Contextual gating of motivationally-relevant stimuli in the mouse nucleus accumbens

Authors: Jimmie M. Gmaz[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 Andrew Alvarenga for the manufacturing of the lickometer and olfactory delivery system, and Jun Ho Lee for mouse husbandry advice. This work was supported by Dartmouth College (Dartmouth Fellowship to JMG, and startup funds to MvdM).

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

Abstract

Neural activity in the nucleus accumbens (NAc) is thought to track fundamentally value-centric quantities such as current or future expected reward, reward prediction errors, the value of work, opportunity cost, and approach vigor. However, the NAc also contributes to flexible behavior in ways that are difficult to explain based on value signals alone, raising the question of if and how non-value signals are encoded in NAc. We recorded NAc neural ensembles while head-fixed mice performed a biconditional discrimination task, and extracted single-unit and population-level correlates of task features. We found coding for context-setting cues that modulate the stimulus-outcome association of subsequently presented reward-predictive cues. This context signal occupied a subspace orthogonal to classic value representations, suggesting that it does not interfere with value-related NAc output. Finally, we show that the context signal is predictive of subsequent value coding, supporting a circuit-level gating model for how the NAc contributes to behavioral flexibility and providing a novel population-level perspective from which to view NAc computations.

Introduction

The nucleus accumbens (NAc) is an important contributor to the motivational control of behavior, acting directly through output pathways involving brainstem motor nuclei ("limbic-motor interface", Graybiel 1975; Gruber & McDonald 2012; Mogenson et al. 1980) and indirectly through return projections within cortico-striatal loops (Haber & Behrens, 2014; Rusu & Pennartz, 2020). Accordingly, leading theories of NAc function, and the mesolimbic dopamine (DA) system it is tightly interconnected with, tend to focus on the processing of reward (and punishment) and its dual role in energizing/directing ongoing actions as well as in learning from feedback (Floresco, 2015; Gruber & McDonald, 2012; Koob & Volkow, 2016; Nicola, 2010; Salamone & Correa, 2012). These proposals attribute to the NAc a role in motivational and reward-related quantities such as incentive salience, value of work, expected future reward, economic value, risk and reward prediction error. In more formal reinforcement learning models, the NAc-dopamine system is typically cast as an "evaluator" or "critic", tracking state values that are useful to set the value of work as well as a source of a teaching signal in the form of reward prediction errors (Averbeck & Costa, 2017; Joel et al., 2002; Khamassi & Humphries, 2012; Stoianov et al., 2018; van der Meer & Redish, 2011a). Although the specifics are the subject of vigorous debate, these prominent theories all share a fundamentally **value-centric** focus: notwithstanding substantial heterogeneity in NAc cell types and circuitry (Cox & Witten, 2019; Floresco, 2015; Humphries & Prescott, 2010) this brain structure as a whole is typically cast as tracking a relatively low-dimensional quantity: a value signal that at its simplest just a single number, reflecting how good or bad the current situation is.

These value-centric accounts are supported by a vast literature that includes the bidirectional control of motivated behaviors such as conditioned responding to reward-predictive cues, and the regulation of how much effort/vigor to exert, with NAc manipulations (Ciano et al., 2001; Corbit & Balleine, 2011; Ghods-Sharifi & Floresco, 2010; Nicola, 2010; Parkinson et al., 2000; Salamone et al., 1994) as well as the demonstration that electrical or optogenetic stimulation of the NAc itself, or dopaminergic terminals in the NAc, is sufficient for inducing behavioral preferences (Cole et al., 2018; Cox & Witten, 2019; Crow, 1972; Mogenson et al.,

1979; Phillips et al., 1975; Prado-Alcalá & Wise, 1984; Tsai et al., 2009). Similarly, unit recording studies in rodents and fMRI work in humans consistently report widespread, sizable value signals in NAc single units, populations, and the NAc blood-oxygen-level-dependent (BOLD) signal (Bissonette et al., 2013; Delgado et al., 2000; FitzGerald et al., 2014; Goldstein et al., 2012; Hollerman et al., 1998; McGinty et al., 2013; Nicola, 2004; Roesch et al., 2009; Roitman et al., 2005; Schultz et al., 1992; Setlow et al., 2003). Thus, there seems to be widespread agreement that the major dimension (principal component) of NAc activity is some form of value signal.

However, in complex, dynamic behavioral tasks, lesions or inactivations of the NAc lead to deficits that are not straightforward to explain from a purely value-centric perspective (Floresco, 2015; Mannella et al., 2013), such as the implementation of conditional rules (Floresco et al., 2018), or switching to a novel behavioral strategy (Floresco et al., 2006). In addition, prominent inputs from brain regions such as orbitofrontal cortex (OFC), medial prefrontal cortex (mPFC), and hippocampus (Gulli et al., 2020; Saez et al., 2015; Zhou et al., 2019) suggest that the NAc has access to non-value signals that would be expected to not only inform its function but help shape its neural activity. Indeed, a study in primates suggests that elements of task structure which are orthogonal to value, but nonetheless crucial for successful behavior on the task, are represented in NAc (Sleezer et al., 2016). In rodents, there have been hints of task structure too, but this has been hard to show conclusively due to the difficulty in cleanly dissociating task structure from value (Atallah et al. 2014; Gmaz et al. 2018; see also related work on dopamine neuron and OFC activity representing task structure, Sadacca et al. 2016; Zhou et al. 2019). Thus, it is currently unknown if, and how, task structure is encoded in rodent NAc, and if found, how such a signal relates to classical value correlates.

To address this issue, we trained mice to perform a biconditional discrimination task using odor cues, in which two different "context" cues determine whether a subsequent "target" cue will be rewarded (Figure 1A). Thus, in context O1, O3 but not O4 is rewarded, whereas in context O2, O4 but not O3 is rewarded. We recorded ensembles of NAc neurons and tested whether there is coding of the (reward-independent) context cues at the single cell and population level. Next, we sought to determine the relationship between

this context signal and classical value coding, using contemporary population analysis tools to test if context coding can be used to inform subsequent value-related processing of target cues.

Results

Mice learn to perform a biconditional discrimination task using odor cues

We sought to test whether NAc encodes information about task structure that is independent of reward. To do this, we used a biconditional discrimination task in which the identity of a "context" cue determines whether a subsequent "target" cue is rewarded or not (Gu & Li, 2019; Han et al., 2018; Zhang et al., 2019). We use the term "context" here to mean a cue that modifies the subsequent outcome value of the target cue (see Discussion). Briefly, a trial initiated with presentation of one of the two context cues for 1 s, followed by a 2 s delay, followed presentation of one of the two target cues for 1 s, followed by an additional 1 s response period (Figure 1). Animals had to make a licking response either during presentation of the target cue or the subsequent response period to get a sucrose reward for rewarded cue pairings. For example, given context cue O1, target cue O3 but not O4 is rewarded, but following context cue O2, O4 but not O3 is rewarded. Thus, rewarded trial types O1-O3 and O2-O4 both had the same outcome value at the time of target cue presentation, while unrewarded trial types O2-O3 and O1-O4 also had the same value. Importantly, a 2 s delay separated the two odor cues in a trial such that mice had to maintain a representation of the context cue while waiting for the target cue.

Mice (n = 4) completed a total of 7 to 28 training sessions to reach criterion before recording sessions began (See Figure 1C for an example learning curve; Figure 1D for number of training sessions for each mouse). During recording sessions, mice licked for a significantly larger proportion of rewarded trials than unrewarded trials (Figure 1E; Proportion of rewarded trials with a lick response: 0.82 +/- 0.08 SD; proportion

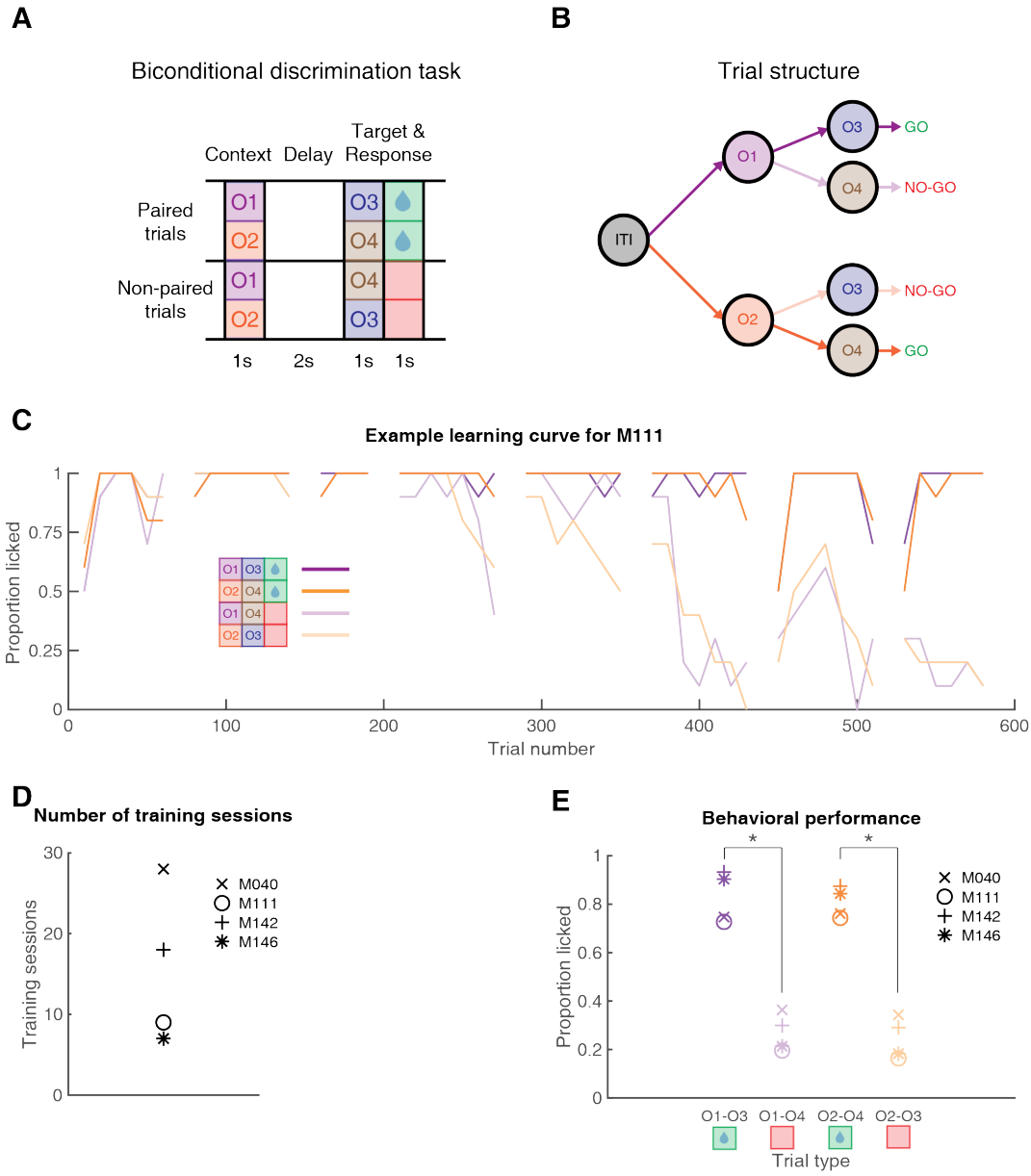

Figure 1: Schematic of the behavioral task and results. **A**: Mice were trained in a head-fixed biconditional discrimination task where they learned to discriminate between different pairings of context and target cues. Mice were first presented with the context cue (1 s), followed by a delay (2 s), followed by target cue presentation (1 s), and an additional response period (1 s). Whether a target cue was rewarded depended upon the identity of the preceding context cue. For instance, licking in response to O3 was rewarded when preceded by O1, but not O2, while for O4 was rewarded when preceded by O2, but not O1. Note, this means that by design, each context cue was rewarded on half of the trials it was presented on. **B**: Learned trial structure of the task. Purple arrows indicate trials with context cue 1 (O1), orange arrows indicate trials with context cue 2 (O2). Dark arrows after context cue presentation indicate rewarded trials, light arrows indicate unrewarded trials. This color scheme is used throughout the text. **C**: Example learning curve showing proportion of trials with a lick over the course of full-task training. Data is shown for each trial type, in 20 trial blocks, with gaps separating individual training sessions. This learning curve shows that by day 6 (trial $\sim$400), there was a clear difference in responding to rewarded versus unrewarded trials. **D**: Number of training sessions before each mouse reached the criterion of 3 consecutive sessions with $>$80% correct responding, after which recordings began. **E**: Behavioral performance during recording sessions showing the average proportion of trials with a licking response for each trial type for each mouse, demonstrating that mice discriminated between rewarded and unrewarded trial types. Asterisks above plot denotes a significant difference.

of unrewarded trials with a lick response: 0.26 +/- 0.08 SD; z-score across mice and sessions: 11.05; $p < 0.001$), but licked similarly for both rewarded target cues (O1-O3: 0.83 +/- 0.11 SD; O2-O4: 0.81 +/- 0.06 SD) and both unrewarded target cues (O1-O4: 0.27 +/- 0.08 SD; O2-O3: 0.25 +/- 0.09 SD; z-score across mice and sessions: 0.51; $p = 0.61$). Furthermore, individual mice showed a similar level of correct responding to the target cues during recording sessions (M040: 70% +/- 9% SD; M111: 78% +/- 4% SD; M142: 80% +/- 5% SD; M146: 83% +/- 8% SD). Therefore, mice learned the appropriate context-target cue associations in the task.

NAc single units signal context

We set out to determine if context is encoded by the NAc, particularly whether the NAc can discriminate between separate context cues that have equal outcome-predictive value, and whether this discrimination persists during the delay period after cue offset. We recorded a total of 320 units with > 200 spikes (out of 386 total) in the NAc from 4 mice over 41 sessions (range: 8 - 12 sessions per mouse) during performance on the task. Initial inspection of the data revealed a diversity of single-unit responses, including units that showed transient responses to all odor cues regardless of their significance in the task, units that that discriminated the various cue identities and their associations, and of particular relevance, units that showed a sustained discrimination between the context cues during the delay period (see Figure 2 for examples). The main cue features of interest in this task were context cue coding, context coding during the delay period after context cue offset, target cue coding, and value coding after target cue onset. To determine whether a given single unit encoded each feature, we focused on comparing the trial-averaged firing rate differences during periods surrounding context and target cue presentations (Figure 3). To determine general coding for odor cues, we compared the 1 s preceding and following cue presentation for all trial types. To investigate coding of cue features we compared the 1 s of cue presentation for context and target cue coding, and the 1 s preceding target cue presentation for context delay coding. Out of all units included in the analysis, 80 (Figure 3B; 26%) discriminated between the two context cues when analyzed during cue presentation, and 49 (Figure

3C; 15%) discriminated between the two contexts when analyzed during the following delay period, showing that individual units within the NAc code for context. Importantly, as there were no differences in behavioral performance for the two context cues (Figure 1), this context coding can not be explained by differences in the perceived value of the cues. Apart from context coding, 216 units (Figure 3A; 68%; 153 increasing, 63 decreasing) showed a change in firing activity in response to any context cue relative to a pre-cue baseline, 199 units (Figure 3D; 62%; 149 increasing, 50 decreasing) showed a change in firing activity in response to any target cue relative to a pre-cue baseline, 87 units (Figure 3E; 27%) discriminated between the two target cues, and 97 units (Figure 3F; 30%; 77 increasing, 20 decreasing) discriminated between rewarded and unrewarded trials during target cue presentation. Finally, to assess the overall relationship in firing rate during progression of a trial, we correlated the trial-averaged firing rate for each unit against itself across all timepoints, and found a general correlation between activity during context and target cue presentation, reflecting the large proportion of units that respond non-discriminately to any cue (Figure 3H). Together, these results suggest that the NAc is coding for the various motivationally-relevant features of the task, with particular interest being NAc units that discriminated between the context cues during the delay period, suggesting that the NAc maintains information about which context the animal is in after offset of the cue itself.

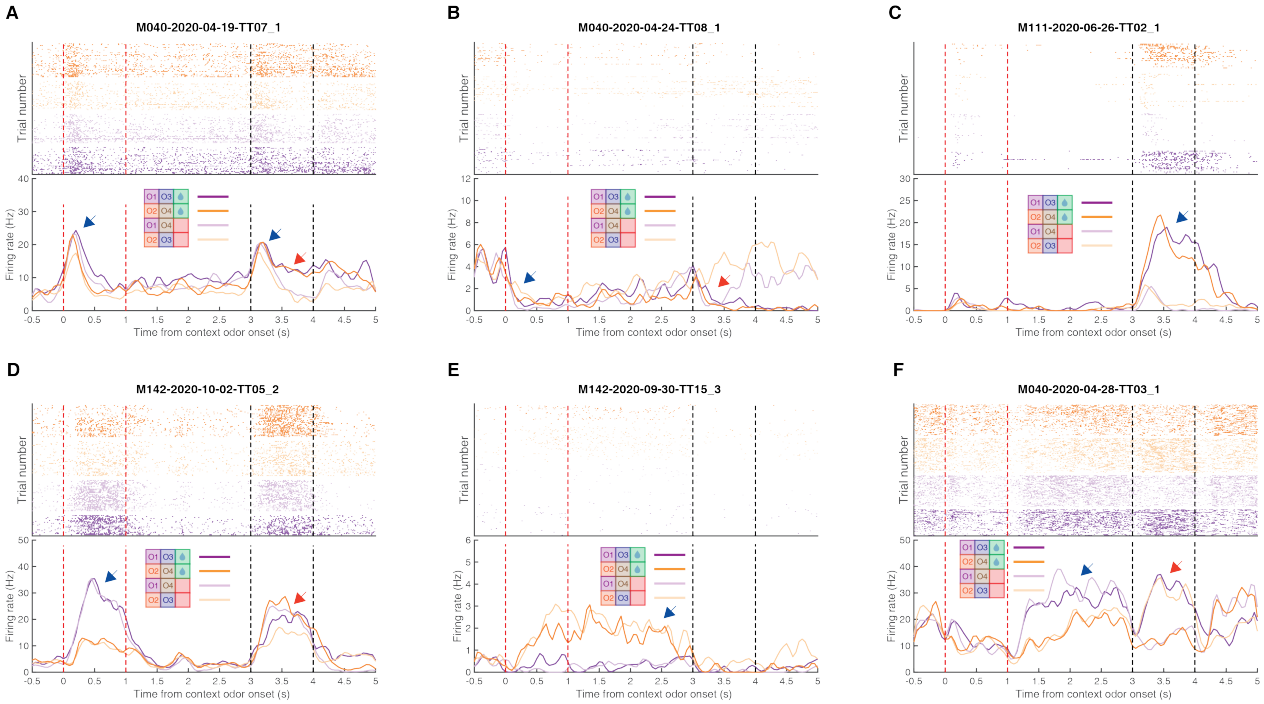

Figure 2: Example single-unit responses for units with value (top) and context (bottom) correlates. Top of each plot shows spike rasters for the four trial types (purple: trials with context cue O1; orange: trials with context cue O2; dark colors: rewarded trials; light colors: unrewarded trials). Bottom half of each panel shows trial-averaged firing rates for each trial type aligned to context cue onset. Context cue presentation (0-1 s) is bordered by red lines, and target cue presentation (3-4 s) is bordered by black lines. **A**: Example unit that shows a general response to cue presentation (blue arrow), as well as a subsequent discrimination between rewarded and unrewarded trials after target cue onset (red arrow). **B**: Example unit showing a dip in firing after context cue onset (blue arrow), followed by a ramping of activity leading up to target cue onset, and a subsequent dip in firing after presentation of the rewarded target cue (red arrow). **C**: Example unit that predominantly responds during presentation of the rewarded target cue (blue arrow). **D**: Example unit that shows transient responses to the cues, showing a discrimination to both context (blue arrow) and target (red arrow) cues. **E**: Example unit that discriminates between context cues, including throughout the delay period (blue arrow). **F**: Example unit that discriminates context cues only during the delay period following offset of the context cue (blue arrow), as well as discriminating the subsequent target cues (red arrow).

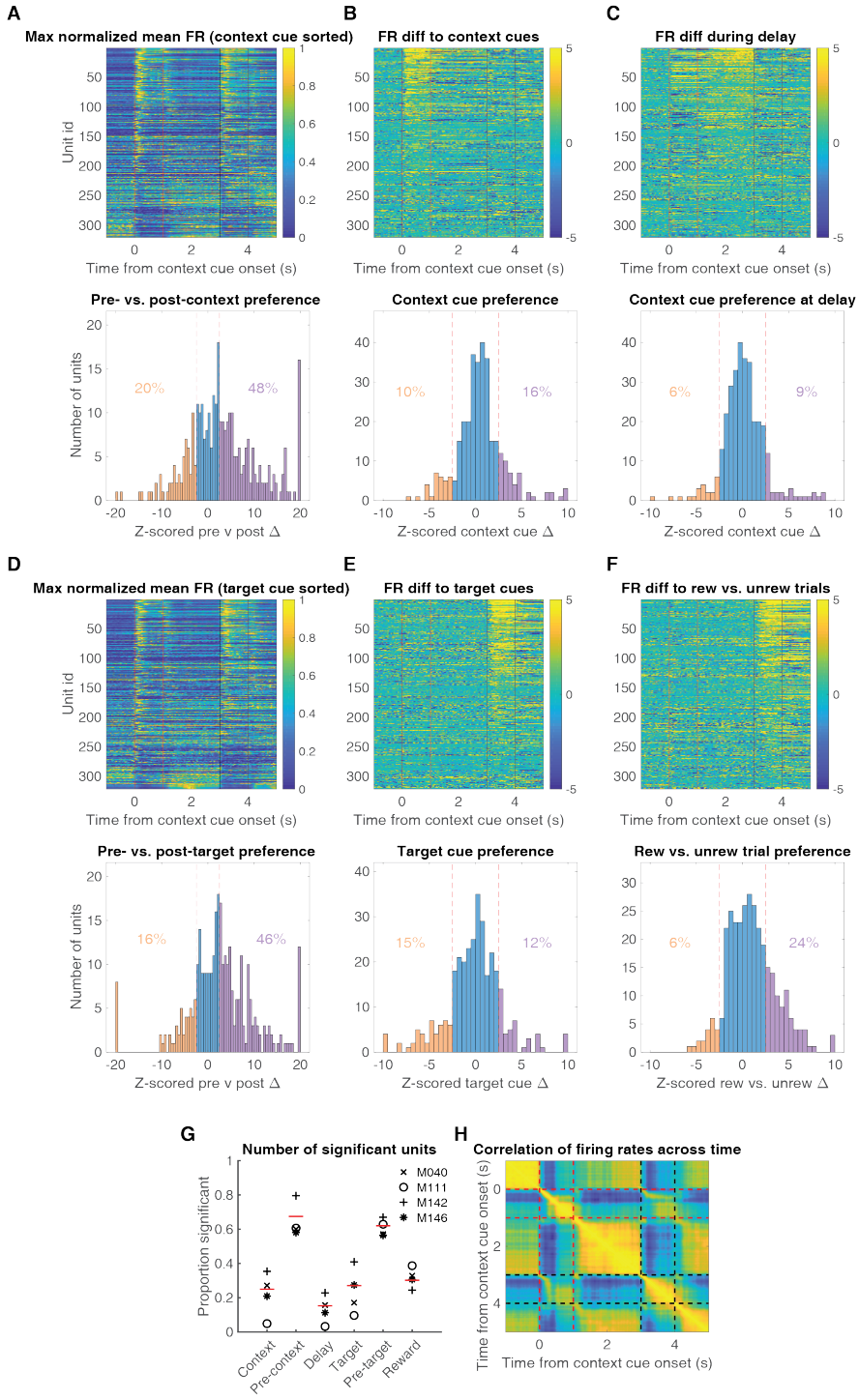

Figure 3: Characterization of single-unit responses to the task. Top of each plot is a heat plot showing either max normalized firing rates or firing rate differences for trial-averaged data for all eligible units, with unit identity sorted according to the peak value for the comparison of interest. Red lines border context cue presentation, and black lines border target cue presentation. Bottom of each plot shows a distribution of units with significant tuning for each task parameter, relative to a shuffled distribution. Red dotted lines signify z-scores of +/- 2.58. **A**: Firing rate profiles for units at 1 s pre- and post-context cue onset, sorted according to maximum value after context cue onset. **B**: Firing rate differences for units across context cues, sorted according to maximum difference during context cue presentation. **C**: Firing rate differences for units across context cues during the delay period, sorted according to maximum difference during the 1 s period preceding target cue presentation. **D**: Firing rate profiles for units at 1 s pre- and post-target cue onset, sorted according to maximum value after target cue onset. **E**: Firing rate differences for units across target cues, sorted according to maximum difference during target cue presentation. **F**: Firing rate differences for units for rewarded and unrewarded trial types during target cue presentation, sorted according to maximum difference during target cue presentation. **G**: Proportion of significant units for each task component. Each mouse is indicated by a symbol, and the average across mice is indicated by the red line. Note the higher context coding for M040 and M142, relative to M111 and M146. Also, note the stronger value coding in M111. General cue, gating, and state value categories represent units that were modulated by a combination of different task components, see methods for classification details. **H**: Correlation of firing rates across time on trial-averaged data across all units. Note that high correlations for periods of time when a cue is present, and the anticorrelations of these cue periods with pre-cue periods. Red lines border context cue presentation, and black lines border target cue presentation.

Across mice, there appeared to be a qualitative relationship between the time spent in training and the strength of context cue coding, with mice that spent more time to acquire the task showing stronger context coding than mice that had a shorter learning curve (Figure 3-supplement 1). For instance, M040 (28 training sessions; 27% context-coding units) and M142 (18 training sessions; 35% context-coding units) showing more context-sensitive units than M111 (9 training sessions; 5% context-coding units) and M146 (7 training sessions; 21% context-coding units). Additionally, M111 which had the least amount of context-coding units, also had the most caudal recording coordinates across mice (Figure 9). Furthermore, this variability across animals was not related to behavioral performance during recording sessions, and a similar relationship was also not seen for target or value coding. Together, this suggests that variability in context coding across mice might be due to differences in training duration or precise recording location.

Multiple signals coexist within the context and delay period

How might value coding be integrated with representations related to the contextual modulation of behavior in the NAc? A common theory of ventral striatal function is that it acts as a switchboard, with corticolimbic inputs gating the efficacy of other inputs (Gruber et al., 2009; Murer & O'Donnell, 2016). For example, studies utilizing in vivo intracellular recordings have found that hippocampal stimulation facilitates subsequent NAc responses to prefrontal cortex (PFC) input, whereas PFC stimulation suppresses the effect of subsequent hippocampal input (Calhoon & O'Donnell, 2013; O'Donnell & Grace, 1995). A possibility is that context-dependent neural activity implements a routing mechanism in the NAc. For instance, depending on the context, which is realized as a distinct network activity pattern in the NAc, a given reward-predictive cue would be associated with a different expected value. Thus, each context-associated network state would gate the flow of subsequent reward-predictive stimuli to generate dynamic value representations, through some previous dopamine-dependent learning mechanism that trains the synaptic weights (Figure 4A).

An alternative possibility for how context may be represented in the NAc is through the framework of rein-

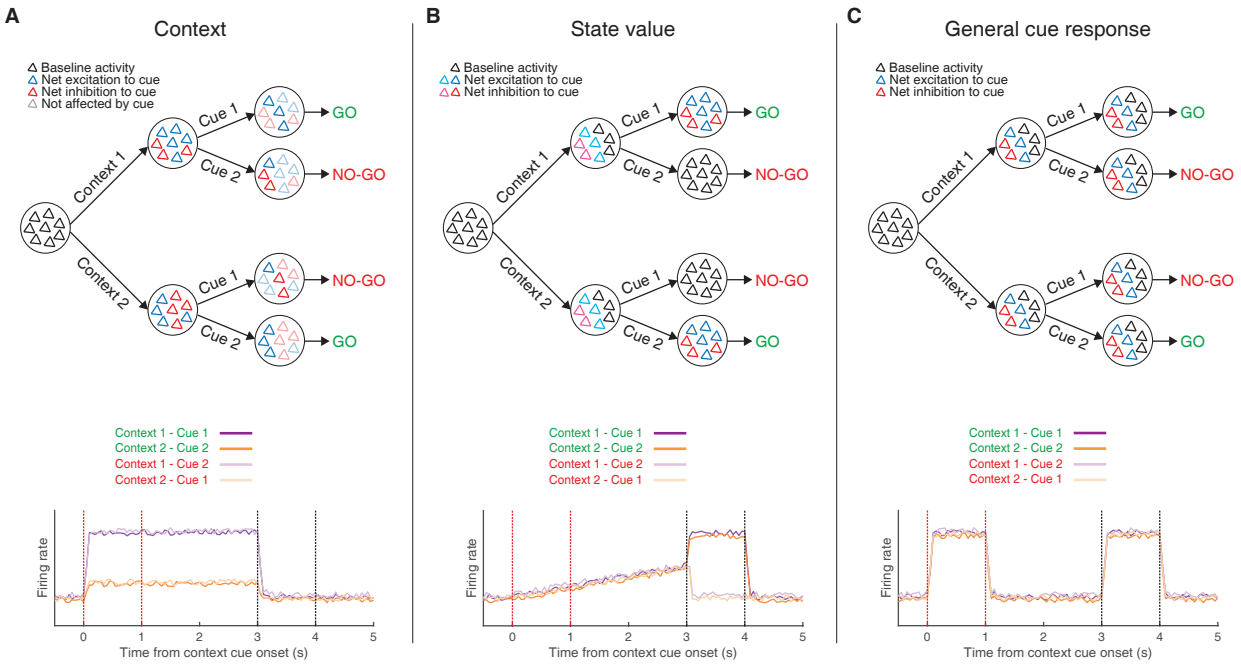

Figure 4: Schematic of hypothetical coding scenarios for context cues. **A**: Context cues function to gate the routing of subsequent target cues. Target cues in this study are reward-predictive cues whose reward-predictive properties are dynamic and not fixed. Top: Schematic of network activity for a pool of neurons in response to a series of motivationally-relevant cues, similar to that presented in 1. In this schematic, a behavioral response to cue 1 is rewarded when preceded by context cue 1 but not by context cue 2, and vice versa for cue 2. After presentation of a context cue, the network shifts to a new activity state, characterized by a change in the firing rates of individual neurons, with each context cue triggering a distinct network state. Furthermore, the difference in the excitability of individual units after presentation of the context cue then determines the receptivity of individual units in the network to subsequent response of the target cue, allowing the generation of dynamic value estimates to facilitate the appropriate Go/No-go response. In this case the network is coding a gating response, that is modulated by the context. Bottom: Hypothetical PETHs for each trial type for a context coding population-level representation. Note, the representation discriminates firing to the two context cues and this difference is sustained until target cue presentation (purple vs. orange lines). Red lines border presentation of the context cue, black lines border presentation of the target cue (see Figure 1 for further details). **B**: Context cues are interpreted in the NAc by their proximity to a rewarded state. Top: Presentation of either context cue elicits a similar network state, that is further amplified upon presentation of the rewarded target cue. This ramp-like activity in the network is coding a state value signal. Bottom: Hypothetical PETHs for a population-level state value coding signal. Note, the peak activity during presentation of the target cue for rewarded trial types (dark colors) but not unrewarded trial types (light colors), and the ramp leading up to this via the context cues. **C**: Context cues are not dissociated from other motivationally-relevant cues in the NAc. Top: Presentation of any of the separate motivationally-relevant cues (context or target) elicits a similar network response in the NAc. The NAc is coding a general cue response. Bottom: Hypothetical PETHs for general cue coding. Note, the example representation responds identically to all cues.

forcement learning. The outcome-predicting signals in the NAc are often interpreted as state value signals, supplying the "expected future reward" part of the equation for the calculation of reward prediction errors (Ito & Doya, 2015; van der Meer & Redish, 2011b). For instance, previous efforts have identified correlates in the NAc such as ramp neurons that increase their firing rate as an animal approaches a reward site, and recent computational work has found success modeling NAc as supplying state values (Stoianov et al. 2018; van der Meer & Redish 2011a). Given that contexts can be thought of as distinct task states, entering a context would modify the current state value relative to the preceding task state, and it is possible that NAc

represents this associated state value, instead of the contextual features themselves. For instance, presentation of a context-setting cue that determines the value of a subsequent cue could be interpreted by the NAc as a state closer to reward, regardless of the identity of the context cue (Figure 4B). Finally, a third possibility is that the NAc may generalize across all motivationally-relevant cues (Figure 4C).

A possible functional role for this context signal is to appropriately gate the response to cues whose relevance is context-dependent. In this case, context-dependent activity preceding target cue presentation should be able to predict the behavioral response for a given target cue. For example, if a unit shows a higher firing rate for context cue O1 over O2, this discrimination would be linked to subsequent behavior if on a trial-by-trial basis it informed whether or not an animal licked in response to O3. In this situation, a licking response to O3 would be predicted on trials where the unit had a higher firing rate preceding target cue presentation. To test for this at the single-unit level, we first reran our firing rate comparisons across the whole trial period, and found that 15-26% of units had firing rates that discriminated between the two context cues during the span of the context cue and delay periods (Figure 5A). We then trained a binomial regression to predict the behavioral response (lick or no lick) for a given target cue using the firing rate of a unit at various time points in a trial. We found that 5-13% of units across time points were able to predict subsequent response to a target cue above chance, suggesting that individual units possess some information about the context that informs future response behavior (Figure 5B).

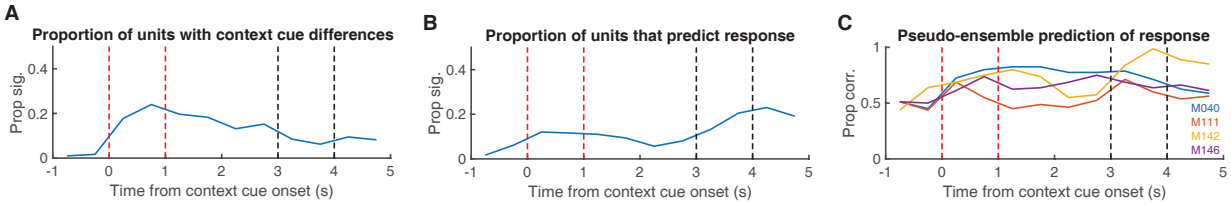

Figure 5: Predicting behavioral response to a given target cue based on firing rate activity. For a given target cue such as O3, this analysis sought to predict whether a lick or no lick response occurred based on activity preceding the target cue during the context and delay period. **A**: Sliding window demonstrating the proportion of units that show discriminatory activity between the two context cues throughout the different periods in the trial, similar to Figure 3B-C. Red lines border context cue presentation, and black lines border target cue presentation. **B**: Proportion of all units whose firing rate at a given point in the trial predicts the behavioral response to a given target cue above chance, according to a binomial regression. **C**: Same as B, but using the firing rates of all units recorded for a mouse to generate a pseudo-ensemble prediction of the behavioral response for a given target cue. Each line denotes the prediction for a various mouse. Note, the variability across mice reflects the variability in single-unit context coding seen in Figure 3G.

While single-unit analyses are informative to get a sense of what information is present within a neural

population, the utility of these responses are dependent upon their position within the broader NAc network, and how they are interpreted by downstream structures. To characterize this population-level activity, we combined across-session data for each mouse to generate 4 pseudo-ensembles, one for each mouse in the dataset. As a first step to test if the population could improve predictions of trial outcome above and beyond that of the best performing single-unit units, we trained a binomial regression to predict the behavioral response for each target cue using the firing rates of these pseudo-ensembles (Figure 5C). This analysis revealed the ability to accurately predict subsequent the behavioral response for a given target cue was above chance for 3 out of the 4 mice (Prediction accuracy at most significant time point, M040: 82%; z-score: 5.82; $p < 0.001$; M111: 69%; z-score: 2.40; $p = 0.016$; M142: 80%; z-score 5.26; $p < 0.001$; M146: 75%; z-score: 4.13; $p < 0.001$). The variability observed across mice closely followed the observations from the single-unit data, with the M111 pseudo-ensemble data being unable to accurately predict above chance the subsequent behavioral response.

To determine which patterns of ensemble activity were driving these behavioral predictions, as well as to test the hypothetical coding scenarios outlined above (Figure 4), we used the dimensionality reduction technique dPCA to extract the task-related latent factors relating to the context cues, target cues, and trial value (Figure 6). dPCA was the method selected as it constrains dimensionality reduction to extract the components that explain the most variance in the data for a given task parameter. dPCA differs from PCA as the latter extracts the components that capture the most variance in the data, agnostic to any aspects of the task. Additionally, dPCA was chosen over LDA, as LDA is focused on reconstructing identities, while dPCA is focused on reconstructing data means, and thus, dPCA is better suited to preserve aspects of the original data. We applied dPCA to the pseudo-ensemble data from each mouse individually, and extracted the top components related to each major task component (see Figure 7 for relative contributions of each component). The strongest component across all animals was a non-specific time-varying signal whose activity and time course was related to the time course of the odor cues, and for this reason we call it the “general cue” signal (Figures 4C, 7A). Another strong non-specific signal was a ramping signal that increased in value from context cue onset to after target cue onset, with a discrimination being observed for rewarded and unrewarded trials for

3 of the 4 mice, consistent with a “state value” signal (Figures 4B, 7B).

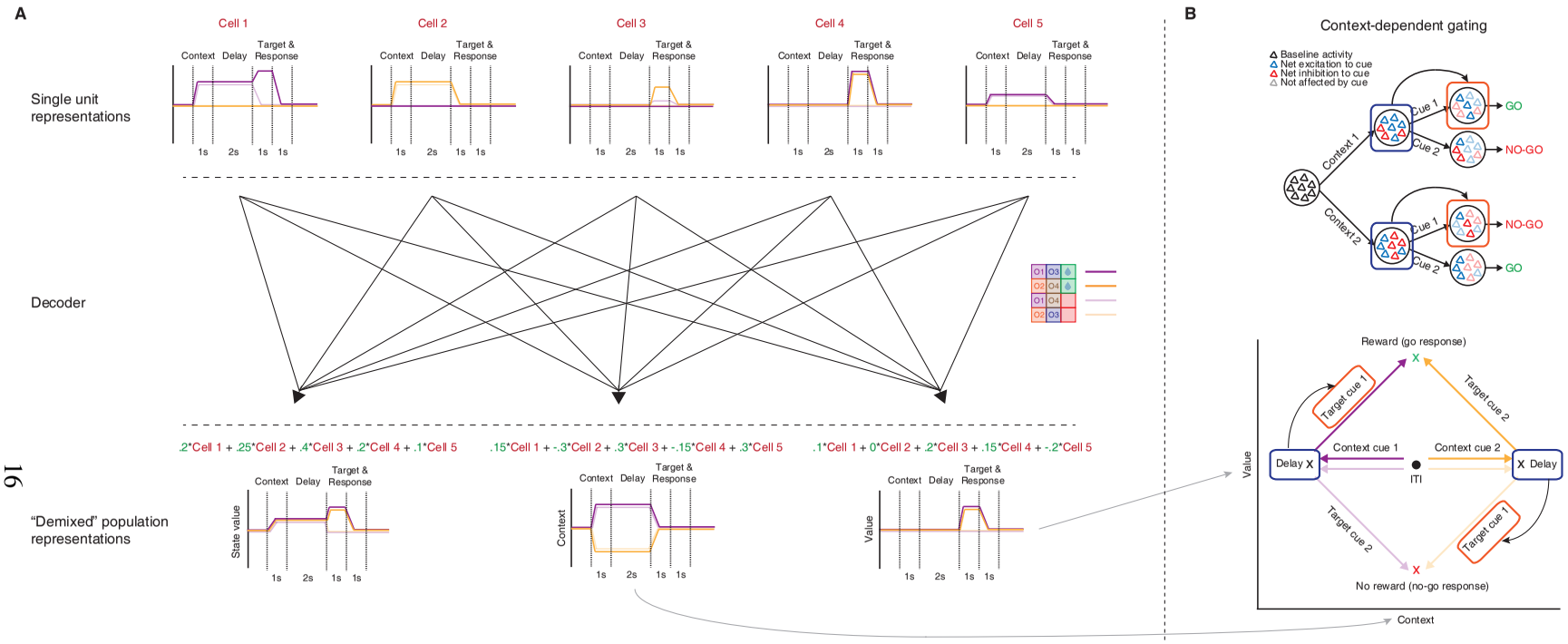

Figure 6: Schematic of workflow for population-level analyses. **A**: To investigate population-level representations of task features, and their relationship to one another, requires a dimensionality reduction technique that can extract latent variables representing individual features of the task, while preserving the structure of the data. In dPCA, pseudo-ensemble activity for each mouse, shown here by five hypothetical single-unit response profiles (top), can be reduced into a few key behaviorally-relevant components (bottom) through a decoder (middle) that seeks to minimize reconstruction error between reconstructed data and task-specific trial-averaged data. Note, there are multiple ways to combine individual units to generate population-level representations. In this example, orthogonal context and value representations are extracted, suggesting that these two patterns of activity occupy separate subspaces in the neural activity space. Green numbers represent the weights of a unit for a given component. **B**: The hypothetical context and value components (bottom) can be used to test the feasibility of the context-dependent gating hypothesis (top). Shown on the bottom is the progression of neural activity through a trial for each trial type in a two-dimensional neural subspace, with the trial-averaged projected activity in the context component on the x-axis, and the trial-averaged projected activity in the value component on the y-axis. If the context-related component brings the network to a distinct state (note the separation along the context-axis from ITI to delay) that modulates the input-output mapping of the subsequent target cue (note the separate paths taken by a target cue in the neural space for each context cue), then a quantifiable relationship should exist between the two components at these timepoints. This can be tested by using linear regression to predict activity (arrow) in the value axis during target cue presentation (red box) from activity in a context axis during the delay period preceding target cue presentation (blue box).

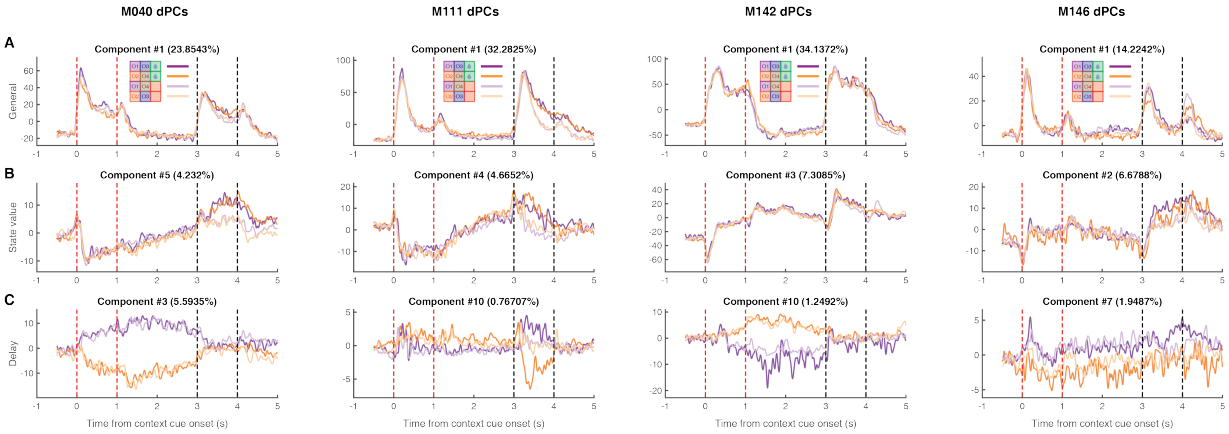

Figure 7: Top extracted components demonstrating the presence of co-existing signals during the context and delay period for each mouse, as outlined in Figure 4. dPCA was applied to the pseudeo-ensemble data from each mouse to extract low-dimensional population representations for various task features. Each plot represents the trial-averaged projected activity onto the component for a given task feature (rows) for each mouse (columns) for each trial type (purple: trials with context cue O1; orange: trials with context cue O2; dark colors: rewarded trials; light colors: unrewarded trials). Plot title denotes the overall ranking of the component, and the amount of variance explained by the component. Red lines border context cue presentation, and black lines border target cue presentation. From left-to-right shows components for M040, M111, M142 and M146. **A**: Top non-specific signal that responded to all odors, and called the general cue component (Figure 4C). Note that this signal is present during presentation of both context and target cues. **B**: The extracted component from each mouse that best represents a state value signal (Figure 4B), with a ramping-like activity after context cue onset, with a separation between rewarded (dark colors) and unrewarded (light colors) trials after target cue onset. Note the variability in this component across mice (greatest rewarded and unrewarded separation for M040 and M111). **C**: The context-related component that best separated context cues during the delay period (Figure 4A). Note that in the mice where this signal is strongest (M040, M142), the strong separation between context O1 (purple) and context O2 (orange) trials from context cue onset until target cue onset.

In terms of components related to the context cues, there are several noteworthy observations. First, there was heterogeneity across mice in terms of the magnitude of the context and delay components (Variance explained for two context-related components, M040: 7.0%; M111: 1.3%; M142: 2.8%; M146: 2.8%), closely following the observations from the single-unit data and pseudo-ensemble predictions. Second, in the mice with the largest degree of separation in context-related neural space, there were two distinct patterns of activity, a signal that was most dominant during the delay period that followed the context cue, the "delay" signal (Figures 4A, 7C), and a signal that followed the time course of the context cue, the "context" component (Figure 7-supplement 1 D). The dot product between delay context-related component and the general cue (M040: 0.07; $p = 0.136$; M111: 0.05; $p = 0.219$; M142: 0.18; $p < 0.001$; M146: 0.24; $p = 0.002$) and state value (M040: -0.07; $p = 0.156$; M111: 0.05; $p = 0.242$; M142: -0.16; $p = 0.001$; M146: 0.45; $p < 0.001$) components did not significantly deviate from zero in some, but not all, mice, suggesting that in some cases these signals are orthogonal and can coexist independently within the same population of NAc units. Finally, there were also clear components related to the identity of the target cue, the "target" component (Figure

7-supplement 1 E), and the behavioral response of the animal, the "value" component (Figure 7-supplement 1 F), which were consistently orthogonal from the delay context-related signals (M040: 0.03; $p = 0.340$; M111: -0.03; $p = 0.340$; M142: 0.04; $p = 0.257$; M146: 0.03; $p = 0.359$). Together, this suggests that all aspects of task-related activity are represented in the population-level activity of the mice, including dominant state value representations that were not generally apparent in the single-unit responses. Furthermore, unlike the single-unit analysis, the presence of distinct population-level representations within a single neural population opens up second order questions that allow the investigation of how individual components are related to each other within the same neuronal population, such as testing for context-dependent gating of value representations, which we do below.

To test which of these population-level representations are responsible for the previously-demonstrated capacity of the pseudo-ensemble activity to accurately predict the behavioral response for a target cue, we ran the same binomial regression, predicting lick or no lick for a given target cue, using the pseudo-ensemble activity from each mouse projected onto the extracted population-level components for each task feature (Figure 7-supplement 2). Unsurprisingly, only the context-related components were able to predict information about the animal's upcoming response to each target cue during the delay period, and the strength of this prediction was related to how strong the component was in a given mouse (Figure 8B for M040 & supplements for M111, M142, M146; Prediction accuracy at end of delay period, M040: 99%; z-score: 3.06; $p = 0.002$; M111: 62%; z-score: 1.69; $p = 0.091$; M142: 92%; z-score 3.98; $p < 0.001$; M146: 83%; z-score: 3.38; $p < 0.001$). In fact, in the two mice that showed the largest separation in the delay context-component, the predictions using this component alone surpassed that of the entire pseudo-ensemble, suggesting that this context-related activity is related to the animal's subsequent response.

Context-specific ensemble states predict the magnitude of the subsequent value response

The findings of clear context-related components that hold information about the trial context during the delay period until presentation of the target cue, and the ability of these components to predict the animals subsequent response, raises the possibility that this activity might be able to predict the behavioral response via gating value-related activity to the target cue. Specifically, if each context cue brings the NAc to a unique network state that possesses distinct input-output transformations for a given target cue to enable the generation of context-appropriate value representations (Figure 4A), then a biomarker for this relationship should be observed in the linking between context-related and value-related population representations. For instance, if the transformation of target cue O3 into a value representation is dependent upon whether the NAc is in the network state related to context cue O1 or context cue O2, then context-related activity during the context period should be informative of subsequent value-related activity to the target cue (See Figure 6B for schematic; Figure 8A & supplements for data trajectories across mice). If, on the other hand, the context-related network state is not relevant for the transformation of the target cue into a value representation, then there should be no relationship between the context-related and value-related components. To test this, we trained a linear regression to predict the activity of the top value-related component during target cue presentation, from the delay context-related component, and found that the delay component could account for a significant amount of variability for the value-related component for a given target cue during the delay period (Figure 8C for M040 & supplements for M111, M142, M146; Proportion of variance explained at end of delay period, M040: 0.45; z-score: 63.16; $p < 0.001$; M111: 0.15; z-score: 7.25; $p < 0.001$; M142: 0.42; z-score: 10.92; $p < 0.001$; M146: 0.22; z-score: 8.89; $p < 0.001$). As a control, we also tried to predict variability in the value-component to a given context cue, but were generally unable to (Figure 8D for M040 & supplements for M111, M142, M146; Proportion of variance explained at end of delay period, M040: 0.09; z-score: 0.08; $p = 0.936$; M111: 0.13; z-score: 1.59; $p = 0.112$; M142: 0.21; z-score: -1.56; $p = 0.119$; M146: 0.11; z-score: 0.14; $p = 0.889$). This suggests that the ability of the delay context-related component to predict variability in the value-related component was specific to the delay component and not a general feature of the population-level representations. Furthermore, to assess whether

this effect was driven by a few highly contributing units, we repeated these predictions while iteratively removing the top 10% of contributors to the component and found that this prediction persisted even after removing the top ∼20% of units (Figure 8E for M040 & supplements for M111, M142, M146). Together, these findings suggest that the NAc ensemble encodes a context-related component during the delay period, and that this activity is linked to subsequent value-coding during the target cue period in a way that supports the context-dependent gating account of context coding. Importantly, this gating feature was unique to the population-level representations, and was not apparent from analysis of single-unit activity.

Discussion

A dominant view of NAc function and what is encoded in its activity is that it tracks fundamentally value-centric quantities (Floresco, 2015; Gruber & McDonald, 2012; Koob & Volkow, 2016; Nicola, 2010; Salamone & Correa, 2012). Here, we present a number of findings that demonstrate this view is too narrow, which we discuss in the following paragraphs. We have shown that both NAc single-unit and population-level representations distinguished between the two context cues in the experiment, even though they did not differ in expected value. Importantly, at both the single-unit and population-level, this coding persisted throughout the delay period when the animal must maintain a representation of this cue to inform subsequent behavior to the target cue. Additionally, at both the single-unit and population-level, activity during the context and delay period could predict the subsequent behavioral response for a given target cue. At the population-level, patterns of activity for different coding scenarios during the context and delay period were found. Furthermore, context-related activity during the context and delay period could be extracted independently from value-related activity after target cue onset. Despite this independence, these context components could explain a significant proportion of variance in the value components in response to a given target cue.

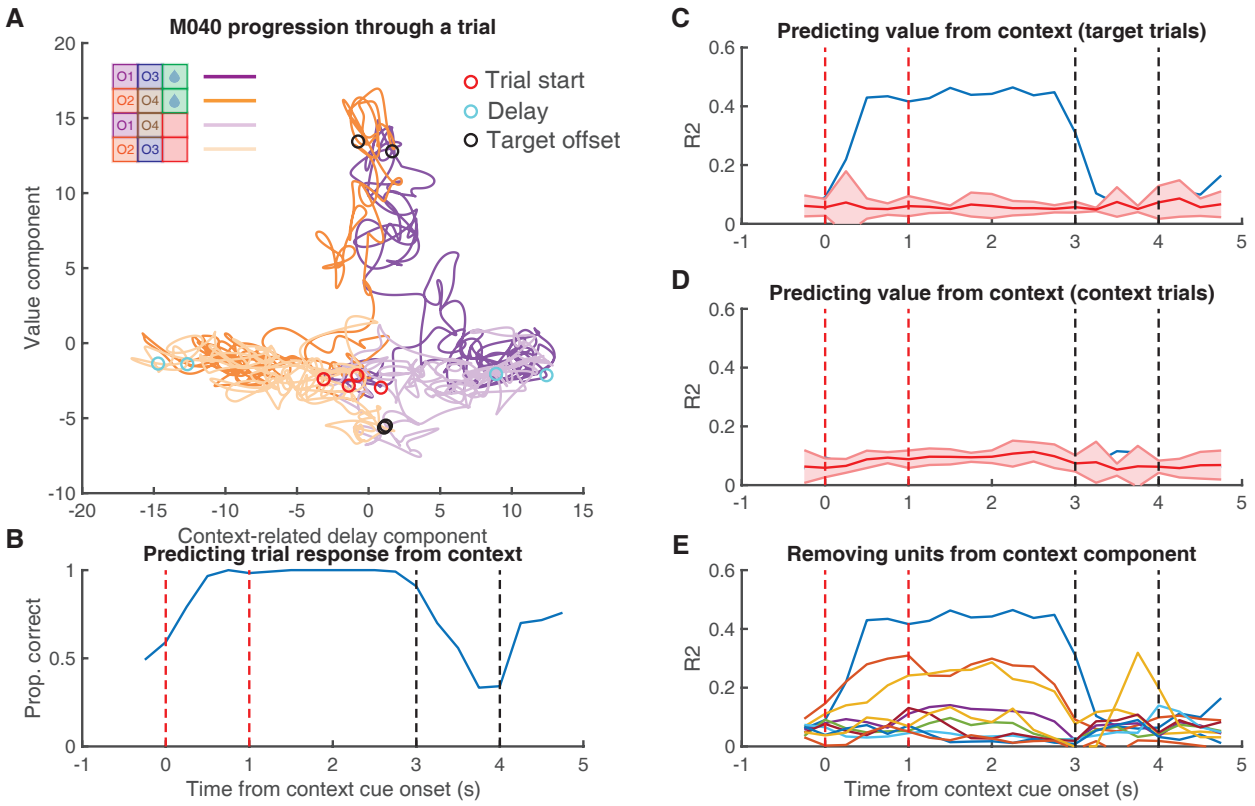

Figure 8: Context-dependent gating of the value-related component, as outlined in Figure 6B. Shown in this Figure is the relationship between the context-related delay component and the value-related component for M040, data for M111, M142 and M146 can be found in the supplements to this Figure. **A**: Progression of neural activity through a trial for each trial type (purple: trials with context cue O1; orange: trials with context cue O2; dark colors: rewarded trials; light colors: unrewarded trials) in a two-dimensional neural subspace, with the trial-averaged projected activity in the context-related delay component (see Figure 7C) on the x-axis, and the trial-averaged projected activity in the value component (see Figure 7-supplement 1 F) on the y-axis. During the context and delay periods a separation is observed along the context axis between context O1 and context O2 trials, after which separation is observed along the value axis following target cue presentation for rewarded versus unrewarded trials. Red circles signal context-cue onset, cyan circles signal delay period 1 s after context cue offset, black circles signal 1 s after target cue onset. **B**: Using a binomial regression to predict the behavioral response (lick or no lick) for a given target cue based on projected activity along the context-related delay component at various timepoints, showing the high accuracy during the context and delay periods. Red lines border context cue presentation, and black lines border target cue presentation. **C**: Using a linear regression to predict projected activity along the value-related axis after target cue onset for a given target cue (black circles from A) based on projected activity in the context-related axis at various timepoints, showing performance above chance levels during the context and delay period. **D**: Control analysis showing the inability of using a linear regression to predict projected activity along the value-related axis after target cue onset for a given context cue based on projected activity in the context-related axis. **E**: Iteratively removing the top 10% of contributors to the context-related delay component and repeating the linear regression-based analysis of predicting value-related activity as in C, showing the ability to achieve above chance perform even after removing the top 20% of single-unit contributors to the context-related delay component.

Context coding in the NAc

To our knowledge, this study is the first to show in rodents that NAc units discriminate between context cues that are not directly tied to reward (26% of units across mice), but instead set the expected value of subsequently presented reward-predictive cues. This finding of context coding is not likely due to unequal

cue salience across context cues as we counterbalanced the odor associations across mice, and observed behavioral performance to be similar across cues for each animal. The present study expands upon our previous work that demonstrated NAc units that distinguished between sets of motivationally-relevant stimuli that had equal outcome predictive properties (Gmaz et al., 2018). Interestingly, a subset of these stimulus set-discriminating units also showed sustained changes in firing during trial periods before the presentation of the cue, suggesting that they encoded an abstract task feature not directly tied to stimulus presentation. However, given that the outcome-predictive properties of the motivationally-relevant stimuli were fixed, and the sets of stimuli were presented in separate continuous blocks, we were unable to further characterize this signal. The present experiment addresses these prior limitations by being the first NAc recording study to present both distinct, temporally precise context cues, and cues with dynamic outcome-predictive properties, demonstrating that the NAc codes information about the context that continues into the delay period. Additionally, our work is comparable to work by Sleezer et al. (2016) that found evidence for rule encoding in the primate NAc, suggesting that this rule coding might be part of a more general NAc computation that primes the network to distinct states to enable behavioral flexibility. Together, our study is the first to demonstrate the presence of context-related correlates in the rodent NAc, providing further support for the burgeoning recognition of the NAc in decision-making outside of pure value processing.

While we use the term "context" throughout the text to refer to the first cues presented during a trial, we acknowledge that since we only used one cue per context we cannot rule out the possibility that these units encoded cue identity as opposed to context. However, even if NAc encoded context cue identity as opposed to context, it would not affect our interpretation of the data through the gating model (see below), as NAc units just need to be in distinct network states during these cues for it to be feasible. Additionally, while we believe that the 2 s of active flushing of the odorant during the delay between cues is sufficient for the context cue odor to disperse from the experimental apparatus, we cannot exclude the possibility that a proportion of the odor remained and combined with the target cue to form a compound cue. However, we think this is unlikely given the presence of single-unit and population-level correlates that only discriminated between context cues during the delay period following context cue offset, as well as the correlates that discriminated

between the target cues, regardless of which context cue was previously presented. Finally, given that the value of a trial and the behavior of the animal were yoked in our task, that is rewarded trials were followed by licking behavior, and unrewarded trials were followed by the absence of licking, we cannot fully separate contributions of value and behavior to our value-related components. However, this does not interfere with the interpretation of our primary finding of context coding, or the linking of context-related activity during the delay period with subsequent, behaviorally-relevant activity during the target cue period.

Although each individual animal had individual units that showed context coding, there was also heterogeneity in the degree of context coding across mice. The clearest relationship between this heterogeneity and any other component of the experiment was in the training duration for the animals, with mice that took less days to reach performance criterion on the task before recording showing less separation in the population-level context component, while those that took more days to reach criterion had a larger separation in the context component. Given the small sample size in each group (n = 2), it is hard to draw any substantial inferences from this data. However, speculatively this finding may suggest that during initial learning, the NAc receives cue information and processes it solely in terms of its motivational relevance, and some other structure is supporting context-dependent behavior, but then after extensive learning, it forms representations of the associative structure. Interestingly, recent work has suggested that motivated approach behavior becomes less NAc-dependent as training progresses (Dobrovitsky et al., 2019). Whether or not this is related to a shift in the role of the NAc as a behavior becomes learned, or related to distinct inputs such as from the hippocampus and cortex, remains to be determined. An additional potential contributor to the variability across mice is recording location. The mouse that contained the lowest number of context-related single-unit responses was also the mouse whose recording coordinates spanned the most caudal aspect of the NAc. Several lines of investigation suggest a heterogeneity in NAc processing along the rostral-caudal gradient (Gill & Grace, 2011; Groenewegen et al., 1982; Ma et al., 2020; Reynolds & Berridge, 2008), largely due to a different distribution of inputs, suggesting that a portion of the observed variability in context coding may be due to differences in recording location.

Relationship between context and value coding in the NAc

In addition to showing context coding using a biconditional task design specifically designed to control for value, a further innovation in this study is the population-level analysis, which allowed us to show evidence for co-existing activity patterns in the population-level representations, as well as a functional link between context coding and subsequent value coding. To determine whether our behavioral predictive power was arbitrary to any population-level component, we ran the binomial regression on all major components, and found that only those containing significant context information had predictive utility during the delay period. These population-level results align with a growing body of work advocating for population-level interpretations of neural data, suggesting that certain neural computations are better understood in terms of their population-level versus single-unit output (Ebitz & Hayden, 2021; Gallego et al., 2017; Saxena & Cunningham, 2019; Vyas et al., 2020). A primary argument for this approach is that there is a high degree of correlation and redundancy across single-unit coding, suggesting that the large neural space occupied by units in a region can be captured by a drastically lower dimensional latent space. Furthermore, the utility of single-unit output is ultimately determined by how it is integrated with other inputs by a receiver network. A proxy for this integration can be assessed by investigating the individual unit weightings of components extracted from dimensionality reduction techniques, as the weights for a component hypothetically represent how units are integrated for a particular output signal. We next discuss interpretations of our findings through this population-level framework below.

The finding of a clear non-value signal supports other work that the NAc is coding for more than a low-dimensional value signal (Gmaz et al., 2018; Sleezer et al., 2016). A potential function for the context coding observed in the present study is implementation of the hypothetical switchboard function of the NAc (Gruber et al., 2009; Murer & O'Donnell, 2016), serving as a routing mechanism to enable dynamic value representations of target cues (gating; Figure 6B). Indeed, the distinct occupancy in the pseudo-ensemble space for each context cue signals that the context cues might be driving the NAc into separate network states, setting an initial state for subsequent input-output flow of the target cue. In addition to the presence

of a context signal, this routing function would also require that the context signal is both functionally linked to the subsequent value signal, while simultaneously not interfering with value-related output. We found support for the former from the observation that activity in this space during the delay period could explain a significant proportion of variance in the value-related component during the target cue period. Similar observations have been observed in the field of motor control and, more recently, economic choice (Elsayed et al., 2016; Kaufman et al., 2014; Yoo & Hayden, 2019). Furthermore, given that the context-related and value-related components were orthogonal from one another, it suggests that context coding does not interfere with the ability of downstream structures to read-out value-related information. Finally, another potential functional role for the observed context coding is in forming the associations between reward-predictive cues and the rewards themselves. Recently, several studies have implicated the cortical-striatal access in credit assignment (Oemisch et al., 2019; Parker et al., 2019), raising the possibility that this context may be used for learning to assign credit to the appropriate state. Whether these context components are generated in the NAc or inherited from inputs, as well as if they represent an internal computation that is locally used to organize NAc activity, or are conveyed downstream, remains to be determined.

Beyond context-dependent gating interpretations of the data, we also found support for reinforcement learning-inspired accounts of NAc function (state value; Figure 4B; Averbeck & Costa 2017; van der Meer & Redish 2011b). In 3 out of 4 mice, a signal that showed both ramping after context cue onset as well a discrimination between rewarded and unrewarded trial types at target cue onset was present, with dynamics that closely mimic what would be expected from a signal conveying state value. Interestingly, this signal co-existed within the same population of units as the context signal, and was orthogonal from the delay context-related component in the 2 mice with the clearest rewarded versus unrewarded discrimination, suggesting that the NAc can process both types of information. This signal is similar to ramping signals observed previously in single-unit studies (van der Meer & Redish, 2011a), and may be the result of the strong hippocampal input to the NAc. Future experiments inactivating hippocampal drive should test the relationship, as well as the necessity of this signal for proper evaluation of outcome-predictive cues.

Interestingly, across all the animals, the strongest extracted component was a general cue signal that signalled the onset and duration of all cues used in this study (general cue; Figure 4C). These condition-independent signals are being found across various domains of systems neuroscience (Kaufman et al., 2016; Raposo et al., 2014; Thura et al., 2020), and their strikingly relative dominance to other task-related signals suggest that they signal general task-related transitions in neural state space, although their exact role is unknown. Given the NAc is part of a broader limbic network that is entrained by respiration, and has strong connections with olfactory-processing regions, it is possible this component is signalling to the NAc the presence of a salient event, and priming it to process the associative content of the cues, perhaps by increasing the excitability of the NAc and opening the aforementioned gate. Regardless of the precise functional relevance of the observed correlates in this study, the finding of clear non-value correlates suggests a revision of the value-centric account of neural activity, encouraging future work to view NAc activity as a richer signal containing more than just reward.

Methods

Subjects

A total of 4 adult female wild-type C57BL-6J mice (Jackson Labs) were used as subjects (data from a 5th mouse was collected, but was not analyzed due to poor behavior). Mice were group housed before being selected for the experiment with a 12/12 hour light-dark cycle, were individually housed once training commenced, and tested during the light cycle. Mice were food restricted to 85 - 90% of their free feeding weight (weight range at start of experiment was 19.7 - 23.8 g), and water restricted for a minimum of 6 hours before testing. All experimental procedures were approved by the Dartmouth College Institutional Animal Care and Use Committee (IACUC).

Overall timeline

Mice initially underwent surgery where craniotomies were marked and a headbar was affixed to the skull. After a 3 day recovery period, mice were food restricted, and acclimated to being handled by the experimenter and being held by the headbar in the experimental room for 3 days. Mice were then habituated to being head-fixed on the apparatus over the course of 3 or more days, starting with 5 minutes and working up to an hour of being head-fixed. During later head-bar habituation sessions, animals were placed on water restriction, and trained to lick a spout for 12% sucrose solution. After learning to lick for sucrose (1 - 2 sessions), mice were then trained to lick in response to the rewarded odors in the task for 1 - 3 days before undergoing full task training. Once behavioral criterion was reached in the full task (6+ days; described below), the first craniotomy was made, and acute recordings commenced after a 24 hour recovery period. Recording sessions were carried out for 5 - 7 days, after which a contralateral craniotomy was made and the process repeated. After sufficient data collection, mice were euthanized and histology was performed to confirm recording sites.

Behavioral task and training

In order to assess whether the NAc codes for information related to context cues, mice were trained to perform a biconditional discrimination task where they were presented with two odors in sequence. The identity of the first odor, the context cue, determined the value of the subsequently presented odor, the target cue, such that each context cue had a rewarded and unrewarded target cue pairing (Figure 1A,B; adapted from Han et al. 2018). The apparatus was a custom built head-fixed mouse behavioral setup, consisting of a running wheel, odor port, lickometer, and headbar holders (Grasshopper Machinewerks LLC). Pressurized air passed through an olfactory delivery system containing 5 distinct tubes, with 1 tube containing mineral oil, and the rest a mixture of mineral oil and a specific odorant. Each tube was connected to an experimentally-controlled valve that then sent the air-odorant mixture to the odor manifold on the head-fixed setup, where

the active line was sent to the mouse. Apart from odorant presentation, mice were continuously presented with unscented air via the mineral oil only line. Odorants used in the study were propyl formate, 1-butanol, propyl acetate and 3-methyl-2-buten-1-ol (Sigma). Odorants were selected based on previous work using this task (Gu & Li, 2019; Han et al., 2018; Zhang et al., 2019). The lickometer detected changes in capacitance from mouse licks, and sent this information to a Digital Lynx acquisition system (Neuralynx). The task was controlled via a custom-written MATLAB script (Mathworks) that triggered TTL pulses from the acquisition system to control the odorant and sucrose valves.

After initial handling and habituation mice were first trained to lick a spout for 12% sucrose solution, until they manually triggered over 100 sucrose water rewards in $<$ 20 mins. Mice were then shaped to lick in response to pseudo-randomized presentation of the 2 rewarded context-target odor pairs, with the context odor determining the outcome predicted by the subsequent target odor. Odor selection and pairing was pseudo-randomized across mice to ensure unique pairings across animals. A single trial consisted of a 1 s presentation of the context odor, followed by a 2 s delay where unscented air was presented to flush out the odorant, followed by a 1 s presentation of the target odor, followed by an additional 1 s response window, followed by a 12 +/- 2 s inter-trial interval (Figure 1A). Licking either during presentation of the target odor or the subsequent response window registered as a correct response. During the shaping phase, sucrose water was delivered pseudo-randomly in 1/3 of the trials in which mice failed to lick. Mice were allowed to complete up to 200 trials in a session, with an individual session being terminated either at 200 trials or if the mouse became sufficiently disengaged by the task, measured by the absence of licking for 10 consecutive trials. This phase of training continued until the mouse licked for $\sim$80% of trials. After the shaping phase, mice then underwent the full task training, where they were presented in pseudo-randomized sequence all 4 context-target odor pairs. Upon reaching criterion of 3 consecutive sessions with $>$80% correct responses (range: 7 - 28 sessions), a craniotomy was made over the first hemisphere, and recordings began after a recovery period.

Surgery

Mice underwent 3 surgeries over the course of the experiment. The first surgery consisted of exposing the skull, marking the location of future craniotomies, and securing the headbar. The second surgery consisted of making the first craniotomy, and installing a posterior reference wire above the cerebellum. The third surgery consisted of making the second craniotomy. In all surgeries, mice were anesthetized with isoflurane, induced with 5% in medical grade oxygen and maintained at 2% throughout the surgery ($\sim$0.8 L/min), and were administered ketoprofen as an analgesic prior to surgery, with a supplementary dose 24 hours after the procedure.

Data acquisition and preprocessing

For recording sessions, 32 (NeuroNexus; A4x2-tet) or 64 (Cambridge NeuroTech; P-1) channel silicon probes were lowered into the NAc (AP: 0.8 - 1.4 mm; ML: +/- 1.0 - 2.0 mm; DV: 4.0+ mm). After letting the probes settle for 30 minutes, single-unit activity was recorded during behavioral performance. NAc signals were acquired using a Digital Lynx data acquisition system with an HS-36 PTB preamplifier (Neuralynx). Putative spikes were recorded as threshold crossings of 600 - 6000 Hz band-pass filtered data with waveforms sampled at 30 kHz. Signals were referenced locally to maximize signal-to-noise of the spiking waveforms. Spike waveforms were clustered with KlustaKwik using energy and the first derivative of energy as features, and subsequently manually sorted using MClust (MClust 3.5, A.D. Redish et al.). Isolated units containing less than 200 spikes during the trial period were excluded from analysis.

Data analysis

Behavior

If mice learned the appropriate associations between context and target cues, then correct behavior on the task would look like a high licking response rate to context-target pairings that are rewarded, and a low licking response rate to context-target pairings that are unrewarded (see Figure 1B for hypothetical learned trial structure). To assess whether mice learned to discriminate between rewarded and unrewarded odor pairs we compared the mean proportion of rewarded and unrewarded trials that the animal made a lick response for a given odor, relative to shuffling the trial type label for the mean proportion of trials licked for a given session. Furthermore, to assess whether mice were not responding differently to individual target cues we also compared the mean proportion of trials with a lick for each target cue, relative to shuffling target cue identity for the mean proportion of trials licked for a given session.

Single-unit coding

To address our question of how the NAc responds to context cues and their relationship to target cues, we compared the mean firing rates for different trial types at different time points for each unit. To determine whether or not a unit responded at all to any context cue, we compared the 1 s pre and 1 s post context cue onset period. To determine whether or not a unit discriminated between the context cues, we compared the mean firing rate for the 1 s presentation of each context cue. To determine whether or not a unit discriminated between the context cues during the delay period, we compared the mean firing rate for each context cue for the 1 s period preceding target cue presentation. Likewise, a similar comparison was performed for general target cue responsiveness (1 s pre vs post target cue onset), target cue selectivity (target cue 1 versus target cue 2 during 1 s target cue presentation period), and outcome-predictive selectivity (rewarded versus

unrewarded trials during 1 s target cue presentation period). All firing rate comparisons were related to a shuffled distribution where the trial identity was shuffled across trials. Overlapping proportions were determined to be significant if they were larger than shuffling the identity of significant units for each task parameter. For all analyses, a value of +/- 2.58 z-scores from the shuffled distribution was used as the threshold value for significance. All analyses were completed in MATLAB 2018a. For single-unit examples, peri-event time histograms (PETHs) were generated by smoothing the trial-averaged data with a Gaussian kernel (σ: 100 ms).

To determine how context is signaled throughout the period between context-cue onset and target-cue onset, we calculated the proportion of units that showed discrimination in firing to the context cues at each time point in the trial in 0.5 s intervals. To determine whether this coding had any utility in informing future behavior we used a binomial regression to predict the animal's behavioral response for a given target based on the firing rate for a unit.

Population-level coding

To assess population level predictions of behavior, we generated pseudo-ensembles for each mouse from the data recorded across sessions. We then trained a binomial regression to predict the behavioral response for each target cue using the firing rates of the pseudo-ensembles. To determine how context cues are represented at the network level, we performed demixed principal component analysis (dPCA) to extract information related to the context and target cue from the population of recorded units (see Kobak et al. 2016 for a detailed description of the methodology). dPCA is an analysis that aims to explain most of the variance in the data as in principal component analysis (PCA), while also separating the data for several task parameters similar to how linear discriminant analysis (LDA) does for a single task parameter (Figure 6A). The dPCA method first took the mean-subtracted, trial-averaged data for all units, and decomposed the population data matrix into a sum of separate data matrices that each represented the contributions of a different aspect of the task,

and noise. These task features are inputs to the analysis set by the experimenter, and in the current experiment the task inputs were context cue type, target cue type, and the interaction between context and target cue signifying cue value. The loss function of dPCA then used the ordinary least squares solution to find the transformation that minimized the reconstruction error between the reconstructed data and the deconstructed data, with the deconstructed data matrix representing the contributions of a given task-parameter to the full trial-averaged data. Dimensionality reduction was then achieved via eigendecomposition of the covariance matrix of the transformed data, and the top components were stored. The explained variance of each component was the fraction of the total variance in the trial-averaged data that could be attributed by the reconstructed data for that component.

In our task, we sought to 'demix' the contributions of the context cue, target cue, and cue value across time, projecting the data using components derived from these task variables, and visualized how the projected neural trajectories evolved throughout a trial in this reduced dPCA space (see Figure 6B for hypothetical trajectories along context and outcome axes). This analysis requires a sample size of 100 neurons to achieve satisfactory demixing, and thus sessions within a mouse were pooled together to run on pseudo-ensembles. Furthermore, given that dPCA does not constrain the components extracted for each parameter to be orthogonal, components were identified as being non-orthogonal if the dot product significantly deviated from zero.

If activity in response to the context cue was indeed constraining subsequent information flow in response to the target cue, then we would expect to be able to predict both behavioral and neural features during the target cue epoch, based on the assumption that a given target cue would possess distinct input-output mappings for each context cue. First, to determine the contributions of the extracted components in predicting subsequent behavioral response to a target cue, we used binomial regressions to predict behavioral response from the projected activity using each component. Next, to more directly test the feasibility of our hypothesis of context-dependent gating of value representations, we used a linear regression to predict the projected activity in the top value component during a particular target cue from the projected activity in the delay component

across the two contexts. As a control, we also attempted to predict activity across target cues for the same context cue. Additionally, to determine whether any effects were driven by a few top contributors to these components, we repeated this analysis while iteratively removing the top 10% of units that had the largest weights.

Histology

Immediately following the final recording session, mice were asphyxiated with carbon dioxide, and transcardial perfusions were performed. Brains were fixed and removed, and then sectioned in 50 mm coronal sections. Sections were then stained with thionin, and visualized under light microscopy to determine probe placement (Figure 9).

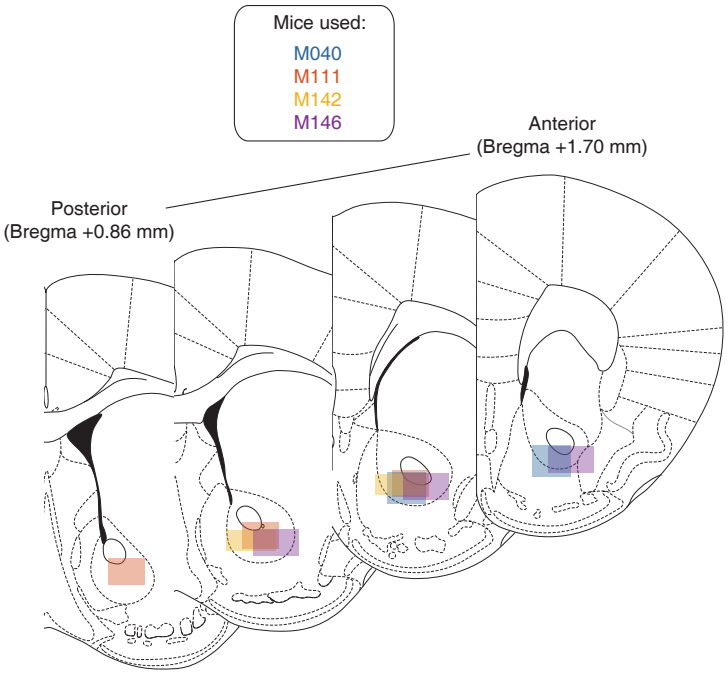

Figure 9: Histology. After the recordings were completed, brains were sectioned and probe placement was identified. Schematic showing recording areas for all subjects.

548 # Figure supplements

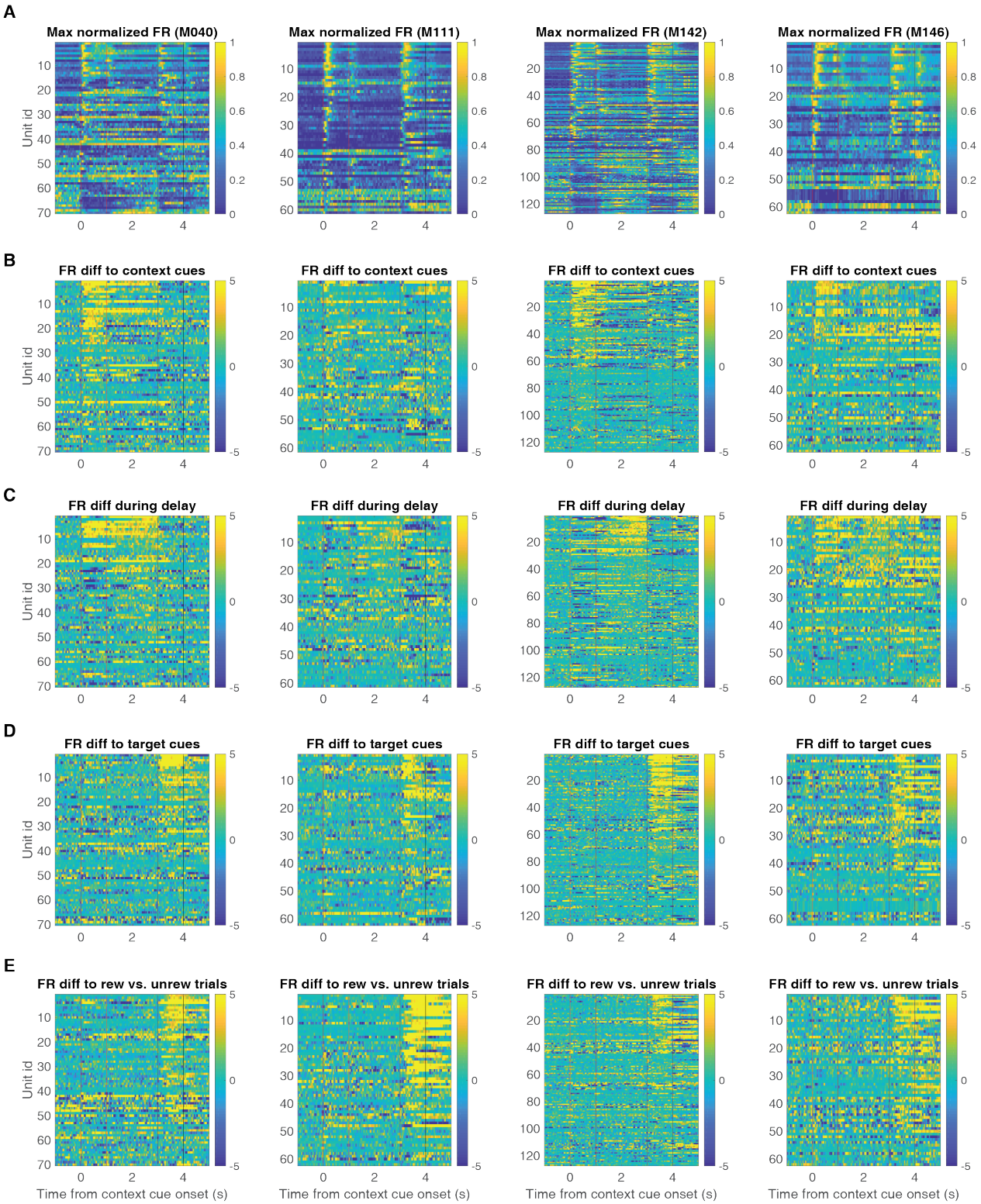

Figure 3 supplement 1: Characterization of single-unit responses to the task for each individual mouse. Each plot is a heat plot showing either max normalized firing rates or firing rate differences for trial-averaged data for all eligible units, with unit identity sorted according to the peak value for the comparison of interest. From left-to-right shows data for M040, M111, M142 and M146. Red lines border context cue presentation, and black lines border target cue presentation. **A**: Firing rate profiles for units at 1 s pre- and post-context cue onset, sorted according to maximum value after context cue onset. **B**: Firing rate differences for units across context cues, sorted according to maximum difference during context cue presentation. **C**: Firing rate differences for units across context cues during the delay period, sorted according to maximum difference during the 1 s period preceding target cue presentation. **D**: Firing rate differences for units across target cues, sorted according to maximum difference during target cue presentation. **E**: Firing rate differences for units for rewarded and unrewarded trial types during target cue presentation, sorted according to maximum difference during target cue presentation.

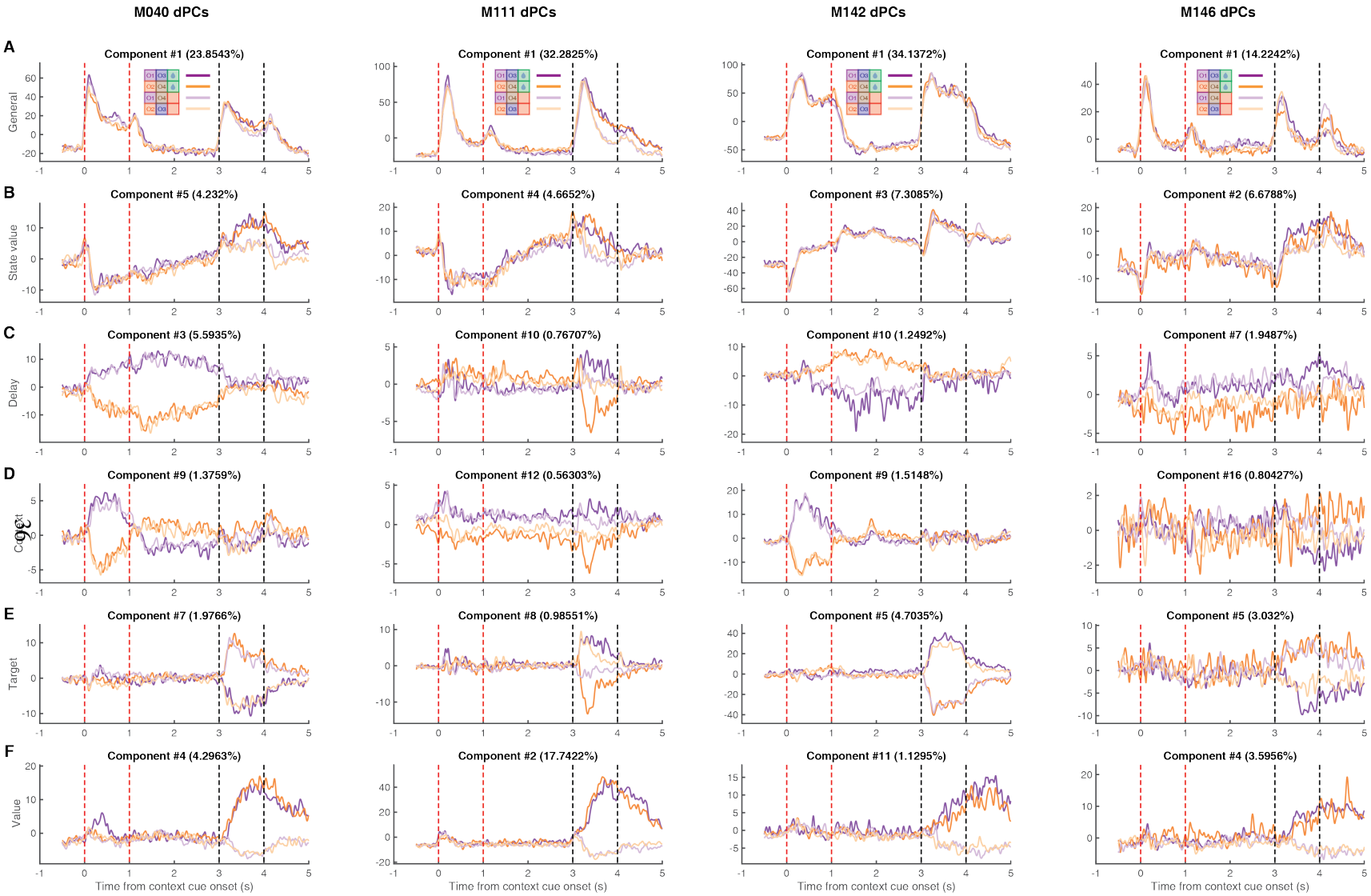

Figure 7 supplement 1: Top behaviorally-relevant components extracted for each mouse. Each plot represents the trial-averaged projected activity onto the component for a given task feature (rows) for each mouse (columns) for each trial type. Plot title denotes the overall ranking of the component, and the amount of variance explained by the component. Red lines border context cue presentation, and black lines border target cue presentation. From left-to-right shows components for M040, M111, M142 and M146. **A**: Top condition-invariant signal that responded to all odors, and called the general cue component. **B**: Another condition-invariant component present in most mice that showed a ramping-like activity after context cue onset, with a separation between rewarded and unrewarded trials after target cue onset. Termed the state value component. **C**: The context-related component that best separated context cues during the delay period. **D**: The context-related component that best separated context cues during cue presentation. **E**: The top target-related component that separated between target cues during target cue presentation. **F**: The top value-related component that separated rewarded and unrewarded trials during target cue presentation.

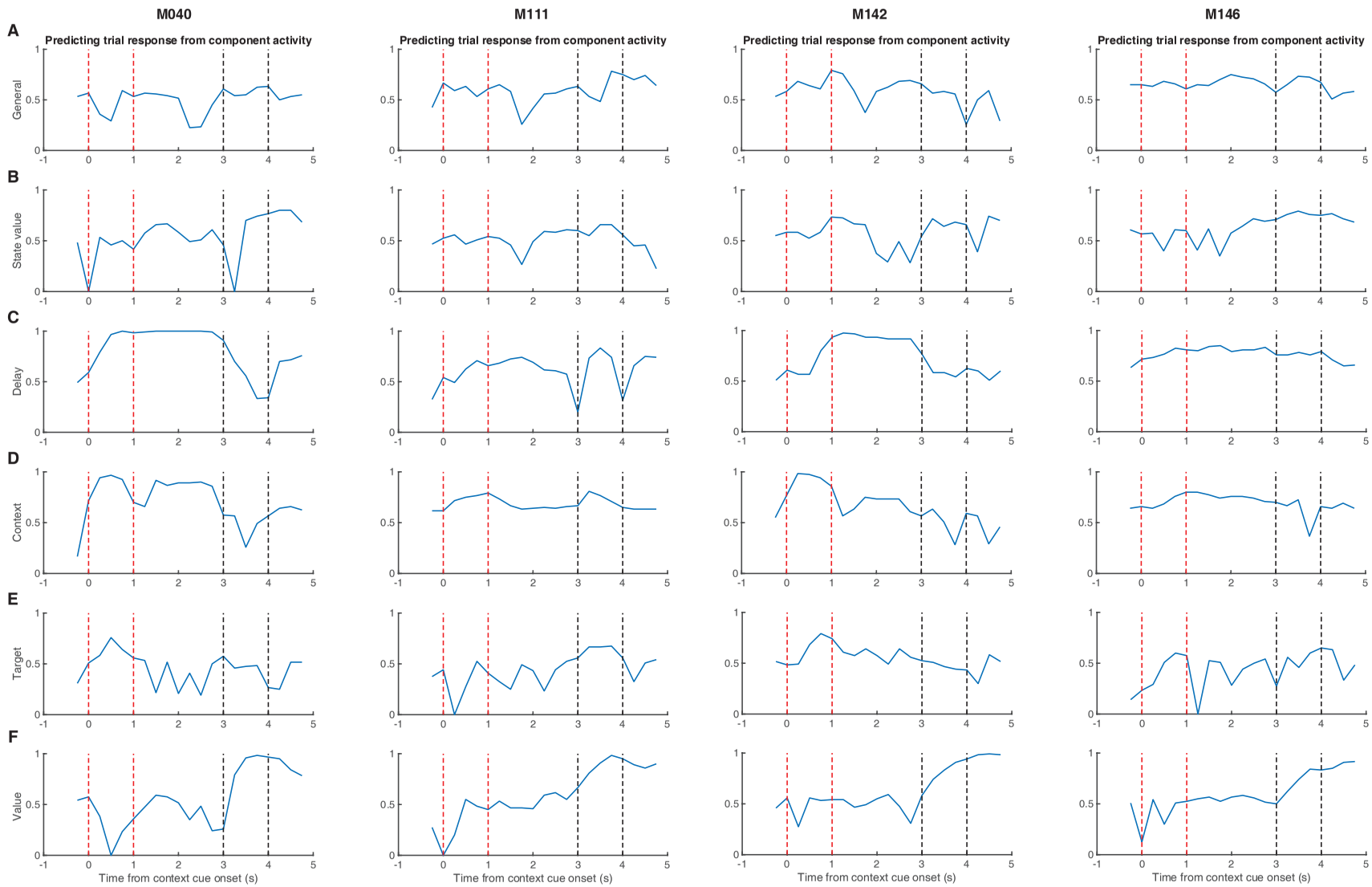

Figure 7 supplement 2: Predicting behavioral response for a given target cue based on projected activity along the components extracted in Figure 7. Each plot represents the accuracy of the behavioral prediction for a given component (rows) for each mouse (columns). Red lines border context cue presentation, and black lines border target cue presentation. From left-to-right shows predictions for M040, M111, M142 and M146. **A**: Prediction accuracy for the general cue component. **B**: Prediction accuracy for the state value component. **C**: Prediction accuracy for the context-related delay component. **D**: Prediction accuracy for the context component. **E**: Prediction accuracy for the target component. **F**: Prediction accuracy for the value component.

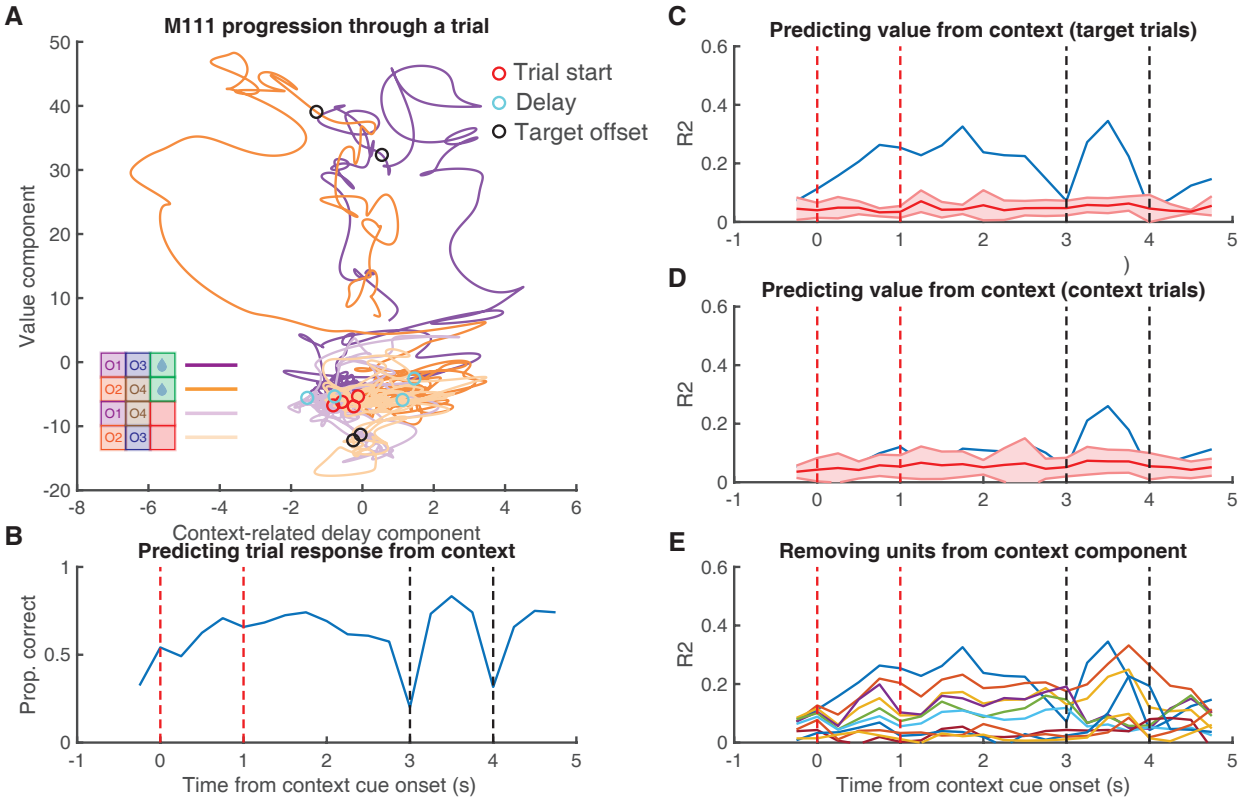

Figure 8 supplement 1: Context-dependent gating of the value-related component. Shown in this Figure is the relationship between the context-related delay component and the value-related component for M111. **A**: Progression of neural activity through a trial for each trial type in a two-dimensional neural subspace, with the trial-averaged projected activity in the context-related delay component (see Figure 7C) on the x-axis, and the trial-averaged projected activity in the value component (see Figure 7-supplement 1 F) on the y-axis. Note the relatively weak structure in the context-related delay axis, compared to M040. Red circles signal context-cue onset, cyan circles signal delay period 1 s after context cue offset, black circles signal 1 s after target cue onset. **B**: Predicting behavioral response for a given target cue based on projected activity along the context-related delay component at various timepoints. Red lines border context cue presentation, and black lines border target cue presentation. **C**: Predicting projected activity along the value-related axis after target cue onset for a given target cue (black circles from A) based on projected activity in the context-related axis at various timepoints. **D**: Control analysis predicting projected activity along the value-related axis after target cue onset for a given context cue based on projected activity in the context-related axis. **E**: Iteratively removing the top 10% of contributors to the context-related delay component and attempting to predict value-related activity as in C.

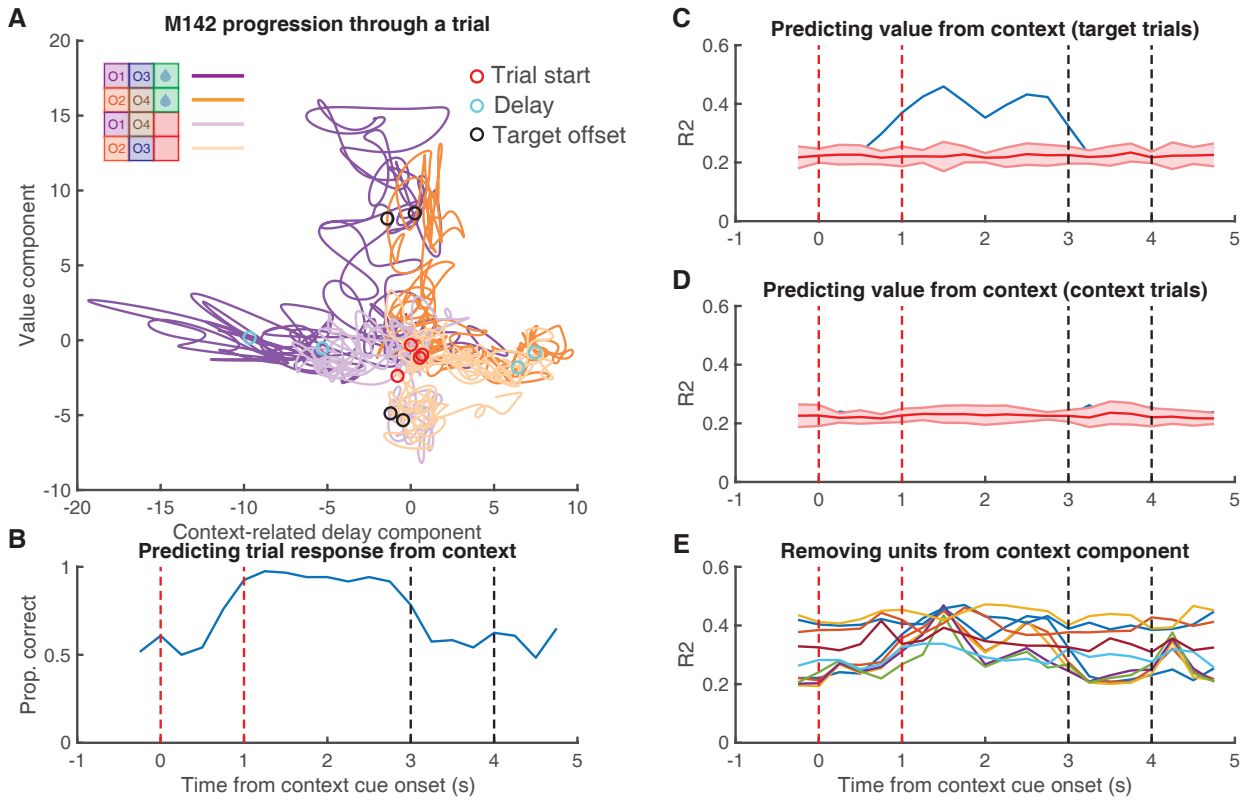

Figure 8 supplement 2: Context-dependent gating of the value-related component. Shown in this Figure is the relationship between the context-related delay component and the value-related component for M142. **A**: Progression of neural activity through a trial for each trial type in a two-dimensional neural subspace, with the trial-averaged projected activity in the context-related delay component (see Figure 7C) on the x-axis, and the trial-averaged projected activity in the value component (see Figure 7-supplement 1 F) on the y-axis. Throughout the progression of a trial a separation is observed along the context axis, which then flows into the value axis after target cue presentation, similar to M040. Red circles signal context-cue onset, cyan circles signal delay period 1 s after context cue offset, black circles signal 1 s after target cue onset. **B**: Predicting behavioral response for a given target cue based on projected activity along the context-related delay component at various timepoints. Red lines border context cue presentation, and black lines border target cue presentation. **C**: Predicting projected activity along the value-related axis after target cue onset for a given target cue (black circles from A) based on projected activity in the context-related axis at various timepoints. **D**: Control analysis predicting projected activity along the value-related axis after target cue onset for a given context cue based on projected activity in the context-related axis. **E**: Iteratively removing the top 10% of contributors to the context-related delay component and attempting to predict value-related activity as in C.

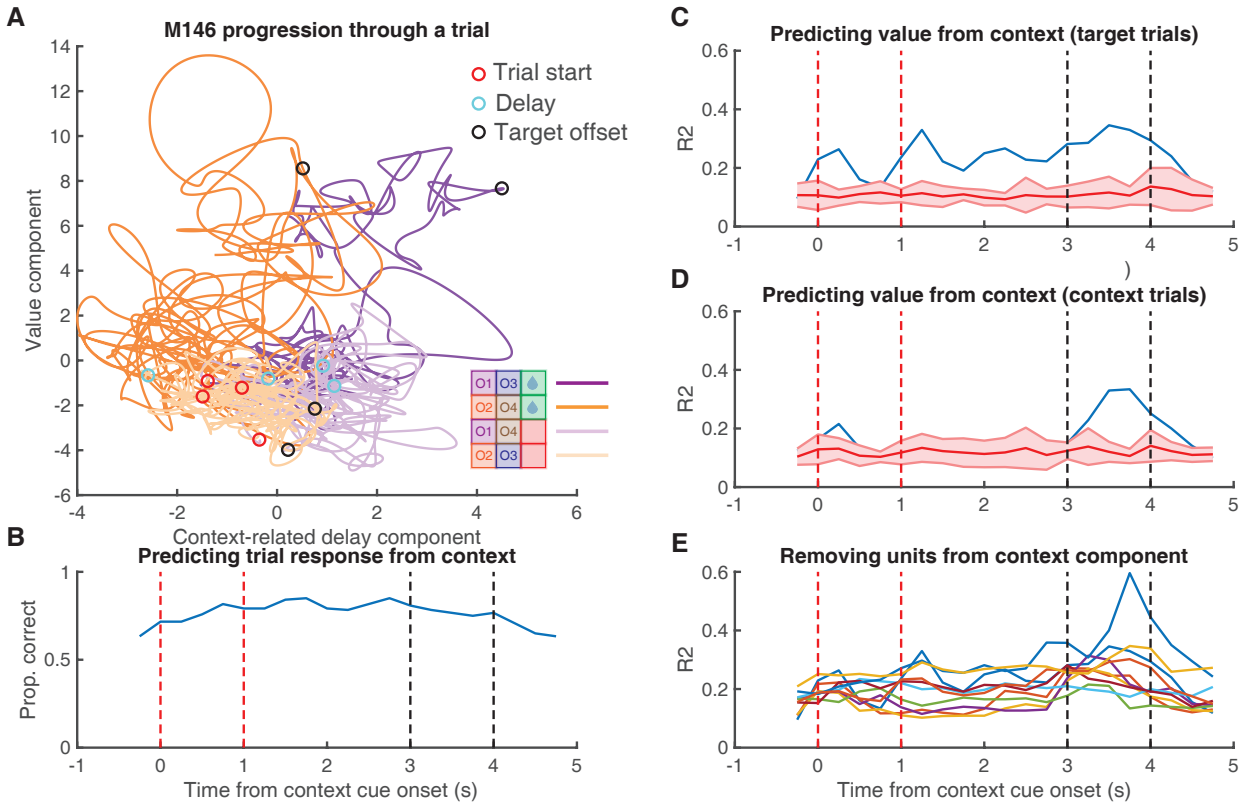

Figure 8 supplement 3: Context-dependent gating of the value-related component. Shown in this Figure is the relationship between the context-related delay component and the value-related component for M146. **A**: Progression of neural activity through a trial for each trial type in a two-dimensional neural subspace, with the trial-averaged projected activity in the context-related delay component (see Figure 7C) on the x-axis, and the trial-averaged projected activity in the value component (see Figure 7-supplement 1 F) on the y-axis. Note the relatively weak structure in the context-related delay axis, compared to M040. Red circles signal context-cue onset, cyan circles signal delay period 1 s after context cue offset, black circles signal 1 s after target cue onset. **B**: Predicting behavioral response for a given target cue based on projected activity along the context-related delay component at various timepoints. Red lines border context cue presentation, and black lines border target cue presentation. **C**: Predicting projected activity along the value-related axis after target cue onset for a given target cue (black circles from A) based on projected activity in the context-related axis at various timepoints. **D**: Control analysis predicting projected activity along the value-related axis after target cue onset for a given context cue based on projected activity in the context-related axis. **E**: Iteratively removing the top 10% of contributors to the context-related delay component and attempting to predict value-related activity as in C.

References

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

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

Calhoon, G. G., & O'Donnell, P. (2013). Closing the Gate in the Limbic Striatum: Prefrontal Suppression of Hippocampal and Thalamic Inputs. *Neuron*, *78*(1), 181–190. doi: 10.1016/j.neuron.2013.01.032

Ciano, P. D., Cardinal, R. N., Cowell, R. A., Little, S. J., Everitt, B. J., Di Ciano, P., . . . Everitt, B. J. (2001). Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. *The Journal of neuroscience*, *21*(23), 9471–9477. doi: 21/23/9471 [pii]

Cole, S. L., Robinson, M. J., & Berridge, K. C. (2018). Optogenetic self-stimulation in the nucleus accumbens: D1 reward versus D2 ambivalence. *PLoS ONE*, *13*(11). doi: 10.1371/journal.pone.0207694

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

Cox, J., & Witten, I. B. (2019). *Striatal circuits for reward learning and decision-making* (Vol. 20) (No. 8). Nature Publishing Group. doi: 10.1038/s41583-019-0189-2

Crow, T. J. (1972). A map of the rat mesencephalon for electrical self-stimulation. *Brain Research*, *36*(2), 265–273. doi: 10.1016/0006-8993(72)90734-2

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology*, *84*(6), 3072–3077. doi: 10.1152/jn.2000.84.6.3072

Dobrovitsky, V., West, M. O., & Horvitz, J. C. (2019). The role of the nucleus accumbens in learned approach behavior diminishes with training. *European Journal of Neuroscience*, *50*(9), 3403–3415. doi: 10.1111/ejn.14523

Ebitz, R. B., & Hayden, B. Y. (2021). The population doctrine revolution in cognitive neurophysiology. *arXiv*.doi: 10.1111/ejn.14523

Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M., & Cunningham, J. P. (2016). Reorganization between preparatory and movement population responses in motor cortex. *Nature communications*, *7*(1), 13239. doi: 10.1038/ncomms13239

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

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

Floresco, S. B., Montes, D. R., Tse, M. M., & van Holstein, M. (2018). Differential contributions of nucleus accumbens subregions to cue-guided risk/reward decision making and implementation of conditional rules. *Journal of Neuroscience*, *38*(8), 1901–1914. doi: 10.1523/JNEUROSCI.3191-17.2018

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). *Neural Manifolds for the Control of Movement* (Vol. 94) (No. 5). Cell Press. doi: 10.1016/j.neuron.2017.05.025

Ghods-Sharifi, S., & Floresco, S. B. (2010). Differential effects on effort discounting induced by inactivations of the nucleus accumbens core or shell. *Behavioral Neuroscience*, *124*(2), 179–191. doi: 10.1037/a0018932

Gill, K. M., & Grace, A. A. (2011). Heterogeneous processing of amygdala and hippocampal inputs in the rostral and caudal subregions of the nucleus accumbens. *The International Journal of Neuropsychopharmacology*, *14*(10), 1301–1314. doi: 10.1017/S1461145710001586

Gmaz, J. M., Carmichael, J. E., & van der Meer, M. A. A. (2018). Persistent coding of outcome-predictive cue features in the rat nucleus accumbens. *eLife*, *7*. doi: 10.7554/eLife.37275

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

Graybiel, A. M. (1975). Input-output anatomy of the basal ganglia. In *Proc. soc. neurosci.* Toronto, Canada.doi: 10.1523/JNEUROSCI.5349-11.2012

Groenewegen, H. J., Room, P., Witter, M. P., & Lohman, A. H. (1982). Cortical afferents of the nucleus accumbens in the cat, studied with anterograde and retrograde transport techniques. *Neuroscience*, *7*(4), 977–996. doi: 10.1016/0306-4522(82)90055-0

Gruber, A. J., Hussain, R. J., & O'Donnell, P. (2009). The nucleus accumbens: A switchboard for goal-directed behaviors. *PLoS ONE*, *4*(4). doi: 10.1371/journal.pone.0005062

Gruber, A. J., & McDonald, R. J. (2012). *Context, emotion, and the strategic pursuit of goals: Interactions among multiple brain systems controlling motivated behavior* (No. AUGUST). doi: 10.3389/fnbeh.2012.00050

Gu, X., & Li, C. T. (2019). Dynamic neuronal activation of a distributed cortico-basal ganglia-thalamus loop in learning a delayed sensorimotor task. *bioRxiv*, 568055. doi: 10.1101/568055

Gulli, R. A., Duong, L. R., Corrigan, B. W., Doucet, G., Williams, S., Fusi, S., & Martinez-Trujillo, J. C. (2020). Context-dependent representations of objects and space in the primate hippocampus during virtual navigation. *Nature Neuroscience*,

23(1), 103–112. doi: 10.1038/s41593-019-0548-3

Haber, S. N., & Behrens, T. E. (2014). *The Neural Network Underlying Incentive-Based Learning: Implications for Interpreting Circuit Disruptions in Psychiatric Disorders* (Vol. 83) (No. 5). Cell Press. doi: 10.1016/j.neuron.2014.08.031

Han, Z., Zhang, X., Zhu, J., Chen, Y., & Li, C. T. (2018). High-throughput automatic training system for odor-based learned behaviors in head-fixed mice. *Frontiers in Neural Circuits*, *12*, 15. doi: 10.3389/fncir.2018.00015

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

Humphries, M. D., & Prescott, T. J. (2010). *The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward* (Vol. 90) (No. 4). doi: 10.1016/j.pneurobio.2009.11.003

Ito, M., & Doya, K. (2015). Parallel Representation of Value-Based and Finite State-Based Strategies in the Ventral and Dorsal Striatum. *PLoS Computational Biology*, *11*(11), e1004540. doi: 10.1371/journal.pcbi.1004540

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

Kaufman, M. T., Churchland, M. M., Ryu, S. I., & Shenoy, K. V. (2014). Cortical activity in the null space: Permitting preparation without movement. *Nature Neuroscience*, *17*(3), 440–448. doi: 10.1038/nn.3643

Kaufman, M. T., Seely, J. S., Sussillo, D., Ryu, S. I., Shenoy, K. V., & Churchland, M. M. (2016). The largest response component in the motor cortex reflects movement timing but not movement type. *eNeuro*, *3*(4), 85–101. doi: 10.1523/ENEURO.0085-16.2016

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

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., … Machens, C. K. (2016). Demixed principal component analysis of neural population data. *eLife*, *5*(APRIL2016). doi: 10.7554/eLife.10989

Koob, G. F., & Volkow, N. D. (2016). *Neurobiology of addiction: a neurocircuitry analysis* (Vol. 3) (No. 8). Elsevier Ltd. doi: 10.1016/S2215-0366(16)00104-8

Ma, L., Chen, W., Yu, D., & Han, Y. (2020). Brain-Wide Mapping of Afferent Inputs to Accumbens Nucleus Core Subdomains and Accumbens Nucleus Subnuclei. *Frontiers in Systems Neuroscience*, *14*, 15. doi: 10.3389/fnsys.2020.00015

Mannella, F., Gurney, K., & Baldassarre, G. (2013). The nucleus accumbens as a nexus between values and goals in goal-directed behavior: a review and a new hypothesis. *Frontiers in Behavioral Neuroscience*, *7*, 135. doi: 10.3389/fnbeh.2013.00135

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

Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). *From motivation to action: Functional interface between the limbic system and the motor system* (Vol. 14) (No. 2-3). doi: 10.1016/0301-0082(80)90018-0

Mogenson, G. J., Takigawa, M., Robertson, A., & Wu, M. (1979). Self-stimulation of the nucleus accumbens and ventral tegmental area of tsai attenuated by microinjections of spiroperidol into the nucleus accumbens. *Brain Research*, *171*(2), 247–259. doi: 10.1016/0006-8993(79)90331-7

Murer, M. G., & O'Donnell, P. (2016). Gating of Cortical Input Through the Striatum. In *Handbook of behavioral neuroscience* (Vol. 24, pp. 439–457). Elsevier B.V. doi: 10.1016/B978-0-12-802206-1.00022-2

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

O'Donnell, P., & Grace, A. A. (1995). Synaptic interactions among excitatory afferents to nucleus accumbens neurons: Hippocampal gating of prefrontal cortical input. *Journal of Neuroscience*, *15*(5 I), 3622–3639. doi: 10.1523/jneurosci.15-05-03622.1995

Oemisch, M., Westendorff, S., Azimi, M., Hassani, S. A., Ardid, S., Tiesinga, P., & Womelsdorf, T. (2019). Feature-specific prediction errors and surprise across macaque fronto-striatal circuits. *Nature Communications*, *10*(1), 1–15. doi: 10.1038/s41467-018-08184-9

Parker, N., Baidya, A., Cox, J., Haetzel, L., Zhukovskaya, A., Murugan, M., … Witten, I. (2019). Choice-selective sequences dominate in cortical relative to thalamic inputs to nucleus accumbens, providing a potential substrate for credit assignment. *bioRxiv*, 725382. doi: 10.1101/725382

Parkinson, J. A., Willoughby, P. J., Robbins, T. W., & Everitt, B. J. (2000). Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs pavlovian approach behavior: Further evidence for limbic cortical-ventral striatopallidal systems. *Behavioral Neuroscience*, *114*(1), 42–63. doi: 10.1037//0735-7044.114.1.42

Phillips, A. G., Brooke, S. M., & Fibiger, H. C. (1975). Effects of amphetamine isomers and neuroleptics on self-stimulation from the nucleus accumbens and dorsal nor-adrenergenic bundle. *Brain Research*, *85*(1), 13–22. doi: 10.1016/0006-8993(75)90998-1

Prado-Alcalá, R., & Wise, R. A. (1984). Brain stimulation reward and dopamine terminal fields. I. Caudate-putamen, nucleus accumbens and amygdala. *Brain Research*, *297*(2), 265–273. doi: 10.1016/0006-8993(84)90567-5

Raposo, D., Kaufman, M. T., & Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nature Neuroscience*, *17*(12), 1784–1792. doi: 10.1038/nn.3865

Reynolds, S. M., & Berridge, K. C. (2008). Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. *Nature Neuroscience*, *11*(4), 423–425. doi: 10.1038/nn2061

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

Rusu, S. I., & Pennartz, C. M. (2020). *Learning, memory and consolidation mechanisms for behavioral control in hierarchically organized cortico-basal ganglia systems* (Vol. 30) (No. 1). John Wiley and Sons Inc. doi: 10.1002/hipo.23167

Sadacca, B. F., Jones, J. L., & Schoenbaum, G. (2016). Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework. *eLife*, *5*(MARCH2016). doi: 10.7554/eLife.13665

Saez, A., Rigotti, M., Ostojic, S., Fusi, S., & Salzman, C. D. (2015). Abstract Context Representations in Primate Amygdala and Prefrontal Cortex. *Neuron*, *87*(4), 869–881. doi: 10.1016/j.neuron.2015.07.024

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

Salamone, J. D., Cousins, M. S., & Bucher, S. (1994). Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit procedure. *Behavioural Brain Research*, *65*(2), 221–229. doi: 10.1016/0166-4328(94)90108-2

Saxena, S., & Cunningham, J. P. (2019). *Towards the neural population doctrine* (Vol. 55). Elsevier Ltd. doi: 10.1016/j.conb.2019.02.002

Schultz, W., Apicella, P., Scarnati, E., & Ljungberg, T. (1992). Neuronal activity in monkey ventral striatum related to the expectation of reward. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *12*(12), 4595–610. doi: 10.1523/JNEUROSCI.12-12-04595.1992

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

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

Stoianov, I. P., Pennartz, C. M., Lansink, C. S., & Pezzulo, G. (2018). Model-based spatial navigation in the hippocampus-ventral striatum circuit: A computational analysis. *PLoS Computational Biology*, *14*(9). doi: 10.1371/journal.pcbi.1006316

Thura, D., Cabana, J. F., Feghaly, A., & Cisek, P. (2020). *Unified neural dynamics of decisions and actions in the cerebral cortex and basal ganglia.* bioRxiv. doi: 10.1101/2020.10.22.350280

Tsai, H. C., Zhang, F., Adamantidis, A., Stuber, G. D., Bond, A., De Lecea, L., & Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*, *324*(5930), 1080–1084. doi: 10.1126/science.1168878

van der Meer, M. A. A., & Redish, A. D. (2011a). 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

van der Meer, M. A. A., & Redish, A. D. (2011b). Ventral striatum: a critical look at models of learning and evaluation. *Current Opinion in Neurobiology*, *21*(3), 387–392. doi: 10.1016/J.CONB.2011.02.011

Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). *Computation through Neural Population Dynamics* (Vol. 43). Annual Reviews Inc. doi: 10.1146/annurev-neuro-092619-094115

Yoo, S. B. M., & Hayden, B. Y. (2019). The Transition from Evaluation to Selection Involves Neural Subspace Reorganization in Core Reward Regions. *Neuron*, *105*(4), 712–724.e4. doi: 10.1016/j.neuron.2019.11.013

Zhang, X., Yan, W., Wang, W., Fan, H., Hou, R., Chen, Y., … Li, C. T. (2019). Active information maintenance in working memory by a sensory cortex. *eLife*, *8*. doi: 10.7554/eLife.43191

Zhou, J., Gardner, M. P., Stalnaker, T. A., Ramus, S. J., Wikenheiser, A. M., Niv, Y., & Schoenbaum, G. (2019). Rat Orbitofrontal Ensemble Activity Contains Multiplexed but Dissociable Representations of Value and Task Structure in an Odor Sequence Task. *Current Biology*, *29*(6), 897–907.e3. doi: 10.1016/j.cub.2019.01.048