

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

**Authors:** Jimmie M. Gmaz<sup>1</sup>, James E. Carmichael<sup>1</sup>, Matthijs A. A. van der Meer<sup>1\*</sup>

<sup>1</sup>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.

**Number of Figures:** 12

**Number of Tables:** 1

**Total Word Count:** ?

**Abstract Word Count:** ?

**Introduction Word Count:** ?

**Discussion Word Count:** ?

**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.



## **1 Abstract**

2 In reinforcement learning, adaptive behavior requires assigning credit to the appropriate reward-predictive  
3 stimuli in the environment. The nucleus accumbens has been shown to be important for learning from  
4 feedback, and biasing and invigorating behavior in response to reward-predictive cues. Much work has  
5 emphasized nucleus accumbens coding of reward features of reward-predictive stimuli such as magnitude  
6 and identity of reward. However, not much is known as to how features of cues themselves are encoded.  
7 To test this we designed a decision making task where rats had to learn multiple sets of reward-predictive  
8 cues, and recorded single-unit activity in the nucleus accumbens during performance. We found that coding  
9 of various cue features occurred alongside coding of expected outcome. Furthermore, we found that this  
10 coding persisted during a delay period, after the rat made a decision and was waiting for an outcome, but not  
11 after the outcome was revealed. This finding of persistent coding of cue features in the nucleus accumbens  
12 could be one way in which the brain maintains an eligibility trace of outcome-predictive stimuli to be used  
13 for updating stimulus-outcome associations to inform future behavior.

## **14 Significance Statement (120 words)**

15 to do

## **16    Introduction**

17   A fundamental problem faced by all reinforcement learning agents is which features and/or actions to credit  
18   and blame for particular outcomes. Adaptive behavior requires associating motivationally relevant outcomes  
19   with the cues that predict them through learning from feedback. Much work has focused on value signals  
20   such as reward prediction errors (RPEs), state values and action values. However, successful learning re-  
21   quires not only a RPE but also a trace of the preceding actions and/or cues. For example, if you become ill  
22   after eating food X in restaurant A, depending on if you remember the identity of the restaurant or the food  
23   at the time of illness, you may learn to avoid all restaurants; A only; X only; all foods; specific pairing of  
24   X-in-A.. Thus a complete understanding of what is learned following feedback requires understanding what  
25   trace is maintained.

26   Theories of nucleus accumbens (NAc) function generally agree that this brain structure contributes to moti-  
27   vated behavior, with some emphasizing a role in learning from reward prediction errors (Averbeck and Costa,  
28   2017; Joel et al., 2002; Khamassi and Humphries, 2012; Lee et al., 2012; Maia, 2009; Schultz, 2016)(see also  
29   the addiction literature on effects of drug rewards; Carelli, 2009; Hyman et al., 2006; Kalivas and Volkow,  
30   2005) and others a role in the modulation of ongoing behavior through stimuli associated with motivation-  
31   ally relevant outcomes (invigorating, directing; Floresco, 2015; Nicola, 2010; Salamone and Correa, 2012).  
32   These proposals echo similar ideas on the functions of the neuromodulator dopamine (Berridge, 2012; Maia,  
33   2009; Salamone and Correa, 2012; Schultz, 2016), with which the NAc is tightly linked functionally as well  
34   as anatomically (Cheer et al., 2007; du Hoffmann and Nicola, 2014; Ikemoto, 2007; Takahashi et al., 2016).

35   Much of our understanding of NAc function comes from studies of how cues that predict motivationally  
36   relevant outcomes (e.g. reward) influence behavior and neural activity in the NAc. Task designs that asso-  
37   ciate such cues with rewarding outcomes provide a convenient access point eliciting conditioned responses  
38   such as sign-tracking and goal-tracking (Hearst and Jenkins, 1974; Robinson and Flagel, 2009), pavlovian-  
39   instrumental transfer (Estes, 1943; Rescorla and Solomon, 1967) and enhanced response vigor (Nicola, 2010;

40 Niv et al., 2007), which tend to be affected by NAc manipulations (Chang et al., 2012; Corbit and Balleine,  
41 2011; Flagel et al., 2011) (although not always straightforwardly (is Hauber appropriate?); Chang and Hol-  
42 land, 2013; Hauber et al., 2000). Similarly, analysis of reward prediction errors typically proceeds by estab-  
43 lishing an association between a cue and subsequent reward, with NAc responses transferring from outcome  
44 to the cue with learning (Day et al., 2007; Roitman et al., 2005; Setlow et al., 2003; Schultz et al., 1997)(Day  
45 2007 is the only paper referenced here that looks at transferring of responses from reward to CS+, other  
46 papers are in VTA, or the development of CS+ responses). **WHAT ABOUT HUMAN WORK**

47 Surprisingly, although substantial work has been done on the coding of outcomes predicted by such cues  
48 (e.g. reward value; Hollerman/Schultz, Roesch, Day; reward identity; Cooch), much less is known about  
49 how reward-predictive cues themselves are encoded in the NAc (Hayden from primate realm). This is an im-  
50 portant issue for at least two reasons. First, in reinforcement learning, motivationally relevant outcomes are  
51 typically temporally delayed relative to the cues that predict them. In order to solve the problem of assigning  
52 credit (or blame) across such temporal gaps, some trace of preceding activity needs to be maintained (Lee  
53 et al., 2012) (Maia?). Since NAc is a primary target of DA signals interpretable as RPEs, and NAc lesions  
54 impair RPEs related to timing, its activity trace will help determine what can be learned when RPEs arrive  
55 (Takahashi).

56 Second, for ongoing behavior, the relevance of cues typically depends on context. In experimental settings,  
57 context may include the identity of a preceding cue (occasion setter, Holland, Kesner), spatial or configurational  
58 arrangements (Good/Honey, Eichenbaum), and unsignaled rules as occurs in set shifting and other cognitive  
59 control tasks (CITE). In such situations, the question arises how selective, context-dependent processing of  
60 reward-predictive cues is implemented. For instance, is there a gate prior to NAc, or are all cues represented  
61 in NAc but their current values dynamically updated (FitzGerald; **WHAT ARE SOME THEORETICAL**  
62 **POSITIONS ON THIS?**)

63 Thus, both from a learning and a flexible performance perspective, it is of interest to determine how cue

64 identity is represented in the brain, with NAc of particular interest given its anatomical and functional posi-  
65 tion at the center of motivational systems. We sought to determine whether cue identity is represented in the  
66 NAc, and if cue identity is represented alongside other motivationally relevant variables, such as cue value  
67 (Figure 1). To address this question, we recorded the activity of NAc neurons as rats performed a task in  
68 which multiple, distinct sets of cues predicted the same outcome.

69 [Figure 1 about here.]

## 70 **Methods**

### 71 **Subjects:**

72 Adult male Long-Evans rats ( $n = 4$ , Charles River, Saint Constant, QC) were used as subjects. Rats were in-  
73 dividually housed with a 12/12-h light-dark cycle, and tested during the light cycle. Rats were food deprived  
74 to 85-90% of their free feeding weight (weight at time of implantation was 440 - 470 g), and water restricted  
75 4-6 hours before testing. All experimental procedures were approved by the the University of Waterloo An-  
76 imal Care Committee (protocol# 11-06) and carried out in accordance with Canadian Council for Animal  
77 Care (CCAC) guidelines.

### 78 **Overall timeline:**

79 Each rat was first handled for seven days during which they were exposed to the experiment room, the su-  
80 crose solution used as a reinforcer, and the click of the sucrose dispenser valves. Rats were then shaped to  
81 run on the task for sucrose reward while being restricted to running in the clockwise direction by presenting  
82 a physical barrier to running counterclockwise. Cues signaling the availability and inavailability of reward

83 were then introduced (described in detail below, and rats were trained for 200 trials per day until they dis-  
84 criminated between the cues. At this point they underwent hyperdrive implantation targeted at the NAc. Rats  
85 were allowed to recover for a minimum of five days before being retrained on the task, and recording began  
86 once performance returned to pre-surgery levels. Upon completion of recording, animals were euthanized  
87 and recording sites were histologically confirmed.

88 **Behavioral task and training:**

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

104 Each session consisted of both a *light block* and a *sound block* with 100 trials each. Within a block, one cue  
105 signaled reward was available on that trial (L1+ or S1+), while the other signaled reward was not available  
106 (L2- or S2-). Light block cues were a flashing white light, and a constant yellow light. Sound block cues

107 were a 2 kHz sine wave and a 8 kHz sine wave whose amplitude was modulated from 0 to maximum by a  
108 2 Hz sine wave. Reward-cue associations were counterbalanced across rats, e.g. for some rats L1+ was the  
109 flashing white light, and for others L1+ was the constant yellow light. The order of cue presentation was  
110 pseudorandomized so that the same cue could not be presented more than twice in a row. Block order within  
111 each day was also pseudorandomized, such that the rat could not begin a session with the same block for more  
112 than two days in a row. Each session consisted of a 5 minute pre-session period on a pedestal (a terracotta  
113 planter filled with towels), followed by the first block, then the second block, then a 5 minute post-session  
114 period on the pedestal. Rats were trained daily until they could distinguish between the reward-available and  
115 reward-unavailable cues for both light and sound blocks for three consecutive days, according to a chi-square  
116 test rejecting the null hypothesis of equal approaches for reward-available and reward-unavailable trials, at  
117 which point they underwent electrode implant surgery.

118 [Figure 2 about here.]

119 **Surgery:**

120 Surgical procedures were as described previously (Malhotra et al., 2015). Briefly, animals were administered  
121 analgesics and antibiotics, anesthetized with isoflurane, induced with 5% in medical grade oxygen and main-  
122 tained at 2% throughout the surgery ( 0.8 L/min). Rats were then chronically implanted with a “hyperdrive”  
123 consisting of 16 independently drivable tetrodes, either all 16 targeted for the right NAc (AP +1.4 mm and  
124 ML +1.6 mm relative to bregma; Paxinos and Watson, 2005), or 12 in the right NAc and 4 targeted at the  
125 mPFC (AP +3.0 mm and ML +0.6 mm, relative to bregma; only data from NAc tetrodes was analyzed).  
126 Following surgery, all animals were given at least five days to recover while receiving post-operative care,  
127 and tetrodes were lowered to the target (DV -6.0 mm) before being reintroduced to the behavioral task.

128 **Data acquisition and preprocessing:**

129 After recovery, rats were placed back on the task for recording. NAc signals were acquired at 20 kHz with a  
130 RHA2132 v0810 preamplifier (Intan) and a KJE-1001/KJD-1000 data acquisition system (Amplipex). Sig-  
131 nals were referenced against a tetrode placed in the corpus callosum above the NAc. Candidate spikes for  
132 sorting into putative single units were obtained by band-pass filtering the data between 600-9000 Hz, thresh-  
133 olding and aligning the peaks (UltraMegaSort2k, Hull et al., 2011). Spike waveforms were then clustered  
134 with KlustaKwik using energy and the first derivative of energy as features, and manually sorted into units  
135 (MClust 3.5, A.D. Redish et al.). Isolated units containing a minimum of 200 spikes within a session were  
136 included for subsequent analysis. Units were classified as fast spiking interneurons (FSIs) by an absence of  
137 interspike intervals (ISIs) > 2 s, while medium spiny neurons (MSNs) had a combination of ISIs > 2 s and  
138 phasic activity with shorter ISIs (Barnes 2005, Atallah 2014).

139 **Data analysis:**

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

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

153 activity, respectively.

154 To determine the relative contribution of different task parameters to firing rate variance (as in Figures 5-6),  
155 a forward selection stepwise general linear model (GLM) was fit to each cue-modulated unit. Cue identity  
156 (light block, sound block), cue location (arm 1, arm 2, arm 3, arm 4), cue outcome (reward-available, reward-  
157 unavailable), behavior (approach, skip), trial length, trial number, and trial history (reward availability on the  
158 previous 2 trials) were used as predictors, and the 1 s post-cue firing rate as the response variable. Units were  
159 classified as being modulated by a given task parameter if addition of the parameter significantly improved  
160 model fit using deviance as the criterion ( $p < .01$ ). A comparison of the R-squared value between the final  
161 model and the final model minus the predictor of interest was used to determine the amount of firing rate  
162 variance explained by the addition of that predictor for a given unit.

163 To better visualize responses to cues and enable subsequent population level analyses (as in Figures 5, 7,  
164 and 8), spike trains were convolved with a Gaussian kernel ( $\sigma = 100$  ms), and peri-event time histograms  
165 (PETHs) were generated by taking the average of the convolved spike trains across trials for a given task  
166 condition. For analysis of population-level responses for cue features (Figure 7), convolved spike trains  
167 for all units where cue identity, cue location, or cue outcome explained a significant portion of firing rate  
168 variance were z-scored. Within a given cue feature, normalized spike trains were then separated according  
169 to the preferred and non-preferred cue condition (e.g. light vs. sound block), and averaged across units to  
170 generate population-level averages. To account for separation that would result from any random selection  
171 of units, cell identity was shuffled and the shuffled average for preferred and non-preferred cue conditions  
172 was generated for 1000 shuffles.

173 To visualize NAc representations of task space within cue conditions, normalized spike trains for all units  
174 were ordered by the location of their maximum or minimum firing rate for a specified cue condition (Figure  
175 8). To compare representations of task space across cue conditions for a cue feature, the ordering of units  
176 derived for one condition (e.g. light block) was then applied to the normalized spike trains for the other

177 condition (e.g. sound block). For control comparisons within cue conditions, half of the trials for a condition  
178 were compared against the other half. To look at the correlation of firing rates of all cells within and across  
179 various cue conditions, trials for each cue condition for a cell were shuffled and divided into two averages,  
180 and averages within and across cue conditions were correlated. A linear mixed effects model was run for  
181 each cue condition to determine if correlations of firing rates within cue conditions were more similar than  
182 correlations across cue conditions.

183 To identify the responsivity of units to different cue features at the time of a nosepoke into a reward re-  
184 ceptacle, and subsequent reward delivery, the same cue-responsive units from the cue-onset analyses were  
185 analyzed at the time of nosepoke and reward receipt using identical analysis techniques (Figures 9, 10, 11,  
186 and 12).

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

194 **Histology:**

195 Upon completion of the experiment, rats were anesthetized with 5% isoflurane, then asphyxiated with carbon  
196 dioxide. Transcardial perfusions were performed, and brains were fixed and removed. Brains were sliced  
197 in 50  $\mu$ m coronal sections and stained with thionin. Slices were visualized under light microscopy, tetrode  
198 placement was determined, and electrodes with recording locations in the NAc were analyzed (Figure 3).

[Figure 3 about here.]

200 **Results**

201 **Behavior**

202 Rats were trained to discriminate between cues signaling the availability and absence of reward on a square  
203 track with four identical arms for two distinct sets of cues (Figure 2). During each session, rats were pre-  
204 sented sequentially with two behavioral blocks containing cues from different sensory modalities, a light and  
205 a sound block, with each block containing a cue that signalled the availability of reward (reward-available),  
206 and a cue that signalled the absence of reward (reward-unavailable). To maximize reward receipt, rats should  
207 approach reward sites on reward-available trials, and skip reward sites on reward-unavailable trials (see Fig-  
208 ure 4A for an example learning curve). All four rats learned to discriminate between the reward-available  
209 and reward-unavailable cues for both the light and sound blocks as determined by reaching significance ( $p <$   
210 .05) on a daily chi-square test comparing approach behavior for reward-available and reward-unavailable  
211 cues for each block, for at least three consecutive days (range for time to criterion: 22 - 57 days). Main-  
212 tenance of behavioral performance during recording sessions was assessed using linear mixed effects models  
213 for both proportion of trials where the rat approached the receptacle, and trial length. Analyses revealed that  
214 the likelihood of a rat to make an approach was influenced by which of the cues was presented (Percentage  
215 approached: light reward-available = 97%; light reward-unavailable = 34%; sound reward-available = 91%;  
216 sound reward-unavailable 35%;  $p < .001$ ), but that cue identity did not significantly affect the length of  
217 time taken to complete a trial (Trial length: light reward-available = 1.85 s; light reward-unavailable = 1.74  
218 s; sound reward-available = 1.91 s; sound reward-unavailable 1.78 s;  $p = .13$ ; Figure 4C,D). Thus, during  
219 recording, rats successfully discriminated the cues according to whether or not they signaled the availability  
220 of reward at the reward receptacle.

**222 NAc neurons encode behaviorally relevant and irrelevant cue features****223 Single unit responses discriminate cue features:**

224 We sought to address which parameters of our task were encoded by NAc activity, specifically whether the  
225 NAc encodes aspects of motivationally relevant cues not directly tied to reward, such as the identity and  
226 location of the cue, and whether this coding is independent or integrated with coding of cue outcome. To do  
227 this we recorded a total of 443 units with > 200 spikes in the NAc from 4 rats over 57 sessions while they  
228 performed a cue discrimination task (Table 1). Units that exhibited a drift in firing rate over the course of  
229 either block were excluded from further analysis, leaving 344 units for further analysis. The activity of 133  
230 (39%) of these 344 units were modulated by the cue, with more showing a decrease in firing ( $n = 103$ ) than  
231 an increase ( $n = 30$ ) around the time of cue-onset (Table 1). Within this group, 24 were classified as FSIs,  
232 while 109 were classified as SPNs. Upon visual inspection, we observed several patterns of firing activity,  
233 including units that discriminated firing upon cue-onset across various cue conditions, showed sustained  
234 differences in firing across cue conditions, had transient responses to the cue, showed a ramping of activity  
235 starting at cue-onset, and showed elevated activity immediately preceding cue-onset, for example (Figure 5).  
236 To characterize more formally whether these cue-evoked responses were modulated by various aspects of  
237 the task, we fit a GLM to each cue-modulated unit. Fitting GLMs revealed that a variety of task parameters  
238 accounted for a significant portion of firing rate variance in NAc cue-modulated units (Figure 6, Table 1).  
239 Notably, there were units that discriminated between whether the rat was performing in the light or sound  
240 block (28% of cue-modulated units, accounting for 6% of variance on average), which arm the rat was  
241 currently on (38% of cue-modulated units, accounting for 6% of variance on average), and whether the rat  
242 was engaged in the common portion of a reward-available or reward-unavailable trial (26% of cue-modulated  
243 units, accounting for 4% of variance on average), suggesting that the NAc independently encodes features

244 of reward-predictive cues separate from expected outcome (Figure 5A-F). Furthermore, overlap of coding of  
245 cue features within units was not different than expected by chance according to chi-square tests, suggesting  
246 for integrated coding across various aspects of a cue (Figure 5G,H). Fitting a GLM to all recorded units  
247 revealed a similar pattern of results (data not shown). Together, these findings show that various cue features  
248 are represented in the NAc, and that this coding has some overlap, but is mostly distinct from expected  
249 outcome (Figure 1; H2,H3).

250 [Table 1 about here.]

251 [Figure 5 about here.]

252 [Figure 6 about here.]

253 **Population level averages reveal characteristic response profiles:**

254 We observed a variety of single unit response profiles around the time of cue onset (Figure 5). To investigate  
255 whether these firing rate patterns were related to what cue features were encoded, we plotted the population  
256 level averages for units that were modulated by each feature. To do this, we normalized firing activity for  
257 each unit that was modulated by a given cue feature, such as light block, then generated the cue-onset aligned  
258 population average firing rate for each of the cue features (Figure 7). Overall, this analysis revealed that cells  
259 that showed an increase upon cue presentation had stronger responses for the preferred cue condition (Figure  
260 7A,C,E). Interestingly, units that were classified as decreasing in response to the cue showed a biphasic  
261 response at the population level, with a small peak at a time in alignment with entry into the arm, followed by  
262 a sustained dip after cue-onset (Figure 7B,D,F). Units that were modulated by cue identity showed a stronger  
263 increase in response to the preferred task block, as well as a higher tonic firing rate to the preferred task block,  
264 most notably in units that decreased in firing rate to the cue (Figure 7A,B). Units that were modulated by cue  
265 location showed a graded response to locations of decreasing preference, with peak firing occurring around

266 cue-onset (Figure 7C,D). Units that were modulated by cue outcome showed a ramping of activity after  
267 cue-onset for their preferred cue type. Additionally, units that exhibited a decrease in firing in response to  
268 the cue and whose activity was modulated by cue outcome, showed a sustained discriminatory response to  
269 reward-available and reward-unavailable cues that extended beyond cue-onset (Figure 7F). Together, these  
270 visualizations of the averaged population responses revealed nuanced differences in the way NAc units are  
271 modulated by cue conditions across cue features.

272 [Figure 7 about here.]

273 **NAc units dynamically segment the task:**

274 Given the varied time courses and response profiles of NAc units to various aspects of the cue, the NAc  
275 may be computing a temporally evolving state value signal (Pennartz., 2011). If this is the case, then the  
276 recruitment of NAc units should vary alongside changes in the environment. To look at the distribution of  
277 responses throughout our task space and see if this distribution is modulated by cue features, we z-scored  
278 the firing rate of each unit and plotted the normalized firing rates of all units aligned to cue-onset and sorted  
279 them according to the time of peak firing rate (Figure 8). We did this separately for both the light and  
280 sound blocks, and found a nearly uniform distribution of firing fields in task space that was not limited to  
281 alignment to the cue (Figure 8A). Furthermore, to determine if this population level activity was similar  
282 across blocks, we also organized firing during the sound blocks according to the ordering derived from the  
283 light blocks. (**JIMMIE EDIT THIS SECTION!**) This revealed that while there was some preservation of  
284 order, the overall firing was qualitatively different across the two blocks. To control for the possibility that  
285 any comparison of trials would produce this effect, we did a within block comparison, comparing half of  
286 the trials in the light block against the other half. This comparison looked similar to our test comparison of  
287 sound block trials ordered by light block trials. Additionally, given that the majority of our units showed  
288 an inhibitory response to the cue, we also plotted the firing rates according to the lowest time in firing, and  
289 again found some maintenance of order, but largely different ordering across the two blocks, and the within

290 block comparison (Figure 8B). To further test this, we divided each block into two halves and looked at the  
291 correlation of the average smoothed firing rates across various combinations of these halves across our cue-  
292 aligned centered epoch. A linear mixed effects model revealed that within block correlations (e.g. one half  
293 of light trials vs other half of light trials) were higher and more similar than across block correlations (e.g.  
294 half of light trials vs half of sound trials), suggesting that activity in the NAc discriminates across various cue  
295 conditions (within block correlations = .381, across block correlations = .342). This process was repeated  
296 for cue location and cue outcome, showing that NAc segmentation of the task is qualitatively different even  
297 during those parts of the task not immediately associated with a specific cue, action, or outcome, although  
298 the within condition comparison of reward-unavailable trials was less correlated than reward-available trials,  
299 and more similar to the across condition comparisons (Figure 8C-F).

300 [Figure 8 about here.]

301 **Encoding of cue features is maintained until outcome:**

302 In order to be useful for credit assignment in reinforcement learning, a trace of the cue must be maintained  
303 until the outcome, so that information about the outcome can be associated with the outcome-predictive cue.  
304 To test whether representations of cue features were maintained post-approach until the outcome was re-  
305 vealed, we fit a GLM to the post-approach firing rates of cue-modulated units aligned to the time of nosepoke  
306 into the reward receptacle. This analysis showed that a variety of units still discriminated firing according  
307 to various cue features, but not other task parameters, showing that NAc activity discriminates various cue  
308 conditions well into a trial (Table 1, Figures 9,10). Additionally, these units were a mix between most of the  
309 units that encoded cue features at cue-onset (observed overlap greater than expected by chance according to  
310 chi-square tests), and those that did not previously have a cue feature as a predictor (29, 48, and 30 units en-  
311 coded both time points for cue identity, location, and outcome, respectively). Fitting a GLM to all recorded  
312 units revealed a similar pattern (data not shown). Population level averages for units that increased to cue-  
313 onset showed a ramping up of activity that peaked upon nosepoke, whereas units that decreased to cue-onset

314 showed a gradual reduction of firing activity that reached a minimum upon nosepoke (Figure 11). Addition-  
315 ally, a peak is seen for preferred cue outcome in decreasing units at 1 second post cue-onset when reward was  
316 received, demonstrating an integration of expected and received reward (Figure 11F). Furthermore, aligning  
317 normalized peak firing rates to nosepoke onset, revealed a clustering of responses around reward receipt for  
318 all cue conditions where the rat would have received reward (Figure 12). To determine whether coding of cue  
319 features was maintained after the outcome was revealed, a GLM was fit to the firing rates of cue-modulated  
320 units at the time of reward receipt, during which the cue was still present. Fitting a GLM revealed 10 units  
321 (8%) where cue outcome accounted for an average of 32% of firing rate variance (Table 1, data not shown).  
322 However, an absence of cue identity or cue location coding was observed, suggesting that the NAc does not  
323 maintain a representation of these cue features once the rat receives behavioral feedback for its decision.

324 [Figure 9 about here.]

325 [Figure 10 about here.]

326 [Figure 11 about here.]

327 [Figure 12 about here.]

## 328 Discussion

329 The present study found evidence for coding of multiple identifying features of motivationally relevant stim-  
330 uli in addition to expected outcome; the sensory modality of the presented cue, as well as its physical location  
331 within the track, and that this coding was maintained on approach trials during a delay period where the rat  
332 held a nosepoke until receipt of outcome. Furthermore, this coding was both independent, and intermixed

333 across cue features and behavioral measures. Coding at the population level revealed qualitatively distinct  
334 comparison profiles across cue features, such as cells that encoded cue modality tended to sustain the dis-  
335 criminatory firing outside of cue presentation, while a subset cells that discriminated reward-available from  
336 reward-unavailable cues maintained this discrimination until after reward receipt. Across all recorded cells,  
337 a tiling of task structure was observed such that all points within our analyzed task space was accounted  
338 for by the ordered peak firing rates of all cells, and this tiling differed between various conditions with a  
339 cue feature, such as light versus sound blocks. Cells that discriminated across blocks were not simply due  
340 to drifting of the signal across trials, as cells that showed a drift in firing between the first and second half  
341 within a block were excluded from the analysis. Furthermore, even though actions were stereotyped during  
342 correct trials, such that the rat always turned left at the decision point to approach for reward, and right to  
343 skip the receptacle and initiate the next trial, cells that were modulated by the expected value of the cue  
344 maintained their specific firing patterns even during error trials where the rat turned left after presentation of  
345 the reward-unavailable cue, suggesting that these signals did not represent action values. Additionally, NAc  
346 signals have been shown to be modulated by response vigor, to detangle this from our results we included  
347 the trial length as a predictor in our GLMs, and found cells with correlates independent of trial length.

348 **Potential functional consequences of maintained coding:**

349 NAc has been shown to encode stimulus-outcome associations. In reinforcement learning, these associa-  
350 tions are commonly framed as a value estimate tied to the reward-predicting stimulus. This view, while  
351 informative, may not be complete in the sense that these value estimates may be more complex than just an  
352 agnostic value signal, for instance, the NAc is tuned to the identity of a reward. Our goal was to expand  
353 our understanding of the informational content of these outcome-predictive signals to include aspects of this  
354 environmental signal expected to be encoded by the NAc based on its corticolimbic inputs (discrete cues -  
355 amygdala input, spatial context - hpc input, task rules/set, behavioral strategy - pfc input). Having an en-  
356 riched (multidimensional?) representation that includes details about the environmental context the animal  
357 finds itself and is maintained online until the outcome is revealed could be useful for assigning credit of a

358 reward to the appropriate elements of an environment, to develop an accurate value function and optimize  
359 long term acquisition of reward (2012 Lee - eligibility trace). From a motivated behavior standpoint, the  
360 presence of these enriched representations can help inform action selection involving Pavlovian behaviors  
361 (conditioned approach), and after a decision has been made the maintained representation can help to hold  
362 the response in the face of competing alternatives while waiting for an outcome (Floresco 2015 summarizes  
363 this literature).

364 If these representations function as an eligibility trace, then there should be a relationship between the con-  
365 tent and robustness of these representations and the degree to which organisms form adaptive associations  
366 between reinforcers and their environment, which could contribute to the presence of prediction errors in the  
367 NAc in those that learned an experimental task in humans versus those that did not (Schonberg 2007). In  
368 the present experiment, understanding what information about the environment the NAc is keeping track of  
369 when reward arrives could be useful for potentially determining what representations are being reinforced.  
370 Would be interesting to see if selectively reward an animal when certain representations are active (e.g. arm  
371 1), would help shape their assignment of reward to that arm. Could be useful for altering a maladaptive pref-  
372 erence back to a more adaptive one, as the loss of certain aspects of coding via BLA and PFC inputs leads to  
373 the loss of outcome-predictive activity to discrete cues in the NAc. Furthermore, non-discriminatory coding  
374 in the NAc could hypothetically be correlated to over-generalization of situations, in which the animal may  
375 not recognize it is in a different situation and perform actions that are inefficient or maladaptive for reward  
376 procurement.

377 **Cue identity:**

378 Our finding that NAc units can discriminate between outcome-predictive stimuli with similar biological  
379 significance, in the current study cues from different sensory modalities, expands upon an extensive literature  
380 examining neural correlates of conditioned stimuli. Perhaps the most comparable work in rodents comes  
381 from a study that found distinct coding for an odor when it predicted separate but equally valued rewards

382 (Coch). The present work is complementary to this as it shows that NAc cells have representations of  
383 identifiable aspects of the cue itself, in addition to the reward it predicts. Another study paired separate  
384 cues with appetitive or aversive outcomes, and found separate populations of cells that encode each cue,  
385 with many switching selectivity after reversal of the associations between the cues and outcomes, providing  
386 evidence that the NAc encodes the biological significance of stimuli. Once again, our study was different  
387 as we recorded neural responses to distinct cues encoding the same anticipated outcome, suggesting that  
388 even when the biological relevance of these stimuli is similar, the NAc dissociates their representations at  
389 the level of the single-unit (Setlow), although this doesn't explain the difference in population averaged firing  
390 rate for the epochs when the cue is not present. Another possibility is that these modality specific cells were  
391 encoding the context, rule, or sequence within a session as some cells responded similar for both reward-  
392 available and reward-unavailable cues within a block. This interpretation is in alignment with a recent paper  
393 from the non-human primate literature that recorded NAc responses during the Wisconsin Card Sorting Task  
394 (WCST), a common set-shifting task used in both the laboratory and clinic, and found cells that preferred  
395 firing to stimuli when a certain rule, or rule category was currently active (Sleazer). Indeed, an encoding of  
396 the current strategy could be an explanation as to why a sustained difference in population averaged firing  
397 was seen across stimulus blocks, as well as a potential explanation for the differentially tiling of task structure  
398 across blocks in the current study. Further support for a modulation of NAc responses by strategy comes from  
399 an fMRI study that examined BOLD levels during a set-shifting task (FitzGerald et al., 2014). In this task,  
400 participants learned two sets of stimulus-outcome contingencies, a visual set and auditory set. During testing  
401 they were presented with both simultaneously, and the stimulus dimension that was relevant was periodically  
402 shifted between the two. Here, they found that bilateral NAc activity reflected value representations of  
403 whatever the currently relevant stimulus dimension was, and not the irrelevant stimulus. The current finding  
404 of separate, but overlapping, populations of cells encoding cue modality and expected value, suggests that  
405 the fMRI finding is generated by the combined activity of several different functional cell types.

406 A caveat of the current study is that rats were never presented with both sets of cues simultaneously, and  
407 thus never had to switch strategies, although extrapolating the data from the primate study, suggests that the

408 activity of the cue identity cells would be modulated by relevance. Keeping along this theme, the current  
409 data set is unable to identify precisely what the modality-sensitive neurons were encoding, that is were they  
410 tracking representations of stimulus identity, a preferred context, or even a macroscale representation of  
411 progress through the session. Furthermore, their relevance for ongoing behavior is also uncertain. NAc core  
412 lesions have been shown to impair shifting between different behavioral strategies, and it is possible that  
413 selectively silencing the cells that prefer responding for a given modality or rule would impair performance  
414 when the animal is required to use that information, or artificial enhancement of those cells would cause  
415 them to use the rule when it is the inappropriate strategy.

416 **Encoding of position:**

417 Our finding that cue-evoked activity was modulated by cue location sides with some of the literature (Lavoie,  
418 1994; Tabuchi, 2000; Strait, 2016). An alternative explanation for a pure spatial representation, is that  
419 these are task segmentation correlates, keeping track of where in the task the rat is. A previous non-human  
420 primate paper has shown that when reward is contingent upon completion of a series of trials, separate  
421 populations of NAc neurons signal the start of a schedule, subsequent trials in the schedule, and the first  
422 trial in extended schedules (Shidara et al., 1998). This signalling of position within a sequence has been  
423 observed in subsequent studies, and it is possible that the our rats were keeping track of which specific arm  
424 they were in as part of a sequence of arms, and not just strictly a spatial representation (Mulder, 2004 and  
425 2005; Khamassi et al., 2008; Berke, 2009). Also, given that our task is pseudo-random, it is possible that the  
426 rats learned which cue to anticipate, and the neural activity could reflect this. However, this is unlikely as  
427 including a previous trial variable in the analysis did not explain a significant amount of firing rate variance  
428 in response to the cue for the vast majority of cells..

429 **Mixed selectivity:**

430 Several other papers have reported unit profiles that integrate different task-related variables. These papers

431 report integrated coding between expected value and subsequent motor responses, expected value and iden-  
432 tity of a reward, and a combination of spatial-, movement-, and reward-related features (Roesch, Lavoie,  
433 Cooch). However, our study is the first to show mixed selectivity among identifying features of a cue and  
434 expected outcome or behavior. The presence of mixed selectivity responses confers a larger number of input-  
435 output relationships that are available to a given neuron. A possible functional consequence of this attribute  
436 of NAc units, is the combination and transformation of various motivationally relevant features into a sig-  
437 nal informing downstream decoders such as the ventral pallidum about appropriate behaviors in obtaining  
438 motivationally relevant goals and biasing action selection towards these behaviors.

439 Mixed selectivity in the NAc could be a consequence of synaptic integration from a variety of anatomically  
440 distinct inputs, as seen in experiments examining the convergence of various NAc afferents at the level of  
441 synaptic transmission and stimulation-induced firing (Goto and Grace 2008). In one such experiment it  
442 was shown that NAc cells that responded to stimulation of either the fornix, amygdala, or PFC, typically  
443 responded to stimulation from all inputs (ODonnell and Grace, 1995). Furthermore, an interaction between  
444 these inputs was observed such that PFC stimulation failed to elicit spiking in the NAc neurons unless they  
445 were in a depolarized UP-state, a state induced by hippocampal stimulation and was dependent on an intact  
446 fornix. Hippocampal-induced suppression of other inputs has also been observed for the BLA (Mulder et al.,  
447 1998). Recently, it has also been shown that train stimulation of PFC afferents reduces hippocampal-evoked  
448 NAc responses, suggesting that there is competition between various inputs (Calhoon and ODonnell, 2013).  
449 These studies suggest that the integration of the variables we saw could be the result of this gating observed  
450 in behaviorally-independent preparations. However, given that we did not systematically manipulate these  
451 various limbic and cortical afferents, comments on the anatomical origins of the observed mixed selectivity  
452 responses are speculative at this point.

453 Integrating cue identity and value, as seen modestly in the present study, could be one neural instantiation of  
454 how value is associated with the appropriate predictive stimuli (credit assignment), keeping in mind that value  
455 encoding is distributed, redundantly in some aspects, across various structures (Hayden Nat Neuro opinion).

456 Indeed, lesions of the NAc impair the ability to learn changes in reward value or identity in an unblocking  
457 experiment, as well as disrupting dopamine RPEs generated by modification of timing of reward (McDannald  
458 2011, Takahashi 2016). Would be interesting to see if uncoupling the integrated coding of stimulus features  
459 and predictive properties of a cue has an effect on the ability of a rat to use reward-predictive cues to pursue  
460 the associated reward.

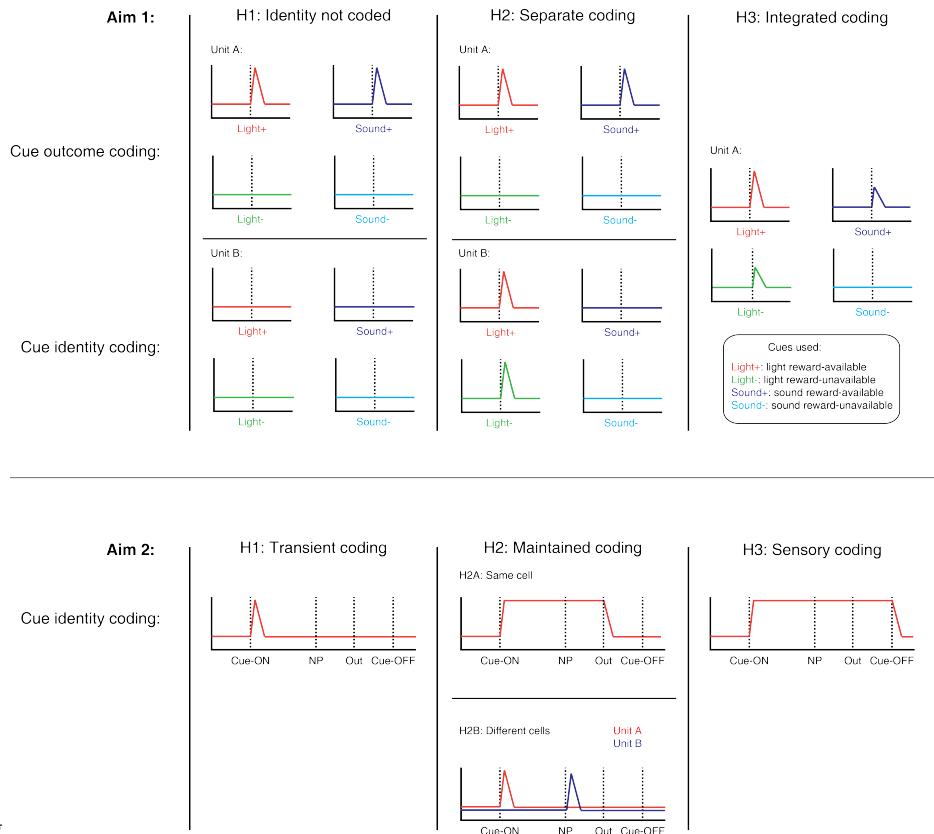
461 **Tiling of task structure:**

462 Additionally, we found that the population of recorded units had a relatively uniform distribution of firing  
463 fields within our task space, similar to what has been reported previously (Shidara, 1998; Berke, 2009;  
464 Lansink, 2012). Uniquely, we found that this representation differed across conditions of a cue features,  
465 such according to whether the rat was currently engaged in the light or sound block, suggesting that this  
466 could be a possible neural correlate for encoding the currently relevant strategy in the NAc. It has been  
467 previously shown that during progress through a predictable trial series, neurons represented state value  
468 of cue (Shidara 1998), and that single-unit responses allowed the monkey to know how it was progressing  
469 throughout the task. Likewise, the tiling we saw could be a consequence of upstream cortical or limbic inputs  
470 informing the striatum of the current task rules. Another possibility is that the NAc not only pays attention  
471 to progress throughout a task within a trial, but also higher-order task information, like blocks. Furthermore,  
472 dopamine levels in the NAc fluctuate through a trial, and it is possible that the observed tiling could be a  
473 NAc-representation of state value related to this temporally evolving dopamine signal. Future experiments  
474 should monitor this mapping of task structure during the application of dopamine antagonists. Finally, the  
475 presence of functional correlates not evident when looking at single-unit responses time-locked to salient  
476 task events emphasizes the need to employ ensemble level analyses across all aspects of a task.

477 **References**

- 478 Averbeck, B. B. and Costa, V. D. (2017). Motivational neural circuits underlying reinforcement learning. *Nature Neuroscience*,  
479 20(4):505–512.
- 480 Berridge, K. C. (2012). From prediction error to incentive salience: mesolimbic computation of reward motivation. *The European  
481 journal of neuroscience*, 35(7):1124–43.
- 482 Carelli, R. (2009). Drug Addiction: Behavioral Neurophysiology. In *Encyclopedia of Neuroscience*, pages 677–682. Elsevier.
- 483 Chang, S. E. and Holland, P. C. (2013). Effects of nucleus accumbens core and shell lesions on autosshaped lever-pressing. *Be-  
484 havioural Brain Research*, 256:36–42.
- 485 Chang, S. E., Wheeler, D. S., and Holland, P. C. (2012). Roles of nucleus accumbens and basolateral amygdala in autosshaped lever  
486 pressing. *Neurobiology of Learning and Memory*, 97(4):441–451.
- 487 Cheer, J. F., Aragona, B. J., Heien, M. L. A. V., Seipel, A. T., Carelli, R. M., and Wightman, R. M. (2007). Coordinated accumbal  
488 dopamine release and neural activity drive goal-directed behavior. *Neuron*, 54(2):237–44.
- 489 Corbit, L. H. and Balleine, B. W. (2011). The General and Outcome-Specific Forms of Pavlovian-Instrumental Transfer Are  
490 Differentially Mediated by the Nucleus Accumbens Core and Shell. *Journal of Neuroscience*, 31(33):11786–11794.
- 491 Day, J. J., Roitman, M. F., Wightman, R. M., and Carelli, R. M. (2007). Associative learning mediates dynamic shifts in dopamine  
492 signaling in the nucleus accumbens. *Nature neuroscience*, 10(8):1020–1028.
- 493 du Hoffmann, J. and Nicola, S. M. (2014). Dopamine Invigorates Reward Seeking by Promoting Cue-Evoked Excitation in the  
494 Nucleus Accumbens. *Journal of Neuroscience*, 34(43):14349–14364.
- 495 Estes, W. K. (1943). Discriminative conditioning. I. A discriminative property of conditioned anticipation. *Journal of Experimental  
496 Psychology*, 32(2):150–155.
- 497 Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czuj, A., Willuhn, I., Akers, C. A., Clinton, S. M., Phillips, P. E. M., and Akil,  
498 H. (2011). A selective role for dopamine in stimulusreward learning. *Nature*, 469(7328):53–57.
- 499 Floresco, S. B. (2015). The Nucleus Accumbens: An Interface Between Cognition, Emotion, and Action. *Annual Review of  
500 Psychology*, 66(1):25–52.
- 501 Hauber, W., Bohn, I., and Giertler, C. (2000). NMDA, but not dopamine D(2), receptors in the rat nucleus accumbens areinvolved  
502 in guidance of instrumental behavior by stimuli predicting reward magnitude. *The Journal of neuroscience : the official journal  
503 of the Society for Neuroscience*, 20(16):6282–8.
- 504 Hearst, E. and Jenkins, H. M. (1974). *Sign-tracking: The stimulus-reinforcer relation and directed action*. Psychonomic Society.
- 505 Hyman, S. E., Malenka, R. C., and Nestler, E. J. (2006). NEURAL MECHANISMS OF ADDICTION: The Role of Reward-Related  
506 Learning and Memory. *Annual Review of Neuroscience*, 29(1):565–598.
- 507 Ikemoto, S. (2007). Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbensolfac-

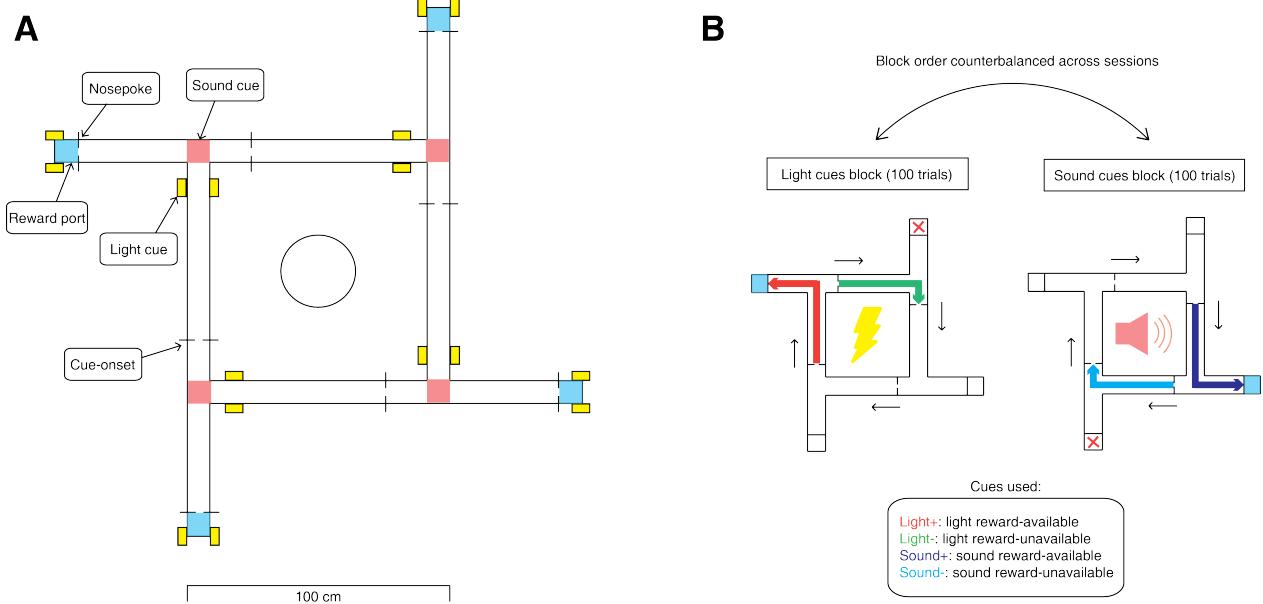
- 508 tory tubercle complex. *Brain Research Reviews*, 56(1):27–78.
- 509 509 Joel, D., Niv, Y., and Ruppin, E. (2002). Actorcritic models of the basal ganglia: new anatomical and computational perspectives.
- 510 510 *Neural Networks*, 15(4-6):535–547.
- 511 511 Kalivas, P. W. and Volkow, N. D. (2005). The Neural Basis of Addiction: A Pathology of Motivation and Choice. *American Journal*
- 512 512 *of Psychiatry*, 162(8):1403–1413.
- 513 513 Khamassi, M. and Humphries, M. D. (2012). Integrating cortico-limbic-basal ganglia architectures for learning model-based and
- 514 514 model-free navigation strategies. *Frontiers in behavioral neuroscience*, 6:79.
- 515 515 Lee, D., Seo, H., and Jung, M. W. (2012). Neural Basis of Reinforcement Learning and Decision Making. *Annu. Rev. Neurosci*,
- 516 516 35:287–308.
- 517 517 Maia, T. V. (2009). Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cognitive, Affective, & Behav-*
- 518 518 *ioral Neuroscience*, 9(4):343–364.
- 519 519 Nicola, S. M. (2010). The flexible approach hypothesis: unification of effort and cue-responding hypotheses for the role of nucleus
- 520 520 accumbens dopamine in the activation of reward-seeking behavior. *The Journal of neuroscience : the official journal of the*
- 521 521 *Society for Neuroscience*, 30(49):16585–16600.
- 522 522 Niv, Y., Daw, N. D., Joel, D., and Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psy-*
- 523 523 *chopharmacology*, 191(3):507–520.
- 524 524 Rescorla, R. A. and Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian conditioning and
- 525 525 instrumental learning. *Psychological Review*, 74(3):151–182.
- 526 526 Robinson, T. E. and Flagel, S. B. (2009). Dissociating the Predictive and Incentive Motivational Properties of Reward-Related Cues
- 527 527 Through the Study of Individual Differences. *Biological Psychiatry*, 65(10):869–873.
- 528 528 Roitman, M. F., Wheeler, R. A., and Carelli, R. M. (2005). Nucleus Accumbens Neurons Are Innately Tuned for Rewarding and
- 529 529 Aversive Taste Stimuli, Encode Their Predictors, and Are Linked to Motor Output. *Neuron*, 45(4):587–597.
- 530 530 Salamone, J. and Correa, M. (2012). The Mysterious Motivational Functions of Mesolimbic Dopamine. *Neuron*, 76(3):470–485.
- 531 531 Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nature Reviews Neuroscience*,
- 532 532 17(3):183–195.
- 533 533 Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)*,
- 534 534 275(5306):1593–9.
- 535 535 Setlow, B., Schoenbaum, G., and Gallagher, M. (2003). Neural encoding in ventral striatum during olfactory discrimination learning.
- 536 536 *Neuron*, 38(4):625–36.
- 537 537 Takahashi, Y., Langdon, A., Niv, Y., and Schoenbaum, G. (2016). Temporal Specificity of Reward Prediction Errors Signaled by
- 538 538 Putative Dopamine Neurons in Rat VTA Depends on Ventral Striatum. *Neuron*, 91(1):182–193.



1 - Schematic neural.png

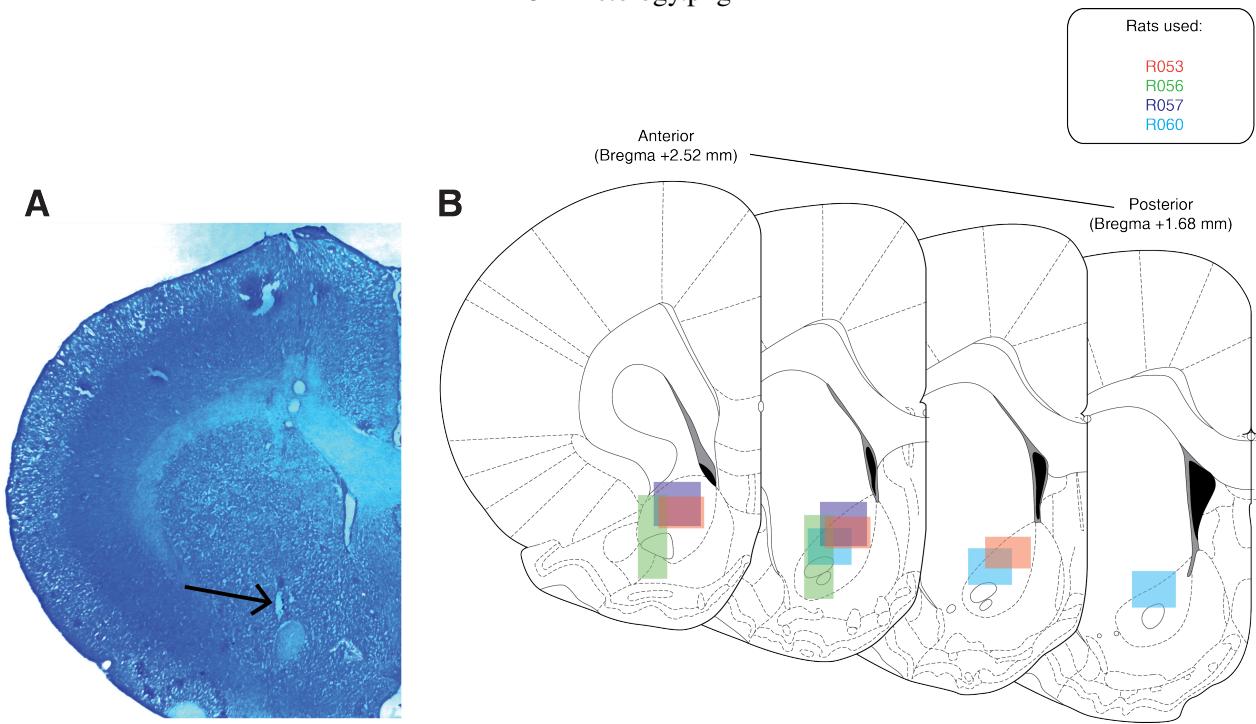
**Figure 1:** Schematic of potential coding strategies for cue identity (light, sound) and cue outcome (reward-available, reward-unavailable) employed by single units in the NAc. **A:** Displayed are schematic PETHs illustrating putative responses to different cues under different hypotheses of how cue identity and value 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 reward-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, 'Cue-OFF' signifies when the cue turns off. H2 (middle panel): Coding of cue identity is maintained 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 is maintained during entire duration of cue, including after outcome is revealed. 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.

2 - Schematic task.png



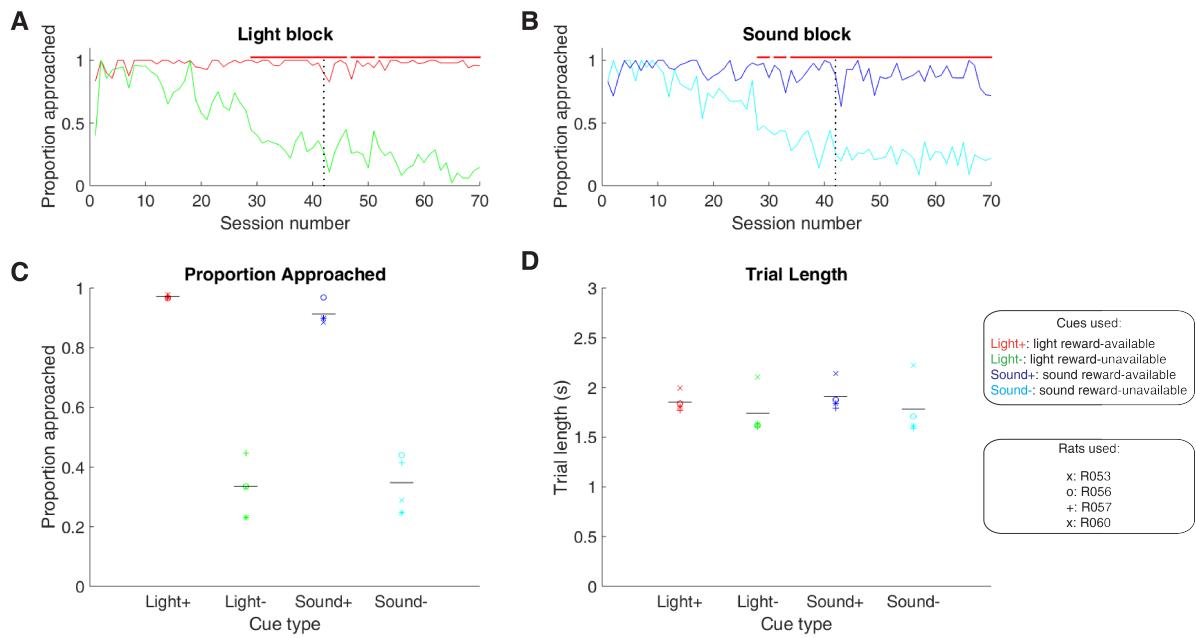
**Figure 2:** Schematic of behavioral task. **A:** To scale depiction of square track consisting of multiple identical T-choice points. At each choice point, the availability of sucrose reward at the nearest reward receptacle (light blue fill) was signaled by one of four possible cues, presented when 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. 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. Light blue fill on ends of arms indicate location of receptacles where sucrose was dispensed on rewarded trials. 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 two blocks of a cue discrimination task on a square track where they had to dissociate between a reward-available and reward-unavailable cue for a set of light cues and a set of sound cues for a target of 100 trials each. Rats ran clockwise on the square track, and upon presentation of the cue could either turn left at the choice point to check if reward was available at the reward receptacle (approach trial; red and navy blue in figure), or turn right at the choice point to initiate the following trial (skip trial; green and light blue in figure). 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. Left in figure depicts a light block, showing an example trajectory for a correct reward-available (red) and reward-unavailable (green) trial. Right in figure depicts a sound block, with a reward-available (navy blue) and reward-unavailable (light blue) trial. Ordering of the light and sound blocks was counterbalanced across sessions. Dashed line within track indicates photobeam that triggered cue onset. Solid line within track indicates photobeam that registered a nosepoke into the reward receptacles. Receptacles containing sucrose solution are present within the squares at the end of each arm. Light blue fill signifies reward receipt on a correct reward-available trial. Arrows outside of track indicate correct running direction. A session ended with another 5 minute recording period on the pedestal.

3 - Histology.png



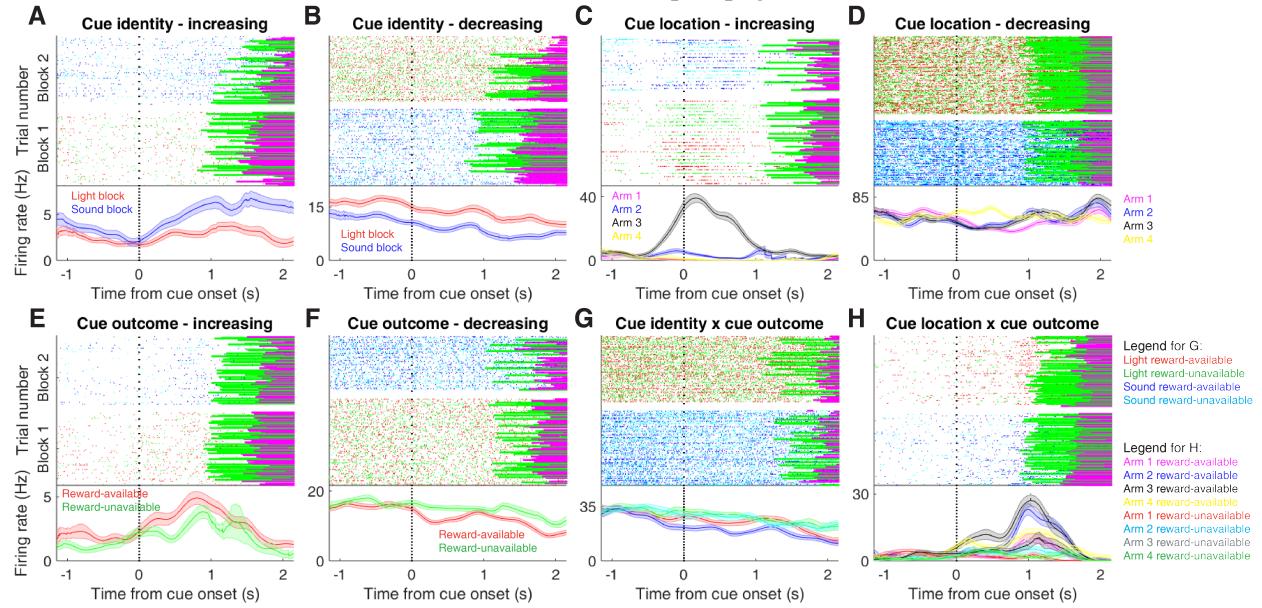
**Figure 3:** Histological verification of recording sites. Upon completion of experiments, brains were sectioned and tetrode placement was confirmed. **A:** Example section from R060 showing a recording site in the NAc core just dorsal to the anterior commissure (white circle). **B:** Schematic showing recording areas for the four rats used.

4 - Behavioral results.png



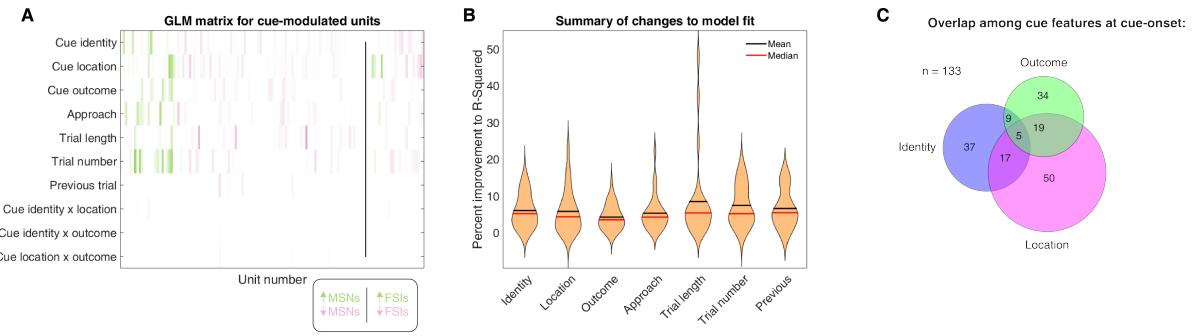
**Figure 4:** Performance on the behavioral task. **A-B.** Example learning curves across sessions from a single subject (R060) showing the proportion approached for reward-available (red line for light block, navy blue line for sound block) and reward-unavailable trials (green line for light block, light blue line for sound block). Fully correct performance corresponds to 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. **C-D:** Summary of performance during recording sessions for each rat. **C:** 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 (~90% approached) and reward-unavailable cues (~30% approached), for both blocks (see Results for statistics). **D:** Average trial length for each cue. Note that the time to complete a trial was comparable for the different cues.

5 - Neural examples.png



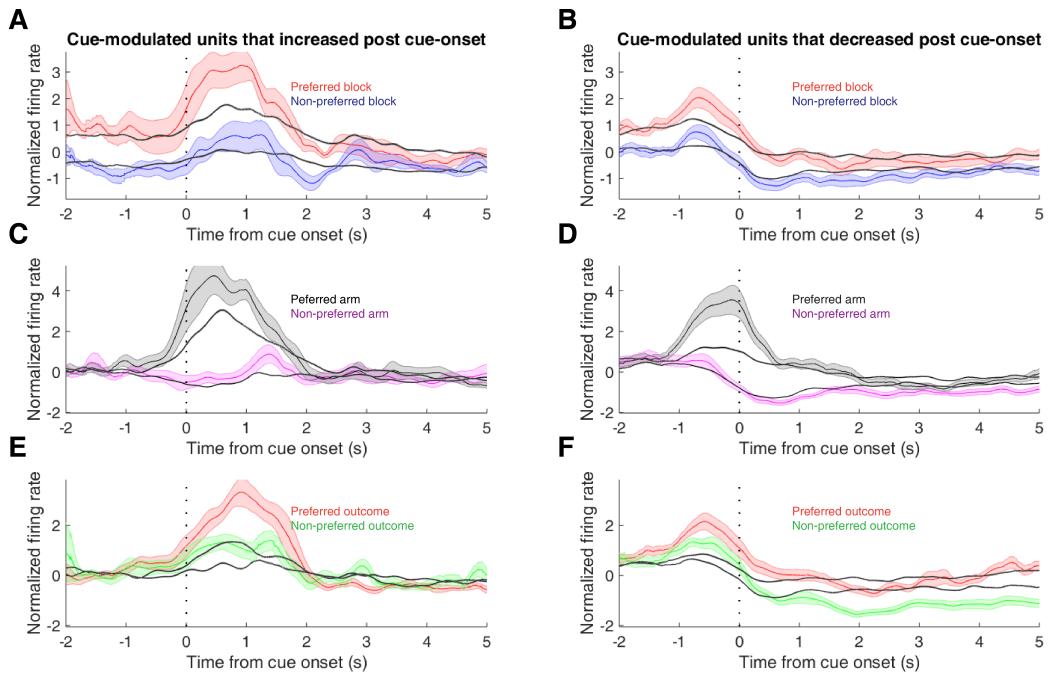
**Figure 5:** Examples of different cue-modulated NAc units influenced by various task parameters. **A:** Example of a cue-modulated NAc unit that showed an increase in firing following the cue, and encoded cue identity. Top: rasterplot showing the spiking activity across all trials aligned to cue-onset. 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.

## 6 - GLM.png

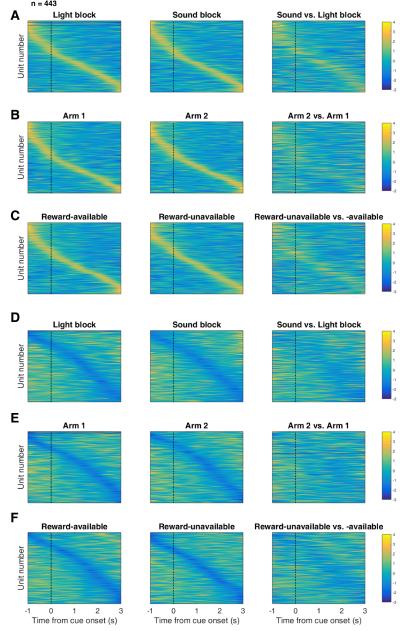


**Figure 6:** Summary of influence of various task parameters on cue-modulated NAc units after cue-onset. **A:** GLM matrix illustrating the contribution of various task parameters to NAc unit firing rates. A stepwise GLM was fit to each unit that showed evidence of cue modulation by a Wilcoxon signed-rank 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. **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.

## 7 - Population averages.png



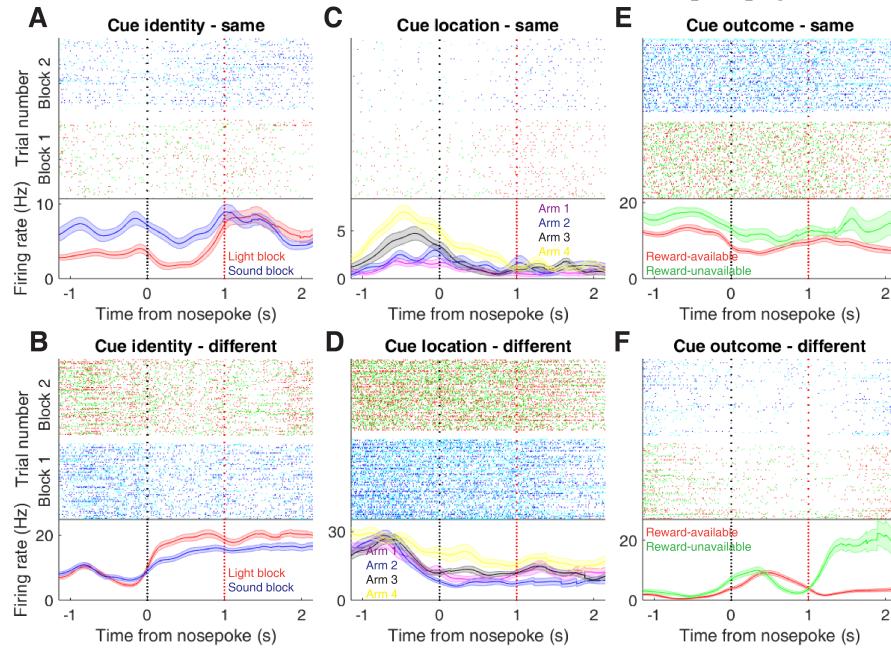
**Figure 7:** Population-level averages of cue feature sensitive NAc units. **A:** Average smoothed normalized (z-score) activity for cue-modulated units where cue identity was a significant predictor in the GLM, aligned to cue-onset. Activity is plotted for preferred stimulus block (red) and non-preferred stimulus block (blue). Dashed line indicates onset of cue. Lightly shaded area indicates standard error of the mean. Note larger increase to preferred stimulus block to 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 cells 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 from most preferred arm (yellow), in decreasing order to least preferred arm (black, navy blue, magenta, respectively). Note the graded response to arms of decreasing preference. **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 cells preferred outcome (E), and the sustained decrease to the nonpreferred outcome (F).



8 - Task tiling.png

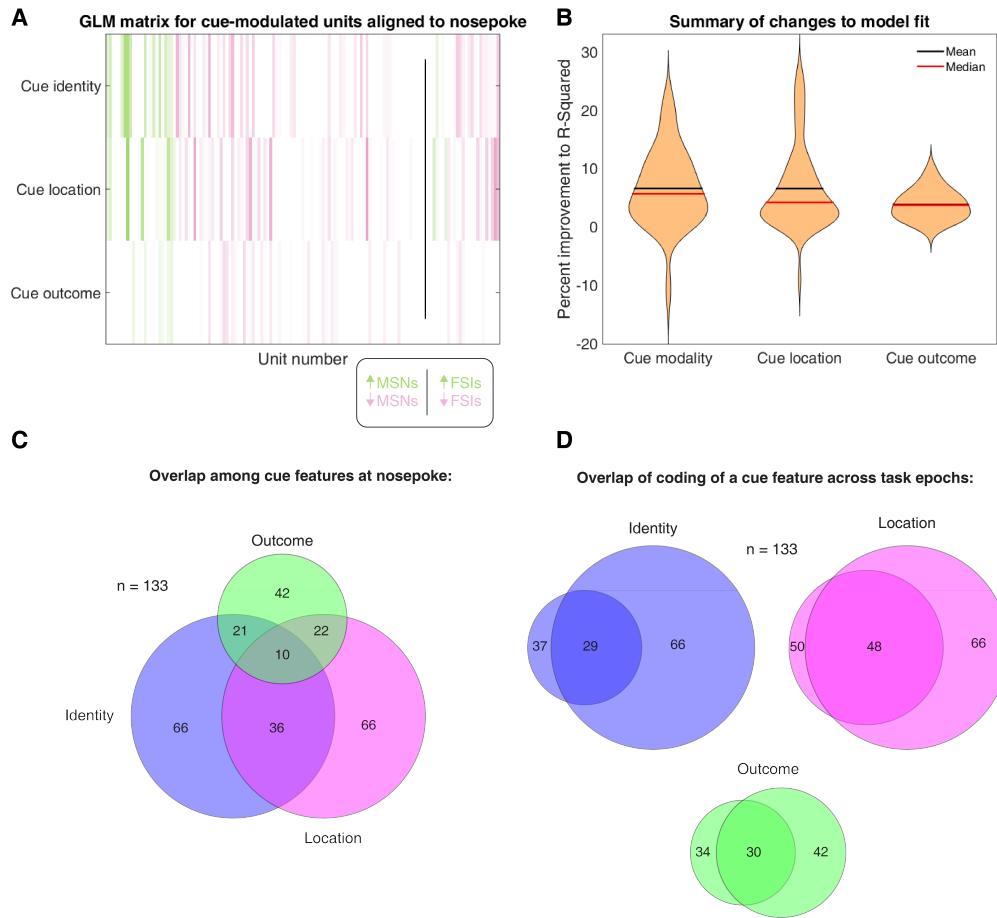
**Figure 8:** Distribution of NAc firing rates across time surrounding cue onset. Each panel shows normalized (z-score) firing rates for all recorded NAc units (each row corresponds to one unit) as a function of time (time 0 indicates cue onset), averaged across all trials for a specific cue type, indicated by text labels. **A-C:** Heat plots aligned to normalized peak firing rates. A, 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. 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 left: 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, middle 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, far right: Unit firing rates taken from half of the light block, ordered according to peak firing rate taken from the other half of light trials. Note similar pattern to middle right suggests 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.

9 - NP Neural examples.png



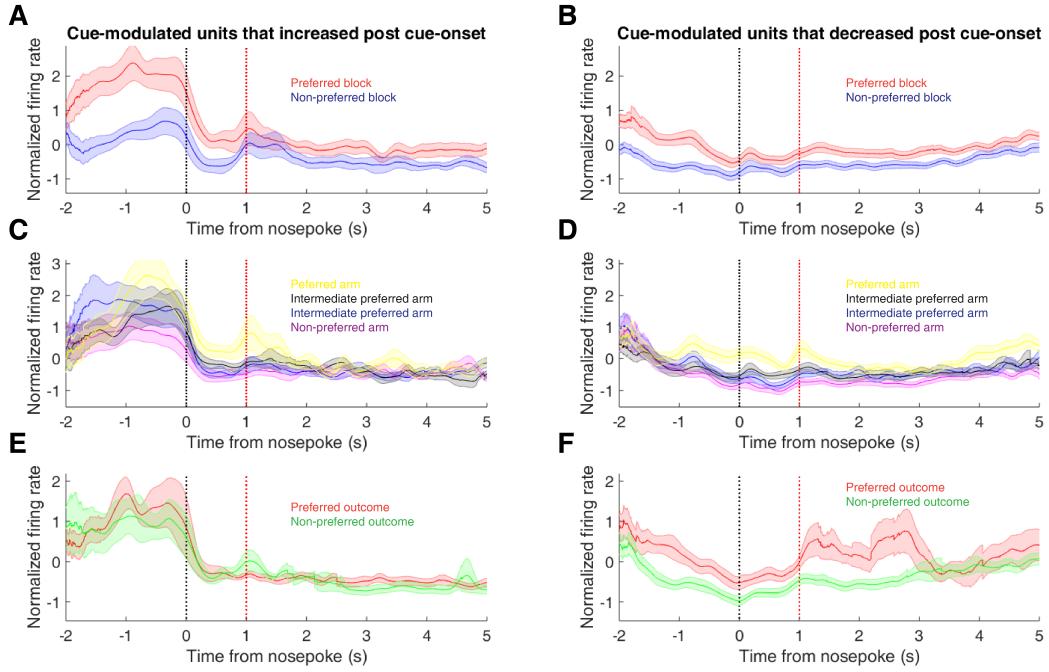
**Figure 9:** Examples of cue-modulated NAc units influenced by various task parameters at time of nosepoke. **A:** Example of a cue-modulated NAc unit that encoded cue identity at both cue-onset and during nosepoke hold. Top: rasterplot showing the spiking activity across all trials aligned to nosepoke. 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 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.

## 10 - NP GLM.png

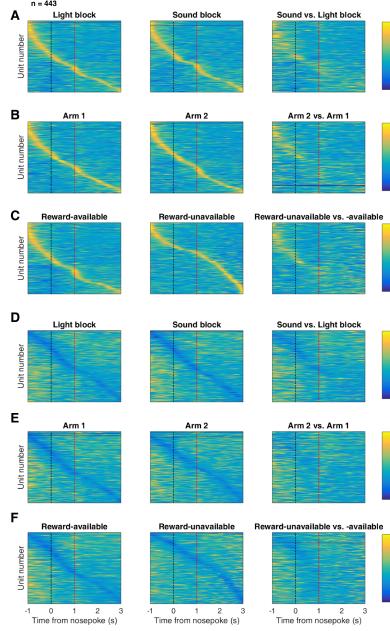


**Figure 10:** Summary of influence of various task parameters of cue-modulated NAc units during nosepoke. **A:** GLM matrix demonstrating impact of various task parameters on NAc firing rates. A stepwise GLM was fit to each unit that showed evidence of cue modulation by a Wilcoxon signed-rank test. Each row represents a given task parameter, and each column is the influence of that task parameter on a given unit. Response variable is 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.

## 11 - NP population averages.png



**Figure 11:** Population-level averages of cue feature sensitive NAc units during a nosepoke. **A:** Average smoothed normalized 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 dashed line indicates nosepoke. Red dashed line indicates reward delivery occurring 1 s after nosepoke for reward-available trials. Lightly shaded area indicates standard error of the mean. Note larger increase leading up to nosepoke to preferred stimulus block to 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 from most preferred arm (yellow), in decreasing order to least preferred arm (black, navy blue, magenta, respectively). Note the subtle but graded response to arms of decreasing preference. **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 reward receipt for preferred outcome in decreasing cells (F).



12 - NP task tiling.png

**Figure 12:** Distribution of NAc firing rates across task space during approach trials. Each panel shows normalized 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**, 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 left: 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 reward receipt. A, middle 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, far right: Unit firing rates taken from half of the light block, ordered according to peak firing rate taken from the other half of light trials. Note similar pattern to middle right suggests 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, but with units ordered according to the time of their minimum firing rate. **C**: 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. **D**: As for C, but with units ordered by their minimum firing rate. **E**: Same layout as in A, except that panels now compare reward-available and reward-unavailable trials. Notice the absence of a cluster of tiling around reward receipt for reward-unavailable trials. **F**: As for A, 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.

Task parameter	Total	↑ MSN	↓ MSN	↑ FSI	↓ FSI
All units	443	155	216	27	45
<b>Rat ID</b>					
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	?	?	?	?
Cue modulated units	133	24	85	6	18
<b>GLM aligned to cue-onset</b>					
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
Recent trial history	5	0	5	0	0
<b>GLM aligned to nosepoke</b>					
Cue identity	66	14	36	2	14
Cue location	66	14	40	3	9
Cue outcome	42	8	29	0	5
<b>GLM aligned to outcome</b>					
Cue outcome	10	0	6	0	4

**Table 1:** Cells overview