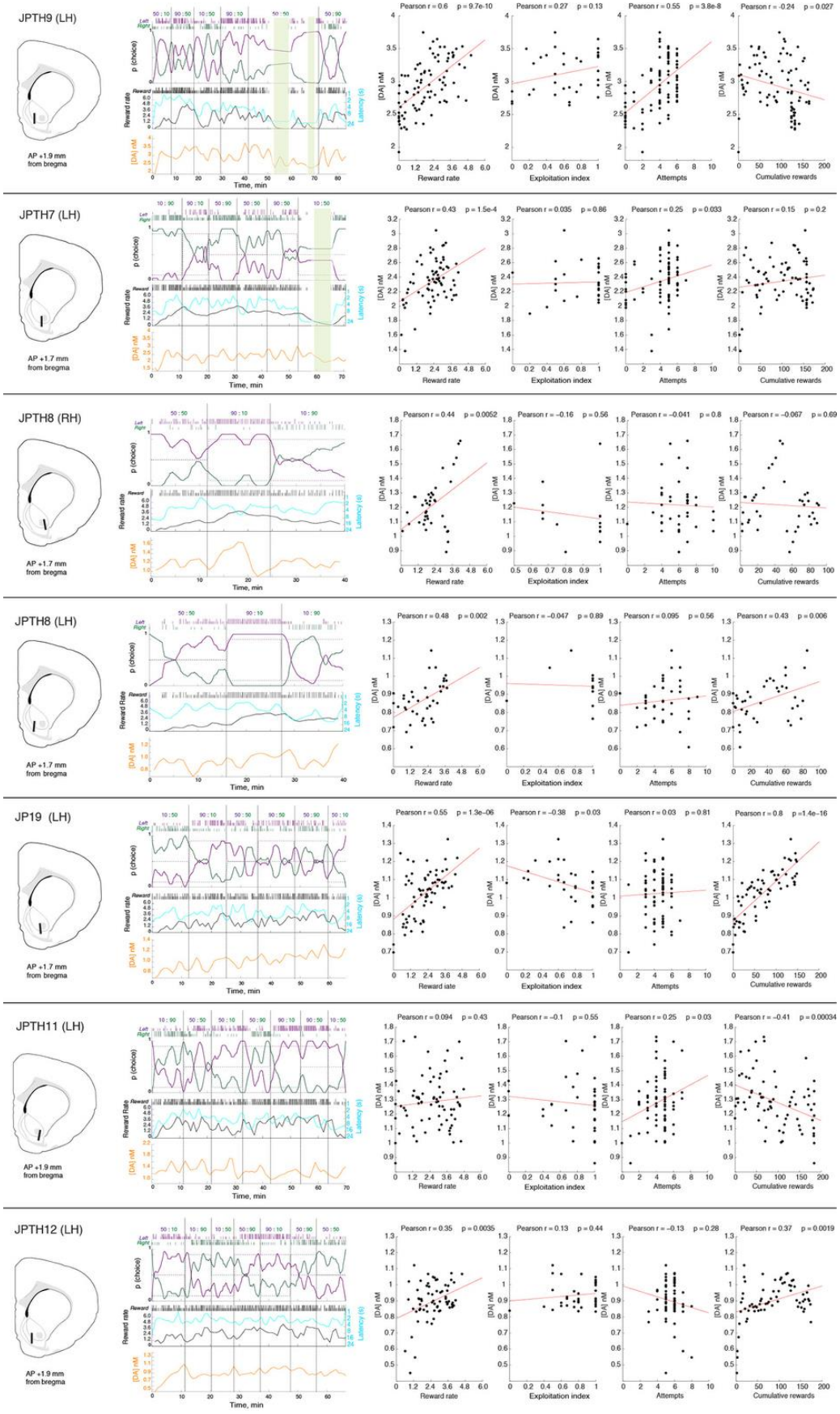

JPTH9 (LH)
AP +1.9 mm from bregma
Pearson r = 0.6 p = 9.7e-10
Pearson r = 0.27 p = 0.13
Pearson r = 0.55 p = 3.8e-8
Pearson r = -0.24 p = 0.027
JPTH7 (LH)
AP +1.7 mm from bregma
Pearson r = 0.43 p = 1.5e-4
Pearson r = 0.035 p = 0.86
Pearson r = 0.25 p = 0.033
Pearson r = 0.15 p = 0.2
JPTH8 (RH)
AP +1.7 mm from bregma
Pearson r = 0.44 p = 0.0052
Pearson r = −0.16 p = 0.56
Perason r = −0.041 p = 0.8
Pearson r = −0.067 p = 0.69
JPTH8 (LH)
AP +1.7 mm from bregma
Pearson r = 0.48 p = 0.002
Pearson r = −0.047 p = 0.89
Pearson r = 0.095 p = 0.56
Pearson r = 0.43 p = 0.006
JP19 (LH)
AP +1.7 mm from bregma
Pearson r = 0.55 p = 1.3e-06
Pearson r = −0.38 p = 0.03
Pearson r = 0.03 p = 0.81
Pearson r = 0.8 p = 1.4e-16
JPTH11 (LH)
AP +1.9 mm from bregma
Pearson r = 0.094 p = 0.43
Pearson r = −0.1 p = 0.55
Pearson r = 0.25 p = 0.03
Pearson r = −0.41 p = 0.00034
JPTH12 (LH)
AP +1.9 mm from bregma
Pearson r = 0.35 p = 0.0035
Pearson r = 0.13 p = 0.44
Pearson r = −0.13 p = 0.28
Pearson r = 0.37 p = 0.0019
p (choice)
Reward rate
[DA] nM
Latency (s)
Time, min
Reward rate
Exploitation index
Attempts
Cumulative rewards

Figure 2.8. Individual microdialysis sessions. Each row shows data for a different session, with indicated rat ID (e.g. IM463) and recording side (LH = left, RH=right). From left: dialysis probe location, behavioral and [DA] time courses, and individual session correlations to behavioral variables. Reward rate is in units of rewards per min. Numbers of microdialysis samples for each of the seven sessions: 86,72,39,39,68,73,67 respectively. The overall relationship between dopamine and reward rate remained highly significant even if excluding periods of inactivity (defined as no trials initiated for >2 minutes, shaded in green; regression $R^2 = 0.12$, $p = 1.4 \times 10^{-13}$).

Figure 2.9. Cross-correlograms for behavioral variables and neurochemicals. Each plot shows cross-correlograms averaged across all microdialysis sessions, all using the same axes (−20min to +20min lags, −0.5 to +1 correlation). Colored lines indicate statistical thresholds corrected for multiple comparisons (see Methods). Many neurochemical pairs show no evidence of covariation, but others display strong relationships including a cluster of glutamate, serine, aspartate and glycine.

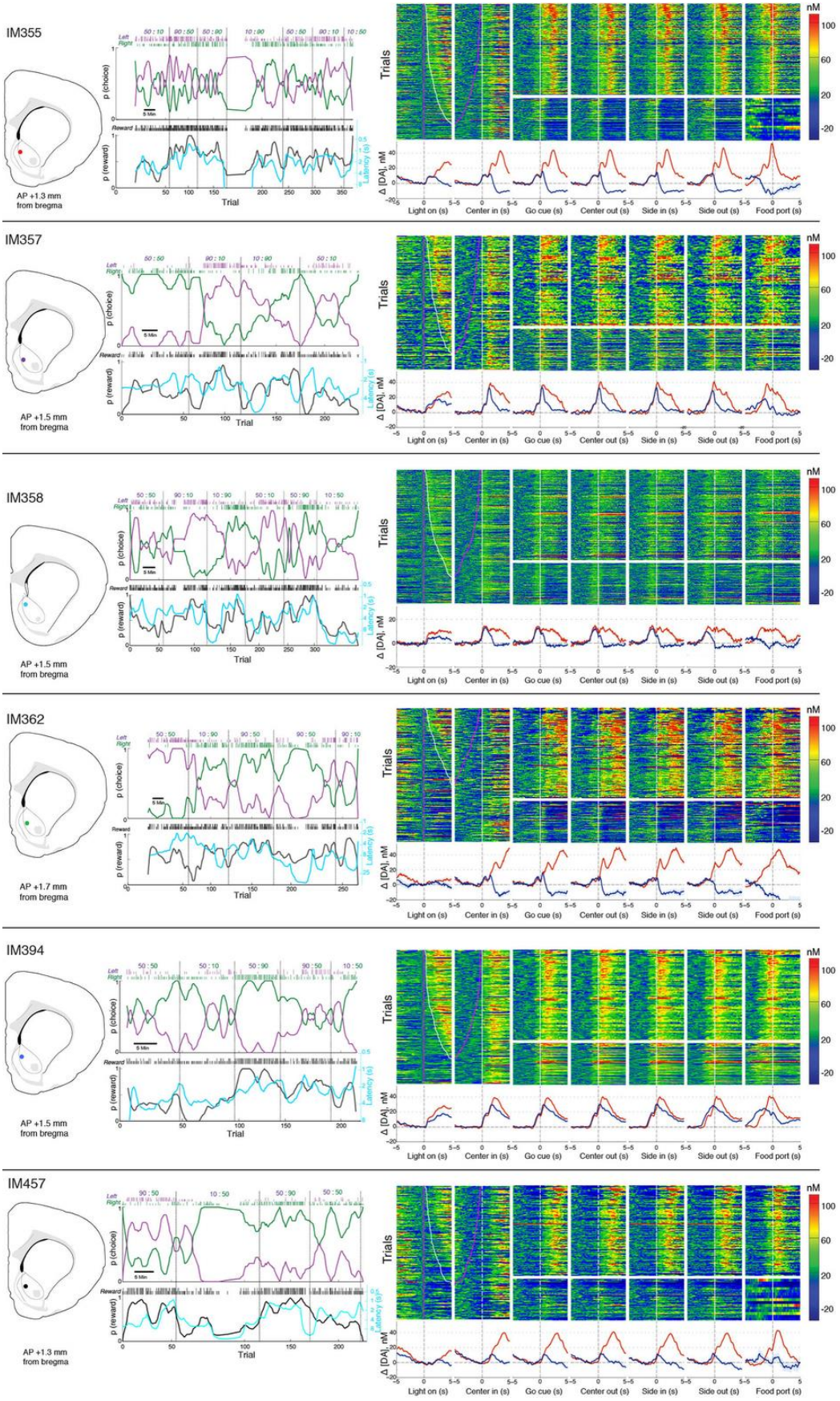

IM355
AP +1.3 mm
from bregma
IM357
AP +1.5 mm
from bregma
IM358
AP +1.5 mm
from bregma
IM362
AP +1.7 mm
from bregma
IM394
AP +1.5 mm
from bregma
IM457
AP +1.3 mm
from bregma
p (choice)
p (reward)
Reward
Trial
Trials
Δ [DA], nM
Light on (s)
Center in (s)
Go cue (s)
Center out (s)
Side in (s)
Side out (s)
Food port (s)
nM

Figure 2.10. Individual voltammetry sessions. Each row shows data for a different rat (e.g. IM355, which was also used as the example in Figs. 2.3, 2.4). At left, recording site within nucleus accumbens. Middle panels show behavioral data for the FSCV session (same format as Fig. 2.1). Right panels show individual FSCV data (same format as Fig. 2.3, but with additional event alignments).

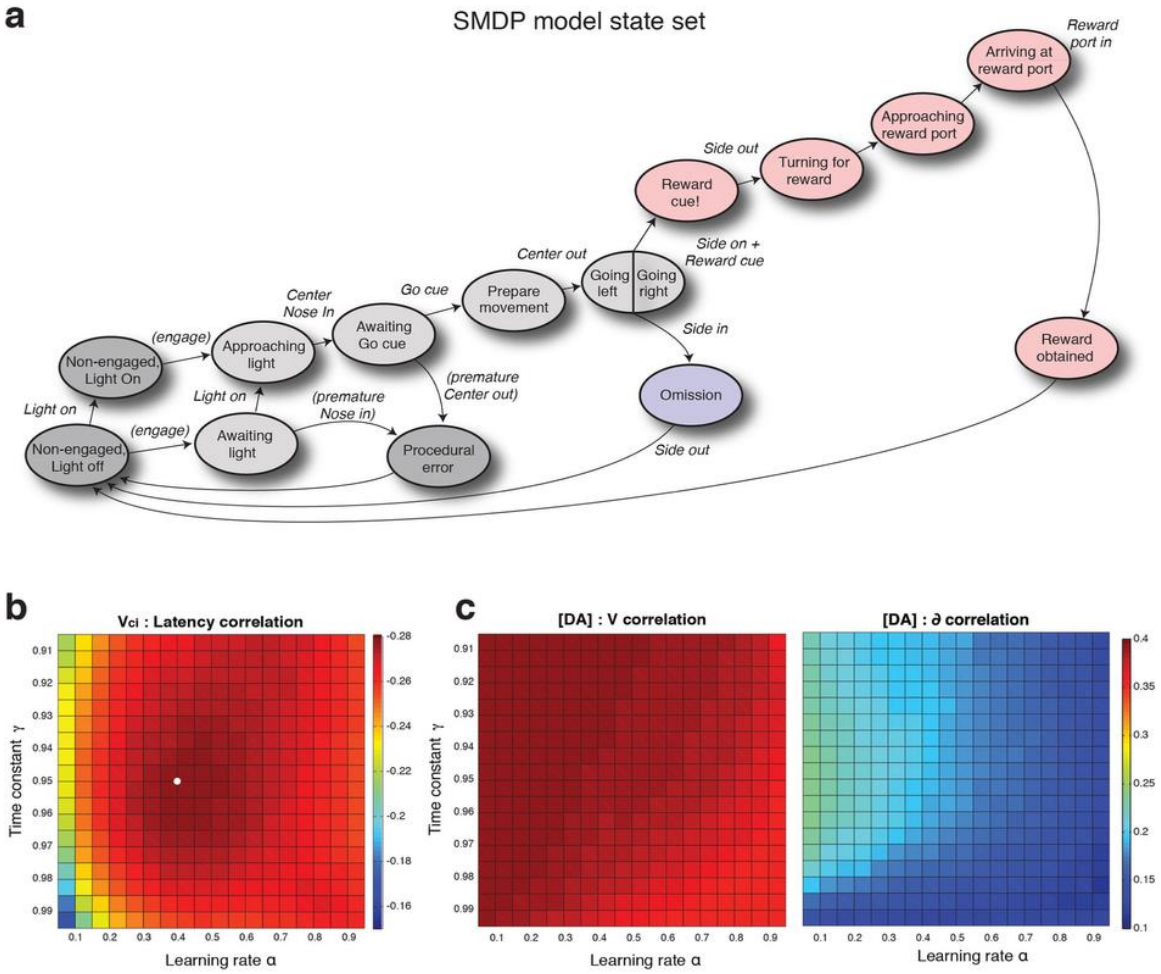
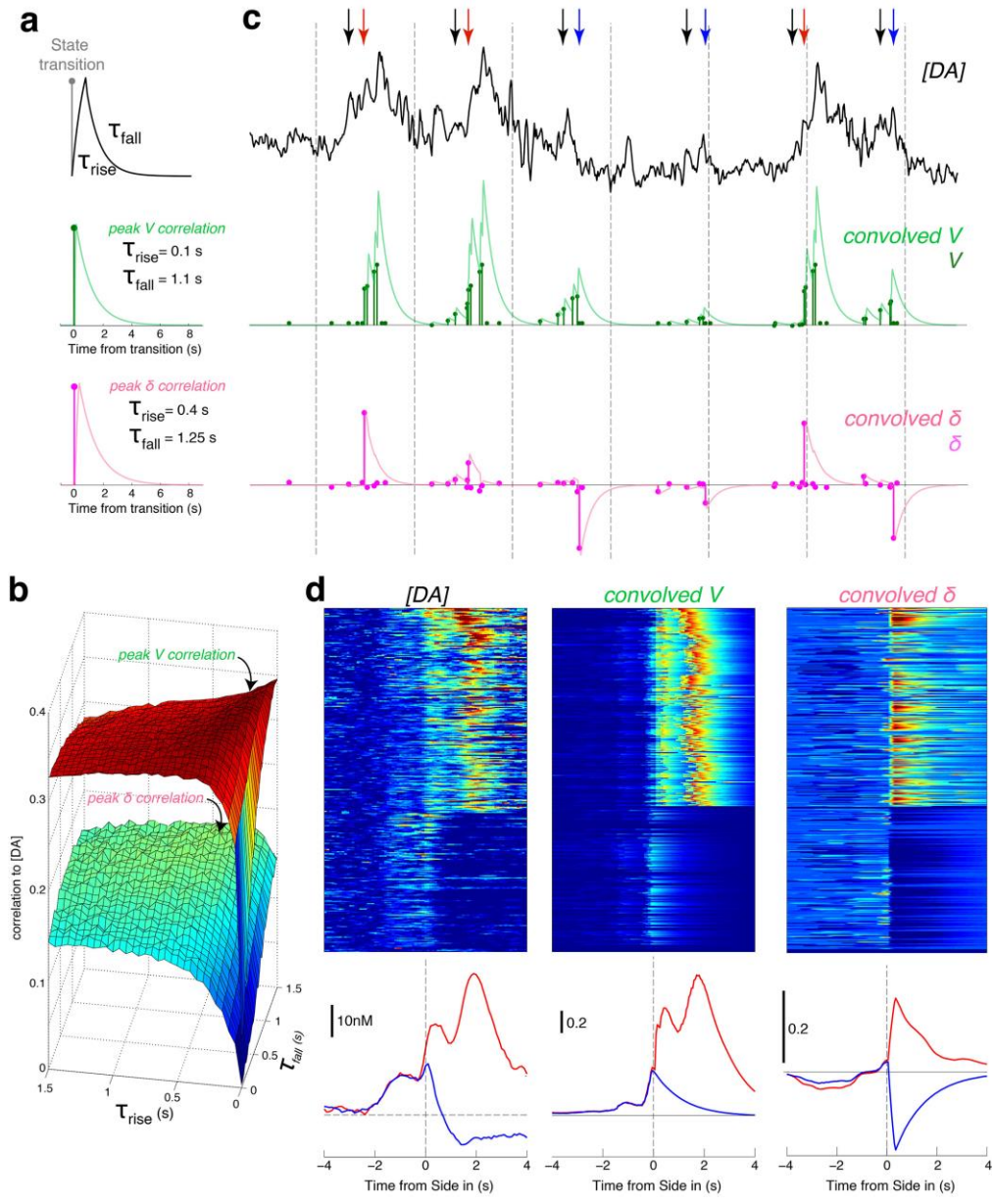


Figure 2.11. SMDP model. (**a**) Task performance was modeled as a sequence of transitions between states of the agent (rat). Each state had a single associated cached value V(s) (rather than, for example, separate state-action (Q) values for leftward and rightward trials). Most state transitions occur at variable times (hence "semi-Markov") marked by observed external events (Center-In, Go-Cue, etc). In contrast, the state sequence between Side-Out and Reward Port In is arbitrarily defined ("Approaching Reward Port" begins 1s before Reward Port In; "Arriving At Reward Port" begins 0.5s before Reward Port In). Changing the number or specific timing of these intermediate states does not materially affect the rising shape of the value function. (**b**) Average correlation (color scale = Spearman's r) between SMDP model state value at Center-In (Vci) and latency across all six FSCV rats, for a range of learning rates α and exponential discounting time constants ɣ. Note that color scale is inverted (red indicates strongest

negative relationship, with higher value corresponding to shorter latency). White dot marks point of strongest relationship (α=0.40, γ=0.95). (**c**) Correlation between [DA] and state value V is stronger than the correlation between [DA] and reward prediction error δ, across the same range of parameters. Color scale at right is the same for both matrices (Spearman's r).



Nature Neuroscience: doi:10.1038/nn.4173

Figure 2.12. Dopamine relationships to temporally-stretched model variables. (**a**) Kernel consisted of an exponential rise (to 50% of asymptote) and an exponential fall, with separate time constants. (**b**) Within-trial correlation coefficients between [DA] and kernel-convolved model variables V and δ, for a range of rise and fall time constants (0 –

1.5s each, in 50ms timesteps, using data from all 6 rats). Regardless of parameter values, [DA] correlations to V were always higher than to δ. (**c**) Same example data as Fig. 2.4E, but also showing convolved V and δ (using time constants that maximized correlation to [DA] in each case). (**d**) Trial-by-trial (top) and average (bottom) [DA], convolved V, and convolved δ, for the same session as Fig. 2.4d,e.

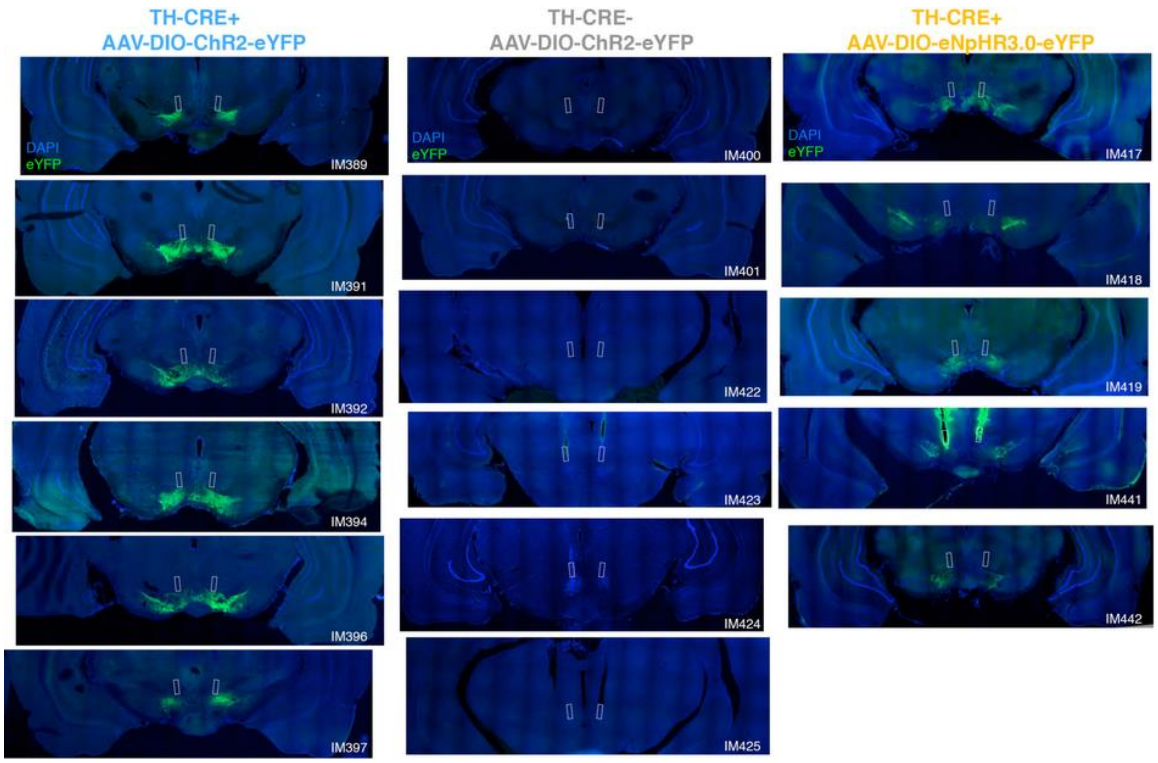


Figure 2.13. Histology for behavioral optogenetic experiments. Identifier (e.g. "IM389") for each rat is given at bottom right corner. Coronal sections shown are within 180μm (anterior-posterior) of the observed fiber tip location. Green indicates expression of eYFP, blue is DAPI counterstain. In a couple of cases (IM423, IM441) autofluorescence of damaged brain tissue is visible along the optic fiber tracts; this was not specific to the green channel.

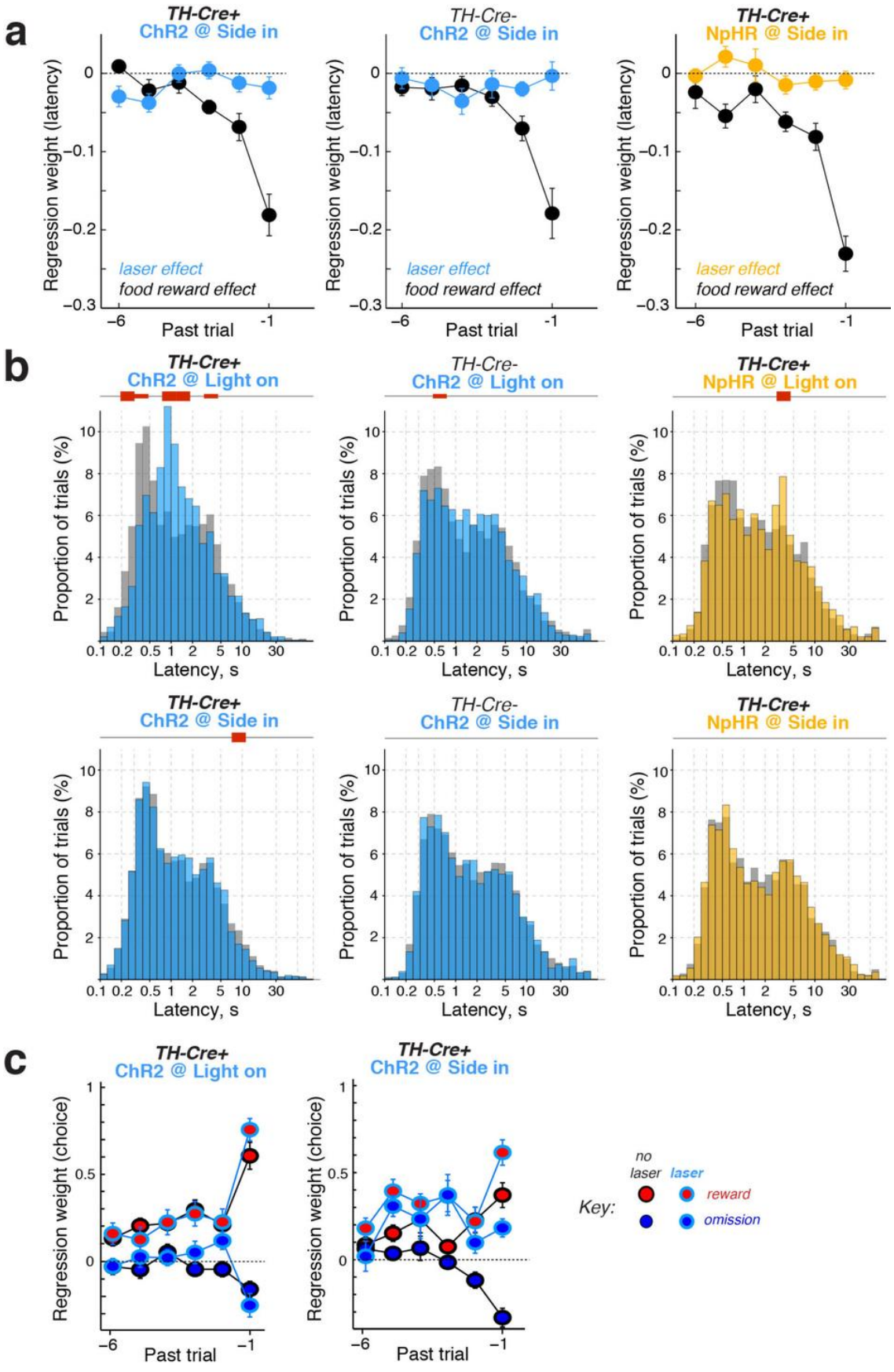

a
TH-Cre+
ChR2 @ Side in
TH-Cre-
ChR2 @ Side in
TH-Cre+
NpHR @ Side in
Regression weight (latency)
laser effect
food reward effect
Past trial
b
TH-Cre+
ChR2 @ Light on
TH-Cre-
ChR2 @ Light on
TH-Cre+
NpHR @ Light on
TH-Cre+
ChR2 @ Side in
TH-Cre-
ChR2 @ Side in
TH-Cre+
NpHR @ Side in
Proportion of trials (%)
Latency, s
c
TH-Cre+
ChR2 @ Light on
TH-Cre+
ChR2 @ Side in
Regression weight (choice)
Past trial
Key:
no laser
laser
reward
omission

Figure 2.14. Further analysis of persistence of optogenetic effects. (**a**) Regression analysis showing substantial effects of recent rewards (black) on latency, but no comparable effect of recent Side-In laser stimulations on latency. (**b**) Effects of Light-On [DA] manipulation on same-trial latency distributions (top), and of Side-In [DA] manipulation on next-trial latency distributions (bottom). Dataset shown is the same as Fig. 2.6c, i.e. all completed trials in TH-Cre^{+} rats with ChR2 (left), TH-Cre^{-} rats with ChR2 (middle) and TH-Cre^{+} rats with halorhodopsin (right). (**c**) Regression analysis of laser stimulation on subsequent left/right choices. Recent food rewards for a given left/right action increase the probability that it will be repeated. Extra [DA] at Light-On has little or no effect on subsequent choices, but extra [DA] at Side-In is persistently reinforcing. For the Side-In data, note especially the positive coefficients for otherwise unrewarded laser trials.

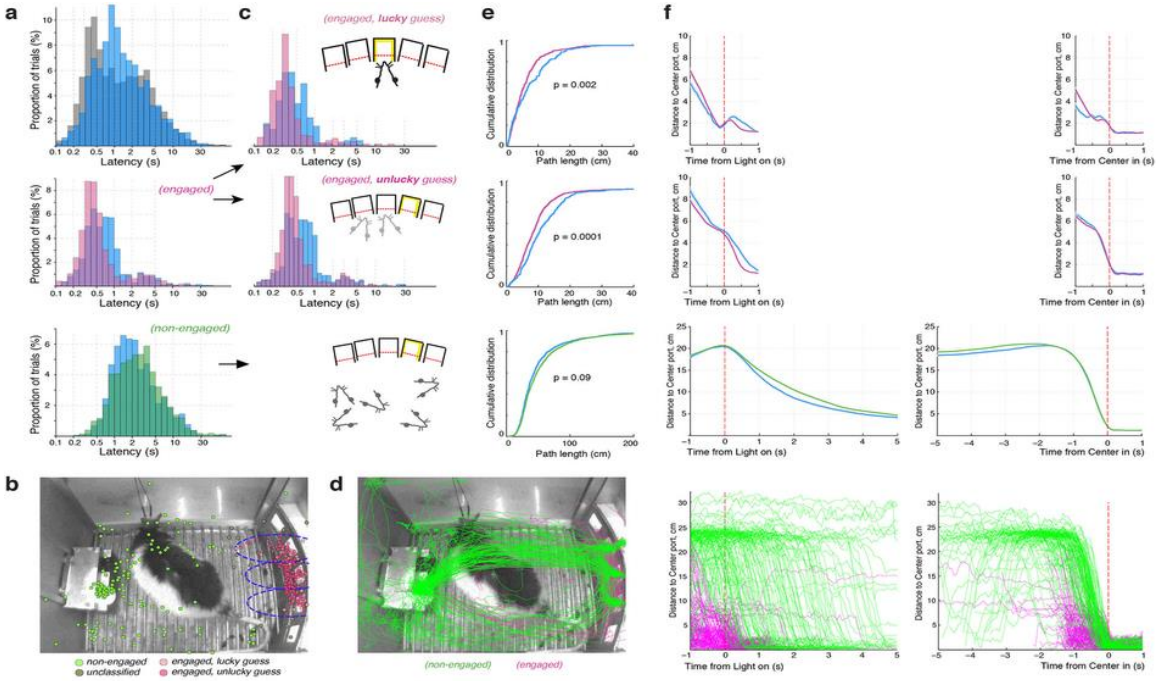


Figure 2.15. Video analysis of optogenetic effects on latency. (a) Extra [DA] at Light-On causes shorter latencies for non-engaged trials, but longer latencies for a subset of engaged trials. Top plot shows all trials (for the n=4 TH-Cre+ rats with ChR2 stimulation at Light-On for which video was recorded; 3 sessions/rat; 3336 no-laser trials in grey; 1335 laser trials in blue). Bottom plots show the breakdown into engaged (n=1975) and non-engaged (n=2696) trials. (b) We examined whether laser-slowed trials might be those in which the rat was waiting at the wrong port (if, for example, DA were to increase the salience of currently attended stimuli). Engaged trials were further broken down into "lucky guesses" (those trials for which the rat was immediately adjacent to the start port as it was illuminated) and "unlucky guesses" (immediately adjacent to one of the other two possible start ports). Blue dashed ellipses indicate zones used to classify trials by guessed port (8.5cm long diameter, 3.4cm short diameter) (c) Laser-slowing was observed for both lucky (n=603) and unlucky (n=1007) guesses. Note that blue

distribution is bimodal in both cases, indicating that only a subset of trials were affected. Video observations suggested that on some trials extra [DA] evokes a small extra head/neck movement, that makes the trajectory to the illuminated port longer and therefore slower. (d) Quantification of trajectories, by scoring rat location on each video frame from 1s before Light-On to 1s after Center-In. Colored lines show all individual trajectories for one example session. Panels at right show the same trajectories plotted as distance remaining from Center-In port, by time elapsed from either Light-On or Center-In. Note that for non-engaged trials (green), the approach to the Center-In port consistently takes ~1-2s. Therefore, the epoch considered as "baseline" in the FSCV analyses (−3 to −1s relative to Center-In) is around the time that rats decide to initiate approach behaviors. (e) Extra [DA] causes longer average trajectories for engaged trials. Cumulative distributions of path-lengths between Light-On and Center-In, for (top-to-bottom) engaged/lucky, engaged/unlucky and non-engaged respectively. Blue lines indicate laser trials, and p-values are from Komolgorov-Smirnov tests comparing laser to no-laser distributions (no-laser/laser trial numbers: top, 292/75; middle, 424/99; bottom, 1897/792). On engaged trials rats often reoriented between the three potential start ports, perhaps checking if they were illuminated; one possibility is that the extra laser-evoked movement on engaged trials reflects dopaminergic facilitation of these orienting movements. If such a movement is already close to execution before Light-On, it may be evoked before the correct start port can be appropriately targeted. (f) Additional trajectory analysis, plotting time courses of rat distance from the illuminated start port. On non-engaged trials extra [DA] tends to make the approach to the illuminated start port occur earlier (note progressive separation of green, blue lines when aligned on Light-On). However, the approach time course is extremely similar (note overlapping lines in the final ~1-2s before Center-In), indicating that extra [DA] did not affect the speed of approach.

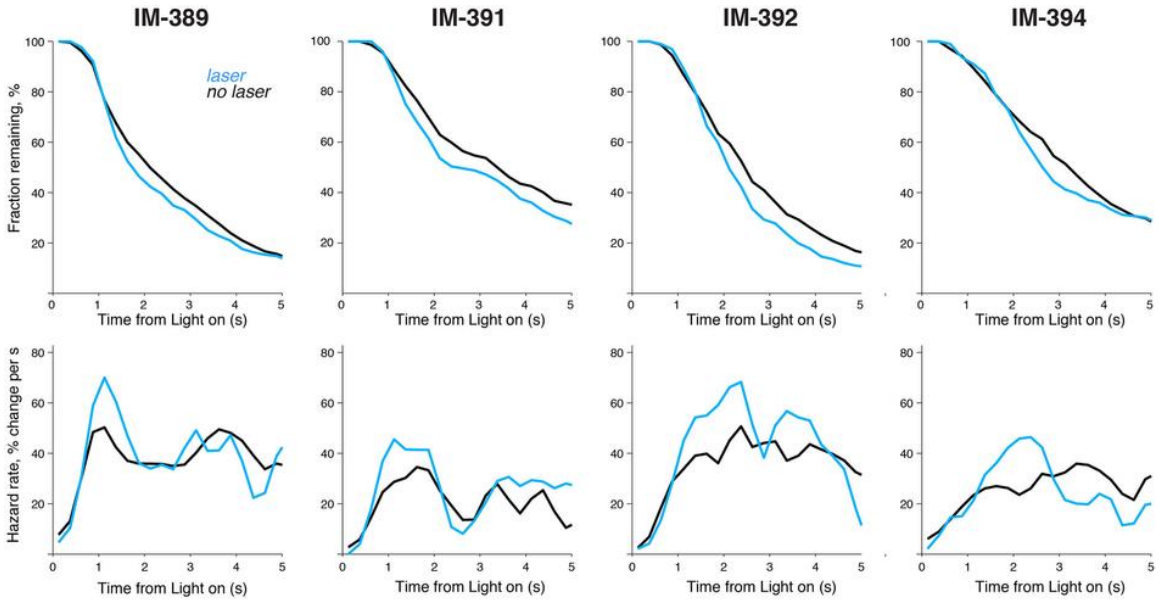


Figure 2.16. Optogenetic effects on hazard rates for individual video-scored rats. Latency survivor plots (top) and corresponding hazard rates (bottom) for each of the four

TH-Cre^{+} rats with ChR2 stimulation at Light-On for which video was recorded (each rat had 3 video sessions that were concatenated for analysis). Only non-engaged trials are included (Numbers of no-laser/laser trials: IM-389, 522/215; IM-391, 294/125; IM-392, 481/191; IM-394, 462/189). For each rat laser stimulation caused an increase in the hazard rate of the Center-In event ~1-2s later (the duration of an approach).

Chapter 3

Dopamine conveys a value signal in cortical and striatal hotspots on a minute-by-minute timescale.

Abstract

We recently demonstrated that dopamine (DA) release in the nucleus accumbens (NAc) signals motivational value, across both fast and slow timescales (Hamid, Pettibone et al. 2015). A critical remaining question is whether this is a global signal that provides similar information within all areas innervated by midbrain DA cells. To investigate this we performed minute-by-minute microdialysis in various subregions of medial frontal cortex and striatum, in rats performing a trial-and-error choice task. We replicated our finding that NAc DA correlates with reward rate, a proxy for the estimated value of work, and found that this value signal is specifically localized to NAc core rather than shell or dorsal-medial striatum. Within frontal cortex we also found a clear relationship between DA and reward rate, but focally within ventral prelimbic cortex rather than more dorsal or ventral subregions. These two "hotspots" of value signaling in rat cortex and striatum show an intriguing correspondence to human brain areas encoding subjective value in fMRI studies (Bartra, McGuire, and Kable 2013). We conclude that, rather than providing a uniform signal in all targets, or providing separate information to cortex and striatum, DA release is tailored to the computational demands of particular cortical-basal ganglia loops.

3.1 Introduction

Drugs of abuse and psychiatric treatments that alter dopamine tone can attenuate or amplify arousal, attention, learning, motivation and risky behavior. Perhaps it is not surprising that there is currently a mottled view of the overall function of DA in the brain. In an attempt to clarify how DA can influence so many aspects of behavior, it has been suggested that parallel streams of information arise from the two distinct modes of dopamine cell firing (Niv et al. 2007). Steady, 'tonic' firing alters extracellular dopamine over a slow time course and may selectively regulate parameters of motivation. Phasic bursting events, on the other hand, are hypothesized to carry a learning signal—specifically, the reward prediction errors (RPEs) of reinforcement learning models (A. A. Grace et al. 2007b).

Much evidence from electrophysiology and voltammetry supports this latter claim. In Pavlovian conditioning paradigms where a cue predicts a reward probabilistically, DA cells scale their response to rewarding events in proportion to the amount of surprise experienced (W Schultz, Dayan, and Montague 1997). Additionally, DA cells can learn to anticipate reward after a fixed delay and will increase or decrease their firing when a reward occurs earlier or later than expected, respectively. This signal has also been observed at DA targets. Sub-second extracellular DA fluctuations in nucleus accumbens correspond to reward predicting cues after repeated pairings but only occur at reward delivery early in training (Day et al. 2007), consistent with a learning/RPE signal.

Much less direct evidence has been gathered about motivation and tonic dopamine, possibly because 'tonic' DA is poorly defined. Motivation is a broad concept and may be difficult to test its many sub-domains. For example, tonic DA has been

proposed to scale overall vigor and arousal. Amphetamine and other drugs of abuse that increase extracellular DA concentration certainly enhance reaction time, motivational 'wanting' (Wyvell and Berridge 2000), and general 'psychomotor activation'. DA may enhance vigor directly by tracking recent reward history, or reward rate (Niv et al. 2007; Guitart-Masip et al. 2011). Reward rate is equivalent to the opportunity cost of sloth, acting to invigorate action.

But 'tonic' DA has also been proposed to provide a directional influence (Beeler et al. 2010; Humphries, Khamassi, and Gurney 2012; Frank et al. 2009). In this elegant account, DA tone can impact *what* is chosen by altering the explore/exploit balance. Here, DA is setting the inverse temperature, or the degree to which differences in value affect choice behavior. This hypothesis is supported by the observation that mice with higher baseline DA (via dopamine transporter knockdown) show a sluggish response to changing reward contingencies, while maintaining similar rates of responding and learning rates compared to their wild-type counterparts (Beeler et al. 2010). In a variant of this hypothesis, it has been suggested that DA regulates 'thrift', or the amount of energy that is expended or conserved in relation to net expected reward (Beeler 2012). This has direct consequences for optimal foraging in that low DA levels should reflect a state of low energy expenditure and promote an exploitative strategy (win-stay) while high DA signals an 'energy rich' state where energy conservation and frugality are of less concern.

In our previous work, we used microdialysis to measure DA on one minute timescales in the nucleus accumbens. Our findings supported the reward rate hypothesis, but did not lend evidence to the ideas that DA modulates a general undirected arousal or shifts the direction of explore/exploit balance. Moreover, this result mirrors our finding

that phasic DA in accumbens signals a moment-by-moment value function, which, like reward rate, estimates expected future reward. While it is striking that accumbens DA correlated with just a single and specific component of motivation, DA cells project to multiple cortical-striatal loops which operate in parallel and facilitate distinct aspects of decision making (Pennartz et al. 2009; Cools 2015). It could be the case that PFC DA is involved in more directional computations, such as setting inverse temperature or adjusting 'win-stay' strategy, while striatal DA is more activational, energizing vigor.

To better understand the contribution of 'tonic' DA more broadly, we repeated our previous dialysis study but sampled from several cortical and striatal targets. First, we replicated our original finding— dopamine concentration in the nucleus accumbens core signals reward rate exclusively. This same signal was found in the ventral portion of the prelimbic prefrontal cortex (pL-PFC), but not in the dorsal portion, the accumbens shell (NACsh), the dorsal medial striatum (DMS), the anterior cingulate (ACC), or the infralimbic cortex (IL). These two 'hotspots' of reward rate coding provide evidence that (i) extracellular DA levels can fluctuate independently, by region, on a minute-by-minute timescale, (ii) the DA signal in one, but not all cortical-stratal loops is consistent and (iii) both cortex and ventral striatum receive a value signal, but may use it for distinct purposes. These results emphasize the importance of appropriate regional specificity in experimental manipulations of DA, and also highlight the implications of psychiatric manipulations that broadly alter DA tone (Kuroki, Meltzer, and Ichikawa 1999).

3.2 Results

3.2.1 Microdialysis during the trial-and-error task

Rats performed a variant of our two-alternative trial-and-error task (Hamid et al., 2015). A single trial in the task proceeded as follows (Figure 3.1a): At Trial Begin the center port was illuminated (Light On) until the rat poked it (Center In), which we defined as 'latency'. The poke was maintained for a brief (500-1500 ms) hold period until a white noise burst played (Go Cue), cueing the rat to withdraw from the center port (Center Out) and poke either of the now illuminated left and right ports (Side In). A food hopper click immediately reported a rewarded trial upon noseport entry and was followed by pellet delivery on the opposite side of the chamber. Unrewarded trials were not signaled other than an absence of the hopper click. Pokes were rewarded stochastically in a block-wise manner, with blocks randomly assigned from the set P(left reward): P(right reward) = [(80:80) (80:20) (20:80) (80:50) (50:80) (50:50) (50:20) (20:50) (20:20)]. Block lengths varied within session from 35-45 trials.

Our dialysis method allows a 1-minute sampling rate of 20 neurochemicals simultaneously. Additionally, we recovered 100 samples in each region, in each session. These features allowed for the unprecedented simultaneous sampling of dopamine targets in cortex and striatum at a time resolution that is physiologically relevant to important variables in our decision making task. For example, rats attempted a range of trials in each minute (0-7 trials) allowing us to gauge their minute-by-minute vigor to perform. Similarly, rats experienced a range of rewards (or omissions) from minute to minute. Our 1 minute time resolution enables us to examine whether changes in extracellular [DA] reflect updates in decision variables that occur rather quickly over a handful of trials or more moderately within a block of trials.

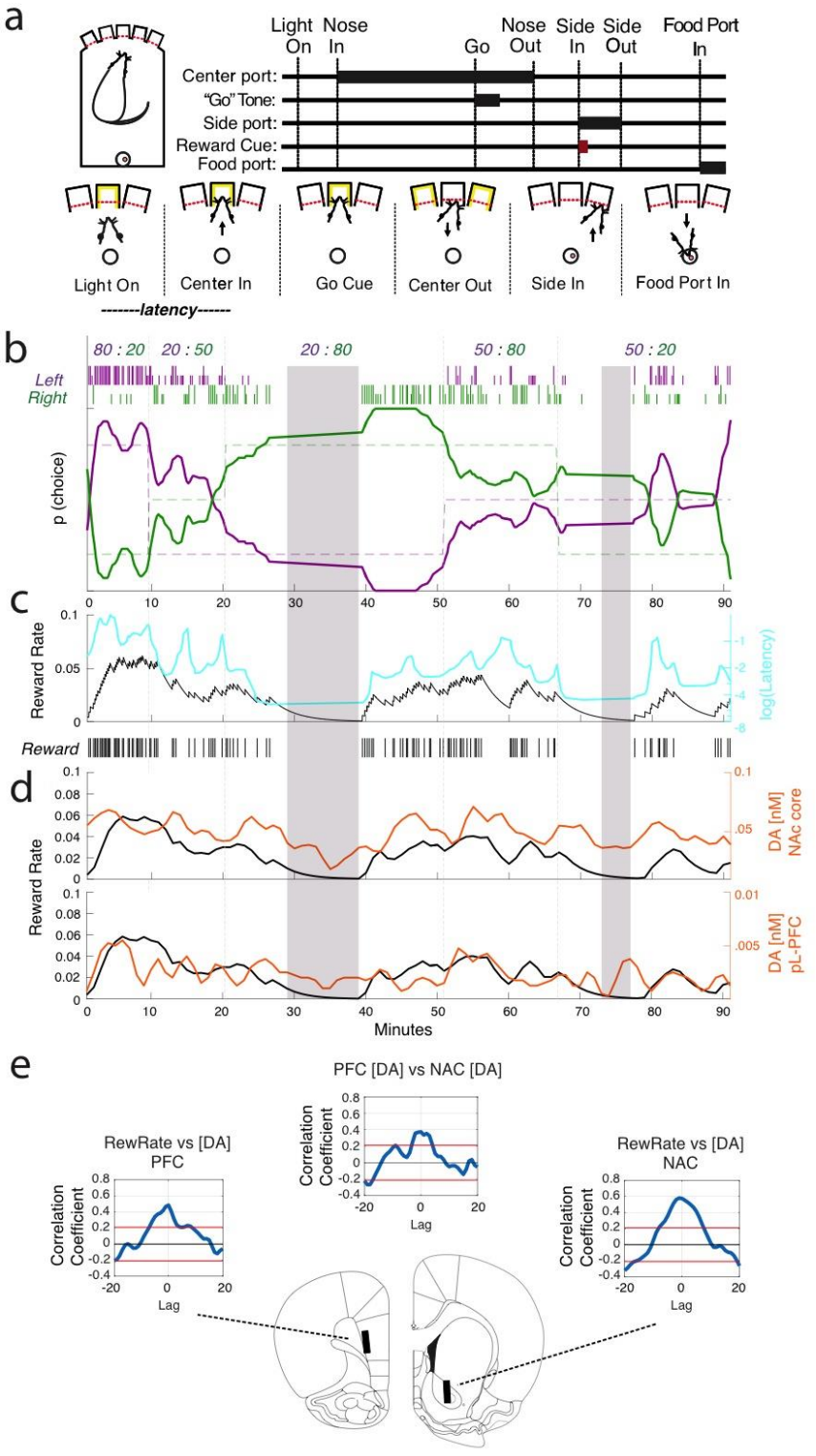


Figure 3.1. Simultaneous microdialysis sampling in cortex and striatum during the trial-and-error task. (a) Schematic diagram of a rewarded trial. (b) Single session of choice

behavior. Reward probabilities are shown at the top of the panel. Each tick mark indicates choice (left = purple, right = green) and outcome (long = rewarded, short = unrewarded). Thick traces show smoothed choice data (seven-trial smoothing) (c) Reward rate vs. latency. Latency data (cyan) is log-transformed and seven-trial smoothed. Reward rate (back) estimate from leaky integrator model shown as a continuous trace. Upward deflections correspond with reward times (black ticks below x-axis). (d) Reward rate vs. simultaneously sampled [DA] in pL-PFC and NAc core. Reward rate (thick black) is averaged across one minute of dialysate collection for comparison to [DA] (orange). (e) Coronal view of probe placements for single session data shown in a-d. Cross-correlograms of reward rate and smoothed [DA] in pL-PFC and NAc core. Middle plot is the cross-correlogram of simultaneously recovered [DA] in pL-PFC and NAc core. Y-axis is Pearson's correlation coefficient and X-axis is the lag in minutes. In each panel, the [DA] signal is referenced to the reward rate signal. Greyed-out areas in b-d represent minutes of inactivity that were excluded from some analyses to rule out the possibility that the strong DA:reward rate correlation was driven by dips in both reward rate and dopamine during these times.

3.2.2 DA encodes reward rate in cortical and striatal 'hotspots'.

The simplest estimate of reward rate in our task would be the number of pellets received per minute. However, this estimate is rather 'forgetful'. Each minute treated as independent from the previous minute and it is unlikely that reward history is represented in this way. A more flexible estimate would allow for a running average with a forgetting parameter that discounts rewards from the distant past. Thus, we employed a leaky integrator model, which applies an exponential decay to each reward beginning at the moment of delivery (Daw, Kakade, and Dayan 2002; Sugrue, Corrado, and Newsome 2004). The value of the decay parameter, 'tau' (range = 1 to 1200s), was selected via a psychometric fit that maximized the negative correlation between trial latency and reward rate, based on our observation that the latency to engage is inversely related to recent reward history (Figure 3.1c). We then averaged the reward rate within each minute of the microdialysis sample for comparison to the DA signal (Figure 3.1d).

We used regression analysis to compare the DA signal to reward rate in each region and found a clear, highly significant relationship between DA and reward rate in

NAc core (R^2= 0.1190 , p= 1.3485e-16, 5/6 rats), replicating our previous finding. This relationship was also observed in an important cortical input to the core— the prelimbic cortex. However, this result was unexpectedly confined to the ventral portion of the pL-PFC (R^2= 0.0953, p= 1.18e-14) with the dorsal portion showing no relationship (R^2= 0.0117, p= 0.0058, 6/7 rats). In a manner consistent with cortical-striatal loops carrying similar signals, we did not see a correlation between DA and reward rate in either NAc shell (R^2= 0.0027, p= 0.2657, 0/8 rats) or its neocortical afferent, the IL (R^2= 0.0307, p= 0.0019, 0/4 rats). Similarly, we found no correspondence between DA and reward rate in our most dorsal targets, the DMS (R^2= 1.23e-05, p = 0.91, 1/10 rats) or the ACC (R^2= 0.012, p= 0.0078, 3/7 rats). To visualize the regional DA:reward rate relationship, we verified the location of the probe membranes and mapped them onto representative atlas sections, color-coded by correlation coefficient (Figure 3.2a). This mapping vividly compliments the findings of our regression analysis.

To get a better sense for the time course of the DA:reward rate relationship and to decouple our finding from a best fit to latency, we next looked at the correlation across all values of Tau (Figure 3.2b). In general, we found that in hotspots DA significantly corresponds with reward rate over a wide, but strikingly similar range of decay values (2-5 minutes). Consistent with our regression analysis, we did not observe this relationship in other regions, at any value of Tau. It is interesting to note that in IL cortex the DA:reward rate relationship is negative across the entire range, however the low number of sessions in the IL prohibits drawing a strong conclusion from this observation.

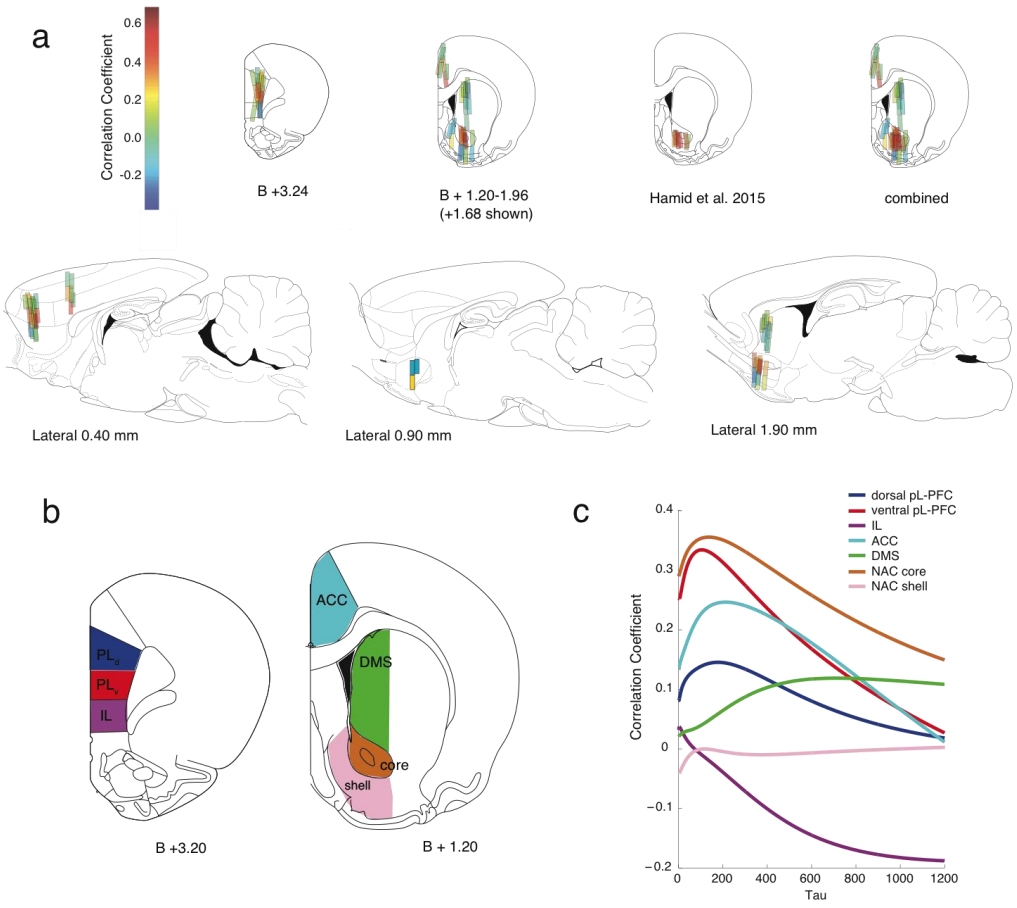


Figure 3.2. DA encodes reward rate in cortical and striatal 'hotspots'. (a) Probe placements for all dialysis sessions, color-coded by DA correlation to reward rate. Chosen atlas sections are representative of average A-P or M-L placements for coronal or sagittal planes, respectively. The two leftmost coronal section panels show results from the current study, and the third coronal panel shows results from our previous study (Hamid, Pettibone et al. 2015). Results from both studies are combined in the rightmost coronal panel. (b) Coronal view of cortical and striatal dialysis probe targets. (c) Summary of DA:reward rate correlation across multiple decay values (Tau), by region. For each session, DA was correlated against leaky integrator reward rate estimates using a range of Tau values (1-1200 seconds). The Pearson correlation at each value was averaged across all sessions in each region.

3.2.3 No relationship with choice

Higher reward rates reflect a higher value for performing the task, and should thus invigorate the decision to work. To test the complimentary hypothesis that DA encodes overall performance vigor, we regressed the DA signal against (i) the number of attempts in each minute and (ii) the averaged latency in each minute. In the model with just attempts we found a similar regional pattern to reward rate with DA significantly correlated to attempts in NAc core (R^2 = 0.073, p= 1.60e-10) and the ventral portion of pL-PFC (R^2 = 0.044, p= 2.27e-07). However, in a model containing both reward rate and attempts as predictor variables, reward rate alone significantly contributed to model fit in both regions (stats). Given that our reward rate estimate was based on a psychometric fit to latency, we expected to see a relationship between latency and DA, at least in regions that showed a relationship between reward rate and latency. This is in contrast to our previous causal manipulation (Hamid, Pettibone et al. 2015), where optogenetically evoked DA immediately shortened latency, specifically on trials where rats were not already engaged (waiting near the noseport to begin the next trial). Our regression analysis did not distinguish between engaged and non-engaged trial types, so by combining them the effect may be too weak to reach significance.

Finally, we sought to uncover whether DA may carry more directional information, especially in targets of the 'dorsal stream' such as DMS and ACC. To test this hypothesis, we regressed DA against several 'choice' related factors. First, we compared DA to the fraction of choices made to contra-versive or ipsi-versive ports, relative to implant side. We found no significant relationship in any region. Next we examined whether DA corresponds with aspects of inverse temperature. In one model, we regressed DA with the fraction of choices made to the port with higher reward

probability in the last 50% of the block. This metric is a simplified framing of 'greedy', 'exploitative' or relatively low-risk choices. In a second model, we regressed DA with the probability of win-stay behavior in each minute, or the likelihood of repeating a choice given that it was rewarded on the previous trial. We did not find a significant fit to DA in any region in either model. For an abbreviated regression analysis summary of several behavioral variables and transmitters, see Figure 3.3.

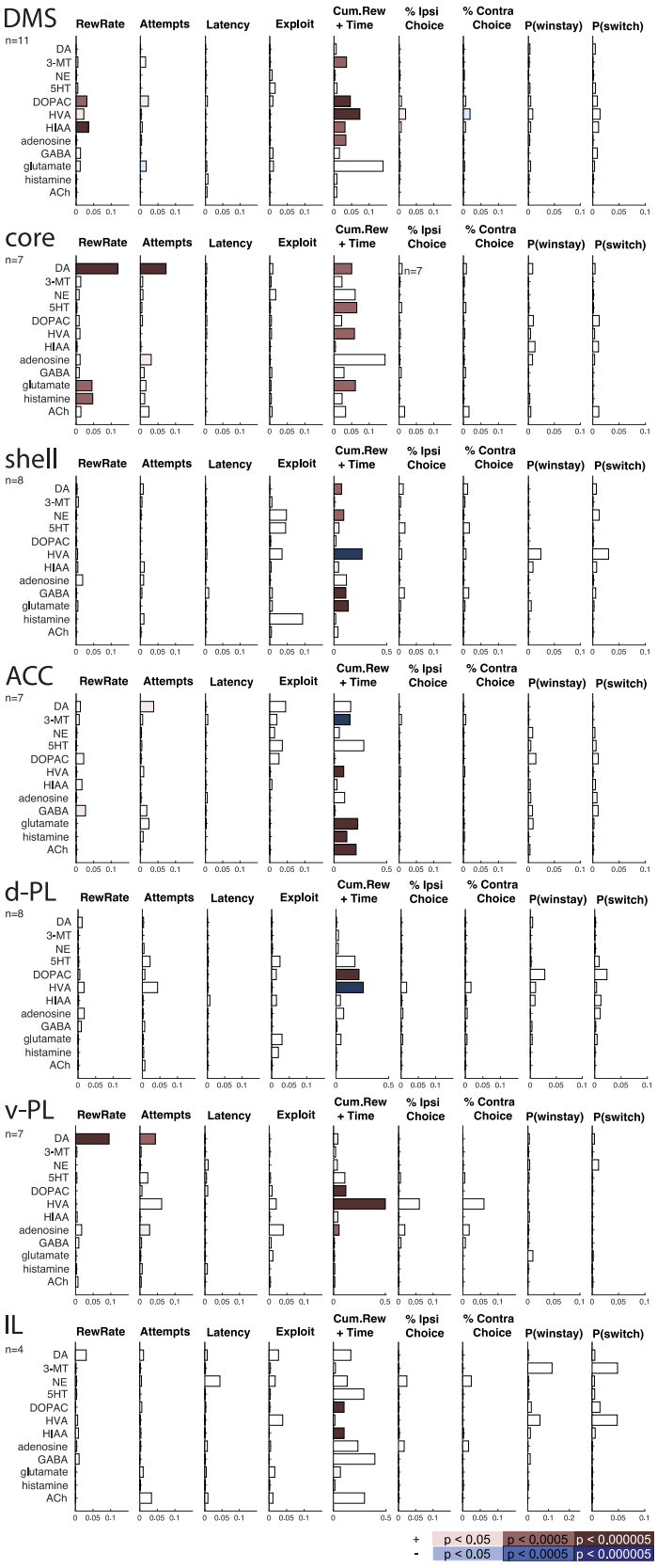

DMS
n=11
core
n=7
shell
n=8
ACC
n=7
d-PL
n=8
v-PL
n=7
IL
n=4
RewRate
Attempts
Latency
Exploit
Cum.Rew + Time
% Ipsi Choice
% Contra Choice
P(winstay)
P(switch)
DA
3-MT
NE
5HT
DOPAC
HVA
HIAA
adenosine
GABA
glutamate
histamine
ACh
+ p < 0.05 p < 0.0005 p < 0.000005
- p < 0.05 p < 0.0005 p < 0.000005

Figure 3.3. Regression analysis. Bars represent R^2 values for linear tests between each analyte (rows) and behavioral measures (columns). Negative relationships are reported in blue and positive relationships are in red. P-values were Bonferroni corrected (20 analytes X 16 measures). To calculate reward rate, we averaged the reward rate in each minute of the dialysis sample. Attempts were calculated as the number of attempted trials (including trials that resulted in an error) in each dialysis minute. Latency was calculated by taking the average of the log10 latency in each minute. Exploitation Index was calculated from the probability of choosing the port with the higher reward probability in the last 50% of the trials in blocks with a superior option (we excluded 20:20, 50:50, and 80:80 blocks from the analysis). Cumulative Reward and Time were included in the same regression model to estimate satiety while accounting for slow timescale increases or decreases in analyte concentration within the session. Cumulative rewards represents the total number of rewards received by the end of the current dialysis minute and Time was simply the number of minutes elapsed since the session began. The bars in this column show color when only the cumulative reward variable was significant. %Ipsi and %Contra represent the fraction of choices to ipsi- or contra-versive ports in each minute, independent of block probability. P(win-stay) was the probability of repeating the previous of choice given the previous choice was rewarded. P(switch) was the probability of not repeating the previous choice, independent of outcome.

3.3 Discussion

3.3.1 Reward rate hotspots and subjective value

Reward rate can be considered an estimate of expected return on work, or temporally discounted value. Estimating value is useful because it can determine whether or not an activity is worth performing, and how vigorously to engage in it. Our present results reaffirm that DA fluctuations signal value in the form of reward rate, but in distinct 'hotspots' rather than ubiquitously throughout the brain. Both of these observations are in sharp contrast to recent electrophysiology recordings of VTA neurons which show a rather uniform and transient RPE response (Eshel et al. 2016). If the terminal [DA] collected by our dialysis probes were a 1:1 reflection of the integrated dopamine neuron activity in (Eshel et al. 2015) we would predict that minute time scale DA would simply be the integration of RPE responses at reward outcome. In this scenario, [DA] would be highest during epochs in the task when rewards are most unexpected/rare and lowest when rewards are regularly occurring and unsurprising. On

the contrary, we observed that [DA] was highest during epochs of regular reward receipt and lowest during reward omission, similar to an online estimate of value.

The 'hotspots' of value coding correspond remarkably well with human fMRI data which reports increased BOLD activity exclusively in vmPFC, ventral striatum, and posterior cingulate cortex during tasks that engage subjective value (Kable and Glimcher 2007; Levy and Glimcher 2012; Bartra, McGuire, and Kable 2013) (see Figure 3.4). Our results suggest that DA could be the principal driving force behind this non-specific signal.

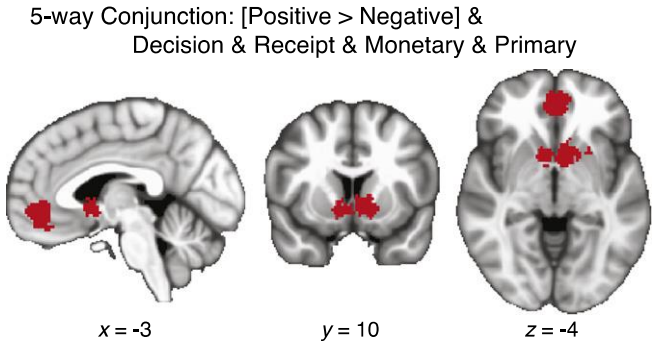


Figure 3.4. Subjective value 'hotspots' in fMRI BOLD signal. Summary of a meta-analysis of subjective value. Red voxels in ventral striatum and vmPFC indicate there was more activity for positive effects compared to negative effects, and above-chance activity during choice and outcome epochs for monetary and primary rewards. Adapted from (Bartra, McGuire, and Kable 2013).

3.3.2 DA signals value in NAC core, but not shell.

We replicated our previous finding that DA in NAc core corresponds with reward rate. However, DA in the shell did not show this relationship. This is an intriguing result, which adds to the accumulating evidence of functionally distinct dopaminergic activity in core and shell related to reward, motivation, and drugs of abuse (S. M. Reynolds and

Berridge 2003; Vander Weele et al. 2014; Saddoris et al. 2015; Ko and Wanat 2016). NAc DA is critical for reward seeking behavior (K. C. Berridge and Robinson 1998). Broad DA depletion in the NAc causes rats to switch from effortful work to less effortful options when foraging (J D Salamone et al. 2007) and decreases responding on tasks with higher fixed ratios (J D Salamone et al. 2001). Ko & Wanat (2016) showed DA in core but not shell was modulated with effortful engagement but not movements unrelated to reward.

3.3.3 A dorsal-ventral divide in prelimbic dopamine function.

The finding that dopamine encodes rate of reward in prelimbic cortex is not new. (St Onge et al. 2012) have shown that DA fluctuations (on 7 minute timescales) reflect reward rate in both instrumental responding and passive reward delivery. Beyond the difference in temporal resolution, one important caveat of this study is the spatial resolution. The membrane length of dialysis probes used in St. Onge et al. were 2mm and spanned the entire prelimbic cortex. Here we report that the special relationship between DA and reward rate is localized to the ventral pL cortex.

Our finding is consistent with important differences in the connectivity and functional roles of these subregions (Heidbreder and Groenewegen 2003). A dorsal-ventral gradient of dopamine neurons in the VTA project to a ventral-dorsal gradient of medial prefrontal cortex (Deutch 1993). Additionally, ventral areas of mPFC provide stronger reciprocal connections to DA cells than their dorsal counterparts (Sesack et al. 1989). The pattern of prefrontal cortex efferent connections also supports a picture of distinct function. Dorsal pL-PFC and ACC send projections to sensorimotor cortex, DMS, and the more lateral NAc core and shell, while ventral pL-PFC and IL-PFC project

to medial NAc shell and core, hippocampus, amygdala, and pyriform cortex (Heidbreder and Groenewegen 2003). One further interesting distinction, with perhaps important functional implications, is that the ventral area of pL cortex more densely projects to the mu-opioid receptor rich 'patch' subcompartments of the dorsal striatum and NAc core (Gerfen 1989). These unique islands project directly to midbrain DA neurons and have been hypothesized to contribute to reward valuation via RPEs (Houk, Joel L. Davis, and Beiser 1995) or, more recently cost/benefit calculations (Friedman et al. 2015). Together, these findings reveal a special role for the ventral pL-PFC in strongly influencing DA cell population activity, directly and indirectly.

Lesions of the ventral prelimbic/infralimbic region impair reversal learning in rats (L. Li and Shao 1998) and 6-OHDA lesions in pL-PFC specifically impair the ability to detect changes in reward contingencies while sparing habitual responding (Naneix et al. 2009). Similarly, focal damage to human vmPFC abolishes reversal learning in both deterministic (Fellows and Farah 2003) and probabilistic (Hornak et al. 2004) task variants. If these regions are indeed involved in choice flexibility, the DA value signal in the prelimbic hotspot may be used to inform the 'stakes' of the current decision. Specifically, in a simple decision of whether to repeat the previous choice or switch to something new, the dopamine reward rate signal may be useful in influencing the probability of repeating the previous action. By comparing information about what is to be lost by switching or, conversely, what is to be gained by repeating to the reward rate, a more accurate risk assesment can be made. In the context of our task we did not see a relationship between DA and explore/exploit behavior, however the task design may be simple enough to not engage the more overt computational processes normally handled by the medial PFC.

3.3.4 Conclusion

Given this large-scale survey of so many regions and transmitter species that facilitate adaptive decision making, it is surprising that more relationships between other transmitters and behavioral variables were not observed. Nevertheless, our primary finding that DA encodes value in cortical and striatal hotspots confirms and extends our previous work. Dopamine uniquely provides a value signal during decision-making and, importantly, this signal is not ubiquitous but rather selective to subregions. These results underline the importance of regional specificity in studies that rely on dopaminergic measurements and manipulations. Future studies of dopamine cells and their targets will be necessary to clarify whether the heterogeneity of the observed dopamine signal arises in the midbrain or is achieved locally by terminal micro-circuitry.

3.4 Methods

3.4.1 Animals

All animal procedures were approved by the University of Michigan Committee on Use and Care of Animals. 10 male Long-Evans rats were bred in-house and maintained on a reverse light-dark cycle (12hr:12hr). During training and experiments, rats were food restricted to 15 g of rat chow daily and given free access to water. Animals were trained 3-6 months in computer-controlled five-hole nose-poke operant chambers (Med Associates).

3.4.2 Trial-and-error task

Autoshaping proceeded in the following sequence: (*Poke Any*, 2 days) Rats were instructed to poke any of five illuminated nose ports, with no limit on hold duration. Pokes were rewarded with a 45mg fruit punch flavored pellet (TestDiet #). (*Poke One*, 1-2 weeks) Rats were instructed to poke and hold their nose in single illuminated port for a variable hold period (500-1500 ms). Poking an unlit port resulted in a wrong start (WS) error and a houselight timeout period (3-5 s). Poking and removing before the termination of the hold period resulted in a false start (FS) error and a houselight timeout period. Training was advanced after rats performed >80% non-error trials for two sessions. *(Probability 50-50*, ~2 weeks) After poking a single lit port and holding for a variable delay (500-1500 ms), a white noise burst (Go Cue) instructed the animal to poke one of two adjacent illuminated ports. Both left and right ports were rewarded randomly at 50% probability. On rewarded trials, food hopper activation produced an audible click ('Reward Cue'). Unrewarded trials were not cued other than a lack of this click and omission of pellet delivery. To encourage responses to both left and right ports, a bias correction was implemented that allowed only three consecutive rewards to either side. The probability of reward to a favored side decreased to 0% until the opposite side was rewarded. Failure to choose either adjacent port within 1 second after the Go Cue resulted in a failure to respond (FR) error and a houselight timeout period. Training was advanced after rats performed >70% non-error trials for two sessions. (*Probability Blocks*, 1-2 months) The trial structure of this final stage is the same as the previous, with the exception of additional left and right reward probability blocks: [(80:80) (80:20) (20:80) (80:50) (50:80) (50:50) (50:20) (20:50) (20:20)]. Block probability order was randomly generated, as was the length of

each block (35-45 trials) within the session. Bias correction was not used in this stage. Rats were ready for implantation after performing >80% non-error trials for two weeks of daily training.

3.4.3 Stereotaxic Surgery

Rats were implanted bilaterally with guide cannula (Part #) in cortex and striatum in two groups. The first group (n=6) received a guide cannula targeting pre limbic and infra limbic cortex at AP +3.2, ML 0.6, DV 1.4 from brain and a guide cannula targeting dorsomedial striatum and Nucleus Accumbens core/shell in the opposite hemisphere at AP +1.3, ML 1.9, DV 3.4. Both implants were angled 5 degrees away from each other along the rostral-caudal plane. A second group received a guide cannula targeting anterior cingulate cortex at AP +1.6, ML 0.8, DV 0.8 and a guide cannula targeting and Nucleus Accumbens core/shell in the opposite hemisphere at AP +1.6, ML 1.4, DV 5.5 (n=2) or AP +1.6, ML 1.9, DV 5.7 (n=2). Implant sides were distributed equally. Animals were allowed to recover 4-7 days prior to retraining.

3.4.4 Microdialysis

Sample Collection. On testing day, animals were placed in the operant chamber with the houselight on. Custom-made concentric polyacrylonitrile membrane microdialysis probes (1 mm dialyzing AN69 membrane; Hospal, Bologna, Italy) were inserted bilaterally into guide cannula and perfused continuously (Chemyx Inc., Fusion 400) with aCSF at 2 μL/min for 90 minutes to allow equilibration. After 5 minutes of baseline collection, the houselight was extinguished, cueing the animal to task availability. Sample collection continued at 1-minute intervals and samples were

immediately derivatized with 1.5 µL sodium carbonate, 100 mM; 1.5 µL BzCl, 2% (v/v) BzCl in acetonitrile; and 1.5 µL isotopically labeled internal standard mixture diluted in 50% (v/v) acetonitrile containing 1% (v/v) sulfuric acid, and spiked with deuterated ACh and Choline (C/D/N isotopes, Pointe-Claire, Canada) to a final concentration of 20 nM. Derivatized samples were analyzed using Thermo Fisher Accela UHPLC system or Thermo Fisher Vanquish UHPLC interfaced to a Thermo Fisher TSQ Quantum Ultra triple quadrupole mass spectrometer fitted with a HESI II ESI probe, operating in multiple reaction monitoring. Five µL samples were injected onto a Phenomenex core-shell biphenyl Kinetex HPLC column (2.1 mm x 100 mm). Mobile phase A was 10 mM ammonium formate with 0.15% formic acid, and mobile phase B was acetonitrile. The mobile phase was delivered an elution gradient at 450 µL/min as follows: initial, 0% B; 0.01 min, 19% B; 1 min, 26% B; 1.5 min, 75% B; 2.5 min, 100% B; 3 min, 100% B; 3.1 min, 5% B; and 3.5 min, 5% B. Thermo Xcalibur QuanBrowser (Thermo Fisher Scientific) was used to automatically process and integrate peaks. Each peak was visually inspected to ensure proper integration…. all 117,000 of them.

Chemicals. Sodium carbonate, benzoyl chloride (BzCl), sulfuric acid, and salts for the small molecule neurochemical analysis were purchased from Sigma Aldrich (St. Louis, MO). Water, methanol, and acetonitrile for mobile phases are Burdick & Jackson HPLC grade purchased from VWR (Radnor, PA). All other chemicals were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Artificial cerebral spinal fluid (aCSF) was comprised of 145 mM NaCl, 2.68 mM KCl, 1.40 mM CaCl-$_2$, 1.01 mM $MgSO_4$, 1.55 mM Na_2HPO_4, and 0.45 mM NaH_2PO_4, adjusted pH to 7.4 with NaOH. To

prevent oxidation of analytes, the aliquot of aCSF that flowed through the lines was spiked with ascorbic acid to a final concentration of 250 nM.

3.4.5 Statistical Analysis

All neurochemical concentration data were smoothed with a 3-point moving average (y' = [0.25*(y-1) + 0.5(y) + 0.25*(y+1)]) and z-score normalized to facilitate between-session comparisons. Only data points from within the session were included in the analysis. Models were generated in a step-wise format using the *regress* function in MATLAB. R values in scatterplots represent Spearman's correlation. Cross-correlograms were generated using the *crosscorr* function in MATLAB. Error bars were generated for each subplot by shuffling one time series 100,000 times and generating a distribution of correlation coefficients for each session. The 5,000th and 95,000th values for each session were averaged across sessions to calculate p = 0.05 error bars and the 1,000th and 99,000th values were similarly averaged to calculate p = 0.01 error bars.

3.4.6 Histology

Within two days of the final sample collection, rats were deeply anesthetized with isoflurane and intercardially perfused with Lactated Ringer's solution followed by 4% PFA. Brains were stored in 30% sucrose at 40 degrees C for ~24 hours and sectioned coronally at 50 um on a cryostat. Sections with stained with cresyl violet (0.5% w/v) and imaged under normal light microscopy. Images were warped to correspond with equivalent atlas sections (Paxinos & Watson, 2005) using SqrlzMorph morphing software to account for any tissue shrinkage and distortion during processing.

Chapter 4

Decision-making and valuation in dorsal striatal microcircuitry

Abstract

The dorsal striatum is fundamentally involved in action selection. Single-unit recordings have shown that the projection cells of the dorsal striatum flexibly encode the action values of reinforcement learning (RL). However, exactly how action selection is achieved by the striatum remains poorly understood. Here, we explore the dorsal striatal micro-circuitry that is implicated in biasing action selection. Specifically, we investigate electrophysiological differences in patch and matrix compartments, which have been hypothesized to encode reward prediction error (RPE) and action value, respectively. Our recent finding that mesolimbic dopamine encodes a temporal-difference value function offers a promising alternative hypothesis: patch cells encode an overall state value and matrix cells encodes specific action values.

4.1 Introduction

Key features of striatal learning are remarkably similar to reinforcement learning models of adaptive decision-making (Daw et al. 2005). In a simple Actor-Critic RL model, action selection is achieved by dividing responsibilities between two agents. The 'actor' selects among candidate actions by arbitrating between a set of *action values* and experiences an outcome. A separate 'critic' then evaluates how different that outcome is

from what was expected, based on the value of the action taken or the *chosen action value*, and generates a prediction error, or RPE. The RPE is used to update a value function, or the *state value*, that is used to estimate the outcomes of future actions. Positive RPEs increase the likelihood of selection the same action in the future, while negative RPEs decrease the likelihood (Sutton and Barto 1998). The actor-critic model, and a popular alternative, the *Q-learning* model, are useful for action selection processes that require little computational processing because only action values/weights are stored, similar to the way that instrumental associations can be stored in cortical-striatal synapses. The prediction error, provided during outcome epochs by the strong dopaminergic input to the striatum, is transient and serves no purpose after action and state values are updated.

The striatum is composed of two anatomically and neurochemically distinct sub-compartments. These compartments readily appear when labeled for mu-opioid receptors. The patch, or striosome, is rich in mu-opioid receptors and receives inputs from limbic and prelimbic neocortical structures (Eblen and Graybiel 1995). The matrix, void of mu-opioid receptors, is receives inputs from sensorimotor cortex and projects onto the direct and indirect pathways of the basal ganglia. Patch neurons project onto midbrain dopamine neurons, which project back onto striatal MSNs (Matsuda et al. 2009; Fujiyama et al. 2011). The observation of patch cells exerting an influence on dopamine cells has led to the hypothesis that patches correspond to the 'critic' in the RL framework, generating RPEs, while matrix cells correspond to the 'actor', selecting actions based on an arbitration between stored action values (Houk, Joel L. Davis, and Beiser 1995).

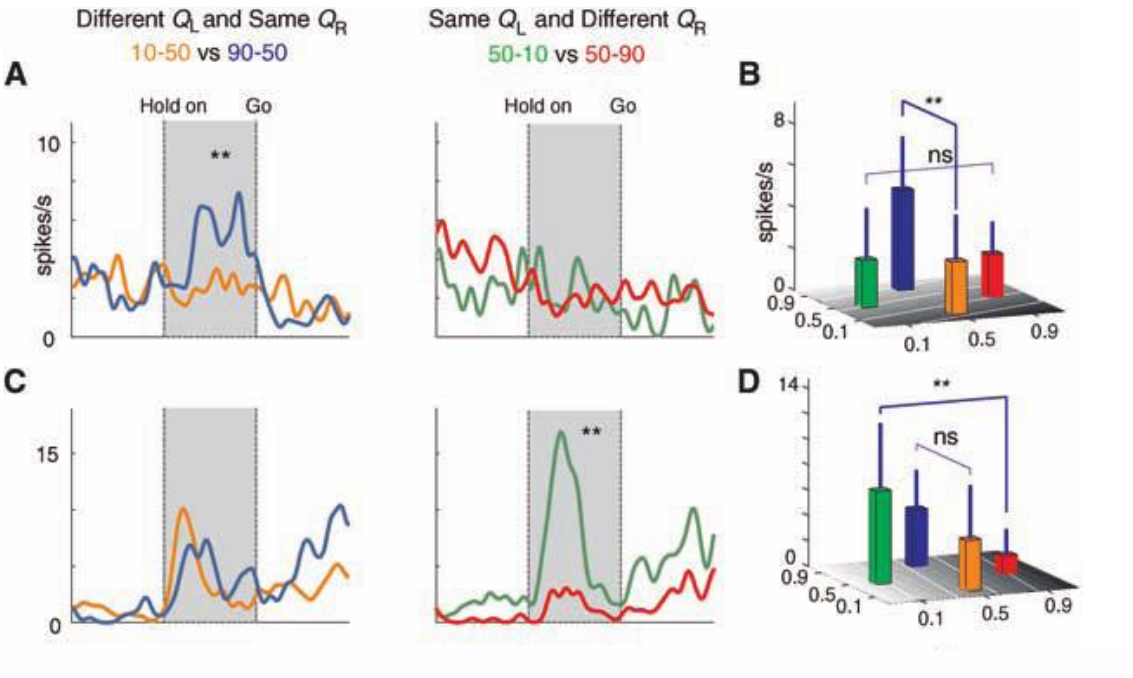


Figure 4.1. Two representative action-value coding neurons in dorsal striatum. Single units showing selectivity for left action values (A,B) and right action values (C,D). Traces and 3D boxplots show averaged firing rates for different block probabilities. The left action value unit is more active during the hold period when the probability of reward for left choices are 90% compared to 10%, independent of the probability of reward for choosing right. The right action value coding unit show increased firing during lower right reward probabilities, independent of left reward probability. Adapted from *Samejima, Ueda, and Doya 2005.*

There are several key findings that support this dissociation. Single-unit recordings in monkey have shown a large population of dorsal striatal cells that scale their firing with action values specific to a left or right choice ('*Q values*' in the Q-learning nomenclature) (Samejima, Ueda, and Doya 2005). Specifically, in a stochastically rewarded two-alternative choice task, these cells show activity prior to action selection and scale their firing with the learned value of that choice while showing no relationship to the learned value of an opponent choice (Figure 4.1). Additional work has since found striatal representations of *chosen action value*, with cells responding during trial outcome in a manner that scales with expected return on the selected action

(H. Kim et al. 2009; Strait, Sleezer, and Hayden 2015).

There is also accumulating evidence that patches contribute a unique evaluative signal. First, animals self-stimulate when stimulating electrodes are placed in patch, but not in matrix (White & Hiroi, 1998). Second, neurotoxic ablations of patch compartments cause deficits in motor learning tasks, but not motor performance (Lawhorn, Smith, and Brown 2009). More recently, putative patch cells have been implicated in cost/benefit evaluation in a task with appetitive and unpleasant stimuli (Friedman et al. 2015). Finally, RPE-like coding has been observed in dorsal striatal neurons during a simple 5 tone classical conditioning experiment (Oyama et al. 2010). Here, animals learned that each tone predicted pellet delivery at a unique, fixed probability. In roughly 9% of recorded cells, the firing rate at tone and reward epochs scaled with the amount reward expectation and surprise, respectively (Figure 4.2). The proportion of cells displaying this activity is congruent with the proportion of dorsal striatum composed of patch.

An alternative hypothesis emerges, however, when considering the findings presented in Chapter 2. We demonstrated that dopamine in the ventral striatum corresponds with a moment-by-moment value function, which is a TDRL *state value*. While this is in contrast to the simple RPE-coding function originally observed, we suggest that an important distinction between our study and previous work may be the task demands. In simple conditioning tasks where reward is not contingent on an instrumental response state value coding may look just like RPE coding. However, in our task, the animal must orchestrate a sequence of responses to progress through several *states* in a single trial, with the final outcome being contingent on successful orchestration of the sequence. Similarly, the RPE-coding neurons found in (Oyama et al. 2010) may actually be coding a state value in a task with just two states. Additionally,

representations of state value have been found in single-unit recordings in vmPFC and ventral striatum (Strait, Sleezer, and Hayden 2015; Lau and Glimcher 2005; H. Kim et al. 2009)—limbic structures which, like patch cells, exert a reciprocal influence on the dopamine system. Collectively, these observations implicate a special role for patches in the online evaluation of performance. In a computational framework, this may manifest as the encoding of prediction errors, or the state value of an actor-critic model.

To examine whether patches participate in signaling a distinct value from matrix, we have been recording from neurons in the dorsal striatum during a probabilistically rewarded two-armed bandit task. Recording specifically from patches has been a longstanding technical challenge. Patches take up a small total volume of striatum and their location is unique from animal to animal. Thus it is difficult to selectively target them. Because of their non-uniform dimensions, it is also difficult distinguish which compartment an electrode was located in with reasonable spatial resolution. The standard practice of electrode localization is to create an electrolytic lesion at the electrode tip. The damage left by electrodes is so small that is often impossible to detect where the electrodes were without these lesions. However, this practice destroys local tissue and obscures immunohistochemical (IHC) labeling of mu-opioid receptors.

To overcome these challenges, we have made a series of advancements in our approach, both in electrode type and histology methods, to recording from patches. Each refinement has pushed us closer toward the goals of clearly defined localization and a high yield of patch recordings.

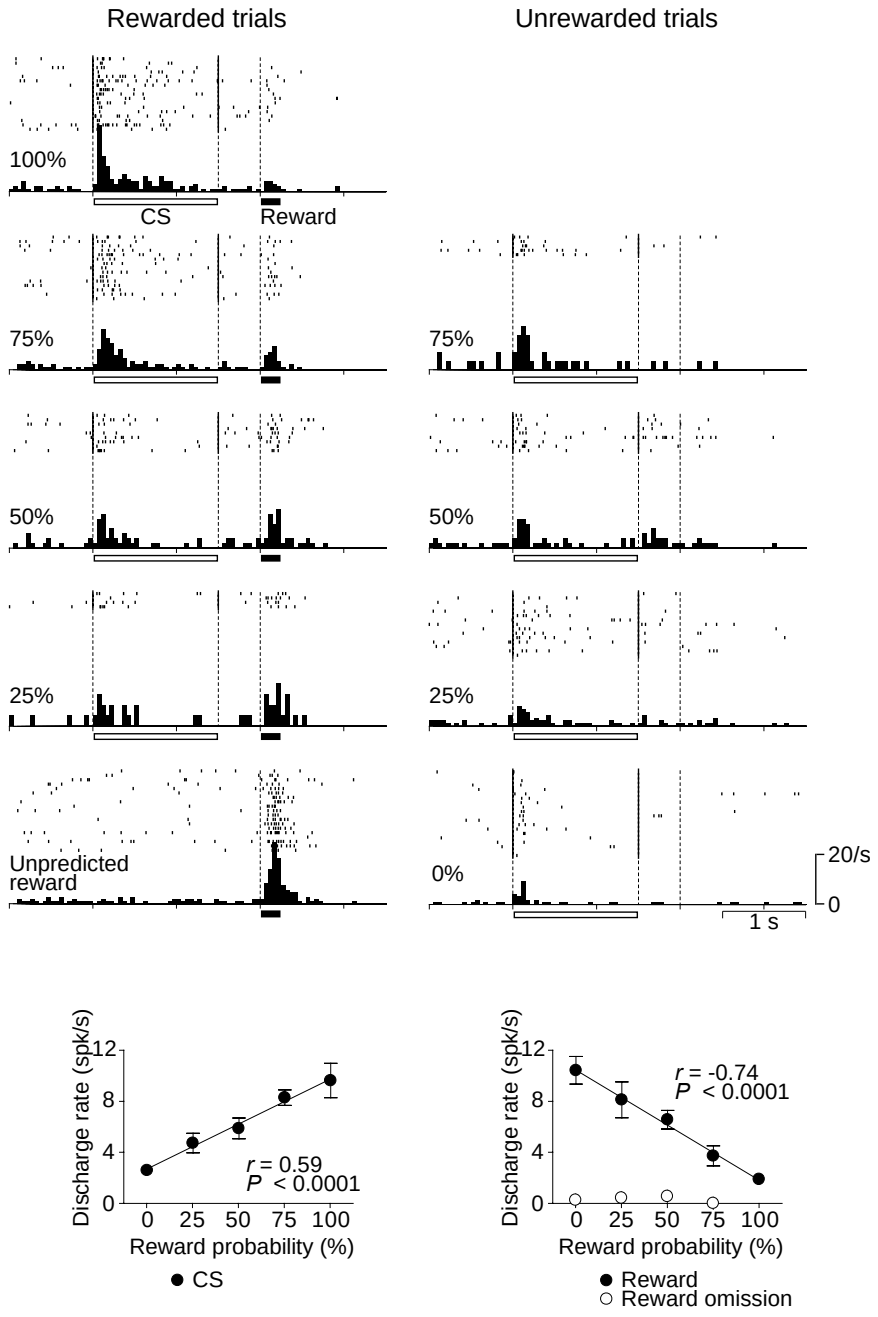


Figure 4.2. RPE coding in dorsal striatum. (Top) Rasters and histograms of spiking activity from a single unit at each of five conditioned tones (rows) aligned to CS onset. The tone that predicted reward with 100% reliability (the least surprising stimulus) elicited a strong response during CS and a weak response during reward, while tones that were weakly predictive elicited a relatively stronger response. (Bottom left) Spiking increases with reward probability during CS and decrease with reward probability during outcome (bottom right), consistent with an RPE signal. Adapted from *Oyama et al. 2010*.

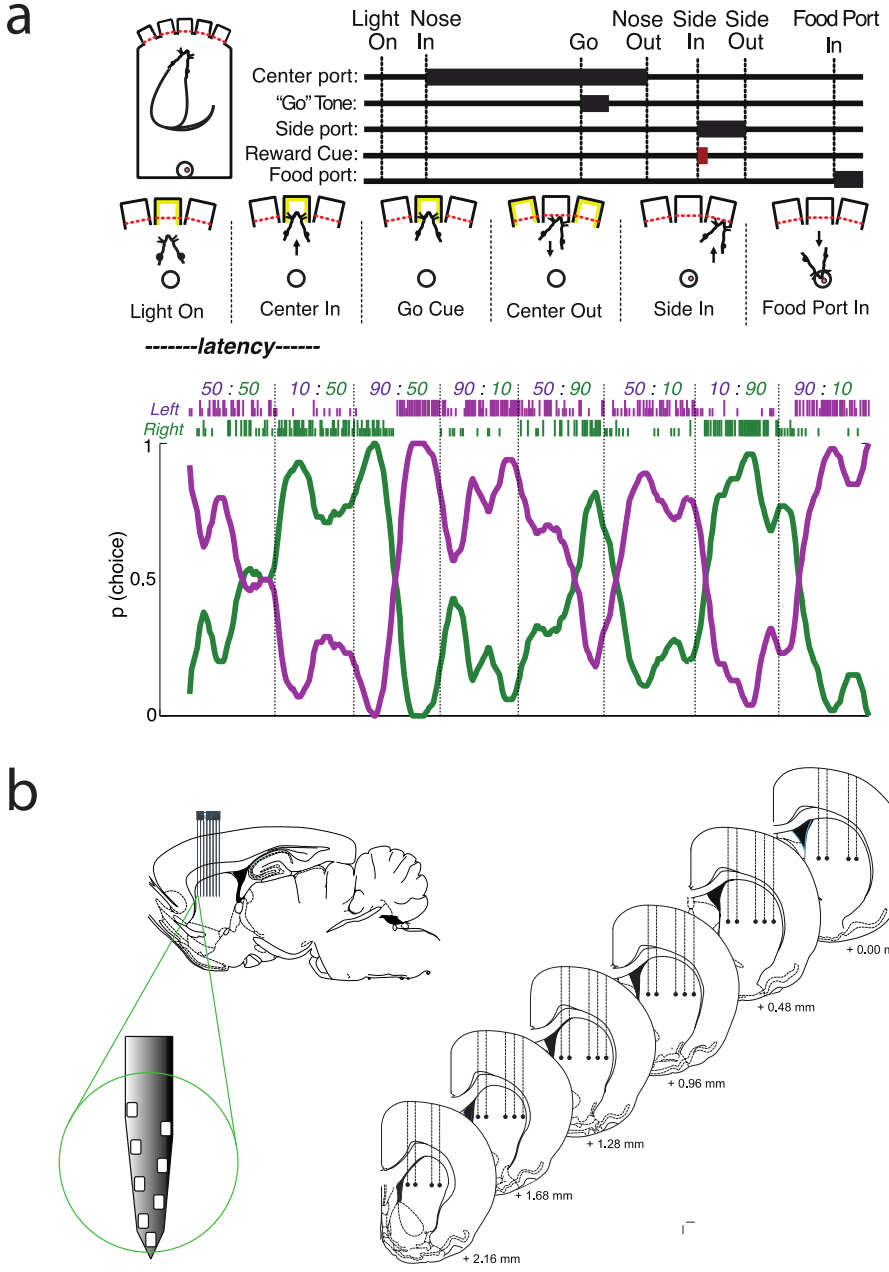

a
Light On
Nose In
Go
Nose Out
Side In
Side Out
Food Port In
Center port:
"Go" Tone:
Side port:
Reward Cue:
Food port:
Light On
Center In
Go Cue
Center Out
Side In
Food Port In
-------latency------
50 : 50
10 : 50
90 : 50
90 : 10
50 : 90
50 : 10
10 : 90
90 : 10
Left
Right
p (choice)
1
0.5
0
b
+ 0.00 mm
+ 0.48 mm
+ 0.96 mm
+ 1.28 mm
+ 1.68 mm
+ 2.16 mm
c
d
Tetrode
60 µV
n= 1100
50 µm
140 µm

Figure 4.3. Summary of behavioral task and electrode placements. (a) (Top) Schematic diagram of trial-and-error task, showing sample of behavior from a single session (bottom). (b) (Left) 64-channel silicon probe array in sagittal section, inset shows tip detail. (Right) Location of electrode placements for 28-tetrode drives. (c) Rat implanted with a chronic tetrode drive in our 5-poke operant chamber. (d) Detail of tetrode construction and use (adapted from Buzsáki 2004). Tetrodes were constructed by twisting four strands of Ni-Chrome wire together.

4.2 Results

4.2.1 Summary of localization techniques and effectiveness

We began our recordings using 64-channel silicon probes. This probe design offered 8 recording sites at the tip of each of 8 shanks (Figure 4.3b). Prior to insertion, probes were coated in the fluorescent lipophilic tracer, DiO, which is taken up by the cell membranes that it contacts, leaving a history of probe location. In the event that a shank entered a patch, the tip of the shank would be easily co-localized with mu-opioid staining. However, while the 4 recordings from these devices resulted in 66 units, we did not find a single well-placed shank in a patch. Two factors contributed to poor tip localization. First, tips were rather wide and often appeared to span multiple histology sections. Second, the DiO signal was absorbed by neighboring tissue in a non-uniform manner, resulting in clearly marked regions and weakly marked regions. This decreased confidence in true electrode location.

To ameliorate the issues of poor localization and low yield, we implanted drivable 28-tetrode arrays (Figure 4.3b-c). Tetrodes are smaller in cross-section than the previously described silicon probes, and the recording sites are at the exposed tip (Figure 4.3d). We reasoned that these finer features would decrease the amount of excessive tissue staining from DiO, decrease the likelihood that the tip would span multiple histology sections, and offer higher resolution when a tip was near the boundary of a

patch. Additionally, we used microglia marker CD11b during IHC to provide another dimension to our localization. Microglia are part of the immune response to tissue damage and provide an excellent visual history of the cell death and inflammation incurred as the electrodes pass through tissue. We used a wide spacing between tetrodes to broadly survey the striatum and increase the chances of hitting a patch. Across 5 tetrode sessions, we recorded a total of 329 units. Of all 84 tetrodes, 2 tetrode tips were localized within a patch (entirely surrounded by patch), with a further 6 on the boundary (within 10-15 um of a patch).

While this was an improvement, it was clear that this approach would not be tenable for generating enough data to test our hypotheses about patch function. To overcome this we have again refined our approach to enhance tip localization and improve yield. We have designed a new generation of silicon probes that have a smaller feature size than the previous generation (much closer to tetrodes). The probes are comb-shaped with 32 shanks and 2 sites per shank, again increasing the chances of hitting a patch. We continue to use the microglia cell marker CD11b during our IHC protocol, because it proved reliable in resolving tetrode locations. Next, we leave the electrodes in the brain during sectioning to keep them in the tissue until imaging. Finally, we are slicing substantially thicker than usual sections (300-500 um) to increase the number of electrode tips in the same section and to minimize movement of the electrodes.

Each of these techniques would likely be insufficient on their own to guarantee confidence in tip localization. When combined, they offer enough information to reasonably resolve patch vs. matrix placement. We are actively recording with these probes and applying the updated histology techniques. Preliminary histology results are promising (Figure 4.4).

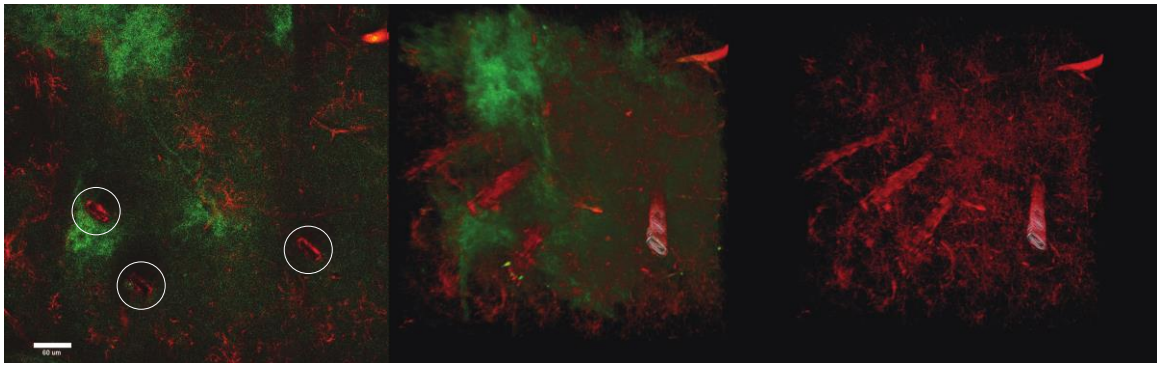


Figure 4.4. New histology method for superior electrode tip localization. Left panel shows overlay of mu-opioid receptors (green) CD11b microglia marker (red) and silicon probe shank locations (white circles). Probe locations are clearly encapsulated by microglia. Middle and right panels display the 3-D reconstruction of a z-stack through the entire 500 um section at a slightly offset angle. Right panel shows CD11b staining only. Scale bar is 60 um.

4.2.2 Summary of electrophysiology recordings

We recorded single-unit activity in dorsal striatum using 64-channel silicon arrays and drivable 28-tetrode arrays. Across 9 sessions, we recorded a total of 395 well-isolated units, with silicon probes (4 sessions) yielding 66 units and 28-tetrode drives (5 sessions) yielding 329 units. Units were further culled by examining whether each unit was active (> 5 Hz firing rate) during at least one task epoch, yielding a total of 145 task-active units (44% of all units).

We examined time bins around the Side-In event because different types of value coding should show distinct responses for choice and outcome (Figure 4.5). In the two example patch units shown, it is clear that firing rate is responsive to trial outcome, with peak firing achieved within 1 second after Side In. The response is inconsistent, however, with one unit preferentially firing for omissions and the other firing for rewarded

outcomes. The reward-preferring unit shows a delayed/shifted response for right rewards compared to left rewards. Interestingly, this unit is nearly silent for an extended period preceding the Side In on all trial types, but remains active for several seconds after all outcomes. There is a similar lateralization in preference in the unit activated by omissions: While the peak firing occurs at nearly the same time for left and right choices, the firing rate is higher for right choices. Of note, there is a slight ramp in firing rate for rewarded outcomes, but it diminishes rather quickly (within a second) compared to the sustained ~2 seconds of activity on omission trials.

Similarly, matrix identified units showed a heterogeneous response to the Side In event. One of the cells showed activity both before and after the Side In event, with higher activity on rewarded trials at outcome. This unit did not display a side preference during choice execution, but a preference emerged ~1 second after the outcome. The other matrix neuron showed no activity until after the outcome when firing rate increased for all trial types. This unit displayed no side preference, but showed a sustained activity for several seconds after omissions.

For a unit to be reasonably labeled 'action value coding' it should show activity in the period before the action is completed, and show a preference for direction. I observed pre-Side In activity in one of the matrix cells shown, but no preference for direction was observed during this epoch. Rather, this unit showed a higher firing rate for rewarded trials than omission trials, independent of choice. Interestingly, there was a second peak in firing for left choices in the post-Side In period, independent of outcome. Together, these observations suggest the cell is sensitive to differences in choice *after the outcome*, consistent with an action-value updating process.

State value, on the other hand, is indifferent to choice and is only concerned with outcome. Thus, choice-preference activity before outcome would rule out state-value coding. While the patch units presented here showed some choice preference, there was a larger distinction in firing during the outcome epoch for both examples. Both units showed activity after choice, similar to an Actor-Critic state value update event.

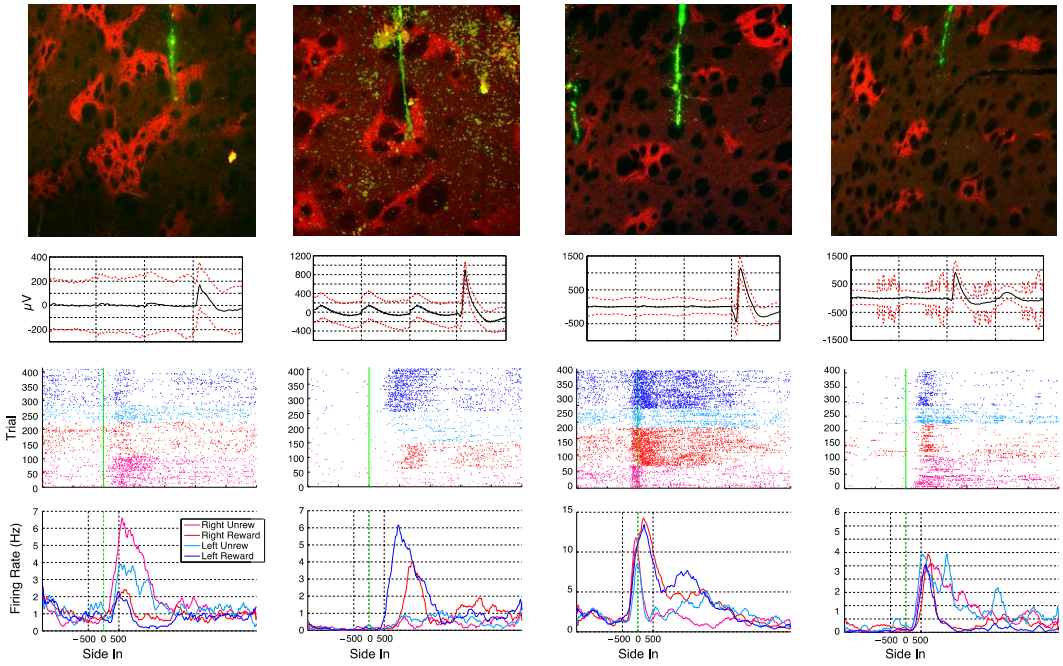


Figure 4.5. Single-unit recordings from identified striatal compartments. Four examples are shown (one per column; left two examples are patch neurons, right two examples are matrix neurons). In each case top panel shows histology of recording site, with mu-opioid-receptors (patch marker) labeled in red and electrode track labeled in green (DiO or CD11b). Other panels show single neuron waveforms (across the four wires of the tetrode) and event-related firing during the trial-and-error task, as raster plots and firing rate histograms. Activity is divided up by chosen movement direction (left/right) and whether the trial was rewarded or not.

4.3 Discussion

The results presented here are necessarily suggestive, rather than supportive, of a particular hypothesis because of a low yield of well defined patch units. Clearly, more recordings will be required before a clear picture of patch function emerges. These results do suggest that the previous approach we were using would require more recordings than are reasonable. Thus we have turned to new approaches. Future recordings will involve probes with increased channel count (256+ recording sites) and improved histological processing techniques. We have collaboratively developed silicon probes with smaller feature size and improved electrode density. We have also developed a refined approach to patch/matrix localization, which shows improvements on multiple fronts.

First, leaving the electrodes through the entire process from implantation until imaging will in place will minimize the possibility of tissue damage and distortion that arises from post-fixation headcap removal and the many stages of immunohistochemistry. In previous experiments with silicon probes, I found inconsistent damage patterns across shanks in the same tissue. Leaving the shank in the tissue should improve confidence that I am looking at the true electrode track rather than a micro tear and allow an additional level of visual co-registration (Figure 4.4).

Next, we have modified our labeling protocol to include the microglia marker CD11b. Previous attempts with the lipophilic tracer were inconsistent from experiement to experiment, with some showing too much labeling to resolve the tip of the electrode and others showing too little. The microglial labeling shows an obvious ‘scar’ around each shank that is consistent along the entire length of the electrode.

Finally, we have begun to perform thick slab immunohistochemistry combined with a tissue clearing protocol. Thick slabs of tissue offer several advantages in this scenario. Since the electrodes are left in the tissue, they must be sliced. Taking thicker sections means taking fewer sections. This minimizes the risk of electrodes falling out of the tissue during processing, minimizes damage to the tissue with fewer cuts, and maximizes the number of shank tips that will appear in the same tissue section. The down side is that thick tissue impedes antibody penetration and increases light scatter during confocal microscopy, where the goal is to scan through small cross-sections of the tissue and reconstruct a stack of images. To overcome these hurdles, we have altered our tissue processing protocol by increasing antibody penetration times and detergent concentrations. Additionally, we now clear the tissue of lipids and replace them with hydrogel to decrease light scattering, using a modified CLARITY (Chung et al. 2013) technique.

4.4 Methods

4.4.1 Experimental Procedures

Adult male Long-Evans rats were group housed on a 12 hour reversed light-dark cycle. Training and testing occurred during the dark phase. Animals were restricted to 15 g of rat chow daily and given access to water *ad libitum*. All procedures were approved by the University of Michigan Committee on the Use and Care of Animals. Training and testing occurred in sound-attenuated operant chambers (Med Associates, St. Albans, VT). Figure 4.1 shows a schematic diagram of the task. Each session lasted 2 hours, divided into blocks of 45-55 trials. At trial begin one of three central nose ports was illuminated

(Light On). The animal then poked and maintained his nose in the port (Center In) until a white-noise burst (Go Cue) was emitted. The animal then withdrew from the central poke and poked into the adjacent left or right nose ports (Side In). On rewarded trials, a food hopper delivered one sucrose pellet into a receptacle on the opposite side of the chamber, followed by an inter-trial interval (ITI; 5-8 s). Unrewarded trials were distinguished by an absence of sucrose pellet delivery, followed by the ITI. Left and right pokes were rewarded independently based upon a set of randomized block probabilities ([0.9L:0.1R], [0.9L:0.5R], [0.5L:0.1R], [0.5L:0.5R]. [0.1L:0.9R], [0.5L:0.9R], [0.1L:0.5R]). Failure to correctly poke the center port, hold for the entire duration, or respond to the Go Cue resulted in a 5-8 second timeout and concurrent houselight illumination. After two weeks of stable behavior (> 80% non-error trials), rats were ready for implantation. Behavioral electrophysiology methods including recording, spike sorting and cell classification have been described elsewhere (Gage et al. 2010; Leventhal et al. 2012; Schmidt et al. 2013).

4.4.2 Histology.

Rats were sacrificed under deep isoflurane anesthesia and perfused with Lactated Ringer's solution, followed by 4% Paraformaldehyde in 1x PBS. After removal, brains were stored in 30% sucrose for 24-48 hours. 50 um sections were taken on a cryostat and stored in 1x PBS at 4 C. Tissue was labeled for mu-opioid receptors (ImmunoStar 24216, 1:2000; Alexa Fluor 594, 1:500) and microglia marker CD11b (1:500 primary, Alexa Fluor 350, 1:500). All IHC steps were performed in dim lighting to prevent photobleaching of the DiO signal.

4.4.3 *Thick section histology with intact electrodes*

We have developed a new histological approach to enhancing electrode localization, which relies upon maintaining co-registration of the electrode in the tissue from the recording session until imaging. Rats are sacrificed under deep isoflurane anesthesia and perfused with Lactated Ringer's solution, followed by 4% Paraformaldehyde in 1x PBS. After perfusion all skin, muscle and fascia are removed from the skull, as well as the entire lower jaw and olfactory turbinates. To minimize movement of the electrodes, all cranial plates surrounding the array are kept intact. The electrode array, including dental acrylic and anchor screws, is left attached. To remove calcium from the bone and facilitate slicing, the skull/drive assembly is next submerged in a chelating solution (20% w/v EDTA in 1x PBS) and shaken in the cold room for 8 days. The chelating solution is changed every 48 hours to maintain a high concentration gradient. After 8 days, skull hardness is tested by attempting to cleanly slice a sliver (~ 500 um) off of a lateral skull ridge with a scalpel blade. If minimal effort is required, the skull is transferred to a 30% sucrose solution for 48 hours.

Prior to sectioning, voids between brain tissue and the skull are filled by a series of vacuum chamber sessions to improve consistency during sectioning. The skull assembly is submerged in a beaker of diluted O.T.C. cutting compound (20% OTC in 30% sucrose solution, followed by 50% OTC in 30% sucrose solution). The beaker is then placed in a vacuum chamber (15-20 psi) and allowed to rest 30 minutes under pressure until all air pockets are displaced. Horizontal sections are taken on a cryostat at 500 um thickness and placed in well plates with 1X PBS. Sections with electrodes were immunostained for mu-opioid receptors and microglia cell marker CD11b.

Chapter 5

Conclusion

In this dissertation, I have developed a further understanding of the neural substrates involved in adaptive decision-making. This work focused prominently on the signal conveyed by dopamine at its many cortical and striatal targets. Here, I have shown data that challenges the widely held view that dopamine provides a prediction error signal. Additionally, I have shown that the dopamine signal is not ubiquitous, but rather relays distinct information within cortical-striatal loops. I would like to conclude this work with a description of how my results fit into the current narrative of how dopamine informs decision-making and how my data fit into the decision-making literature more generally, with some observations and insights collected during the course of the experiments presented in these chapters.

5.1 Do my results contradict the RPE hypothesis?

A standard RPE account of dopamine function predicts that dopamine exclusively signals changes in reward expectation. Indeed, putative dopamine neuron recordings have shown firing activity that scales during cue presentation with the probability that the cue predicts a reward. At reward delivery these cells display activity that scales with the amount of unexpectedness (Oyama et al. 2010). It is worth noting that these cells do not display a sustained activity from cue onset until reward delivery, but rather two distinct pulses at cue and outcome. Following from these observations, and the dopamine = RPE

hypothesis more broadly, it would be predicted that during our task we would see, at most, these same two pulses at Trial Begin and at Side In represented in the phasic dopamine signal. The changes in dopamine at Trial Begin would reflect the likelihood of a reward and the changes at Side In would reflect the corresponding amount of unexpectedness of reward outcome. Alternatively, it may be predicted that a single phasic pulse would arise at trial outcome, signaling an RPE. Our measured dopamine response was far from these observations in several respects.

First, observed dopamine fluctuations early on in the trial but not at the earliest predictive cue (Center Light On), as a standard RPE account would predict. Rather, the dopamine consistently responded to the Center Nose In event, as the rat self-initiated each trial. Further, this dopamine rise during approach to the Center In was related to trial latency. As latency shortened, the dopamine ramped sooner, suggestive of a motivational drive to begin the trial. This interpretation may throw a flag to those more inclined toward an RPE interpretation. Indeed, it might be argued that short latencies correspond with higher rates of reward which likely co-occur with sequences of unexpected positive RPEs, thus driving elevated dopamine levels early in the trial. However, when dopamine was increased via optogenetic stimulation at outcome on the previous trial, there was no detectable effect on latency on the following trial. This result suggests that even immodest increases that mimic the largest RPE-like responses (completely unexpected reward delivery) at trial outcome are insufficient to promote enhanced vigor.

Secondly, if dopamine indeed signals RPE, sequential rewards should generate progressively smaller RPEs and, consequently, smaller phasic dopamine release events at Side In. Rather than observing changes in the peak level of dopamine at time of outcome on sequentially rewarded trials, we observed an equally high peak but a progressively

smaller difference between baseline and peak as rewards were sequentially achieved. In other words, the relative baseline dopamine level changed rather than the relative peak dopamine level. Baseline shifts were bidirectional, contingent on the sequence of outcomes (consecutive omissions resulted in lower baseline dopamine, while consecutive rewards resulted in higher baseline dopamine), consistent with shifts in the value of performing a trial rather than shifts in the impact of reward.

Finally, a major distinction between our observed dopamine signal and that predicted by the RPE hypothesis is the overall time course of the signal within a trial. The increases seen at Center In do not decay, but rather remain elevated until outcome. Consistent with an RPE hypothesis, a second increase is observed at reward delivery while a decrease occurs for omissions. Finally, on rewarded trials, the dopamine increases again until the rat has reached the food delivery port, at which time it immediately begins to decrease. The overall pattern is that of a ramp, punctuated by key task events, until reward is reached. The final increase until the reward port is not predicted by a standard RPE account. As the rat approaches the port, nothing new is being learned about the task and reward is certain. The dopamine should decay after the food hopper click, or perhaps even stay level, if there is no RPE. Yet there is an upward ramp, suggesting that dopamine is still useful in the movement required to approach the reward. Contrary to the RPE explanation, this observation fits nicely with a value function explanation because at the time just before reward acquisition, value is at a maximum. Indeed, it would be foolish for the rat to suddenly become disinterested in obtaining reward when it is most imminent. We suggest that this epoch of time when dopamine is maximal is precisely the time when the value of expected reward is maximal.

5.2 *RPE ramps?*

Alternative interpretations of ramping dopamine have been suggested by Gershman (2014) and Morita & Kati (2014). These interpretations are nested within the RPE framework and mathematically resolve how RPEs can behave as ramps rather than transient responses classically seen in Schultz experiments. Gershman's (2014) effort was made strictly to address the dopamine ramping observed in Howe et al. (2013) as rats approached a food pellet by traversing a maze arm. His account suggests that RPEs can in fact ramp as an agent's proximity to a deterministic reward decreases. This ramping is achieved through a quadratic transformation of the proximity variable. Importantly, the model makes no predictions for (i) encountering cues which would suddenly impact the value of the ramping RPE or (ii) how the ramping RPE behaves in probabilistic tasks more generally or (iii) what the corresponding value function would look like, besides recognizing that "[the transformation] implies a spatial compression of the value function similar to Weber's law, such that values of locations far from the goal are closer together than values of locations near the goal."

Because of the limitations noted in point (i) and (ii), it is difficult to predict what Gershman's model would predict for our task, which integrates spatial and non-spatial cues within a trial. The new ramping RPE is dependent upon proximity to reward only, so it may be the case that non-spatial cues (such as the food hopper click) would either have no effect on the ramping RPE or add an RPE to the ramping RPE. Unfortunately, the hopper click doesn't provide information about spatial proximity— it only signals reward availability. As for point (iii), I did observe a nonlinear DA ramp. It was much less linear than that observed by Howe et al. but more like the value function Gershman suggests. In the averaged within-trial dopamine ramp, dopamine showed an exponential increase as

reward grew linearly closer in time. Thus, sequential state values were similar early in the trial and grew increasingly disparate as the trial progressed. Ultimately, the new formulation of a ramping RPE signal does little to address the motivational component of dopamine. It rather suggests that dopamine is exclusively concerned with learning, which does not connect with our observation that increased dopamine at the Trial Begin immediately invigorates work.

5.3 Revising RPE

While none of these observations on their own is sufficient to dismiss the RPE hypothesis, together they provide substantive grounds for a revised view of dopamine function— a view that incorporates and explains many of the observed motivational phenomenon that occur as a result of normal and altered dopamine *as well as* dopamine's effect on learning. Schultz himself may very well agree to such a revision, yet he would not likely consider our account accurate. Schultz describes the dopamine response as having two components (W Schultz 2016), rather than one continuous value component as we suggest. The early component at a cue is 'activational' and signals surprise and motivational salience. The later component is evaluative, takes the form of an RPE, and persists until reward occurs. Thus, Schultz would likely say our observed dopamine fluctuations are entirely predicted and have already been explained by existing accounts. Some of his recent writing even suggests that the RPEs that dopamine signals give rise to a temporally-discounted subjective value estimate (Stauffer et al. 2016; W Schultz, Carelli, and Wightman 2015). This subjective value can be used to inform a utility function, and RPEs are now 'utility RPEs'.

The two component view may sufficiently describe the dopamine response in

Pavlovian tasks that alter reward amount or type, or simple instrumental tasks with a variable delay between response and reward. However, in self-paced instrumental tasks with multiple sequential responses the two-component view becomes cumbersome/unstable. When does the 'physical impact of the cue' end and the 'evaluation period' begin in such a task? While there is no denying the presence of two dopamine events, it is quite possible that they are not carrying different information. The value function account offered in these chapters does not make such a distinction, and thus offers a more parsimonious explanation of the data. The value function describes expected reward proximity and changes in this expectation simultaneously and continuously. In this way, it can account for the devaluation of distant rewards compared to closer ones, the learning that occurs when more or less reward unexpectedly appears or is omitted, and the motivational drives required to initiate work and to recover the reward when the work is completed.

5.4 Is there functional difference between phasic and tonic dopamine?

The term 'tonic', as it applies to dopamine, has come to be used in a rather loose manner. It is used at once to describe a method of measurement and a physiological time course. In the literature, 'tonic' has been applied to slow non-synchronous firing (~ 3-5 Hz), slowly-evolving changes in extracellular dopamine concentration, and dopamine fluctuations measured by dialysis, usually over several minutes (A. Grace 1991; A. Grace et al. 2007). A couple of fundamental assumptions are being made when using the term 'tonic' in this manner.

Firstly, it seems implicitly assumed that the dopamine measured at the terminal is roughly correlated with the dopamine cell population activity measured in the midbrain.

Thus, high levels of 'tonic' dopamine in the terminal must arise from increases in activity at the cell body and an increased proportion of non-silent dopamine neurons (Chéramy et al. 1990). This assumption is manifest in manipulations (or indeed models like Humphries et al. (2012)) of dopamine that broadly increase terminal concentrations via amphetamine or dopamine transporter knockdown (Beeler et al. 2010). These manipulations are said to mimic changes in tonic dopamine by specifically manipulating tone while leaving bursting intact, but it is usually unclear whether the authors are suggesting that the manipulations are mimicking the activity of dopamine neurons or simply dopamine tone. It has become clear that dopamine release can be modulated at the synapse, independent of cell firing. Dopamine release can be triggered by presynaptic glutamatergic inputs from neocortex (Floresco et al. 2003) or synchronous local cholinergic input (Threlfell et al. 2015). Clearly, the term 'tonic' needs to be clarified when discussing measured dopamine vs. dopamine cell activity, rather than presuming they are equivalent.

Secondly, it is assumed that microdialysis measures 'tonic' dopamine exclusively. The assertion is that the dopamine signal measured by dialysis receives a negligible contribution from phasic burst events, because the dopamine released by phasic bursts is rapidly removed from the extracellular space by either transport or metabolism, or confined mostly to the synapse (A. Grace 1991; Nirenberg et al. 1997; Floresco et al. 2003). However, this idea may be a byproduct of the fact that 'tonic' dopamine has been traditionally measured over tens of minutes or hours (Nieoullon, Cheramy, and Glowinski 1978; Chéramy et al. 1990; A. Grace 1991). On the other hand, if the critique collapses to whether or not a dialysis probe is measuring dopamine inside the synaptic cleft, the same critique could be applied to voltammetry, the currently favored technique for

measuring phasic dopamine (Phillips and Wightman 2004). This is not to say that there are not two unique modes of firing, but rather that the relative contribution of each to the overall measured dopamine signal at the target may not be clearly distinguishable.

In their 2007 work, Niv et al. similarly suggest this may be the case— that 'tonic' dopamine may in fact be the time-integrated average of the phasic release events, filtered by re-uptake. Thus, the distinction may simply be a matter of the timescale chosen for measurement. While my observations are consistent with the hypothesis that 'tonic' dopamine (terminal dopamine concentration measured over a slower timescale) signals the rate of reward, they are distinct from the hypothesis as originally proposed by Niv et al. (2007) in a crucial way. Specifically, the authors proposed that tonic dopamine is an accumulation of phasic RPE events that integrate into a reward rate. While the RPEs are reactive, the reward rate is predictive in that it estimates future reward and thus modulates vigor. One testable prediction that arises from their logic is that particularly large RPEs, provoking correspondingly large increases in phasic dopamine, can be used to enhance vigor. Our optogenetic increase of dopamine at reward outcome, selected to mimic the dopamine transients observed during unexpected reward delivery, had no impact on latency on the following trial. In other words, large RPEs on the previous trial do not enhance vigor on the next trial. Yet, it is the case that latencies are faster when reward rate is higher. Why?

We suggest that the moment-by-moment dopamine is encoding a value function rather than RPEs. As rewards accumulate, the baseline dopamine concentration increases and the expected value of doing the task increases. The baseline increase may be sufficient to drive vigor while the transient unexpected shifts in value (TD-RPEs) are not. On a slower timescale, this value and reward rate look very similar. Specifically, when

the sMDP model estimate of state values are averaged across one-minute time bins, the minute-by-minute value estimate is strikingly similar to the minute-by-minute reward rate estimate. Thus, I conclude that tonic dopamine is likely the time-smeared phasic dopamine, but that on both timescales dopamine is signaling value, rather than RPE. □

5.5 Measuring dopamine: Insights and Limitations

The data support the hypothesis that tonic dopamine is nothing more than the time-integrated phasic signal (Niv et al. 2007), however this comes with a couple of caveats. While I show that mesolimbic dopamine (on what is normally considered a tonic timescale) is likely the averaged phasic signal, I also show that this underlying phasic signal is not RPE. Rather, it is a continuous value function, estimating moment-by-moment future reward, in which sub-second fluctuations report deviations in the value function. These deviations correspond to delta, or the temporal difference error, in a TDRL model. Of course, to completely validate the claim that tonic [DA] is the same as phasic [DA], the most conclusive study would involve simultaneous voltammetry and fast-timescale microfluidic measurements. This is the other caveat, and it is not readily addressed with the technology at my disposal.

One real hurdle was overcoming the limitations of measurement in each technique. The ideal measurement would provide a true baseline and offer a resolution that spans from sub-second to tens of minutes. The dopamine signal from voltammetry measurements 'drifts' over tens of seconds in a way that cannot be compensated for with filtering. As a result, the signal must be baselined at intervals, so a true baseline cannot be detected. On the other hand, it is increasingly difficult to reliably detect signal in increasingly small volumes of dialysate. One problem is that small deviations in the

pipetting of the three reagents into each sample can result in larger sample-to-sample measurement deviation (most especially deviations in the internal standard volume). The other problem is hitting a floor effect in the number of ions that can be detected by the mass spectrometer when trying to detect so many analytes. The 2 uL/min sampling rate I used was optimized to give the best sample-to-sample reliability (stability in total fraction volume) while staying at least an order of magnitude above the limits of detection for our analytes of interest.

5.6 Dopamine and work

The idea that dopamine is involved in the regulation of motivational effort has a solid experimental basis. For example, Aberman, Ward, and Salamone (1998) showed that accumbens dopamine depletion decreases the amount of lever-pressing rats are willing to perform in a work-dependent manner. Under the dopamine depletion condition, rats will press at a fixed-ratio requirement of 1 at a similar rate to control animals. However, as the fixed-ratio increases to 4, 16 and 64 the dopamine-depleted rats are increasingly less willing to lever-press. This general result has been replicated and expanded several times (J D Salamone et al. 2001; Ishiwari et al. 2004; J D Salamone et al. 2007). Before our experiments, the exact decision variable being carried by dopamine that enhances effort was not previously well defined. For example, in the context of a utility function, dopamine could perceivably influence effort by conveying information about the net benefit, the net cost, or the overall utility of an action. In this framework, value would represent the net benefit of engaging in the task.

While I did not parametrically manipulate the work requirements, as is traditionally done by altering lever-pressing requirements via number of presses or force

required (J D Salamone et al. 2001; Beeler et al. 2010), the amount of work/effort nevertheless varies throughout the session. There are epochs of low reward probability and epochs of high reward probability that arise from the regularly changing block probabilities that the rats experience. A worst-case scenario, assuming no procedural errors, would be attempting 10 trials to recover zero rewards while a best case would be ten rewards in ten trials. I observed that rats decreased latency to begin a trial and increased the number of attempted trials in each minute as the rate of reward increased. Our finding that dopamine fluctuates with the minute-by-minute rate of reward and the moment-by-moment value is evidence that dopamine exclusively encodes net expected *benefits* of work, as opposed to a combination of costs and benefits. This is consistent with prior similar reports measuring phasic dopamine during lever-pressing (Gan, Walton, and Phillips 2010) and the finding that dopamine cell firing during an instrumental decision-making task scales with net expected return (Morris et al. 2006).

5.7 A distributed valuation system

Optimal decision-making has many components. There are costs and benefits, risks and uncertainty, and preferences that change based on internal and external factors. Evaluating how to allocate effort and utilize learned information can be a tall order. A wealth of previous work has uncovered traces of various aspects of this process, leading to speculation that specific subregions and neuromodulators offer unique contributions to a broadly distributed decision-making system. In light of this, the expansive microdialysis dataset I generated in Chapters 2 and 3 remains largely unexploited, but not necessarily unexplored. While I tested several behavioral variables related to performance (Reaction Time, Movement Time, Latency), outcome (reward rate, omissions, cumulative

reward, reward volatility, RPE, value), and choice (% ipsi, % contra, probability of win-stay, probability of switch, and a Bayesian estimate of uncertainty of action values), the most prominent and replicated relationship was the dopamine:reward rate correlation. However, while the task I use throughout this dissertation is rich in decision-making variables, some variables may be under-expressed in the task and thus difficult to detect with the methods employed. I will next survey a sample of unresolved hypotheses and examine how my data fit into the broader literature of decision-making.

5.8 Dopamine and adenosine: Complimentary roles in effort?

There is a strong overlap in the mapping of dopamine targets and adenosine targets in nucleus accumbens, with a special co-localization of D2 dopamine and A2A adenosine receptor subtypes (Hillion et al. 2002). There is also a strong similarity in the impact of dopamine and adenosine on effort (John D Salamone et al. 2016). On its own, A2A receptor agonism reduces lever pressing but increases the consumption of less-preferred and easily obtained food (Font et al. 2008). Additionally, A2A agonism selectively decreases effort in tasks with high work requirements but not low work requirements (Mingote et al. 2008). Interestingly, decreases in effort that occur under dopamine antagonism can be rescued by following up with A2A antagonism (Worden et al. 2009; Ishiwari et al. 2004), suggesting that adenosine and dopamine in the accumbens may interact to flexibly drive effortful behavior in a manner similar to a gas pedal and a brake pedal.

With respect to my own data, adenosine in the accumbens did not reliably correlate with my measures of effort (attempts/min, average latency). However, this does not definitively rule out its role in shaping effort. Many of the tasks in which adenosine

receptor antagonism decreased effortful engagement directly challenged metabolic energy expenditure in a non-trivial manner. Voluntary physical exertion was required to climb the barriers in Mott et al. (2009) or run the wheels in Correa et al. (2016). It may be precisely this component of the effort equation that adenosine mediates. Of note, dopamine antagonism has been shown to alter physical effort in a different manner from cognitive effort (Hosking, Floresco, and Winstanley 2015). My task does not specifically challenge voluntary physical effort and rats show no signs of fatigue within a session such as gradual slowing of latencies or number of trials attempted. Importantly, there was a significant correlation between minute-by-minute accumbens core adenosine and dopamine levels (as well as between adenosine and dopamine metabolites), suggestive of the naturally interactive relationship hypothesized (Farrar et al. 2010).

5.9 Serotonin and the cost of vigor

Serotonin has been proposed to provide an opponent role to dopamine, signaling an estimate of aversive outcomes and the opportunity cost of vigor (Daw, Kakade, and Dayan 2002; Boureau and Dayan 2011; Niv et al. 2007). In other words, as evidence against taking an action accumulates, the metabolic cost of repeating actions increases is reflected in concomitant fluctuations in serotonin. In this framework, a total utility function is generated by serotonin signaling the cost of vigor and dopamine signaling the cost of sloth. In the trial-and-error task, I would expect to see serotonin reflect the number of omissions in each minute, as the benefits of effortful work decline. Similarly, the measured serotonin should change with latency as motivation decreases. Serotonin could also fluctuate with the probability of switching, as evidence against selecting the current option accumulates. The measured serotonin reflected none of these metrics. One

reason may be the lack of a pronounced affective component in our task, such as a mild footshock punishment for incorrect responses. Several studies that manipulate serotonin have seen no performance effects in tasks that lack an affective component (Clark et al. 2005; Evers et al. 2007). Another possibility is that the specific opponent effects of serotonin occur closer to the dopamine cell bodies, where serotonin has prominent, albeit mixed effects (Tsai; Gervais and Rouillard 2000). The lack of support for the opponent interaction hypothesis only highlights the importance of selecting tasks with an appropriate specificity for behavioral demands.

There are many other lines of questioning that can be pursued with this dataset in relation to other transmitters, including where and to what extent glutamate (Gleich et al. 2015) and acetylcholine (Threlfell et al. 2012; Rice, Patel, and Cragg 2011) can shape dopamine efflux. The set of analyses presented in Chapters 2 and 3 are by no means exhaustive, however, they do offer a glance across multiple decision-making circuits and neuromodulators. I propose that our trial-and-error task challenges the dopamine system, as evidenced by the reliable and robust relationships observed between value-related task variables, dopamine, and dopamine metabolites.

5.10 Valuation in the anterior cingulate cortex and orbitofrontal cortex

Much recent work has placed a spotlight on the role of the ACC in optimal foraging. While traditionally well-observed in error monitoring, it has been argued that the ACC may in fact have a wider scope in processing a history of outcomes based on positive and negative evidence (Rushworth et al. 2007; Kolling et al. 2012). Keeping track of a history is decidedly distinct from maintaining and updating a single decision

variable (like an action value). In the context of foraging, this would manifest as series of events, each contributing a single piece of evidence for or against an action, with each piece actively contributing weight to the decision. Tracking both positive and negative evidence is similar net utility function, which is useful in determining which action is worthwhile to engage in. Indeed, monkeys doing a reversal learning task with ACC lesions appropriately switched during the reversal phase but displayed an overall worse performance than controls (Kennerley et al. 2006). This suggests an inability to actively maintain information about reward history.

Imaging studies have revealed that the ACC and OFC are commonly activated together during decision-making tasks but are reported to participate in distinctive forms of evaluation (Walton et al. 2007; Kennerley, Behrens, and Wallis 2011). Activity in the ACC has been shown to contrast recent outcomes against predicted outcomes, similar to performing subjective value estimations. The OFC, on the other hand, has been shown to encode chosen action values with activity seen during choice and outcome epochs (Furuyashiki and Gallagher 2007; Y. Li et al. 2016; Kennerley, Behrens, and Wallis 2011). Additionally, the ACC and OFC diverge in the types of cost-benefit functions they perform in tasks that challenge effort allocation. Lesions of the ACC resulted in preference toward the smaller reward/less effort arm of a T-maze while sparing the ability to outwait a delay for a larger reward. Lesions of the OFC resulted in impulsive choices on a delay task while sparing the effort required for the larger reward in the T-maze (Rudebeck et al. 2006).

The idea that the ACC is critical for maintaining outcome history or some form net utility is certainly intriguing. In my regression analysis I used a model that contained both rewards and omissions to see whether any of my analyte concentrations in the ACC

could be accounted for. I did not find such a relationship, but this model is admittedly a rather crude way to approach utility. Alternatively, it may be that the task is rather simple and does not rely on maintaining a reward history of more than one or two trials back, or possibly no single transmitter gives rise to these functions. Even though I did not see a consistent value coding in my anterior cingulate placements, 3 out of 7 did show a significant correlation between reward rate and dopamine. From the pattern of results I describe in Chapter 3, it is clear that the possibility of 'hotspots' should not be overlooked. It has been suggested that subregions of the anterior cingulate (Walton et al. 2007), as well as the orbitofrontal cortex (Walton et al. 2011), may well handle different components of valuation. Indeed, in a meta-anlaysis of subjective value coding regions in human fMRI data, the posterior cingulate (but not anterior cingulate) emerges during trial outcome for positive outcomes only, and scales with subjective value estimates of available alternatives (Bartra, McGuire, and Kable 2013). It is possible that a distinct hotspot may emerge from a more exhaustive scan of this area.

5.11 Behavioral flexibility and 'sudden insights'

There is a breadth of evidence that the medial prefrontal cortex, the anterior cingulate, and dorsal striatum mediate the behaviors which regularly fall under the definition of 'behavioral flexbility', including reversal learning, set-shifting/task switching, and strategy switching (Durstewitz et al. 2010). In particular, dopamine has been implicated in facilitating these behaviors (Floresco et al. 2005; Durstewitz and Seamans 2008; Tai et al. 2012). Across the many behavioral variables tested in my dialysis experiments, there was a decisive lack of evidence in support of dopamine fluctuations driving behavioral flexibility.

One possible explanation for this result is that signals that mark behavioral switches are extremely transient and thus difficult to detect at the time resolution I measured with. Ensemble recordings in the prefrontal cortex have shown that neural networks undergo abrupt transitions rather than gradual shifts during rule learning (Durstewitz et al. 2010). Similarly, sudden changes in task structure that increase uncertainty about reward contingency alter PFC neural activity for only a handful of trials (Karlsson, Tervo, and Karpova 2012). Finally, in a reversal learning task, measured dopamine during a series of three reversals only increased during the first reversal, indicating that a relationship between dopamine and switching behavior is not only transient within a session, but may be difficult to detect in general (Van Der Meulen et al. 2007). These results allude to the idea that during reversal learning there are transient 'sudden insights' that prompt a shift in strategy. If this were the case, our analyses would not have captured such a relationship.

5.12 Concluding remarks

In this thesis, I sought to gain a better understanding of the circuitry underlying decision-making. This thesis challenges important ideas about dopamine function in the context of adaptive behavior. First, I challenge the dopamine= RPE hypothesis. I demonstrate that mesolimbic dopamine plays a role in learning and motivation by conveying a single signal—a value function. I show that this signal is not confined to the ventral striatum, but is shared by its afferent circuitry in the medial prefrontal cortex. This confirms that dopamine does not provide unique signals to cortical and striatal circuitry. The work presented here contributes to a more refined understanding of the

timing and functional connectivity of the dopamine system, which offers important implications for clinical manipulations of dopamine.

References

Aberman, JE, SJ Ward, and JD Salamone. 1998. "Effects of Dopamine Antagonists and Accumbens Dopamine Depletions on Time-Constrained Progressive-Ratio Performance." *Pharmacology Biochemistry and Behavior* 61 (4):341-45.

Adamantidis, Antoine R, Hsing-Chen Tsai, Benjamin Boutrel, Feng Zhang, Garret D Stuber, Evgeny A Budygin, Clara Touriño, Antonello Bonci, Karl Deisseroth, and Luis de Lecea. 2011. "Optogenetic Interrogation of Dopaminergic Modulation of the Multiple Phases of Reward-Seeking Behavior." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 31 (30): 10829–35.

Ainslie, George. 2005. "Précis of Breakdown of Will." *The Behavioral and Brain Sciences* 28 (5): 635–50; discussion 650–73.

Bartra, Oscar, Joseph T. McGuire, and Joseph W. Kable. 2013. "The Valuation System: A Coordinate-Based Meta-Analysis of BOLD fMRI Experiments Examining Neural Correlates of Subjective Value." *NeuroImage* 76. Elsevier Inc.: 412–27.

Bayer, Hannah M, Brian Lau, and Paul W Glimcher. 2007. "Statistics of Midbrain Dopamine Neuron Spike Trains in the Awake Primate." *Journal of Neurophysiology* 98 (3): 1428–39.

Beeler, Jeff a, Nathaniel Daw, Cristianne R M Frazier, and Xiaoxi Zhuang. 2010. "Tonic Dopamine Modulates Exploitation of Reward Learning." *Frontiers in Behavioral Neuroscience* 4 (November): 170.

Beeler, Jeff a, Cristianne R M Frazier, and Xiaoxi Zhuang. 2012. “Putting Desire on a Budget: Dopamine and Energy Expenditure, Reconciling Reward and Resources.” *Frontiers in Integrative Neuroscience* 6 (July): 49.

Beeler, Jeff a. 2012. “Thorndike’s Law 2.0: Dopamine and the Regulation of Thrift.” *Frontiers in Neuroscience* 6 (AUG): 116.

Beierholm, Ulrik, Marc Guitart-Masip, Marcos Economides, Rumana Chowdhury, Emrah Düzel, Ray Dolan, and Peter Dayan. 2013. “Dopamine Modulates Reward-Related Vigor.” *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* 38 (8): 1495–1503.

Berridge, K C, and T E Robinson. 1998. “What Is the Role of Dopamine in Reward: Hedonic Impact, Reward Learning, or Incentive Salience?” *Brain Research. Brain Research Reviews* 28 (3): 309–69.

Berridge, Kent. 2007. “The Debate over Dopamine’s Role in Reward: The Case for Incentive Salience.” *Psychopharmacology* 191 (3): 391–431.

Boureau, Y-Lan, and Peter Dayan. 2011. “Opponency Revisited: Competition and Cooperation Between Dopamine and Serotonin.” *Neuropsychopharmacology* 36 (1). Nature Publishing Group: 74–97.

Bromberg-Martin, Ethan S, Masayuki Matsumoto, and Okihide Hikosaka. 2010a. “Distinct Tonic and Phasic Anticipatory Activity in Lateral Habenula and Dopamine Neurons.” *Neuron* 67 (1): 144–55.

Bromberg-Martin, Ethan S., Masayuki Matsumoto, and Okihide Hikosaka. 2010b. “Dopamine in Motivational Control: Rewarding, Aversive, and Alerting.” *Neuron*

68 (5). Elsevier Inc.: 815–34.

Buzsáki, György. 2004. “Large-Scale Recording of Neuronal Ensembles.” *Nature Neuroscience* 7 (5). Nature Publishing Group: 446–51.

Cagniard, Barbara, Peter D Balsam, Daniela Brunner, and Xiaoxi Zhuang. 2006. “Mice with Chronically Elevated Dopamine Exhibit Enhanced Motivation, but Not Learning, for a Food Reward.” *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* 31 (7): 1362–70.

Cannon, Claire Matson, and Richard D Palmiter. 2003. “Reward without Dopamine.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 23 (34): 10827–31.

Chéramy, A, L Barbeito, G Godeheu, J M Desce, A Pittaluga, T Galli, F Artaud, and J Glowinski. 1990. “Respective Contributions of Neuronal Activity and Presynaptic Mechanisms in the Control of the in Vivo Release of Dopamine.” *Journal of Neural Transmission. Supplementum* 29 (January): 183–93.

Chung, Kwanghun, Jenelle Wallace, Sung-Yon Kim, Sandhiya Kalyanasundaram, Aaron S Andalman, Thomas J Davidson, Julie J Mirzabekov, et al. 2013. “Structural and Molecular Interrogation of Intact Biological Systems.” *Nature* 497 (7449): 332–37.

Clark, L, J P Roiser, R Cools, D C Rubinsztein, B J Sahakian, and T W Robbins. 2005. “Stop Signal Response Inhibition Is Not Modulated by Tryptophan Depletion or the Serotonin Transporter Polymorphism in Healthy Volunteers: Implications for the 5-HT Theory of Impulsivity.” *Psychopharmacology* 182 (4): 570–78.

Coizet, V, E J Dommett, P Redgrave, and P G Overton. 2006. “Nociceptive Responses of

Midbrain Dopaminergic Neurones Are Modulated by the Superior Colliculus in the Rat." *Neuroscience* 139 (4): 1479–93.

Cools, Roshan. 2015. "The Cost of Dopamine for Dynamic Cognitive Control." *Current Opinion in Behavioral Sciences* 4. Elsevier Ltd: 1–8.

Correa, Mercè, Marta Pardo, Pilar Bayarri, Laura López-Cruz, Noemí San Miguel, Olga Valverde, Catherine Ledent, and John D Salamone. 2016. "Choosing Voluntary Exercise over Sucrose Consumption Depends upon Dopamine Transmission: Effects of Haloperidol in Wild Type and Adenosine A_2 AKO Mice." *Psychopharmacology* 233 (3): 393–404.

Crittenden, Jill R, and Ann M Graybiel. 2011. "Basal Ganglia Disorders Associated with Imbalances in the Striatal Striosome and Matrix Compartments." *Frontiers in Neuroanatomy* 5 (September): 59.

Daw, Nathaniel D, Aaron C Courville, David S Tourtezky, and David S Touretzky. 2006. "Representation and Timing in Theories of the Dopamine System." *Neural Computation* 18 (7): 1637–77.

Daw, Nathaniel D, and Kenji Doya. 2006. "The Computational Neurobiology of Learning and Reward." *Current Opinion in Neurobiology* 16 (2): 199–204.

Daw, Nathaniel D, Yael Niv, and Peter Dayan. 2005. "Uncertainty-Based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control." *Nature Neuroscience* 8 (12): 1704–11.

Daw, Nathaniel D., Sham Kakade, and Peter Dayan. 2002. "Opponent Interactions between Serotonin and Dopamine." *Neural Networks* 15 (4-6): 603–16.

Day, Jeremy J, Mitchell F Roitman, R Mark Wightman, and Regina M Carelli. 2007. "Associative Learning Mediates Dynamic Shifts in Dopamine Signaling in the Nucleus Accumbens." *Nature Neuroscience* 10 (8): 1020–28.

Demanuele, Charmaine, Peter Kirsch, Christine Esslinger, Mathias Zink, Andreas Meyer-Lindenberg, and Daniel Durstewitz. 2015. "Area-Specific Information Processing in Prefrontal Cortex during a Probabilistic Inference Task: A Multivariate fMRI BOLD Time Series Analysis." *PloS One* 10 (8): e0135424.

Deutch, A Y. 1993. "Prefrontal Cortical Dopamine Systems and the Elaboration of Functional Corticostriatal Circuits: Implications for Schizophrenia and Parkinson's Disease." *Journal of Neural Transmission. General Section* 91 (2-3): 197–221.

Dreyer, J. K., K. F. Herrik, R. W. Berg, and J. D. Hounsgaard. 2010. "Influence of Phasic and Tonic Dopamine Release on Receptor Activation." *Journal of Neuroscience* 30 (42): 14273–83.

du Hoffmann, Johann, and Saleem M Nicola. 2014. "Dopamine Invigorates Reward Seeking by Promoting Cue-Evoked Excitation in the Nucleus Accumbens." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 34 (43): 14349–64.

Durstewitz, Daniel, and Jeremy K Seamans. 2008. "The Dual-State Theory of Prefrontal Cortex Dopamine Function with Relevance to Catechol-O-Methyltransferase Genotypes and Schizophrenia." *Biological Psychiatry* 64 (9). Elsevier: 739–49.

Durstewitz, Daniel, Nicole M Vittoz, Stan B Floresco, and Jeremy K Seamans. 2010. "Abrupt Transitions between Prefrontal Neural Ensemble States Accompany Behavioral Transitions during Rule Learning." *Neuron* 66 (3): 438–48.

Eblen, F, and A M Graybiel. 1995. “Highly Restricted Origin of Prefrontal Cortical Inputs to Striosomes in the Macaque Monkey.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 15 (9): 5999–6013.

Eshel, Neir, Michael Bukwich, Vinod Rao, Vivian Hemmelder, Ju Tian, and Naoshige Uchida. 2015. “Arithmetic and Local Circuitry Underlying Dopamine Prediction Errors.” *Nature* 525 (7568): 243–46.

Eshel, Neir, Ju Tian, Michael Bukwich, and Naoshige Uchida. 2016. “Dopamine Neurons Share Common Response Function for Reward Prediction Error.” *Nature Neuroscience* 19 (3): 479–86.

Euston, David R, Aaron J Gruber, and Bruce L McNaughton. 2012. “The Role of Medial Prefrontal Cortex in Memory and Decision Making.” *Neuron* 76 (6): 1057–70.

Evers, E A T, F M van der Veen, D Fekkes, and J Jolles. 2007. “Serotonin and Cognitive Flexibility: Neuroimaging Studies into the Effect of Acute Tryptophan Depletion in Healthy Volunteers.” *Current Medicinal Chemistry* 14 (28): 2989–95.

Farrar, A M, K N Segovia, P A Randall, E J Nunes, L E Collins, C M Stopper, R G Port, et al. 2010. “Nucleus Accumbens and Effort-Related Functions: Behavioral and Neural Markers of the Interactions between Adenosine A2A and Dopamine D2 Receptors.” *Neuroscience* 166 (4): 1056–67.

Fellows, Lesley K, and Martha J Farah. 2003. “Ventromedial Frontal Cortex Mediates Affective Shifting in Humans: Evidence from a Reversal Learning Paradigm.” *Brain : A Journal of Neurology* 126 (Pt 8): 1830–37. doi:10.1093/brain/awg180.

Fiorillo, Christopher D, Philippe N Tobler, and Wolfram Schultz. 2003. “Discrete Coding

of Reward Probability and Uncertainty by Dopamine Neurons." *Science (New York, N.Y.)* 299 (5614): 1898–1902.

Floresco, Stan B, Orsolya Magyar, Sarvin Ghods-Sharifi, Claudia Vexelman, and Maric T L Tse. 2005. "Multiple Dopamine Receptor Subtypes in the Medial Prefrontal Cortex of the Rat Regulate Set-Shifting." *Neuropsychopharmacology* 31 (2): 297–309.

Floresco, Stan B, Anthony R West, Brian Ash, Holly Moore, and Anthony A Grace. 2003. "Afferent Modulation of Dopamine Neuron Firing Differentially Regulates Tonic and Phasic Dopamine Transmission." *Nature Neuroscience* 6 (9): 968–73.

Font, Laura, Susana Mingote, Andrew M Farrar, Mariana Pereira, Lila Worden, Colin Stopper, Russell G Port, and John D Salamone. 2008. "Intra-Accumbens Injections of the Adenosine A2A Agonist CGS 21680 Affect Effort-Related Choice Behavior in Rats." *Psychopharmacology* 199 (4): 515–26.

Frank, Michael J, Bradley B Doll, Jen Oas-Terpstra, and Francisco Moreno. 2009. "Prefrontal and Striatal Dopaminergic Genes Predict Individual Differences in Exploration and Exploitation." *Nature Neuroscience* 12 (8): 1062–68.

Freed, C R, and B K Yamamoto. 1985. "Regional Brain Dopamine Metabolism: A Marker for the Speed, Direction, and Posture of Moving Animals." *Science (New York, N.Y.)* 229 (4708): 62–65.

Friedman, Alexander, Daigo Homma, Leif G. Gibb, Ken-ichi Amemori, Samuel J. Rubin, Adam S. Hood, Michael H. Riad, and Ann M. Graybiel. 2015. "A Corticostriatal Path Targeting Striosomes Controls Decision-Making under Conflict." *Cell* 161 (6). Elsevier Inc.: 1320–33.

Fujiyama, Fumino, Jaerin Sohn, Takashi Nakano, Takahiro Furuta, Kouichi C Nakamura, Wakoto Matsuda, and Takeshi Kaneko. 2011. "Exclusive and Common Targets of Neostriatofugal Projections of Rat Striosome Neurons: A Single Neuron-Tracing Study Using a Viral Vector." *The European Journal of Neuroscience* 33 (4): 668–77.

Furuyashiki, Tomoyuki, and Michela Gallagher. 2007. "Neural Encoding in the Orbitofrontal Cortex Related to Goal-Directed Behavior." *Annals of the New York Academy of Sciences* 1121 (December): 193–215.

Gage, Gregory J, Colin R Stoetzner, Alexander B Wiltschko, and Joshua D Berke. 2010. "Selective Activation of Striatal Fast-Spiking Interneurons during Choice Execution." *Neuron* 67 (3): 466–79.

Gan, Jerylin O, Mark E Walton, and Paul E M Phillips. 2010. "Dissociable Cost and Benefit Encoding of Future Rewards by Mesolimbic Dopamine." *Nature Neuroscience* 13 (1): 25–27.

Gerfen, C R. 1989. "The Neostriatal Mosaic: Striatal Patch-Matrix Organization Is Related to Cortical Lamination." *Science (New York, N.Y.)* 246 (4928): 385–88.

Gershman, Samuel J. 2014. "Dopamine Ramps Are a Consequence of Reward Prediction Errors." *Neural Computation* 26 (3): 467–71.

Gervais, J, and C Rouillard. 2000. "Dorsal Raphe Stimulation Differentially Modulates Dopaminergic Neurons in the Ventral Tegmental Area and Substantia Nigra." *Synapse (New York, N.Y.)* 35 (4): 281–91.

Gleich, Tobias, Lorenz Deserno, Robert Christian Lorenz, Rebecca Boehme, Anne Pankow, Ralph Buchert, Simone Kühn, Andreas Heinz, Florian Schlagenhauf, and

Jürgen Gallinat. 2015. “Prefrontal and Striatal Glutamate Differently Relate to Striatal Dopamine: Potential Regulatory Mechanisms of Striatal Presynaptic Dopamine Function?” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 35 (26): 9615–21.

Grace, AA. 1991. “Phasic versus Tonic Dopamine Release and the Modulation of Dopamine System Responsivity: A Hypothesis for the Etiology of Schizophrenia.” *Neuroscience*.

Grace, Anthony A, Stan B Floresco, Yukiori Goto, and Daniel J Lodge. 2007a. “Regulation of Firing of Dopaminergic Neurons and Control of Goal-Directed Behaviors.” *Trends in Neurosciences* 30 (5): 220–27.

Grace, Anthony A, Stan B Floresco, Yukiori Goto, and Daniel J Lodge. 2007b. “Regulation of Firing of Dopaminergic Neurons and Control of Goal-Directed Behaviors.” *Trends in Neurosciences* 30 (5): 220–27.

Graybiel, Ann M. 1998. “Preferential Localization of Self-Stimulation Sites in Striosomes ¯Patches in the Rat Striatum” 95 (May): 6486–91.

Guitart-Masip, Marc, Ulrik R Beierholm, Raymond Dolan, Emrah Duzel, and Peter Dayan. 2011. “Vigor in the Face of Fluctuating Rates of Reward: An Experimental Examination.” *Journal of Cognitive Neuroscience* 23 (12): 3933–38.

Haith, Adrian M, Thomas R Reppert, and Reza Shadmehr. 2012. “Evidence for Hyperbolic Temporal Discounting of Reward in Control of Movements.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 32 (34): 11727–36.

Hamid, Arif A, Jeffrey R Pettibone, Omar S Mabrouk, Vaughn L Hetrick, Robert Schmidt, Caitlin M Vander Weele, Robert T Kennedy, Brandon J Aragona, and Joshua D Berke. 2015. "Mesolimbic Dopamine Signals the Value of Work." *Nature Neuroscience* 19 (1). Nature Publishing Group: 117–26.

Hart, Andrew S, Robb B Rutledge, Paul W Glimcher, and Paul E M Phillips. 2014. "Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 34 (3): 698–704.

Heidbreder, Christian a., and Henk J. Groenewegen. 2003. "The Medial Prefrontal Cortex in the Rat: Evidence for a Dorso-Ventral Distinction Based upon Functional and Anatomical Characteristics." *Neuroscience and Biobehavioral Reviews* 27 (6): 555–79.

Heien, Michael L A V, Amina S Khan, Jennifer L Ariansen, Joseph F Cheer, Paul E M Phillips, Kate M Wassum, and R Mark Wightman. 2005. "Real-Time Measurement of Dopamine Fluctuations after Cocaine in the Brain of Behaving Rats." *Proceedings of the National Academy of Sciences of the United States of America* 102 (29): 10023–28.

Hillion, Joelle, Meritxell Canals, Maria Torvinen, Vicent Casado, Rizaldy Scott, Anton Terasmaa, Anita Hansson, et al. 2002. "Coaggregation, Cointernalization, and Codesensitization of Adenosine A2A Receptors and Dopamine D2 Receptors." *The Journal of Biological Chemistry* 277 (20): 18091–97.

Hornak, J, J O'Doherty, J Bramham, E T Rolls, R G Morris, P R Bullock, and C E Polkey. 2004. "Reward-Related Reversal Learning after Surgical Excisions in

Orbito-Frontal or Dorsolateral Prefrontal Cortex in Humans." *Journal of Cognitive Neuroscience* 16 (3): 463–78.

Hosking, Jay G, Stan B Floresco, and Catharine A Winstanley. 2015. "Dopamine Antagonism Decreases Willingness to Expend Physical, but Not Cognitive, Effort: A Comparison of Two Rodent Cost/benefit Decision-Making Tasks." *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* 40 (4): 1005–15.

Houk, James C., Joel L. Davis, and David G. Beiser. 1995. *Models of Information Processing in the Basal Ganglia*. MIT Press.

Howe, Mark W, Patrick L Tierney, Stefan G Sandberg, Paul E M Phillips, and Ann M Graybiel. 2013. "Prolonged Dopamine Signalling in Striatum Signals Proximity and Value of Distant Rewards." *Nature* 500 (7464). Nature Publishing Group: 575–79.

Hull, C. L. 1932. "The Goal-Gradient Hypothesis and Maze Learning."

Humphries, Mark D., Mehdi Khamassi, and Kevin Gurney. 2012. "Dopaminergic Control of the Exploration-Exploitation Trade-off via the Basal Ganglia." *Frontiers in Neuroscience* 6 (FEB): 9.

Ikemoto, S, and J Panksepp. 1999. "The Role of Nucleus Accumbens Dopamine in Motivated Behavior: A Unifying Interpretation with Special Reference to Reward-Seeking." *Brain Research. Brain Research Reviews* 31 (1): 6–41.

Ishiwari, Keita, Suzanne M Weber, Susana Mingote, Mercè Correa, and John D Salamone. 2004. "Accumbens Dopamine and the Regulation of Effort in Food-Seeking Behavior: Modulation of Work Output by Different Ratio or Force

Requirements." *Behavioural Brain Research* 151 (1-2): 83–91.

Juarez, Barbara, and Ming-Hu Han. 2016. "Diversity of Dopaminergic Neural Circuits in Response to Drug Exposure." *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, March.

Kable, Joseph W, and Paul W Glimcher. 2007. "The Neural Correlates of Subjective Value during Intertemporal Choice." *Nature Neuroscience* 10 (12): 1625–33.

Kacelnik, A. 1997. "Normative and Descriptive Models of Decision Making: Time Discounting and Risk Sensitivity." *Ciba Foundation Symposium* 208 (January): 51–67; discussion 67–70.

Karlsson, Mattias P, Dougal G R Tervo, and Alla Y Karpova. 2012. "Network Resets in Medial Prefrontal Cortex Mark the Onset of Behavioral Uncertainty." *Science (New York, N.Y.)* 338 (6103): 135–39.

Kehagia, Angie A, Graham K Murray, and Trevor W Robbins. 2010. "Learning and Cognitive Flexibility: Frontostriatal Function and Monoaminergic Modulation." *Current Opinion in Neurobiology* 20 (2): 199–204.

Kelley, A E, and J M Delfs. 1991. "Dopamine and Conditioned Reinforcement. I. Differential Effects of Amphetamine Microinjections into Striatal Subregions." *Psychopharmacology* 103 (2): 187–96.

Kennerley, Steven W, Timothy E J Behrens, and Jonathan D Wallis. 2011. "Double Dissociation of Value Computations in Orbitofrontal and Anterior Cingulate Neurons." *Nature Neuroscience* 14 (12): 1581–89.

Kennerley, Steven W, Mark E Walton, Timothy E J Behrens, Mark J Buckley, and

Matthew F S Rushworth. 2006. “Optimal Decision Making and the Anterior Cingulate Cortex.” *Nature Neuroscience* 9 (7): 940–47.

Kile, Brian M, Paul L Walsh, Zoé A McElligott, Elizabeth S Bucher, Thomas S Guillot, Ali Salahpour, Marc G Caron, and R Mark Wightman. 2012. “Optimizing the Temporal Resolution of Fast-Scan Cyclic Voltammetry.” *ACS Chemical Neuroscience* 3 (4): 285–92.

Kim, Hoseok, Jung Hoon Sul, Namjung Huh, Daeyeol Lee, and Min Whan Jung. 2009. “Role of Striatum in Updating Values of Chosen Actions.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (47): 14701–12.

Kim, Kyung Man, Michael V Baratta, Aimei Yang, Doheon Lee, Edward S Boyden, and Christopher D Fiorillo. 2012. “Optogenetic Mimicry of the Transient Activation of Dopamine Neurons by Natural Reward Is Sufficient for Operant Reinforcement.” *PloS One* 7 (4): e33612.

Ko, D., and M. J. Wanat. 2016. “Phasic Dopamine Transmission Reflects Initiation Vigor and Exerted Effort in an Action- and Region-Specific Manner.” *Journal of Neuroscience* 36 (7): 2202–11.

Kobayashi, Shunsuke, and Wolfram Schultz. 2008. “Influence of Reward Delays on Responses of Dopamine Neurons.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 28 (31): 7837–46.

Kolling, Nils, Timothy E J Behrens, Rogier B Mars, and Matthew F S Rushworth. 2012. “Neural Mechanisms of Foraging.” *Science (New York, N.Y.)* 336 (6077): 95–98.

Kuroki, T, H Meltzer, and J Ichikawa. 1999. “Effects of Antipsychotic Drugs on Extracellular Dopamine Levels in Rat Medial Prefrontal Cortex and Nucleus Accumbens.” *The Journal of Pharmacology and Experimental Therapeutics* 288 (2): 774–81.

Lammel, Stephan, Andrea Hetzel, Olga Häckel, Ian Jones, Birgit Liss, and Jochen Roeper. 2008. “Unique Properties of Mesoprefrontal Neurons within a Dual Mesocorticolimbic Dopamine System.” *Neuron* 57 (5): 760–73.

Lau, Brian, and Paul W Glimcher. 2005. “Dynamic Response-by-Response Models of Matching Behavior in Rhesus Monkeys.” *Journal of the Experimental Analysis of Behavior* 84 (3): 555–79.

Lawhorn, C, D M Smith, and L L Brown. 2009. “Partial Ablation of Mu-Opioid Receptor Rich Striosomes Produces Deficits on a Motor-Skill Learning Task.” *Neuroscience* 163 (1): 109–19.

Leventhal, Daniel K, Gregory J Gage, Robert Schmidt, Jeffrey R Pettibone, Alaina C Case, and Joshua D Berke. 2012. “Basal Ganglia Beta Oscillations Accompany Cue Utilization.” *Neuron* 73 (3): 523–36.

Leventhal, Daniel K, Colin Stoetzner, Rohit Abraham, Jeff Pettibone, Kayla DeMarco, and Joshua D Berke. 2014. “Dissociable Effects of Dopamine on Learning and Performance within Sensorimotor Striatum.” *Basal Ganglia* 4 (2): 43–54.

Levy, Dino J, and Paul W Glimcher. 2012. “The Root of All Value: A Neural Common Currency for Choice.” *Current Opinion in Neurobiology* 22 (6): 1027–38.

Li, L, and J Shao. 1998. “Restricted Lesions to Ventral Prefrontal Subareas Block

Reversal Learning but Not Visual Discrimination Learning in Rats." *Physiology & Behavior* 65 (2): 371–79.

Li, Yansong, Giovanna Vanni-Mercier, Jean Isnard, François Mauguière, and Jean-Claude Dreher. 2016. "The Neural Dynamics of Reward Value and Risk Coding in the Human Orbitofrontal Cortex." *Brain : A Journal of Neurology* 139 (Pt 4): 1295–1309.

Lloyd, Kevin, and Peter Dayan. 2015. "Tamping Ramping: Algorithmic, Implementational, and Computational Explanations of Phasic Dopamine Signals in the Accumbens." *PLoS Computational Biology* 11 (12): e1004622.

Maia, Tiago V, and Michael J Frank. 2011. "From Reinforcement Learning Models to Psychiatric and Neurological Disorders." *Nature Neuroscience* 14 (2). Nature Publishing Group: 154–62.

Matsuda, Wakoto, Takahiro Furuta, Kouichi C Nakamura, Hiroyuki Hioki, Fumino Fujiyama, Ryohachi Arai, and Takeshi Kaneko. 2009. "Single Nigrostriatal Dopaminergic Neurons Form Widely Spread and Highly Dense Axonal Arborizations in the Neostriatum." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (2): 444–53.

Matsumoto, Masayuki, and Okihide Hikosaka. 2009. "Two Types of Dopamine Neuron Distinctly Convey Positive and Negative Motivational Signals." *Nature* 459 (7248). Nature Publishing Group: 837–41.

Mazur, J E. 1986. "Fixed and Variable Ratios and Delays: Further Tests of an Equivalence Rule." *Journal of Experimental Psychology. Animal Behavior Processes* 12 (2): 116–24.

McClure, Samuel M, Nathaniel D Daw, and P Read Montague. 2003. “A Computational Substrate for Incentive Salience.” *Trends in Neurosciences* 26 (8): 423–28.

Mingote, Susana, Laura Font, Andrew M Farrar, Regina Vontell, Lila T Worden, Colin M Stopper, Russell G Port, et al. 2008. “Nucleus Accumbens Adenosine A2A Receptors Regulate Exertion of Effort by Acting on the Ventral Striatopallidal Pathway.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 28 (36): 9037–46.

Montague, P R, P Dayan, and T J Sejnowski. 1996. “A Framework for Mesencephalic Dopamine Systems Based on Predictive Hebbian Learning.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 16 (5): 1936–47.

Morita, Kenji, and Ayaka Kato. 2014. “Striatal Dopamine Ramping May Indicate Flexible Reinforcement Learning with Forgetting in the Cortico-Basal Ganglia Circuits.” *Frontiers in Neural Circuits* 8 (April): 36.

Morris, Genela, Alon Nevet, David Arkadir, Eilon Vaadia, and Hagai Bergman. 2006. “Midbrain Dopamine Neurons Encode Decisions for Future Action.” *Nature Neuroscience* 9 (8): 1057–63.

Mott, Allison M, Eric J Nunes, Lyndsey E Collins, Russell G Port, Kelly S Sink, Jörg Hockemeyer, Christa E Müller, and John D Salamone. 2009. “The Adenosine A2A Antagonist MSX-3 Reverses the Effects of the Dopamine Antagonist Haloperidol on Effort-Related Decision Making in a T-Maze Cost/benefit Procedure.” *Psychopharmacology* 204 (1): 103–12.

Nagano-Saito, Atsuko, Paul Cisek, Andrea S Perna, Fatemeh Z Shirdel, Chawki Benkelfat, Marco Leyton, and Alain Dagher. 2012. “From Anticipation to Action,

the Role of Dopamine in Perceptual Decision Making: An fMRI-Tyrosine Depletion Study." *Journal of Neurophysiology* 108 (2): 501–12.

Naneix, F., A. R. Marchand, G. D. Scala, J.-R. Pape, and E. Coutureau. 2009. "A Role for Medial Prefrontal Dopaminergic Innervation in Instrumental Conditioning." *Journal of Neuroscience* 29 (20): 6599–6606.

Nicola, Saleem M. 2010. "The Flexible Approach Hypothesis: Unification of Effort and Cue-Responding Hypotheses for the Role of Nucleus Accumbens Dopamine in the Activation of Reward-Seeking Behavior." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 30 (49): 16585–600.

Nieoullon, A., A. Cheramy, and J. Glowinski. 1978. "Release of Dopamine Evoked by Electrical Stimulation of the Motor and Visual Areas of the Cerebral Cortex in Both Caudate Nuclei and in the Substantia Nigra in the Cat." *Brain Research* 145 (1): 69–83.

Nirenberg, M J, J Chan, A Pohorille, R A Vaughan, G R Uhl, M J Kuhar, and V M Pickel. 1997. "The Dopamine Transporter: Comparative Ultrastructure of Dopaminergic Axons in Limbic and Motor Compartments of the Nucleus Accumbens." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 17 (18): 6899–6907.

Niv, Yael. 2013. "Neuroscience: Dopamine Ramps Up." *Nature* 500 (7464): 533–35.

Niv, Yael, Nathaniel D Daw, and Peter Dayan. 2006. "How Fast to Work : Response Vigor , Motivation and Tonic Dopamine."

Niv, Yael, Nathaniel D Daw, Daphna Joel, and Peter Dayan. 2007. "Tonic Dopamine:

Opportunity Costs and the Control of Response Vigor." *Psychopharmacology* 191 (3): 507–20.

Niyogi, Ritwik K, Yannick-Andre Breton, Rebecca B Solomon, Kent Conover, Peter Shizgal, and Peter Dayan. 2014. "Optimal Indolence: A Normative Microscopic Approach to Work and Leisure." *Journal of the Royal Society, Interface / the Royal Society* 11 (91): 20130969.

Oyama, Kei, István Hernádi, Toshio Iijima, and Ken-Ichiro Tsutsui. 2010. "Reward Prediction Error Coding in Dorsal Striatal Neurons." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 30 (34): 11447–57.

Pennartz, Cyriel M a, Joshua D Berke, Ann M Graybiel, Rutsuko Ito, Carien S Lansink, Matthijs van der Meer, a David Redish, Kyle S Smith, and Pieter Voorn. 2009. "Corticostriatal Interactions during Learning, Memory Processing, and Decision Making." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (41): 12831–38.

Phillips, Paul E. M., Garret D. Stuber, Michael L. A. V. Heien, R. Mark Wightman, and Regina M. Carelli. 2003. "Subsecond Dopamine Release Promotes Cocaine Seeking." *Nature* 422 (6932): 614–18.

Ragozzino, Michael E. 2003. "Acetylcholine Actions in the Dorsomedial Striatum Support the Flexible Shifting of Response Patterns." *Neurobiology of Learning and Memory* 80 (3): 257–67.

Rapoport, J L, M S Buchsbaum, H Weingartner, T P Zahn, C Ludlow, and E J Mikkelsen. 1980. "Dextroamphetamine. Its Cognitive and Behavioral Effects in Normal and Hyperactive Boys and Normal Men." *Archives of General Psychiatry* 37 (8): 933–43.

Reynolds, John N J, Brian I Hyland, and Jeffery R Wickens. 2001. “A Cellular Mechanism of Reward-Related Learning,” 67–70.

Reynolds, Sheila M, and Kent C Berridge. 2003. “Glutamate Motivational Ensembles in Nucleus Accumbens: Rostrocaudal Shell Gradients of Fear and Feeding.” *The European Journal of Neuroscience* 17 (10): 2187–2200.

Rice, M E, J C Patel, and S J Cragg. 2011. “Dopamine Release in the Basal Ganglia.” *Neuroscience* 198 (December): 112–37.

Robbins, T W, M Cador, J R Taylor, and B J Everitt. 1989. “Limbic-Striatal Interactions in Reward-Related Processes.” *Neuroscience and Biobehavioral Reviews* 13 (2-3): 155–62.

Roitman, Mitchell F, Garret D Stuber, Paul E M Phillips, R Mark Wightman, and Regina M Carelli. 2004. “Dopamine Operates as a Subsecond Modulator of Food Seeking.” *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 24 (6): 1265–71.

Rudebeck, Peter H, Mark E Walton, Angharad N Smyth, David M Bannerman, and Matthew F S Rushworth. 2006. “Separate Neural Pathways Process Different Decision Costs.” *Nature Neuroscience* 9 (9). Nature Publishing Group: 1161–68.

Rushworth, Matthew F S, Mark J Buckley, Timothy E J Behrens, Mark E Walton, and David M Bannerman. 2007. “Functional Organization of the Medial Frontal Cortex.” *Current Opinion in Neurobiology* 17 (2): 220–27.

Saddoris, M., F. Cacciapaglia, R. M. Wightman, and R. M. Carelli. 2015. “Differential Dopamine Release Dynamics in the Nucleus Accumbens Core and Shell Reveal

Complementary Signals for Error Prediction and Incentive Motivation." *Journal of Neuroscience* 35 (33): 11572–82.

Saka, Esen, and Ann M Graybiel. 2003. "Pathophysiology of Tourette's Syndrome: Striatal Pathways Revisited." *Brain & Development* 25 Suppl 1 (December): S15–19.

Salamone, J D, M Correa, A Farrar, and S M Mingote. 2007. "Effort-Related Functions of Nucleus Accumbens Dopamine and Associated Forebrain Circuits." *Psychopharmacology* 191 (3): 461–82.

Salamone, J D, M S Cousins, and B J Snyder. 1997. "Behavioral Functions of Nucleus Accumbens Dopamine: Empirical and Conceptual Problems with the Anhedonia Hypothesis." *Neuroscience and Biobehavioral Reviews* 21 (3): 341–59.

Salamone, J D, A Wisniecki, B B Carlson, and M Correa. 2001. "Nucleus Accumbens Dopamine Depletions Make Animals Highly Sensitive to High Fixed Ratio Requirements but Do Not Impair Primary Food Reinforcement." *Neuroscience* 105 (4): 863–70.

Salamone, John D, and Mercè Correa. 2012. "The Mysterious Motivational Functions of Mesolimbic Dopamine." *Neuron* 76 (3): 470–85.

Salamone, John D, Merce Correa, Samantha Yohn, Laura Lopez Cruz, Noemi San Miguel, and Luisa Alatorre. 2016. "The Pharmacology of Effort-Related Choice Behavior: Dopamine, Depression, and Individual Differences." *Behavioural Processes* 127 (February): 3–17.

Samejima, Kazuyuki, Yasumasa Ueda, and Kenji Doya. 2005. "Representation of Action-Specific Reward Values in the Striatum" 310 (November): 1337–41.

Satoh, Takemasa, Sadamu Nakai, Tatsuo Sato, and Minoru Kimura. 2003. “Correlated Coding of Motivation and Outcome of Decision by Dopamine Neurons” 23 (30): 9913–23.

Schlinger, Henry D, Adam Derenne, and Alan Baron. 2008. “What 50 Years of Research Tell Us about Pausing under Ratio Schedules of Reinforcement.” *The Behavior Analyst / MABA* 31 (1): 39–60.

Schmidt, Robert, Daniel K Leventhal, Nicolas Mallet, Fujun Chen, and Joshua D Berke. 2013. “Canceling Actions Involves a Race between Basal Ganglia Pathways.” *Nature Neuroscience* 16 (8): 1118–24.

Schultz, W, P Dayan, and P R Montague. 1997. “A Neural Substrate of Prediction and Reward.” *Science (New York, N.Y.)* 275 (5306): 1593–99.

Schultz, Wolfram. 2016. “Dopamine Reward Prediction-Error Signalling: A Two-Component Response.” *Nature Reviews. Neuroscience* 17 (3): 183–95.

Schultz, Wolfram, Regina M Carelli, and R Mark Wightman. 2015. “Phasic Dopamine Signals: From Subjective Reward Value to Formal Economic Utility.” *Current Opinion in Behavioral Sciences* 5 (October): 147–54.

Schultz, Wolfram, Peter Dayan, and P Read Montague. 1997. “A Neural Substrate of Prediction and Reward” 275 (June 1994): 1593–1600.

Seger, Carol A. 2009. *The Basal Ganglia IX*. Springer Science & Business Media.

Sesack, S R, A Y Deutch, R H Roth, and B S Bunney. 1989. “Topographical Organization of the Efferent Projections of the Medial Prefrontal Cortex in the Rat: An Anterograde Tract-Tracing Study with Phaseolus Vulgaris Leucoagglutinin.”

The Journal of Comparative Neurology 290 (2): 213–42.

Shenhav, Amitai, Matthew M Botvinick, and Jonathan D Cohen. 2013. "The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function." *Neuron* 79 (2): 217–40.

Simen, Patrick, Jonathan D. Cohen, and Philip Holmes. 2006. "Rapid Decision Threshold Modulation by Reward Rate in a Neural Network." *Neural Networks* 19: 1013–26.

Song, Peng, Neil D Hershey, Omar S Mabrouk, Thomas R Slaney, and Robert T Kennedy. 2012. "Mass Spectrometry 'sensor' for in Vivo Acetylcholine Monitoring." *Analytical Chemistry* 84 (11): 4659–64.

Song, Peng, Omar S. Mabrouk, Neil D. Hershey, and Robert T. Kennedy. 2012. "In Vivo Neurochemical Monitoring Using Benzoyl Chloride Derivatization and Liquid Chromatography-Mass Spectrometry." *Analytical Chemistry* 84 (1): 412–19.

St Onge, Jennifer R, Soyon Ahn, Anthony G Phillips, and Stan B Floresco. 2012. "Dynamic Fluctuations in Dopamine Efflux in the Prefrontal Cortex and Nucleus Accumbens during Risk-Based Decision Making." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 32 (47): 16880–91.

Stauffer, William R, Armin Lak, Shunsuke Kobayashi, and Wolfram Schultz. 2016. "Components and Characteristics of the Dopamine Reward Utility Signal." *The Journal of Comparative Neurology* 524 (8): 1699–1711.

Steinberg, Elizabeth E, Ronald Keiflin, Josiah R Boivin, Ilana B Witten, Karl Deisseroth, and Patricia H Janak. 2013. "A Causal Link between Prediction Errors, Dopamine Neurons and Learning." *Nature Neuroscience* 16 (7). Nature Publishing Group:

966–73.

Stephens, David W. 1986. *Foraging Theory*. Princeton University Press.

Strait, Caleb E., Brianna J. Sleezer, and Benjamin Y. Hayden. 2015. "Signatures of Value Comparison in Ventral Striatum Neurons." Edited by Matthew F. S. Rushworth. *PLOS Biology* 13 (6). Public Library of Science: e1002173.

Stuber, Garret D, Mitchell F Roitman, Paul E M Phillips, Regina M Carelli, and R Mark Wightman. 2005. "Rapid Dopamine Signaling in the Nucleus Accumbens during Contingent and Noncontingent Cocaine Administration." *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* 30 (5): 853–63.

Sugrue, Leo P, Greg S Corrado, and William T Newsome. 2004. "Matching Behavior and the Representation of Value in the Parietal Cortex." *Science (New York, N.Y.)* 304 (5678): 1782–87.

Sutton, Richard S., and Andrew G. Barto. 1998. *Reinforcement Learning: An Introduction*. MIT Press.

Tachibana, Yoshihisa, and Okihide Hikosaka. 2012. "The Primate Ventral Pallidum Encodes Expected Reward Value and Regulates Motor Action." *Neuron* 76 (4). Elsevier Inc.: 826–37.

Tai, Lung-Hao, A Moses Lee, Nora Benavidez, Antonello Bonci, and Linda Wilbrecht. 2012. "Transient Stimulation of Distinct Subpopulations of Striatal Neurons Mimics Changes in Action Value." *Nature Neuroscience* 15 (9): 1281–89.

Thorndike, Edward Lee. 1911. *Animal Intelligence: Experimental Studies*. Macmillan.

Threlfell, Sarah, Tatjana Lalic, Nicola J Platt, Katie A Jennings, Karl Deisseroth, and Stephanie J Cragg. 2012. "Striatal Dopamine Release Is Triggered by Synchronized Activity in Cholinergic Interneurons." *Neuron* 75 (1): 58–64.

Tsai, C T. "Involvement of Serotonin in Mediation of Inhibition of Substantia Nigra Neurons by Noxious Stimuli." *Brain Research Bulletin* 23 (1-2): 121–27.

Turner, Robert S, and Michel Desmurget. 2010. "Basal Ganglia Contributions to Motor Control: A Vigorous Tutor." *Current Opinion in Neurobiology* 20 (6): 704–16.

van der Meer, Matthijs A A, and A David Redish. 2011. "Ventral Striatum: A Critical Look at Models of Learning and Evaluation." *Current Opinion in Neurobiology* 21 (3): 387–92.

Van Der Meulen, J. a J, R. N J M a Joosten, J. P C De Bruin, and M. G P Feenstra. 2007. "Dopamine and Noradrenaline Efflux in the Medial Prefrontal Cortex during Serial Reversals and Extinction of Instrumental Goal-Directed Behavior." *Cerebral Cortex* 17 (6): 1444–53.

Vander Weele, Caitlin M, Kirsten a Porter-Stransky, Omar S Mabrouk, Vedran Lovic, Bryan F Singer, Robert T Kennedy, and Brandon J Aragona. 2014. "Rapid Dopamine Transmission within the Nucleus Accumbens: Dramatic Difference between Morphine and Oxycodone Delivery." *The European Journal of Neuroscience* 40 (7): 3041–54.

Venton, B Jill, Kevin P Troyer, and R Mark Wightman. 2002. "Response Times of Carbon Fiber Microelectrodes to Dynamic Changes in Catecholamine Concentration." *Analytical Chemistry* 74 (3): 539–46.

Voorn, Pieter, Louk J M J Vanderschuren, Henk J. Groenewegen, Trevor W. Robbins, and Cyriel M a Pennartz. 2004. "Putting a Spin on the Dorsal-Ventral Divide of the Striatum." *Trends in Neurosciences* 27 (8): 468–74.

Walton, Mark E, Timothy E J Behrens, MaryAnn P Noonan, and Matthew F S Rushworth. 2011. "Giving Credit Where Credit Is Due: Orbitofrontal Cortex and Valuation in an Uncertain World." *Annals of the New York Academy of Sciences* 1239 (December): 14–24.

Walton, Mark E, Paula L Croxson, Timothy E J Behrens, Steven W Kennerley, and Matthew F S Rushworth. 2007. "Adaptive Decision Making and Value in the Anterior Cingulate Cortex." *NeuroImage* 36 Suppl 2 (January): T142–54.

Wang, Alice Y, Keiji Miura, and Naoshige Uchida. 2013. "The Dorsomedial Striatum Encodes Net Expected Return, Critical for Energizing Performance Vigor." *Nature Neuroscience* 16 (5). Nature Publishing Group: 639–47.

Wardle, Margaret C, Michael T Treadway, Leah M Mayo, David H Zald, and Harriet de Wit. 2011. "Amping up Effort: Effects of D-Amphetamine on Human Effort-Based Decision-Making." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 31 (46): 16597–602.

White, N M, and N Hiroi. 1998. "Preferential Localization of Self-Stimulation Sites in Striosomes/patches in the Rat Striatum." *Proceedings of the National Academy of Sciences of the United States of America* 95 (11): 6486–91.

Witten, Ilana B, Elizabeth E Steinberg, Soo Yeun Lee, Thomas J Davidson, Kelly A Zalocusky, Matthew Brodsky, Ofer Yizhar, et al. 2011. "Recombinase-Driver Rat Lines: Tools, Techniques, and Optogenetic Application to Dopamine-Mediated

Reinforcement." *Neuron* 72 (5): 721–33.

Worden, Lila T, Mona Shahriari, Andrew M Farrar, Kelly S Sink, Jörg Hockemeyer, Christa E Müller, and John D Salamone. 2009. "The Adenosine A2A Antagonist MSX-3 Reverses the Effort-Related Effects of Dopamine Blockade: Differential Interaction with D1 and D2 Family Antagonists." *Psychopharmacology* 203 (3): 489–99.

Wyvell, C L, and K C Berridge. 2000. "Intra-Accumbens Amphetamine Increases the Conditioned Incentive Salience of Sucrose Reward: Enhancement of Reward 'wanting' without Enhanced 'liking' or Response Reinforcement." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 20 (21): 8122–30.