PREFERENCE ENCODING IN VENTRAL PALLIDUM MEDIATES REWARD-SEEKING BEHAVIOR

by
David Joshua Ottenheimer

A dissertation submitted to the Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March 2020

Abstract

An essential function of the nervous system is to direct reward-seeking behavior in order to maximize the acquisition of preferred rewards. This process requires a method for evaluating all available outcomes on a common scale and using these valuations to organize the appropriate behavioral response. The ventral pallidum (VP) is a key node in a basal ganglia circuit hypothesized to convert limbic information, like reward values, into reward-seeking actions. Previous work has linked VP neural activity to the availability and palatability of rewards, and VP has been functionally implicated in the motivation to pursue rewards. Open questions include how VP encodes the values of multiple available rewards and whether its activity contributes to preference-driven behaviors. For this dissertation, we conducted a series of electrophysiological and optogenetic experiments to characterize the role of VP in navigating scenarios with multiple rewarding outcomes. First, we demonstrated that, following reward delivery, the activity of a majority VP neurons reflected the value of the delivered outcome relative to the locally available options; notably, this activity preceded and outnumbered reward-specific activity in nucleus accumbens, the most frequently studied input to VP. Further analysis of VP activity revealed that, consistent with a reward prediction error signal, a subset of neurons' reward-evoked activity incorporated the outcomes from the most recent previous trials. The prediction error hypothesis was further supported by optogenetic manipulations of VP activity during this epoch, which altered rats' engagement in the reward-seeking task according to changes in their estimate of the task's value. In a final set of experiments, we linked VP neural activity to the evolution of rats' choice behavior under changing physiological conditions and demonstrated a causal role for VP outcome

signals in driving behavioral preference. Our results not only establish VP as a crucial site for encoding reward preferences; they also provide insight into fundamental principles of reward signaling in the nervous system, with particular consideration for the interface between prediction errors and preference, both static and dynamic.

Thesis advisor: Patricia H. Janak, Ph.D.

Reader: Marshall G. Hussain Shuler, Ph.D.

Acknowledgments

What happens during the Ph.D. years to transform a curious budding scientist into a full-blown Doctor of Philosophy? Probably a lot of different things to different people. But I can reflect a little on my journey and the people who made it possible.

The first time I felt like a scientist was during my very first rotation in grad school, which happened to be with the lab I joined, the Janak lab. Surprisingly, it wasn't conducting experiments that helped me find confidence in my professional identity; it was writing my NSF GRFP proposal. This two page grant was the first time I had ever been asked to come up with a scientific question. When I met with Tricia to talk about the proposal, I came in with a rough idea, but I had little confidence in my ability to identify a good project. In fact, I fully expected that during the course of the meeting she'd steer me in a more certain direction–maybe give me a couple of options and some associated papers to read. Instead, she told me my idea sounded good, and I should give writing it a shot. I left our (pretty brief) meeting thinking, "Okay, here goes nothing." A few weeks later, I had completed a draft of my proposal. And it was pretty good! The best part was, I had read the literature, identified a question, and designed experiments almost entirely on my own. I realized I enjoyed the process, and I started to believe I could be a scientist.

Working with Tricia, I was able to develop confidence in my abilities to think creatively and critically and to conduct rigorous science. And it was not just because Tricia granted me the right balance of guidance and independence; it was also because she demonstrated faith that I could succeed even before I saw it for myself. I couldn't ask for a better mentor for myself, nor could I ask for a better role model. Tricia exemplifies kindness, community,

scholarship, honesty, and generosity in a field that doesn't always hold these values in highest esteem. It's been an honor working with her.

Studying ventral pallidum was actually one big serendipitous mistake. I originally planned to start recording in nucleus accumbens and then add in medial prefrontal cortex. After somewhat disappointing results from initial accumbens recordings, I finally started seeing robust reward-specific neural responses in one of the rats from my third cohort. I remember thinking, "I finally found the accumbens neurons that encode reward preference!" Then, I checked the placement of the electrodes and saw that I was NOT recording from that rat's accumbens. My best guess was that I was in ventral pallidum, so I sheepishly showed the histology to our lab's resident VP expert, Jocelyn Richard. "Yup, that looks like VP," she said. Since the data were interesting, I followed it up with another rat, this time aiming for VP, and the rest is history.

Since that moment, Jocelyn played an instrumental (or is it Pavlovian?) role in shaping my thesis work. It was an incredible opportunity to work alongside an expert in the (pretty tiny) VP field and to learn firsthand from a very talented electrophysiologist and behavioralist. I couldn't be happier with the surprising way things turned out.

My thesis committee–Marshall Hussain Shuler, Jeremiah Cohen, and Michela Gallagher–played a very important role in steering and motivating me. Some of my most successful experiment ideas came out of conversations during those meetings. Perhaps most importantly, I could tell they set high expectations for me, and that knowledge pushed me to hold myself to a high standard of rigor and creativity. I'll still be thinking "What would Marshall/Jeremiah/Michela say about this experiment?" for many years to come.

Marshall played another important role in my Ph.D., along with Kristina Nielsen, as co-director of our program's Systems Journal Club. Our long discussions, despite being perhaps too philosophical sometimes, and too nit-picky at others, immersed me in a world of critical thought that has been immensely useful when evaluating my own work as much as others'. Their curation of journal club is a major reason why Hopkins is one of the best places to do

a Ph.D. in systems neuroscience.

Science is a team sport, and my thesis would not be the same without the following people. First, my lab. I was lucky to have many role models in the lab; in addition to Tricia and Jocelyn, I had the pleasure of working with Ben Saunders, Ron Keiflin, Zayra Millan, and Youna Vandaele. And then there were my fellow grad students, Kurt Fraser and Tabitha Kim (and recent addition Emma Chaloux-Pinette), with whom I've sat in the same office essentially every day for four years. We shared a lot of fun memories as we developed along our own Ph.D. trajectories, including running 5Ks, taking holiday photos, and herding an escape artist rat with a broom. And we got to collaborate a little bit, too. I've had the amazing opportunity to mentor rotation and undergraduate students, as well, including Elissa Sutlief, Karen Wang, Xiao Tong, and Yasmin Padovan-Hernandez, who all contributed substantially to the work included here. Beyond lab, I benefited immensely from conversations and collaborations with too many people to list, but I'd especially like to acknowledge Bilal Bari (my co-author for the project in Chapter 3), Raina D'Aleo, and Cooper Grossman.

The Janak lab is part of both the Psychology and Neuroscience departments, and it's been a lot of fun getting to know both communities very well. I got the best of both worlds with opportunities for talks to attend and give and amazing students to meet and befriend. I gave a lot of my energy to the Neuroscience department over the years, serving on many committees. It was an honor to be able to make an impact on a community I gained so much from. Special thank you to Rita Ragan and Beth Wood-Roig for helping make it such a friendly environment and helping us graduate students clear the hurdles along the way.

One of the joys of graduate school was living near some of my favorite people: my aunt, uncle, and cousins. Getting to be close by for birthdays, holidays, weddings, babies, and everything in between (more 5Ks!) was wonderful, and I'm so grateful for all the happy times we shared. Special shout out to my cousin Sarah Gebauer, who also has a Ph.D. She provided me with one of my first research experiences 9 years ago and has imparted vital

wisdom ever since.

As I write this, I am sitting in my parents' house, just a few feet from where I did most of my homework growing up. I am so lucky to have parents who have supported me every step of the way, in both my academic and non-academic pursuits. I'm proud to say that they are two of the world's top ventral pallidum experts thanks to listening to my practice talks and asking me about my work. They've gone above and beyond to be an active part of this period of my life, and I couldn't be more grateful. And of course I need to thank Luna, the best and most beautiful dog in the world. Having a therapy dog for a sister sure came in handy while writing this dissertation.

The one person who has been by my side for this whole journey is Lionel Rodriguez. Applying to graduate school, moving to Baltimore, attending conferences around the country (plus Portugal), and becoming part of the global neuroscience community are adventures we undertook together. From our conversations about science to the not-at-all-science-related times we've shared; it all had an immeasurable impact on my work. I can't wait to see everything he accomplishes in his Ph.D. and beyond.

Contents

List of Figures

Chapter 1

Introduction

How do we choose which rewards to pursue? The evaluation of outcomes in the environment is critical for adaptive behavioral responding, but the incredible complexity of the reward space makes this a difficult task. This point can be illustrated by the conundrum of deciding which restaurant to go to for dinner. Often, hungry individuals must choose between dozens or hundreds of options that vary on many dimensions, such as food served, distance away, price, and probability of being open. With so many factors to consider, it becomes clear that assigning a single value to each option is not trivial. Unsurprisingly, given this complexity, numerous brain regions have been implicated in reward value processing (Louie and Glimcher, 2012; Schultz, 2015). How value is represented in the nervous system and how these representations instruct individual's choices are areas of active investigation.

For this dissertation, we explored the role of an understudied basal ganglia nucleus, the ventral pallidum (VP), in preference-based value signaling and reward-seeking behavior. VP has been linked with reward-seeking behavior (Smith et al., 2009; Root et al., 2015), especially within the context of a major input, the nucleus accumbens (NAc); however, despite many functional demonstrations of a role for VP in reward processing, *in vivo* observations of VP neural activity during reward-seeking tasks are lacking. Throughout the course of the experiments presented here, we characterize a prominent relative value signal in VP that tracks shifting preference across different task and physiological conditions. We also demonstrate that VP activity contributes to value-based behavioral responding. To fully

appreciate these results, we first summarize the historical context of research on VP and on value signaling, broadly:

1.1 The ventral striatopallidal system.

The ventral pallidum (VP), given its name by Heimer and Wilson (1975), was first identified anatomically as a major site of efferent connections from the nucleus accumbens (NAc) (Heimer and Wilson, 1975; Nauta et al., 1978). The prevailing theory for NAc function was that it integrated limbic information from inputs like amygdala, hippocampus, ventral tegmental area (VTA), and frontal cortex and linked it with motor output (Mogenson et al., 1980). VP quickly became a likely candidate to relay these limbic-related motor outputs from NAc as many studies identified functional connections from hippocampus, amygdala, and dopamine inputs, through NAc, and onto VP (and from there onto mediodorsal thalamus), largely using locomotion and (anesthetized) electrophysiological recordings as readouts (Jones and Mogenson, 1980; Yim and Mogenson, 1983; Mogenson and Nielsen, 1984; Yang and Mogenson, 1985; Swerdlow and Koob, 1987; Root et al., 2015; Smith et al., 2009). Thus, NAc and VP, collectively known as the ventral striatopallidal system, were canonized as the most ventral portion of the highly parallel basal ganglia circuits, specializing in limbic processing (due to their connectivity with 'limbic' regions) (Alexander et al., 1986; de Olmos and Heimer, 1999; Groenewegen et al., 1999).

A growing number of studies have probed the role of this circuit in reward-related behaviors by manipulating the connectivity between these two regions. For instance, normal connectivity between NAc and VP was necessary for cues paired with reward delivery to invigorate reward-seeking actions in a Pavlovian instrumental transfer protocol (Leung and Balleine, 2013); similarly, inhibition of VP-projecting NAc cells reduced attraction to reward-predicting cues (Smedley et al., 2019). On the other hand, disconnection of NAc and VP enhanced the attribution of motivational salience to a reward-predicting cue (Chang et al., 2018), demonstrating that connections between NAc and VP have important but varying

roles in the valuation of and responding to reward-related stimuli in different tasks. There is also evidence for the importance of NAc-VP connectivity in reward consumption. Normalizing plasticity between NAc D2 MSNs and their downstream targets in VP reversed deficits in hedonic responses and motivation to work for natural reward following cocaine exposure (Creed et al., 2016); accordingly, pharmacological inhibition of NAc D2 MSN terminals in VP increased motivation to work for food reward (Gallo et al., 2018). Additionally, NAc and VP contain reciprocally connected μ-opioid-agonist-responsive hotspots that readily alter rats' reward intake and expression of pleasure (Smith and Berridge, 2007; Richard et al., 2013a). This collection of findings is consistent with the notion that VP is a crucial downstream mediator of NAc reward-related functions.

One area of inquiry that is lacking from studies of NAc and VP is *in vivo* recordings of activity in each region during tasks that require integration of reward information and execution of reward-related behavior. More precisely, how and when reward-related information is represented by neurons in each region, and when connectivity between NAc and VP is necessary for proper reward processing, remain unclear. In fact, a number of existing *in vivo* recording studies challenge the idea that VP simply inherits limbic information from NAc. In an instrumental task with cues indicating reward availability, cue-evoked excitations in VP individual neurons frequently preceded cue-evoked activity in NAc neurons, indicating that NAc cannot be the sole source of these VP responses (Richard et al., 2016). Similar patterns emerged when comparing VP to other striatal regions–the rostromedial caudate (Fujimoto et al., 2019) and the internal-capsule-bordering dorsal striatum (White et al., 2019). Additional studies of neural activity in behaving animals are required to clarify the contributions of NAc and VP to specific features of reward processing. In this dissertation, we examine the activity of neurons in this circuit in tasks with differentially preferred rewards and, in turn, how this activity relates to reward-seeking actions.

Figure 1.1

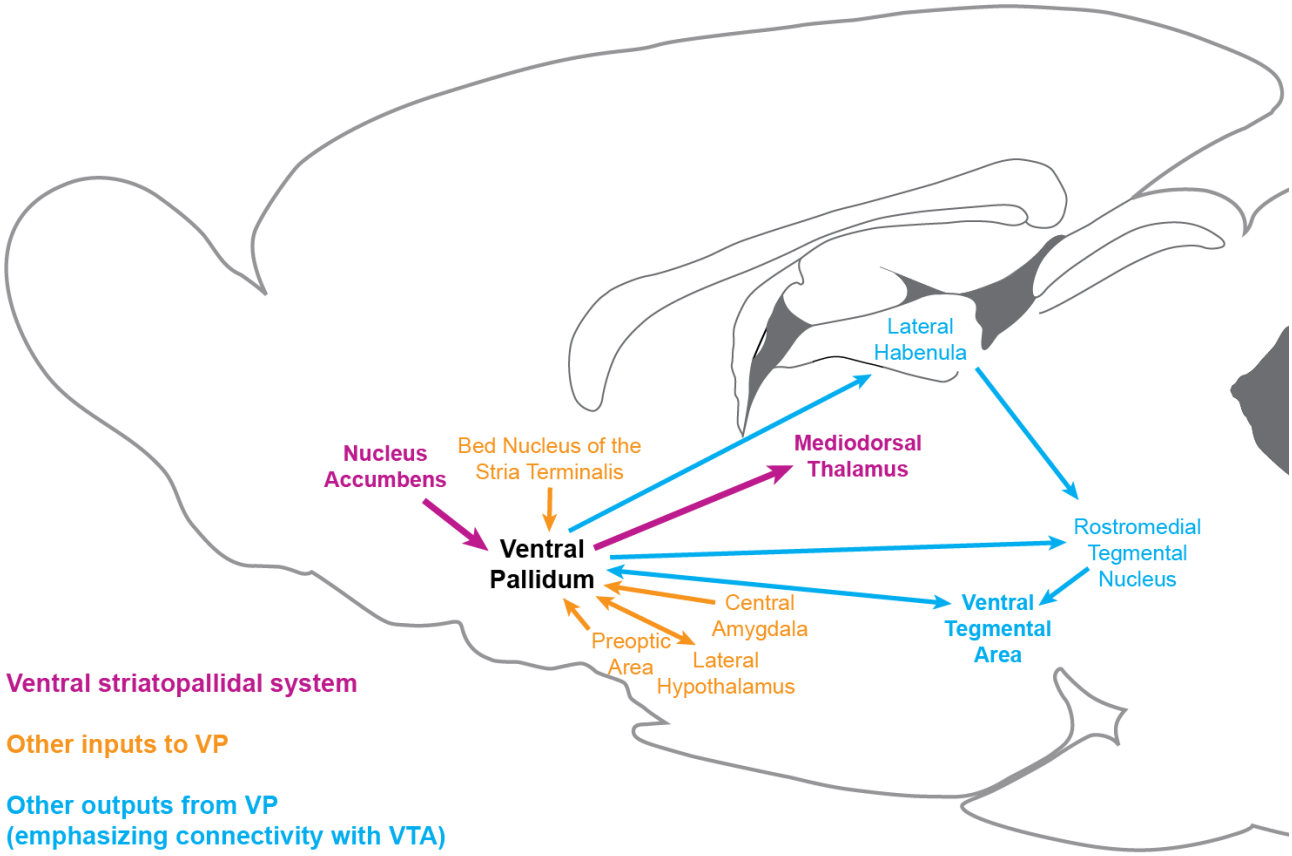

Figure 1.1. VP inputs and outputs.

1.2 Evolving views of ventral pallidum.

Since the recognition of VP as a distinct region, a considerable amount of research has been conducted to characterize VP function, extending beyond its connectivity with NAc (Smith et al., 2009; Root et al., 2015). In recent years, there has been increasing emphasis on identifying inputs to and outputs from VP beyond the classic ventral striatopallidothalamic pathway (Fig. 1.1). Notable non-NAc inputs to VP neurons include central amygdala, the bed nucleus of the stria terminalis, preoptic area, lateral hypothalamus, and VTA (Hnasko et al., 2012; Knowland et al., 2017; Tooley et al., 2018; Stephenson-Jones et al., 2020). Mean-

while, work on VP outputs has focused on lateral habenula (Knowland et al., 2017; Tooley et al., 2018; Faget et al., 2018; Stephenson-Jones et al., 2020), lateral hypothalamus (Prasad et al., 2020), rostromedial tegmenetal nucleus (Tooley et al., 2018), and VTA (Mahler et al., 2014; Knowland et al., 2017; Tooley et al., 2018; Faget et al., 2018), including direct innervation of dopamine neurons (Tian et al., 2016). These findings encourage reinterpretation of some initial findings on VP (Smith et al., 2009; Root et al., 2015) within the context of a broader circuit, with special focus on VP contributions to reward value processing given the known functions of the newly emphasized interconnected regions (Schultz, 2015; Keiflin and Janak, 2015).

Another layer of heterogeneity has been added to the understanding of VP anatomy by characterizing different cell types in ventral pallidum. Unlike other pallidal structures, VP contains a small proportion of glutamatergic cells in addition to GABAergic cells, and these populations are mostly non-overlapping (Knowland et al., 2017; Tooley et al., 2018; Faget et al., 2018). Interestingly, there is a strong divergence between the effects of manipulations of VP GABAergic cells and glutamatergic cells on behavior, with GABAergic activity generally linked to appetitive behaviors and positive reinforcement whereas glutamatergic activity is linked to avoidance (Tooley et al., 2018; Faget et al., 2018; Stephenson-Jones et al., 2020; Prasad et al., 2020). Other notable cell types include cholinergic and parvalbumin-positive neurons, which contribute to depressive-like phenotypes (Tooley et al., 2018; Faget et al., 2018). The fact that subsets of VP neurons mediate distinct behaviors has garnered increasing interest in VP as a site for bidirectional control of motivation and learning; these findings also support the theory that VP could serve as an interface between value signaling and reward-seeking behavior, a concept we explore in this dissertation.

1.3 Preference and relative value.

How are neural representations of value measured? One of the most straightforward methods of studying neural correlates of reward value is to present individuals with rewards whose

values differ along one, easily interpretable dimension and look for neurons whose activity scales accordingly. This approach has been exploited by many studies that have provided fundamental insight into value coding in VP and closely related regions, including NAc and VTA. One of the most common dimensions to manipulate is the size of the reward. In NAc (Bissonette et al., 2013; Cooch et al., 2015; Goldstein et al., 2012; Roesch et al., 2009; Webber et al., 2016) and in VP (Tachibana and Hikosaka, 2012; Avila and Lin, 2014a,a; Fujimoto et al., 2019; Stephenson-Jones et al., 2020), neural responses evoked by rewards and reward-predicting cues reflect the differences in magnitude of the outcomes. This approach has also been useful for characterizing the firing of dopamine neurons in the VTA and their adherence to a reward prediction error-style of encoding (Takahashi et al., 2011; Roesch et al., 2007; Tobler et al., 2005; Cohen et al., 2012; Eshel et al., 2015). Another approach is to vary the concentration of sucrose solution, which affects the palatability (and caloric content) of the outcomes without affecting the volume; this approach has revealed neurons in NAc whose activity scales with palatability (Taha and Fields, 2005; Wheeler et al., 2005; Villavicencio et al., 2018).

In addition to these manipulations of aspects of the primary reward, many studies have examined how the conditions for obtaining the reward impact signaling in this circuit. In particular, cue- and reward-evoked activity of dopamine neurons (or dopamine release) has been examined when rewards are delivered after varying delays (Roesch et al., 2007; Kobayashi and Schultz, 2008; Day et al., 2010; Takahashi et al., 2011, 2016), requiring different amounts of effort (Day et al., 2010; Gan et al., 2010), and with varying probabilities of being delivered (Fiorillo et al., 2003; Nakahara et al., 2004; Sugam et al., 2012; Lak et al., 2014; Eshel et al., 2015; Tian et al., 2016). To some extent, the impact of these reward features on firing in NAc has also been explored (Roesch et al., 2009; Day et al., 2011; Tian et al., 2016), but only minimally in VP (Tian et al., 2016). One theme that emerges is that there are cells across the entire circuit that are sensitive to each of these manipulations.

Of the above results, some of the most compelling data on value signaling come from

studies that vary multiple reward features in the same task. From these, a common observation is that the firing of many neurons reflects an integration across all of the varying features, placing each outcome onto a common scale, termed subjective value (Lak et al., 2014; Louie and Glimcher, 2012; Schultz, 2015). Subjective value allows comparison of each reward *relative* to the other available rewards; these relative values then inform individuals' behavioral preferences when choosing a reward. The fact that neural correlates of subjective value have been found in humans, including ventral striatum (Kable and Glimcher, 2007), has lent credence to this coding scheme underlying preference.

One implication of a subjective valuation system is that the exact same reward can have different subjective values in different scenarios. Thus, looking for differences in neural activity for the same stimulus in changing task conditions is a key test for relative value coding. One implementation of this phenomenon is through the contrast effect (Flaherty, 1999), a behavioral phenomenon where the amount of consumption (or seeking) of a first reward is altered when a better (or worse) second reward is also offered (compared to individuals who only receive the first reward). This style of experiment has been adapted for electrophysiology studies that have found individual neurons whose activity for the same reward depends on its ranking among available options. There have been several reports of this in NAc (Taha and Fields, 2005; Webber et al., 2016; Cromwell et al., 2005; Wheeler et al., 2005) as well as amygdala (Bermudez and Schultz, 2010; Saez et al., 2017), orbitofrontal cortex (Tremblay and Schultz, 1999; Saez et al., 2017), and parietal cortex (Louie et al., 2011). We use a similar approach to characterize VP activity in Chapter 2.

Another key component of subjective value is the physiological state of the individual. A popular theory is that the purpose of rewards is to reduce homeostatic drive (Hull, 1943); a complementary theory is that the pleasantness of a reward depends on internal metrics of satiety in a process known as alliesthesia (Cabanac, 1971). It follows that neural representations of a reward's value will track the physiological need for that reward (Keramati and Gutkin, 2014; Schultz, 2015). Across the brain, there have been many observations of

individual neurons whose reports of reward value track changes in satiety (de Araujo et al., 2006; Bouret and Richmond, 2010; Burgess et al., 2016; Livneh et al., 2017; Allen et al., 2019; Livneh et al., 2020), including in VP (Fujimoto et al., 2019; Stephenson-Jones et al., 2020).

While these results demonstrate that reduced hunger or thirst changes representations of food and liquid outcomes, respectively, a more nuanced question is how physiological state can differentially impact the *relative* values of different outcomes. One example of this phenomenon is sensory-specific satiety (Rolls et al., 1981). Feeding an individual to satiety on one reward reduced the activity of individual neurons in orbitofrontal cortex and hypothalamus to that reward but preserved the responses to other rewards that were not fed to satiety (Rolls et al., 1986, 1989); this finding also extended to reward-predicting stimuli in orbitofrontal cortex (Critchley and Rolls, 1996). Whether sensory-specific satiety similarly impacts firing of individual neurons in VP is a question we address in Chapter 4 of this dissertation.

Another example of reward-specific changes in neural encoding due to physiological state was demonstrated by manipulating salt appetite. Although highly salty solutions typically produce aversive behavioral responses, induction of a salt appetite through sodium deprivation causes an increase in hedonic responses to salty solutions (Berridge et al., 1984). This type of procedure was used to probe physiological state-dependent neural firing in VP (Tindell et al., 2006, 2009). Before deprivation, VP activity to oral infusions of salty water was much lower than following sucrose; after deprivation, when both salt and sucrose solutions produced hedonic responses, the salt-evoked activity actually exceeded sucrose (Tindell et al., 2006). In a similar experiment with the addition of specific reward-predicting cues, there was an increase in the number of neurons with responses for the salt-predicting cue following deprivation, even before presentation of salty water in the newly deprived condition (Tindell et al., 2009). Importantly, in both studies, the neural response to sucrose (and associated cues) was unchanged. These data demonstrate that VP value representations (both

for primary reward and for predictive stimuli) are sensitive to reward-specific physiological changes.

The studies on VP mentioned above have found promising evidence that VP encodes the value of available outcomes, but many questions remain. Does the reward-evoked activity of individual VP neurons shift for the same reward when the value of the other available rewards changes (as seen with the contrast effect)? Within an individual session, can individual VP neurons track the relative values of rewards that are differentially impacted by a changing physiological state, or does this shift occur on a slower timescale on a population level? These questions motivate our work in this dissertation, with the ultimate goal of clarifying both the robustness of relative value coding in VP and the general properties of relative value signals in the nervous system.

1.4 Expectation and reward prediction error.

In much of the work summarized above, the primary goal was to find the neural correlates of a reward's value. Perhaps at odds with this experimental goal, the subject (rodent, primate, or otherwise) is constantly updating its understanding of the experimental environment. Knowledge of the parameters of a task is a serious confound to the interpretation of neural responses to rewards and reward-predicting stimuli because this neural activity often incorporates the animal's expectations. One common solution is to conduct extensive training before beginning recordings in hopes that the neural representations (and the animal's understanding) of the task have stabilized. Another solution is to intentionally engage with the learning aspects of the task and include (or even focus on) them in the course of analysis.

One popular framework that formalizes the way individuals adapt to their environment is reinforcement learning (Sutton and Barto, 1998). Reinforcement learning posits that animals are able to exhibit flexible behavior by integrating information about past interactions with the environment to make predictions about the future that guide their behavioral responses. Predictions are updated when the achieved outcome is different than what was expected;

these deviations are known as reward prediction errors (RPEs), and they are used to iteratively update future predictions. This type of iterative updating of estimates of task value was used in Rescorla and Wagner (1972) to describe how animals learn about Pavlovian conditioned stimuli, particularly compound stimuli.

In the simplest version of Pavlovian conditioning, a neutral stimulus such as a light or tone (referred to as the conditioned stimulus) will begin to produce conditioned responding in an individual when it is reliably followed by a rewarding (or averisve) unconditioned stimulus, like food (Pavlov, 1927). In the phenomenon known as blocking (Kamin, 1968), when A is a conditioned stimulus that has been paired with an unconditioned stimulus and reliably produces conditioned responding, and X is a novel stimulus, simultaneous presentation of A and X prior to the unconditioned stimulus will not cause X to acquire any ability to promote conditioned responding; learning about X is 'blocked.' This process can be explained mathematically within an RPE framework, as first demonstrated by Rescorla and Wagner (1972):

$$\Delta V_A = \alpha_A \beta(\lambda - V_{AX})$$
$$\Delta V_X = \alpha_X \beta(\lambda - V_{AX})$$

where ΔV_A and ΔV_X are the changes in the strengths of the associations between each cue and the unconditioned stimulus, α_A and α_X are cue-specific learning rates (how quickly the associations are updated), β is the learning rate common to both cues, λ is the asymptote of learning (basically, the value of the unconditioned stimulus), and V_{AX} is the total associative strength of both cues (estimated as $V_A + V_X$). If A has already been paired extensively with the unconditioned stimulus, then $V_A = \lambda$, so $V_{AX} = \lambda$ and ΔV_A and ΔV_X (the prediction errors) will both be 0. Thus, the associative strength of both cues remains unchanged across trials, so learning to X is blocked.

This equation can be simplified to describe trial-based learning more generally:

$$\delta(t) = o(t) - V(t)$$

$$\Delta V = \alpha \cdot \delta(t)$$

where the change in the reward-associated value of the task (ΔV) is equal to the prediction error on the current trial ($\delta(t)$), or the difference between the value of the outcome received on that trial ($o(t)$) and the expected value for that trial ($V(t)$), scaled by a learning parameter that determines how quickly the association is updated (α). This model does a good job of capturing trial-by-trial dynamics of the reward-evoked activity of dopamine neurons (Bayer and Glimcher, 2005), and it provides a potential explanation for changes in relative value signaling across evolving reward conditions (Padoa-Schioppa, 2009; Saez et al., 2017). We use this equation to characterize a role for expectation in VP relative value signaling in Chapter 3.

The Rescorla–Wagner model describes changes in expected value across trials, but it does not capture within-trial dynamics. Temporal difference (TD) learning is a key component of reinforcement learning that recasts predictions in terms of time steps rather than trials as the interval of learning (Sutton, 1988; Sutton and Barto, 1998). This means that there is a prediction error (δ) for each time step (Schultz et al., 1997):

$$\delta(t) = o(t) + \gamma V(t+1) - V(t)$$

where $o(t)$ is the value of any outcome received at the current time step, $V(t)$ and $V(t+1)$ are the value estimates for the current and subsequent time steps, and γ is a discounting factor that allows rewards sooner in the future to have a greater impact on the estimate.

TD learning captures many of the key observations of dopamine neurons (Schultz et al., 1997; Nakahara et al., 2004; Pan et al., 2005). Most famously, an expected delivery of fruit juice produces a phasic burst of activity in dopamine neurons corresponding to a positive

prediction error. After extensive exposure to a conditioned stimulus that predicts the juice reward, the burst of firing corresponding to a positive prediction error transfers to the cue, and no change in firing occurs at juice delivery. Finally, omission of juice following presentation of the conditioned stimulus produces a pause in firing corresponding to a negative prediction error. TD learning provides a framework for describing why each of these error signals occur at each phase of the trial, and it has inspired scores of studies examining the precise nature of neural encoding of predictive cues and their associated outcomes, including many of the studies of value we noted above.

Because the activity of dopamine neurons resembles a full TD prediction error, one area of inquiry has been to determine the source of each component of the error calculation, with the idea that dopamine neurons compute RPEs locally by integrating distinct elements of the signal relayed from distinct input regions (Keiflin and Janak, 2015; Tian et al., 2016; Watabe-Uchida et al., 2017). Work on upstream contributors to dopamine neuron RPE calculation has revealed that individual regions are critical for different elements of the RPE: lateral habenula contributes to the coding of negative RPEs (Matsumoto and Hikosaka, 2007; Tian and Uchida, 2015) via the rostromedial tegmental nucleus (Jhou et al., 2009; Hong et al., 2011), orbitofrontal cortex contributes to expectancy-related RPE signaling, including adapation following repeated reward presentations (Takahashi et al., 2011), and ventral striatum appears necessary for proper temporal specificity of RPEs (Takahashi et al., 2016). Locally, GABAergic neurons in the ventral tegmental area contribute to the computation by providing an expectancy-related subtraction (Eshel et al., 2015).

A pioneering study on the neural activity of monosynaptic inputs to dopamine neurons revealed a mixture of reward and expectation signals across many brain regions (including VP) rather than distinct components (like value or outcome) in each region, and, notably, there were very few upstream neurons encoding full RPEs (Tian et al., 2016), maintaining the idea that, by and large, RPE is calculated within dopamine neurons themselves (Watabe-Uchida et al., 2017). On the other hand, given a recent report of TD error-like signaling in

VP (Stephenson-Jones et al., 2020), and the fact that Tian et al. (2016) only examined VP neurons that synapse directly onto dopamine neurons, the existence of TD learning signals (and RPE signals more generally) in VP needs further clarification. Our data in Chapter 3 and 4 address this possibility.

1.5 Main objectives

As summarized above, there are a number of questions about the role of VP within the ventral striatopallidal circuit and within the context of value signaling that need to be resolved. Our main goal in this dissertation was to design a series of well-controlled behavioral tasks that would allow us to explore the influence of preference, relative value, and expectation on reward-related neural activity in VP, with comparisons to signaling in NAc. Furthermore, with the addition of optogenetic manipulations inspired by our electrophysiological findings, we aimed to link the reward-related activity we observed in VP with a behavioral output. The experiments are presented as follows:

For *Chapter 2*, we recorded from NAc and VP in a task contrasting two calorically equivalent and similarly palatable rewards. This approach allowed us to characterize the influence of preference (rather than reward magnitude or caloric value) on reward-evoked neural activity in each region and compare the timing of these representations. We went on to examine the relative nature of VP value signaling by recording during sessions with additional reward combinations.

For *Chapter 3*, we developed a modeling approach within the Rescorla–Wagner prediciton error framework to identify the influence of recent outcome history on the VP reward-evoked signaling. Our findings motivated an optogenetic test of a role for this VP signal in updating value estimates as they pertain to task engagement.

For *Chapter 4*, we implemented two novel behavioral tasks that assessed shifting behavioral preference across evolving physiological state and compared this preference to VP neural reports of relative value at the time of the cue and the reward outcome. We also

investigated the link between VP activity at these time points and choice behavior.

In *Chapter 5*, we summarize our results, consider limitations of our work, and discuss open questions for the future.

1.6 General Methods

The experiments presented in this thesis use a combination of behavioral, electrophysiological, and optogenetic approaches. Here, we present the methods that were common across these experiments. Specific task details are presented in the respective chapters.

Animals.

Subjects were male (and female, where noted) Long-Evans rats from Envigo weighing 200-275g at arrival and single-housed on a 12hr light/dark cycle. Rats were given free access to food and water in their home cages for the duration of the experiment (unless otherwise noted). All experimental procedures were performed in strict accordance with protocols approved by the Animal Care and Use Committee at Johns Hopkins University.

Reward solutions.

Reward solutions were 10% solutions by weight of sucrose (Thermo Fisher Scientific, MA) and maltodextrin (SolCarb, Solace Nutrition, CT) in tap water, or tap water alone. Before behavioral training, rats were given free access to sucrose and/or maltodextrin solution in their home cages depending on the rewards used for the experiment.

Surgical procedures.

Rats were anesthetized with isoflurane (5%) and maintained under anesthesia for the duration of the surgery (1-2%). Rats received injections of carprofen (5 mg/kg) and cefazolin (70 mg/kg) prior to incision. **Electrophysiology.** Drivable electrode arrays were prepared with custom-designed 3D-printed plastic pieces assembled with metal tubing, screws, and

nuts. 16 insulated tungsten wires and 2 silver ground wires were soldered to an adapter that permitted interfacing with the headstage (Plexon Inc, TX). The drives were surgically implanted in trained rats. Using a stereotactic arm, electrodes were aimed at either NAc (AP +1.5 mm, ML +1.2 mm, DV -7 mm) or VP (AP +0.5mm, ML +2.4mm ML, DV -8mm). The base of the drive and the adapter were secured to the skull with 7 screws and cement. The ground wire was wrapped around a screw and placed superficially in brain tissue in a separate craniotomy posterior to the recording electrodes. **Optogenetics.** First, 0.7 μL of virus containing the archaerhodopsin gene construct (Han et al., 2011) (AAV5-CamKIIa-eArchT3.0-eYFP, $7 * 10^{12}$ viral particles/mL from the University of North Carolina Vector Core), channelrhodopsin (AAV5-hsyn-hChR2(H134R)-EYFP, $1.7 * 10^{13}$ viral particles/mL from Addgene, gift from Karl Deisseroth) or their respective control virus (AAV5-CamKIIa-eYFP, $7.4 * 10^{12}$ viral particles/mL from the University of North Carolina Vector Core, or AAV5-hsyn-EGFP, $1.2 * 10^{13}$ viral particles/mL from Addgene, gift from Bryan Roth) was delivered bilaterally to VP through 31 gauge gastight Hamilton syringes at a rate of 0.1 μL per min for 7 minutes controlled by a Micro4 Ultra Microsyringe Pump 3 (World Precision Instruments). Injectors were left in place for 10 min following the infusion to allow virus to diffuse away from the infusion site. Injector tips were aimed at the following coordinates in relation to Bregma: +0.5 mm AP, +/-2.5 mm mediolateral, -8.2 mm dorsoventral. Then, rats were implanted with 300 micron diameter optic fibers constructed in house, aimed 0.3 mm above the center of the virus infusion. Optic fiber implants were secured to the skull with 4 screws and dental cement.

Histology.

Animals were anesthetized with pentobarbital. Rats were perfused intracardially with 0.9% saline followed by 4% paraformaldehyde, after which brains were extracted and post-fixed in 4% paraformaldehyde for 24hrs. Brains were then transferred to 25% sucrose for at minimum 24hr before being frozen on dry ice and sectioned into 50um slices on a cryostat.

Electrophysiology. Following the injection with pentobarbital, electrode sites were labeled by passing a DC current through each electrode. Following sectioning, slices were stained with cresyl violet to determine recording sites. **Optogenetics.** For slices from rats from the optogenetic inhibition experiments, we performed immunohistochemistry for GFP and substance P (SP), in order to identify the localization of virus expression and fiber placement within the borders of VP. Sections were washed in PBS with bovine serum albumin and triton (PBST) for 20 minutes, and incubated in 10% normal donkey serum in PBST for 30 minutes, before incubating in primary antibody (mouse anti-GFP 1:1500 Thermo Fisher #A11120, RRID: AB_221568; rabbit anti-SP 1:6500 Immunostar #20064, RRID: AB_572266) in PBST overnight at 4°C. Sections were then washed with PBST 3-times, incubated in 2% normal donkey serum in PBS for 10 minutes, and incubated for 2 hours in secondary antibody in PBS (Alexa Fluor 488 donkey anti-mouse 1:200 Thermo Fisher #A21202, RRID: AB_141607; Alexa Fluor 594 donkey anti-rabbit 1:200 Thermo Fisher #A21207, RRID: AB_141637). Sections were then washed with PBS 3-times, mounted on coated glass slides in PBS, air-dried, and coverslipped with Vectashield mounting medium with DAPI. For slices from rats from the optogenetic stimulation experiments, we determined there was enough native fluorescence from the virus, so we mounted directly on coated glass slides and coverslipped with Vectashield mounting medium with DAPI.

Recording and spike sorting.

During recording sessions, rats were tethered via a cable from their headstage to a commutator in the center of the chamber ceiling. Electrical signals and behavioral events were collected using the OmniPlex system (Plexon) with a 40kHz sampling rate. Spikes were sorted into units using offline sorter (Plexon); following initial manual selection of units based on clustering of waveforms along the first two principal components, units were separated and refined using waveform energy and waveform heights at various times relative to threshold crossing ("slices"). Any units that were not detectable for the entire session

were discarded. Event creation and review of individual neurons' responses were conducted in NeuroExplorer (Nex Technologies, AL). Cross-correlation was plotted for simultaneously recorded units to identify and remove any neurons that were recorded on multiple channels.

Optogenetic manipulations.

At least 5 weeks after surgery and completion of operant training, rats were habituated to patch cord connections. Rats were connected bilaterally (inhibition) or unilaterally (stimulation) via ceramic mating sleeves to a 200 μm core patch cord (bifurcated for inhibition group), which was then connected through a fiber-optic rotary joint (Doric), to another patch cord which interfaced with either a 532nm (inhibition) or 473nm (stimulation) DPSS laser (Opto-Engine LLC). The time of laser delivery was initiated by TTL pulses from MedPC SmartCTRL cards to a Master9 Stimulus Controller (AMPI) which dictated the duration of stimulation.

1.7 Disclosures

One published manuscript and one preprint have been reformatted to comply with the requirements of this thesis (Ottenheimer et al., 2018, 2019a).

Chapter 2

Ventral pallidum encodes relative reward value earlier and more robustly than nucleus accumbens

This chapter is adapted from Ottenheimer et al. (2018).

2.1 Introduction

A collection of anatomical (Alexander et al., 1986; Groenewegen et al., 1999; de Olmos and Heimer, 1999) and experimental (Chang et al., 2018; Creed et al., 2016; Leung and Balleine, 2013; Smith and Berridge, 2007) studies have led to the prevailing view that VP serves primarily as a major output of NAc within the striatopallidal system (Root et al., 2015; Smith et al., 2009). A serious limitation thus far in the characterization of this circuit is the lack of comparative observations in NAc and VP on a timescale relevant for the reward-related processing in which the circuit is functionally implicated. Understanding the transformation of reward-related information across the ventral striatopallidal system during a seconds-long behavioral response requires temporally precise measurements of neural activity in each region. Previous work has identified neurons in both NAc (Ambroggi et al., 2011; Bissonette et al., 2013; Cooch et al., 2015; Goldstein et al., 2012; Nicola et al., 2004; Roesch et al., 2009; Setlow et al., 2003; Taha and Fields, 2005; Wheeler et al., 2005; Roitman et al., 2005;

Villavicencio et al., 2018) and VP (Avila and Lin, 2014a,b; Itoga et al., 2016; Lin and Nicolelis, 2008; Richard et al., 2016, 2018; Smith et al., 2011; Stephenson-Jones et al., 2020; Tindell et al., 2006, 2009) with phasic responses to rewards (and their predictive stimuli) that track outcome value, but it is unclear how reward-related neural responses in VP result from activity in NAc, as the classic model of ventral striatopallidal function would predict. In fact, a recent comparison of activity in NAc and VP in the same behavioral task found that the onset of cue responses in VP frequently precedes their onset in NAc (Richard et al., 2016), leaving the question of whether VP acts exclusively downstream of NAc in this reward processing circuit.

To further interrogate the respective roles of VP and NAc in reward processing, we set out to measure neural activity in a task with multiple reward outcomes. In addition to permitting a comparison of the onset of phasic activity in response to reward outcome, this approach allowed us to track over time the reward-specific information contained within the spiking activity of individual neurons and neural ensembles in each region. Surprisingly, we found that a much greater proportion of VP neurons was reward-selective than neurons in NAc. Moreover, the reward-specific information signaled by both individual neurons and ensembles in VP preceded that signaled by NAc neurons. Further, we found that VP neurons reliably and rapidly tracked relative value across a variety of reward conditions. The flexibility of this VP value signal and its abundance within the neural population establish VP as a robust value signaler and suggest it does so at least partly independently of its classical input, NAc, encouraging consideration of VP as an important reward processing center rather than simply a relay for reward-related information to motor outputs.

2.2 Materials and Methods

Behavioral task.

Rats were trained to respond to a 10s white noise cue by making an entry into the reward port. The cue terminated upon port entry, and 500ms following port entry, 110μl of either reward was delivered into the metal cup within the reward port. Sucrose and maltodextrin trials were pseudorandomly interspersed throughout the session such that rats could not detect the identity of the reward until it was delivered. Individual licks were recorded with a custom-built arduino-based lickometer using a capacitance sensor (MPR121, Adafruit Industries, NY) with a 1kHz sampling rate. Each cue was separated by a variable intertrial interval (ITI) that averaged 45s. During the ITI, the reward cup was evacuated via vacuum pump, flushed with 110μl of water, and evacuated again. Maltodextrin, sucrose, and water were each delivered via separate infusion pumps (Med Associates, VT) and separate metal tubes entering the cup. There were a total of 60 trials per session. After we started recording, on some sessions, we presented the rewards in blocks of 30 trials each (data presented in Chapter 3).

Preference test.

To assay rats' preference for sucrose or maltodextrin, we performed two 60-minute two-bottle choice tests, during which rats had free access to 10% solutions of each reward. Bottles were weighed before and after to determine the amount of each solution consumed by each rat. The first test was following recovery from surgery and prior to recording. The second was at least a day after the final session with sucrose and maltodextrin and prior to any subsequent sessions with different reward outcomes (see below).

Recording.

Following a week of recovery in their home cages (and the first two-bottle choice test), rats were trained on the task again until they became accustomed to performing the task while tethered. Once they responded on at least 40 of 60 trials, recording sessions began. We continued to record from the same location for multiple sessions if new neurons appeared on previously unrecorded channels; if multiple sessions from the same location were included in analysis, the same channel was never included more than once. If no neurons were detectable or following successful recording, the drive was advanced 160μm, and recording resumed in the new location at minimum two days later to ensure settling of the tissue around the wires.

Additional sessions with altered reward outcomes.

For 3 of the rats with electrodes in VP (VP2, VP3, and VP5), we conducted an additional session with water (replacing sucrose) and maltodextrin (VP3 did not complete the session, likely due to low motivation to pursue the new reward outcomes). The session was otherwise unchanged from those with sucrose and maltodextrin. Subsequently, all three rats were tested in sessions with all three outcomes available. The three trial types were pseudorandomly interspersed throughout the session. The total number of trials was expanded to 90 to permit equivalent amounts of each reward to the previous sessions.

Initial neural analysis.

Analysis of neural activity was performed in MATLAB (MathWorks, MA). Event-related responses were found by constructing peristimulus time histograms (PSTHs) for spikes following each event. Neurons were determined to be modulated by an event if the spike rate in a custom window following each presentation of the event significantly differed from a 10s window prior to cue onset according to a Wilcoxon signed-rank test ($p < 0.05$, two-tailed). For these tests, we analyzed activity 500ms after the cue, the 1000ms centered on port entry, and 1000ms after reward delivery.

Optimal bin size for averaged PSTH activity was determined as described previously Ambroggi et al. (2011). Briefly, the optimal bin size for each neuron was found using Akaike Information Criteria (AIC). For our data, we used the smallest possible bin size that showed less than a 10% change from the optimal AIC value. This bin size, referred to as the deflection point, typically ranged from 20 to 100ms. The spiking activity across these bins was smoothed with a LOWESS function.

To visualize the normalized activity of neurons, the mean activity within each of the smoothed, optimally-sized bins of the PSTH plots for each neuron was transformed to a z-score with the equation (Fi – Fmean)/Fsd, where Fi is the firing rate of the ith bin of the PSTH, and Fmean and Fsd are the mean and the standard deviation of the firing rate during the 10s baseline period. Color-coded maps of individual neurons' activity and average activity traces were constructed based on these z-scores.

Licking analysis.

PSTHs for visualizing licking activity around reward solution delivery were constructed as for neurons (above) with a fixed bin size of 100ms and LOWESS smoothing. To test for differences in the duration of the licking bout and the number of licks on sucrose and maltodextrin trials, we ran a three-way ANOVA on the raw licking data for the fixed effect of reward and the random effects of session and subject, with session nested within subject (with trial as our n). We also ran this test on the number of licks 1-4.5s post reward delivery, an epoch in which we noticed a visible difference in the average lick rat (Fig. 2.1d). We further characterized this difference in licking activity by finding the mean duration of the interlick intervals following the first 30 licks of each reward. We ran a three-way ANOVA for the fixed effects of reward and interval no. and the random effect of subject (with trial as our n).

Classification of neurons as reward-selective.

For the analysis of reward-selective activity during reward consumption, we segmented the time surrounding reward delivery into overlapping 600ms bins advanced by 100ms. We only included trials in which the rat began licking within 2s of reward delivery to ensure the rat sampled the reward on each included trial. Neurons were significantly reward-modulated for a given bin if they there was a significant interaction ($p<0.01$) for that neuron between the effect of baseline (-22 to -12s from reward delivery) vs bin firing and the effect of reward solution (with trial as our n) in two consecutive bins. This approach minimized the amount of noise in the classification (measurable as the number of neurons classified as reward selective prior to reward delivery) while still permitting relatively brief reward-specific responses to register. We then further classified these reward-selective neurons by the reward for which they had greater normalized firing in that bin, found with the equation (Fb – Fmean)/Fsd, where Fb is the firing rate of each bin, and Fmean and Fsd are the mean and the standard deviation of the firing rate during the 10s baseline period. This same analysis was used to classify neurons from the sessions comparing water and maltodextrin.

To choose which bins best captured the population of reward-selective neurons across both regions, we plotted the cumulative onset of reward selectivity for all neurons as a fraction of the total population (Fig. 2.6a,b). We chose to include all neurons that were reward-selective in any of the bins from 0.4 to 3s, which captured the majority of phasic reward-specific responses following reward delivery in both regions. To determine which of these neurons were significantly excited or inhibited by either reward (Fig. 2.6g,j,m,p), we performed a Wilcoxon signed-rank test comparing on each trial the (raw) firing rate during the -22 to -12s baseline window from reward delivery to the firing rate in each of the bins centered 0.4-3s for each reward ($p < 0.05$ cutoff, two-tailed). A neuron was considered excited or inhibited by a given reward if it had a significant increase or decrease in spikes for any of the bins 0.4-3s post reward delivery. We also plotted the cumulative onsets of reward selective neurons as a fraction of total reward selective neurons in each region to compare the

timing of the onsets in each region and compared the distributions with a two-way ANOVA with the main effect of region and the random effect of subject (Fig. 2.6d).

To classify neurons as reward-selective with three reward outcomes, we performed the same ANOVA analysis as before with the water condition added to the effect of reward, looking for an interaction between the effects of reward and baseline vs bin firing (with trial as our n). We then further classified reward-selective neurons by the outcome for which they had the greatest spiking in that bin and found, as before, if a neuron was significantly inhibited or excited by any of the outcomes in any bin 0.4-3s with Wilcoxon signed-rank tests ($p < 0.05$ cutoff, two-tailed). Because there were three outcomes, we also performed a two-way ANOVA on the effect of reward outcome (and random effect of subject) on the average normalized firing 0.8-1.4s post reward delivery (the bin with the most number of reward-selective neurons) of all selective neurons with greatest firing for sucrose (Fig. 2.12c) as well as pairwise comparisons between the three rewards (Tukey test, correcting for multiple comparisons).

Quantification of average activity based on current and previous reward.

To examine how average activity in each region was affected by previous reward, we normalized the average activity of all neurons in each region to their baseline firing rate in a window -22 to -12s from reward delivery. We chose to quantify the average activity 0.8-1.3s post reward delivery (marked with blue lines in Fig. 2.10a,c), a period we visually identified as having the best evidence of previous reward-modulated activity. Thus, the activity of neurons was normalized with the equation (Fr – Fmean)/Fsd, where Fr is the mean firing rate 0.8-1.3s following reward delivery for each of the four current/previous reward combination, and Fmean and Fsd are the mean and the standard deviation of the firing rate during the 10s baseline period on all trials. We then performed ANOVAs testing the effects of reward and previous reward (and random effect of subject) on the normalized activity of neurons in that window for each region (with neuron as our n). To compare the regions, we also

performed a test on all the neurons from both regions with the added factor of region.

Linear models.

To find the impact of previous trials' outcomes on current trial firing, we fit linear models ("fitlm" in MATLAB) to the firing rate of each neuron on each trial according to the outcomes on the current trial and the previous 6 trials. For this analysis, we used the same window as above, 0.8-1.3s post reward delivery, and normalized the activity for each neuron on each trial to the activity of that neuron during baseline period -22 to -12s from reward delivery on all trials. The normalized activity on each trial was paired with a corresponding vector of seven 0s and 1s indicating the reward outcome (0 for maltodextrin and 1 for sucrose) on the current and previous six trials (this required exclusion of all trials preceding the seventh completed trial). This convention caused positive coefficients to indicate a positive influence of receiving sucrose rather than maltodextrin on firing rate for that trial and vice versa. We then found the coefficients for each of the seven relative trials for each neuron as well as whether there was a significant impact of that relative trial on firing rate ($p < 0.05$, two-tailed t-test). We then did the same analysis but shuffled the trial outcomes to find what values would be expected by chance. For each region, we performed ANOVAs testing the main effects of shuffled vs true data and trial relative to current (and the random effect of subject) on coefficient and then performed Tukey tests correcting for multiple comparisons to find differences on each trial between the coefficients and their shuffled data ($p < 0.05$). We tested for significant differences in the proportion of neurons with significant coefficients between true and shuffled data from both regions, as well as between the true data from each region, with chi-squared tests for each relative trial ($p < 0.05$).

Emergence of responses to water and maltodextrin.

To track how the average activity of reward-selective neurons changed on water and maltodextrin trials across the session with those two reward outcomes, we normalized the mean

activity of the reward-selective neurons identified in Fig. 2.11c-e on each trial to their baseline activity in the 10s window -22 to -12s from reward delivery. We focused our analysis on each neuron's normalized activity 0.8-1.8s following reward delivery, an epoch we visually identified as representative of the maltodextrin excitations and water inhibitions. Thus, neurons were normalized with the equation (Ft – Fmean)/Fsd, where Ft is the mean firing rate 0.8-1.8s following reward delivery on each trial, and Fmean and Fsd are the mean and the standard deviation of the firing rate during the 10s baseline period on all trials. We then plotted the average activity according the number of trials the rat had completed (Fig. 2.11f). We performed a two-way ANOVA (reward X trials of reward) on the normalized activity of the neurons from each rat across each respective trial of each reward (with each neuron's normalized activity on each trial as our n). This approach required capping the total number of trials included in the test at the maximum number of trials for the reward with the least number of completed trials.

Decoding.

For single unit decoding, a linear discriminant analysis (LDA) model (the "fitcdiscr" function in MATLAB) was trained on each neuron's spike activity for one 600ms bin on 80% of trials. This model was then used to classify the remaining 20% of trials as sucrose or maltodextrin. We performed this 5 times in a 5-fold cross-validation approach and averaged performance across all 5 repetitions to find that unit's accuracy. We also conducted the analysis with the trial identities shuffled to determine the accuracy on shuffled data. We then repeated this analysis for every neuron in each region for each bin. If there were fewer than 7 spikes across all sucrose or maltodextrin trials, we excluded that neuron for that bin to avoid errors from creating an LDA model on a dataset with too little variance. To determine when accuracy in each region improved over shuffled data, we found all bins when the mean accuracy of the true data exceeded the 99% confidence interval of the shuffled data for at minimum 2 consecutive bins. To ensure that our results were not affected by the greater number of

neurons in VP (423 versus 182), we took 20 random selections of 182 of the unit models from VP and recalculated the confidence intervals to evaluate if it would affect the results (by and large it did not; see Results). To compare accuracy in our standard window of 0.4-3s after reward delivery (Fig. 2.6) across regions, we performed an ANOVA testing the effects of shuffled vs true data (whether or not the accuracy came from a shuffled data model or a true data model), region, and bins (and the random effect of subject) with each neuron model's true or shuffled accuracy in each bin as our n. We also performed an ANOVA testing the effects of shuffled vs true data and region (and the random effect of subject) to compare the accuracy of the most accurate bin in each region (with the shuffled and true data from each neuron in the respective bin from each region as our n). To compare only reward-selective neurons (Fig. 2.9), we performed the same tests but included only neurons classified as reward-selective in Fig. 2.6.

To look at how model classification accuracy increased with additional units, we pooled together separately recorded units. This approach requires matched numbers of trials, so we only included neurons recorded during sessions with at least 20 trials of each reward. Subsequently, when training our pseudoensemble LDA models, we restricted the analysis to 20 (randomly selected) trials of each reward. We found the 5-fold cross-validated accuracy for models trained on the activity of randomly selected levels of 10, 25, 50, 100, and 150 units from each region. For each level, we performed the analysis 50 times. We then performed a two-way ANOVA on the effects of pseudoensemble size and region on the accuracy at each level's peak bin (with each repetition at that peak bin for each level as our n). We also performed a two-way ANOVA on the effects of pseudoensemble size and region on the time of most accurate bin for each LDA model replicate (with each repetition's peak bin time at each level as our n). We also performed these analyses on pseudoensembles containing only reward-selective neurons as classified in Fig. 2.6 (Fig. 2.9).

Statistical analysis.

We used analysis of variance (ANOVA) tests (the "anovan" function in MATLAB) to test for main effects and interactions, Tukey tests for pairwise comparisons corrected for multiple comparisons ("multcompare" in MATLAB), and chi-squared tests for contingencies ("crosstab" in MATLAB). For all ANOVAs testing behavioral and neural data across subjects, we included the random effect of subject to account for non-independence in the data.

2.3 Results

To test encoding of multiple rewards in NAc and VP, we chose to compare responses to 10% solutions of sucrose and maltodextrin, two palatable carbohydrates with equivalent caloric value but distinct tastes (Nissenbaum and Sclafani, 1987; Sako et al., 1994; Treesukosol et al., 2011). After multiple days of free access to the sucrose and maltodextrin solutions in their home cages, rats began training on the behavioral task. On each trial, 110μL of reward solution was delivered into a metal bowl contingent upon rats' entry into the reward port during a 10s white noise cue (Fig. 2.1a). Trials with presentation of a given solution were pseudorandomly interspersed throughout the session, an approach that obscured the reward identity from the rat until the solution was delivered into the reward cup. Once rats responded to the cue on 80% of trials, we implanted drivable tungsten electrode arrays in either NAc or VP (Figs. 2.1b, 2.2). To evaluate reward preference, we conducted 60-minute two-bottle choice tests prior to and following the first and last recording sessions with sucrose and maltodextrin; rats consistently showed a preference for sucrose (Fig. 2.1c).

Neural recording sessions began after rats recovered from surgery. To monitor consumption during the task, we recorded each rat's individual licks during each recording session. The overall licking pattern was similar for both rewards; there was no significant main effect of reward on the total number of licks ($F(1,3142) = 1.24$, $p = 0.29$) or the total duration of licking ($F(1,3142) = 0.303$, $p = 0.59$) within the 15 seconds following reward delivery.

However, rats licked slightly, but significantly, more for the preferred reward, sucrose, during the period 1-4.5s following reward delivery (23.2 vs 22.3 licks; $F(1,3142) = 66.0$, $p = 5.3E\text{-}6$; Figs. 2.1d, 2.3). Complementarily, the interlick intervals following the first 30 licks of each trial were significantly shorter on sucrose trials (Fig. 2.1e; $F(1,3000) = 33.3$, $p = 0.000084$). This accelerated consumption of sucrose echoes the rats' preference for sucrose over maltodextrin in the two-bottle choice test (Fig. 2.1c).

2.3.1 More Neurons in VP Fire Reward-selectively than in NAc

We collected neural activity from 6 rats with electrodes in NAc (182 neurons, 4-49 per rat, median 32, 36 sessions) and 5 rats with electrodes in VP (436 neurons, 32-137 per rat, median 86, 25 sessions). Neurons in both regions responded to reward-related events: cue onset, port entry (PE), and reward delivery (RD) (Fig. 2.4), consistent with prior findings (NAc: Bissonette et al. (2013); Cooch et al. (2015); Goldstein et al. (2012); Nicola et al. (2004); Roesch et al. (2009); Setlow et al. (2003); Taha and Fields (2005); Carelli et al. (2000); Janak et al. (2004) VP: Avila and Lin (2014a,b); Lin and Nicolelis (2008); Richard et al. (2016); Tindell et al. (2006); Ahrens et al. (2016). In order to evaluate reward selectivity, we focused on the neural activity following reward delivery, when the rats first detected the identity of the reward and consumed it. Initial inspection of the peri-event histograms of individual neurons' spiking reveal instances of reward-specific firing in each region (Fig. 2.5).

To more precisely determine the presence and onset of reward-selective responding, we divided the time surrounding reward delivery into overlapping bins with a sliding window of 600ms advanced by 100ms. For each bin, we found the number of neurons whose firing rates were significantly differentially modulated across sucrose and maltodextrin trials (see Methods) and further categorized these neurons by which reward elicited greater firing. We conducted this analysis for all neurons from each region (Fig. 2.6a,b), as well as for each individual rat to ensure general consistency across subjects and recording locations (Fig. 2.7). Notably, the peak number of reward-selective neurons in any given bin was greater

in VP than in NAc (33% vs 10%, χ^2 = 34.3, p = 4.7E-9) and this bin with peak reward selectivity was earlier in VP (centered at 1.1s) than in NAc (centered at 1.9s) (Fig. 2.6a,b). We compared the time course of selectivity in each region by subtracting the proportion of selective neurons in VP from the proportion in NAc in each bin (Fig. 2.6c), revealing that at no point was there more reward selectivity in NAc than in VP, as might be expected if the reward-specific information originated in NAc. We also compared the onset of reward-selective responses in each region and found that the distribution of onsets was earlier in VP than in NAc (Fig. 2.6d).

To characterize the activity of reward-selective neurons in each region, we identified neurons that met our criteria for reward selectivity in any of the bins centered 0.4-3s after reward delivery, a period of time that captured the majority of phasic reward-selective responses across both regions (Fig. 2.6a,b). We then plotted these neurons' individual and averaged activity on sucrose and maltodextrin trials (Fig. 2.6e-p). Within this time window, 24% of neurons in NAc and 52% of neurons in VP were at one point reward-selective, a significantly greater proportion in VP (χ^2 = 39.9, p = 2.6E-10). In both regions, we found that most of the reward-selective responses were excitations for sucrose; some of these cells were also inhibited for maltodextrin (Fig. 2.6e-g,k-m). A smaller subset of reward-selective cells in each region had greater firing rates for maltodextrin, often due to an inhibition for sucrose (Fig. 2.6h-j,n-p). Thus, despite only minimal differences in licking behavior for each reward, a substantial proportion of neurons in both VP and NAc fire in a reward-selective manner, and these reward-selective responses are represented in a greater proportion of the recorded population in VP than in NAc.

2.3.2 VP Units and Ensembles Decode Trial Type Earlier and More Accurately than NAc

The analysis above indicates differential encoding of two rewards, sucrose and maltodextrin, with most selective units showing greater responding for sucrose, the preferred reward, over

maltodextrin. To complement this analysis, we used linear discriminant analysis (LDA) to test when and to what extent neural activity in each region could be used to predict reward identity. Using 5-fold cross-validation, we determined how accurately LDA models trained on the spike activity of individual neurons could classify trials as sucrose or maltodextrin. We conducted the analysis for each neuron in each region across each of the 600ms bins we used in Fig. 2.6. We compared these results to LDA models trained on the same data with the trial identities randomly shuffled. Models trained on single unit activity classified trial type at rates above chance in both VP and NAc (Fig. 2.8a). Focusing on our window of interest from Fig. 2.6 (0.4-3s post reward delivery), we found that VP single unit accuracy improved over shuffled data more than NAc (shuffled vs true X region: $F(1, 31150) = 11.5$, $p = 0.0019$). When comparing the most accurate bin in NAc (centered at 1.4s) to that in VP (centered at 1s), there was a noticeable shift in classification accuracy in the cumulative distribution function (CDF) (Fig. 2.8b), corresponding to a significantly greater improvement in accuracy over shuffled data in VP (shuffled vs true X region: $F(1,1154) = 13.6$, $p = 0.00037$). Notably, VP single units first improved over shuffled data for the bin centered at 0.5s, whereas NAc single units first improved over shuffled data at 0.9s (purple and green lines in Fig. 2.8a). To control for the possibility that the earlier improvement over shuffled data in VP relative to NAc was due to the larger number of neurons recorded in VP, we conducted the analysis 20 more times with 182 randomly chosen VP units. The first bin significantly more accurate than shuffled data ranged from 0.4-0.6s (median 0.5s), consistently earlier than 0.9s in NAc.

Although the data from individual neurons points to more reward-selective activity in VP than NAc, an alternate explanation is that reward-specific information is more distributed across neurons in NAc than in VP. If so, including additional neurons in the model should improve the accuracy of the NAc decoders relative to VP. To overcome the limited number of sessions in NAc with greater than 5 neurons, we pooled neurons together into pseudoensembles to compare how much information is contained within larger groups of neurons in each

region. We ran the same analysis as before using LDA models trained with the spiking activity of 10, 25, 50, 100, and 150 neurons randomly selected from each region. Increasing the number of neurons improved accuracy in both regions (Fig. 2.8c,d), contributing to a significant main effect of ensemble size on peak bin accuracy ($F(4,490) = 237$, $p = 4.3E\text{-}113$). Pseudoensembles in VP had greater peak accuracy than those in NAc across all levels, evident in a main effect of region on peak bin accuracy ($F(1,490) = 212$, $p = 3.3E\text{-}40$; Fig. 2.8e). Notably, pseudoensembles in VP reliably reached 100% decoding accuracy with 100 neurons; NAc pseudoensembles reached at most 97% with 150 neurons, the upper bound we could test with our dataset (Fig. 2.8e). The smaller difference in accuracy between the two regions with 150 neurons was reflected in a significant interaction between ensemble size and region on decoder accuracy ($F(4,490) = 8.73$, $p = 8.2E\text{-}7$). Nevertheless, although there was comparable accuracy across regions with increasing neurons, VP pseudoensembles continued to achieve peak accuracy earlier than those in NAc (Fig. 2.8f; main effect of region: $F(1,490) = 289$, $p = 2.5E\text{-}51$). Overall, our results from these decoding analyses confirm that VP neurons more reward-specific information than NAc neurons and indicate that this information arises and peaks earlier in VP than in NAc.

While our initial decoding analysis included all neurons from each region regardless of their status as reward-selective or not, we were also interested in directly comparing the amount of reward-specific information contained in the reward-selective population in each region, so we conducted the same decoding analyses but restricted our sample to those neurons classified as reward-selective in Fig. 2.6. The accuracy of single unit models trained exclusively on reward-selective neurons was much closer across regions (Fig. 2.9a); VP no longer improved over shuffled data more than NAc in the window 0.4-3s post reward delivery (shuffled vs true X region: $F(1, 13612) = 0.0952$, $p = 0.7617$) nor when comparing the peak bin in each region (shuffled vs true X region: $F(1, 504) = 3.46$, $p = 0.080$). Nevertheless, there continued to be a noticeably earlier rise and peak in accuracy for VP models than for NAc (Fig. 2.9a). For pseudoensemble models, we were limited by the number of reward-selective

neurons in NAc, but we found that with groups of 10 and 25 neurons, there was a main effect of region on peak bin accuracy ($F(1,196) = 38.9$, $p = 2.7E\text{-}9$; Fig. 2.9e) and time of peak accuracy ($F(1,196) = 54.8$, $p = 3.8E\text{-}12$; Fig. 2.9f), indicating that VP pseudoensembles consisting of reward-selective neurons are more predictive and achieve peak accuracy earlier than NAc reward-selective pseudoensembles. Overall, these data provide some evidence that, even among the reward-selective population, VP neurons represent reward-specific information earlier and more strongly than NAc.

2.3.3 VP Reward Signal Reflects Previous Outcome

Because the predominant reward-selective response in both regions was increased spiking for the preferred reward (sucrose) relative to maltodextrin, and given the results from previous recording studies (Taha and Fields, 2005; Wheeler et al., 2005; Tindell et al., 2006, 2009; Webber et al., 2016), we hypothesized that this reward-specific signal reflects relative reward value. If so, we would predict that the report of relative value would depend on recent reward history, which we can approximate by analyzing trials according to both the current and previously received reward. For instance, the relative value of sucrose would be greater following trials where rats received maltodextrin, and the relative value of maltodextrin would be lesser following sucrose trials. To look for evidence of such a scheme, we plotted the mean activity of all neurons in NAc and VP for each of the four combinations of previous and current reward (Fig. 2.10). While there was some evidence for previous-reward modulation of reward response across the population of neurons in NAc (Fig. 2.10a,b), VP neurons showed very prominent modulation of the reward-related response according to our prediction: greater firing for sucrose following maltodextrin trials and lesser firing to maltodextrin following sucrose (Fig. 2.10c,d). When analyzing the contribution of reward and previous reward to the neural activity in each region 0.8-1.3s following reward delivery (marked with vertical blue lines in Fig. 2.10a,c), we found a significant main effect for previous reward in VP ($F(1,1724) = 10.1$, $p = 0.022$) but not in NAc ($F(1,704) = 0.0167$, $p = 0.90$), though a

test including data from both regions did not find a significant interaction between previous reward and region ($F(1,2428) = 3.89$, $p = 0.055$).

We next sought to more quantitatively assess the impact of previous outcomes on the reward-evoked signals in each region by using a linear model approach that predicts a neuron's firing rate based on the reward outcomes on the current trial and each of the prior six completed trials (Bayer and Glimcher, 2005). The weights of the coefficients assigned to each trial reveal how strongly the outcome from that trial factors into the neuron's firing rate on the current trial. For both the current trial and the previous trial, only VP models showed, on average, coefficients that deviated from chance (Fig. 2.10e). Consistent with our relative value hypothesis and with our observations in Fig. 2.10c,d, the direction of the coefficients indicated a strong positive impact of receiving sucrose on the firing rate in the current trial and a negative impact of sucrose received on the previous trial. We also found that more neurons in VP had significant coefficients than in NAc for both the current and most recent trial (Fig. 2.10f), reflecting the stronger impact of reward outcome on VP firing. We found no impact of previous trials beyond the most recent on reward-related firing in either region. Thus, firing in VP at the time of reward reflects the previously received outcome.

2.3.4 VP Signals Reward Value Relative to Currently Available Options

Our results comparing VP neural responses to sucrose and maltodextrin are consistent with a relative value signal, but a stronger test of this hypothesis requires changing the relative value of the reward outcomes and looking for a corresponding change in neural activity. Such an approach has demonstrated that neurons in NAc report relative value (Taha and Fields, 2005; Wheeler et al., 2005; Webber et al., 2016), but it is unclear whether VP neural reports of a reward's value are relative to other currently available outcomes. We tested for relative value by conducting an additional session for VP rats in which sucrose was replaced with water, an outcome much less rewarding (in rats that are not water-restricted) than both sucrose and maltodextrin solutions (Treesukosol et al., 2011; Sclafani et al., 1987). We

predicted that, if VP neural activity reflects relative value, then the predominant reward-specific neural response would be excitations for maltodextrin, which in this scenario is the preferred outcome. Alternatively, if VP activity reflects absolute value, then the neural responses would remain suppressed to maltodextrin as in the sessions with sucrose and maltodextrin (Fig. 2.6k).

Two rats successfully completed this session type, contributing a total of 125 neurons (79 and 46, respectively). Water was much less preferred than maltodextrin, evident in the mean lick rate for each outcome (Fig. 2.11a). By calculating the number of reward-selective neurons across 600ms bins, we saw an even greater proportion of neurons in VP showed reward-specific responses for any given bin (59% at 1.3s) (Fig. 2.11b). We plotted the activity of neurons that met our criteria for reward selectivity during any bin within the 0.4-3s window we used in Fig. 2.6 (Fig. 2.11c,d); 70% of neurons were reward-selective during this time, a significantly greater proportion than the 53% in sessions these two rats completed with sucrose and maltodextrin ($\chi^2 = 9.68$, $p = 0.0019$); this higher proportion may reflect the considerable difference in value between water and maltodextrin compared to the similar value of the two appetitive reward outcomes, sucrose and maltodextrin. Consistent with our first prediction, the vast majority of these reward-specific neurons were excited by the preferred reward, maltodextrin, and most were also inhibited by the less preferred outcome, water (Fig. 2.11e).

Because this session was the first time rats experienced water and maltodextrin together, we were able to observe the emergence of the excitations for maltodextrin delivery, which previously produced a reduction in firing in reward-selective cells (Fig. 2.6k), and the emergence of inhibitions for the novel outcome, water. By averaging together the normalized activity of the reward-selective cells with greater firing for maltodextrin (the same group of neurons from Fig. 2.11c-e), we tracked the population's responses across each trial of each reward (Fig. 2.11f). In both rats, there was a noticeable increase in firing for maltodextrin and decrease in firing for water among this population of reward-selective neurons through-

out the session, reflected in a significant interaction between the effects of reward and the number of trials in both rats (VP2: $F(14,1500) = 17.0$, $p = 2.1E\text{-}39$; VP5: $F(27,1960) = 22.7$, $p = 6.2E\text{-}96$). In fact, despite being classified as having greater firing for maltodextrin than water, in neither rat did these neurons start out with greater firing for maltodextrin. These data demonstrate that neurons in VP modulate their responses within minutes to reflect the relative value of available outcomes in an altered reward landscape.

2.3.5 VP Activity Orders Three Outcomes by Relative Value

Finally, to test whether reward-selective neurons in VP can reflect the relative value of more than two options, we conducted additional sessions for VP rats where we reintroduced sucrose along with maltodextrin and water for a total of three possible reward outcomes. We recorded activity from 254 neurons in 3 rats (83, 104, and 67 neurons, respectively) across 4 total sessions. As before, we looked for neurons with significant reward-selective responses across the three outcomes for each 600ms bin and classified them by the reward that elicited the greatest firing. At most, 77% of the population was significantly modulated by reward outcome (for the bin at 1.1s), the majority of which had greatest firing for sucrose (Fig. 2.12b). We then looked at the activity of reward-selective neurons with greatest firing for sucrose during any bin in our standard 0.4-3s window. Remarkably, this population showed on average a large excitation for sucrose, a smaller excitation for maltodextrin, and an inhibition for water, consistent with the rats' relative preference for the three rewards (Fig. 2.12c,d). As a whole, this population had significantly different mean normalized firing rates for the time period 0.8-1.4s after reward delivery for all three reward outcomes ($F(2,561) = 441$, $p = 0.000014$; all pairs of rewards: $p < 1E\text{-}6$, Tukey test correcting for multiple comparisons). Therefore, rather than simply indicating good and bad options, VP can reliably report the relative value of multiple outcomes in a complex reward space.

2.4 Discussion

Our data here demonstrate that neurons in both NAc and VP fire in a reward-selective manner, but this reward-specific firing is much more prevalent in the VP neural population. This relation is evident in both the larger number of neurons in VP that fire selectively for sucrose and maltodextrin, as well as the greater decoding accuracy of LDA models trained on the spiking data of neurons in VP. Moreover, both the onset and peak of reward-specific information in VP precedes those in NAc. We also found that neurons in VP tracked the relative value of the reward outcomes across three different conditions: on a trial-by-trial basis in sessions contrasting sucrose and maltodextrin, in a new session replacing sucrose with water where maltodextrin became the preferred outcome, and, finally, in sessions with all three outcomes. Thus, our data demonstrate a robust reward valuation signal in VP that is unlikely to be fully explained by its classical NAc input.

2.4.1 Reward-selective Encoding Despite Highly Controlled Stimuli

Previous work has shown that neural responses in NAc for orally consumed rewards and their predictive stimuli are modulated by the location (Robinson and Carelli, 2008), motor response (Roitman et al., 2005), size (Bissonette et al., 2013; Cooch et al., 2015; Goldstein et al., 2012; Roesch et al., 2009; Webber et al., 2016), and concentration (Taha and Fields, 2005; Wheeler et al., 2005; Villavicencio et al., 2018) of the reward outcomes. In VP, reward-related neural responses are known to be modulated by reward size (Avila and Lin, 2014a,a; Stephenson-Jones et al., 2020) and the rat's physiological need for a given reward (Tindell et al., 2006, 2009). Here, we controlled for all of these factors by choosing two reward solutions (sucrose and maltodextrin) with equivalent caloric value that were delivered in the same location and elicited nearly identical motor responses in rats in a normal physiological state. Thus, aside from their chemosensory properties, the two rewards differed only in the rats' preference for each, suggesting that the reward selectivity reported here was based on

preference or identity. Because the dominant response in both regions was greater firing rate for the preferred reward, sucrose (Fig. 2.6e,k), it is likely that preference is the major contributor to the reward-selective responses we observed in each region. If NAc and VP neural activity primarily coded reward identity, we would expect equivalent numbers of biased responses for each reward, along with greater rigidity of reward-specific coding in VP across changing reward contexts, two conditions that were not met. Still, the existence of a small proportion of cells in both regions with greater firing for maltodextrin (Fig. 2.6h,n) could be indicative of the presence of some identity-based encoding.

2.4.2 VP As a Reward Processor Independent of NAc

Due to its well-defined anatomical role as an output of NAc within the ventral striatopallidal pathway (Alexander et al., 1986; Groenewegen et al., 1999; de Olmos and Heimer, 1999), most studies on the role of VP in reward processing have been within the context of NAc function. Such experiments have established an important role for this pathway in reward-related behavior (Chang et al., 2018; Creed et al., 2016; Gallo et al., 2018; Leung and Balleine, 2013; Richard et al., 2013a; Smith and Berridge, 2007); however, because these studies use longer timescale manipulations, they do not clarify how reward-related information arrives in each region and when NAc and VP connectivity is necessary for proper reward processing, questions readily answered with in vivo observations of neural activity in each region during a reward-processing task.

A recent study of NAc and VP activity in vivo found that the onset of VP excitatory neural responses to a cue indicating reward availability typically precedes the onset of cue-evoked neural responses in NAc, demonstrating that NAc cannot be the primary source of excitatory VP cue responses and, therefore, that VP does not act exclusively downstream of NAc in the processing of cues predicting reward (Richard et al., 2016). Likewise, in the present work, we found that reward-specific information arises and peaks earlier in VP than in NAc. This reward-specific information is largely contained within phasic excitations to

the preferred reward; therefore, given that NAc inputs to VP are predominantly inhibitory (or produce a biphasic response) (Root et al., 2015; Chrobak and Napier, 1993; Hakan et al., 1992; Maurice et al., 1997; Mitrovic and Napier, 1998), it is unlikely that this reward-specific excitation originates in NAc. Nevertheless, our data do not exclude the possibility that certain aspects of NAc activity, such as the inhibitions we observe around the time of port entry and reward delivery (Fig. 2.4b,c), are permissive of the reward-selective responses in VP, which do not arise until 0.4s following reward delivery (Figs. 2.6b,2.8a); additionally, later-occurring inhibitions to sucrose in VP (Fig. 2.6b,n,o) could originate from earlier sucrose-specific excitations in NAc (Fig. 2.6a,e,f). Together, our findings support the notion that VP processes certain aspects of reward independently of NAc, and they highlight the importance of studying other inputs to VP, such as amygdala (Mitrovic and Napier, 1998; Maslowski-Cobuzzi and Napier, 1994), lateral hypothalamus (Baldo et al., 2003; Ho and Berridge, 2013), and prefrontal cortex, which, in addition to direct projections (Zaborszky et al., 1997), could provide input via the subthalamic nucleus (Turner et al., 2001; Maurice et al., 1999), a route that is reported to be faster than through striatum (Nambu et al., 2002; Ryan and Clark, 1991).

One caveat to our conclusions is that the neural data from each region were collected from separate animals. This approach introduces the possibility that variations in each subject's task performance and reward preference and subtle changes in the experimental conditions could contribute to the differences observed between these two groups. We were careful to control for as many of these elements as possible, but future recordings performed in the same animal would provide definitive evidence that reward-specific information arises in VP prior to NAc and features more prominently in the VP neural population.

In our recordings, we sampled a large proportion of the anterior-posterior extent of medial NAc shell and core and the majority of the anterior-posterior and medial-lateral axes of VP (Fig. 2.2). Despite previous evidence in NAc and VP for subregion heterogeneity in reward-related function (Root et al., 2015; Smith et al., 2009; Richard et al., 2016; Castro and

Berridge, 2014; Kelley, 1999; Pecina and Berridge, 2005; Richard et al., 2013b; Smith and Berridge, 2005), we saw no meaningful differences in reward selectivity across our recorded location (Figs. 2.2, 2.7), which is consistent with a previous report of uniformly distributed relative value responses in NAc (Taha and Fields, 2005), although high density recordings in NAc and VP subregions are required to make definitive conclusions. Given the current data, our observations on the timing and magnitude of reward-selective signaling in NAc and VP appear to hold true across subregions in both structures, but the data do not preclude differences in lateral NAc shell and more rostral portions of ventrolateral VP, which we did not record from in our study, nor do they preclude different functions for a relative value signal dependent on local and long-range connectivity.

2.4.3 A Relative Value Signal in VP

Our data show that VP neurons can flexibly signal a reward's value relative to the other currently available outcomes. A similar scheme has been shown for a small fraction of reward-selective neurons in NAc by varying the concentrations of available sucrose solutions (Taha and Fields, 2005; Wheeler et al., 2005) or the magnitude of reward (Webber et al., 2016). Previous work has shown that VP can signal differences in value based on size (Avila and Lin, 2014a,a), physiological need (Tindell et al., 2006, 2009), and associative learning (Itoga et al., 2016). Of particular interest is the finding that the VP neural responses to heavily salinated water (normally an aversive stimulus) is greater than that of sucrose when rats are salt-deprived (Tindell et al., 2006); however, in that study, there was no significant reduction in firing for sucrose once it became the less preferred reward, perhaps because salt water and sucrose were administered in separate blocks, hindering a direct comparison. In our experiments, we have shown that the VP neural response to the same reward (maltodextrin) in the same physiological state is altered when that reward's value relative to the other available outcomes changes (Fig. 2.11), the hallmark of a relative value signal. The robustness of this signal across the population invites consideration of the (to

our knowledge, unexplored) role of ventral pallidum in the contrast effect (Flaherty, 1999). Despite multiple demonstrations of neural correlates of negative and positive contrast in both rat and primate NAc (Taha and Fields, 2005; Webber et al., 2016; Cromwell et al., 2005; Wheeler et al., 2005), NAc lesions affect only instrumental but not consummatory contrast effects (Eagle et al., 1999; Leszczuk and Flaherty, 2000); the strong relative value signal in VP makes it an appealing candidate to contribute to both of these effects.

Figure 2.1

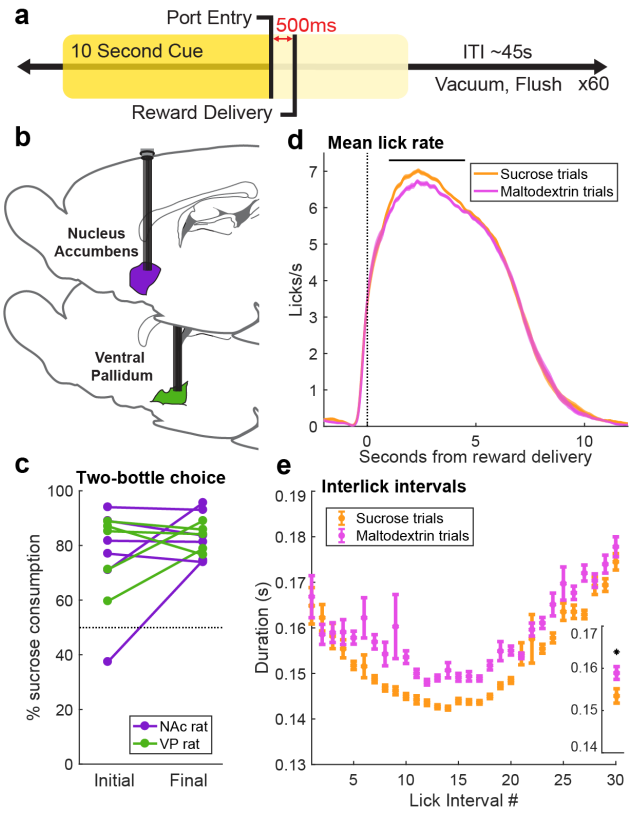

Figure 2.1. Experimental design and reward preference results.

(a) Sucrose or maltodextrin solution was delivered 500ms following rats' entry into the reward port during a 10s white noise cue. Trials of each reward were randomly interspersed throughout the session such that reward identity was unpredictable to the rat.

(b) After training, drivable 16 electrode arrays were implanted in either nucleus accumbens (n=6) or ventral pallidum (n=5).

(c) Rats' preference (percentage sucrose consumption of total consumption) during 1hr free access to 10% solutions of sucrose and maltodextrin. Tests were after surgical recovery (Initial) and after final session with sucrose and maltodextrin (Final).

(d) Average lick rate on sucrose (orange) and maltodextrin (pink) trials during the task. Shading is SEM. Black bar indicates greater number of licks on sucrose trials 1-4.5s post reward delivery ($F(1,3142) = 66.0$, $p = 5.3E\text{-}6$). See also Fig. 2.3.

(e) Interlick interval duration following the first 30 licks on sucrose (orange) and maltodextrin (pink) trials. Inset: mean interlick interval duration across all 30 intervals. Asterisk indicates significant main effect of reward on duration ($F(1,3000) = 33.3$, $p = 0.000084$).

Figure 2.2

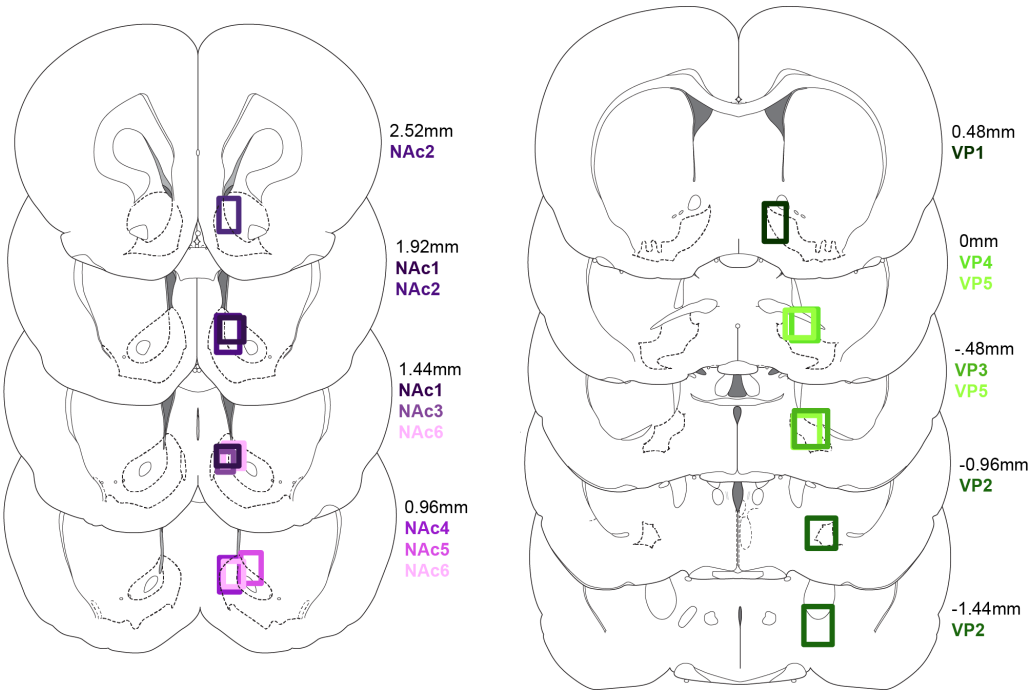

Figure 2.2. Recording locations.

(a) Dashed lines demarcate nucleus accumbens shell and core and ventral pallidum. Placements are color coded by rat. The posterior portion of VP2's placement included extended amygdala.

Figure 2.3

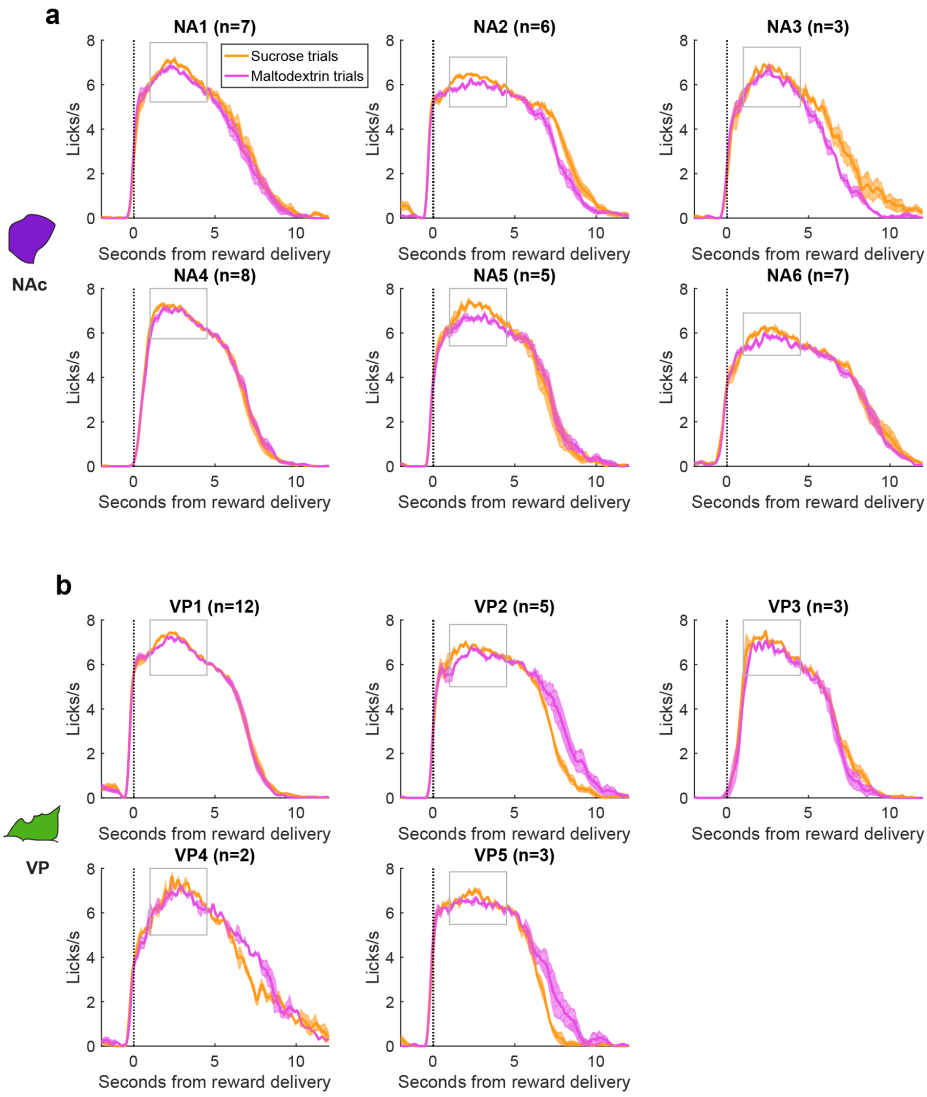

Figure 2.3. Elevated licking for sucrose across all rats.

(a) Mean lick rate for each individual NAc rat on sucrose (orange) and maltodextrin (pink) trials (n is each included session). Gray box indicates time of interest (1-4.5s following reward delivery) when licking is consistently greater for sucrose than for maltodextrin.

(b) As in (a), for VP rats.

Figure 2.4

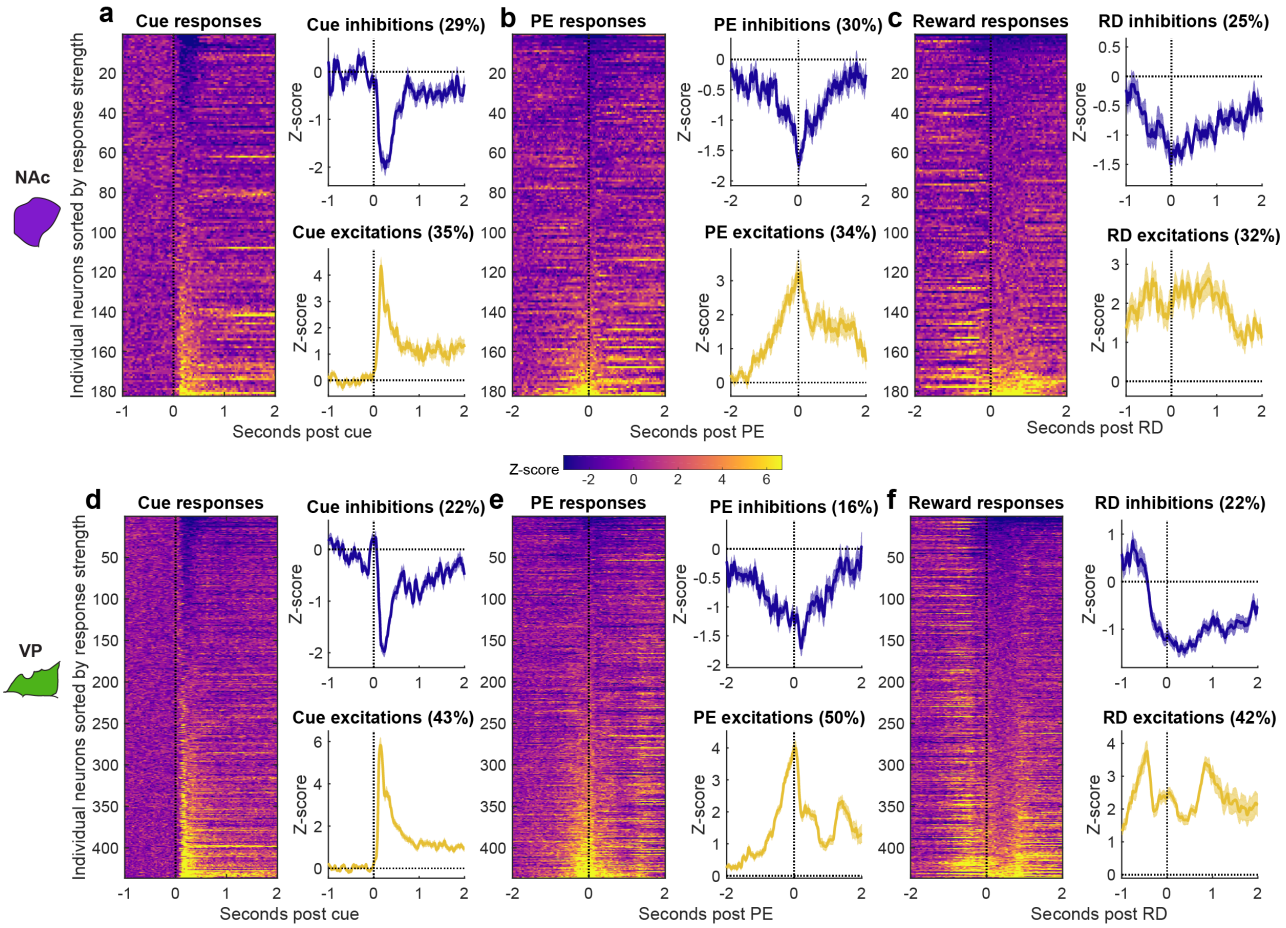

Figure 2.4. Event-evoked responses in NAc and VP.

(a) Left: All NAc neurons sorted by firing rate 500ms following cue onset. Right: Neurons significantly inhibited (blue) and excited (yellow) by cue onset. Shading is SEM.

(b) As in (a), for the 1000ms centered on port entry (PE).

(c) As in (a), for the 1000ms following reward delivery (RD).

(d-f) As in (a-c), for neurons from VP.

Figure 2.5

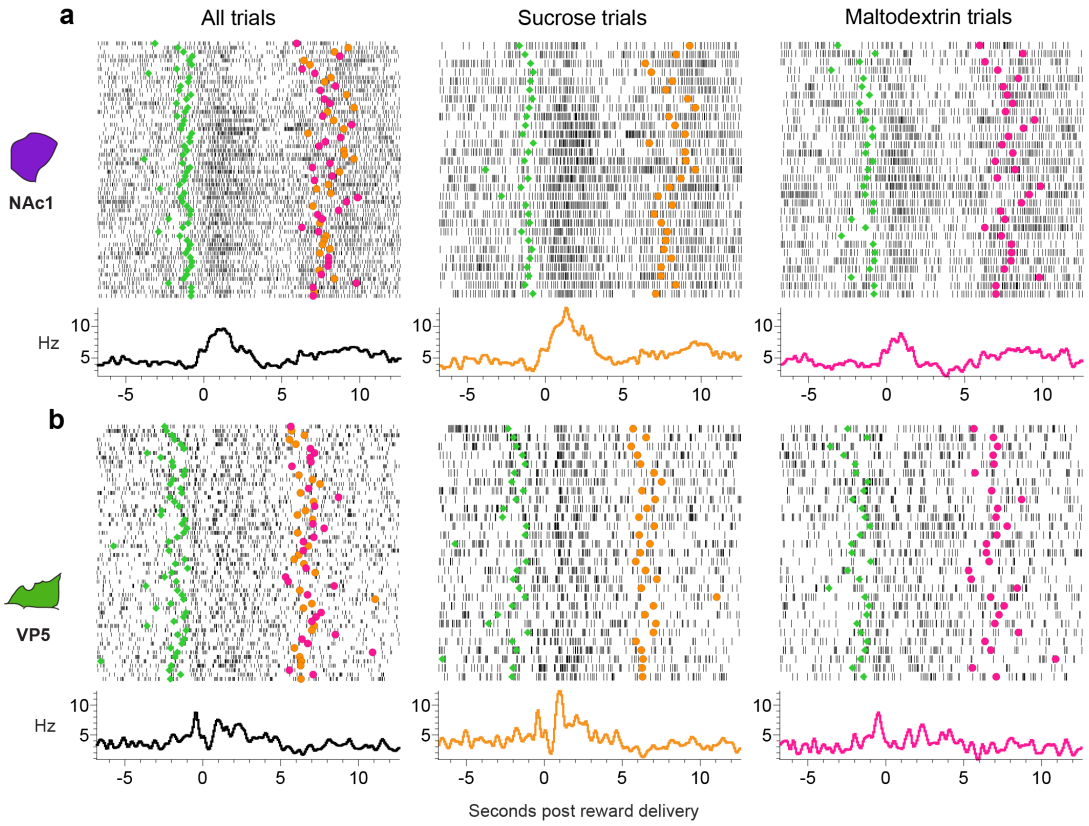

Figure 2.5. Example reward-selective neurons.

(a) Perievent raster (top) and histogram (bottom) of a reward-selective neuron from NAc1 aligned to reward delivery for all (left), sucrose (center) and maltodextrin (right) trials. Green diamond is cue, orange circle is final sucrose lick, pink circle is final maltodextrin lick. Perievent histogram constructed with 75ms bins and smoothed with a Gaussian filter over 3 bins.

(b) As in (a), for a reward-selective neuron from VP5.

Figure 2.6

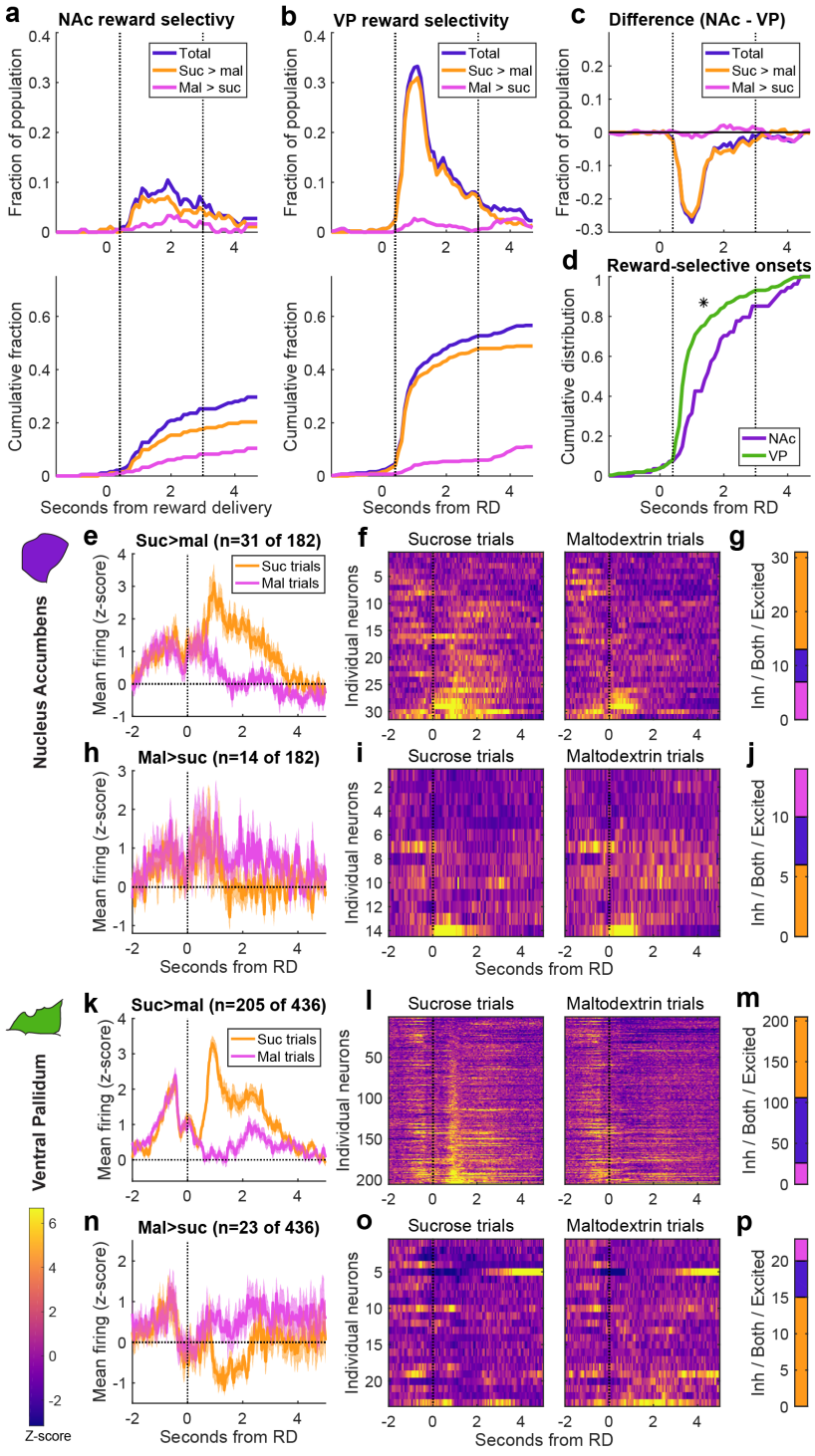

Figure 2.6. More neurons in VP fire selectively for sucrose and maltodextrin than in NAc.

Figure 2.6. More neurons in VP fire selectively for sucrose and maltodextrin than in NAc.

(a) Top panel: fraction of NAc neurons meeting criteria for reward selectivity as a function of time after reward delivery. Plotted are total fraction of reward-selective neurons (blue) and, of those, neurons with greater firing for sucrose (orange) and greater firing for maltodextrin (pink). Bottom panel: Cumulative distribution of reward selectivity over time after reward delivery.

(b) Same as (a), for VP.

(c) Subtraction of VP reward selectivity from NAc in each bin. Negative values indicate more selectivity in VP.

(d) Cumulative distribution of reward selectivity onsets as a fraction of total reward-selective neurons. Asterisk indicates significantly earlier onsets in VP ($F(1,290) = 12.7$, $p = 0.00071$).

(e) Mean normalized firing rate for sucrose-selective neurons (neurons with greater firing for sucrose in any bin centered at 0.4-3s) on sucrose (orange) and maltodextrin (pink) trials. Shading is SEM.

(f) Heat maps of the normalized activity of individual sucrose-selective neurons on sucrose and maltodextrin trials.

(g) Number of neurons with maltodextrin inhibitions (pink), sucrose excitations (orange), or both (blue).

(h) Mean normalized firing rate for maltodextrin-selective neurons (neurons with greater firing for maltodextrin in any bin centered at 0.4-3s) on sucrose (orange) and maltodextrin (pink) trials. Shading is SEM.

(i) Heat maps of the normalized activity of individual maltodextrin-selective neurons on sucrose and maltodextrin trials.

(j) Number of neurons with maltodextrin inhibitions (pink), sucrose excitations (orange), or both (blue).

(k-p) As in (e-j), for VP neurons.

Figure 2.7

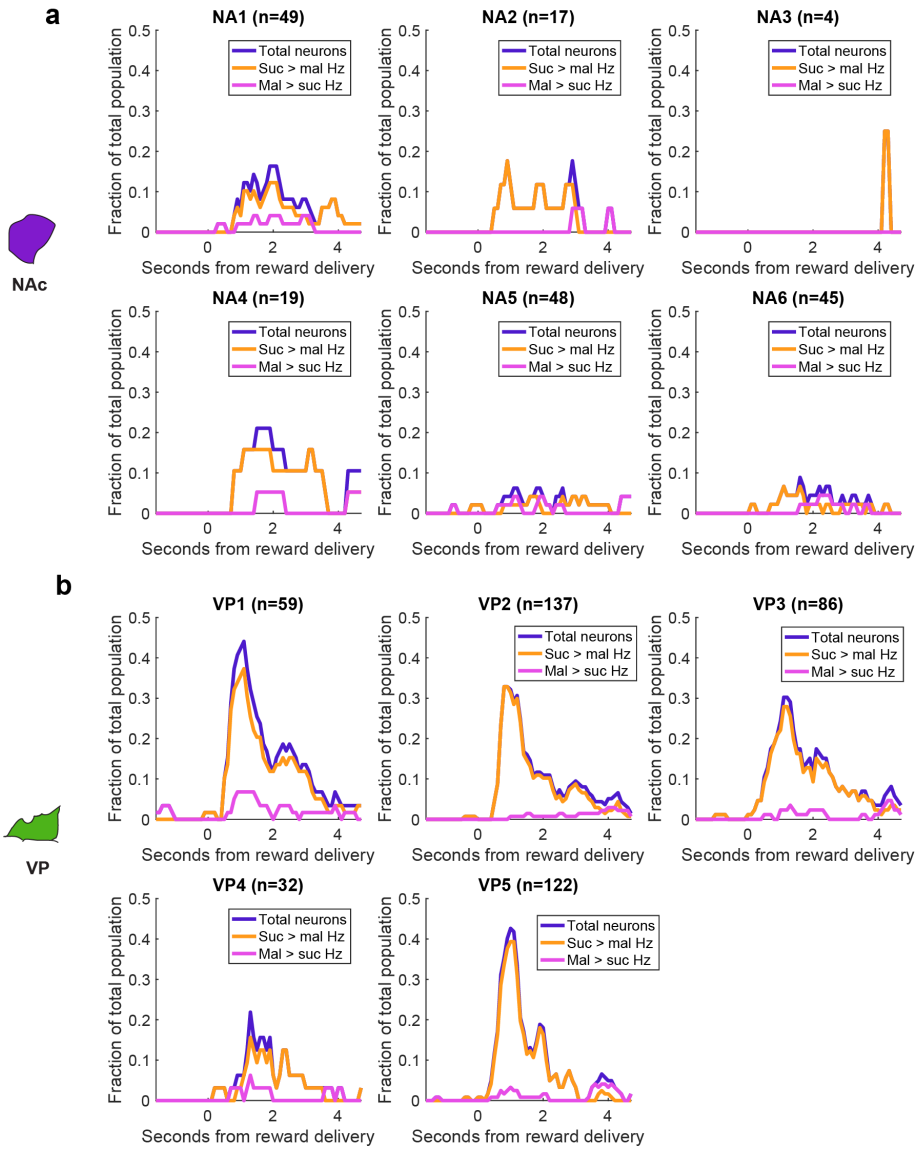

Figure 2.7. Reward-selective neurons in each rat.

(a) Histogram of the fraction of neurons in each NAc rat that meet criteria for reward selectivity in overlapping 600ms bins (advanced by 100ms) (see Figure 2a). Plotted are the total fraction of reward-selective neurons (blue) and, of those, neurons with greater firing for sucrose (orange) and greater firing for maltodextrin (pink).

(b) As in (a), for VP rats.

Figure 2.8

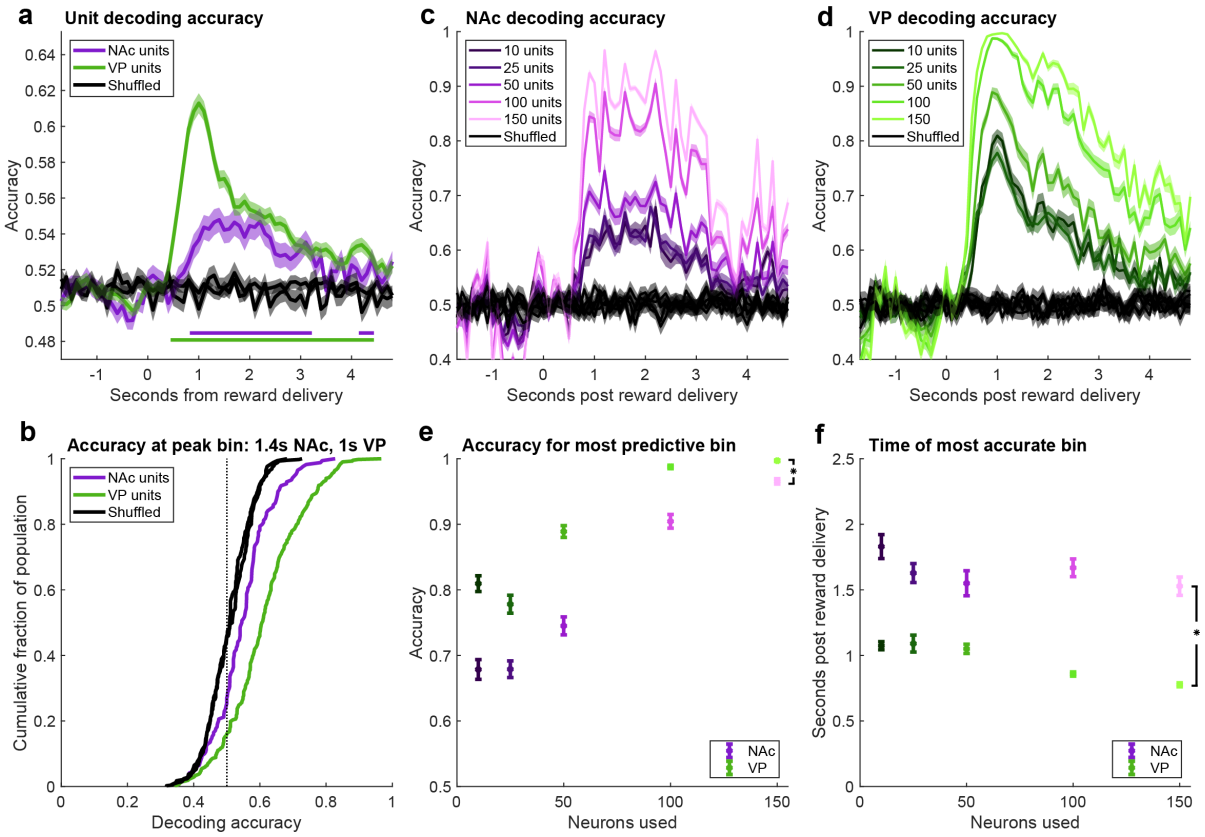

Figure 2.8. VP activity decodes trial identity earlier and more accurately than NAc activity.

Figure 2.8. VP activity decodes trial identity earlier and more accurately than NAc activity.

(a) Average cross-validated decoding across 600ms overlapping bins relative to reward delivery. Decoding accuracy for NAc (purple), VP (green), and data with shuffled trial identity from each region (black). Shading is SEM. Purple (NAc) and green (VP) lines indicate consecutive bins where accuracy exceeds 99% confidence interval of corresponding shuffled data.

(b) Cumulative distribution of accuracies in the bin with the greatest average accuracy in each region (centered at 1.4s in NAc and 1s in VP) and the corresponding shuffled data from that bin in each region.

(c) Average cross-validated decoding accuracy relative to reward delivery time of linear discriminant analysis models trained on spiking data of 20 randomly selected groups of 10, 25, 50, 100, or 150 neurons in NAc and corresponding models trained on data with trial identity shuffled. Shading is SEM.

(d) Same as (c) for VP pseudoensemble models.

(e) Average accuracy of each replicate for the bin with peak accuracy for each pseudoensemble size in each region. Asterisk indicates significant main effect of region on accuracy ($F(1,490) = 212$, $p = 3.3E\text{-}40$).

(f) Average peak accuracy time post-reward for each replicate of each pseudoensemble size in each region. Asterisk indicates significant main effect of region on peak accuracy time ($F(1,490) = 289$, $p = 2.5E\text{-}51$).

Figure 2.9

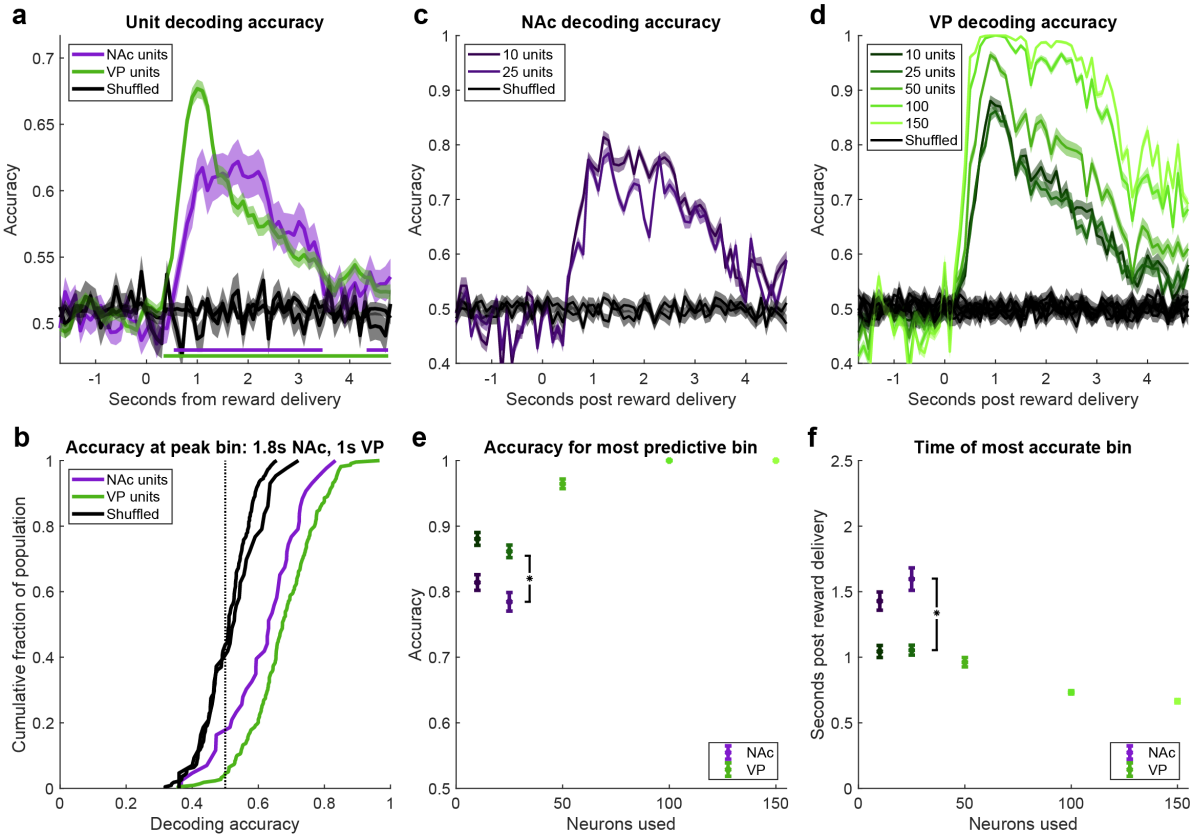

Figure 2.9. Decoding trial identity with only reward-selective neurons.

Figure 2.9. Decoding trial identity with only reward-selective neurons.

(a) Average cross-validated decoding using individual reward-selective neurons (from Fig. 2.6) across 600ms overlapping bins. Decoding accuracy for NAc (purple), VP (green), and data with shuffled trial identity from each region (black). Shading is SEM. Purple (NAc) and green (VP) lines indicate consecutive bins where accuracy exceeds 99% confidence interval of corresponding shuffled data.

(b) Cumulative distribution of decoding accuracies in the bin with the greatest average accuracy in each region (centered at 1.6s in NAc and 1s in VP) and the corresponding shuffled data.

(c) Average cross-validated decoding accuracy relative to reward delivery time of linear discriminant analysis models trained on spiking data of 20 randomly selected groups of 10 or 25 reward-selective neurons in NAc and corresponding models trained on data with trial identity shuffled. Shading is SEM.

(d) Same as (c), but for for VP reward-selective neurons.

(e) Average accuracy of each replicate for the bin with peak accuracy for each pseudoensemble size in each region. Asterisk indicates significant main effect of region on accuracy for 10 and 25 neuron ensembles ($F(1,196) = 38.9$, $p = 2.7E\text{-}9$).

(f) Average peak accuracy time post-reward for each replicate of each pseudoensemble size in each region. Asterisk indicates significant main effect of region on time of peak accuracy for 10 and 25 neuron ensembles ($F(1,196) = 54.8$, $p = 3.8E\text{-}12$).

Figure 2.10

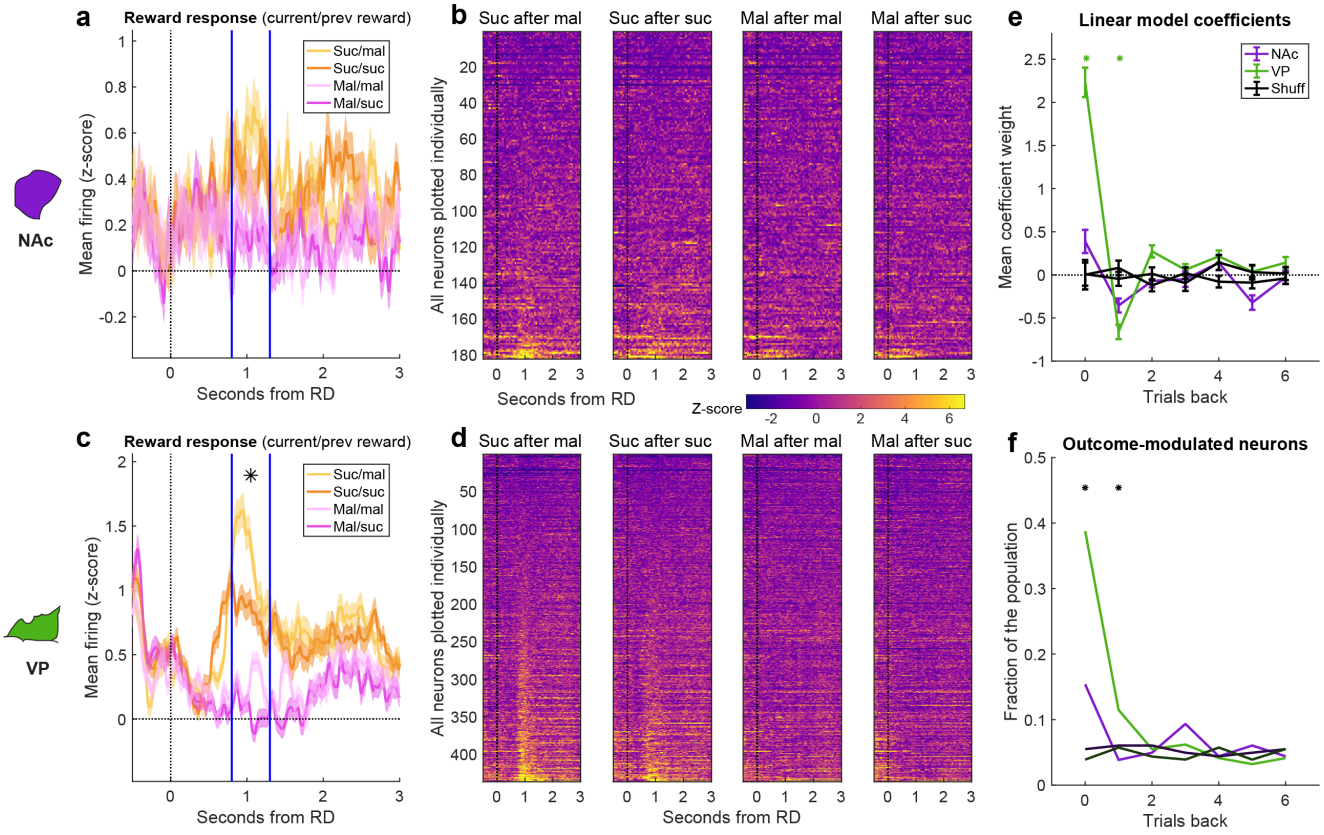

Figure 2.10. Previous reward outcome impacts current reward firing.

(a) Normalized activity of all neurons in NAc on sucrose (orange) and maltodextrin (pink) trials of each reward, separated by reward outcome on preceding trial; darker lines indicate that sucrose was the prior trial's reward.

(b) Normalized reward-related activity of every individual neuron in NAc on trials with each combination of previous and current reward.

(c) Same as (a), for VP. Asterisk indicates a significant main effect of previous reward on normalizing firing rate in VP ($F(1,1724) = 10.1$, $p = 0.022$) between the vertical blue lines.

(d) Same as (b), for VP.

(e) Mean coefficient weights for the impact of the current and previous 6 trials on normalized firing rate in the same epoch as (a,c) for each neuron in NAc (purple), VP (green), and corresponding data for each neuron with the outcomes shuffled (black) for each region. Error bars are SEM. Asterisks are $p < 0.05$ for Tukey tests comparing VP coefficients to shuffled data, corrected for multiple comparisons.

(f) Proportion of the neural populations in VP (green), NAc (purple), and corresponding shuffled neurons (black) with significant coefficients for each of the relative trials. Asterisks represent $p < 0.05$ for chi-square tests on both the distribution of neurons across all four conditions (true and shuffled data from each region) and across the true data from each region.

Figure 2.11

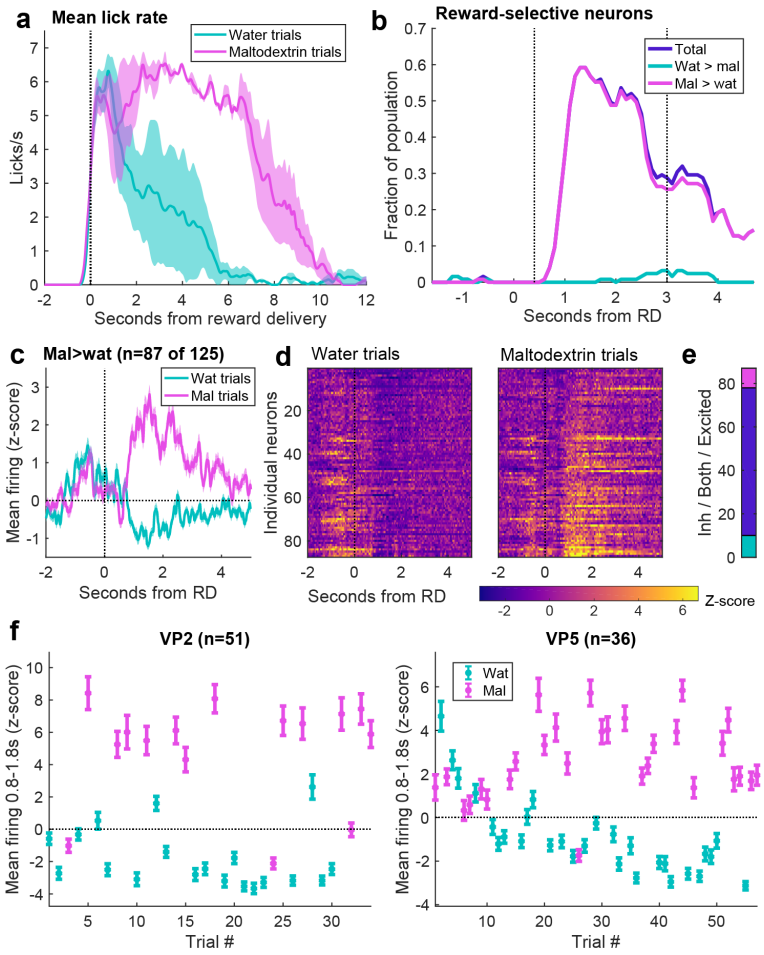

Figure 2.11. VP reward-selective activity adjusts to reflect relative value of new outcomes.

(a) Lick rate on water (blue) and maltodextrin (pink) trials. Shading is SEM.

(b) Fraction of VP neurons that meet criteria for reward selectivity relative to reward delivery time: all reward-selective neurons (dark blue) and, of those, neurons with greater firing for maltodextrin (pink) and greater firing for water (light blue). Dashed lines indicate the window (0.4-3s) for which reward-selective neurons were selected for (c) and (d).

(c) Average normalized firing rate for neurons with greater firing for maltodextrin on water (blue) and maltodextrin (pink) trials. Shading is SEM.

(d) Heat maps of the normalized activity of individual neurons on water and maltodextrin trials.

(e) Number of neurons with maltodextrin excitations (pink), water inhibitions (light blue), or both (dark blue).

(f) Emergence of maltodextrin (pink) excitations and water (blue) inhibitions among reward-selective neurons across each completed trial of the session. Plotted as mean normalized activity 0.8-1.8s post reward delivery; error bars are SEM.

Figure 2.12

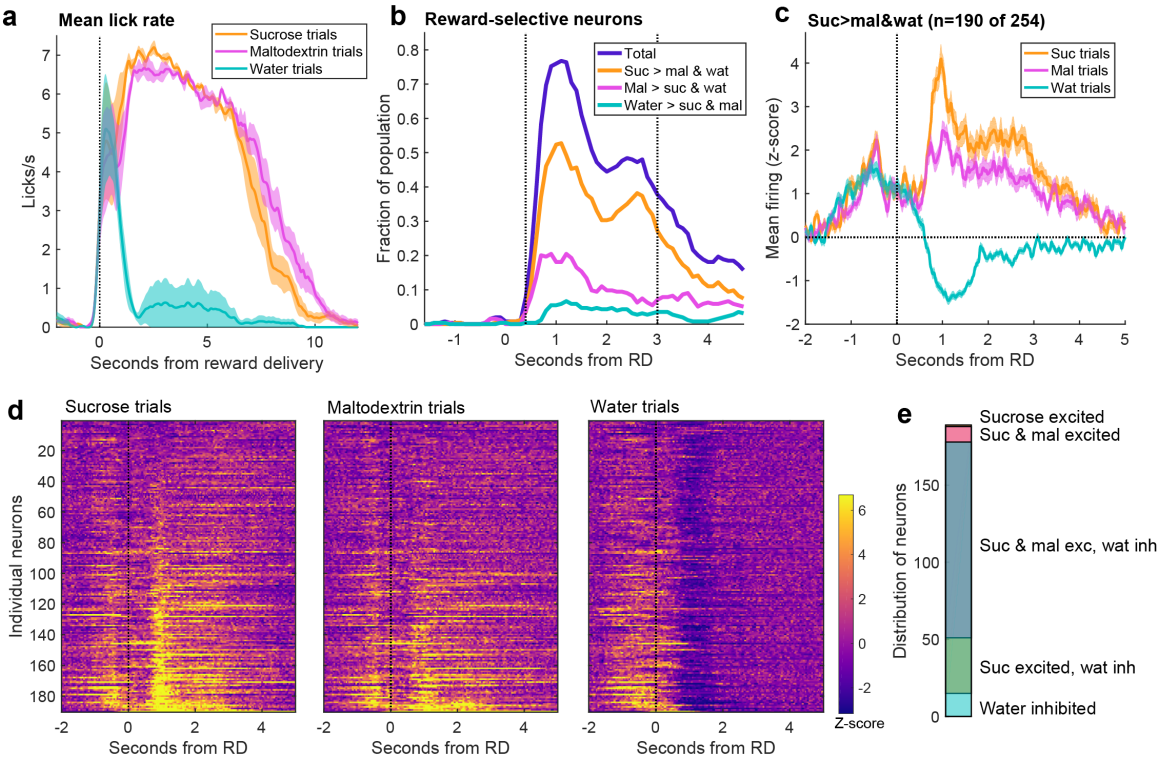

Figure 2.12. VP neurons report the relative value of three reward outcomes.

(a) Lick rate on sucrose (orange), maltodextrin (pink), and water (blue) trials. Shading is SEM.

(b) Fraction of neurons in VP that meet criteria for reward selectivity relative to reward delivery time: all reward-selective neurons (dark blue) and, of those, neurons with greatest firing for sucrose (orange), maltodextrin (pink), or water (light blue). Dashed lines indicate the window (0.4-3s) for which reward-selective neurons were selected for (c-e).

(c) Mean normalized firing rate of neurons that are reward-selective for any bin 0.4-3s post reward delivery and have greatest firing rate for sucrose. Shading is SEM.

(d) Normalized firing rate of individual neurons included in (c). Neurons in all three plots are sorted by amount of firing on sucrose trials in the bin with the most number of neurons with greatest firing for sucrose.

(e) Distribution of neurons in (c) and (d) according to sucrose excitation, maltodextrin excitation, and water inhibition.

Chapter 3

A quantitative reward prediction error signal in ventral pallidum

This chapter is adapted from Ottenheimer et al. (2019a).

3.1 Introduction

In Chapter 2, we described a flexible relative value signal in ventral pallidum (VP). We observed this both in the ability of the signal to rescale to reflect the relative values of a new set of rewards (Figs. 2.11, 2.12) and a more local adjustment in response to the previously received reward (Fig. 2.10). Flexible representations of the reward landscape are a key component of reinforcement learning, a well-established approach for describing how individuals interact with environments to maximize reward (Sutton and Barto, 1998). Reinforcement learning frameworks formalize the notion that individuals integrate information about past rewards to make predictions about the future. Deviations from these predictions, known as reward prediction errors (RPEs), are used to iteratively update future predictions (Rescorla and Wagner, 1972).

One remarkable extension of reinforcement learning to neuroscience was the discovery that midbrain dopamine neurons encode RPEs (Schultz et al., 1997) and do so over local timescales (Bayer and Glimcher, 2005). Despite the influence of the discovery of dopamine neuron RPE signaling, little is known about how related brain regions contribute to the

calculation of RPEs to drive learning. VP has dense reciprocal connectivity with dopamine neurons in the ventral tegmental area (Watabe-Uchida et al., 2012; Beier et al., 2015; Root et al., 2015; Tian et al., 2016; Faget et al., 2018). Recent studies show some evidence for RPE encoding in VP (Tian et al., 2016; Stephenson-Jones et al., 2020). With this collection of results in mind, we reexamined the data from Chapter 1 through the lens of RPE signaling. We also conducted additional electrophysiology studies to test the adherence of VP activity to an RPE signal in different scenarios. By adapting and fitting computational models to predict the spike counts of individual neurons, we were able to demonstrate that an RPE model predicts key features of VP neural activity. We also found that the activity of VP outcome-sensitive cells predicted subsequent task engagement, and optogenetic manipulation of VP altered task engagement in a manner consistent with the model, providing a possible explanation for the purpose of this signal.

3.2 Materials and Methods

Behavioral tasks.

In addition to the data from the tasks in Chapter 2 (which here are called **random sucrose/maltodextrin** and **random sucrose/maltodextrin/water**), this chapter includes data from two additional tasks. **Blocked sucrose/maltodextrin:** For the same group of rats as the random sucrose/maltodextrin task, additional blocked sessions were performed on roughly alternating days with the random sucrose/maltodextrin task. In blocked sessions, sucrose and maltodextrin were presented 30 trials in a row for a total of 60 trials. The order of the rewards switched each blocked session. **Predictable and random sucrose/maltodextrin:** A new group of rats (n=4) was trained on a task with the same trial structure (with a shorter ITI of 30s) but with three possible auditory cues. One predicted sucrose delivery with 100% probability (30 trials), one maltodextrin with 100% probability (30 trials), and one, as in the random sucrose/maltodextrin task, predicted each reward with

a 50% probability (60 trials). For these rats, we maintained the electrode wires in the same position for the duration of the experiment. Each wire from these rats only contributed to the included dataset once.

Optogenetic manipulations.

Note: the rats from the inhibition and excitation experiments here performed the respective experiments from Chapter 4 first. **Inhibition.** For this experiment, rats were trained on a variation of the random sucrose/maltodextrin task where instead of maltodextrin delivery, rats received sucrose + bilateral continuous (5 sec, 15-20 mW) photoinhibition of VP. For these sessions the reward volume was reduced to 55μL and the total number of trials was increased to 90. In our analysis, we only included rats who completed at least 30 trials and had both fibers and viral expression in VP. This resulted in 7 rats in each group: 4 males and 3 females in the ArchT3.0 group, and 2 males and 5 females in the YFP group. **Excitation.** For rats with ChR2, on half of trials, rats received sucrose + unilateral 40Hz pulsed photoexcitation of VP for 2 sec (10ms pulse width, 10-12mW). Because rats were implanted bilaterally, we stimulated the side with maximal effect and minimal off-target effects as determined in prior experiments. We only included rats who completed at least 30 trials and had their stimulated fiber and viral expression in VP. This resulted in 5 males and 5 females in the ChR2 group, and 3 males and 4 females in the GFP group.

PSTH creation.

Peri-stimulus time histograms (PSTHs) were constructed using 0.01ms bins surrounding the event of interest (generally, reward delivery). PSTHs were smoothed using a half-normal filter ($\sigma = 6.6$) that only used activity in previous, but not upcoming, bins. Each bin of the PSTH was z-scored by subtracting the mean firing rate across 10s windows before each trial and dividing by the standard deviation across those windows (n = no. of trials). PSTHs for licking were created in the same manner (without z-scoring) using 0.05ms bins and $\sigma = 8$.

Model fitting.

For each neuron, we took the spike count, $s(t)$, within the 0.75-1.95s post-reward delivery time bin for each trial and fit Poisson spike count models. For the random and blocked sucrose/maltodextrin tasks, we fit the following three models.

RPE model

$$\delta(t) = o(t) - V(t)$$
$$V(t+1) = V(t) + \alpha \cdot \delta(t)$$
$$s(t) \sim \text{Poisson}(\exp(a \cdot \delta(t) + b))$$

where $V(t)$ is the expected value, $\delta(t)$ is the RPE, $o(t)$ is the outcome and α is the learning rate. For the tasks with sucrose and maltodextrin outcomes, we coded $o(t) = 0$ for maltodextrin, and 1 for sucrose. For the tasks with sucrose, maltodextrin, and water outcomes, we coded $o(t) = 0$ for water, 1 for sucrose, and ρ for maltodextrin, a free parameter we estimated during model fitting. To map RPEs to spike counts, we used a as a slope (gain) and b as an intercept (offset) parameter. This affine-transformed RPE was mapped through an exponential function, to avoid negative values, and used as the rate parameter for a Poisson distribution.

Current outcome model

$$s(t) \sim \text{Poisson}(\exp(a \cdot o(t) + b))$$

Unmodulated model

$$s(t) \sim \text{Poisson}(\exp(\ln(\overline{s})))$$

where $\overline{s}$ is the mean firing rate.

For the predictable and random sucrose/maltodextrin task, we added the following three models

RPE + cue model

$$\delta(t) = o(t) - V(t)$$

$$V(t+1) = V(t) + \alpha \cdot \delta(t)$$

If a sucrose-predicting cue was given

$$s(t) \sim \text{Poisson}(\exp(a \cdot \delta(t) + b - V_{sucrose}))$$

If a maltodextrin-predicting cue was given

$$s(t) \sim \text{Poisson}(\exp(a \cdot \delta(t) + b - V_{maltodextrin}))$$

If a non-predictive cue was given

$$s(t) \sim \text{Poisson}(\exp(a \cdot \delta(t) + b))$$

where $V_{sucrose}$ and $V_{maltodextrin}$ are free parameters for the values of the sucrose- and maltodextrin-predicting cues, respectively.

Current outcome + cue model

If a sucrose-predicting cue was given

$$s(t) \sim \text{Poisson}(\exp(a \cdot o(t) + b - V_{sucrose}))$$

If a maltodextrin-predicting cue was given

$$s(t) \sim \text{Poisson}(\exp(a \cdot o(t) + b - V_{maltodextrin}))$$

If a non-predictive cue was given

$$s(t) \sim \text{Poisson}(\exp(a \cdot o(t) + b))$$

Unmodulated + cue model

If a sucrose-predicting cue was given

$$s(t) \sim \text{Poisson}(\exp(\ln(\overline{s}) - V_{sucrose}))$$

If a maltodextrin-predicting cue was given

$$s(t) \sim \text{Poisson}(\exp(\ln(\overline{s}) - V_{maltodextrin}))$$

If a non-predictive cue was given

$$s(t) \sim \text{Poisson}(\exp(\ln(\overline{s}))$$

To estimate predictive cue effects on firing at the time of the cue, we fit the *Unmodulated model* and *Unmodulated + cue model*, with $V_{sucrose}$ and $V_{maltodextrin}$ sign-flipped.

We also considered RPE models in which the predictive cue allowed for partial to full cancellation of RPEs.

If a sucrose-predicting cue was given

$$\eta(t) = o(t) - ((1 - w) \cdot V(t) + w \cdot V_{sucrose})$$

If a maltodextrin-predicting cue was given

$$\eta(t) = o(t) - ((1 - w) \cdot V(t) + w \cdot V_{maltodextrin})$$

If a non-predictive cue was given

$$\eta(t) = o(t) - V(t)$$

$$s(t) \sim \text{Poisson}(\exp(a \cdot \eta(t) + b))$$

We fixed $V_{sucrose} = 1$ and $V_{maltodextrin} = 0$ and set w as a free parameter. If $w = 0$, this is equivalent to the *RPE model*, and if $w = 1$, the predictive cues allow for full cancellation of the RPE ($\eta(t) = 0$). Intermediate values of w allow the predictive cues to partially cancel the history-based RPE. This model was best for a negligible number of neurons.

We only analyzed trials in which the rat licked within the first two seconds of reward delivery, to ensure that they sampled the outcome. For all RPE models, $V(1)$ was initialized to 0.5. For all models with a slope parameter, we constrained the slope, a, to be > 0, as our work in Chapter 2 found only a trivial fraction of VP neurons preferentially encoding maltodextrin. We found maximum likelihood estimates for each model and selected the best model using Akaike information criterion (lower AIC indicates a better fit, after taking into account the number of parameters). We used 10 randomly-selected starting initial values for each parameter to avoid finding local minima.

Correlation and RPE tuning curves for real and simulated neurons.

For neurons best fit by the RPE model, we report correlations between real and predicted spike trains, as well as RPE tuning curves for real and predicted spikes. For each neuron, we estimated the Pearson correlation coefficient between real spikes and 501 independent model-generated spike count trains, using parameters estimated from the same neuron, and report the median correlation. The median-correlated spike count is plotted in Figure 3.2e, 3.5i. We also compared the mean and standard deviations of real vs simulated spike counts in Figure X. To generate RPE tuning curves for real spikes, we took z-scored spike counts and binned according to estimated RPEs. We performed this procedure for all RPE neurons and report the average tuning curve. To generate tuning curves for predicted spikes, we simulated spike trains using neuron-derived parameter estimates and followed the same procedure.

Outcome history-based linear regression.

To estimate how the outcome of the current and previous trials affected the firing rate of the current trial, we conducted a complete-pooling linear regression analysis. We z-scored the firing rate of each neuron using the baseline activity across the set of 10s bins prior to each trial and combined the firing rates of all neurons of interest. Similarly, our design matrix included the current and ten previous trial outcomes for all neurons of interest. For the random sucrose/maltodextrin task, we gave maltodextrin a value of 0 and sucrose a value of 1. For the random sucrose/maltodextrin/water task, water was given a value of 0, sucrose was given a value of 1, and maltodextrin was given a value of 0.75 for RPE cells and 0.8 for current outcome cells, the values which achieved the maximum R^2 for the linear regression.

Video analysis.

During recording sessions, videos were taken at 30 frames per second of the rats as they performed the task. During the optogenetic sessions, videos were taken at 6-8 frames per second. These videos permitted analysis of movement around the behavioral chamber. We

used DeepLabCut (Mathis et al., 2018; Nath et al., 2019) in Python to determine the location of the rat's head in each frame. DeepLabCut generates a likelihood for the location of each feature in each frame, and we discarded any frames below 0.95. We further processed the X-coordinate and Y-coordinate traces to remove outliers above 2 standard deviations of the median across moving 1s bins. These traces were used to calculate the location of the rat within a 0.2s window surrounding each cue onset and the locations of the rat in 0.2s bins from the last lick within the first 15s after reward delivery (or 15s even if rats were still licking) until the next cue onset for rats from the recording sessions, or from the final port exit within the first 10s after reward delivery (or 10s even if the rats were still in the port) until the next cue onset for rats from the optogenetic sessions. To find the average distance from the port during this time period, we found the area under the curve for distance from the port and divided by the total time. To compare this measure across sucrose and maltodextrin (or sucrose + laser trials), we found the average across all trials of each type for each rat and compared the two groups with a Wilcoxon signed-rank test. For the electrophysiology experiment, this measure was then correlated (Spearman's) to the activity of each RPE neuron in our bin of interest on each trial. To compare to shuffled data, we produced 1000 correlations for each neuron with shuffled trial order and compared the true mean to the distribution of means from the 1000 shuffled populations. For the optogenetic experiment, we calculated for each rat the fractional change in distance from the port produced by the laser by dividing the difference (laser - no laser) by the no laser value. We compared the values from these two groups with a Wilcoxon rank-sum test.

Evolution of activity across session.

To visualize how the reward-evoked activity of neurons changed across each reward block in the blocked task, we plotted the mean activity within our bin of interest (0.75s-1.95s post-reward delivery) for 5 groups of 3 trials at a time, equally spaced throughout the completed trials of each reward (and applied the same approach to the random sucrose/maltodextrin

task, as well). To assess the impact of session progress on firing rate, we pooled the activity on each trial for all neurons of interest (say, RPE cells in sessions with sucrose block first) and the proportional progress throughout the session (of total completed trials) for the respective trial and performed a linear regression.

Statistical analysis.

Data are presented as mean +/− s.e.m. unless otherwise noted. Statistical analyses were performed in MATLAB (MathWorks) on unsmoothed data. Specific tests are noted in the text, figure legends, and throughout the methods.

3.3 Results

3.3.1 Ventral pallidum neurons signal prediction errors according to reward preference

We began by analyzing the activity of the VP neurons (n = 436) presented in Figs. 2.6-2.10. To recap, on each trial, rats responded to a 10-second white noise cue that indicated the availability of 10% solutions of either sucrose or maltodextrin contingent upon entry into the reward port (Figure 3.1a). In this task ("random sucrose/maltodextrin"), there was only one cue, which predicted sucrose or maltodextrin reward with equal probability. This task design ensured rats could not accurately predict upcoming rewards (Figure 3.1b). In addition to the observation that VP neurons reflected the rats' preference for sucrose with increased firing rates (Figure 3.1e), we also saw that the previous outcome modulated the reward signal in a direction consistent with reward prediction error (RPE) coding (Figure 3.1f). For example, receiving sucrose on the previous trial increased expectation of future sucrose, leading to decreased firing when sucrose was delivered on the current trial. The expected trend held true for all combinations of past/current outcomes, suggesting that VP neural activity might contain an RPE signal.

Intrigued by the possibility of RPE signaling in VP, we expanded upon our prior findings by quantifying the impact of current and previous outcomes on reward-evoked firing in VP. We applied a linear regression that has previously been used to quantify the effect of reward history on dopamine neuron firing (Bayer and Glimcher, 2005). Consistent with our findings in Fig. 2.10, across all neurons, only the current trial and previous trial significantly impacted firing rates at the time of the outcome (Figure 3.1g). While this pattern of regressors is consistent with RPE coding, it is on a much shorter timescale than has been observed for dopamine neurons (Bayer and Glimcher, 2005; Mohebi et al., 2019), and is shorter than typical history effects in other brain regions (Kepecs et al., 2008; Padoa-Schioppa, 2009; Asaad and Eskandar, 2011; Cai and Padoa-Schioppa, 2012). One limitation of our linear regression approach is it assumes that VP is largely homogeneous, which risks introducing bias into coefficient estimates. This leaves open the possibility that VP contains subsets of neurons that encode reward history on a longer timescale.

To identify neurons in VP sensitive to reward history, we developed three models to fit the firing rates of individual neurons, corresponding to three potential patterns of neuronal activity. The first model, 'RPE,' fit spike counts as a function of estimated RPEs (Figure 3.1h). This model generated trial-by-trial value estimates (V) which constituted reward predictions. On each trial, an RPE was generated by the difference between actual and predicted rewards, and this RPE was multiplied by a learning rate (α) before updating V for the next trial (Rescorla and Wagner, 1972). Small values of the learning rate allow for integration of reward history multiple trials into the past. We also fit two additional models to serve as controls, one in which the spike count was determined only by the current outcome ('Current outcome'), and one with no impact of outcome ('Unmodulated'). We used maximum likelihood estimation to fit the models to each neuron and selected the most parsimonious model using the Akaike information criterion (AIC), which selects the best-fit model after penalizing for model complexity. This classification process revealed that 17% of neurons were best described by the RPE model, and another 29% were best fit by the

current outcome only (Fig. 3.1h); notably, of the 47% of neurons we had previously classified as sucrose-preferring (Fig. 2.6), 74% were classified as either RPE or current outcome here, suggesting the modeling approach relatively faithfully captured reward preference-encoding neurons.

We plotted the mean activity of each subset of neurons for each combination of previous and current outcome and found agreement between firing rate and the predictions of each model (Figure 3.1i). We then performed the same reward-history linear regression on each subset of neurons rather than the entire VP population; this revealed an exponential decay-like influence of multiple previous trials on firing of neurons best fit by the RPE model, indicating that VP neurons modulated by reward history were in fact integrating information over a more extended period of time (Figure 3.1j). Indeed, the mean (median) learning rate across all neurons was 0.56 (0.52); this corresponds to an exponential learning process with a half-life of 0.84 (0.94) trials, indicating that neurons accumulate information over ~ 4.22 (4.72) trials to reach a steady-state value estimate. Thus, given the closely matched caloric value and motor responses to each reward, these data indicate that some VP neurons signal a history-based RPE according to reward preference.

3.3.2 VP encodes reward preference RPEs more robustly than nucleus accumbens, a key input structure

We next asked how faithfully VP neurons encoded RPEs. Our fitting procedure allowed us to recover trial-by-trial estimates of RPEs, based on parameter estimates for that individual neuron as well as the outcome history for that session. We found that the activity of both individual neurons (Fig 3.2a) and the average across all RPE neurons (Fig 3.2b) were strongly correlated with model-derived RPEs. Importantly, this approach revealed a finer dynamic range of firing than was revealed by only looking at current and previous outcomes (Fig. 3.1i). We next generated RPE tuning curves for these neurons and, as expected, found a strong monotonic relationship ($t_{3,961} = 40.3$, $p < 10^{-10}$, linear relationship between RPEs

and z-scored firing rates). As a stronger test, we used parameters estimated for each neuron to simulate RPE-correlated spike counts and generated an 'ideal' RPE tuning curve. We observed a clear overlap between real and simulated tuning curves (Fig. 3.2d). Finally, we quantified the correlation between predicted spikes and real spikes and found good agreement (Pearson's correlation coefficient: mean - 0.34, median - 0.31; Fig. 3.2e-f).

To contextualize the robustness of the RPE responses in VP, we ran the same analysis on neurons (n = 183) recorded during the same task in nucleus accumbens (NAc) (Fig. 2.6i), a relevant comparison region because NAc is a major input to VP and, like VP, has reciprocal connections with dopamine neurons in the ventral tegmental area (Groenewegen and Russchen, 1984; Lu et al., 1997; Watabe-Uchida et al., 2012; Beier et al., 2015). We found fewer cells whose activity was fit best by the RPE model in NAc than in VP (8% versus 17%, $\chi^2 = 8.3$, $p < 0.01$) and by the current outcome model (14% in NAc versus 29% in VP, $\chi^2 = 13.6$, $p < 0.001$) (Figure 3.2c). Moreover, NAc neurons classified as RPE-signaling were described less well by the model than similarly classified VP neurons. This was evident by a poorer match between real and simulated neuron tuning curves (mean squared error between real and simulated tuning curves; bootstrapped 95% confidence intervals: [1.23 1.38] in VP, [1.45 1.78] in NAc; Figure 3.2d) and in poorer correlation between model-predicted and actual spiking for individual RPE neurons (Pearson's correlation coefficient: mean - 0.18, median - 0.15; Wilcoxon rank-sum test $p < 0.001$; Figure 3.2e-f). Since striatal activity has been a focus for studies examining the influence of reward history on outcome-evoked signaling (Asaad and Eskandar, 2011; Stalnaker et al., 2012; Kim et al., 2013; Bloem et al., 2017; Shin et al., 2018), it is notable that VP, typically thought of as inheriting its firing from NAc, has more robust RPE signaling than NAc.

3.3.3 VP RPE activity mediates trial-by-trial task engagement

According to reinforcement learning, the function of an RPE is to update the estimate of the current state's value (V) (Sutton and Barto, 1998). The existence of RPE-like signals in this

task raised the question of whether there were corresponding changes in value estimates which could be read out through rats' behavior. Because the rats were freely moving, the decision to participate in the task represented a trade-off between reward seeking and competing interests, including rest, grooming, and exploring the behavioral chamber (Niyogi et al., 2014). Changes in the estimated value of the task could impact rats' engagement in the task. To determine whether rats adjusted their behavior in response to reward outcomes, we analyzed videos (n = 13) from the recording sessions of four of our five VP rats (Mathis et al., 2018; Nath et al., 2019). To estimate task engagement on a trial-by-trial basis, we calculated the average distance from the port in each intertrial interval (ITI). This analysis revealed instances where rats traveled far from the reward port and, in some cases, remained far from the reward port at the beginning of the next trial (Figure 3.3a). Rats typically moved further from the port during the ITI following maltodextrin (Figure 3.3b).

The next question was whether VP activity is related to this measure of engagement. Because there were only two possible outcomes, essentially every sucrose delivery is a positive prediction error, and every maltodextrin delivery is a negative prediction error. Thus, the activity of both Current Outcome and RPE cells would correspond with an increase or decrease in value estimate on sucrose and maltodextrin trials, respectively. To see if the activity of these cells related to task engagement, we calculated the correlation between the reward-evoked activity of VP cells and the distance from the port during the following ITI on each trial. There was, on average, a negative correlation between the activity of VP Current Outcome (n = 111 in these sessions) and RPE cells (n = 60) following reward delivery and distance from the port during the following ITI ($p < 0.01$ compared to shuffled data) but not for Unmodulated cells ($p = 0.92$, figure 3.3c). The negative correlation indicates rats traveled around the chamber and remained far from the reward port after activity of these neurons was low (i.e. negative prediction errors); conversely, they remained closer after high activity.

The correlation between VP neuron activity and task engagement suggested that VP neu-

rons could influence this measure, a possibility we explored with an optogenetic approach. In a new group of rats, we injected virus containing either the inhibitory opsin ArchT3.0-eYFP (n = 7) or eYFP alone, as a control (n = 7), into VP and implanted an optic fiber aimed at VP. We then trained these rats on a similar task; port entry during a 10s cue earned a sucrose reward, but on half of trials, we inhibited VP for 5s beginning at onset of sucrose delivery, mimicking a negative prediction error (Figure 3.3d-e). Much like maltodextrin delivery (the less-preferred option), optogenetic inhibition of VP increased rats' typical distance from port during the following ITI ($p < 0.02$, Wilcoxon signed-rank test); however, this was not true in control rats ($p = 0.81$) (Figure 3.3f-h). We then performed the complementary experiment by injecting channelrhodopsin-containing virus (n = 10) or GFP control (n = 7) into another group of rats. Rats were trained on the same task, and on half of trials, we stimulated VP for 2s at 40Hz, approximating a positive prediction error (Fig. 3.3j). Stimulation of VP increased subsequent task engagement, decreasing distance from port during the following ITI ($p < 0.001$, Wilcoxon signed-rank test), but not in control rats ($p = 0.30$). Thus, VP activity is instructive of task engagement-related behavior, suggesting that outcome-related signals in VP are used to update an estimate of value that motivates task performance.

3.3.4 An expanded value space reveals stronger RPE signaling in VP

One shortcoming of our previous experiment contrasting sucrose and maltodextrin is that the similar palatability of the outcomes may not fully probe the limits of value signaling and would thus constrain our ability to identify RPE neurons; maltodextrin delivery does not typically strongly inhibit responses at the time of reward (Figure 3.1e). In Chapter 2, we found that delivering water, an outcome that was less rewarding than maltodextrin, more strongly inhibited firing rates ('random sucrose/maltodextrin/water' task; 3.5a-c). We hypothesized that this expansion of the dynamic range of firing would reveal additional RPE neurons. We applied the same models as before to neurons recorded during this task (n=254) to identify cells with firing that reflected outcome history-based RPEs, current outcome only,

or no modulation, with an additional free parameter to estimate the value associated with maltodextrin (on the scale of water (0) to sucrose (1)). As we hypothesized, a greater proportion of neurons was best fit by the RPE model than in the random sucrose/maltodextrin task (31% versus 17%, $\chi^2 = 20.0$, $p < 0.00001$; Figure 3.5d). Trial history regressions revealed an impact of many previous trials on these neurons (Figure 3.5e). We observed graded changes in firing rates as a function of estimated RPEs for individual neurons (Figure 3.5f); this relationship was consistent in the population-average PSTH (Figure 3.5g). The firing rates of these RPE neurons monotonically increased as a function of estimated RPEs, and this relationship was consistent with tuning curves for simulated RPE neurons (Figure 3.5h). Moreover, the model's predictions of trial-by-trial spiking for each neuron was robust and stronger than we found in the random sucrose/maltodextrin task (Pearson's correlation coefficient: mean - 0.49, median - 0.48; Wilcoxon rank-sum test between VP-RPE correlation in 'random sucrose/maltodextrin' vs 'random sucrose/maltodextrin/water' task, $p < 0.00001$; Figure 3.2e-f). Thus, with outcomes spanning an expanded value space, we found more neurons that encode RPEs, and do so more robustly.

3.3.5 VP RPE neuron firing adapts to repeated reward presentations

Repeated presentation of the same reward (or sets of rewards) can produce an adaptation in neural responses as the outcome becomes expected (Roesch et al., 2007; Takahashi et al., 2011, 2016), a phenomenon consistent with reinforcement learning. We investigated whether VP neurons also attenuate their reward-evoked firing to repeated outcomes by analyzing the activity of neurons ($n = 348$) recorded during a variation of the sucrose and maltodextrin task where each reward was presented in blocks of 30 trials (Figure 3.6a). We fit the neural activity to the same three models and found a similar number of RPE neurons during this task as in the random sucrose/maltodextrin task (Figure 3.6c). Compared to activity in the random task (Figure 3.6e), RPE neurons in the blocks task had noticeably elevated firing for sucrose trials relative to maltodextrin at the time of cue onset and port entry, consistent with an

acquired reward-specific expectation after repeated trials, and a slightly attenuated difference in firing for the two rewards following reward delivery (Figure 3.6d). To determine how the reward-evoked activity evolved across each block, we plotted the activity in 3-trial bins evenly spaced throughout the session (Figure 3.6f-h). RPE neurons demonstrated notable reward-specific adaptations: a reduction in activity within sucrose blocks ($t_{804} = -5.7$, $p < 10^{-7}$ for a linear model fitting neural activity to session progress for RPE neurons recorded with sucrose block presented first; $t_{882} = -8.5$, $p < 10^{-10}$ for RPE neurons when sucrose block was second) and an increase within the maltodextrin block when maltodextrin was second ($t_{697} = 4.3$, $p < 0.0001$) although not when it was first ($t_{821} = 0.38$,$p = 0.71$), resulting in a significant interaction between the effects of session progress and outcome on the firing rates of RPE neurons in both session types (sucrose first: $t_{1501} = -6.8$, $p < 10^{-10}$, sucrose second: $t_{1703} = -6.4$, $p < 10^{-9}$); this was in contrast to neurons best fit by the Current outcome and Unmodulated models (all $p > 0.05$ for interaction between session progress and outcome). This interaction was also not present in RPE neurons from random sucrose/maltodextrin sessions where rewards were not presented repeatedly within a block ($t_{3959} = 1.7$, $p > 0.05$). The same reinforcement learning model, therefore, that describes neurons sensitive to trial history when rewards are randomly interspersed can also identify neurons in VP that exhibit adaptation across blocks.

3.3.6 Cued information impacts VP firing separately from outcome history-derived information

RPEs are frequently modulated by specific cue-reward associations, as is the case for the midbrain dopamine system (Fiorillo et al., 2003; Eshel et al., 2015, 2016; Tian et al., 2016). To assess whether VP neurons that encode outcome history are also sensitive to cue-based modulation, we trained rats to associate a 'non-specific' cue with unpredictable sucrose/maltodextrin (like the random sucrose/maltodextrin task), and two 'specific' cues, the first which predicted sucrose with 100% probability and the second which predicted

maltodextrin with 100% probability (Figure 3.7a.) As before, consumption of sucrose and maltodextrin was nearly identical across conditions (Figure 3.7d), as was the latency to go to the reward port following cue onset (Figure 3.7b).

We recorded from VP neurons (n = 487) during this task and found once again a prominent difference in activity on sucrose and maltodextrin trials (Figure 3.7e). To quantify how specific predictive cues modulated outcome-evoked firing rates, we augmented the RPE, Current outcome, and Unmodulated models with two free parameters to estimate the contribution of the new cues; thus, each neuron was fit with six models (Figure 3.7f). Remarkably, we found that most neurons ($\sim 78\%$) were best fit by the cue-agnostic models (Figure 3.7g). We alternatively considered a model in which the specific cues allowed for partial to full cancellation of the RPE, but this model was best for a negligible number of cells (see Methods).

We first focused on the neurons without significant cue effects on their outcome signaling to compare with our previous findings. Outcome history regression revealed that these RPE neurons, like previously, incorporated information from multiple previous trials (Figure 3.7h). Firing rates varied smoothly as a function of that trial's estimated RPE (Figure 3.7i), and RPE tuning curves closely matched those of simulated neurons (Figure 3.7j).

With the existence of history-dependent RPE neurons established in this task, we next sought to determine how the specific cues impacted firing. RPE cells were no more likely than Current Outcome or Unmodulated cells to exhibit significant impact of cues on outcome signaling, suggesting RPE responses and predictive-cue responses are independent effects ($\chi^2_2 = 3.86$, $p > 0.14$; Figure 3.7g). In our models, cues were permitted to take on any value; therefore, to better understand how cue information impacted outcome signaling, we plotted the weights for each predictive cue for all neurons with cue effects (Figure 3.7k); we included all neurons because RPE neurons with cue effects were rare (7 of 487 neurons). If cue information is incorporated into outcome signaling in a traditional RPE fashion, sucrose cues should have a positive value and maltodextrin cues should have a negative value. We observed

that, although this particular combination (positive sucrose and negative maltodextrin) was the most common parameter estimate, it did not differ from chance (exact binomial test, 0.321 [0.235-0.417] mean [95% CI], $p > 0.09$ compared to null of 0.25), and in fact there was a whole variety of values assigned to each cue.

A possible reason for the lack of a robust relative value effect of the cues is that rats may not have properly learned the cue-reward associations present in this task. To verify that the significance of the cues was properly learned, we estimated the effect of the specific cues on activity at the time of cue onset, an epoch known to be sensitive to reward value (Tindell et al., 2009; Richard et al., 2016, 2018; Tachibana and Hikosaka, 2012; Fujimoto et al., 2019; Ottenheimer et al., 2019b). Here we found that 29% of neurons were impacted by cue identity (Figure 3.7m). Cells whose cue-evoked firing reflected a positive value for sucrose and a negative value for maltodextrin were the most common category (Figure 3.7n), and this distribution of parameters significantly differed from chance (exact binomial test, 0.411 [0.329-0.497] mean [95% CI], $p < 0.0001$ compared to null of 0.25). This finding indicates that the relative values of sucrose and maltodextrin are represented in the firing evoked by their respective predictive cues, providing evidence of neural learning of cue significance. Therefore, we interpret the diverse impact of predictive cues on outcome signaling that is largely distinct from the impact of outcome history as evidence that these two sources of prediction are separately represented in the VP neural population in this task where the cues do not overtly influence behavior (Fig. 3.7b).

3.4 Discussion

We investigated the influence of outcome history on reward-evoked firing in ventral pallidum (VP) through the lens of reward prediction error (RPE) signaling. Random presentations of two highly palatable outcomes, sucrose and maltodextrin, revealed a subset of neurons in VP that reflected an RPE generated from previously received outcomes and consistent with a preference for sucrose. This RPE signal correlated with measures of task engagement

in the following trial, and optogenetic manipulation of VP following reward delivery altered subsequent task engagement. When a third outcome with a much lower preference (water) was introduced, the expanded range in outcome values revealed additional RPE-encoding neurons in VP. We further found that RPE neurons in VP demonstrate adaptation when the same reward is presented repeatedly. This series of findings is strong evidence for encoding of history-based RPEs by VP neurons that contribute to task performance. We then introduced a task where sucrose and maltodextrin were on some trials preceded by faithfully predictive cues. We found that a smaller subset of neurons, largely distinct from the reinforcement learning RPE population, incorporated the predicted information into their outcome-evoked signaling. The data suggest that, in tasks with highly palatable outcomes, VP contributes to outcome history-based RPE computations that guide engagement of task-related reward-seeking behavior.

A longstanding view is that RPEs are rare outside of dopamine neurons because dopamine neurons compute RPEs locally by integrating distinct elements of the signal relayed from distinct input regions (Keiflin and Janak, 2015; Tian et al., 2016; Watabe-Uchida et al., 2017). In particular, a study of the neural activity of monosynaptic inputs to dopamine neurons revealed a mixture of reward and expectation signals across many brain regions (including VP) rather than distinct components (like value or outcome) in each region, but, notably, there were very few upstream neurons encoding full RPEs (Tian et al., 2016), maintaining the idea that, by and large, RPE is calculated within dopamine neurons themselves (Watabe-Uchida et al., 2017).

Here, we describe a robust RPE signal in VP. Our innovation in the present work (beyond our initial findings in Chapter 2) was implementing a computational modeling approach that allowed us to identify individual neurons whose firing was consistent with RPEs. We found that these neurons integrated reward history over several trials (Figure 3.1j). Learning over a long timescale is consistent with normative Bayesian learning models that predict a reduction in learning rates (equivalently, an increase in reward history integration) in stable

environments, like the ones we used (Behrens et al., 2007). Notably, this kind of outcome history-modulated signal is similar to that observed in dopamine neurons (Tobler et al., 2005; Bayer and Glimcher, 2005; Mohebi et al., 2019).

How can we reconcile our current findings with previous observations that RPEs are rare outside of ventral tegmental area (Watabe-Uchida et al., 2017), even in VP (Tian et al., 2016)? An important distinction is that we focused our analysis on outcome history-based prediction errors rather than cue-based prediction errors. Outcome history-based modulation depends on an expectation of value derived from the combination of previously received outcomes, whereas cue-based modulation depends on the expected value associated with a specific predictive cue and has close ties to temporal difference learning (Sutton, 1988; Sutton and Barto, 1998). Previous work in VP has focused exclusively on cue-based RPEs and found mixed evidence of such encoding (Tindell et al., 2004; Tian et al., 2016; Stephenson-Jones et al., 2020). In our present work, we also saw limited evidence cue-based modulations of outcome signaling (Figure 3.7). None of the prior studies reported analysis of outcome history-based RPEs in the recorded population, so it remains unknown whether VP inputs to dopamine neurons are enriched for history-based RPEs relative to other inputs, and how the VP signal compares to the dopamine neuron signal.

Another intriguing possibility is that VP does not update the values of particular cues, but rather updates estimates of average environmental reward over behaviorally-relevant timescales. Theories and experiments have suggested that average environmental reward signals are critical for invigorating behavior (Niv et al., 2007; Yoon et al., 2018; Bari et al., 2019; Wang et al., 2013; Hamid et al., 2016). Intriguingly, subtle manipulations of VP slow response vigor (Richard et al., 2016) and gross manipulations are typically associated with motivational deficits (Farrar et al., 2008; Smith et al., 2009). Both of these effects are consistent with a role for VP in computing average reward. Our finding that the activity of VP RPE and Current Outcome cells correlates with subsequent task engagement, and that optogenetic manipulations of VP alter subsequent task engagement, additionally supports

this idea (Figure 3.3). Future experiments that incorporate tasks with greater motivation and learning demands will inform more definitive conclusions.

Figure 3.1. A subset of ventral pallidum neurons signal preference-based reward prediction errors.

(a) Reward-seeking task, in which entering the reward port during a 10s white noise cue earned rats a reward.

(b) The white noise cue indicated 50/50 probability of receiving sucrose or maltodextrin solutions, as seen in example session (right).

(c) Percentage sucrose of total solution consumption in a two-bottle choice, before ("Initial") and after ("Final") recording.

(d) Mean(+/−SEM) lick rate relative to pump onset.

(e) Mean(+/−SEM) activity of all recorded neurons on sucrose (Suc) and maltodextrin (Mal) trials.

(f) Mean(+/−SEM) activity of all recorded neurons on trials sorted by previous and current outcome. Dashed lines indicate window used for analysis in (g-h,j) and all equivalent analysis in subsequent figures.

(g) Coefficients from a linear regression fit to the activity of all neurons and the outcomes on the current and preceding 10 trials.

(h) Schematic of model-fitting and neuron classification process. For each neuron, the reward outcome and spike count following reward delivery on each trial were used to fit three models: RPE, Current outcome, and Unmodulated. Akaike information criterion (AIC) was used to select which model best fit each neuron's activity (right).

(i) Mean(+/−SEM) activity of neurons best fit by each of the three models, plotted according to previous and current outcome.

(j) Trial history linear regression for each class of neuron.

Figure 3.1

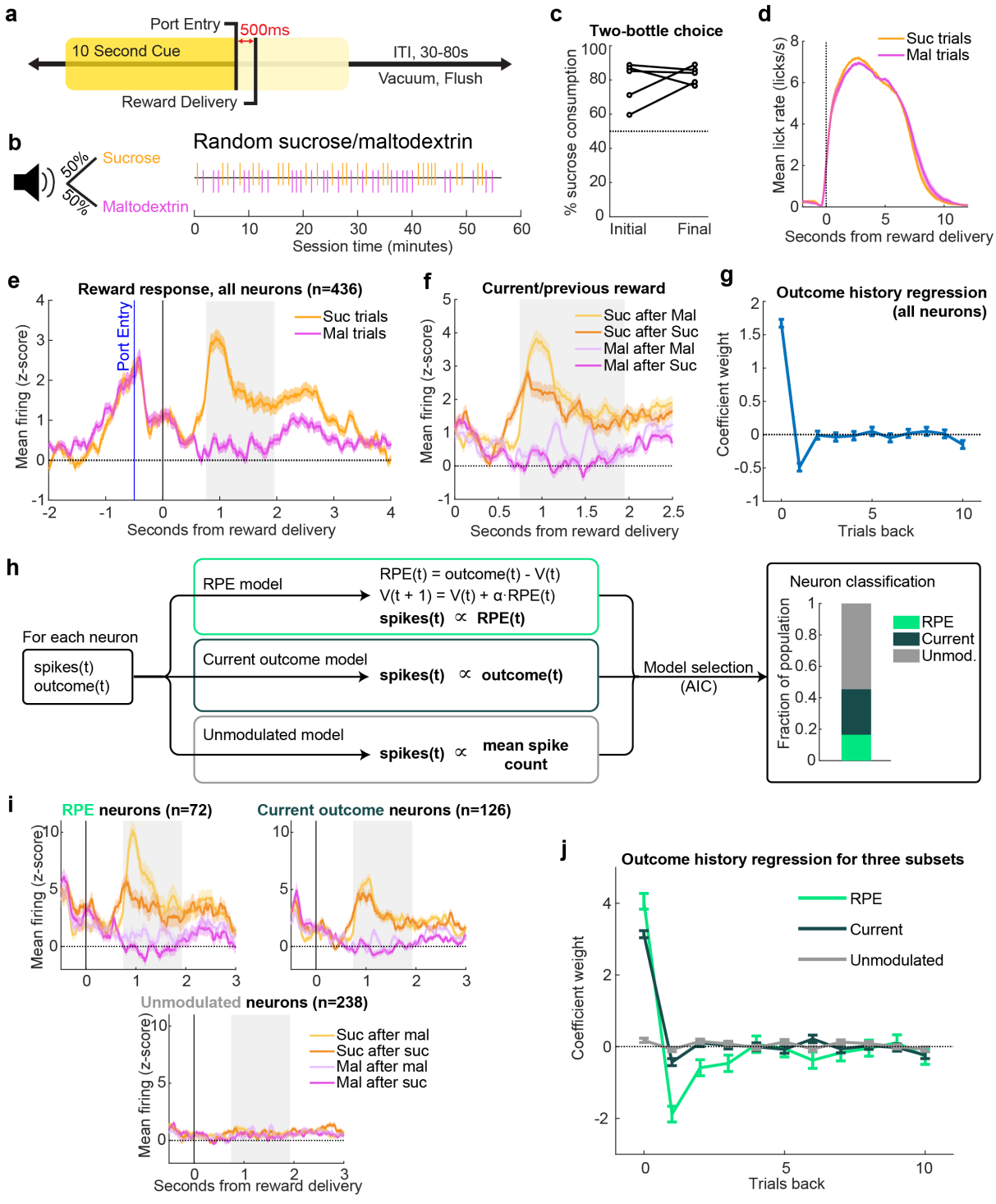

Figure 3.1. A subset of ventral pallidum neurons signal preference-based reward prediction errors.

Figure 3.2

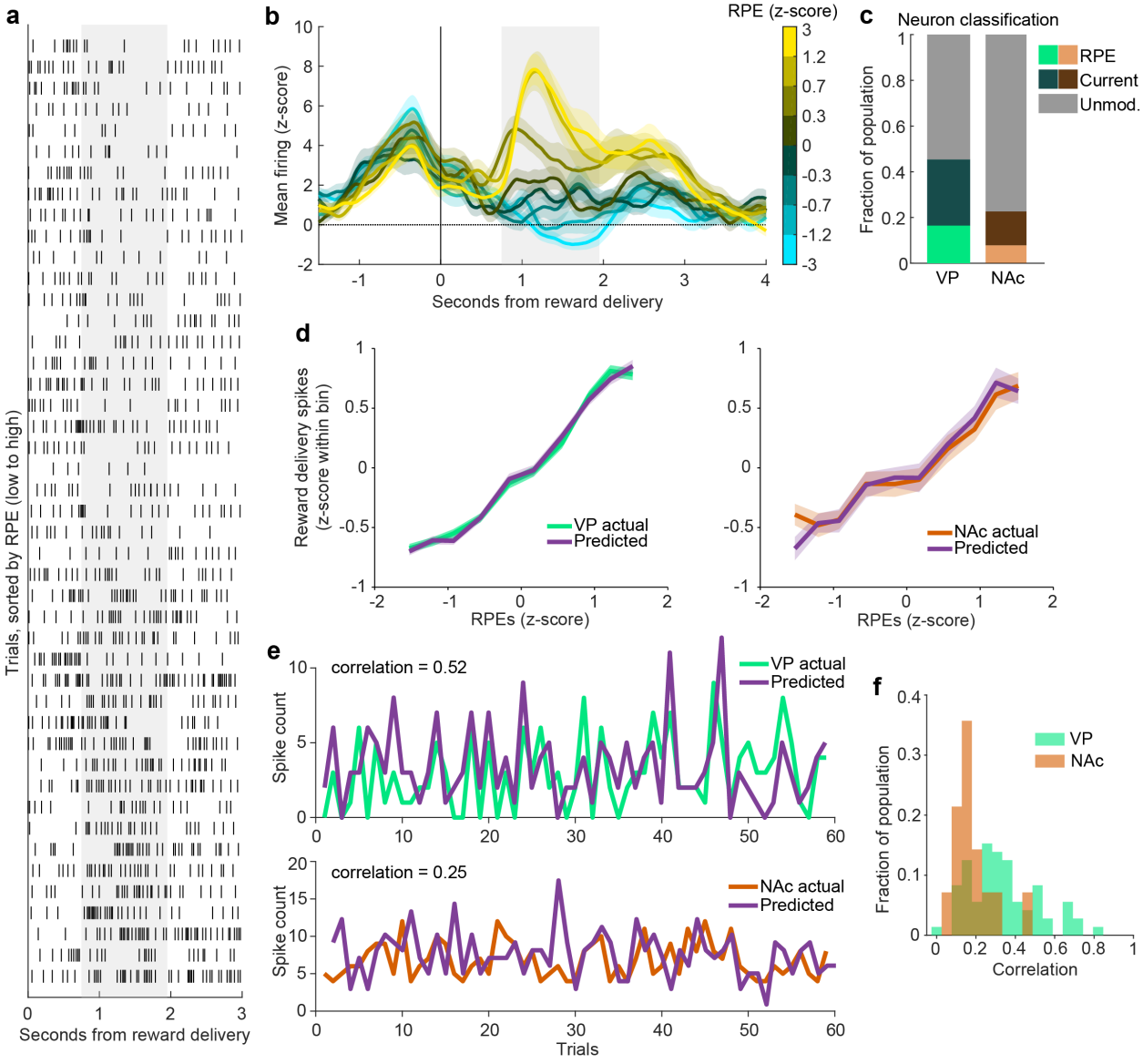

Figure 3.2. RPE encoding is more prevalent and robust in VP than in NAc.

Figure 3.2. RPE encoding is more prevalent and robust in VP than in NAc.

(a) Raster of an individual VP neuron's spikes on each trial, aligned to reward delivery, and sorted by the model-derived RPE value for each trial. Green shaded region indicates window used for analysis.

(b) Population averages of all VP RPE neurons identified in Fig. 1. The trials for each neuron are binned according to their model-derived RPE.

(c) Proportion of the population in VP and NAc classified as RPE, Current outcome, or Unmodulated. There were more RPE ($p < 0.01$) and Current outcome ($p < 0.001$) cells in VP than in NAc.

(d) Mean population activity of simulated and actual RPE neurons according to each trial's RPE value for VP (top) and NAc (bottom).

(e) The model-predicted and actual spikes on each trial for one RPE neuron each from VP (top) and NAc (bottom). These neurons were the 85th percentile for correlation for each respective region.

(f) Distribution of correlations between model-predicted and actual spiking for all RPE neurons from each region.

Figure 3.3

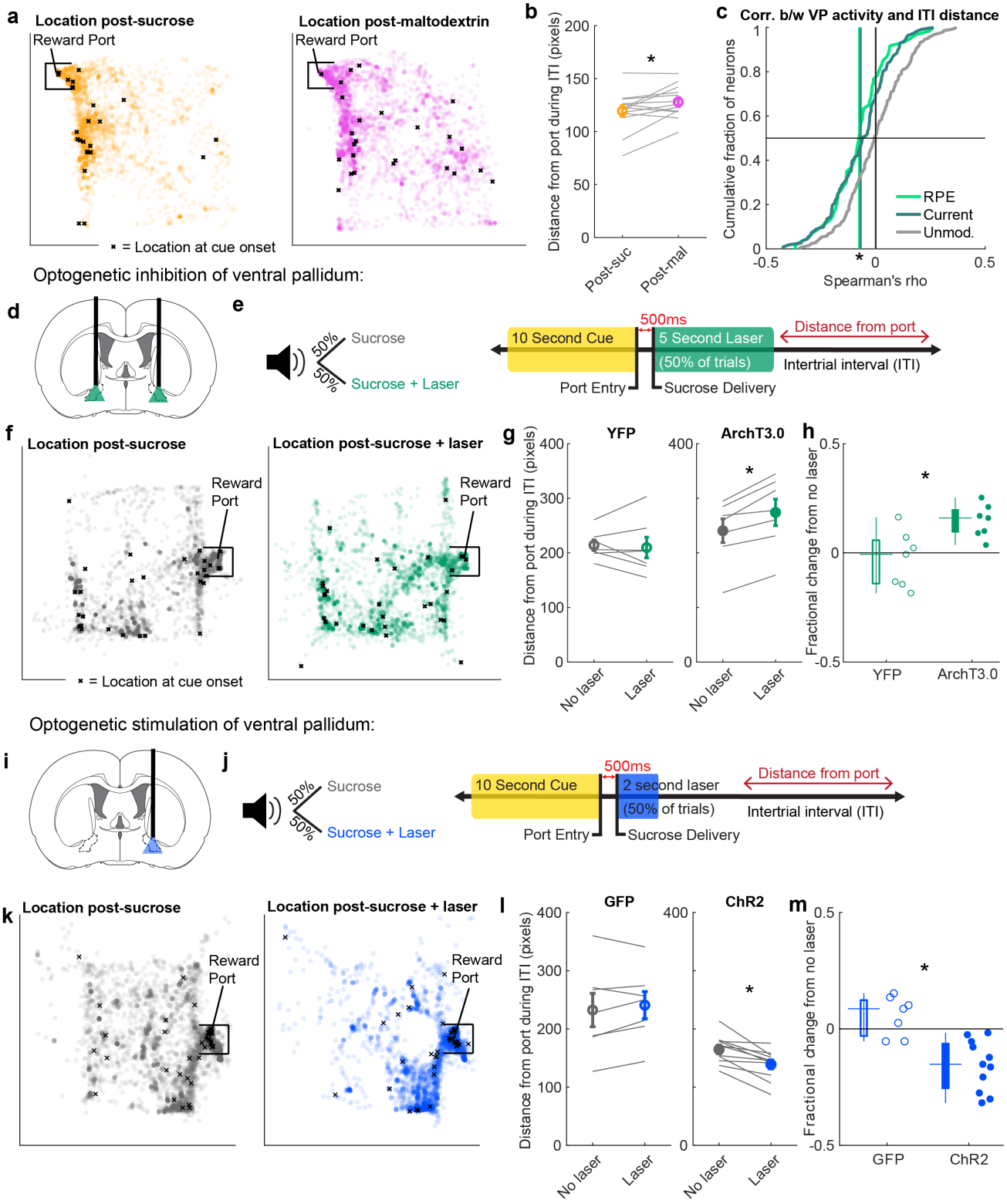

Figure 3.3. VP activity mediates trial-by-trial task engagement.

Figure 3.3. VP activity mediates trial-by-trial task engagement.

(a) All locations of a rat from an example session during the intertrial interval (ITI) following sucrose delivery (left) and maltodextrin delivery (right). Each circle is one location during a 0.2s bin. X marks the location at cue onset for the subsequent trial. Chamber is 32.4cm x 32.4cm (approximately 306 x 306 pixels).

(b) Average distance from the port during ITI following sucrose (orange) and maltodextrin (pink) trials. Gray lines represent average for one subject in one session. * = $p < 0.05$, Wilcoxon signed-rank test.

(c) Distribution of correlations between individual VP neurons' firing rates on each trial and the distance from the port during the subsequent ITI. * = $p < 0.01$ for significant negative shift in mean correlation coefficient (vertical lines) compared to 1000 shuffles of data for both RPE and Current Outcome neurons.

(d) Optogenetic inhibition of VP with ArchT3.0.

(e) Experimental approach to evaluate the contribution of VP to task engagement. Rats received a sucrose reward on every completed trial; on 50% of trials, they also received laser inhibition (left). Specifically, entry into the reward port during the 10s cue triggered delivery of sucrose 500ms later and 5s of constant green laser (right). We then evaluated the rats' distance from the port in the subsequent ITI.

(f) All locations of a rat from an example session during the intertrial interval (ITI) following sucrose delivery without laser (left) and with laser (right). Each circle is one location during a 0.2s bin. X marks the location at cue onset for the subsequent trial. Chamber is 29.2cm x 24.4cm (approximately 542 x 460 pixels).

(g) Average distance from the port in the ITI following sucrose with and without laser for animals receiving a control virus (YFP, left) or the ArchT3.0 virus (right). Individual rats' data shown in gray lines. * = $p < 0.02$, Wilcoxon signed-rank test.

(h) Fractional change in ITI distance from port for each group of rats, displayed as a box plot (median, 25th and 75th percentiles) and for individual rats. * = $p < 0.02$, Wilcoxon rank-sum test.

(i) Optogenetic stimulation of VP with ChR2.

(j) Like (e), but entry into the reward port during the cue triggered delivery of 2s of blue laser at 40Hz, 10ms pulse width (right).

(k) All locations of a rat from an example session during the intertrial interval (ITI) following sucrose delivery without laser (left) and with laser (right).

(l) Average distance from the port in the ITI following sucrose with and without laser for animals receiving a control virus (GFP, left) or the ChR2 virus (right). Individual rats' data shown in gray lines. * = $p < 0.001$, Wilcoxon signed-rank test.

(m) Fractional change in ITI distance from port for each group of rats. * = $p < 0.001$, Wilcoxon rank-sum test.

Figure 3.4

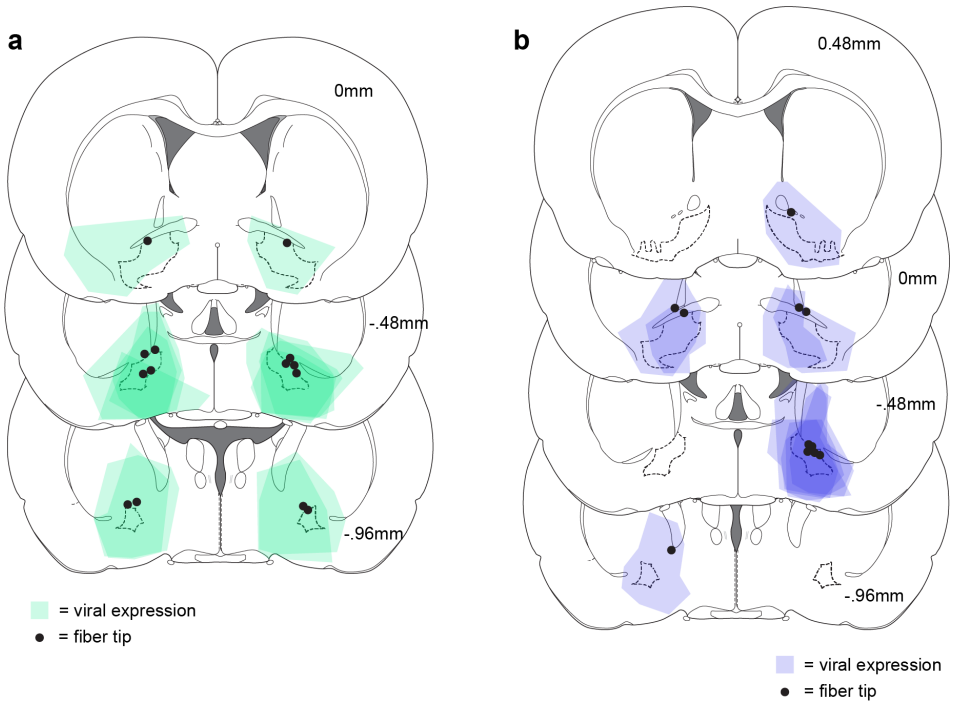

Figure 3.4. Placements for optogenetic experiments.

(a) Expression of ArchT3.0:YFP and fiber tip placement for the rats included in the ArchT3.0 group for the optogenetic experiment.

(b) Expression of ChR2:GFP and fiber tip placement for the rats included in the ChR2 group for the optogenetic experiment.

Figure 3.5

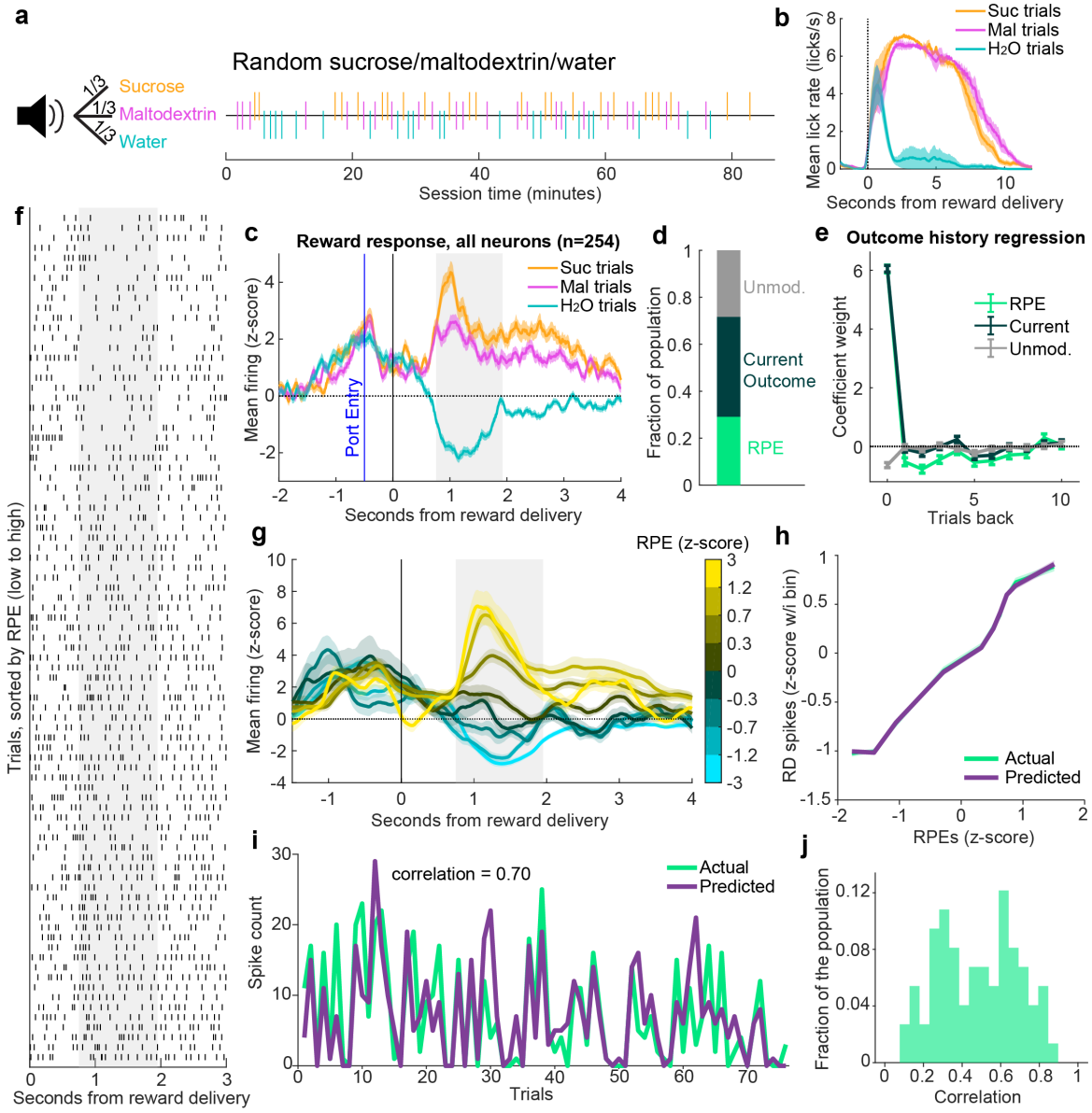

Figure 3.5. An expanded value space reveals stronger RPE signaling in VP.

Figure 3.5. An expanded value space reveals stronger RPE signaling in VP.

(a) A white noise cue indicated 1/3 probability each of receiving sucrose, maltodextrin, or water, as seen in the example session (right).

(b) Mean(+/−SEM) lick rate relative to reward delivery.

(c) Mean(+/−SEM) activity of all recorded neurons on sucrose, maltodextrin, and water trials.

(d) Fraction of the population of neurons recorded in this task best fit by each of the three models.

(e) Trial history regression for each of the three classes of neurons.

(f) Raster of an individual neuron's spikes on each trial, aligned to reward delivery, and sorted by the model-derived RPE value for each trial. Green shaded region indicates window used for analysis.

(g) Population average of all RPE neurons. The trials for each neuron are binned according to their model-derived RPE.

(h) Mean population activity of simulated and actual VP RPE neurons according to each trial's RPE value.

(i) The model-predicted and actual spikes on each trial for the RPE neuron with the 85th percentile correlation.

(j) Distribution of correlations between model-predicted and actual spiking for all RPE neurons.

Figure 3.6

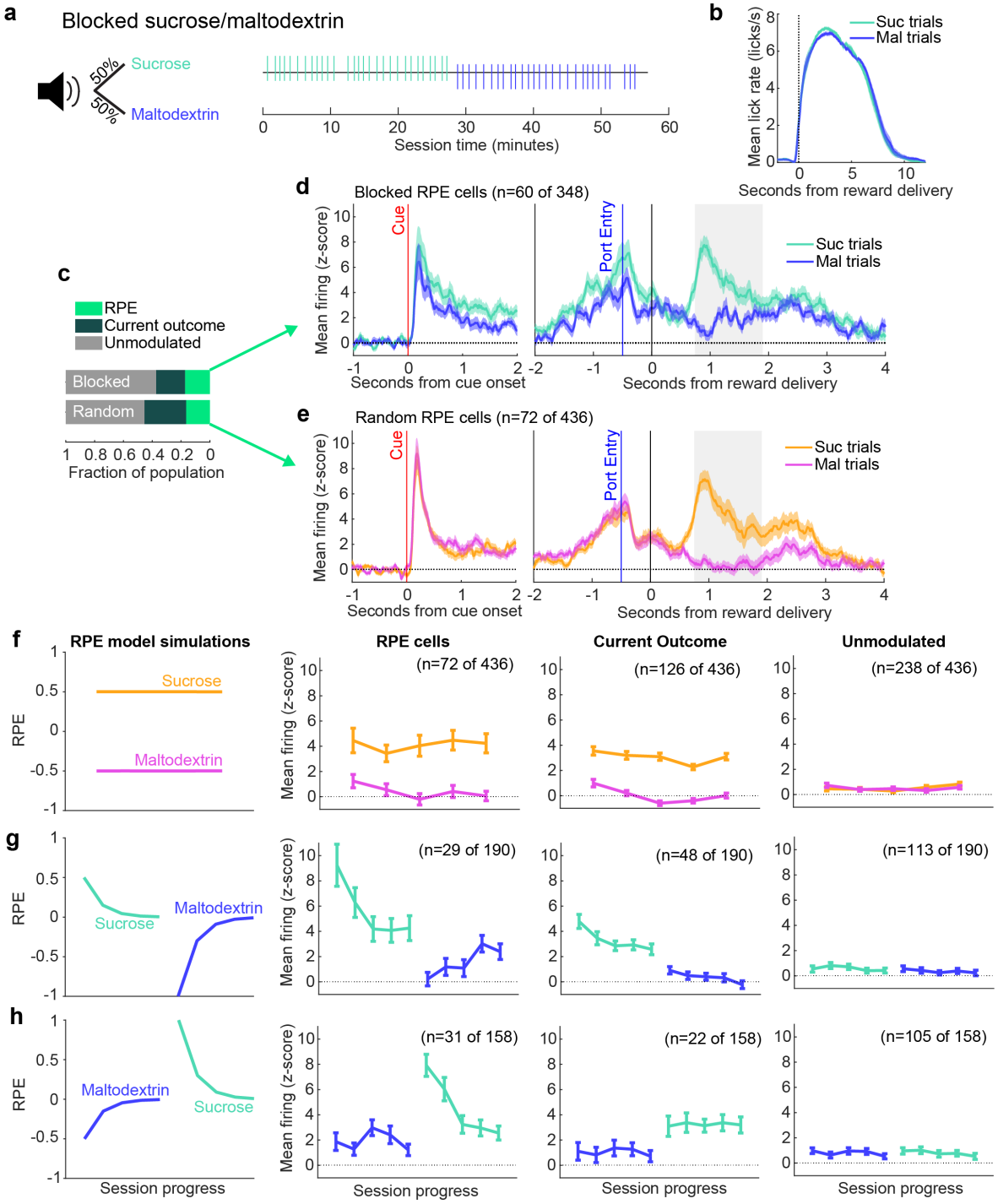

Figure 3.6. VP RPE neuron signaling adapts across reward blocks.

Figure 3.6. VP RPE neuron signaling adapts across reward blocks.

(a) A white noise cue indicated an overall 50/50 probability of receiving sucrose or maltodextrin solutions, but the order of trials was structured into blocks of thirty trials, as seen in example session (right).

(b) Mean(+/−SEM) lick rate relative to pump onset.

(c) Proportion of neurons best fit by each of the three models in the random and blocked sucrose/maltodextrin tasks.

(d) Mean(+/−SEM) activity of all RPE neurons from the blocks tasks aligned to cue onset and to reward delivery.

(e) Mean(+/−SEM) activity of all RPE neurons from the random sucrose/maltodextrin task aligned to cue onset and to reward delivery.

(f) RPE model simulations (left) and mean(+/−SEM) activity of RPE, Current Outcome, and Unmodulated cells from the random sucrose/maltodextrin task, plotted in bins of three trials evenly spaced throughout all completed sucrose and maltodextrin trials.

(g) As in (f), for blocked sessions with sucrose first.

(h) As in (f) and (g) for blocked sessions with maltodextrin first.

Figure 3.7

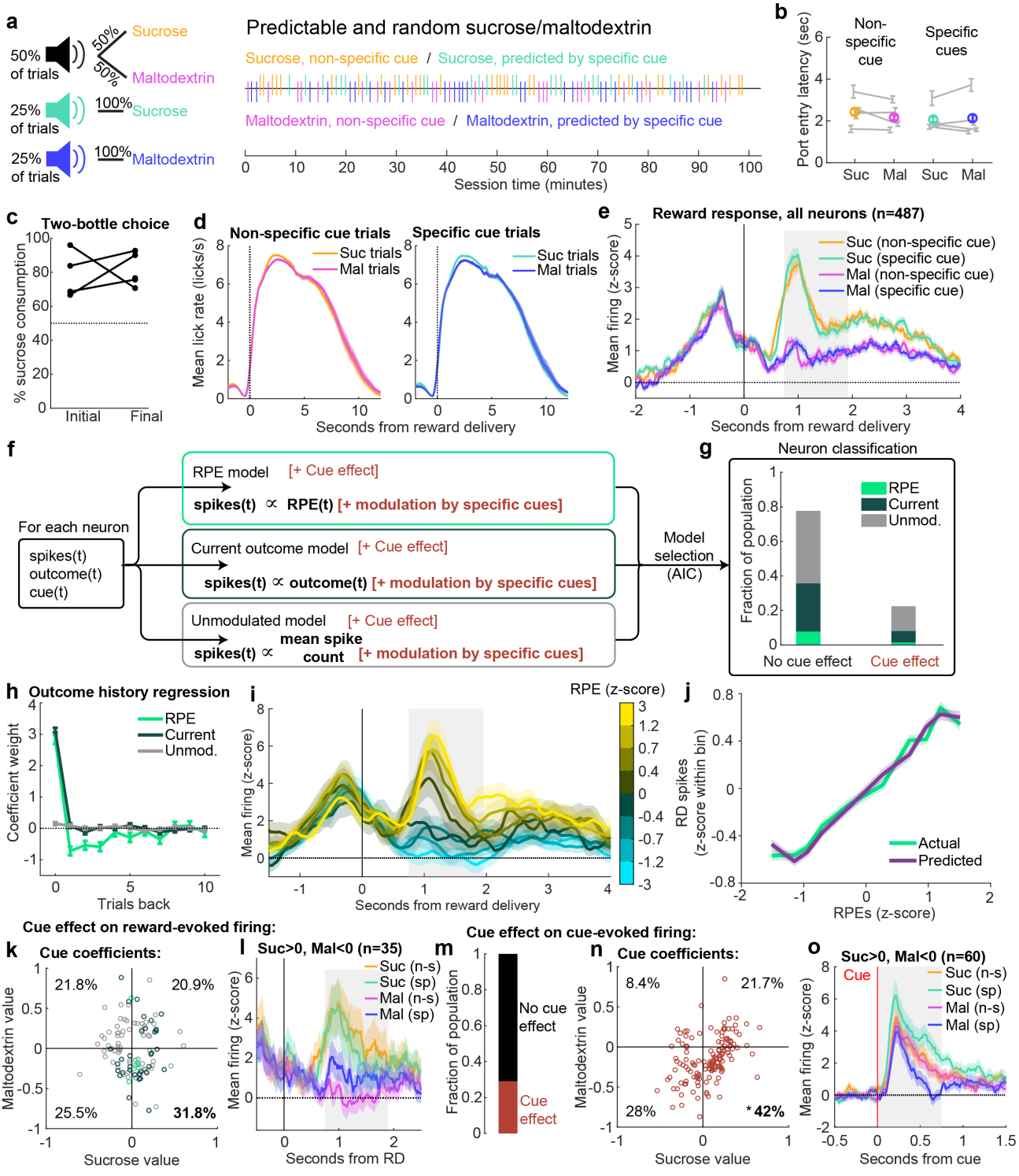

Figure 3.7. Cue- and history-derived information are processed separately by VP neurons.

Figure 3.7. Cue- and history-derived information are processed separately by VP neurons.

(a) Three distinct auditory cues indicated three trial types: a 50/50 probability of receiving sucrose or maltodextrin solutions, a 100% probability of receiving sucrose, or a 100% probability of receiving maltodextrin, as seen in the example session (right).

(b) Median latency to enter reward port following onset of cue for each trial type, plotted as the mean(+/−SEM) across all sessions for each rat (gray) and the overall mean.

(c) Percentage sucrose of total solution consumption in a two-bottle choice, before ("Initial") and after ("Final") recording.

(d) Mean(+/−SEM) lick rate relative to pump onset for each trial type.

(e) Mean(+/−SEM) activity of all neurons recorded in the predictable and random sucrose/maltodextrin task, aligned to reward delivery.

(f) Schematic of cue model-fitting and neuron classification process. The reward outcome and spike count from each trial were used to fit six models: RPE, Current outcome, and Unmodulated with and without the cue effect, which allowed a different weight for the impact of each cue. Neurons were classified according to Aikake information criterion.

(g) Fraction of the population best fit by each model.

(h) Outcome history regression for each class of neurons with no cue effect.

(i) Mean(+/−SEM)activity of all RPE neurons with no cue effect. The trials for each neuron are binned according to their model-derived RPE.

(j) Population activity of simulated and actual VP RPE neurons with no cue effect according to each trial's RPE value.

(k) Scatterplot of each cue effect neuron's weight for sucrose cues and maltodextrin cues. The percentage of neurons falling in each quadrant is indicated.

(l) Mean(+/−SEM) activity of neurons with sucrose values > 0 and maltodextrin values < 0, consistent with a value-based cued expectation modulation.

(m) Neurons with cue effects for cue-evoked signaling, rather than reward-evoked signaling, as in (g).

(n) As in (k), for activity at the time of the cue rather than time of reward. * = $p < 0.0001$ for exact binomial test compared to null of 25%.

(o) As in (l), for activity at the time of the cue rather than time of reward

Figure 3.8

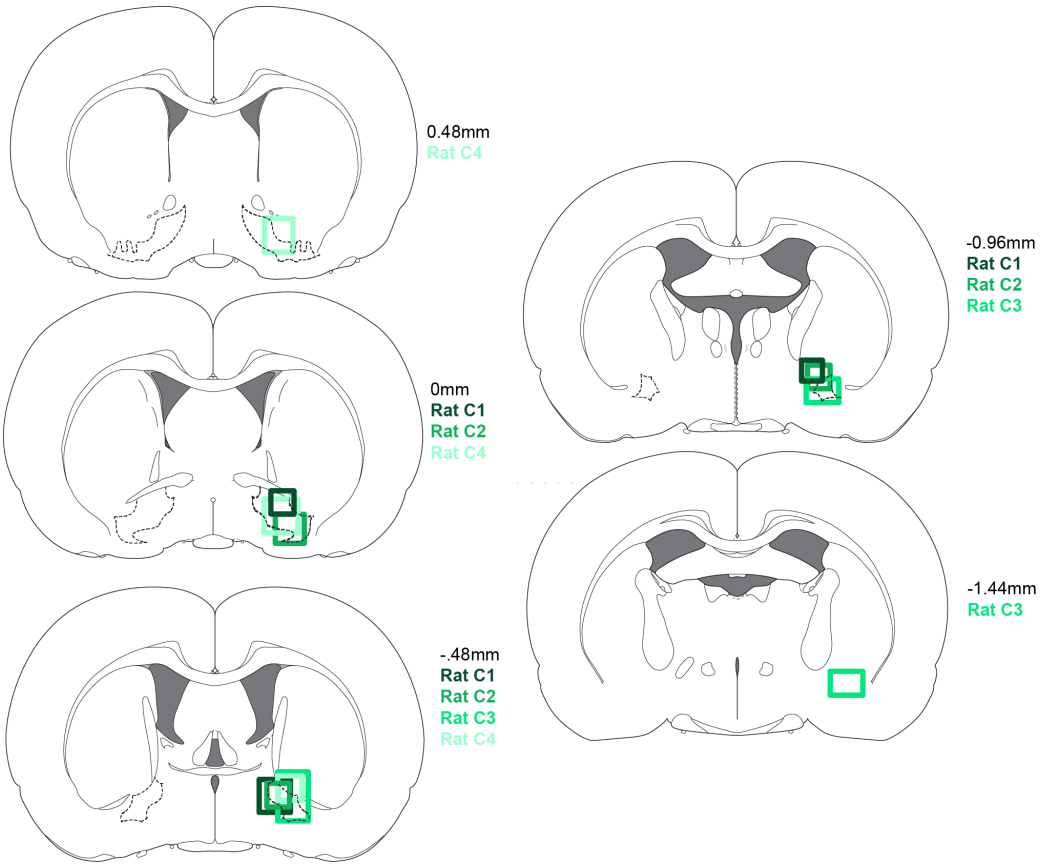

Figure 3.8. Placements for predictable and random sucrose/maltodextrin rats.

(a) Recording locations for rats from predictable and random sucrose/maltodextrin experiment.

Chapter 4

Dynamic preference encoding in ventral pallidum guides choice behavior

4.1 Introduction

In Chapters 2-3, we investigated VP encoding of stable reward preferences (sucrose > maltodextrin > water) and how this signal may impact preference-driven reward-seeking. In many circumstances, however, individuals' valuations of individual reward outcomes can vary across contexts, including physiological state (Hull, 1943; Berridge, 2004; Keramati and Gutkin, 2014; Schultz, 2015). For instance, when rats are salt-deprived, heavily salinated water, which is normally aversive, becomes as palatable sucrose, and VP neural activity reflects this change (Tindell et al., 2006, 2009). Despite this and other work (Fujimoto et al., 2019; Stephenson-Jones et al., 2020) demonstrating an impact of physiological state on VP activity, this has yet to be tested in a scenario when rats are working for multiple distinct outcomes that are differentially impacted by physiological state as it evolves across an individual session (Rolls et al., 1986, 1989; Critchley and Rolls, 1996).

For the present experiments, we had two goals. First was to determine whether the VP relative value signal we previously characterized would reflect a dynamic preference for two rewarding outcomes. Second, building upon our prediction error findings in Chapter 3, was to determine whether the timing of information about the identity of the reward outcome

determined when relative value signaling occurred. By designing two tasks where thirsty rats chose between a small volume of sucrose and a large volume of water, we were able to show that rats' behavioral preferences evolved with physiological state and that a majority of VP neurons track these changes. Moreover, consistent with a reward prediction error framework, the timing of this dynamic relative value signal depended on when information about the outcome was revealed. We further demonstrated that VP activity is sufficient to alter choice behavior but not necessary for online execution of a choice.

4.2 Materials and Methods

Water restriction.

Rats were water restricted overnight prior to behavioral sessions. Rats were given at least 3 hours of access to water each day, and restriction never continued for more than 10 consecutive days. Rats were maintained above 90% baseline body weight. The restriction (and all behavior) was conducted in strict accordance with protocols approved by the Animal Care and Use Committee at Johns Hopkins University.

Behavioral tasks (electrophysiology).

The behavioral apparatus consisted of two retractable levers (Med Associates), one on each side of a reward port. Rats were trained to associate each lever with a distinct reward: 55μL of sucrose or 110μL of water. The pairing of rewards with levers was counterbalanced across rats but remained the same for each rat across days. Rats were first trained on FR1 with both levers present then were moved to a mixture of forced- and free-choice trials with lever retracting upon successful press. Once rats were trained, the levers remained extended for the duration of the session. For the **'Specific Cues'** task, trial types (forced sucrose–30% of trials, forced water–30% of trials, or choice–40% of trials) were announced by distinct auditory cues (white noise, pure tones, or siren, assignments counterbalanced). Trials were randomly

interspersed throughout session. If the rat selected the incorrect lever on forced trials, both levers retracted for 10s, and then the rat could correct its mistake upon reinsertion. Cues remained on until the rat selected a correct lever, which triggered vacuum-mediated evacuation of any residual liquid in the reward cup and delivery of the lever-associated reward 2 sec later. There was a 20-45 sec inter-trial interval following reward delivery before the next cue onset. Rats had experienced >= 10 sessions with final contingencies when recording started, and all performed above chance on forced choice trials (57%-97% accuracy, median 65%). After we completed recordings from these sessions, rats were trained on the second task, **'Uncertain Outcome.'** In this task, there were two trial types–forced (60% of trials) and choice (40% of trials). Instead of distinct auditory cues announcing forced sucrose and water trials, a novel auditory cue (lower frequency siren) indicated that the rat should go directly to the reward port, which terminated the cue and triggered random delivery of sucrose or water 2 sec later. Lever presses had no effect on these trials. Choice trials remained the same as Specific Cues. Sessions were self-paced; we generally stopped the session after 90 minutes. In total, 5 rats completed these sessions and had electrodes successfully targeted to VP. We analyzed one session from each rat for each of the two tasks, selected to have water preference in first quarter of trials, sucrose preference in final quarter, and relatively monotonic transition. Electrodes remained in the same location for the duration of the experiment.

Optogenetic manipulations.

Inhibition. For this experiment, rats were trained on a variation of the sucrose/water choice task where all trials were choice trials. Both levers extended at trial start concurrent with the onset of a white noise cue. Both levers retracted and cue terminated after the choice press, which triggered reward delivery 2 sec later. The rewards were 55μL of sucrose or 165μL of water. Lever assignments were counterbalanced. Rats received 20 days of training on the final task before test. On the test session, on half of trials, rats received bilateral continuous (15 sec, 15-20 mW) photoinhibition of VP, beginning 5s prior to cue. In our

analysis, we included all rats with both fibers and viral expression in VP. This resulted in 8 rats with ArchT3.0 (5M, 3F) and 8 rats with YFP (3M, 5F). **Excitation.** For this experiment, a new group of rats were trained on a sucrose/maltodextrin choice task where they earned 55μL of either reward. Trial frequency was 30% forced sucrose, 30% forced maltodextrin, and 40% choice, randomly interspersed. The available levers extended at trial onset, and all extended levers retracted after a press was made, triggering reward delivery 2 sec later. Lever assignments were counterbalanced. Rats received 12 days of training on the final task before testing. On test session, maltodextrin delivery was paired with unilateral 40Hz pulsed photoexcitation of VP for 5 sec (10ms pulse width, 10-12mW), parameters we selected for being maximally reinforcing (Faget et al., 2018). Although rats were implanted bilaterally, we stimulated the right hemisphere only in all rats for this test for consistency and for a stronger test of sufficiency. We only included rats who had their right fiber and viral expression in VP. This resulted in 11 rats with ChR2 (5M, 6F) and 9 rats with GFP (4M, 5F). For intracranial self-stimulation (ICSS), the same rats were given access to two previously occluded nosepoke ports in the same behavioral chambers. Entry into one port triggered 1s of 40Hz stimulation (10ms pulse width, 10-12mW) of the right hemisphere.

Analyzing trials in session quarters.

Because the sessions were self-paced, each rat completed a different number of trials (88-149, median 99 for Specific Cues; 126-175, median 157 for Uncertain Outcome). To analyze changes across the session, we elected to group rats' trials into quarters of total completed trials rather than dividing sections by a set number of trials. Since individual rats could have different levels of thirst and overall motivation, this method of grouping should ensure that motivation level is roughly matched within quarters across rats performing the same task.

Behavioral analysis.

We estimated preference across choice trials by smoothing the rats' choices (0 for water and 1 for sucrose) with a Gaussian filter ($\sigma = 5$). We then estimated preference on forced trials by assigning each forced trial the smoothed preference of the nearest choice trial. Preference within a given quarter was calculated by finding the fraction of choices from that quarter that were sucrose; this was then linearly transformed from -1 to 1. For analysis of lever press latency, we log-transformed the time interval between cue onset and first correct lever press. The mean latency and licking per quarter was calculated per session, so each session only contributed one point to each quarter.

PSTH creation.

Peri-stimulus time histograms (PSTHs) were constructed using 0.01ms bins surrounding the event of interest. PSTHs were first smoothed on an individual trial basis using a half-normal filter ($\sigma = 3$) that only used activity in previous, but not upcoming, bins. Then, the PSTH across all trials was smoothed with another half-normal filter ($\sigma = 8$). Each bin of the PSTH was z-scored by subtracting the mean firing rate across 10s windows before each trial and dividing by the standard deviation across those windows (n = no. of trials). PSTHs for licking were created in the same manner (without z-scoring) with only one round of smoothing after PSTH creation, $\sigma = 25$.

GLM.

To determine the influence of time and outcome on cue- and reward-evoked firing, we fit a generalized linear model (GLM) with a Poisson distribution to the unsmoothed, binned activity of each neuron on forced trials ('fitglm' in MATLAB). For cue activity, we used a bin 0-0.75s following cue onset, which captured the majority of the phasic response to the cue (Fig. 4.3). For reward activity, we used a bin 0.75-1.95s following reward delivery to remain consistent with Chapter 2, where we saw this was a bin particularly sensitive to

modulation by previous outcome. We only included trials where the rat was in the reward port during reward delivery. For the GLM's predictors, we used trial number as a proxy for time, the outcome on each trial, and the interaction between these two predictors ('Outcome X Time'). Significant predictors were determined by the fitglm function in MATLAB with a cutoff of $p < 0.05$.

Correlations with latency and preference.

We calculated correlations for cue activity with lever press latency and reward activity with preference using the non-parametric Spearman's ρ. We chose this test because we wanted to assess co-variance among the variables of interest without assuming a linear relationship. For cue activity and lever press latency, we included all trials (including choice trials). For reward activity and preference, we only looked at forced trials.

Model fitting.

For each neuron, we took the spike count, $s(t)$, within the 0.75-1.95s post-reward delivery time bin for each forced trial and fit the following four Poisson spike count models. We only included trials where the rat was in the reward port during reward delivery. For all but the Unmodulated model, we used a as a slope (gain) and b as an intercept (offset) parameter to map the model values to spike counts.

Unmodulated model

$$s(t) \sim \text{Poisson}(\exp(\ln(\overline{s})))$$

where $\overline{s}$ is the mean firing rate across all trials.

Satiety model

$$\text{Sat}(t) = 1 - t/t_{end}$$

$$s(t) \sim \text{Poisson}(\exp(a \cdot \text{Sat}(t) + b))$$

where t_{end} is the total number of trials.

Preference model

For sucrose trials

$$\text{Pref}(t) \sim \frac{1}{1 + e^{-k \cdot (t - t_0)}}$$

For water trials

$$\text{Pref}(t) \sim 1 - \frac{1}{1 + e^{-k \cdot (t - t_0)}}$$

$$s(t) \sim \text{Poisson}(\exp(a \cdot \text{Pref}(t) + b))$$

where k is the steepness of the curve, and t_0 is the midpoint (in trials). The convention for the logistic function was to approximate sucrose preference (increasing throughout the session), so it needed to be inverted for water trials. Pref was then normalized from 0 to 1 across all trials in the session before being transformed into spikes.

Mixed model

$$\text{SWP}(t) = w \cdot \text{Pref}(t) + (1 - w) \cdot \text{Sat}(t)$$

$$s(t) \sim \text{Poisson}(\exp(a \cdot \text{SWP}(t) + b))$$

where Pref and Sat are found as above, and w determines the relative contribution of

each to the satiety-weighted preference (SWP).

For all models with a slope parameter, we constrained the slope, a, to be > 0, as our previous work demonstrated the majority of outcome-selective VP neurons are positively correlated with value. We found maximum likelihood estimates for each model and selected the best model using Akaike information criterion (lower AIC indicates a better fit, after taking into account the number of parameters). We used 20 randomly-selected starting initial values for each parameter to avoid finding local minima.

When plotting the reward value estimates from the model fits (Fig. 4.6), we used Sat, Pref, and SWP, respectively. To find the logistic function estimate of behavioral preference (rather than neural activity), we used 'nlinfit' in MATLAB and the same equation we used to find Pref above, but applied to the choices on each trial. To calculate the correlation between behavioral and neural estimates of preference, we used the Pref component from fitting the Mixed model to each neuron. The t_0 from the Mixed model gave us the neural estimates of midpoint. To compare indifference point accuracy within session versus across sessions we took the median distance to behavioral indifference points from the sessions each neuron was not recorded from for all neurons best fit by the Preference or Mixed model (which we call 'preference-encoding' neurons). To compare the similarity of indifference point estimates within session versus across session, we took the median distance to all other preference-encoding neurons' estimates from the same session and the median distance to all other preference-encoding neurons' estimates from the other sessions for each preference-encoding neuron.

4.3 Results

4.3.1 Dynamic preference driven by physiological state.

With the goal of evaluating how dynamic preferences are encoded by ventral pallidum, we designed two tasks where thirsty rats earned either a 55μL sucrose reward or a 110μL water

reward (Fig. 4.1). The choice component of both tasks was the same: on 40% of trials, a 'choice' auditory cue indicated that rats could press either of the available levers, triggering delivery of the associated reward into the reward port 2 sec later. These trials allowed us to assess the rats' preference for sucrose versus water across the session (Fig. 4.1c-d). In the 'Specific Cues' task (Fig. 4.1a), the remaining 60% of trials were either forced sucrose or forced water trials, each indicated with a distinct auditory cue and requiring the rat to press the correct associated lever, triggering delivery of that reward 2 sec later. Because the outcome on forced trials was indicated by the cue, we could evaluate how cue-evoked behavior and neural activity evolved as rats' preferences shifted. In the 'Uncertain Outcome' task (Fig. 4.1b), the remaining 60% of trials were forced trials where the rats responded to a single auditory cue by going directly to the reward port, which triggered delivery of either sucrose or water 2 sec later with 50/50 probability (much like the tasks from Chapters 2-3). Because the outcome was obscured from the rat until delivery, we could evaluate how the reward-evoked neural activity we characterized previously (in Chapters 2-3) would track the changing preference.

In both tasks, rats (n=5) demonstrated dynamic preference, initially preferring water when thirsty at the beginning of the session and switching to preferring sucrose by the end of the session (Fig. 4.1d). This was largely driven by a reduced motivation to consume water, evident in maintained licking for sucrose across the session but consistently decreasing licking for water (Fig. 4.1e-f). Thus, we succeeded in training rats on a task where, despite unchanging task conditions, the rats' demonstrated a preference that shifted according to physiological state, allowing characterization of VP encoding of internally-driven changes in preference.

4.3.2 Dynamic reward encoding occurs when outcome identity is revealed.

We recorded the activity of individual neurons from VP while rats performed the Specific Cues (n=164) and the Uncertain Outcome (n=210) tasks. We were particularly interested

in analyzing the activity of these neurons following the onset of the reward-predicting cues and following delivery of the rewards, two time points sensitive to reward prediction error (Schultz et al., 1997) and demonstrating strong reward value correlates in VP (Tindell et al., 2004, 2006, 2009; Smith et al., 2011; Richard et al., 2016, 2018; Ottenheimer et al., 2019b; Stephenson-Jones et al., 2020) (also, Chapters 2-3) (Figure 4.2a). To evaluate how (or if) this VP activity changed across the session, we implemented a generalized linear model (GLM) that assessed the impact of outcome, time (in numbers of trials), and the interaction between these two predictors ('Outcome X Time') on the activity of individual neurons on forced trials at each time point (Figure 4.2b). We then asked how many neurons from each task were significantly modulated by each predictor (Fig. 4.2c). We were especially interested in neurons whose activity was predicted by Outcome X Time and had more positive slopes for sucrose and more negative slopes for water; that is, neurons whose activity tended to increase for sucrose and decrease for water as the session progressed. Remarkably, we found that the the timing of this activity within a given trial was highly dependent on the task. In the task with Specific Cues, 35% of neurons had cue-evoked activity that followed this pattern and 22% had reward-evoked activity with this pattern. In the task with Uncertain Outcome, 4% of neurons had cue-evoked activity in this pattern (essentially noise, since there was only one, non-specific cue) and 71% had reward-evoked activity with this pattern. These proportions follow a prediction error framework, where outcome-specific responses are encoded by the earliest predictive stimulus; Specific Cues increased the number of VP neurons with outcome-specific activity at the time of cue ($p < 1e-14$) and decreased the number of neurons with outcome-specific activity at the time of reward ($p < 1e-20$, chi-squared test) compared to the task with Uncertain Outcome. Unlike typical demonstrations of prediction error signals, these outcome-specific representations evolved across the session as the relative values of the outcomes shifted. Our next step was to characterize this evolving cue- and reward-evoked firing in the Specific Cues and Uncertain Outcome tasks, respectively, and determine whether it aligned with the shifting behavioral preference.

4.3.3 Activity evoked by specific cues tracks reward-specific task performance.

We first examined the activity of the cue-evoked Outcome X Time neurons in the Specific Cues task (Fig. 4.3). These neurons were notable for their pronounced excitations (above baseline activity) to the water cue at the beginning of the session, which decreased and eventually became inhibitions (below baseline activity) by the end of the session (Fig. 4.3a-c). On the other hand, sucrose cue-evoked activity remained stable across the session (Fig. 4.3a-c). The ranking of the average activity evoked by each cue switched midway through the session, echoing the switch in behavioral preference (Fig. 4.1d). These data are noteworthy for demonstrating that individual VP neurons' cue-evoked representations are not just dependent on physiological state (Fujimoto et al., 2019; Stephenson-Jones et al., 2020) but also specific to the cue's associated reward (and the impact of physiological state on that reward), a phenomenon previously seen across days (Tindell et al., 2009) but not on a per neuron basis as physiological state changes within session.

When applying the prediction error framework to behavior, the transfer of reward signaling from the outcome to the cue is thought to be accompanied by the development of a conditioned behavioral response specific to value of the conditioned cue (Schultz et al., 1997; Fiorillo et al., 2003; Cohen et al., 2012). Therefore, we wondered if rats' behavioral responses in the Specific Cues task would parallel the changes in cue-evoked value representations we observed in these VP neurons. Specifically, we looked for a link between the neural activity evoked by each cue and the latency for the rats to press the appropriate lever, a measure previously used to infer the individual's motivation to respond to the cue (Richard et al., 2016). First, we examined how the latency to press the lever evolved across the session (Fig. 4.3d-e). Within individual sessions, the latencies evolved independently for each reward, with a notable increase in latency to press the water lever on forced water trials; while rats were quicker to respond to the water cue than the sucrose cue at the beginning of the session, this ranking switched midway. Broadly, these changes in latency mirrored the change in cue-evoked activity from the Outcome X Time neurons (Fig. 4.3c).

These data suggested that the specific cue-evoked value representations in VP could invigorate behavioral responding. To explore this idea more closely, we asked, across all trial types, how well the activity of these neurons predicted the latency to lever press (Fig. 4.3f). Many individual neurons had strong negative correlations between cue-evoked firing and latency, indicating greater motivation to press the lever following high cue-evoked activity (Fig. 4.3g). Across the population, Outcome X Time neurons were particularly enriched for negative correlations with latency compared to the other neurons ($p < 0.0000001$, Wilcoxon rank-sum test), demonstrating that this population of neurons with cue-specific value representations was closely linked with the motivation to respond to each cue. This finding provides a possible link between prediction error signals and flexible behavioral responding.

4.3.4 Ventral pallidal activity is necessary for normal cue-triggered responding but not choosing a reward.

The results from the cue-evoked Outcome X Time neurons in the Specific Cues task left us with two testable questions: is this neural activity necessary for normal behavioral responding to the cue – this had been shown previously in a similar task (Richard et al., 2016) – and is VP necessary for executing preference-generated reward choices? We trained a new group of rats on a modified version of the task where all trials were choice trials (Fig. 4.4a). As before, rats increasingly preferred sucrose across the session (Fig. 4.4d). Prior to training, these rats were implanted bilaterally with optic fibers and virus containing either the inhibitory opsin ArchT3.0 (n=8) or a control construct (eYFP, n=8) in VP (Fig. 4.4b). During the test session, on half of trials, we photoinhibited VP bilaterally, beginning 5 sec before cue onset and terminating 10 sec after, disrupting normal VP activity during the execution of rats' choices (Fig. 4.4c). As seen previously (Richard et al., 2016), inhibition of VP during cue presentation increased the latency to respond to the cue (Fig. 4.4e-f). Despite this disruption of the rats' behavioral response, photoinhibition did not affect the choices the rat made (Fig. 4.4a); there was a significant change in preference across the session ($F_{3,56} = 30.5, p < 1e-11$) but

no difference between trials with and without laser ($F_{1,56} = 0.4, p = 0.54$) nor interaction ($F_{3,56} = 0.16, p = 0.92$). These data suggest that the VP reward-specific cue representations we observed (Fig. 4.3) motivate specific reward-seeking actions but are not necessary for making a choice.

4.3.5 With uncertain outcomes, reward-evoked activity closely matches behavioral preference.

We next examined the reward-evoked activity of Outcome X Time neurons in the Uncertain Outcome task. On forced trials, receiving the reward resolves the uncertainty of which reward will be delivered; within a prediction error framework, this signal should reflect a positive error when receiving the preferred reward and a negative error when receiving the the non-preferred reward. Therefore, water-evoked activity should decrease across the session as it becomes less preferred, and sucrose-evoked activity should increase. Interestingly, in contrast to cue-evoked activity, this was indeed the case for reward-evoked activity in these neurons; there were increases for sucrose and decreases for water across the session (Fig. 4.5a-b). Again, the ranking of the activity of these neurons for the respective rewards switched during the session (Fig. 4.5c), mirroring the switch in behavioral preference from these sessions (Fig. 4.1d). To see how closely the neural activity tracked behavioral preference, we compared the activity of individual neurons on forced water and sucrose trials with the choice-derived preference for the respective reward, revealing a strong relationship between the two (Fig. 4.5d). We computed the correlation between these two measures and found, on average, a stronger positive correlation between Outcome X Time neurons' activity and preference on both sucrose ($p < 1e-13$) and water ($p < 1e-24$, Wilcoxon rank-sum test) trials in comparison to other neurons (Fig. 4.5e). Therefore, these neurons followed the general pattern of a preference-derived error signal.

Nevertheless, we noticed that, on average, the neural responses to sucrose and water do not change symmetrically across the session (Fig. 4.5b-c) as might be expected for an

error signal driven purely by preference. Since the overall value of the task declines as the rats become sated, we hypothesized that the reward-evoked activity in VP may reflect a combination of satiety and preference. To evaluate how well satiety and preference capture VP activity, we designed a series of models incorporating these features that could be fit to the activity of individual neurons (similar to our approach in chapter 3) (Fig. 4.6). The first, 'Unmodulated,' has no reward-specific or satiety-related modulation. The second, 'Satiety,' is a linear approximation of declining motivation that decreases uniformly with each trial. The third, 'Preference,' is a logistic function with midpoint and steepness as free parameters. The final, 'Mixed,' is a linear combination of Satiety and Preference with an additional free parameter determining their relative weights. We fit all four models to all neurons from the reward-evoked activity in both tasks and determined which best described each neuron using a Maximum Likelihood Estimation (MLE) approach with Akaike information criterion (AIC) as our selection metric. The example neuron in Fig. 4.6a shows value estimates from the best fit of each model; the Mixed model clearly captures the asymmetric reward-specific firing changes best. Across the population, the Mixed model best described the most number of neurons (Fig. 4.6b), accounting for 56% of the neurons from the Uncertain Outcome task. Together with the neurons best fit by the Preference model (which did have symmetric reward-evoked firing, as seen in Fig. 4.6c), 71% of neurons encoded preference in some capacity in the Uncertain Outcome task, 96% of which were also classified as Outcome X Time by the GLM. In comparison, 30% of neurons in the Specific Cues task encoded preference, of which 66% were also classified as Outcome X Time (but 92% of Outcome X Time neurons were classified as preference encoding with MLE, suggesting MLE was a more sensitive classification approach than the GLM).

In addition to establishing the influence of satiety and independently confirming the greater prevalence of reward-evoked preference encoding in the Uncertain Outcome task, the MLE approach also provided the opportunity to compare neural estimates of preference, which were provided by the fit of the logistic function component of the Preference and

Mixed models, to behavioral preference. To acquire behavioral estimates of preference, we fit the same logistic function from our neural models to the choices from each session (Fig. 4.6d). Visual inspection revealed that preference estimates from the behavior agreed well with the mean estimates from the preference-encoding neurons from their respective sessions (Fig. 4.6e), and this was also evident in the distribution of correlation coefficients comparing each neuron's preference estimate to the behavioral estimate from that session (Fig. 4.6f). Notably, the estimates from preference-encoding neurons were better correlated with the behavioral estimate than the estimates from Unmodulated and Satiety neurons (Fig. 4.6h, $p < 1e-20$, Wilcoxon rank-sum test). The logistic function also explicitly estimates the indifference point, when the preference for the two rewards is ambivalent; there was generally high agreement between neural and behavioral estimates of indifference point among preference encoding neurons (Fig. 4.6g), which outperformed the remaining neurons (Fig. 4.6i, $p < 1e-14$, Wilcoxon rank-sum test). Additionally, the indifference point estimates tended to be more similar among preference-encoding neurons from the same session than between sessions (Fig. 4.6k, $p < 1e-8$, Wilcoxon signed-rank test), demonstrating consistency among the predictions from simultaneously recorded neurons. These metrics indicate that, despite being derived completely independently of the choice behavior, the preference-encoding neurons' estimates of preference approximated the rats' choice behavior well, lending credence to the idea that VP reward-evoked signals are derived from preference.

4.3.6 Optogenetic simulation of a positive prediction error at reward delivery biases choice behavior.

If VP reward-evoked activity signals a preference-based error signal, then manipulations of VP activity at the time of the outcome, mimicking a prediction error, should be able to artificially update the value of the preceding actions. We wondered whether simulation of a positive prediction error after receiving a less-preferred outcome would shift rats' preference

to that option in the future. Previous work in mice has linked stimulation of VP GABAergic cells to place preference, intracranial self-stimulation, and the choice to participate in a task with both appetitive and aversive outcomes (Faget et al., 2018; Stephenson-Jones et al., 2020), setting a precedent to reinforce behavior with VP stimulation. To test our hypothesis, we trained a new group of rats on a modified version of the Specific Cues task with sucrose and maltodextrin as rewards, paralleling the experiments in Chapters 2-3. In this version of the task, available rewards were indicated by lever insertion, and lever pressing led to lever retraction and delivery of the associated reward 2 sec later (Fig. 4.8a). Prior to training, rats were implanted with optic fibers and virus containing the excitatory opsin ChR2 (n=11), permitting stimulation of VP, or a control virus (EGFP, n=9) (Fig. 4.8c). Once rats established a behavioral preference for sucrose, we ran a test session where we stimulated VP unilaterally for 5 sec (40Hz, 10ms pulse width) concurrent with maltodextrin delivery, or whenever the rat first entered the reward port thereafter (Fig. 4.8b). Impressively, pairing maltodextrin with VP stimulation shifted rats' preference from the sucrose lever to the maltodextrin lever on choice trials (Fig. 4.8d, $p < 0.000001$ relative to controls, Tukey test, corrected for multiple comparisons); this shift could be tracked within the session as the rats experienced additional laser-paired maltodextrin trials (Fig. 4.8e). Surprisingly, this shift in preference persisted for at least one additional day when laser stimulation was withheld ($p < 0.00001$, Tukey test), demonstrating that VP stimulation produced long-lasting learning.

How well do these results map onto a prediction error framework? According to the predictions of TD(λ) reinforcement learning (Sutton and Barto, 1998), this fictive positive prediction error should impact the most recent preceding events according to their eligibility traces, which decay with time. Since port entry is the most recent event preceding stimulation, in addition to the observed effect of stimulation on lever pressing, we should also see some impact on port entry behavior if VP stimulation approximates a TD prediction error. Indeed, there was an increase in port entry rate on test session for ChR2 rats relative to con-

trols (Fig. 4.9a). We also found that stimulation of VP supports intracranial self-stimulation in these rats (Fig. 4.8f), confirming the reinforcing properties of VP stimulation. Overall, our data establish that VP activity induces persistent behavioral preferences through its reinforcing properties in a manner consistent with reinforcement learning.

4.4 Discussion

In this chapter, we addressed two questions that were raised in the course of the prior experiments. We previously observed, at a population level, that VP reward-evoked neural activity can flexibly report relative value when the available rewards change (Figs. 2.11, 2.12). It remained unclear whether individual VP neurons would track the relative values of the same rewards as they changed within an individual session. Second, our observation that the outcome-evoked signal in VP many ways resembled a reward prediction error (Figs. 2.10, 3.1, 3.2) left open the possibility that, with sufficiently learned specific cues, the error signal we observed would transfer to the predictive stimuli. In this chapter, we also investigated a link between this VP value signal and value-based decision-making by introducing a choice component to the task and manipulating VP activity optogenetically.

4.4.1 VP reports of relative value vary with physiological state and behavioral preference

We previously found that the reward-evoked activity of a majority of VP neurons reports the relative value of available options in a task with random presentations of rewarding outcomes (Chapter 2). In those experiments, the individual values of the rewards was not (necessarily) changing; rather, the value of the reward relative to the other outcomes in the session changed. Here, we tested how VP would report the relative values of rewards whose value changed within the session due to evolving physiological state. We accomplished this by using a small volume of sucrose solution and larger volume of water as our outcomes, and then we water restricted our rats prior to the session. Importantly, we wanted to be

able to track preference across the session, so we incorporated choice trials that permitted a behavioral readout of rats' preference as well as an opportunity to link VP activity with choice behavior.

Remarkably, we found the 71% of the neurons we recorded during the Uncertain Outcome task tracked the changing values of sucrose and water across the session (Fig. 4.2c). In particular, the reversing of sucrose- and water-evoked firing (Fig. 4.5b-c) and the positive correlations between these neurons and behavioral preference demonstrated that ventral pallidum signaling flexibly tracks the relative values of rewards across physiological states. This result is reminiscent of findings that satiety impacts ventral pallidum event-related firing (Fujimoto et al., 2019; Stephenson-Jones et al., 2020; Tindell et al., 2006, 2009). Here, we build upon these studies by demonstrating that individual VP neurons have distinct representations of the relative values of two rewards as they are differentially impacted by physiological state and the behavioral preference for them switches. Moreover, the strong agreement between neural and behavioral estimates of the rats' preference for sucrose versus water (Fig. 4.6) and the ability of optogenetic stimulation of VP to alter a sucrose versus maltodextrin preference (Fig. 4.8) suggest the VP signaling helps direct value-guided choices.

Our data are reminiscent of reports of individual neurons in OFC and hypothalamus whose responses diminish for rewards (and their predictive cues) fed to satiety but remain intact for other rewards (Rolls et al., 1986, 1989; Critchley and Rolls, 1996). Our data build upon these findings by demonstrating a link between this physiologically sensitive reward-specific encoding and choice behavior. Future work should examine the interplay between homeostatic, value, and decision-making circuits in this setting.

4.4.2 Does VP activity reflect a temporal difference prediction error?

Across our data sets, there were many instances where the VP outcome activity resembled a trial-based prediction error (Rescorla and Wagner, 1972) where the expected value of the trial is updated iteratively with each outcome; the error between this expected value

and the value of the achieved outcome is what we observe in many VP neurons (Chapter 3). Within the reward-learning field, particularly regarding dopamine activity, temporal difference (TD) learning is a more prominent framework to describe neural error signals (Sutton, 1988; Sutton and Barto, 1998; Schultz et al., 1997; Nakahara et al., 2004; Pan et al., 2005). The key advance of the TD model over Rescorla-Wagner is that it allows predictions to update within a trial, not just across trials. Within this framework, the reward-related neural response should transfer to the earliest predictive stimulus. Previous reports of TD learning signals in VP are mixed (Tindell et al., 2004; Ito and Doya, 2009; Tian et al., 2016; Stephenson-Jones et al., 2020). In the majority of our experiments, there was a cue that announced reward availability but not the identity of the reward, so the cue could not acquire a reliable prediction of the outcome, and we were not able to assess adherence to TD learning. In one experiment (Fig. 3.7), we included a mixture of ambiguous and specific cues in hopes of testing whether the reward-specific signal would transfer to the specific cues. Although we did see reward-specific predictive signaling at the time of cue in some neurons (Fig. 3.7m-o), the reward-evoked error signal was largely intact (Fig. 3.7g-l). This result is inconsistent with TD learning, which predicts no error signal at the time of reward delivery when the outcome is faithfully predicted by cues. Nevertheless, this interpretation is confounded by the fact that rats did not appear to be using the cues to guide their behavior (Fig. 3.7b).

In Chapter 4, we revisited the possibility that VP signaling reflects a TD prediction error, this time with two separate tasks in one group of rats. Rats were first trained on the Specific Cues task, where there were three trial types, all indicated by distinct auditory cues. In this task, the cues were much more salient than in Fig. 3.7 because they informed the rats which levers they could press (and incorrect presses led to a 10 sec timeout). With this new task, we found that more neurons reflected the changing value of the outcomes at the time of cue than at the time of reward (Fig. 4.2c). Afterwards, the rats were trained on the Uncertain Outcome version of the task, where lever-pressing on forced trials was replaced

with an entry into the reward port (same as Chapters 2-3), and the cue was no longer distinct for each forced trial type. Here, there was no outcome value tracking at the time of the cue (since the cues were the same for both outcomes), but, remarkably, the number of neurons whose reward-evoked activity reflected the changing values of sucrose and water dramatically increased compared to the Specific Cues task (Fig 4.2c). Our results demonstrate that, in the presence of faithfully predictive (and behaviorally relevant) stimuli, error signaling in VP transfers to the earliest stimulus, consistent with the predictions of TD learning.

One major caveat to the TD learning interpretation of VP signaling is that cue-evoked VP activity appears to be model-based rather than model-free; that is, the cue-evoked activity tracks the value of the predicted outcome as it changes due to physiological state (Fig. 4.3, see also Tindell et al. (2009)). A strict interpretation of TD learning as a model-free algorithm would require the value of the cue to be updated only when there were prediction errors at the outcome; although there was less prediction error signaling at the time of outcome in the Specific Cues task, it is possible that the value of the cue could be reduced iteratively each time the rat receives water in a sated state. TD learning cannot, however, explain the robust activity for the water cue at the beginning of the session since the last time the rat experienced the water cue was in the sated state the previous session. Dopamine neurons also signal a mixture of model-based and model-free errors, and a combination of the two systems is a likely explanation in both cases (Schultz, 2013; Langdon et al., 2018).

Some aspects of TD learning, however, can help us interpret the effects of the optogenetic manipulation in Fig. 4.8. In this experiment, we stimulated VP concurrently with maltodextrin delivery if rats were in the reward port or as soon as they entered the reward port thereafter. Therefore, according to TD(λ) learning (Sutton and Barto, 1998), this fictive positive prediction error should affect both preceding actions: the port entry and the maltodextrin lever press. Indeed, the port entry rate on test session increased for ChR2 rats relative to controls (Fig. 4.9a). The next most recent event, and the most reliable predictor, was pressing the maltodextrin lever, a measure which also increased in test session (Fig.

4.8d). The persistence of the maltodextrin lever preference but not the increased port entry rate on the first recovery day could reflect a) the greater reliability of maltodextrin lever as a predictor of stimulation in the test session and b) slower updating due to its more distal position and decayed eligibility trace. Overall, these data are consistent with many aspects of TD learning and suggest that VP signals behaviorally instructive prediction errors.

Figure 4.1

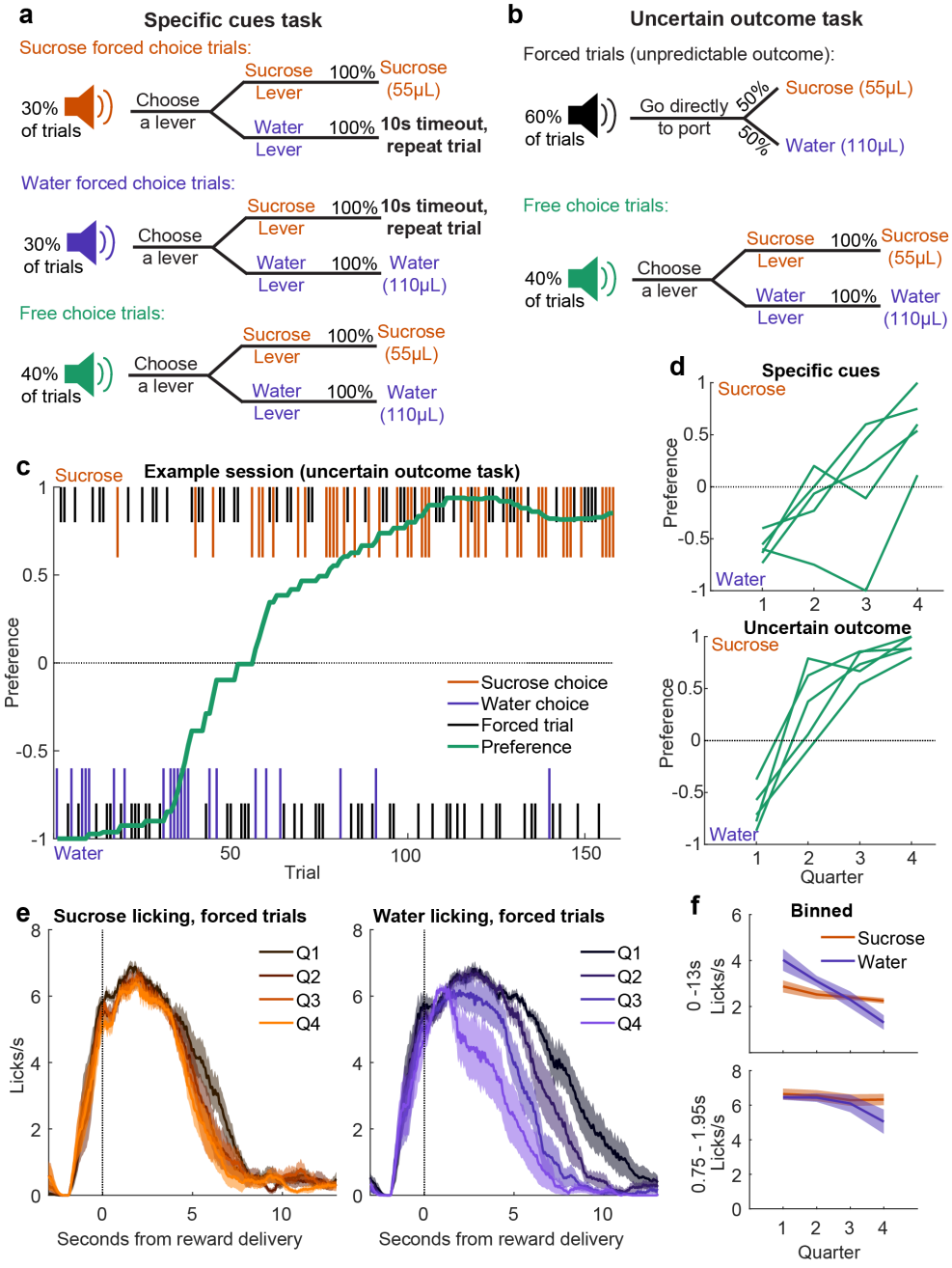

Figure 4.1. Dynamic preference driven by physiological state.

Figure 4.1. Dynamic preference driven by physiological state.

(a) Schematic of 'Specific Cues' task, where there were three trial types, each with a unique auditory cue. Correct lever presses on forced choice trials led to delivery of the associated reward.

(b) Schematic of 'Uncertain Outcome' task, which rats were trained on after the Specific Cues task. The choice trials (and cue) were the same, but the forced trials had a novel auditory cue and required entry into the reward port rather than a lever press, after which either reward was delivered.

(c) Example Uncertain Outcome session, depicting choice trials (colored, longer lines) and forced trials (black, shorter lines) for sucrose (top) and water (bottom), overlayed with preference (green), found by smoothing across choice trials.

(d) Preference in each task for each of the 5 rats, found from the proportion of sucrose choices in each of the four quarters of trials. We divided the session into quarters to account for different numbers of completed trials from each rat.

(e) Mean(+/−SEM) lick rate relative to reward delivery across the four quarters of trials, split into forced sucrose trials (left) and forced water trial (left). Sessions from both tasks are combined here.

(f) Mean(+/−SEM) lick rate across 13 sec, capturing nearly all of the reward-related licking (top), and within the bin used for neural analysis (0.75-1.95s post-delivery).

Figure 4.2

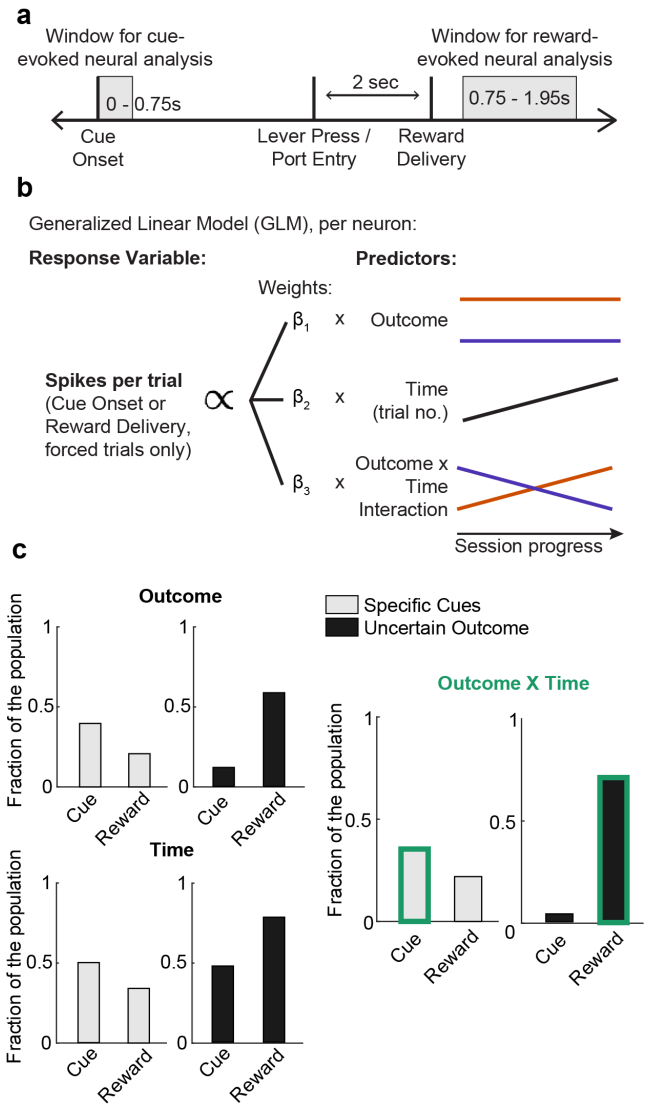

Figure 4.2. Dynamic reward encoding occurs when outcome identity is revealed.

Figure 4.2. Dynamic reward encoding occurs when outcome identity is revealed.

(a) Windows for used for neural analysis relative to task events.

(b) Schematic of the generlized linear model (GLM) used to predict individual neurons' activity at time of cue and reward.

(c) Proportion of neurons from each task with significant impacts of Outcome, Time, and Outcome X Time at the time of cue and time of reward. We focused our additional analysis on the Outcome X Time neurons at time of cue for the Specific Cues task and at the time of reward for the Uncertain Outcome task (marked in green).

Figure 4.3

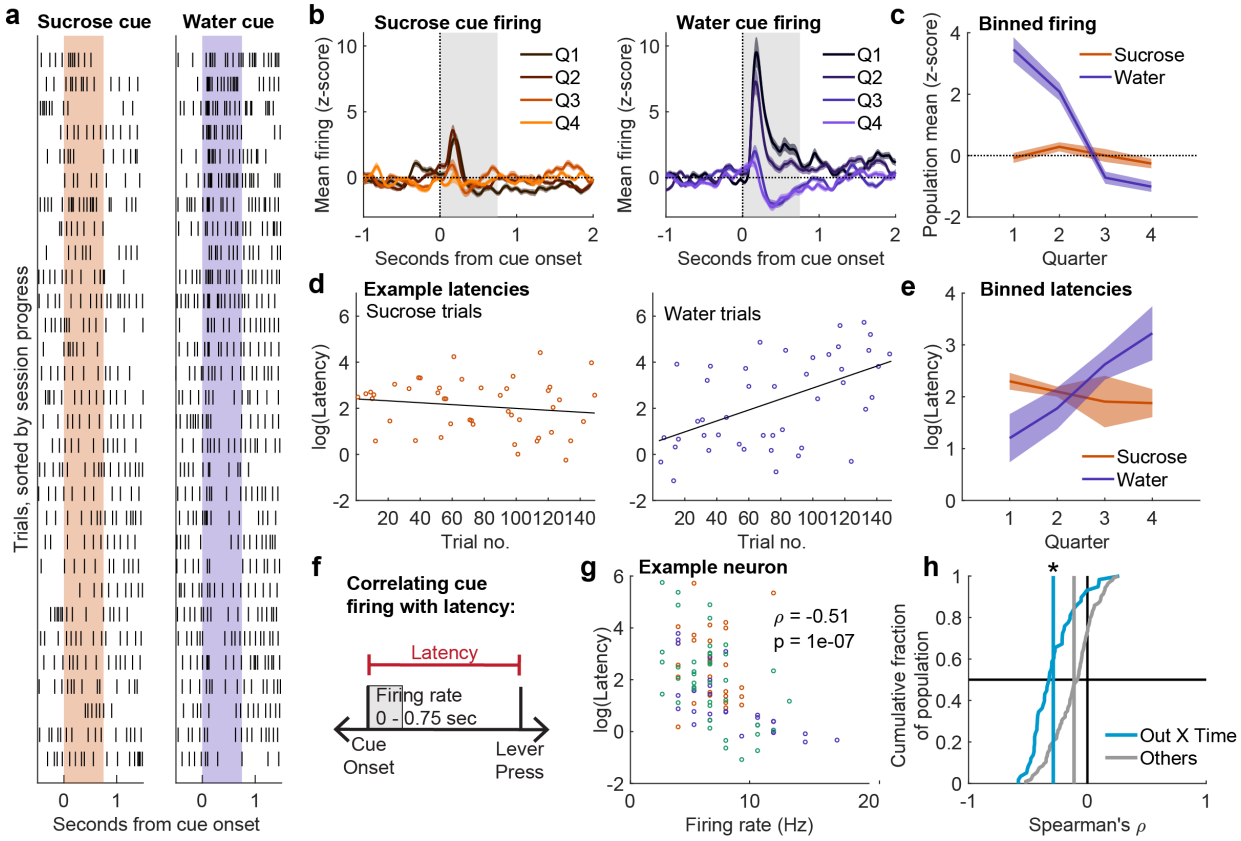

Figure 4.3. Cue-evoked activity tracks reward-specific task performance.

(a) Raster from example Outcome X Time neuron from the Specific Cues task, aligned to sucrose (left) and water (right) forced trial cues. Shading indicates window for neural analysis.

(b) Mean(+/−SEM) sucrose cue- (left) and water cue-evoked (right) firing for all Outcome X Time neurons from the Specific Cues task across the four quarters of trials. Gray shading indicates window for neural analysis (including in (c)).

(c) Mean(+/−SEM) binned firing for these neurons.

(d) Latency to press lever (log-transformed) across all sucrose (left) and water (right) forced trials from an example Specific Cues session.

(e) Mean(+/−SEM) binned firing for Specific Cues sessions.

(f) Approach for calculating correlation between cue-evoked firing and latency to lever press.

(g) Correlation between firing rate for sucrose (orange), water (blue), and choice (green) trials and log(Latency) for an example Outcome X Time neruon.

(h) Distribution of correlation coefficients for Outcome X Time (blue) and other (gray) neurons. Mean of each group marked with vertical line. * indicates $p < 0.0000001$ for Wilcoxon rank-sum test comparing the groups.

Figure 4.4

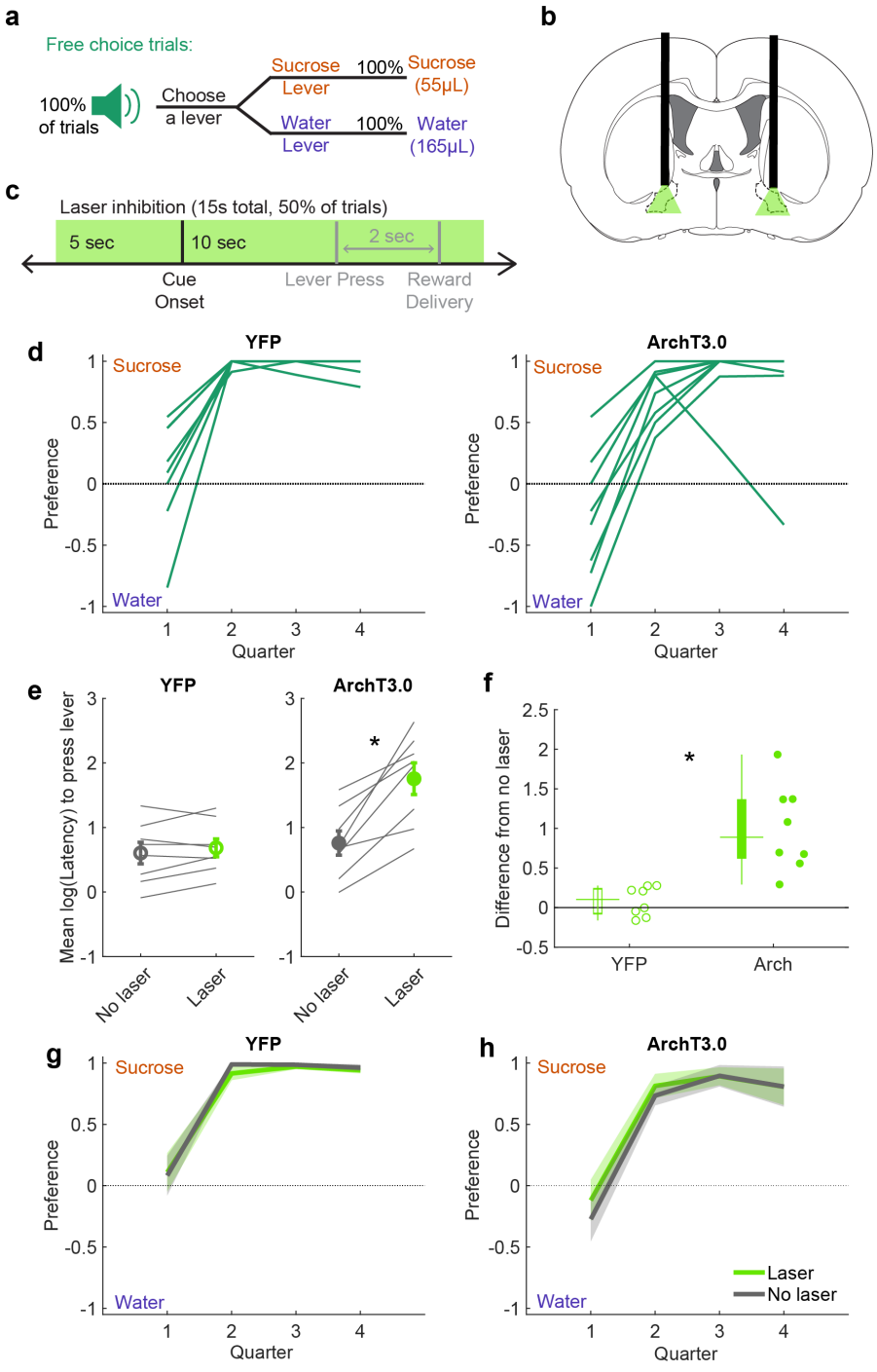

Figure 4.4. Ventral pallidal activity is necessary for normal behavioral responding but not choosing a reward.

Figure 4.4. Ventral pallidal activity is necessary for normal behavioral responding but not choosing a reward.

(a) Task design for optogenetic inhibition experiment. All trials were choice trials, indicated my lever insertion and white noise cue. Levers were retracted and cue terminated after choice was made.

(b) Optic fibers and virus containing ArchT3.0 (or YFP control) were implanted bilaterally in VP.

(c) During the test session, on half of trials, VP was photoinhibited bilaterally for 15 sec, beginning 5 sec before cue onset.

(d) Preference for each of the 8 rats from each group, found from the proportion of sucrose choices in each of the four quarters of trials.

(e) Mean log(Latency) on trials with and without laser for YFP (left) and ArchT3.0 (right) rats. There was no effect of laser on YFP rats ($p = 0.31$) but there was for Arch rats ($p < 0.01$, Wilcoxon signed-rank test).

(f) Difference in log(Latency) between trials with and without laser for YFP (left) and ArchT3.0 (right) rats. The difference was greater in ArchT3.0 rats ($p < 0.001$, Wilcoxon rank-sum test).

(g) Preference across the session for YFP rats, split into trials with and without laser. Although there was a significant change in preference across the four quarters ($F_{3,56} = 54.0, p < 1e-15$), there was no effect of laser ($F_{1,56} = 0.13, p = 0.72$) nor interaction ($F_{3,56} = 0.12, p = 0.95$).

(h) As in (g) for rats with ArchT3.0 in VP. Quarter: $F_{3,56} = 30.5, p < 1e-11$, laser: $F_{1,56} = 0.4, p = 0.54$, interaction: $F_{3,56} = 0.16, p = 0.92$.

Figure 4.5

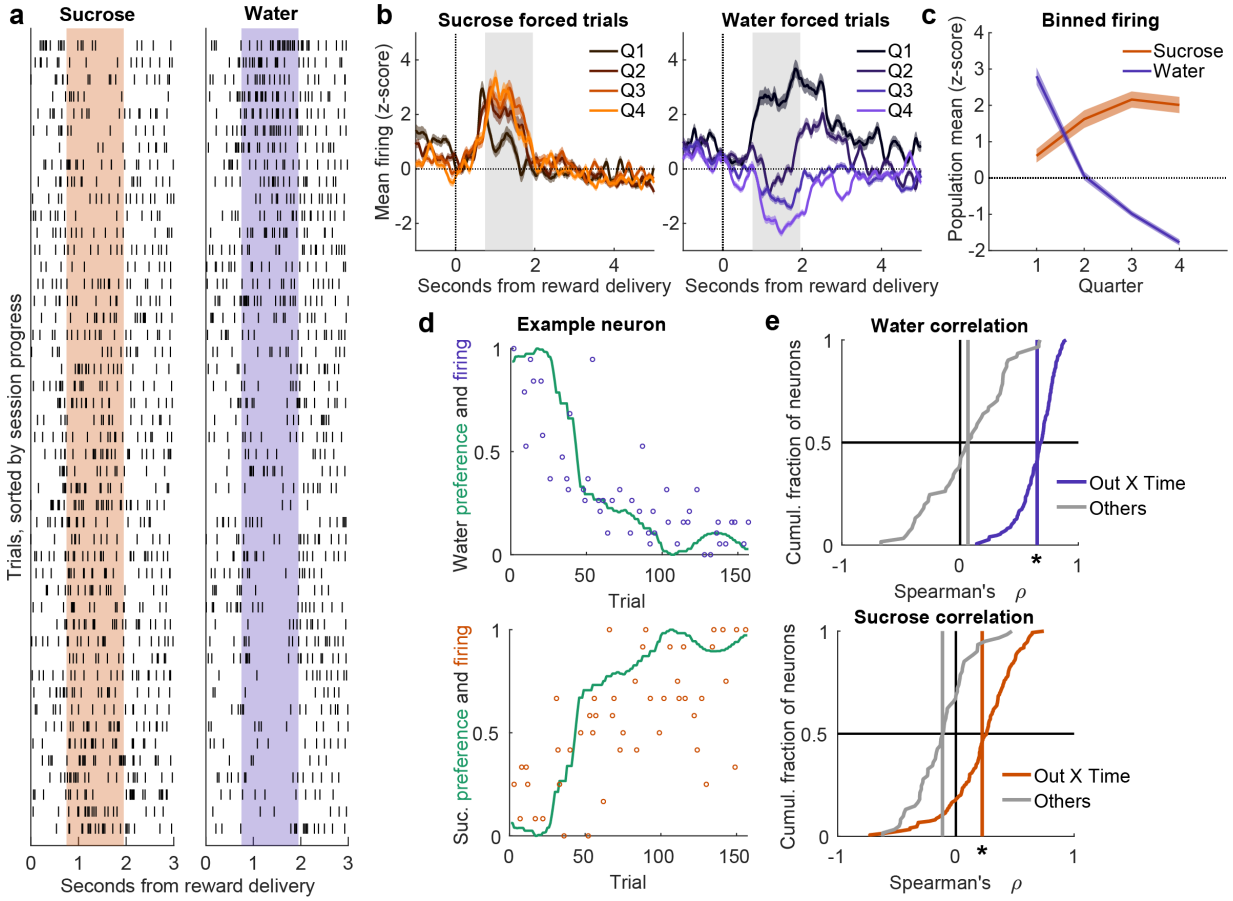

Figure 4.5. Reward-evoked activity closely matches behavioral preference.

(a) Raster from example Outcome X Time neuron from the Uncertain Outcome task, aligned to sucrose (left) and water (right) delivery on forced trials. Shading indicates window for neural analysis.

(b) Mean(+/−SEM) sucrose- (left) and water-evoked (right) firing for all Outcome X Time neurons from the Uncertain Outcome task across the four quarters of trials. Gray shading indicates window for neural analysis (including in (c)).

(c) Mean(+/−SEM) binned firing for these neurons.

(d) Normalized firing of an example Outcome X Time neuron on forced water trials overlayed with water preference (top) and the same neuron on sucrose trials with sucrose preference (bottom).

(e) Distribution of correlation coefficients between firing rate and preference for Outcome X Time (blue or orange) and other (gray) neurons on forced water trials (top) or forced sucrose trials (bottom). Mean of each group marked with vertical line. * indicates $p < 1e-24$ for Wilcoxon rank-sum test comparing the groups on water trials, and $p < 1e-13$ for sucrose trials.

Figure 4.6

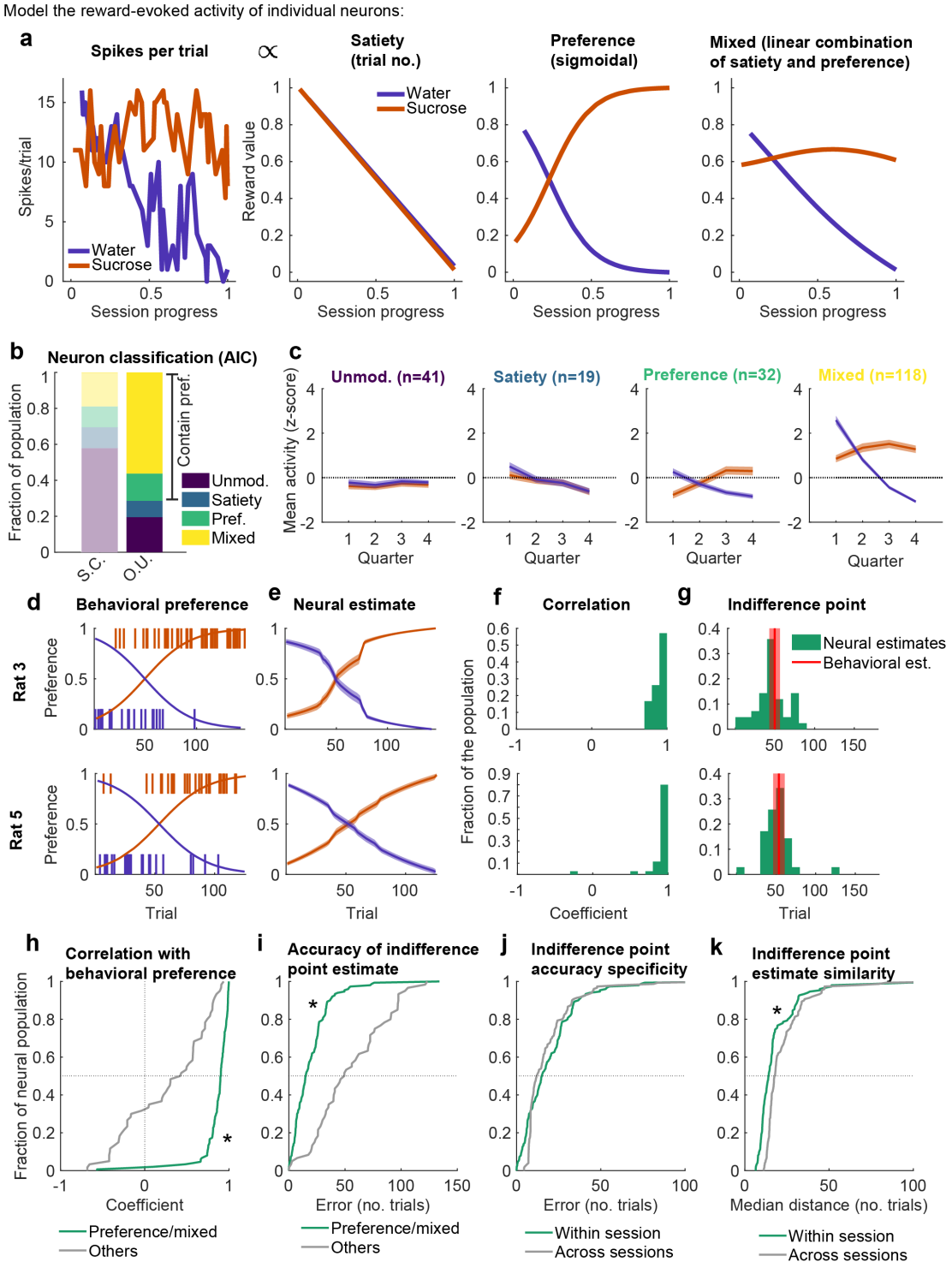

Figure 4.6. Models of VP reward-evoked activity accurately predict behavioral preference.

Figure 4.6. Models of VP reward-evoked activity accurately predict behavioral preference.

(a) Example fits for the three models we considered to describe the activity of VP neurons in the Uncertain Outcome task at time of reward delivery: Satiety, Preference, and Mixed, which linearly combined Satiety and Preference (a fourth model, Unmodulated, stayed at the mean firing rate for the whole session). Mixed was best able to capture the changes in activity in this example neuron.

(b) Classification of neurons based on their reward-evoked activity, split by task. Classification was achieved by finding the model that best described the activity of each individual neuron, determined with Akaike information criterion (AIC). Neurons classified as Preference or Mixed were considered preference-encoding.

(c) The mean(+/−SEM) activity across quarters on sucrose (orange) and water (blue) trials for neurons from the Uncertain Outcome task best fit by each model.

(d) From two example sessions, the choices of the rats across the session and the preference estimate from fitting a logistic function (the same one contributing to the Preference neural model).

(e) The mean(+/−SEM) estimate of preference from the preference-encoding neurons from these sessions.

(f) Correlation between neural estimate and behavioral estimate of preference for each neuron from each of the example sessions.

(g) Estimates of the indifference point (sucrose and water equally preferred) from the neural and behavioral models (+/−SE).

(h) Across all Uncertain Outcome sessions, preference-encoding neurons had preference estimates with higher correlations with the behavioral estimate than estimates from the remaining neurons ($p < 1e-20$, Wilcoxon rank-sum test).

(i) Preference-encoding neurons' estimates of indifference point were also closer to the behavioral estimate than estimates from non-preference-encoding neurons ($p < 1e-14$, Wilcoxon rank-sum test).

(j) Among preference-encoding neurons, the estimate of their parent session's indifference point was not better than the estimate of other sessions' indifference point ($p = 0.16$, Wilcoxon signed-rank test), perhaps reflecting the similar indifference points across sessions.

(k) Among preference-encoding neurons, the estimate of indifference point was more similar to other neurons from the same session than neurons from other sessions ($p < 0.00000001$, Wilcoxon signed-rank test).

Figure 4.7

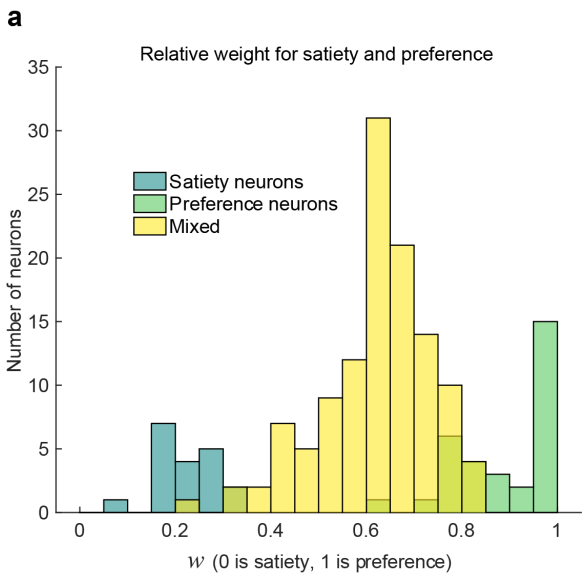

Figure 4.7. Weights for Mixed model fits.

(a) Distribution of weights (w) for Satiety, Preference, and Mixed neurons when fit with the Mixed model, revealing separate clusters for each group.

Figure 4.8

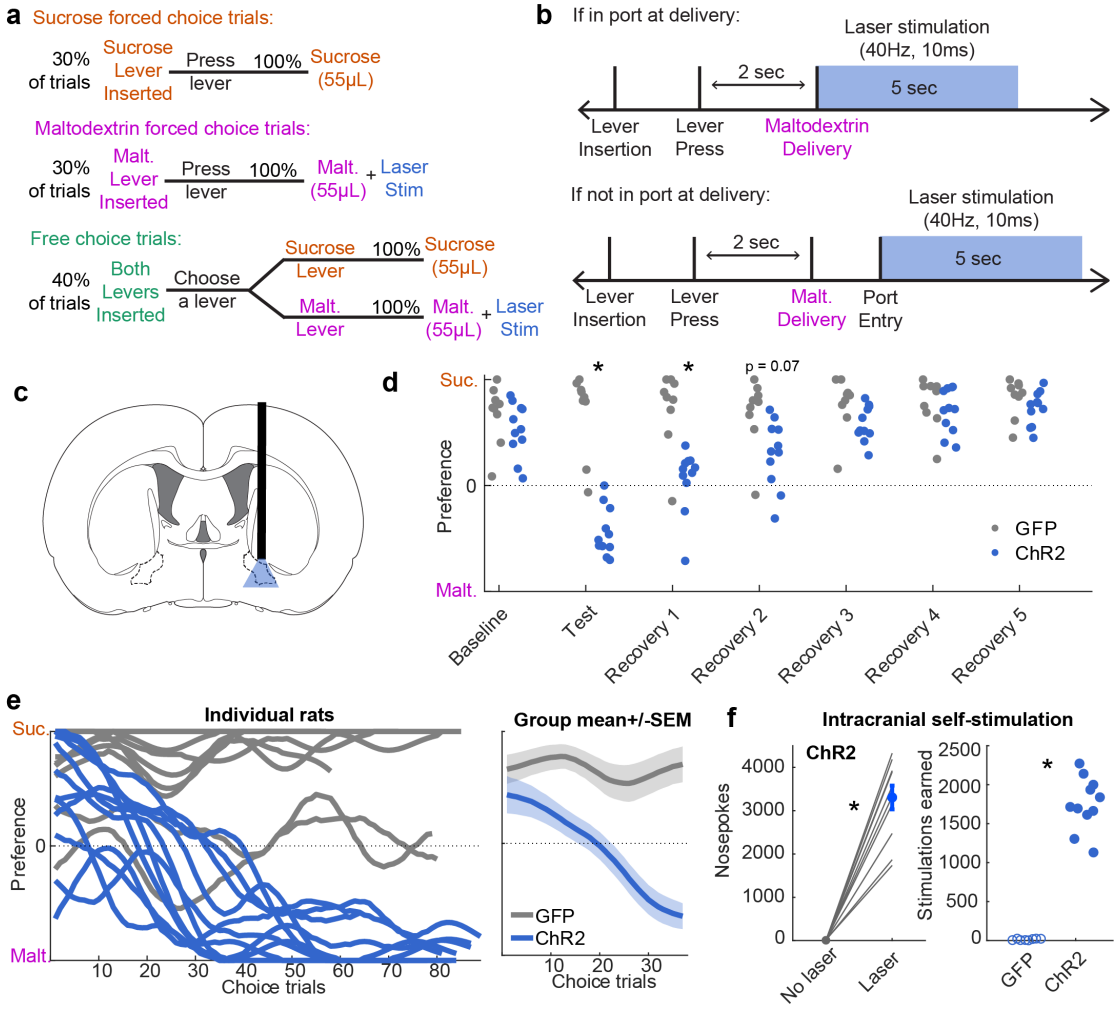

Figure 4.8. Stimulation of VP at reward delivery biases choice behavior.

Figure 4.8. Stimulation of VP at reward delivery biases choice behavior.

(a) Task design for optogenetic stimulation experiment. Rats chose between sucrose and maltodextrin. Trial type was indicated by lever insertion. Levers were retracted after press.

(b) During the test session, on maltodextrin trials (both forced and choice), VP was photostimulated unilaterally for 5 sec at 40Hz, beginning with maltodextrin delivery, or whenever the rat first entered the port thereafter.

(c) Optic fiber and virus containing ChR2 (or GFP control) were implanted bilaterally in VP, but only the right hemisphere was stimulated in this experiment.

(d) Preference for sucrose versus maltodextrin on choice trials at baseline (after training), on test session, and for 5 recovery days after without laser. There was a significant interaction between day and group across these 7 sessions ($F_{6,126} = 10.6, p < 0.00000001$). Post-hoc Tukey tests (corrected for multiple comparisons) revealed a significant difference between groups on test day ($p < 0.000001$) and the first recovery day ($p < 0.00001$).

(e) Preference (smoothed) on choice trials across the session for individual rats (left) and averaged for each group (right).

(f) Intracrancial self-stimulation (1s, 40Hz) of VP via nosepoke port. Rats made more nosepokes on laser-paired port than on the unpaired port during the 1 hr session ($p < 0.001$, Wilcoxon signed-rank test).

(g) ChR2 rats earned more stimulations during the 1 hr session than control rats ($p < 0.0001$, Wilcoxon rank-sum test).

Figure 4.9

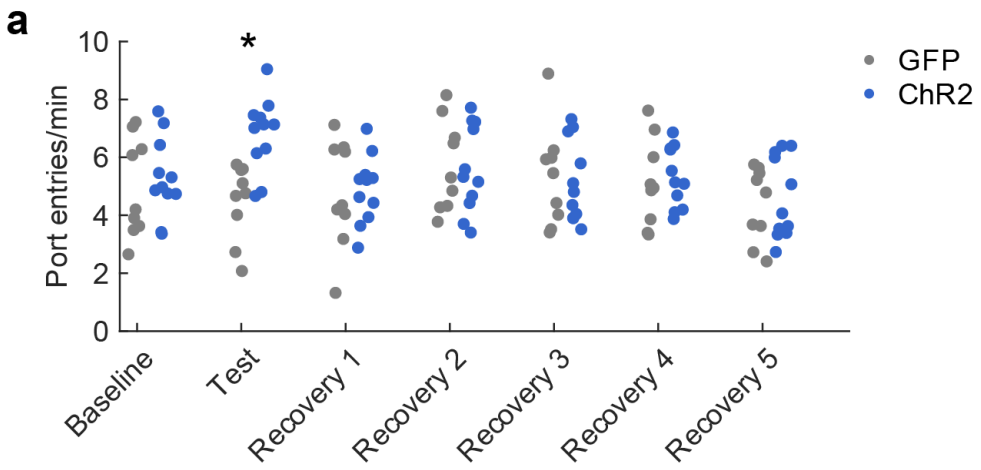

Figure 4.9. Impact of VP stimulation on port entries.

(a) Port entries per minute for all rats from the baseline, test, and recovery sessions. * = $p < 0.03$, Tukey test corrected for multiple comparisons, for difference between GFP and ChR2 groups on test day.

Figure 4.10

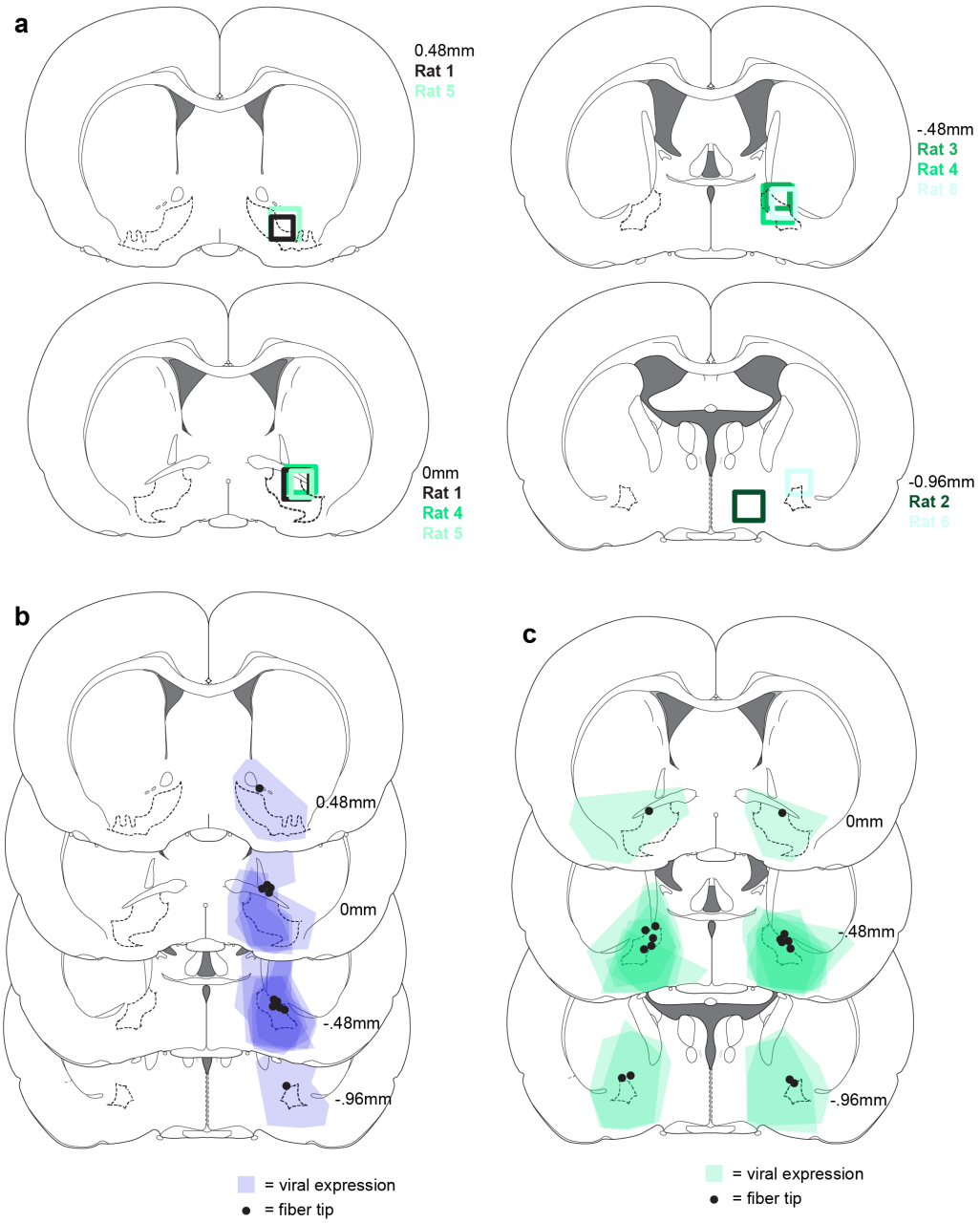

Figure 4.10. Placement for electrodes, fibers, and virus.

(a) Placements for electrophysiology recordings. Rat 2 was excluded from analysis.

(b) Placements for ChR2 experiments.

(c) Placements for ArchT3.0 experiments.

Chapter 5

General discussion

Our initial goal for this thesis was to examine how reward processing systems in the brain represent information about highly palatable outcomes. We chose to study the ventral striatopallidal system because of its well-theorized role in integrating information about rewards in order to direct an appropriate behavioral response. An initial question was whether nucleus accumbens (NAc) and ventral pallidum (VP) would have distinct neural responses to sucrose and maltodextrin at all given their equivalent caloric content, volume, and associated motor responses in our rats. In NAc, the answer was mixed; although many neurons were modulated during reward consumption (Fig. 2.4), only a fraction of these responses were specific to sucrose or maltodextrin (Fig. 2.6). On the other hand, a majority of neurons in VP had responses that were reward-specific, and these responses tended to occur earlier than in NAc (Figs. 2.6, 2.8). Remarkably, nearly all of these responses were phasic excitations to sucrose and, often, inhibitions to maltodextrin. This robust, uniform response that matched the rats' preference for sucrose motivated additional exploration of value signaling in VP.

The next step was to test whether VP signaling fit the properties of a relative value signal. We were able to do this by conducting additional recording sessions in the same rats where water was added as another outcome, first replacing sucrose and then all together. We were particularly interested in seeing whether the VP activity following maltodextrin delivery would change when maltodextrin was no longer the least preferred option. In both

session types this was the case; maltodextrin now evoked increases in VP activity. We were even able to see this shift within the first session with water instead of sucrose (Fig. 2.11). It was also notable that VP activity reflected the relative value of all three outcomes when they were presented in the same session (Fig. 2.12), which spoke to the robustness and flexibility of the signal.

We next examined the role of expectation in the signal we characterized. In the task the rats performed, the trials were demarcated by a single white noise cue that announced reward availability but did not indicate which reward the rats would receive. We chose this setup to limit the rats' expectation of which reward they would receive. Nevertheless, we observed that VP activity was sensitive to the previous outcome (Fig. 2.10), indicating that rats may have formed an expectation of upcoming reward based off of the recent outcome history rather than cued information. We wondered whether the activity of neurons in VP was consistent with a reward prediction error across trials. We developed a model fitting and classification approach that identified VP neurons whose activity was consistent with a prediction error integrating the outcomes on multiple previous trials (Fig. 3.1). Moreover, we found VP neurons followed this pattern of a trial-level prediction error in several conditions (Figs. 3.5, 3.6, 3.7).

Since we found neural evidence of trial-to-trial updates of expectation, we wondered whether the rats were adjusting their behavior accordingly. We analyzed the movements of the rats following reward delivery and preceding the onset of the cue for the next trial and found that their task engagement during this inter-trial interval was influenced by the just-received outcome, a measure that correlated with the activity of VP neurons (Fig. 3.3). We next asked if the activity of VP neurons drives this behavioral adaptation. We transiently inhibited or stimulated VP neurons following reward delivery using optogenetics and found a similar modulation of task engagement, suggesting that the reward-evoked activity in VP does influence adaptive behavior.

Although all of these experiments harnessed rats' varying preferences for the delivered

outcomes, none of them actually assessed their behavioral preference during the recording session. Moreover, they focused on stable preferences, whereas many preferences vary across different conditions. To address these points, we designed a task that allowed a more rigorous test of VP preference encoding. Thirsty rats chose between a large volume of water and a small volume of sucrose in a mixture of forced and free choice trials. Rats demonstrated an evolving preference by switching from choosing water to choosing sucrose as they became less thirsty. This changing preference was reflected in the activity of VP neurons, with remarkable agreement between behavioral and neural reports of the relative values of water and sucrose (Figs. 4.5, 4.6). Additionally, the timing of this preference-sensitive signal depended on when the identity of the outcome was revealed, following the pattern of a reward prediction error framework (Fig. 4.2).

With a tight link between VP activity and preference established, we wondered if VP activity informs rats' choices. When we optogenetically inhibited VP during the execution of the choice, we saw no impact on the choices the rats made (Fig. 4.4), indicating VP is not necessary for the expression of preference. On the other hand, optogenetic stimulation of VP following choice execution, at the time of reward consumption, was able to reverse rats' preference from sucrose to maltodextrin over the course of a single session, an effect that persisted to the following session (Fig. 4.8). These results suggest that VP activity at the time of the outcome instructs rats' behavioral preferences.

Our work here helps to clarify the role of VP in encoding reward preference and driving certain reward-seeking behaviors. There are some limitations of our studies that should be considered when interpreting our results, and there are many open questions remaining. We discuss these here.

5.1 Behavioral tasks.

In all of the experiments we conducted, the outcomes on the majority (if not all) of the trials were outside of the rats' control. Moreover, the behavioral response (at least for sucrose and

maltodextrin) was nearly identical regardless of the outcome. This leaves the question: why have an outcome-specific neural response at all? And why bother keeping track of which outcomes have been delivered?

One explanation could be that the reward-evoked activity is related to the experience of consuming the reward. In fact, VP is one of the few regions functionally linked to the experience of pleasure that comes from consuming palatable rewards (Smith and Berridge, 2007; Smith et al., 2009). This explanation is not consistent with the changes in VP firing we observed according to recent history, and when the available options were changed, unless the hedonic experience of consuming the rewards is also shifting. If so, we were unable to detect this just by examining lick traces; one potential future approach would be to video record the mouth movements of the rats, which are the best-established readouts of hedonia, while recording from VP.

Another explanation is that, even though rats have (little or) no control over which rewards they receive in these tasks, they still keep track of the statistics of reward availability because of how useful this information is generally for adaptive behavioral responding. We do see some evidence of adaptive behavioral responding in the rats' proximity to the reward port following sucrose and maltodextrin delivery, even though the currently received reward is not predictive of the subsequent reward. An important next question, then, would be to see whether VP value estimates guide behavior when the animal needs to adjust its behavior to changing reward statistics in order to maximize reward (Parker et al., 2016; Bari et al., 2019). Our data that optogenetic stimulation of VP can alter choice preference suggest that VP activity could be important for choice behavior in a dynamic decision-making task.

Another instance where task design limited our ability to interpret the results was in the predictable and random sucrose/maltodextrin task, where we aimed to explore an influence of a specific cue representation on outcome signaling (Fig. 3.7). Overall, we did not observe a consistent effect of specific predictive cues on outcome signaling in this task. There are considerable confounds for the conclusion that cue-mediated predictions do not impact VP

outcome signaling, however. First, there is no behavioral evidence that these cues were used by the rats; the latency to respond to each cue was similar. This may have been due to the similar behavioral relevance of the two highly palatable outcomes (sucrose and maltodextrin). Indeed, in tasks where the outcomes elicit different behavioral responses, VP cue-evoked activity readily discriminates among trial types (Tindell et al., 2009; Tachibana and Hikosaka, 2012; Richard et al., 2016, 2018; Ottenheimer et al., 2019b; Stephenson-Jones et al., 2020), whereas we only saw a modest difference in cue-evoked firing for each cue. A version of this task where the outcomes differed more in value might reveal a larger influence of cue-based expectation on VP signaling, both at the time of cue onset and outcome.

In the tasks where rats were able to choose between water and sucrose, we were able to see a shift in the timing of VP value encoding depending on whether the cues predicted a specific outcome. In some ways, this addressed the question we asked with the predictable and random sucrose/maltodextrin task. What allowed us to see an effect of cue-based expectation in this task? For one, the rats were first trained on a version of the task with only specific cues; there were no trials intermixed where the outcome was uncertain. Moreover, the behavioral response for each trial type was distinct–the cue indicated which lever the rat needed to press. Finally, because the value of the rewards changed independently throughout the session due to the effects of satiety, the outcome may have been more behaviorally relevant than sucrose versus maltodextrin.

Because the outcome was never uncertain in the first choice task, we had to train rats on a new task without specific cues (or lever presses) to analyze neural activity in uncertain outcome conditions; in this second task with no specific cues, there was no possibility for cue-mediated expectation and, consequently, we saw a robust outcome-specific signal at the time of reward delivery. Unfortunately, because these were separate tasks (on separate days), we could not compare the activity of individual neurons on trials with certain or uncertain outcome. Thus, one future direction would be to take the strengths from the sucrose versus water task that permitted a robust cue-mediated expectation and apply them to a new task

that would allow a within-session comparison of the effects of specific versus non-specific cues. This would permit more rigorous conclusions on temporal difference error encoding in single neurons in VP.

5.2 RPE encoding in this circuit.

One major question this work leaves open is how the activity we observed in VP might relate to the activity of dopamine neurons in VTA given their smilarity in coding schemes. In our data, we saw more RPE-like activity in VP than in NAc, another input to midbrain dopamine neurons, leaving the possibility that VP is a privileged provider of outcome history-based predictions to dopamine neurons. There are multiple routes for VP activity to reach VTA given multiple demonstrations of VP synapses not only onto VTA neurons but also onto other VTA input nuclei like lateral habenula and rostromedial tegemental nucleus (Hong and Hikosaka, 2013; Tian et al., 2016; Knowland et al., 2017; Tooley et al., 2018; Faget et al., 2018). Stimulation of VP GABAergic neurons increases the number of putative midbrain dopamine neurons expressing Fos, consistent with an indirect mechanism for modulating features of dopamine neuron RPE signaling (Faget et al., 2018). In songbird, VP has been shown to send performance-related error signals to the ventral tegmental area during singing (Chen et al., 2019; Kearney et al., 2019). On the other hand, VTA could be the source of VP RPE signals; in addition to dopaminergic innvervation of VP (Root et al., 2015), VTA also has dense glutamatergic projections to VP, which could provide the early onset phasic responses we observed in VP (Hnasko et al., 2012). Future work should untangle a uni- or bidirectional role of error-related signals in these regions, as well as possible unique roles of each population in adaptive behavior. A combination of projection-specific recordings and manipulations (perhaps simultaneously in both regions) would help clarify this question. Additionally, because NAc is not the likely source of the signals we characterized in VP, the role of other VP inputs like central amygdala, lateral hypothalamus, preoptic area, and the bed nucleus of the striatum should be investigated (Knowland et al., 2017; Tooley et al., 2018;

Stephenson-Jones et al., 2020), and a broadening of the conceptualization of information flow in this circuit is in order.

5.3 Ventral pallidum anatomy.

Prior to being termed the ventral pallidum, this region of the brain was known as substantia innominata ('unnamed substance'), which also included the region now known as extended amygdala. These regions were designated from the expression patterns of certain receptors and connectivity patterns with other regions, but their separate identities have been debated (de Olmos and Heimer, 1999). Our work has some relevance to this debate. There were instances, because they are neighboring regions, when we recorded from or stimulated extended amygdala as well as ventral pallidum (Figs. 2.2, 3.4, 3.8). In all instances, the patterns of neural activity and effects from the stimulation were the same as data from rats with placements fully in VP. It is possible that drawing a precise line between these regions masks a gradient of responses and functions.

As an increasingly complicated picture of cell types with distinct connectivity patterns emerge in mouse VP (Knowland et al., 2017; Tooley et al., 2018; Faget et al., 2018; Stephenson-Jones et al., 2020), it will be important to see whether the same anatomy applies to rat VP (and beyond). One particularly impactful finding is the opposing effects of stimulating GABAergic and glutamatergic cells in VP on behavior (Faget et al., 2018; Stephenson-Jones et al., 2020). It would be useful to know whether those trends hold in rats and whether the functional subtypes of neurons we classified (reward-selective, RPE-encoding, etc.) map onto a certain neurotransmitter type (or projection pattern). Given one report that the activity of GABAergic neurons in VP tends to correlate positively with value, we could speculate that the neurons we studied are likely to be GABAergic (Stephenson-Jones et al., 2020). With increasing numbers of transgenic rats and viral tools available, these questions should be answerable in the near future.

5.4 Beyond reward.

In our experiments, the outcomes ranged from palatable to neutral orally ingested stimuli. There are many reinforcing stimuli in the environment beyond the reward landscape we explored, and it is possible the kinds of value signals we characterized in VP could integrate them as well. For instance, a role for VP has been established in social behaviors, maternal behaviors, aversion, and drug abuse (Smith et al., 2009; Root et al., 2015). Initial reports demonstrate that VP activity scales with aversive stimuli along with reward (Tian et al., 2016; Stephenson-Jones et al., 2020). Thus, VP activity is a compelling neural substrate for continued exploration of representations of competing (or synergistic) reinforcing stimuli of different modalities.

5.5 Cell classification.

A number of conclusions in this dissertation are derived from the numbers of neurons classified into a given category – for instance, reward-selective cells in the task contrasting sucrose and maltodextrin. For that particular metric, the inclusion criteria were selected to minimize noisy classification (cells counting as reward-selective before rewards were delivered, which could only arise due to chance); this entailed meeting a $p < 0.01$ cutoff for two consecutive time bins. As with many statistical cutoffs, neurons classified as selective or not are not necessarily two separate populations of neurons. Rather, the classification is judgment of how well the selectivity of those neurons can be distinguished from noise. There are many factors that could impact the ability to distinguish reward selectivity from noise, including the magnitude of the test stimuli. For this reason, it is hard to make the distinction between more neurons being recruited to report relative value in the sessions with water, or whether it is just easier to statistically identify the neurons when the stimuli span a wider range in value.

This issue is also relevant for the model fitting and classification procedure. Model selec-

tion with Akaike information criterion makes a judgment about which model most parsimoniously describes the activity of a neuron based on the error between the model's predictions of a neuron's spikes on each trial and the actual spikes of that neuron. Thus, RPE neurons and Current Outcome neurons are not necessarily different populations of neurons; it is just a distinction of whether the RPE model did enough of a better job than the Current Outcome model in describing those neurons' spikes. Again, this leads to a difficult interpretation of there being more neurons classified as RPE-encoding in the task with water. This not a unique problem to our data; many studies looking for RPE correlates demonstrate a whole distribution of responses with no clear distinct subpopulation of RPE-encoding cells (Takahashi et al., 2011; Engelhard et al., 2019). The approach of classifying neurons into distinct subgroups is a practice that should be carefully considered across the field.

One place we explored the distribution of model fits more carefully was for the Satiety, Preference, and Mixed models (Fig. 4.7). By looking at the relative weight of satiety and preference in the Mixed model fits for neurons from all three categories, we were able to see whether these neurons were really distinct categories. Interestingly, these populations did separate fairly well, with separate peaks in the distributions, suggesting there are separate sets of neurons that encode satiety, preference, or a mixture between them, although the exact mapping of these firing patterns onto more loaded concepts like preference and satiety could be debated, and other explanations are possible. The application of these models to the firing rates of neurons from other regions in similar tasks could lend additional support for this interpretation.

Bibliography

Ahrens AM, Meyer PJ, Ferguson LM, Robinson TE, Aldridge JW. Neural activity in the ventral pallidum encodes variation in the incentive value of a reward cue. *Journal of Neuroscience* 36: 7957–7970, 2016.

Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience* 9: 357–381, 1986.

Allen WE, Chen MZ, Pichamoorthy N, Tien RH, Pachitariu M, Luo L, Deisseroth K. Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science* 364: 253–253, 2019.

Ambroggi F, Ghazizadeh A, Nicola SM, Fields HL. Roles of nucleus accumbens core and shell in incentive-cue responding and behavioral inhibition. *Journal of Neuroscience* 31: 6820–6830, 2011.

Asaad WF, Eskandar EN. Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. *Journal of Neuroscience* 31: 17772–17787, 2011.

Avila I, Lin SC. Distinct neuronal populations in the basal forebrain encode motivational salience and movement. *Frontiers in behavioral neuroscience* 8: 421, 2014a.

Avila I, Lin SC. Motivational salience signal in the basal forebrain is coupled with faster and more precise decision speed. *PLoS biology* 12, 2014b.

Baldo BA, Daniel RA, Berridge CW, Kelley AE. Overlapping distributions of orexin/hypocretin-and dopamine-β-hydroxylase immunoreactive fibers in rat brain regions mediating arousal, motivation, and stress. *Journal of Comparative Neurology* 464: 220–237, 2003.

Bari BA, Grossman CD, Lubin EE, Rajagopalan AE, Cressy JI, Cohen JY. Stable Representations of Decision Variables for Flexible Behavior. *Neuron* , 2019.

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

Behrens TE, Woolrich MW, Walton ME, Rushworth MF. Learning the value of information in an uncertain world. *Nature neuroscience* 10: 1214, 2007.

Beier KT, Steinberg EE, DeLoach KE, Xie S, Miyamichi K, Schwarz L, Gao XJ, Kremer EJ, Malenka RC, Luo L. Circuit architecture of VTA dopamine neurons revealed by systematic input-output mapping. *Cell* 162: 622–634, 2015.

Bermudez MA, Schultz W. Reward magnitude coding in primate amygdala neurons. *Journal of neurophysiology* 104: 3424–3432, 2010.

Berridge KC. Motivation concepts in behavioral neuroscience. *Physiology & behavior* 81: 179–209, 2004.

Berridge KC, Flynn FW, Schulkin J, Grill HJ. Sodium depletion enhances salt palatability in rats. *Behavioral neuroscience* 98: 652, 1984.

Bissonette GB, Burton AC, Gentry RN, Goldstein BL, Hearn TN, Barnett BR, Kashtelyan V, Roesch MR. Separate populations of neurons in ventral striatum encode value and motivation. *PLoS One* 8, 2013.

Bloem B, Huda R, Sur M, Graybiel AM. Two-photon imaging in mice shows striosomes and

matrix have overlapping but differential reinforcement-related responses. *Elife* 6: e32353, 2017.

Bouret S, Richmond BJ. Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *Journal of Neuroscience* 30: 8591–8601, 2010.

Burgess CR, Ramesh RN, Sugden AU, Levandowski KM, Minnig MA, Fenselau H, Lowell BB, Andermann ML. Hunger-dependent enhancement of food cue responses in mouse postrhinal cortex and lateral amygdala. *Neuron* 91: 1154–1169, 2016.

Cabanac M. Physiological role of pleasure. *Science* 173: 1103–1107, 1971.

Cai X, Padoa-Schioppa C. Neuronal encoding of subjective value in dorsal and ventral anterior cingulate cortex. *Journal of Neuroscience* 32: 3791–3808, 2012.

Carelli RM, Ijames SG, Crumling AJ. Evidence that separate neural circuits in the nucleus accumbens encode cocaine versus "natural"(water and food) reward. *Journal of Neuroscience* 20: 4255–4266, 2000.

Castro DC, Berridge KC. Opioid hedonic hotspot in nucleus accumbens shell: mu, delta, and kappa maps for enhancement of sweetness "liking" and "wanting". *Journal of Neuroscience* 34: 4239–4250, 2014.

Chang SE, Todd TP, Smith KS. Paradoxical accentuation of motivation following accumbens-pallidum disconnection. *Neurobiology of learning and memory* 149: 39–45, 2018.

Chen R, Puzerey PA, Roeser AC, Riccelli TE, Podury A, Maher K, Farhang AR, Goldberg JH. Songbird Ventral Pallidum Sends Diverse Performance Error Signals to Dopaminergic Midbrain. *Neuron* , 2019.

Chrobak J, Napier T. Opioid and GABA modulation of accumbens-evoked ventral pallidal activity. *Journal of Neural Transmission/General Section JNT* 93: 123–143, 1993.

Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *nature* 482: 85, 2012.

Cooch NK, Stalnaker TA, Wied HM, Bali-Chaudhary S, McDannald MA, Liu TL, Schoenbaum G. Orbitofrontal lesions eliminate signalling of biological significance in cue-responsive ventral striatal neurons. *Nature communications* 6: 7195, 2015.

Creed M, Ntamati NR, Chandra R, Lobo MK, Lüscher C. Convergence of reinforcing and anhedonic cocaine effects in the ventral pallidum. *Neuron* 92: 214–226, 2016.

Critchley HD, Rolls ET. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *Journal of neurophysiology* 75: 1673–1686, 1996.

Cromwell HC, Hassani OK, Schultz W. Relative reward processing in primate striatum. *Experimental Brain Research* 162: 520–525, 2005.

Day JJ, Jones JL, Carelli RM. Nucleus accumbens neurons encode predicted and ongoing reward costs in rats. *European journal of neuroscience* 33: 308–321, 2011.

Day JJ, Jones JL, Wightman RM, Carelli RM. Phasic nucleus accumbens dopamine release encodes effort-and delay-related costs. *Biological psychiatry* 68: 306–309, 2010.

de Araujo IE, Gutierrez R, Oliveira-Maia AJ, Pereira Jr A, Nicolelis MA, Simon SA. Neural ensemble coding of satiety states. *Neuron* 51: 483–494, 2006.

de Olmos JS, Heimer L. The concepts of the ventral striatopallidal system and extended amygdala. *Annals of the New York Academy of Sciences* 877: 1–32, 1999.

Eagle DM, Humby T, Howman M, Reid-Henry A, Dunnett SB, Robbins TW. Differential effects of ventral and regional dorsal striatal lesions on sucrose drinking and positive and negative contrast in rats. *Psychobiology* 27: 267–276, 1999.

Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw ND, Tank DW, et al. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* p. 1, 2019.

Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N. Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* 525: 243, 2015.

Eshel N, Tian J, Bukwich M, Uchida N. Dopamine neurons share common response function for reward prediction error. *Nature neuroscience* 19: 479, 2016.

Faget L, Zell V, Souter E, McPherson A, Ressler R, Gutierrez-Reed N, Yoo JH, Dulcis D, Hnasko TS. Opponent control of behavioral reinforcement by inhibitory and excitatory projections from the ventral pallidum. *Nature communications* 9: 849, 2018.

Farrar AM, Font L, Pereira M, Mingote S, Bunce JG, Chrobak JJ, Salamone JD. Forebrain circuitry involved in effort-related choice: Injections of the GABAA agonist muscimol into ventral pallidum alter response allocation in food-seeking behavior. *Neuroscience* 152: 321–330, 2008.

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

Flaherty CF. *Incentive relativity*, vol. 15. Cambridge University Press, 1999.

Fujimoto A, Hori Y, Nagai Y, Kikuchi E, Oyama K, Suhara T, Minamimoto T. Signaling incentive and drive in the primate ventral pallidum for motivational control of goal-directed action. *Journal of Neuroscience* 39: 1793–1804, 2019.

Gallo EF, Meszaros J, Sherman JD, Chohan MO, Teboul E, Choi CS, Moore H, Javitch JA, Kellendonk C. Accumbens dopamine D2 receptors increase motivation by decreasing inhibitory transmission to the ventral pallidum. *Nature communications* 9: 1–13, 2018.

Gan JO, Walton ME, Phillips PE. Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nature neuroscience* 13: 25–27, 2010.

Goldstein BL, Barnett BR, Vasquez G, Tobia SC, Kashtelyan V, Burton AC, Bryden DW, Roesch MR. Ventral striatum encodes past and predicted value independent of motor contingencies. *Journal of Neuroscience* 32: 2027–2036, 2012.

Groenewegen HJ, Russchen FT. Organization of the efferent projections of the nucleus accumbens to pallidal, hypothalamic, and mesencephalic structures: a tracing and immunohistochemical study in the cat. *Journal of Comparative Neurology* 223: 347–367, 1984.

Groenewegen HJ, Wright CI, Beijer AV, Voorn P. Convergence and segregation of ventral striatal inputs and outputs. *Annals of the New York Academy of Sciences* 877: 49–63, 1999.

Hakan RL, Berg GI, Henriksen SJ. Electrophysiological evidence for reciprocal connectivity between the nucleus accumbens septi and ventral pallidal region. *Brain research* 581: 344–350, 1992.

Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT, Aragona BJ, Berke JD. Mesolimbic dopamine signals the value of work. *Nature neuroscience* 19: 117, 2016.

Han X, Chow BY, Zhou H, Klapoetke NC, Chuong A, Rajimehr R, Yang A, Baratta MV, Winkle J, Desimone R, et al. A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. *Frontiers in systems neuroscience* 5: 18, 2011.

Heimer L, Wilson R. The subcortical projections of the allocortex: Similarities in the neural associations of the hippocampus, the piriform cortex, and the neocortex. *Golgi Centennial Symposium* pp. 177–193, 1975.

Hnasko TS, Hjelmstad GO, Fields HL, Edwards RH. Ventral tegmental area glutamate neurons: electrophysiological properties and projections. *Journal of Neuroscience* 32: 15076–15085, 2012.

Ho CY, Berridge KC. An orexin hotspot in ventral pallidum amplifies hedonic 'liking'for sweetness. *Neuropsychopharmacology* 38: 1655–1664, 2013.

Hong S, Hikosaka O. Diverse sources of reward value signals in the basal ganglia nuclei transmitted to the lateral habenula in the monkey. *Frontiers in human neuroscience* 7: 778, 2013.

Hong S, Jhou TC, Smith M, Saleem KS, Hikosaka O. Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. *Journal of Neuroscience* 31: 11457–11471, 2011.

Hull CL. *Principles of behavior*, vol. 422. Appleton-century-crofts New York, 1943.

Ito M, Doya K. Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *Journal of Neuroscience* 29: 9861–9874, 2009.

Itoga CA, Berridge KC, Aldridge JW. Ventral pallidal coding of a learned taste aversion. *Behavioural brain research* 300: 175–183, 2016.

Janak PH, Chen MT, Caulder T. Dynamics of neural coding in the accumbens during extinction and reinstatement of rewarded behavior. *Behavioural brain research* 154: 125–135, 2004.

Jhou TC, Fields HL, Baxter MG, Saper CB, Holland PC. The rostromedial tegmental

nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. *Neuron* 61: 786–800, 2009.

Jones D, Mogenson G. Nucleus accumbens to globus pallidus GABA projection subserving ambulatory activity. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 238: R65–R69, 1980.

Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal choice. *Nature neuroscience* 10: 1625–1633, 2007.

Kamin LJ. Attention-like processes in classical conditioning. *Miami Symposium on the prediction of behavior: Aversive stimulation* pp. 177–193, 1968.

Kearney MG, Warren TL, Hisey E, Qi J, Mooney R. Discrete Evaluative and Premotor Circuits Enable Vocal Learning in Songbirds. *Neuron* , 2019.

Keiflin R, Janak PH. Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. *Neuron* 88: 247–263, 2015.

Kelley AE. Functional specificity of ventral striatal compartments in appetitive behaviors. *Annals of the New York Academy of Sciences* 877: 71–90, 1999.

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

Keramati M, Gutkin B. Homeostatic reinforcement learning for integrating reward collection and physiological stability. *Elife* 3: e04811, 2014.

Kim H, Lee D, Jung MW. Signals for previous goal choice persist in the dorsomedial, but not dorsolateral striatum of rats. *Journal of Neuroscience* 33: 52–63, 2013.

Knowland D, Lilascharoen V, Pacia CP, Shin S, Wang EHJ, Lim BK. Distinct ventral pallidal neural populations mediate separate symptoms of depression. *Cell* 170: 284–297, 2017.

Kobayashi S, Schultz W. Influence of reward delays on responses of dopamine neurons. *Journal of neuroscience* 28: 7837–7846, 2008.

Lak A, Stauffer WR, Schultz W. Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proceedings of the National Academy of Sciences* 111: 2343–2348, 2014.

Langdon AJ, Sharpe MJ, Schoenbaum G, Niv Y. Model-based predictions for dopamine. *Current Opinion in Neurobiology* 49: 1–7, 2018.

Leszczuk MH, Flaherty CF. Lesions of nucleus accumbens reduce instrumental but not consummatory negative contrast in rats. *Behavioural Brain Research* 116: 61–79, 2000.

Leung BK, Balleine BW. The ventral striato-pallidal pathway mediates the effect of predictive learning on choice between goal-directed actions. *Journal of Neuroscience* 33: 13848–13860, 2013.

Lin SC, Nicolelis MA. Neuronal ensemble bursting in the basal forebrain encodes salience irrespective of valence. *Neuron* 59: 138–149, 2008.

Livneh Y, Ramesh RN, Burgess CR, Levandowski KM, Madara JC, Fenselau H, Goldey GJ, Diaz VE, Jikomes N, Resch JM, et al. Homeostatic circuits selectively gate food cue responses in insular cortex. *Nature* 546: 611–616, 2017.

Livneh Y, Sugden AU, Madara JC, Essner RA, Flores VI, Sugden LA, Resch JM, Lowell BB, Andermann ML. Estimation of Current and Future Physiological States in Insular Cortex. *Neuron* , 2020.

Louie K, Glimcher PW. Efficient coding and the neural representation of value. *Annals of the New York Academy of Sciences* 1251: 13–32, 2012.

Louie K, Grattan LE, Glimcher PW. Reward value-based gain control: divisive normalization in parietal cortex. *Journal of Neuroscience* 31: 10627–10639, 2011.

Lu XY, Ghasemzadeh MB, Kalivas P. Expression of D1 receptor, D2 receptor, substance P and enkephalin messenger RNAs in the neurons projecting from the nucleus accumbens. *Neuroscience* 82: 767–780, 1997.

Mahler SV, Vazey EM, Beckley JT, Keistler CR, McGlinchey EM, Kaufling J, Wilson SP, Deisseroth K, Woodward JJ, Aston-Jones G. Designer receptors show role for ventral pallidum input to ventral tegmental area in cocaine seeking. *Nature neuroscience* 17: 577, 2014.

Maslowski-Cobuzzi R, Napier T. Activation of dopaminergic neurons modulates ventral pallidal responses evoked by amygdala stimulation. *Neuroscience* 62: 1103–1119, 1994.

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* , 2018.

Matsumoto M, Hikosaka O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 447: 1111, 2007.

Maurice N, Deniau J, Menetrey A, Glowinski J, Thierry A. Position of the ventral pallidum in the rat prefrontal cortex–basal ganglia circuit. *Neuroscience* 80: 523–534, 1997.

Maurice N, Deniau JM, Glowinski J, Thierry AM. Relationships between the prefrontal cortex and the basal ganglia in the rat: physiology of the cortico-nigral circuits. *Journal of Neuroscience* 19: 4674–4681, 1999.

Mitrovic I, Napier TC. Substance P attenuates and DAMGO potentiates amygdala glutamatergic neurotransmission within the ventral pallidum. *Brain research* 792: 193–206, 1998.

Mogenson GJ, Jones DL, Yim CY. From motivation to action: functional interface between the limbic system and the motor system. *Progress in neurobiology* 14: 69–97, 1980.

Mogenson GJ, Nielsen M. A study of the contribution of hippocampal—accumbens—subpallidal projections to locomotor activity. *Behavioral and neural biology* 42: 38–51, 1984.

Mohebi A, Pettibone JR, Hamid AA, Wong JMT, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD. Dissociable dopamine dynamics for learning and motivation. *Nature* 570: 65–70, 2019.

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

Nambu A, Tokuno H, Takada M. Functional significance of the cortico–subthalamo–pallidal 'hyperdirect'pathway. *Neuroscience research* 43: 111–117, 2002.

Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature protocols* , 2019.

Nauta W, Smith G, Faull R, Domesick VB. Efferent connections and nigral afferents of the nucleus accumbens septi in the rat. *Neuroscience* 3: 385–401, 1978.

Nicola SM, Yun IA, Wakabayashi KT, Fields HL. Cue-evoked firing of nucleus accumbens neurons encodes motivational significance during a discriminative stimulus task. *Journal of neurophysiology* 91: 1840–1865, 2004.

Nissenbaum JW, Sclafani A. Qualitative differences in polysaccharide and sugar tastes in the rat: a two-carbohydrate taste model. *Neuroscience & Biobehavioral Reviews* 11: 187–196, 1987.

Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* 191: 507–520, 2007.

Niyogi RK, Breton YA, Solomon RB, Conover K, Shizgal P, Dayan P. Optimal indolence:

a normative microscopic approach to work and leisure. *Journal of The Royal Society Interface* 11: 20130969, 2014.

Ottenheimer D, Richard JM, Janak PH. Ventral pallidum encodes relative reward value earlier and more robustly than nucleus accumbens. *Nature communications* 9: 4350, 2018.

Ottenheimer DJ, Bari BA, Sutlief E, Fraser KM, Kim TH, Richard JM, Cohen JY, Janak PH. A history-derived reward prediction error signal in ventral pallidum. *bioRxiv* p. 807842, 2019a.

Ottenheimer DJ, Wang K, Haimbaugh A, Janak PH, Richard JM. Recruitment and disruption of ventral pallidal cue encoding during alcohol seeking. *European Journal of Neuroscience* , 2019b.

Padoa-Schioppa C. Range-adapting representation of economic value in the orbitofrontal cortex. *Journal of Neuroscience* 29: 14004–14014, 2009.

Pan WX, Schmidt R, Wickens JR, Hyland BI. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *Journal of Neuroscience* 25: 6235–6242, 2005.

Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, Witten IB. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nature neuroscience* 19: 845, 2016.

Pavlov IP. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford University Press London, 1927.

Pecina S, Berridge KC. Hedonic hot spot in nucleus accumbens shell: where do μ-opioids cause increased hedonic impact of sweetness? *Journal of neuroscience* 25: 11777–11786, 2005.

Prasad AA, Xie C, Chaichim C, Nguyen JH, McClusky HE, Killcross S, Power JM, McNally GP. Complementary roles for ventral pallidum cell types and their projections in relapse. *Journal of Neuroscience* 40: 880–893, 2020.

Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory* 2: 64–99, 1972.

Richard JM, Ambroggi F, Janak PH, Fields HL. Ventral pallidum neurons encode incentive value and promote cue-elicited instrumental actions. *Neuron* 90: 1165–1173, 2016.

Richard JM, Castro DC, DiFeliceantonio AG, Robinson MJ, Berridge KC. Mapping brain circuits of reward and motivation: in the footsteps of Ann Kelley. *Neuroscience & Biobehavioral Reviews* 37: 1919–1931, 2013a.

Richard JM, Plawecki AM, Berridge KC. Nucleus accumbens GABA ergic inhibition generates intense eating and fear that resists environmental retuning and needs no local dopamine. *European Journal of Neuroscience* 37: 1789–1802, 2013b.

Richard JM, Stout N, Acs D, Janak PH. Ventral pallidal encoding of reward-seeking behavior depends on the underlying associative structure. *Elife* 7: e33107, 2018.

Robinson DL, Carelli RM. Distinct subsets of nucleus accumbens neurons encode operant responding for ethanol versus water. *European Journal of Neuroscience* 28: 1887–1894, 2008.

Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature neuroscience* 10: 1615, 2007.

Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G. Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *Journal of Neuroscience* 29: 13365–13376, 2009.

Roitman MF, Wheeler RA, Carelli RM. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron* 45: 587–597, 2005.

Rolls BJ, Rolls ET, Rowe EA, Sweeney K. Sensory specific satiety in man. *Physiol Behav* 27: 137–142, 1981.

Rolls ET, Murzi E, Yaxley S, Thorpe S, Simpson S. Sensory-specific satiety: food-specific reduction in responsiveness of ventral forebrain neurons after feeding in the monkey. *Brain research* 368: 79–86, 1986.

Rolls ET, Sienkiewicz ZJ, Yaxley S. Hunger modulates the responses to gustatory stimuli of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey. *European Journal of Neuroscience* 1: 53–60, 1989.

Root DH, Melendez RI, Zaborszky L, Napier TC. The ventral pallidum: Subregion-specific functional anatomy and roles in motivated behaviors. *Progress in neurobiology* 130: 29–70, 2015.

Ryan L, Clark K. The role of the subthalamic nucleus in the response of globus pallidus neurons to stimulation of the prelimbic and agranular frontal cortices in rats. *Experimental brain research* 86: 641–651, 1991.

Saez RA, Saez A, Paton JJ, Lau B, Salzman CD. Distinct roles for the amygdala and orbitofrontal cortex in representing the relative amount of expected reward. *Neuron* 95: 70–77, 2017.

Sako N, Shimura T, Komure M, Mochizuki R, Matsuo R, Yamamoto T. Differences in taste responses to Polycose and common sugars in the rat as revealed by behavioral and electrophysiological studies. *Physiology & behavior* 56: 741–745, 1994.

Schultz W. Updating dopamine reward signals. *Current opinion in neurobiology* 23: 229–238, 2013.

Schultz W. Neuronal reward and decision signals: from theories to data. *Physiological reviews* 95: 853–951, 2015.

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

Sclafani A, Hertwig H, Vigorito M, Feigin MB. Sex differences in polysaccharide and sugar preferences in rats. *Neuroscience & Biobehavioral Reviews* 11: 241–251, 1987.

Setlow B, Schoenbaum G, Gallagher M. Neural encoding in ventral striatum during olfactory discrimination learning. *Neuron* 38: 625–636, 2003.

Shin JH, Kim D, Jung MW. Differential coding of reward and movement information in the dorsomedial striatal direct and indirect pathways. *Nature communications* 9: 404, 2018.

Smedley EB, DiLeo A, Smith KS. Circuit directionality for motivation: lateral accumbens-pallidum, but not pallidum-accumbens, connections regulate motivational attraction to reward cues. *Neurobiology of learning and memory* 162: 23–35, 2019.

Smith KS, Berridge KC. The ventral pallidum and hedonic reward: neurochemical maps of sucrose "liking" and food intake. *Journal of neuroscience* 25: 8637–8649, 2005.

Smith KS, Berridge KC. Opioid limbic circuit for reward: interaction between hedonic hotspots of nucleus accumbens and ventral pallidum. *Journal of neuroscience* 27: 1594–1605, 2007.

Smith KS, Berridge KC, Aldridge JW. Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proceedings of the National Academy of Sciences* 108: E255–E264, 2011.

Smith KS, Tindell AJ, Aldridge JW, Berridge KC. Ventral pallidum roles in reward and motivation. *Behavioural brain research* 196: 155–167, 2009.

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G. Reward prediction error signaling in posterior dorsomedial striatum is action specific. *Journal of Neuroscience* 32: 10296–10305, 2012.

Stephenson-Jones M, Bravo-Rivera C, Ahrens S, Furlan A, Xiao X, Fernandes-Henriques C, Li B. Opposing Contributions of GABAergic and Glutamatergic Ventral Pallidal Neurons to Motivational Behaviors. *Neuron* , 2020.

Sugam JA, Day JJ, Wightman RM, Carelli RM. Phasic nucleus accumbens dopamine encodes risk-based decision-making behavior. *Biological psychiatry* 71: 199–205, 2012.

Sutton RS. Learning to predict by the methods of temporal differences. *Machine learning* 3: 9–44, 1988.

Sutton RS, Barto AG. *Introduction to reinforcement learning*, vol. 2. MIT press Cambridge, 1998.

Swerdlow N, Koob G. Lesions of the dorsomedial nucleus of the thalamus, medial prefrontal cortex and pedunculopontine nucleus: effects on locomotor activity mediated by nucleus accumbens-ventral pallidal circuitry. *Brain research* 412: 233–243, 1987.

Tachibana Y, Hikosaka O. The primate ventral pallidum encodes expected reward value and regulates motor action. *Neuron* 76: 826–837, 2012.

Taha SA, Fields HL. Encoding of palatability and appetitive behaviors by distinct neuronal populations in the nucleus accumbens. *Journal of Neuroscience* 25: 1193–1202, 2005.

Takahashi YK, Langdon AJ, Niv Y, Schoenbaum G. Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. *Neuron* 91: 182–193, 2016.

Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'donnell P, Niv Y, Schoenbaum G. Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature neuroscience* 14: 1590, 2011.

Tian J, Huang R, Cohen JY, Osakada F, Kobak D, Machens CK, Callaway EM, Uchida N, Watabe-Uchida M. Distributed and mixed information in monosynaptic inputs to dopamine neurons. *Neuron* 91: 1374–1389, 2016.

Tian J, Uchida N. Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. *Neuron* 87: 1304–1316, 2015.

Tindell AJ, Berridge KC, Aldridge JW. Ventral pallidal representation of pavlovian cues and reward: population and rate codes. *Journal of Neuroscience* 24: 1058–1069, 2004.

Tindell AJ, Smith KS, Berridge KC, Aldridge JW. Dynamic computation of incentive salience:"wanting" what was never "liked". *Journal of Neuroscience* 29: 12220–12228, 2009.

Tindell AJ, Smith KS, Peciña S, Berridge KC, Aldridge JW. Ventral pallidum firing codes hedonic reward: when a bad taste turns good. *Journal of neurophysiology* 96: 2399–2409, 2006.

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

Tooley J, Marconi L, Alipio JB, Matikainen-Ankney B, Georgiou P, Kravitz AV, Creed MC. Glutamatergic ventral pallidal neurons modulate activity of the habenula–tegmental circuitry and constrain reward seeking. *Biological psychiatry* 83: 1012–1023, 2018.

Treesukosol Y, Smith KR, Spector AC. Behavioral evidence for a glucose polymer taste receptor that is independent of the T1R2+ 3 heterodimer in a mouse model. *Journal of Neuroscience* 31: 13527–13534, 2011.

Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. *Nature* 398: 704, 1999.

Turner MS, Lavin A, Grace AA, Napier TC. Regulation of limbic information outflow by the subthalamic nucleus: excitatory amino acid projections to the ventral pallidum. *Journal of Neuroscience* 21: 2820–2832, 2001.

Villavicencio M, Moreno MG, Simon SA, Gutierrez R. Encoding of Sucrose's Palatability in the Nucleus Accumbens Shell and Its Modulation by Exteroceptive Auditory Cues. *Frontiers in neuroscience* 12: 265, 2018.

Wang AY, Miura K, Uchida N. The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nature neuroscience* 16: 639, 2013.

Watabe-Uchida M, Eshel N, Uchida N. Neural circuitry of reward prediction error. *Annual review of neuroscience* 40: 373–394, 2017.

Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74: 858–873, 2012.

Webber ES, Mankin DE, Cromwell HC. Striatal activity and reward relativity: Neural signals encoding dynamic outcome valuation. *Eneuro* 3, 2016.

Wheeler RA, Roitman MF, Grigson PS, Carelli RM. Single neurons in the nucleus accumbens track relative reward. *International Journal of Comparative Psychology* 18, 2005.

White JK, Bromberg-Martin ES, Heilbronner SR, Zhang K, Pai J, Haber SN, Monosov IE. A neural network for information seeking. *Nature communications* 10: 1–19, 2019.

Yang C, Mogenson G. An electrophysiological study of the neural projections from the hippocampus to the ventral pallidum and the subpallidal areas by way of the nucleus accumbens. *Neuroscience* 15: 1015–1024, 1985.

Yim C, Mogenson GJ. Response of ventral pallidal neurons to amygdala stimulation and its modulation by dopamine projections to nucleus accumbens. *Journal of neurophysiology* 50: 148–161, 1983.

Yoon T, Geary RB, Ahmed AA, Shadmehr R. Control of movement vigor and decision making during foraging. *Proceedings of the National Academy of Sciences* 115: E10476–E10485, 2018.

Zaborszky L, Gaykema R, Swanson D, Cullinan W. Cortical input to the basal forebrain. *Neuroscience* 79: 1051–1078, 1997.

Curriculum Vitae

David Joshua Ottenheimer

Education

Aug. 2015 - Present	Johns Hopkins University, Baltimore, MD, USA Ph.D. candidate, Department of Neuroscience
Aug. 2010 - May 2014	Yale University, New Haven, CT, USA B.S. Psychology, neuroscience track

Research Experience

Aug. 2015 - Present — Ph.D. Candidate at the Johns Hopkins University, Baltimore

Advisor: Patricia Janak

Studied neural representations of relative value, expectation, and dynamic preference in ventral pallidum using *in vivo* electrophysiology and optogenetics in freely moving rats. Additional techniques included behavioral design, microcontrollers, 3D-printing, computer programming, computational analysis and modeling.

June 2014 - July 2015 — Research Assistant at Yale University, New Haven

Advisor: Ralph DiLeone

Developed a model of anorexia with optogenetic self-stimulation. Learned rodent surgery. Responsible for cloning, cell culture, virus packaging, animal husbandry, and day-to-day lab operations.

Feb. 2013 - May 2014 Undergraduate Researcher at Yale University, New Haven

Advisor: Amy Arnsten

Studed the effects of mGluR2/3 ligands in rat prefrontal cortex on working memory. Learned drug infusions and rodent behavioral methods.

June 2012 - Aug. 2012 Undergraduate Researcher at Georgetown University, Washington, D.C.

Advisor: Kathleen Maguire-Zeiss

Investigated the role of N-Cadherin in microglial activation. Learned cell culture, ELISA assays, western blot, co-immunoprecipitation.

June 2011 - Aug. 2011 Biological Science Aid at the United States Department of Agriculture, Beltsville

Advisor: David Baer and Janet Novotny

Conducted research on antioxidant content of fruit processed by various juicers. Assisted with clinical research on the health effects of whole grains.

Mar. 2010 - May 2010 Research Intern at Harvard University, Cambridge

Advisor: Daniel Janes

Investigated sex determination mechanisms in turtle, emu, and alligator.

Fellowships

2017-2020 National Science Foundation Graduate Research Fellowship Program

Honors

2020	Michael A. Shanoff Award, Young Investigators' Day, Johns Hopkins School of Medicine
2018	Tianqiao and Chrissy Chen Fellowship, 83rd Cold Spring Harbor Laboratory Symposium on Quantitative Biology
2018	2nd place (senior students), Johns Hopkins School of Medicine Graduate Student Association Poster Session
2016	1st place, Lasker Essay Contest
2014	Inducted into Phi Beta Kappa (Yale University)
2014	Magna Cum Laude (Yale University)
2014	Distinction in the Neuroscience Track of Psychology (Yale University)
2012	Invited to Psi Chi: National Honor Society in Psychology
2010-2014	IBM Thomas J. Watson Memorial Scholarship
2010	National Merit Scholar

Courses

2018	Janelia Junior Scientist Workshop on Mechanistic Cognition, Janelia Research Campus

Publications

1. Ottenheimer, D.J., Bari, B.A., Sutlief, E., Fraser, K.M., Kim, T.H., Richard, J.M., Cohen, J.Y. and Janak, P.H., 2019. A history-derived reward prediction error signal in ventral pallidum. bioRxiv, and *in revision*.

2. Vandaele, Y., Mahajan, N.R., Ottenheimer, D.J., Richard, J.M., Mysore, S.P. and Janak, P.H., 2019. Distinct recruitment of dorsomedial and dorsolateral striatum

erodes with extended training. eLife, 8, p.e49536.

3. Ottenheimer, D.J., Wang, K., Haimbaugh, A., Janak, P.H. and Richard, J.M., 2019. Recruitment and disruption of ventral pallidal cue encoding during alcohol seeking. European Journal of Neuroscience, 50(9), pp.3428-3444.

4. Ottenheimer, D., Richard, J.M. and Janak, P.H., 2018. Ventral pallidum encodes relative reward value earlier and more robustly than nucleus accumbens. Nature communications, 9(1), pp.1-14.

5. Conant, K., Daniele, S., Bozzelli, P.L., Abdi, T., Edwards, A., Szklarczyk, A., Ottenheimer, D. and Maguire-Zeiss, K., 2017. Matrix metalloproteinase activity stimulates N-cadherin shedding and the soluble N-cadherin ectodomain promotes classical microglial activation. Journal of neuroinflammation, 14(1), p.56.

6. Jin, L.E., Wang, M., Yang, S.T., Yang, Y., Galvin, V.C., Lightbourne, T.C., Ottenheimer, D., Zhong, Q., Stein, J., Raja, A. and Paspalas, C.D., 2017. mGluR2/3 mechanisms in primate dorsolateral prefrontal cortex: evidence for both presynaptic and postsynaptic actions. Molecular psychiatry, 22(11), pp.1615-1625.

7. Zhu, X., Ottenheimer, D. and DiLeone, R.J., 2016. Activity of D1/2 receptor expressing neurons in the nucleus accumbens regulates running, locomotion, and food intake. Frontiers in behavioral neuroscience, 10, p.66.

Presentations

Talks

Aug. 2019	GRS Catecholamines. Newry, ME.
Mar. 2019	Baltimore Brain Series. Baltimore, MD.
Oct. 2018	Janelia Junior Scientist Workshop. Ashburn, VA.
June 2018	83rd Symposium on Quantitative Biology. Cold Spring Harbor, NY.
Oct. 2017	Johns Hopkins Neuroscience Lab Lunch. Baltimore, MD.
Feb. 2014	Berkeley Commonplace Mellon Forum. New Haven, CT.

Posters

Mar. 2020	COSYNE. Denver, CO.
Oct. 2019	Society for Neuroscience. Chicago, IL.
Sep. 2019	Johns Hopkins Department of Neuroscience Retreat. Ashburn, VA.
Aug. 2019	GRC Catecholamines. Newry, ME.
Mar. 2019	COSYNE. Lisbon, Portugal.
Nov. 2018	Society for Neuroscience. Washington D.C.
Sep. 2018	Johns Hopkins Department of Neuroscience Retreat. Cambridge, MD.
June 2018	83rd Symposium on Quantitative Biology. Cold Spring Harbor, NY.
May 2018	Johns Hopkins Graduate Student Association poster session. Baltimore, MD.
Nov. 2017	Society for Neuroscience. Washington D.C.
Sep. 2017	Johns Hopkins Department of Neuroscience Retreat. Cambridge, MD.

Service and Leadership

Nov. 2018 - May 2020	Member, Johns Hopkins Ph.D. Advisory Committee
Sep. 2017 - Jun. 2019	Contributor, Johns Hopkins Biomedical Odyssey Blog
Jul. 2017 - May 2020	Founder/Member, Neuroscience Dept. Student Diversity Committee
Jun. 2017 - Oct. 2019	Co-leader, Neuroscience Dept. NSF GRFP Workshop
Apr. 2017 - Oct. 2019	Member, Neuroscience Dept. Committee on Diversity & Inclusion
Jan. - May 2017	Teaching assistant, Neuroscience and Cognition II graduate course
Jul. 2016 - Sep. 2018	Co-chair, Neuroscience Department Retreat Committee
Aug. 2016	Workshop Co-leader, NIH-RISE at Morgan State University
June 2016 - May 2020	Neuroscience Department Ph.D. Recruitment Committee
June 2016 - May 2019	Neuroscience Department Student Representative
Sep. 2015 - May 2018	Lead Mentor, STEM Achievement in Baltimore Elementary Schools