

Striatal Representations Reflect Task Demands

Nathaniel D. Daw^{1,3}, David S. Touretzky^{1,3}, & William E. Skaggs^{2,3}

¹Department of Computer Science, Carnegie Mellon University

²Department of Neuroscience, University of Pittsburgh

³Center for the Neural Basis of Cognition

8th February 2002

Abstract

Temporal-difference learning models of the dopamine system have led modelers to postulate that the striatum — the chief input to the system and a key dopamine target — performs reward prediction or action selection. We report on a striatal recording study in rats performing a T-maze task, which suggests that the striatum instead may provide a task-specialized representation of the information that downstream areas would need to perform these functions. Animals trained on different versions of the task showed differing striatal representations apparently specialized to task demands, and, as a group, their neurons did not tend to anticipate upcoming action choices or rewards.

1 Introduction

Recent models [2, 3, 1] based on the temporal-difference learning algorithm [5] have had some success at explaining the responses of dopamine neurons in the primate midbrain as carrying an error signal for the prediction of reward and the selection of actions. This has led modelers to postulate that the striatum — which is both the chief input to the dopamine neurons and one of their key targets — is responsible for either or both of these functions [1, 4]. We report the results of a striatal recording study in rats performing a cued choice on a T-maze, which cast some doubt on both of these possibilities. Instead, the data suggest a

model in which striatal neurons represent the information from which downstream areas could make decisions or predictions. Such a representation could be specialized to the task, and might include not just current sensory data but other information that is relevant to choices, such as memories of past events.

2 Methods

Two rats performed cued decisions for food reward in the T-maze pictured in Figure 1. After running up the center path to start a trial, the animal received a light or tone cue indicating whether he should turn left or right at the T-junction. If he then made the correct choice, a food pellet was delivered from a dispenser at the corner of the maze. Finally, he could return to the base of the maze to start a new trial.

One rat received the same stimulus contingencies — light for left turn; tone for right turn — throughout the experiment, and performed near 100% correctly during recording. The other subject faced repeated reversals of the stimulus-action associations, alternating between light→left, tone→right and light→right, tone→left. A reversal occurred about a third of the way through each day’s session of roughly 450 trials. After much training, the rat was usually able to return to at least 80% accuracy within the remainder of the session. The recordings presented here were taken from the portion of the session prior to the reversal, during which the behavior was more stable. The reported effects look similar throughout the session, however.

Data were recorded using a Neuralynx Cheetah system from an array of twelve tetrodes made from 25 micron nichrome wire insulated with polyimide. Reward delivery was controlled by a clock sequence running on a DataWave Discovery system. Electrodes were positioned above the dorsal striatum (1.5mm anterior to bregma, 1.5mm right of the midline) and incrementally lowered into more ventral portions of striatum.

Between the two animals, 357 cells were marked as clean and well-isolated; the rest were excluded. Since trials were self-paced, the intervals between task events (dotted lines in Figure 1) varied. After discarding a small number of trials with unusually long pauses, we warped the remaining trials onto a common timeline by dividing the intervals between task events into a fixed number of equal-duration bins, chosen to average around 150 ms/bin over all trials. Firing rates within each bin were then averaged across all warped trials to generate an "average timeline" resembling a peri-event time histogram aligned on multiple events.

3 Results

Approximately 70% of the units appeared to be medium spiny neurons according to the criteria of low mean firing rates and bursty spike trains. Approximately 15% appeared to be tonically active neurons according to the criteria of tonic activity and mean firing rates between 3 and 10 Hz. The remaining cells were very diverse.

Figure 2 displays the task-related tuning of the 240 units from both animals whose firing rate exceeded 0.5 Hz in some bin and was less than 10 Hz in some bin. (These criteria excluded neurons which were unusually active or inactive.) Each row of the figure represents the average firing of a single unit over the course of a trial; warmer colors represent bins with higher firing rates, and the area under the rate curve for each neuron is normalized so that the warmest colors occur in the most sharply tuned neurons. The neurons are sorted by where in the task their peak firing occurs. From this depiction, it is clear that the majority of neurons show phasic task-related firing. There may be slightly elevated numbers of units tuned to times between the cue and the decision, and to times near the reward, but, as a population, the neurons tile the timeline of the task fairly evenly.

We explored more specific correlates for the units, constructing and comparing separate histograms for trials meeting different criteria. The cell in Figure 3 typifies an interesting class of neurons: it is excited during the period between the cue and the reward, but more so when the choice on the *previous* trial had been a right turn than a left. In contrast, it shows very little preference for left or right turns on the *current* trial, the action the animal is currently executing. Such neurons are also generally better correlated with the previous *action* than the previous *stimulus*, exemplified by the neuron in Figure 4 (this is measurable because the animal did not always make the correct choice). Figure 5 shows more examples of similar neurons, including a tonically active one and one whose firing varies depending on whether the previous trial was performed correctly or incorrectly.

We tested the prevalence and significance of this behavior using a two-tailed T-test on firing rates during the interval between the decision point and reward delivery. (Earlier firing was not included to ensure that any activity related to persistent behavioral reflections of the previous choice was excluded.) We counted the

number of neurons whose firing rates were highly significantly different ($p < .001$) when the previous choice was left versus right, when the current choice was left versus right, when the previous and current stimuli were lights versus tones, and when the previous choice was correct versus incorrect. In the animal who received frequent reversals in task contingencies (Figure 6, top) many more cells (20/197) significantly reflected his behavior on the *previous* trial than reflected his behavior on the current trial (4/197). Also, more neurons were correlated with the previous action choice than the previous stimulus.

In contrast, the animal who performed the task with consistent contingencies showed more cells sensitive to his current choice (13/160) than his previous choice (6/160) (Figure 6, bottom). Because this animal performed the task with great accuracy, it was difficult to identify cells that distinguished failure from success and to determine whether neurons were better correlated with stimuli or action choices.

4 Discussion

These results raise the question of why the animals differed in their striatal representations. One answer is that the task differences engendered different neuronal representations specialized for solving each version. Frequent reversals in task contingencies may promote richer representation of past events because such information is useful in detecting and responding to environmental shifts. More specifically, a flexible solution to the serial reversal problem chooses an action by comparing the previous stimulus and response, and whether they were rewarded, to the current stimulus, rather than simply learning mappings from the current cue to the appropriate action, which have to be relearned after each reversal. Though the animal faced with the reversals was not fully using this higher-order rule, he seemed to employ some elements of this strategy. If so, his striatal neurons carried some of the information about past events that he needed to choose the current action.

These data are not particularly consonant with the hypothesis arising out of several modeling efforts [1, 4] that the striatum performs reward prediction or action selection *per se*; in one animal, very few neurons reflected the animal's *current* action choice, and in both animals, the neurons' temporal tuning was not strongly concentrated around rewards but instead tiled the task's full timeline fairly uniformly. Together

with the way the striatal representation seems to have adapted to the task differences between animals, these observations suggest an alternative, that the neurons instead represent what is known in temporal-difference models as the *state* of the task: the information necessary to make action decisions or reward predictions. For many tasks, this information would include not just immediate sensory input, but also remembered or inferred information that is relevant to prediction and action.

Acknowledgments

Supported by NSF IIS-9978403 and DGE-9987588. We are grateful to Judith Joyce Balcita-Pedicino for significant technical assistance, and to Mark Albert for assistance with rat training.

References

- [1] J.C. Houk, J.L. Adams, and A.G. Barto, A model of how the basal ganglia generate and use neural signals that predict reinforcement, in: J.C. Houk, J.L. Davis, and D.G. Beiser, eds., *Models of Information Processing in the Basal Ganglia* (MIT Press, Cambridge, 1995) 249-270.
- [2] P.R. Montague, P. Dayan, and T.J. Sejnowski, A framework for mesencephalic dopamine systems based on predictive Hebbian learning, *J. Neurosci.* 16 (1996) 1936-1947.
- [3] W. Schultz, P. Dayan and P.R. Montague, A neural substrate of prediction and reward, *Science* 275 (1997) 1593-1599.
- [4] R.E. Suri and W. Schultz, Temporal difference model reproduces predictive neural activity, *Neural Computation* 13 (2001) 841-862.
- [5] R.S. Sutton, Learning to predict by the method of temporal differences, *Machine Learning* 3 (1988) 9-44.

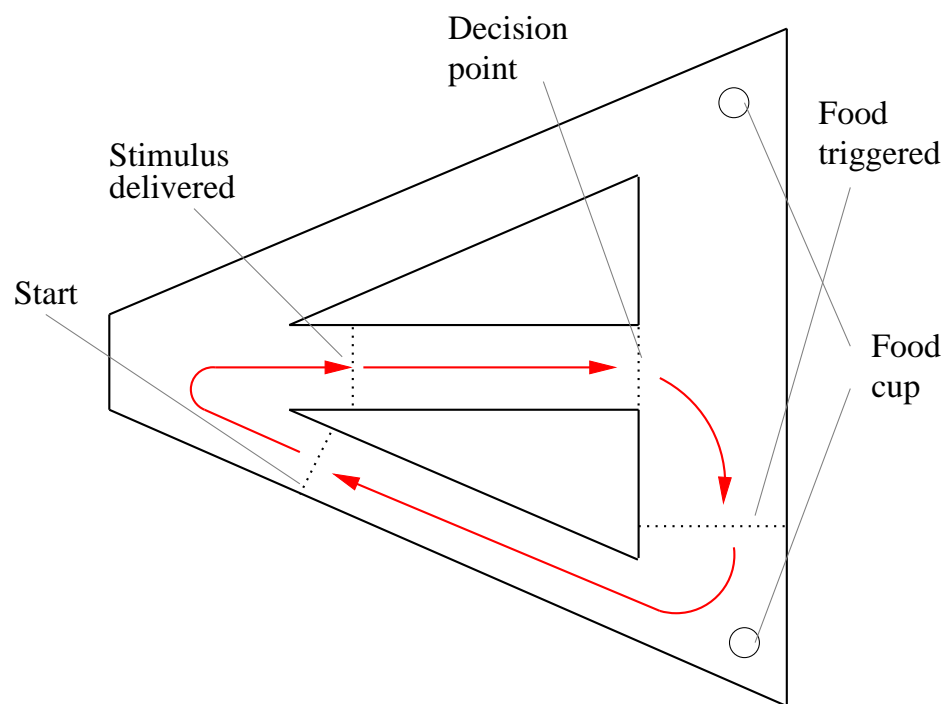


Figure 1: T-maze apparatus used in study.

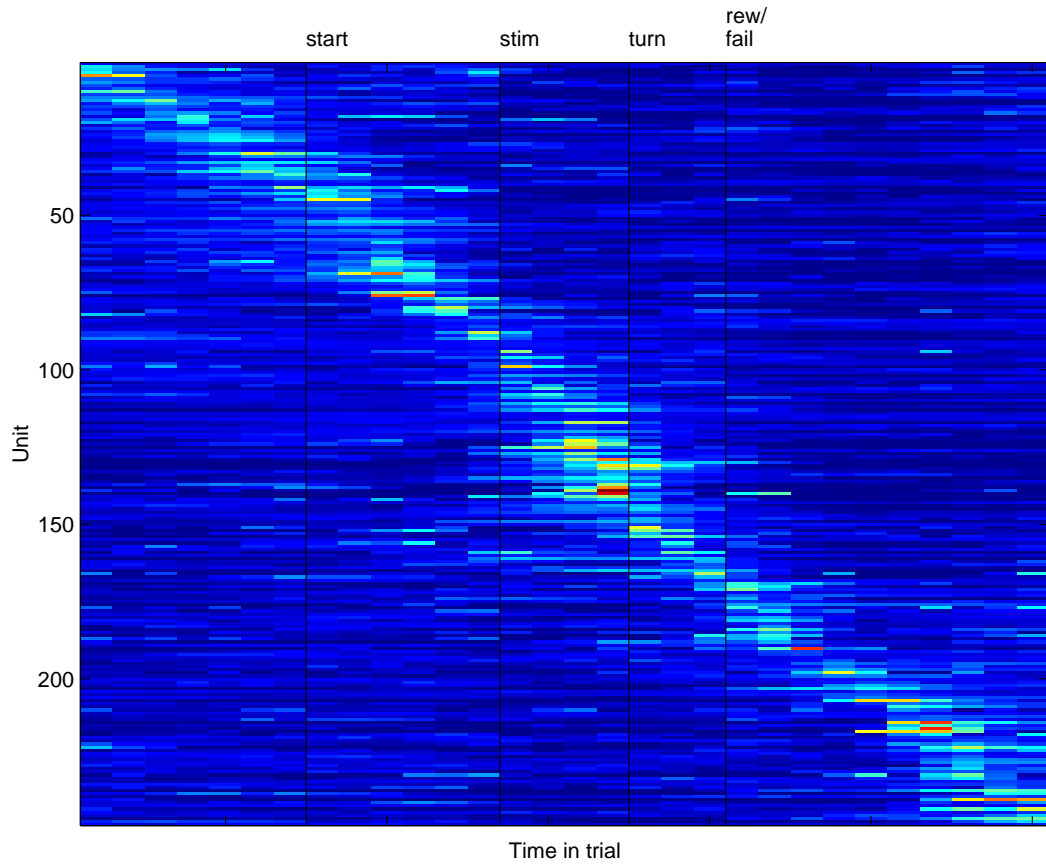


Figure 2: Task-related firing of 240 striatal neurons. Each row represents the firing of a single striatal neuron; normalized firing rate is indicated by color, with warmer colors signifying bins with higher firing rates. Neurons are sorted by the time of peak firing.

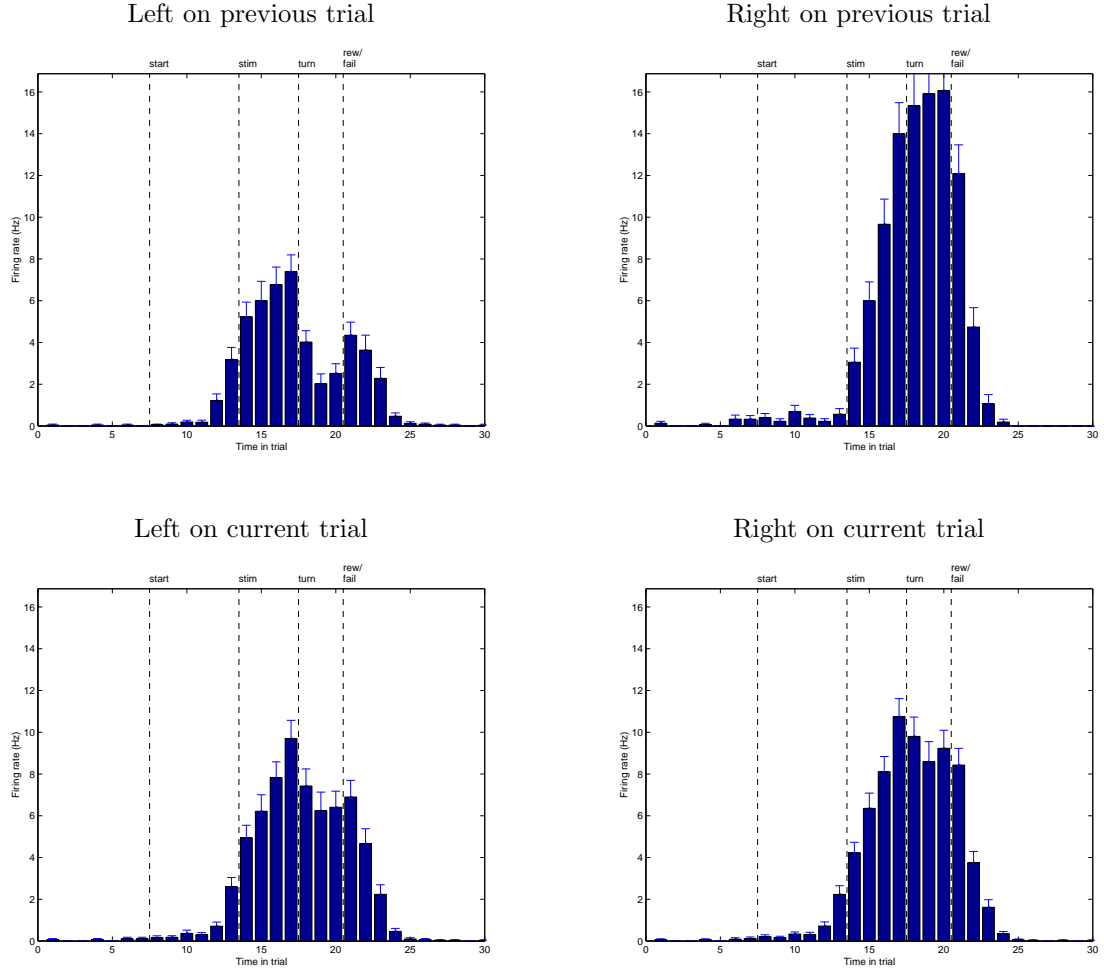


Figure 3: Firing histograms for a neuron broken down by whether the choices on the previous and current trials were left or right turns, showing that the neuron distinguishes the action chosen on the previous trial.

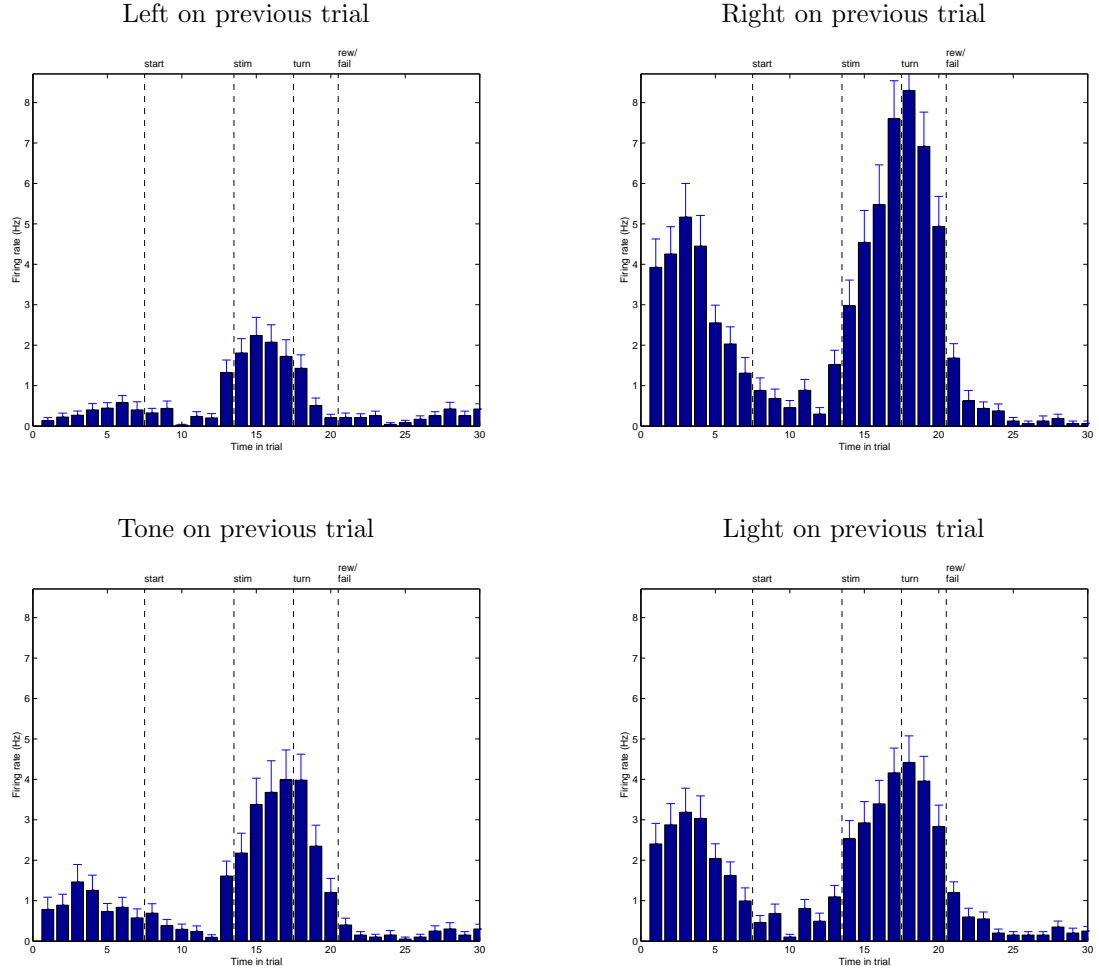


Figure 4: Firing histograms for a neuron broken down by whether the choice on the previous trial was a left or right turn, and whether the stimulus on the previous trial was a tone or a light, showing that the neuron is better correlated with the previous action than the previous stimulus.

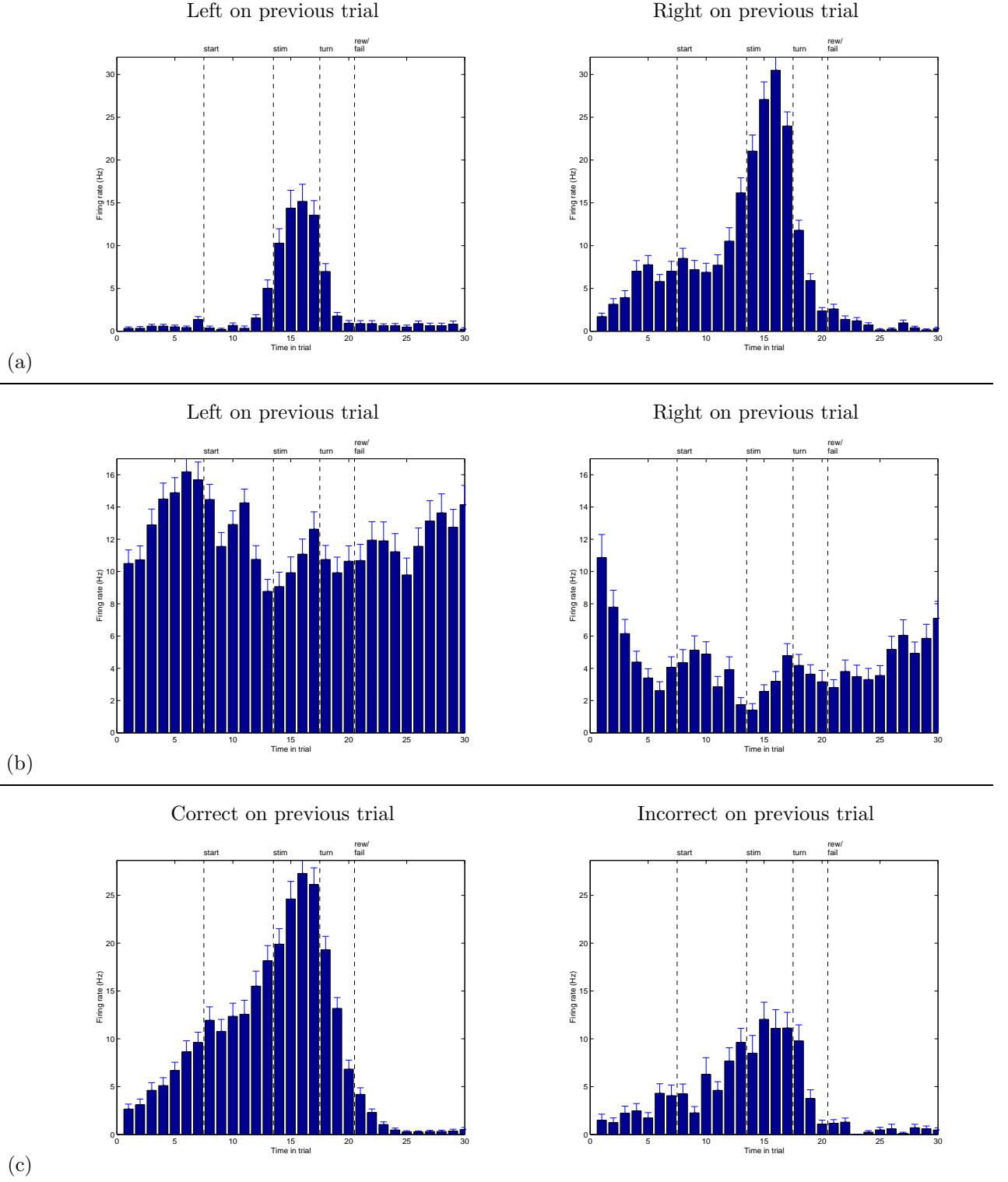


Figure 5: Firing histograms for three neurons. (a) Another neuron with phasic activity modulated by the previous choice. (b) A neuron with tonic activity modulated by the previous choice. (c) A neuron whose activity is modulated by whether the previous choice was correct or incorrect.

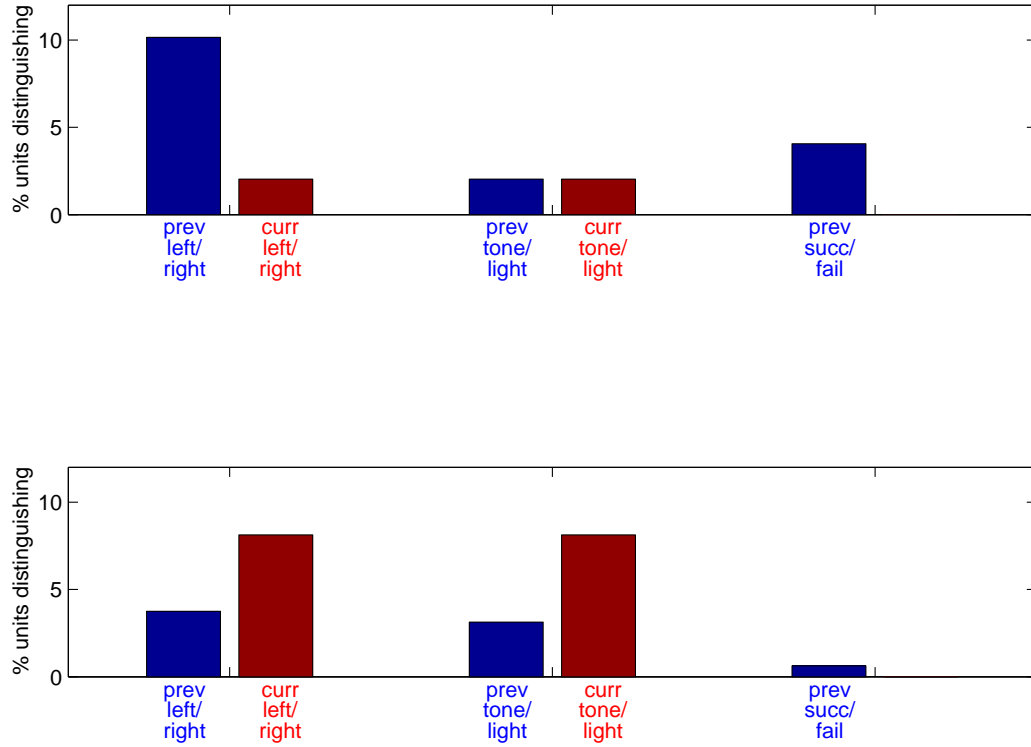


Figure 6: Numbers of neurons whose firing significantly distinguished various task events (two tailed T-tests, $p < 0.001$). Top: In the animal who faced repeated reversals in task contingencies, more neurons distinguished whether the *previous* choice was left or right than distinguished whether the current choice was left or right. Fewer neurons distinguished the current and previous stimuli and whether the previous choice was rewarded. Bottom: The animal who faced consistent task contingencies showed a contrasting pattern: more neurons significantly distinguished the current than the previous choice.