Persistent coding of outcome-predictive cue features in the rat nucleus accumbens.

Authors: Jimmie M. Gmaz[1], James E. Carmichael[1], Matthijs A. A. van der Meer[1*]

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover NH 03755

[*]Correspondence should be addressed to MvdM, Department of Psychological and Brain Sciences, Dartmouth College, 3 Maynard St, Hanover, NH 03755. E-mail: mvdm@dartmouth.edu.

Acknowledgments: We thank Nancy Gibson, Martin Ryan and Jean Flanagan for animal care, and Min-Ching Kuo and Alyssa Carey for technical assistance. This work was supported by Dartmouth College (Dartmouth Fellowship to JMG and JEC, and start-up funds to MvdM) and the Natural Sciences and Engineering Research Council (NSERC) of Canada (Discovery Grant award to MvdM, Canada Graduate Scholarship to JMG).

Conflict of Interest: The authors declare no competing financial interests.

Abstract

The nucleus accumbens (NAc) has been shown to be important for learning from feedback, and biasing and invigorating behavior in response to outcome-predictive cues. NAc encodes outcome-related cue features such as the magnitude and identity of reward. However, not much is known about how features of cues themselves are encoded. We designed a decision making task where rats learned multiple sets of outcome-predictive cues, and recorded single-unit activity in the NAc during performance. We found that coding of various cue features occurred alongside coding of expected outcome. Furthermore, this coding persisted both during a delay period, after the rat made a decision and was waiting for an outcome, and after the outcome was revealed. Encoding of cue features in the NAc may enable contextual modulation of ongoing behavior, and provide an eligibility trace of outcome-predictive stimuli for updating stimulus-outcome associations to inform future behavior.

Introduction

Theories of nucleus accumbens (NAc) function generally agree that this brain structure contributes to motivated behavior, with some emphasizing a role in learning from reward prediction errors (RPEs) (Averbeck & Costa 2017; Joel, Niv, & Ruppin 2002; Khamassi & Humphries 2012; Lee, Seo, & Jung 2012; Maia 2009; Schultz 2016; see also the addiction literature on effects of drug rewards; Carelli 2010; Hyman, Malenka, & Nestler 2006; Kalivas & Volkow 2005) and others a role in the modulation of ongoing behavior through stimuli associated with motivationally relevant outcomes (invigorating, directing; Floresco, 2015; Nicola, 2010; Salamone & Correa, 2012). These proposals echo similar ideas on the functions of the neuromodulator dopamine (Berridge, 2012; Maia, 2009; Salamone & Correa, 2012; Schultz, 2016), with which the NAc is tightly linked functionally as well as anatomically (Cheer et al., 2007; du Hoffmann & Nicola, 2014; Ikemoto, 2007; Takahashi, Langdon, Niv, & Schoenbaum, 2016).

Much of our understanding of NAc function comes from studies of how cues that predict motivationally relevant outcomes (e.g. reward) influence behavior and neural activity in the NAc. Task designs that associate such cues with rewarding outcomes provide a convenient access point eliciting conditioned responses such as sign-tracking and goal-tracking (Hearst & Jenkins, 1974; Robinson & Flagel, 2009), pavlovian-instrumental transfer (Estes, 1943; Rescorla & Solomon, 1967) and enhanced response vigor (Nicola, 2010; Niv, Daw, Joel, & Dayan, 2007), which tend to be affected by NAc manipulations (Chang, Wheeler, & Holland 2012; Corbit & Balleine 2011; Flagel et al. 2011; although not always straightforwardly; Chang & Holland 2013; Giertler, Bohn, & Hauber 2004). Similarly, analysis of RPEs typically proceeds by establishing an association between a cue and subsequent reward, with NAc responses transferring from outcome to the cue with learning (Day, Roitman, Wightman, & Carelli, 2007; Roitman, Wheeler, & Carelli, 2005; Schultz, Dayan, & Montague, 1997; Setlow, Schoenbaum, & Gallagher, 2003).

Surprisingly, although substantial work has been done on the coding of outcomes predicted by such cues (Atallah, McCool, Howe, & Graybiel, 2014; Bissonette et al., 2013; Cooch et al., 2015; Day, Wheeler,

Roitman, & Carelli, 2006; Goldstein et al., 2012; Hollerman, Tremblay, & Schultz, 1998; Lansink et al., 2012; McGinty, Lardeux, Taha, Kim, & Nicola, 2013; Nicola, 2004; Roesch, Singh, Brown, Mullins, & Schoenbaum, 2009; Roitman et al., 2005; Saddoris, Stamatakis, & Carelli, 2011; Setlow et al., 2003; Sugam, Saddoris, & Carelli, 2014; West & Carelli, 2016), much less is known about how outcome-predictive cues themselves are encoded in the NAc (but see; Sleezer, Castagno, & Hayden, 2016). This is an important issue for at least two reasons. First, in reinforcement learning, motivationally relevant outcomes are typically temporally delayed relative to the cues that predict them. In order to solve the problem of assigning credit (or blame) across such temporal gaps, some trace of preceding activity needs to be maintained (Lee et al., 2012; Sutton & Barto, 1998). For example, if you become ill after eating food X in restaurant A, depending on if you remember the identity of the restaurant or the food at the time of illness, you may learn to avoid all restaurants, restaurant A only, food X only, or the specific pairing of X-in-A. Therefore, a complete understanding of what is learned following feedback requires understanding what trace is maintained. Since NAc is a primary target of DA signals interpretable as RPEs, and NAc lesions impair RPEs related to timing, its activity trace will help determine what can be learned when RPEs arrive (Hamid et al., 2015; Hart, Rutledge, Glimcher, & Phillips, 2014; Ikemoto, 2007; McDannald, Lucantonio, Burke, Niv, & Schoenbaum, 2011; Takahashi et al., 2016).

Second, for ongoing behavior, the relevance of cues typically depends on context. In experimental settings, context may include the identity of a preceding cue, spatial or configural arrangements (Bouton, 1993; Holland, 1992; Honey, Iordanova, & Good, 2014), and unsignaled rules as occurs in set shifting and other cognitive control tasks (Cohen & Servan-Schreiber, 1992; Floresco, Ghods-Sharifi, Vexelman, & Magyar, 2006; Grant & Berg, 1948; Sleezer et al., 2016). In such situations, the question arises how selective, context-dependent processing of outcome-predictive cues is implemented. For instance, is there a gate prior to NAc such that only currently relevant cues are encoded in NAc, or are all cues represented in NAc but their current values dynamically updated (FitzGerald, Schwartenbeck, & Dolan, 2014; Goto & Grace, 2008; Sleezer et al., 2016). Representation of cue identity would allow for context-dependent mapping of outcomes predicted by specific cues.

Thus, both from a learning and a flexible performance perspective, it is of interest to determine how cue identity is represented in the brain, with NAc of particular interest given its anatomical and functional position at the center of motivational systems. We sought to determine whether cue identity is represented in the NAc, if cue identity is represented alongside other motivationally relevant variables, such as cue value, and if these representations are maintained after a behavioral decision has been made (Figure 1). To address these questions, we recorded the activity of NAc units as rats performed a task in which multiple, distinct sets of cues predicted the same outcome.

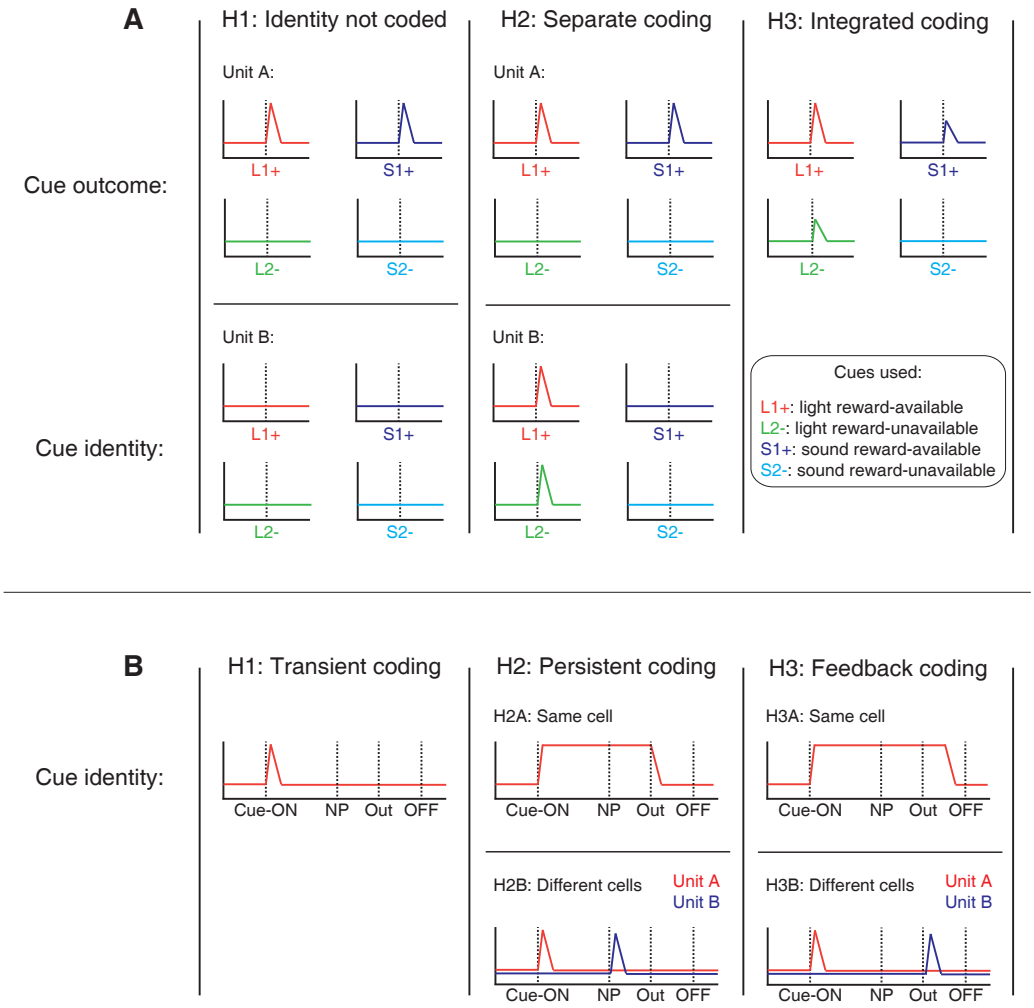

Figure 1: Schematic of potential coding strategies for cue identity (light, sound) and cue outcome (reward-available, reward-unavailable) employed by single units in the NAc across different units (A) and in time (B). **A**: Displayed are schematic PETHs illustrating putative responses to different cues under different hypotheses of how cue identity and outcome are coded. H1 (left panel): Coding of cue identity is absent in the NAc. Top: Unit A encodes a motivationally relevant variable, such as expected outcome, similarly across other cue features, such as cue identity or physical location. Hypothetical plot is firing rate across time. L1+ (red) signifies a reward-available light cue, S1+ (navy blue) a reward-available sound cue, L2- (green) a reward-unavailable light cue, S2- (light blue) a reward-unavailable sound cue. Dashed line indicates onset of cue. Bottom: No units within the NAc discriminate their firing according to cue identity. H2 (middle panel): Coding of cue identity occurs independently of encoding of motivationally relevant variables such as expected outcome or subsequent vigor. Top: Same as H1, with unit A discriminating between reward-available and reward-unavailable cues. Bottom: Unit B discriminates firing across stimulus modalities, depicted here as firing to light cues but not sound cues. H3 (right panel): Coding of cue identity is integrated with coding of other motivationally relevant variables. Hypothetical example demonstrating a unit that responds to outcome-predictive cues, but firing rate is also modulated by cue identity, firing most for the reward-available light cue. **B**: Displayed are schematic PETHs illustrating potential ways in which cue identity signals may persist over time. H1 (left panel): Cue-onset triggers a transient response to a unit that codes for cue identity. Dashed lines indicate time of a behavioral or environmental event. 'Cue-ON' signifies onset of cue, 'NP' signifies when the rat holds a nosepoke at a reward receptacle, 'Out' signifies when the outcome is revealed, 'OFF' signifies when the cue turns off. H2 (middle panel): Coding of cue identity persists during a nosepoke hold period until outcome is revealed. Coding can either be maintained by the same unit as during cue-onset (H2A) or by a sequence of units (H2B). H3 (right panel): Coding of cue identity persists after the outcome is received when the rat gets feedback about his decision, by either the same unit as during cue-onset (H3A) or by a sequence of units (H3B). The same hypotheses apply to other information-containing aspects of the environment when the cue is presented, such as the physical location of the cue.

Results

Behavior

Rats were trained to discriminate between cues signaling the availability and absence of reward on a square track with four identical arms for two distinct set of cues (Figure 2). During each session, rats were presented sequentially with two behavioral blocks containing cues from different sensory modalities, a light and a sound block, with each block containing a cue that signalled the availability of reward (reward-available), and a cue that signalled the absence of reward (reward-unavailable). To maximize reward receipt, rats should approach reward sites on reward-available trials, and skip reward sites on reward-unavailable trials (see Figure 3A for an example learning curve). All four rats learned to discriminate between the reward-available and reward-unavailable cues for both the light and sound blocks as determined by reaching significance ($p <$ .05) on a daily chi-square test comparing approach behavior for reward-available and reward-unavailable cues for each block, for at least three consecutive days (range for time to criterion: 22 - 57 days). Maintenance of behavioral performance during recording sessions was assessed using linear mixed effects models for both proportion of trials where the rat approached the receptacle, and trial length. Analyses revealed that the likelihood of a rat to make an approach was influenced by whether a reward-available or reward-unavailable cue was presented, but was not significantly modulated by whether the rat was presented with a light or sound cue (Percentage approached: light reward-available = 97%; light reward-unavailable = 34%; sound reward-available = 91%; sound reward-unavailable 35%; cue identity p $=$ .115; cue outcome p$<$ .001; Figure 3B). A similar trend was seen with the length of time taken to complete a trial (Trial length: light reward-available = 1.85 s; light reward-unavailable = 1.74 s; sound reward-available = 1.91 s; sound reward-unavailable 1.78 s; cue identity p $=$ .106; cue outcome p $<$ .001; Figure 3C). Thus, during recording, rats successfully discriminated the cues according to whether or not they signaled the availability of reward at the reward receptacle.

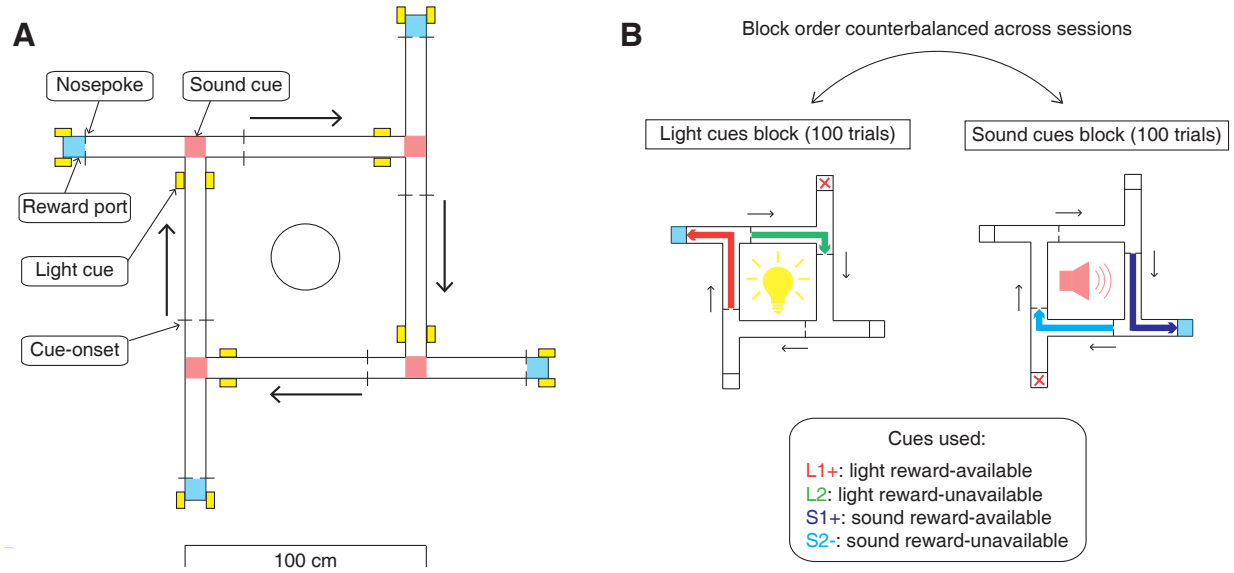

Figure 2: Schematic of behavioral task. **A**: To scale depiction of square track consisting of multiple identical T-choice points. At each choice point, the availability of 12% sucrose reward at the nearest reward receptacle (light blue fill) was signaled by one of four possible cues, presented when the rat initiated a trial by crossing a photobeam on the track (dashed lines). Photobeams at the ends of the arms by the receptacles registered Nosepokes (solid lines). Rectangular boxes with yellow fill indicate location of LEDs used for light cues. Speakers for tone cues were placed underneath the choice points, indicated by magenta fill on track. Arrows outside of track indicate correct running direction. Circle in the center indicates location of pedestal during pre- and post-records. Scale bar is located beneath the track. **B**: Progression of a recording session. A session was started with a 5 minute recording period on a pedestal placed in the center of the apparatus. Rats then performed the light and sound blocks of the cue discrimination task in succession for 100 trials each, followed by another 5 minute recording period on the pedestal. Left in figure depicts a light block, showing an example trajectory for a correct reward-available (approach trial; red) and reward-unavailable (skip trial; green) trial. Right in figure depicts a sound block, with a reward-available (approach trial; navy blue) and reward-unavailable (skip trial; light blue) trial. Ordering of the light and sound blocks was counterbalanced across sessions. Reward-available and reward-unavailable cues were presented pseudo-randomly, such that not more than two of the same type of cue could be presented in a row. Location of the cue on the track was irrelevant for behavior, all cue locations contained an equal amount of reward-available and reward-unavailable trials.

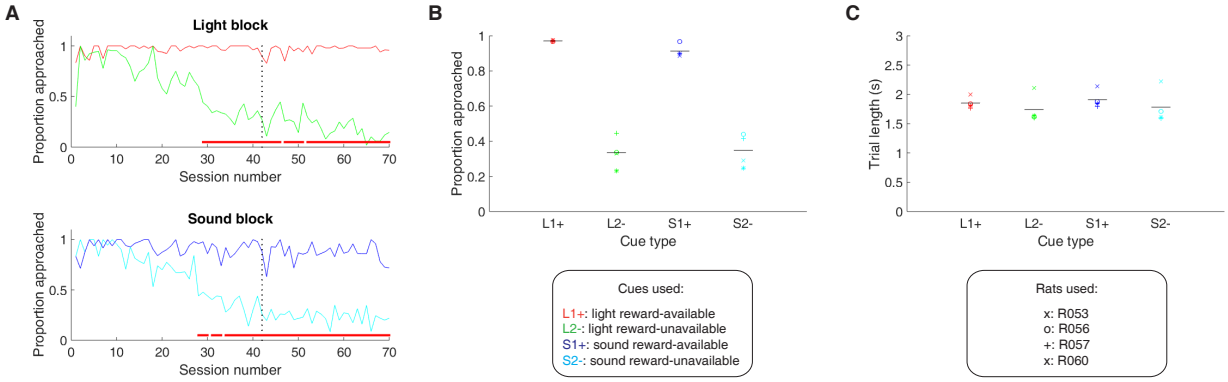

Figure 3: Performance on the behavioral task. **A**. Example learning curves across sessions from a single subject (R060) showing the proportion approached for reward-available (red line for light block, navy blue line for sound block) and reward-unavailable trials (green line for light block, light blue line for sound block) for light (top) and sound (bottom) blocks. Fully correct performance corresponds to an approach proportion of 1 for reward-available trials and 0 for reward-unavailable trials. Rats initially approach on both reward-available and reward-unavailable trials, and learn with experience to skip non-rewarded trials. Red bars indicate days in which a rat statistically discriminated between reward-available and reward-unavailable cues, determined by a chi square test. Dashed line indicates time of electrode implant surgery. **B-C**: Summary of performance during recording sessions for each rat. **B**: Proportion approached for all rats, averaged across all recording sessions. Different columns indicate the different cues (reward-available (red) and reward-unavailable (green) light cues, reward-available (navy blue) and reward-unavailable (light blue) sound cues). Different symbols correspond to individual subjects; horizontal black line shows the mean. All rats learned to discriminate between reward-available and reward-unavailable cues, as indicated by the clear difference of proportion approached between reward-available ($\sim$90% approached) and reward-unavailable cues ($\sim$30% approached), for both blocks (see Results for statistics). **C**: Average trial length for each cue. Note that the time to complete a trial was comparable for the different cues.

NAc units encode behaviorally relevant and irrelevant cue features

Single unit responses discriminate cue features:

We sought to address which parameters of our task were encoded by NAc activity, specifically whether the NAc encodes aspects of motivationally relevant cues not directly tied to reward, such as the identity and location of the cue, and whether this coding is independent or integrated with coding of cue outcome. To do this we recorded a total of 443 units with > 200 spikes in the NAc from 4 rats over 57 sessions (range: 12 - 18 sessions per rat) while they performed a cue discrimination task (Table 1). Units that exhibited a drift in firing rate over the course of either block were excluded from further analysis, leaving 344 units for further analysis. The activity of 133 (39%) of these 344 units were modulated by the cue, with more showing a decrease in firing (n = 103) than an increase (n = 30) around the time of cue-onset (Table 1). Within this group, 24 were classified as FSIs, while 109 were classified as SPNs. Upon visual inspection, we observed several patterns of firing activity, including units that discriminated firing upon cue-onset across various cue conditions, showed sustained differences in firing across cue conditions, had transient responses to the cue, showed a ramping of activity starting at cue-onset, and showed elevated activity immediately preceding cue-onset, for example (Figure 4). To characterize more formally whether these cue-evoked responses were modulated by various aspects of the task, we fit a GLM to each cue-modulated unit. Fitting GLMs revealed that a variety of task parameters accounted for a significant portion of firing rate variance in NAc cue-modulated units (Figure 5, Table 1). Notably, there were units that discriminated between whether the rat was performing in the light or sound block (28% of cue-modulated units, accounting for 6% of variance on average), which arm the rat was currently on (38% of cue-modulated units, accounting for 6% of variance on average), and whether the rat was engaged in the common portion of a reward-available or reward-unavailable trial (26% of cue-modulated units, accounting for 4% of variance on average), suggesting that the NAc encodes features of reward-predictive cues separate from expected outcome (Figure 4A-F). Furthermore, overlap of coding of cue features within units was not different than expected by chance according to chi-square tests, suggesting for integrated coding across various aspects of a cue (Figure 4G,H). Additionally, a sliding window GLM

centered on cue-onset revealed that cue identity and cue location contributed to the activity of a significant proportion of cue-modulated units throughout this epoch, whereas an increase in units encoding cue outcome became apparent after cue-onset (Figure 5D). Together, these findings show that various cue features are represented in the NAc, and that this coding is both integrated and separate from expected outcome (Figure 1; H2,H3).

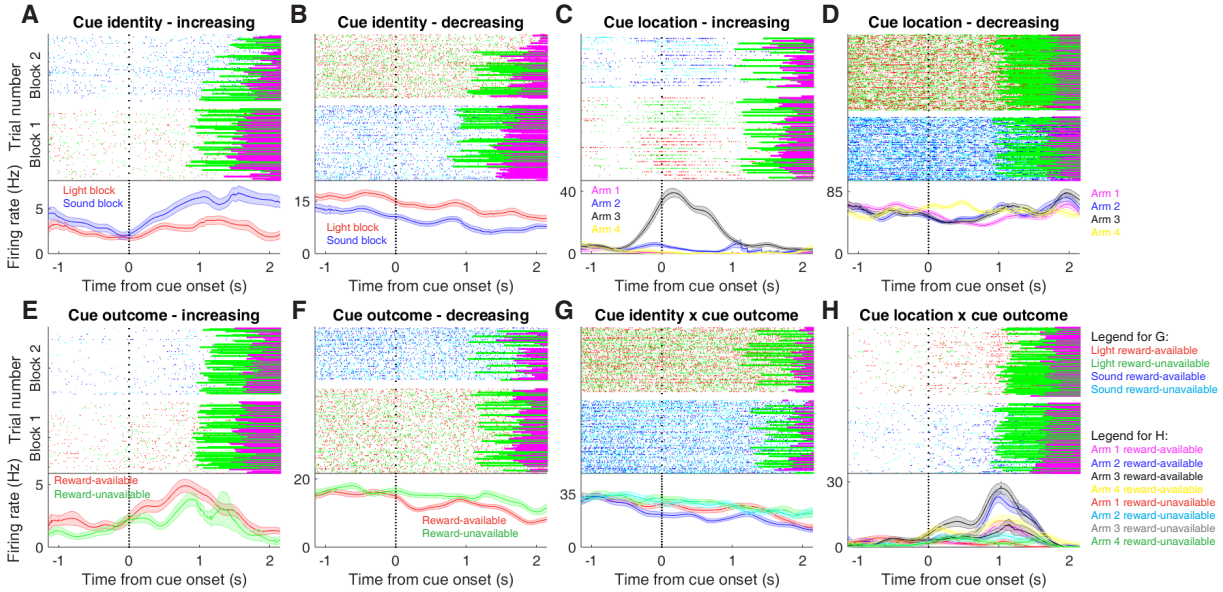

Figure 4: Examples of different cue-modulated NAc units influenced by various task parameters. **A**: Example of a cue-modulated NAc unit that showed an increase in firing following the cue, and encoded cue identity. Top: rasterplot showing the spiking activity across all trials aligned to cue-onset. Spikes across trials are color-coded according to cue type (red: reward-available light; green: reward-unavailable light; navy blue: reward-available sound; light blue: reward-unavailable sound). Green and magenta bars indicate trial termination when a rat initiated the next trial or made a nosepoke, respectively. White space halfway up the rasterplot indicates switching from one block to the next. Dashed line indicates cue onset. Bottom: PETHs showing the average smoothed firing rate for the unit for trials during light (red) and sound (blue) blocks, aligned to cue-onset. Lightly shaded area indicates standard error of the mean. Note this unit showed a larger increase in firing to sound cues. **B**: An example of a unit that was responsive to cue identity as in A, but for a unit that showed a decrease in firing to the cue. Note the sustained higher firing rate during the light block. **C-D**: Cue-modulated units that encoded cue location, each color in the PETHs represents average firing response for a different cue location. **C**: The firing rate of this unit only changed on arm 3 of the task. **D**: Firing decreased for this unit on all arms but arm 4. **E-F**: Cue-modulated units that encoded cue outcome, with the PETHs comparing reward-available (red) and reward-unavailable (green) trials. **E**: This unit showed a slightly higher response during presentation of reward-available cues. **F**: This unit showed a dip in firing when presented with reward-available cues. **G-H**: Examples of cue-modulated units that encoded multiple cue features. **G**: This unit integrated cue identity and outcome. **H**: An example of a unit that integrated cue identity and location.

Task parameter	**Total**	**↑ MSN**	**↓ MSN**	**↑ FSI**	**↓ FSI**
All units	443	155	216	27	45
Rat ID					
R053	145	51	79	4	11
R056	70	12	13	17	28
R057	136	55	75	3	3
R060	92	37	49	3	3
Analyzed units	344	117	175	18	34
Cue modulated units	133	24	85	6	18
GLM aligned to cue-onset					
Cue identity	37	7	21	1	8
Cue location	50	13	27	3	7
Cue outcome	34	10	18	0	6
Approach behavior	31	8	18	1	4
Trial length	25	5	18	0	2
Trial number	32	11	12	1	8
Previous trial	5	0	5	0	0
GLM aligned to nosepoke					
Cue identity	66	14	36	2	14
Cue location	66	14	40	3	9
Cue outcome	42	8	29	0	5
GLM aligned to outcome					
Cue outcome	10	0	6	0	4

Table 1: Overview of recorded NAc units and their relationship to task variables at various time epochs.

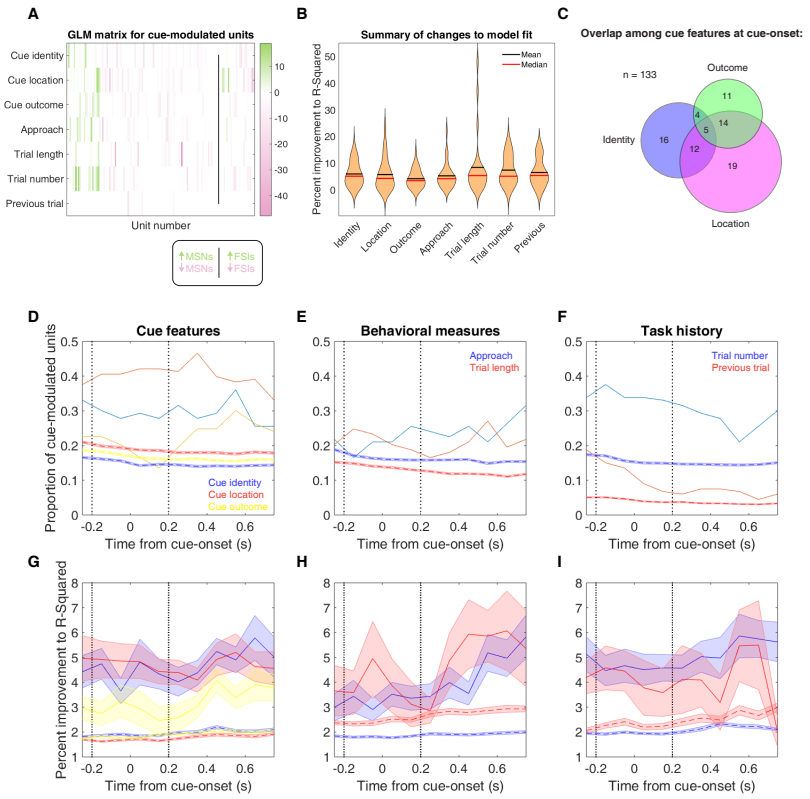

Figure 5: Summary of influence of various task parameters on cue-modulated NAc units after cue-onset. **A**: GLM matrix illustrating the contribution of various task parameters to NAc unit firing rates. A stepwise GLM was fit to each unit that showed evidence of cue modulation by a Wilcoxon signed-rank test. Each row represents a given task parameter, and each column corresponds to a single unit. Colors indicate how much of the firing rate variance an individual predictor contributed to the model, as measured by differences in R-squared between the final model and the model minus the predictor of interest. Ordering from left to right: MSNs that increased firing in response to the cue (green, left of line), MSNs with a decreasing response (red, left of line), FSIs with an increasing response (green, right of line), FSIs with a decreasing response (red, right of line). Darker shades indicate more firing rate variance explained by a given predictor. Black line indicates separation of MSNs and FSIs. Scale bar indicates range of improvements to model fit for units with an increasing (green) and decreasing (red) response to the cue. **B**: Violin plots demonstrating changes in R-squared values with the addition of each of the individual predictors. The mean, median, and distribution of changes in R-squared values is plotted for each of the seven task parameters used in the GLM. **C**: Venn diagram illustrating the number of cue-modulated units encoding cue identity (blue circle), cue location (green circle), cue outcome (pink circle), as well as the overlap among units that encoded multiple cue features. **D-F**: Sliding window GLM illustrating the proportion of cue-modulated units influenced by various predictors around time of cue-onset. **D**: Sliding window GLM (bin size: 500 ms; step size: 100 ms) demonstrating the proportion of cue-modulated units where cue identity (blue solid line), cue location (red solid line), and cue outcome (yellow solid line) significantly contributed to the model at various time epochs relative to cue-onset. Dashed colored lines indicate the average of shuffling the firing rate order that went into the GLM 100 times. Points in between the two vertical dashed lines indicate bins where both pre- and post-cue-onset time periods where used in the GLM. **E**: Same as D, but for approach behavior and trial length. **F**: Same as D, but for trial number and previous trial. **G-I**: Average improvement to model fit. **G**: Average percent improvement to R-squared for units where cue identity, cue location, or cue outcome were significant contributors to the final model for time epochs surrounding cue onset. Shaded area around mean represents the standard error of the mean. **H**: Same as G, but for approach behavior and trial length. **I**: Same G, but for trial number and previous trial.

Population level averages reveal characteristic response profiles:

We observed a variety of single unit response profiles around the time of cue onset (Figure 4). To investigate whether these firing rate patterns were related to what cue features were encoded, we plotted the population level averages for units that were modulated by each feature. To do this, we normalized firing activity for each unit that was modulated by a given cue feature, such as light block, then generated the cue-onset aligned population average firing rate for each of the cue features (Figure 6). Overall, this analysis revealed that cells that showed an increase upon cue presentation had stronger responses for the preferred cue condition (Figure 6A,C,E). Interestingly, units that were classified as decreasing in response to the cue showed a biphasic response at the population level, with a small peak at a time in alignment with entry into the arm, followed by a sustained dip after cue-onset (Figure 6B,D,F). Units that were modulated by cue identity showed a stronger increase in response to the preferred task block, as well as a higher tonic firing rate to the preferred task block, most notably in units that decreased in firing rate to the cue (Figure 6A,B). Units that were modulated by cue location showed a graded response to locations of decreasing preference, with peak firing occurring around cue-onset (Figure 6C,D). Units that were modulated by cue outcome showed a ramping of activity after cue-onset for their preferred cue type. Additionally, units that exhibited a decrease in firing in response to the cue and whose activity was modulated by cue outcome, showed a sustained discriminatory response to reward-available and reward-unavailable cues that extended beyond cue-onset (Figure 6F). Together, these visualizations of the averaged population responses revealed nuanced differences in the way NAc units are modulated by cue conditions across cue features.

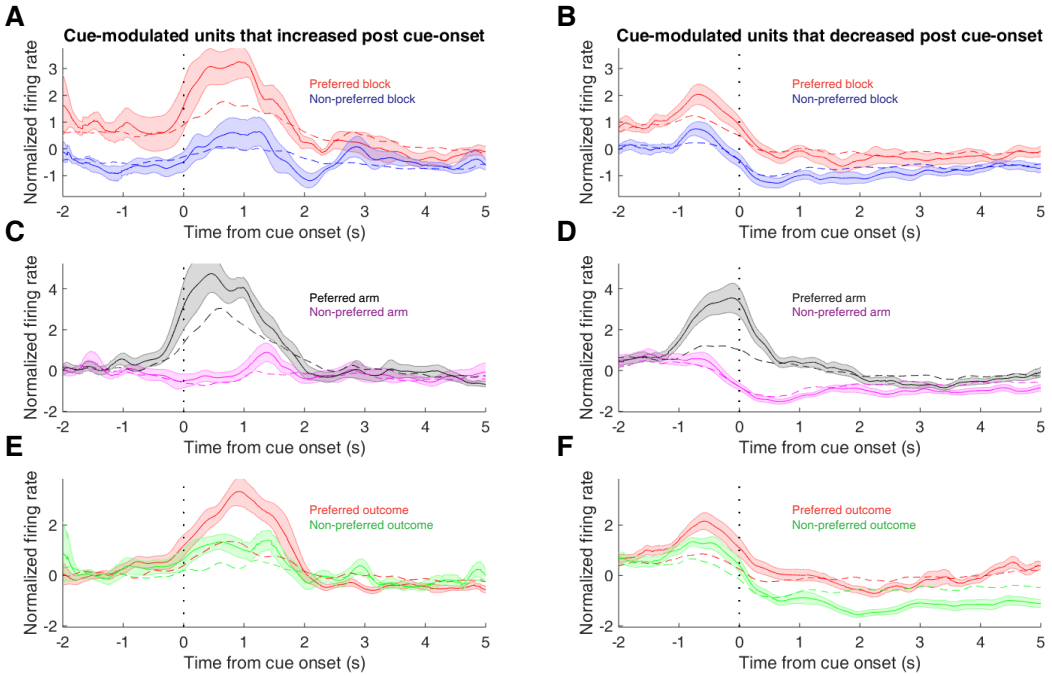

Figure 6: Population-level averages of cue feature sensitive NAc units. **A**: Average smoothed normalized (z-score) activity for cue-modulated units where cue identity was a significant predictor in the GLM, aligned to cue-onset. Activity is plotted for preferred stimulus block (red) and nonpreferred stimulus block (blue). Dashed vertical line indicates onset of cue. Dashed color lines indicate the result of shuffling the identity of the units used for this average 1000 times. Lightly shaded area indicates standard error of the mean. Note larger increase to preferred stimulus block over nonpreferred stimulus block. Black lines indicate the average of 1000 rounds of random sampling of units from the non-drifting population for the preferred and non-preferred blocks. **B**: Same as A but for units that decreased in firing. Note population level activity reveals units classified as decreasing in response to cue show a biphasic response at the population level, with a transient increase around the time the rat starts on the arm, followed by a minimum after cue onset. Also, note the sustained difference in firing between the two blocks. **C-D**: Same as A-B for cue location. Activity is plotted for most preferred arm (black) and least preferred arm (magenta). **E-F**: Same as A-B for cue outcome. Activity is plotted for preferred expected outcome (red), and nonpreferred outcome (green). Note the larger increase to the cue representing the units preferred outcome (E), and the sustained decrease to the nonpreferred outcome (F).

NAc units dynamically segment the task:

Given the varied time courses and response profiles of NAc units to various aspects of the cue, the NAc may be computing a temporally evolving state value signal (Pennartz., 2011). If this is the case, then the recruitment of NAc units should vary alongside changes in the environment. To look at the distribution of responses throughout our task space and see if this distribution is modulated by cue features, we z-scored the firing rate of each unit and plotted the normalized firing rates of all units aligned to cue-onset and sorted them according to the time of peak firing rate (Figure 7). We did this separately for both the light and sound blocks, and found a nearly uniform distribution of firing fields in task space that was not limited to alignment to the cue (Figure 7A). Furthermore, to determine if this population level activity was similar across blocks, we also organized firing during the sound blocks according to the ordering derived from the light blocks. This revealed that while there was some preservation of order, the overall firing was qualitatively different across the two blocks, implying that population activity distinguishes between light and sound blocks. To control for the possibility that any comparison of trials would produce this effect, we did a within block comparison, comparing half of the trials in the light block against the other half. This comparison looked similar to our test comparison of sound block trials ordered by light block trials. Additionally, given that the majority of our units showed an inhibitory response to the cue, we also plotted the firing rates according to the lowest time in firing, and again found some maintenance of order, but largely different ordering across the two blocks, and the within block comparison (Figure 7D). To further test this, we divided each block into two halves and looked at the correlation of the average smoothed firing rates across various combinations of these halves across our cue-aligned centered epoch. A linear mixed effects model revealed that within block correlations (e.g. one half of light trials vs other half of light trials) were higher and more similar than across block correlations (e.g. half of light trials vs half of sound trials) suggesting that activity in the NAc discriminates across various cue conditions (within block correlations = .383 (light), .379 (sound); across block correlations = .343, .338, .337, .348; within block vs. within block comparison = $p = .934$; within block vs. across block comparisons = $p < .001$). This process was repeated for cue location (Figure 7B,E; within block correlations = .369 (arm 1), .350 (arm 2); across block correlations = .290, .286, .285,

.291; within block vs. within block comparison = $p = .071$; within block vs. across block comparisons = $p < .001$) and cue outcome (Figure 7C,F; within block correlations = .429 (reward-available), .261 (reward-unavailable); across block correlations = .258, .253, .255, .249; within block vs. within block comparison = $p < .001$; within block vs. across block comparisons = $p < .001$), showing that NAc segmentation of the task is qualitatively different even during those parts of the task not immediately associated with a specific cue, action, or outcome, although the within condition comparison of reward-unavailable trials was less correlated than reward-available trials, and more similar to the across condition comparisons, potentially due do the greater behavioral variability for the reward-unavailable trials.

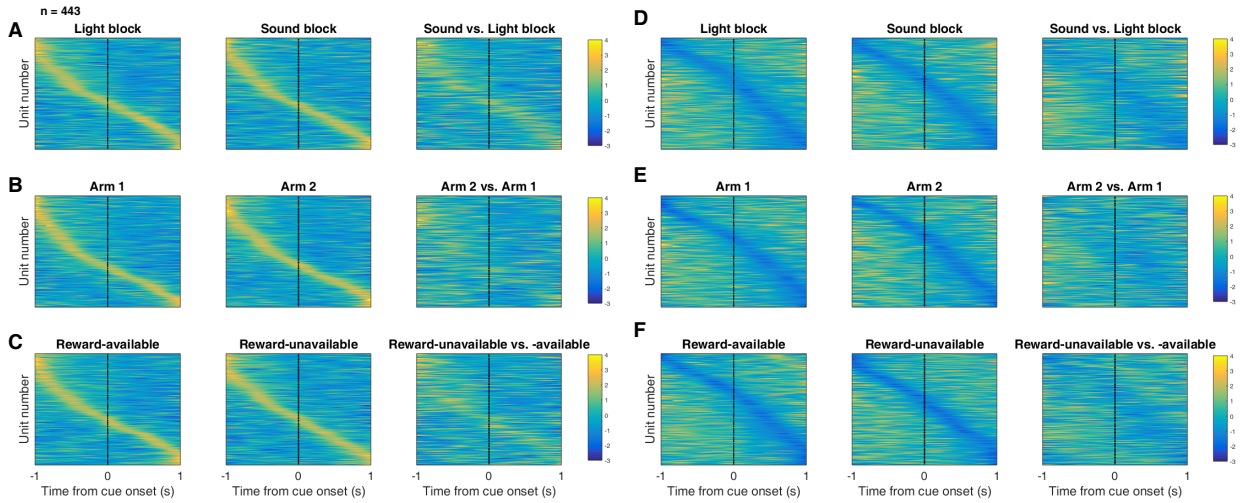

Figure 7: Distribution of NAc firing rates across time surrounding cue onset. Each panel shows normalized (z-score) firing rates for all recorded NAc units (each row corresponds to one unit) as a function of time (time 0 indicates cue onset), averaged across all trials for a specific cue type, indicated by text labels. **A-C**: Heat plots aligned to normalized peak firing rates. **A**, left: Heat plot showing smoothed normalized firing activity of all recorded NAc units ordered according to the time of their peak firing rate during the light block. Each row is a units average activity across time to the light block. Dashed line indicates cue onset. Notice the yellow band across time, indicating all aspects of visualized task space were captured by the peak firing rates of various units. A, middle: Same units ordered according to the time of the peak firing rate during the sound block. Note that for both blocks, units tile time approximately uniformly with a clear diagonal of elevated firing rates. A, right: Unit firing rates taken from the sound block, ordered according to peak firing rate taken from the light block. Note that a weaker but still discernible diagonal persists, indicating partial similarity between firing rates in the two blocks. A similar pattern exists for within-block comparisons suggesting that reordering any two sets of trials produces this partial similarity, however correlations within blocks are more similar than correlations across blocks (see text). **B**: Same layout as in A, except that the panels now compare two different locations on the track instead of two cue modalities. As for the different cue modalities, NAc units clearly discriminate between locations, but also maintain some similarity across locations, as evident from the visible diagonal in the right panel. Two example locations were used for display purposes; other location pairs showed a similar pattern. **C**: Same layout as in A, except that panels now compare reward-available and reward-unavailable trials. **D-F**: Heat plots aligned to normalized minimum firing rates. **D**: Responses during different stimulus blocks as in A, but with units ordered according to the time of their minimum firing rate. **E**: Responses during trials on different arms as in B, but with units ordered by their minimum firing rate. **F**: Responses during cues signalling different outcomes as in C, but with units ordered by their minimum firing rate. Overall, NAc units "tiled" experience on the task, as opposed to being confined to specific task events only. Units from all sessions and animals were pooled for this analysis.

Encoding of cue features persists until outcome:

In order to be useful for credit assignment in reinforcement learning, a trace of the cue must be maintained until the outcome, so that information about the outcome can be associated with the outcome-predictive cue. To test whether representations of cue features persisted post-approach until the outcome was revealed, we fit a GLM to the post-approach firing rates of cue-modulated units aligned to the time of nosepoke into the reward receptacle. This analysis showed that a variety of units still discriminated firing according to various cue features, but not other task parameters, showing that NAc activity discriminates various cue conditions well into a trial (Table 1, Figures 8,9). Additionally, these units were a mix between most of the units that encoded cue features at cue-onset (observed overlap greater than expected by chance according to chi-square tests), and those that did not previously have a cue feature as a predictor (29, 48, and 30 out of 133 cue-modulated units encoded both time points for cue identity, cue location, and cue outcome, respectively). Population level averages for units that increased to cue-onset showed a ramping up of activity that peaked upon nosepoke, whereas units that decreased to cue-onset showed a gradual reduction of firing activity that reached a minimum upon nosepoke (Figure 10). Additionally, a peak is seen for preferred cue outcome in decreasing units at 1 second post cue-onset when reward was received, demonstrating an integration of expected and received reward (Figure 10F). Furthermore, aligning normalized peak firing rates to nosepoke onset, revealed a clustering of responses around outcome receipt for all cue conditions where the rat would have received reward (Figure 11), in addition to the same trend of higher within- vs across-block correlations for cue identity (Figure 11A,C; within block correlations = .560 (light), .541 (sound); across block correlations = .487, .481, .483, .486; within block vs. within block comparison = $p = .112$; within block vs. across block comparisons = $p < .001$) and cue location (Figure 11B,E; within block correlations = .474 (arm 1), .461 (arm 2); across block correlations = .416, .402, .416, .415; within block vs. within block comparison = $p = .810$; within block vs. across block comparisons = $p < .001$), but not cue outcome (Figure 11C,F; within block correlations = .620 (reward-available), .401 (reward-unavailable); across block correlations = .418, .414, .390, .408; within block vs. within block comparison = $p < .001$; within block vs. across block comparisons = $p < .001$). To determine whether coding of cue features persisted after

the outcome was revealed, a GLM was fit to the firing rates of cue-modulated units at the time of outcome receipt, during which the cue was still present. Fitting a GLM revealed 10 units (8%) where cue outcome accounted for an average of 32% of firing rate variance (Table 1, data not shown). An absence of cue identity or cue location coding at this level of analysis was observed, but looking at the data more closely with a sliding window GLM revealed that cue identity, cue location, and cue outcome were encoded throughout time epochs surround cue-onset, nosepoke hold, and outcome receipt, suggesting that the NAc maintains a representation of these cue features once the rat receives behavioral feedback for its decision (Figure 9E-J).

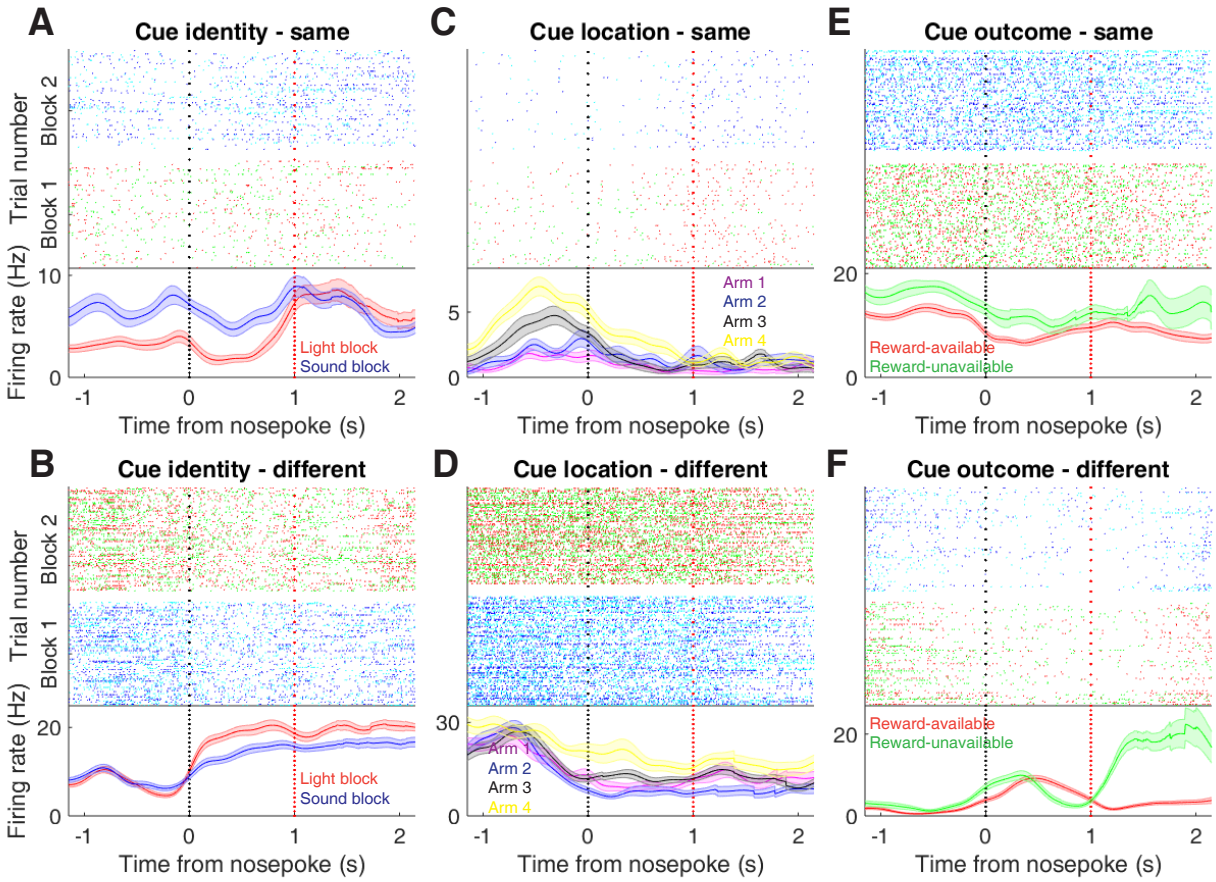

Figure 8: Examples of cue-modulated NAc units influenced by various task parameters at time of nosepoke. **A**: Example of a cue-modulated NAc unit that encoded cue identity at both cue-onset and during nosepoke hold. Top: rasterplot showing the spiking activity across all trials aligned to nosepoke. Spikes across trials are color coded according to cue type (red: reward-available light; green: reward-unavailable light; navy blue: reward-available sound; light blue: reward-unavailable sound). White space halfway up the rasterplot indicates switching from one block to the next. Black dashed line indicates nosepoke. Red dashed line indicates receipt of outcome. Bottom: PETHs showing the average smoothed firing rate for the unit for trials during light (red) and sound (blue) blocks, aligned to nosepoke. Lightly shaded area indicates standard error of the mean. Note this unit showed a sustained increase in firing to sound cues during the trial. **B**: An example of a unit that was responsive to cue identity at time of nosepoke but not cue-onset. **C-D**: Cue-modulated units that encoded cue location, at both cue-onset and nosepoke (C), and only nosepoke (D). Each color in the PETHs represents average firing response for a different cue location. **E-F**: Cue-modulated units that encoded cue outcome, at both cue-onset and nosepoke (E), and only nosepoke (F), with the PETHs comparing reward-available (red) and reward-unavailable (green) trials.

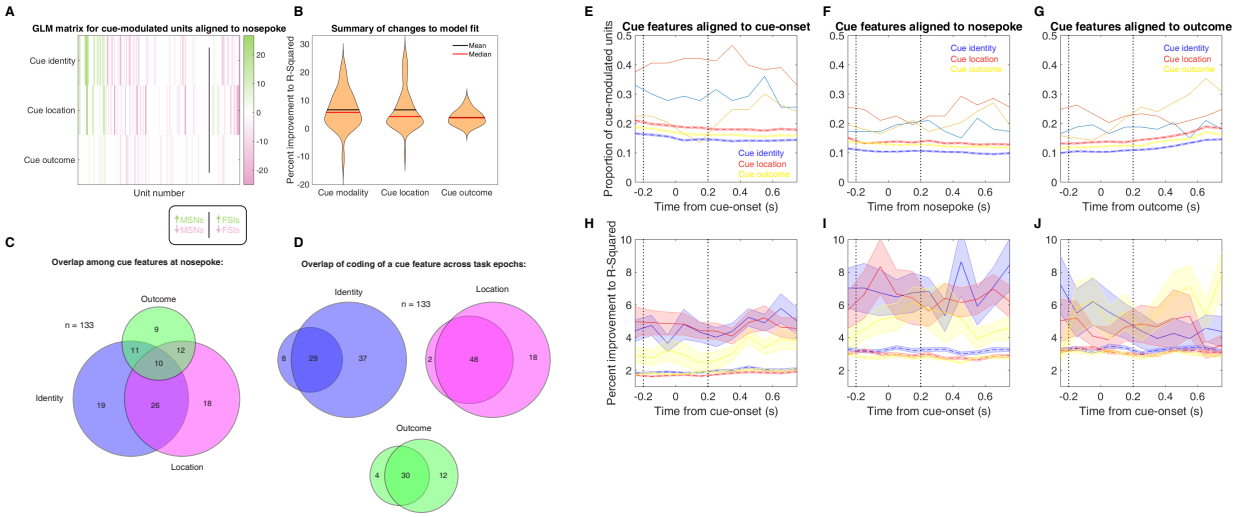

Figure 9: Summary of influence of various task parameters on cue-modulated NAc units during nosepoke. **A**: GLM matrix illustrating the contribution of various task parameters to NAc unit firing rates. A stepwise GLM was fit to each unit that showed evidence of cue modulation by a Wilcoxon signed-rank test. Each row represents a given task parameter, and each column corresponds to a single unit. Colors indicate how much of the firing rate variance an individual predictor contributed to the model, as measured by differences in R-squared between the final model and the model minus the predictor of Interest. Ordering from left to right: MSNs that increased firing in response to the cue (green, left of line), MSNs with a decreasing response (red, left of line), FSIs with an increasing response (green, right of line), FSIs with a decreasing response (red, right of line). Darker shades indicate more firing rate variance explained by a given predictor. Black line indicates separation of MSNs and FSIs. **B**: Violin plots demonstrating changes in R-squared values with the addition of each of the individual predictors. The mean, median, and distribution of changes in R-squared values is plotted for each of the three task parameters that were significant predictors in the GLM. **C**: Venn diagram illustrating the number of cue-modulated units encoding cue identity (blue circle), cue location (green circle), cue outcome (pink circle), as well as the overlap among units that encoded multiple cue features. **D**: Venn diagram illustrating the number of cue-modulated units encoding cue identity, cue location, and cue outcome during cue-onset (left circle), during nosepoke hold (right circle), and during both epochs (overlap). **E-G**: Sliding window GLM illustrating the proportion of cue-modulated units influenced by various predictors around time of cue-onset (E), nosepoke (F), and outcome (G). **E**: Sliding window GLM (bin size: 500 ms; step size: 100 ms) demonstrating the proportion of cue-modulated units where cue identity (blue solid line), cue location (red solid line), and cue outcome (yellow solid line) significantly contributed to the model at various time epochs relative to cue-onset. Dashed colored lines indicate the average of shuffling the firing rate order that went into the GLM 100 times. Points in between the two vertical dashed lines indicate bins where both pre- and post-cue-onset time periods where used in the GLM. **F**: Same as E, but for time epochs relative to nosepoke where the rat waited for the outcome. **G**: Same as E, but for time epochs relative to receipt of outcome after the rat got feedback about his approach. **H-J**: Average improvement to model fit. **H**: Average percent improvement to R-squared for units where cue identity, cue location, or cue outcome were significant contributors to the final model for time epochs relative to cue-onset. Shaded area around mean represents the standard error of the mean. **I**: Same as H, but for time epochs relative to nosepoke. **J**: Same H, but for time epochs relative to receipt of outcome.

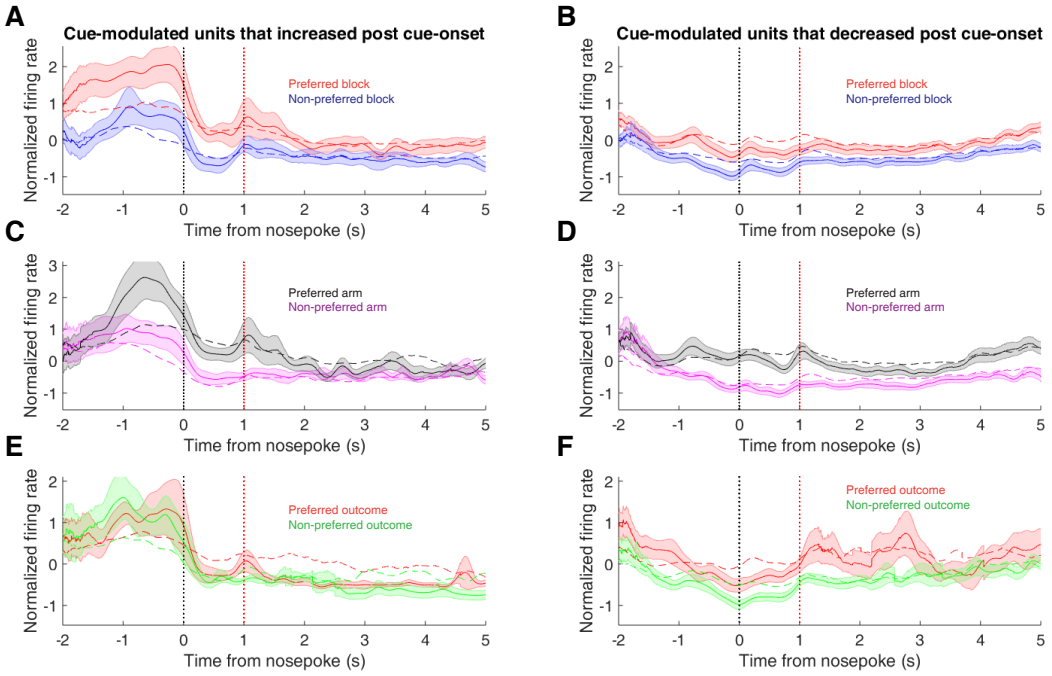

Figure 10: Population-level averages of cue feature sensitive NAc units during a nosepoke. **A**: Average smoothed normalized (z-score) activity for cue-modulated units where cue identity was a significant predictor in the GLM, aligned to nosepoke with reward delivery occurring 1 s after nosepoke. Activity is plotted for preferred stimulus block (red) and nonpreferred stimulus block (blue). Black vertical dashed line indicates nosepoke. Red vertical dashed line indicates reward delivery occurring 1 s after nosepoke for reward-available trials. Dashed color lines indicate the result of shuffling the identity of the units used for this average 1000 times. Lightly shaded area indicates standard error of the mean. Note larger increase leading up to nosepoke to preferred stimulus block over nonpreferred stimulus block. **B**: Same as A but for units that decreased in firing. Note the sustained difference in firing between the two blocks. **C-D**: Same as A-B for cue location. Activity is plotted for most preferred arm (black) and least preferred arm (magneta). **E-F**: Same as A-B for cue outcome. Activity is plotted for preferred expected outcome (red), and nonpreferred outcome (green). Note the peak after outcome receipt for preferred outcome in decreasing units (F).

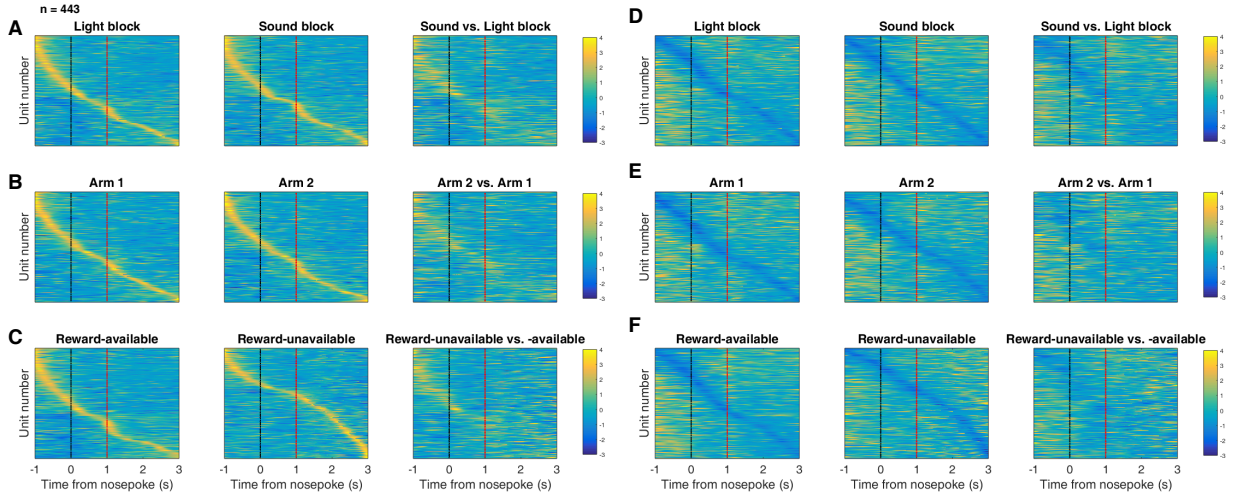

Figure 11: Distribution of NAc firing rates across time surrounding nosepoke for approach trials. Each panel shows normalized (z-score) firing rates for all recorded NAc units (each row corresponds to one unit) as a function of time (time 0 indicates nosepoke), averaged across all approach trials for a specific cue type, indicated by text labels. **A-C**: Heat plots aligned to normalized peak firing rates. **A**, far left: Heat plot showing smoothed normalized firing activity of all recorded NAc units ordered according to the time of their peak firing rate during the light block. Each row is a units average activity across time to the light block. Black dashed line indicates nosepoke. Red dashed line indicates reward delivery occurring 1 s after nosepoke for reward-available trials. Notice the yellow band across time, indicating all aspects of visualized task space were captured by the peak firing rates of various units. A, middle: Same units ordered according to the time of the peak firing rate during the sound block. Note that for both blocks, units tile time approximately uniformly with a clear diagonal of elevated firing rates, and a clustering around outcome receipt. A, right: Unit firing rates taken from the sound block, ordered according to peak firing rate taken from the light block. Note that a weaker but still discernible diagonal persists, indicating partial similarity between firing rates in the two blocks. A similar pattern exists for within-block comparisons suggesting that reordering any two sets of trials produces this partial similarity, however correlations within blocks are more similar than correlations across blocks (see text). **B**: Same layout as in A, except that the panels now compare two different locations on the track instead of two cue modalities. As for the different cue modalities, NAc units clearly discriminate between locations, but also maintain some similarity across locations, as evident from the visible diagonal in the right panel. Two example locations were used for display purposes; other location pairs showed a similar pattern. **C**: Same layout as in A, except that panels now compare correct reward-available and incorrect reward-unavailable trials. The disproportionate tiling around outcome receipt for reward-available, but not reward-unavailable trials suggests encoding of reward receipt by NAc units. **D-F**: Heat plots aligned to normalized minimum firing rates. **D**: Responses during different stimulus blocks as in A, but with units ordered according to the time of their minimum firing rate. **E**: Responses during trials on different arms as in B, but with units ordered by their minimum firing rate. **F**: Responses during cues signalling different outcomes as in C, but with units ordered by their minimum firing rate. Overall, NAc units ”tiled” experience on the task, as opposed to being confined to specific task events only. Units from all sessions and animals were pooled for this analysis.

Discussion

The main result of the present study is that NAc units encode not only the expected outcome of outcome-predictive cues, but also the identity of such cues. Importantly, this identity coding was maintained on approach trials both during a delay period where the rat held a nosepoke until the outcome was received, after immediately after outcome receipt (H2 and H3 in Figure 1B). Units coding for cue identity showed partial overlap with those coding for expected outcome (H3 in Figure 1A). Units that coded different cue features (identity, outcome, location) exhibited different temporal profiles as a whole, although across all recorded units a tiling of task structure was observed such that all points within our analyzed task space was accounted for by the ordered peak firing rates of all units. Furthermore, this tiling differed between various conditions with a cue feature, such as light versus sound blocks. We discuss these observations and their implications below.

Cue identity:

Our finding that NAc units can discriminate between different outcome-predictive stimuli with similar motivational significance (i.e. encodes cue identity) expands upon an extensive rodent literature examining NAc correlates of conditioned stimuli (Ambroggi, Ishikawa, Fields, & Nicola, 2008; Atallah et al., 2014; Bissonette et al., 2013; Cooch et al., 2015; Day et al., 2006; Dejean et al., 2017; Goldstein et al., 2012; Ishikawa, Ambroggi, Nicola, & Fields, 2008; Lansink et al., 2012; McGinty et al., 2013; Nicola, 2004; Roesch et al., 2009; Roitman et al., 2005; Saddoris et al., 2011; Setlow et al., 2003; Sugam et al., 2014; West & Carelli, 2016; Yun, Wakabayashi, Fields, & Nicola, 2004). Perhaps the most comparable work in rodents comes from a study that found distinct coding for an odor when it predicted separate but equally valued rewards (Cooch et al., 2015). The present work is complementary to such *outcome identity* coding as it shows that NAc units encode *cue identity*, both separately and integrated with the reward it predicts (H2 and H3 in Figure 1A). Similarly, Setlow et al. (2003) paired distinct cues with appetitive or aversive outcomes, and found separate populations of units that encoded each cue. Once again, our study was different in asking how

distinct cues encoding the same anticipated outcome are encoded. Such cue identity encoding suggests that even when the biological relevance of these stimuli is similar, NAc dissociates their representations at the level of the single-units. A possible interpretation of this coding of cue features alongside expected outcome is that these representations are used to associate reward with relevant features of the environment, so-called credit assignment in the reinforcement learning literature (Sutton & Barto, 1998). A burgeoning body of human and non-human primate work has started to elucidated neural correlates of credit assignment in the PFC, particularly in the lateral orbitofrontal cortex (Akaishi, Kolling, Brown, & Rushworth, 2016; Asaad, Lauro, Perge, & Eskandar, 2017; Chau et al., 2015; Noonan, Chau, Rushworth, & Fellows, 2017). Given the importance of cortical inputs in NAc associative representations, it is possible that information related to credit assignment is relayed from the cortex to NAc (Cooch et al., 2015; Ishikawa et al., 2008).

A different possible function for cue identity coding is to support contextual modulation of the motivational relevance of specific cues. A context can be understood as a particular mapping between specific cues and their outcomes: for instance, in context 1 cue A but not cue B is rewarded, whereas in context 2 cue B but not cue A is rewarded. Successfully implementing such contextual mappings requires representation of the cue identities. Indeed, Sleezer et al. (2016) recorded NAc responses during the Wisconsin Card Sorting Task, a common set-shifting task used in both the laboratory and clinic, and found units that preferred firing to stimuli when a certain rule, or rule category was currently active. Further support for a modulation of NAc responses by strategy comes from an fMRI study that examined BOLD levels during a set-shifting task (FitzGerald et al., 2014). In this task, participants learned two sets of stimulus-outcome contingencies, a visual set and an auditory set. During testing they were presented with both simultaneously, and the stimulus dimension that was relevant was periodically shifted between the two. Here, they found that bilateral NAc activity reflected value representations for the currently relevant stimulus dimension, and not the irrelevant stimulus. The current finding of separate, but overlapping, populations of units encoding cue identity and expected outcome, suggests that the fMRI finding is generated by the combined activity of several different functional cell types.

Our analyses were designed to eliminate several potential alternative interpretations to cue identity coding. Because the different cues were separated into different blocks, units that discriminated between cue identities could instead be encoding time or other slowly-changing quantities. We excluded this possible confound by excluding units that showed a drift in firing between the first and second half within a block. However, the possiblity remains that instead of or in addition to stimulus identity, these units encode a preferred context, or even a macroscale representation of progress through the session. Indeed, encoding of the current strategy could be an explanation for the sustained difference in population averaged firing across stimulus blocks (Figure 6), as well as a potential explanation for the differentially tiling of task structure across blocks in the current study (Figure 7).

A different potential confound is that between outcome and action value coding. We discriminated between these possiblities by analzying error trials, where the rat appraoched reward (left turn) after presentation of the reward-unavailable cue. Units that were modulated by the expected outcome of the cue maintained their specific firing patterns even during error trials, as expected from outcome value coding but not action value coding. Additionally, NAc signals have been shown to be modulated by response vigor (McGinty et al., 2013); to detangle this from our results we included trial length (i.e. latency to arrival at the reward site) as a predictor in our GLMs, and found units with cue feature correlates independent of trial length.

An overall limitation of the current study is that rats were never presented with both sets of cues simultaneously, and were not required to switch strategies between multiple sets of cues. Thus, it is unknown to what extent the cue identity encoding we observed is behaviorally relevant, although extrapolating data from other work (Sleezer et al., 2016) suggests that cue identity coding would be modulated by relevance. NAc core lesions have been shown to impair shifting between different behavioral strategies (Floresco et al., 2006), and it is possible that selectively silencing the units that prefer responding for a given modality or rule would impair performance when the animal is required to use that information, or artificial enhancement of those units would cause them to use the rule when it is the inappropriate strategy.

Encoding of position:

Our finding that cue-evoked activity was modulated by cue location is in alignment with several previous reports (Lavoie & Mizumori, 1994; Mulder, Shibata, Trullier, & Wiener, 2005; Strait et al., 2016; Wiener et al., 2003). The NAc receives inputs from the hippocampus, and the communication of place-reward information across the two structures suggests that the NAc tracks locations associated with reward (Lansink et al., 2008; Lansink, Goltstein, Lankelma, McNaughton, & Pennartz, 2009; Lansink et al., 2016; Pennartz, 2004; Sjulson, Peyrache, Cumpelik, Cassataro, & Buzsáki, 2017; Tabuchi, Mulder, & Wiener, 2000; van der Meer & Redish, 2011). NAc units can also signal progress through a sequence of cues and/or actions (Atallah et al., 2014; Berke, Breck, & Eichenbaum, 2009; Khamassi, Mulder, Tabuchi, Douchamps, & Wiener, 2008; Lansink et al., 2012; Mulder, Tabuchi, & Wiener, 2004; Shidara, Aigner, & Richmond, 1998). Given that the current task was pseudo-random, it is possible that the rats learned the structure of sequential cue presentation, and the neural activity could reflect this. However, this is unlikely as including a previous trial variable in the analysis did not explain a significant amount of firing rate variance in response to the cue for the vast majority of units. In any case, NAc units on the present task continued to distinguish between different locations, even though location, and progress through a sequence, were explicitly irrelevant in predicting reward. We speculate that this persistent coding of location in NAc may represent a bias in credit assignment, and associated tendency for rodents to associate motivationally relevant events with the locations where they occur.

Implications:

Maladaptive decision making, as occurs in schizophrenia, addiction, Parkinson's, among others, can result from dysfunctional RPE and value signals (Frank, Seeberger, & O'Reilly, 2004; Gradin et al., 2011; Maia & Frank, 2011). This view has been successful in explaining both positive and negative symptoms in schizophrenia, and deficits in learning from feedback in Parkinson's (Frank et al., 2004; Gradin et al., 2011). However, the effects of RPE and value updating are contingent upon encoding of preceding action and cue

features, the eligibility trace (Lee et al., 2012; Sutton & Barto, 1998). Value updates can only be performed on these aspects of preceding experience that are encoded when the update occurs. Therefore, maladaptive learning and decision making can result from not only aberrant RPEs but also from altered cue feature encoding. For instance, on this task the environmental stimulus that signaled the availability of reward was conveyed by two distinct cues that were presented in four locations. While in our current study, the location and identity of the cue did not require any adjustments in the animals behavior, we found coding of these features alongside the expected outcome of the cue that could be the outcome of credit assignment computations computed upstream. Identifying neural coding related to an aspect of credit assignment is important as inappropriate credit assignment could be a contributor to conditioned fear overgeneralization seen in disorders with pathological anxiety such as generalized anxiety disorder, post traumatic stress disorder, and obsessive-compulsive disorder (Kaczkurkin et al., 2017; Kaczkurkin & Lissek, 2013; Lissek et al., 2014), and delusions observed in disorders such as schizophrenia, Alzheimer's and Parkinson's (Corlett, Taylor, Wang, Fletcher, & Krystal, 2010; Kapur, 2003). Thus, our results provide a neural window into the process of credit assignment, such that the extent and specific manner in which this process fails in syndromes such as schizophrenia, obsessive-compulsive disorder, etc. can be experimentally accessed.

Methods

Subjects:

A sample size of 4 adult male Long-Evans rats (Charles River, Saint Constant, QC) from an apriori determined sample of 5 were used as subjects (1 rat was excluded from the data set due to poor cell yield). Rats were individually housed with a 12/12-h light-dark cycle, and tested during the light cycle. Rats were food deprived to 85-90% of their free feeding weight (weight at time of implantation was 440 - 470 g), and water restricted 4-6 hours before testing. All experimental procedures were approved by the the University of Waterloo Animal Care Committee (protocol# 11-06) and carried out in accordance with Canadian Council for

Animal Care (CCAC) guidelines.

Overall timeline:

Each rat was first handled for seven days during which they were exposed to the experiment room, the sucrose solution used as a reinforcer, and the click of the sucrose dispenser valves. Rats were then trained on the behavioral task (described in the next section) until they reached performance criterion. At this point they underwent hyperdrive implantation targeted at the NAc. Rats were allowed to recover for a minimum of five days before being retrained on the task, and recording began once performance returned to pre-surgery levels. Upon completion of recording, animals were gliosed, euthanized and recording sites were histologically confirmed.

Behavioral task and training:

The behavioral apparatus was an elevated, square-shaped track (100 x 100 cm, track width 10 cm) containing four possible reward locations at the end of track "arms" (Figure 2). Rats initiated a *trial* by triggering a photobeam located 24 cm from the start of each arm. Upon trial initiation, one of two possible light cues (L1, L2), or one of two possible sound cues (S1, S2), was presented that signaled the presence (*reward-available trial*, L1+, S1+) or absence (*reward-unavailable trial*, L2-, S2-) of a 12% sucrose water reward (0.1 mL) at the upcoming reward site. A trial was classified as an *approach trial* if the rat turned left at the decision point and made a nosepoke at the reward receptacle (40 cm from the decision point), while trials were classified as a *skip trial* if the rat instead turned right at the decision point and triggered the photobeam to initiate the next trial. A trial is labeled *correct* if the rat approached (i.e. nosepoked) on reward-available trials, and skipped (i.e. did not nosepoke) on reward-unavailable trials. On reward-available trials there was a 1 second delay between a nosepoke and subsequent reward delivery. *Trial length* was determined by measuring the length of time from cue onset until nosepoke (for approach trials), or from cue onset until the start of the following trial (for skip trials). Trials could only be initiated through clockwise progression

through the series of arms, and each entry into the subsequent arm on the track counted as a trial.

Each session consisted of both a *light block* and a *sound block* with 100 trials each. Within a block, one cue signaled reward was available on that trial (L1+ or S1+), while the other signaled reward was not available (L2- or S2-). Light block cues were a flashing white light, and a constant yellow light. Sound block cues were a 2 kHz sine wave and a 8 kHz sine wave whose amplitude was modulated from 0 to maximum by a 2 Hz sine wave. Outcome-cue associations were counterbalanced across rats, e.g. for some rats L1+ was the flashing white light, and for others L1+ was the constant yellow light. The order of cue presentation was pseudorandomized so that the same cue could not be presented more than twice in a row. Block order within each day was also pseudorandomized, such that the rat could not begin a session with the same block for more than two days in a row. Each session consisted of a 5 minute pre-session period on a pedestal (a terracotta planter filled with towels), followed by the first block, then the second block, then a 5 minute post-session period on the pedestal. For approximately the first week of training, rats were restricted to running in the clockwise direction by presenting a physical barrier to running counterclockwise. Cues signaling the availability and unavailability of reward, as described above, were present from the start of training. Rats were trained for 200 trials per day (100 trials per block) until they discriminated between the reward-available and reward-unavailable cues for both light and sound blocks for three consecutive days, according to a chi-square test rejecting the null hypothesis of equal approaches for reward-available and reward-unavailable trials, at which point they underwent electrode implant surgery.

Surgery:

Surgical procedures were as described previously (Malhotra, Cross, Zhang, & van der Meer, 2015). Briefly, animals were administered analgesics and antibiotics, anesthetized with isoflurane, induced with 5% in medical grade oxygen and maintained at 2% throughout the surgery (0.8 L/min). Rats were then chronically implanted with a “hyperdrive” consisting of 16 independently drivable tetrodes, either all 16 targeted for the right NAc (AP +1.4 mm and ML +1.6 mm relative to bregma; Paxinos & Watson 1998), or 12 in the right

NAc and 4 targeted at the mPFC (AP +3.0 mm and ML +0.6 mm, relative to bregma; only data from NAc tetrodes was analyzed). Following surgery, all animals were given at least five days to recover while receiving post-operative care, and tetrodes were lowered to the target (DV -6.0 mm) before being reintroduced to the behavioral task.

Data acquisition and preprocessing:

After recovery, rats were placed back on the task for recording. NAc signals were acquired at 20 kHz with a RHA2132 v0810 preamplifier (Intan) and a KJE-1001/KJD-1000 data acquisition system (Amplipex). Signals were referenced against a tetrode placed in the corpus callosum above the NAc. Candidate spikes for sorting into putative single units were obtained by band-pass filtering the data between 600-9000 Hz, thresholding and aligning the peaks (UltraMegaSort2k, Hill, Mehta, & Kleinfeld, 2011). Spike waveforms were then clustered with KlustaKwik using energy and the first derivative of energy as features, and manually sorted into units (MClust 3.5, A.D. Redish et al., http://redishlab.neuroscience.umn.edu/MClust/MClust.html). Isolated units containing a minimum of 200 spikes within a session were included for subsequent analysis. Units were classified as fast spiking interneurons (FSIs) by an absence of interspike intervals (ISIs) > 2 s, while medium spiny neurons (MSNs) had a combination of ISIs > 2 s and phasic activity with shorter ISIs (Atallah et al., 2014; Barnes, Kubota, Hu, Jin, & Graybiel, 2005).

Data analysis:

Behavior. To determine if rats distinguished behaviorally between the reward-available and reward-unavailable cues (*cue outcome*), we generated linear mixed effects models to investigate the relationships between cue type and our behavioral variables, with *cue outcome* (reward available or not) and *cue identity* (light or sound) as fixed effects, and the addition of an intercept for rat identity as a random effect. For each cue, the average proportion of trials approached and trial length for a session were used as response variables. Contribution of cue outcome to behavior was determined by comparing the full model to a model with cue

outcome removed for each behavioral variable.

Neural data. To investigate the contribution of different cue features cue identity and cue outcome) on the firing rates of NAc single units, we first determined whether firing rates for a unit were modulated by the onset of a cue by collapsing across all cues and comparing the firing rates for the 1 s preceding cue-onset with the 1 s following cue-onset. Single units were considered to be *cue-modulated* if a Wilcoxon signed-rank test comparing pre- and post-cue firing was significant at $p < .01$. Cue-modulated units were then classified as either increasing or decreasing if the post-cue activity was higher or lower than the pre-cue activity, respectively.

To determine the relative contribution of different task parameters to firing rate variance (as in Figures 4-5), a forward selection stepwise general linear model (GLM) was fit to each cue-modulated unit. Cue identity (light block, sound block), cue location (arm 1, arm 2, arm 3, arm 4), cue outcome (reward-available, reward-unavailable), behavior (approach, skip), trial length, trial number, and trial history (reward availability on the previous 2 trials) were used as predictors, and the 1 s post-cue firing rate as the response variable. Units were classified as being modulated by a given task parameter if addition of the parameter significantly improved model fit using deviance as the criterion ($p < .01$). A comparison of the R-squared value between the final model and the final model minus the predictor of interest was used to determine the amount of firing rate variance explained by the addition of that predictor for a given unit. To investigate more finely the temporal dynamics of the influence of task parameters to unit activity, we then fit a sliding window GLM with the same task parameters using 500 ms bins and 100 ms steps, starting 500 ms before cue-onset, up to 500 ms after cue-onset, and measured the proportion of units and average R-squared value for a given time bin where a particular predictor contributed significantly to the final model. To control for the amount of units that would be affected by a predictor by chance, we shuffled the trial order of firing rates for a particular unit within a time bin, and took the average of this value over 100 shuffles.

To better visualize responses to cues and enable subsequent population level analyses (as in Figures 4, 6,

and 7), spike trains were convolved with a Gaussian kernel (σ = 100 ms), and peri-event time histograms (PETHs) were generated by taking the average of the convolved spike trains across all trials for a given task condition. For analysis of population-level responses for cue features (Figure 6), convolved spike trains for all units where cue identity, cue location, or cue outcome explained a significant portion of firing rate variance were z-scored. Within a given cue feature, normalized spike trains were then separated according to the preferred and non-preferred cue condition (e.g. light vs. sound block), and averaged across units to generate population-level averages. To account for separation that would result from any random selection of units, unit identity was shuffled and the shuffled average for preferred and non-preferred cue conditions was generated for 1000 shuffles.

To visualize NAc representations of task space within cue conditions, normalized spike trains for all units were ordered by the location of their maximum or minimum firing rate for a specified cue condition (Figure 7). To compare representations of task space across cue conditions for a cue feature, the ordering of units derived for one condition (e.g. light block) was then applied to the normalized spike trains for the other condition (e.g. sound block). For control comparisons within cue conditions, half of the trials for a condition were compared against the other half. To look at the correlation of firing rates of all units within and across various cue conditions, trials for each cue condition for a unit were shuffled and divided into two averages, and averages within and across cue conditions were correlated. A linear mixed effects model was run for each cue condition to determine if correlations of firing rates within cue conditions were more similar than correlations across cue conditions.

To identify the responsivity of units to different cue features at the time of a nosepoke into a reward receptacle, and subsequent reward delivery, the same cue-responsive units from the cue-onset analyses were analyzed at the time of nosepoke and outcome receipt using identical analysis techniques (Figures 8, 9, 10, and 11).

Given that some of our analyses compare firing rates across time, particularly comparisons across blocks,

we sought to exclude units with unstable firing rates that would generate spurious results reflecting a drift in firing rate over time unrelated to our task. To do this we ran a Mann-Whitney U test comparing the cue-evoked firing rates for the first and second half of trials within a block, and excluded 99 of 443 units from analysis that showed a significant change for either block, leaving 344 units for further analyses. All analyses were completed in MATLAB R2015a, the code is available on our public GitHub repository (http://github.com/vandermeerlab/papers), and the data can be accessed through DataLad.

Histology:

Upon completion of the experiment, recording channels were gliosed by passing 10 μA current for 10 seconds and waiting 5 days before euthanasia, except for rat R057 whose implant detached prematurely. Rats were anesthetized with 5% isoflurane, then asphyxiated with carbon dioxide. Transcardial perfusions were performed, and brains were fixed and removed. Brains were sliced in 50 μm coronal sections and stained with thionin. Slices were visualized under light microscopy, tetrode placement was determined, and electrodes with recording locations in the NAc were analyzed (Figure 12).

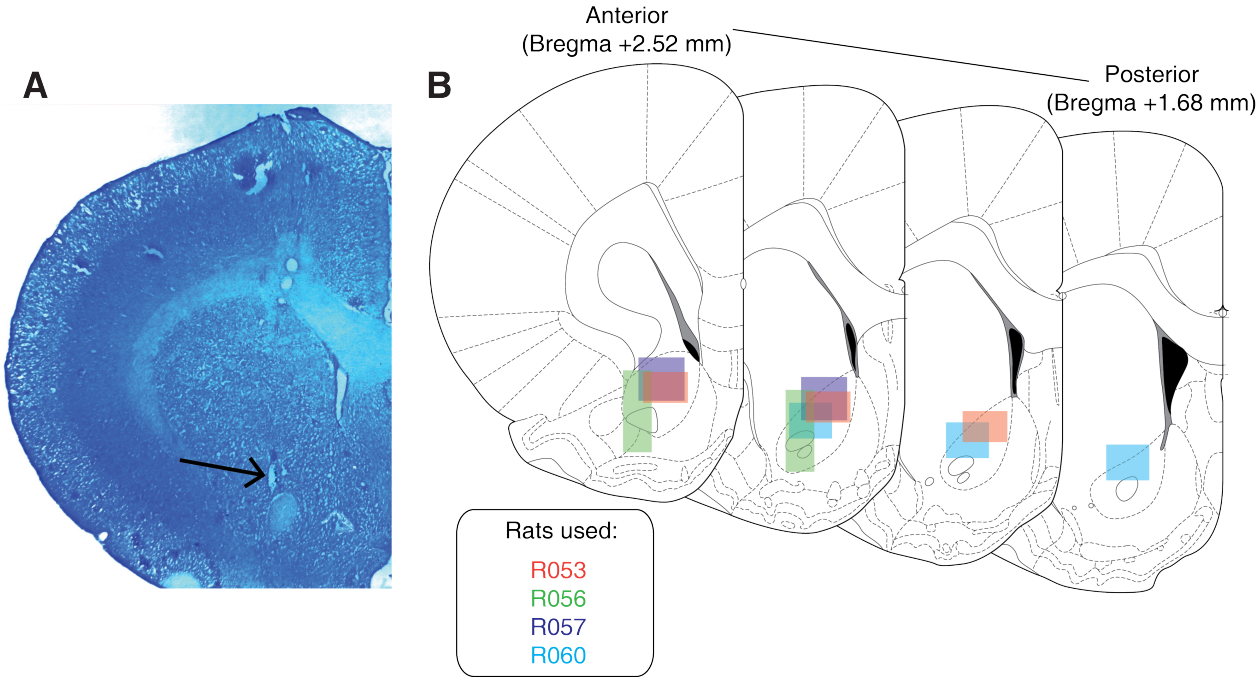

Figure 12: Histological verification of recording sites. Upon completion of experiments, brains were sectioned and tetrode placement was confirmed. **A**: Example section from R060 showing a recording site in the NAc core just dorsal to the anterior commissure (arrow). **B**: Schematic showing recording areas for all subjects.

References

Akaishi, R., Kolling, N., Brown, J. W., & Rushworth, M. (2016). Neural Mechanisms of Credit Assignment in a Multicue Environment. *Journal of Neuroscience*, *36*(4), 1096–1112. doi: 10.1523/JNEUROSCI.3159-15.2016

Ambroggi, F., Ishikawa, A., Fields, H. L., & Nicola, S. M. (2008). Basolateral Amygdala Neurons Facilitate Reward-Seeking Behavior by Exciting Nucleus Accumbens Neurons. *Neuron*, *59*(4), 648–661. doi: 10.1016/j.neuron.2008.07.004

Asaad, W. F., Lauro, P. M., Perge, J. A., & Eskandar, E. N. (2017). Prefrontal Neurons Encode a Solution to the Credit-Assignment Problem. *The Journal of Neuroscience*, *37*(29), 6995–7007. doi: 10.1523/JNEUROSCI.3311-16.2017

Atallah, H. E., McCool, A. D., Howe, M. W., & Graybiel, A. M. (2014). Neurons in the ventral striatum exhibit cell-type-specific representations of outcome during learning. *Neuron*, *82*(5), 1145–1156. doi: 10.1016/j.neuron.2014.04.021

Averbeck, B. B., & Costa, V. D. (2017). Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience*, *20*(4), 505–512. doi: 10.1038/nn.4506

Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z., & Graybiel, A. M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, *437*(7062), 1158–1161. doi: 10.1038/nature04053

Berke, J. D., Breck, J. T., & Eichenbaum, H. (2009). Striatal Versus Hippocampal Representations During Win-Stay Maze Performance. *Journal of Neurophysiology*, *101*(3), 1575–1587. doi: 10.1152/jn.91106.2008

Berridge, K. C. (2012). From prediction error to incentive salience: Mesolimbic computation of reward motivation. *European Journal of Neuroscience*, *35*(7), 1124–1143. doi: 10.1111/j.1460-9568.2012.07990.x

Bissonette, G. B., Burton, A. C., Gentry, R. N., Goldstein, B. L., Hearn, T. N., Barnett, B. R., … Roesch, M. R. (2013). Separate Populations of Neurons in Ventral Striatum Encode Value and Motivation. *PLoS ONE*, *8*(5), e64673. doi: 10.1371/journal.pone.0064673

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*(1), 80–99.doi: 10.1371/journal.pone.0064673

Carelli, R. M. (2010). Drug Addiction: Behavioral Neurophysiology. In *Encyclopedia of neuroscience* (pp. 677–682). Elsevier. doi: 10.1016/B978-008045046-9.01546-1

Chang, S. E., & Holland, P. C. (2013). Effects of nucleus accumbens core and shell lesions on autoshaped lever-pressing. *Behavioural Brain Research*, *256*, 36–42. doi: 10.1016/j.bbr.2013.07.046

Chang, S. E., Wheeler, D. S., & Holland, P. C. (2012). Roles of nucleus accumbens and basolateral amygdala in autoshaped lever pressing. *Neurobiology of Learning and Memory*, *97*(4), 441–451. doi: 10.1016/j.nlm.2012.03.008

Chau, B. K. H., Sallet, J., Papageorgiou, G. K., Noonan, M. A. P., Bell, A. H., Walton, M. E., & Rushworth, M. F. S. (2015). Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit Assignment and Learning in Macaques. *Neuron*, *87*(5), 1106–1118. doi: 10.1016/j.neuron.2015.08.018

Cheer, J. F., Aragona, B. J., Heien, M. L. A. V., Seipel, A. T., Carelli, R. M., & Wightman, R. M. (2007). Coordinated Accumbal Dopamine Release and Neural Activity Drive Goal-Directed Behavior. *Neuron*, *54*(2), 237–244. doi: 10.1016/j.neuron.2007.03.021

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, *99*(1), 45.doi: 10.1016/j.neuron.2007.03.021

Cooch, N. K., Stalnaker, T. A., Wied, H. M., Bali-Chaudhary, S., McDannald, M. A., Liu, T. L., & Schoenbaum, G. (2015). Orbitofrontal lesions eliminate signalling of biological significance in cue-responsive ventral striatal neurons. *Nature Communications*, *6*, 7195. doi: 10.1038/ncomms8195

Corbit, L. H., & Balleine, B. W. (2011). The General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer Are Differentially Mediated by the Nucleus Accumbens Core and Shell. *Journal of Neuroscience*, *31*(33), 11786–11794. doi: 10.1523/JNEUROSCI.2711-11.2011

Corlett, P. R., Taylor, J. R., Wang, X.-J., Fletcher, P. C., & Krystal, J. H. (2010). Toward a neurobiology of delusions. *Progress in Neurobiology*, *92*, 345–369. doi: 10.1016/j.pneurobio.2010.06.007

Day, J. J., Roitman, M. F., Wightman, R. M., & Carelli, R. M. (2007). Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature Neuroscience*, *10*(8), 1020–1028. doi: 10.1038/nn1923

Day, J. J., Wheeler, R. A., Roitman, M. F., & Carelli, R. M. (2006). Nucleus accumbens neurons encode Pavlovian approach behaviors: Evidence from an autoshaping paradigm. *European Journal of Neuroscience*, *23*(5), 1341–1351. doi: 10.1111/j.1460-9568.2006.04654.x

Dejean, C., Sitko, M., Girardeau, P., Bennabi, A., Caillé, S., Cador, M., … Le Moine, C. (2017). Memories of Opiate Withdrawal Emotional States Correlate with Specific Gamma Oscillations in the Nucleus Accumbens. *Neuropsychopharmacology*, *42*(5), 1157–1168. doi: 10.1038/npp.2016.272

du Hoffmann, J., & Nicola, S. M. (2014). Dopamine Invigorates Reward Seeking by Promoting Cue-Evoked Excitation in the Nucleus Accumbens. *Journal of Neuroscience*, *34*(43), 14349–14364. doi: 10.1523/JNEUROSCI.3492-14.2014

Estes, W. K. (1943). Discriminative conditioning. I. A discriminative property of conditioned anticipation. *Journal of Experimental Psychology*, *32*(2), 150–155. doi: 10.1037/h0058316

FitzGerald, T. H. B., Schwartenbeck, P., & Dolan, R. J. (2014). Reward-Related Activity in Ventral Striatum Is Action Contingent and Modulated by Behavioral Relevance. *Journal of Neuroscience*, *34*(4), 1271–1279. doi: 10.1523/JNEUROSCI.4389-13.2014

Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., … Akil, H. (2011). A selective role for dopamine in stimulus-reward learning. *Nature*, *469*(7328), 53–59. doi: 10.1038/nature09588

Floresco, S. B. (2015). The Nucleus Accumbens: An Interface Between Cognition, Emotion, and Action. *Annual Review of Psychology*, *66*(1), 25–52. doi: 10.1146/annurev-psych-010213-115159

Floresco, S. B., Ghods-Sharifi, S., Vexelman, C., & Magyar, O. (2006). Dissociable roles for the nucleus accumbens core and shell

in regulating set shifting. *Journal of Neuroscience*, *26*(9), 2449–2457. doi: 10.1523/JNEUROSCI.4431-05.2006

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science*, *306*(5703), 1940–1943. doi: 10.1126/science.1102941

Giertler, C., Bohn, I., & Hauber, W. (2004). Transient inactivation of the rat nucleus accumbens does not impair guidance of instrumental behaviour by stimuli predicting reward magnitude. *Behavioural Pharmacology*, *15*(1), 55–63.doi: 10.1126/science.1102941

Goldstein, B. L., Barnett, B. R., Vasquez, G., Tobia, S. C., Kashtelyan, V., Burton, A. C., … Roesch, M. R. (2012). Ventral Striatum Encodes Past and Predicted Value Independent of Motor Contingencies. *Journal of Neuroscience*, *32*(6), 2027–2036. doi: 10.1523/JNEUROSCI.5349-11.2012

Goto, Y., & Grace, A. A. (2008). Limbic and cortical information processing in the nucleus accumbens. *Trends in Neurosciences*, *31*(11), 552–558. doi: 10.1016/j.tins.2008.08.002

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., … Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, *134*(6), 1751–1764. doi: 10.1093/brain/awr059

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. doi: 10.1037/h0059831

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., … Berke, J. D. (2015). Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, *19*(1), 117–126. doi: 10.1038/nn.4173

Hart, A. S., Rutledge, R. B., Glimcher, P. W., & Phillips, P. E. M. (2014). Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. *The Journal of Neuroscience*, *34*(3), 698–704. doi: 10.1523/JNEUROSCI.2489-13.2014

Hearst, E., & Jenkins, H. M. (1974). *Sign-tracking: the stimulus-reinforcer relation and directed action*. Psychonomic Society.doi: 10.1523/JNEUROSCI.2489-13.2014

Hill, D. N., Mehta, S. B., & Kleinfeld, D. (2011). Quality Metrics to Accompany Spike Sorting of Extracellular Signals. *Journal of Neuroscience*, *31*(24), 8699–8705. doi: 10.1523/JNEUROSCI.0971-11.2011

Holland, P. C. (1992). Occasion setting in pavlovian conditioning. *Psychology of Learning and Motivation*, *28*(C), 69–125. doi: 10.1016/S0079-7421(08)60488-0

Hollerman, J. R., Tremblay, L., & Schultz, W. (1998). Influence of Reward Expectation on Behavior-Related Neuronal Activity in Primate Striatum. *Journal of Neurophysiology*, *80*(2), 947–963. doi: 10.1152/jn.1998.80.2.947

Honey, R. C., Iordanova, M. D., & Good, M. (2014). Associative structures in animal learning: Dissociating elemental and configural processes. *Neurobiology of Learning and Memory*, *108*, 96–103. doi: 10.1016/j.nlm.2013.06.002

Hyman, S. E., Malenka, R. C., & Nestler, E. J. (2006). NEURAL MECHANISMS OF ADDICTION: The Role of Reward-Related Learning and Memory. *Annual Review of Neuroscience*, *29*(1), 565–598. doi: 10.1146/annurev.neuro.29.051605.113009

Ikemoto, S. (2007). Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Research Reviews*, *56*(1), 27–78. doi: 10.1016/j.brainresrev.2007.05.004

Ishikawa, A., Ambroggi, F., Nicola, S. M., & Fields, H. L. (2008). Dorsomedial Prefrontal Cortex Contribution to Behavioral and Nucleus Accumbens Neuronal Responses to Incentive Cues. *Journal of Neuroscience*, *28*(19), 5088–5098. doi: 10.1523/JNEUROSCI.0253-08.2008

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, *15*(4-6), 535–547. doi: 10.1016/S0893-6080(02)00047-3

Kaczkurkin, A. N., Burton, P. C., Chazin, S. M., Manbeck, A. B., Espensen-Sturges, T., Cooper, S. E., ... Lissek, S. (2017). Neural substrates of overgeneralized conditioned fear in PTSD. *American Journal of Psychiatry*, *174*(2), 125–134. doi: 10.1176/appi.ajp.2016.15121549

Kaczkurkin, A. N., & Lissek, S. (2013). Generalization of Conditioned Fear and Obsessive-Compulsive Traits. *Journal of Psychology & Psychotherapy*, *7*, 3. doi: 10.4172/2161-0487.S7-003

Kalivas, P. W., & Volkow, N. D. (2005). The neural basis of addiction: A pathology of motivation and choice. *American Journal of Psychiatry*, *162*(8), 1403–1413. doi: 10.1176/appi.ajp.162.8.1403

Kapur, S. (2003). Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry*, *160*(1), 13–23. doi: 10.1176/appi.ajp.160.1.13

Khamassi, M., & Humphries, M. D. (2012). Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies. *Frontiers in Behavioral Neuroscience*, *6*, 79. doi: 10.3389/fnbeh.2012.00079

Khamassi, M., Mulder, A. B., Tabuchi, E., Douchamps, V., & Wiener, S. I. (2008). Anticipatory reward signals in ventral striatal neurons of behaving rats. *European Journal of Neuroscience*, *28*(9), 1849–1866. doi: 10.1111/j.1460-9568.2008.06480.x

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., Joosten, R. N. J. M. A., McNaughton, B. L., & Pennartz, C. M. A. (2008). Preferential Reactivation of Motivationally Relevant Information in the Ventral Striatum. *Journal of Neuroscience*, *28*(25), 6372–6382. doi: 10.1523/JNEUROSCI.1054-08.2008

Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L., & Pennartz, C. M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biology*, *7*(8), e1000173. doi: 10.1371/journal.pbio.1000173

Lansink, C. S., Jackson, J. C., Lankelma, J. V., Ito, R., Robbins, T. W., Everitt, B. J., & Pennartz, C. M. A. (2012). Reward Cues in Space: Commonalities and Differences in Neural Coding by Hippocampal and Ventral Striatal Ensembles. *Journal of Neuroscience*, *32*(36), 12444–12459. doi: 10.1523/JNEUROSCI.0593-12.2012

Lansink, C. S., Meijer, G. T., Lankelma, J. V., Vinck, M. A., Jackson, J. C., & Pennartz, C. M. A. (2016). Reward Expectancy Strengthens CA1 Theta and Beta Band Synchronization and Hippocampal-Ventral Striatal Coupling. *Journal of Neuroscience*, *36*(41), 10598–10610. doi: 10.1523/JNEUROSCI.0682-16.2016

Lavoie, A. M., & Mizumori, S. J. (1994). Spatial, movement- and reward-sensitive discharge by medial ventral striatum neurons of rats. *Brain Research*, *638*(1-2), 157–168. doi: 10.1016/0006-8993(94)90645-9

Lee, D., Seo, H., & Jung, M. W. (2012). Neural Basis of Reinforcement Learning and Decision Making. *Annual Review of Neuroscience*, *35*(1), 287–308. doi: 10.1146/annurev-neuro-062111-150512

Lissek, S., Kaczkurkin, A. N., Rabin, S., Geraci, M., Pine, D. S., & Grillon, C. (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. *Biological Psychiatry*, *75*(11), 909–915. doi: 10.1016/j.biopsych.2013.07.025

Maia, T. V. (2009). Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cognitive, Affective and Behavioral Neuroscience*, *9*(4), 343–364. doi: 10.3758/CABN.9.4.343

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154–162. doi: 10.1038/nn.2723

Malhotra, S., Cross, R. W., Zhang, A., & van der Meer, M. A. A. (2015). Ventral striatal gamma oscillations are highly variable from trial to trial, and are dominated by behavioural state, and only weakly influenced by outcome value. *European Journal of Neuroscience*, *42*(10), 2818–2832.doi: 10.1038/nn.2723

McDannald, M. A., Lucantonio, F., Burke, K. A., Niv, Y., & Schoenbaum, G. (2011). Ventral Striatum and Orbitofrontal Cortex Are Both Required for Model-Based, But Not Model-Free, Reinforcement Learning. *Journal of Neuroscience*, *31*(7), 2700–2705. doi: 10.1523/JNEUROSCI.5499-10.2011

McGinty, V. B., Lardeux, S., Taha, S. A., Kim, J. J., & Nicola, S. M. (2013). Invigoration of reward seeking by cue and proximity encoding in the nucleus accumbens. *Neuron*, *78*(5), 910–922. doi: 10.1016/j.neuron.2013.04.010

Mulder, A. B., Shibata, R., Trullier, O., & Wiener, S. I. (2005). Spatially selective reward site responses in tonically active neurons of the nucleus accumbens in behaving rats. *Experimental Brain Research*, *163*(1), 32–43. doi: 10.1007/s00221-004-2135-3

Mulder, A. B., Tabuchi, E., & Wiener, S. I. (2004). Neurons in hippocampal afferent zones of rat striatum parse routes into multi-pace segments during maze navigation. *European Journal of Neuroscience*, *19*(7), 1923–1932. doi: 10.1111/j.1460-9568.2004.03301.x

Nicola, S. M. (2004). Cue-Evoked Firing of Nucleus Accumbens Neurons Encodes Motivational Significance During a Discriminative Stimulus Task. *Journal of Neurophysiology*, *91*(4), 1840–1865. doi: 10.1152/jn.00657.2003

Nicola, S. M. (2010). The Flexible Approach Hypothesis: Unification of Effort and Cue-Responding Hypotheses for the Role of Nucleus Accumbens Dopamine in the Activation of Reward-Seeking Behavior. *Journal of Neuroscience*, *30*(49), 16585–16600. doi: 10.1523/JNEUROSCI.3958-10.2010

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520. doi: 10.1007/s00213-006-0502-4

Noonan, M. P., Chau, B. K., Rushworth, M. F., & Fellows, L. K. (2017). Contrasting Effects of Medial and Lateral Orbitofrontal Cortex Lesions on Credit Assignment and Decision-Making in Humans. *The Journal of Neuroscience*, *37*(29), 7023–7035. doi: 10.1523/JNEUROSCI.0692-17.2017

Paxinos, G., & Watson, C. (1998). *The Rat Brain in Stereotaxic Coordinates* (4th ed.). San Diego: Academic Press.doi:

10.1523/JNEUROSCI.0692-17.2017

Pennartz, C. M. A. (2004). The Ventral Striatum in Off-Line Processing: Ensemble Reactivation during Sleep and Modulation by Hippocampal Ripples. *Journal of Neuroscience*, *24*(29), 6446–6456. doi: 10.1523/JNEUROSCI.0575-04.2004

Rescorla, R. A., & Solomon, R. L. (1967). Two-Process Learning Theory: Relationships Between Pavlovian Conditioning and Instrumental Learning. *Psychological Review*, *74*(3), 151–182. doi: 10.1037/h0024475

Robinson, T. E., & Flagel, S. B. (2009). Dissociating the Predictive and Incentive Motivational Properties of Reward-Related Cues Through the Study of Individual Differences. *Biological Psychiatry*, *65*(10), 869–873. doi: 10.1016/j.biopsych.2008.09.006

Roesch, M. R., Singh, T., Brown, P. L., Mullins, S. E., & Schoenbaum, G. (2009). Ventral Striatal Neurons Encode the Value of the Chosen Action in Rats Deciding between Differently Delayed or Sized Rewards. *Journal of Neuroscience*, *29*(42), 13365–13376. doi: 10.1523/JNEUROSCI.2572-09.2009

Roitman, M. F., Wheeler, R. A., & Carelli, R. M. (2005). Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron*, *45*(4), 587–597. doi: 10.1016/j.neuron.2004.12.055

Saddoris, M. P., Stamatakis, A., & Carelli, R. M. (2011). Neural correlates of Pavlovian-to-instrumental transfer in the nucleus accumbens shell are selectively potentiated following cocaine self-administration. *European Journal of Neuroscience*, *33*(12), 2274–2287. doi: 10.1111/j.1460-9568.2011.07683.x

Salamone, J. D., & Correa, M. (2012). The Mysterious Motivational Functions of Mesolimbic Dopamine. *Neuron*, *76*(3), 470–485. doi: 10.1016/j.neuron.2012.10.021

Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, *18*(1), 23–32. doi: 10.1038/nrn.2015.26

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. doi: 10.1126/science.275.5306.1593

Setlow, B., Schoenbaum, G., & Gallagher, M. (2003). Neural encoding in ventral striatum during olfactory discrimination learning. *Neuron*, *38*(4), 625–636. doi: 10.1016/S0896-6273(03)00264-2

Shidara, M., Aigner, T. G., & Richmond, B. J. (1998). Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *Journal of Neuroscience*, *18*(7), 2613–25.doi: 10.1016/S0896-6273(03)00264-2

Sjulson, L., Peyrache, A., Cumpelik, A., Cassataro, D., & Buzsáki, G. (2017). Cocaine place conditioning strengthens location-specific hippocampal inputs to the nucleus accumbens. *bioRxiv*, 1–10. doi: 10.1101/105890

Sleezer, B. J., Castagno, M. D., & Hayden, B. Y. (2016). Rule Encoding in Orbitofrontal Cortex and Striatum Guides Selection. *Journal of Neuroscience*, *36*(44), 11223–11237. doi: 10.1523/JNEUROSCI.1766-16.2016

Strait, C. E., Sleezer, B. J., Blanchard, T. C., Azab, H., Castagno, M. D., & Hayden, B. Y. (2016). Neuronal selectivity for spatial positions of offers and choices in five reward regions. *Journal of Neurophysiology*, *115*(3), 1098–1111. doi: 10.1152/jn.00325.2015

Sugam, J. A., Saddoris, M. P., & Carelli, R. M. (2014). Nucleus accumbens neurons track behavioral preferences and reward outcomes during risky decision making. *Biological Psychiatry*, *75*(10), 807–816. doi: 10.1016/j.biopsych.2013.09.010

Sutton, R., & Barto, A. (1998). *Reinforcement Learning: An Introduction* (Vol. 9) (No. 5). MIT Press, Cambridge, MA. doi: 10.1109/TNN.1998.712192

Tabuchi, E. T., Mulder, A. B., & Wiener, S. I. (2000). Position and behavioral modulation of synchronization of hippocampal and accumbens neuronal discharges in freely moving rats. *Hippocampus*, *10*(6), 717–728. doi: 10.1002/1098-1063(2000)10:6¡717::AID-HIPO1009¿3.0.CO;2-3

Takahashi, Y. K., Langdon, A. J., Niv, Y., & Schoenbaum, G. (2016). Temporal Specificity of Reward Prediction Errors Signaled by Putative Dopamine Neurons in Rat VTA Depends on Ventral Striatum. *Neuron*, *91*(1), 182–193. doi: 10.1016/j.neuron.2016.05.015

van der Meer, M. A. A., & Redish, A. D. (2011). Theta Phase Precession in Rat Ventral Striatum Links Place and Reward Information. *Journal of Neuroscience*, *31*(8), 2843–2854. doi: 10.1523/JNEUROSCI.4869-10.2011

West, E. A., & Carelli, R. M. (2016). Nucleus Accumbens Core and Shell Differentially Encode Reward-Associated Cues after Reinforcer Devaluation. *Journal of Neuroscience*, *36*(4), 1128–1139. doi: 10.1523/JNEUROSCI.2976-15.2016

Wiener, S. I., Shibata, R., Tabuchi, E., Trullier, O., Albertin, S. V., & Mulder, A. B. (2003). Spatial and behavioral correlates in nucleus accumbens neurons in zones receiving hippocampal or prefrontal cortical inputs. *International Congress Series*, *1250*(C), 275–292. doi: 10.1016/S0531-5131(03)00978-6

Yun, I. A., Wakabayashi, K. T., Fields, H. L., & Nicola, S. M. (2004). The Ventral Tegmental Area Is Required for the Behavioral and Nucleus Accumbens Neuronal Firing Responses to Incentive Cues. *Journal of Neuroscience*, *24*(12), 2923–2933. doi: 10.1523/JNEUROSCI.5282-03.2004