



Published in final edited form as:

*Nat Neurosci.* 2016 June ; 19(6): 845–854. doi:10.1038/nn.4287.

## Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target

Nathan F. Parker<sup>1</sup>, Courtney M. Cameron<sup>1</sup>, Joshua P. Taliaferro<sup>1</sup>, Junuk Lee<sup>1</sup>, Jung Yoon Choi<sup>1</sup>, Thomas J. Davidson<sup>2</sup>, Nathaniel D. Daw<sup>1</sup>, and Ilana B. Witten<sup>1,3</sup>

<sup>1</sup>Princeton Neuroscience Institute and Department of Psychology, Princeton University, Princeton, NJ, 08544

<sup>2</sup>Howard Hughes Medical Institute, University of California, San Francisco, CA, 94158

### Abstract

Dopaminergic (DA) neurons in the midbrain provide rich, topographic innervation of the striatum and are central to learning and to generating actions. Despite the importance of this DA innervation, it remains unclear if and how DA neurons are specialized based on the location of their striatal target. Thus, we sought to compare the function of subpopulations of DA neurons that target distinct striatal subregions in the context of an instrumental reversal learning task. We identified key differences in the encoding of reward and choice in dopamine terminals in dorsal versus ventral striatum: DA terminals in ventral striatum responded more strongly to reward consumption and reward-predicting cues, whereas DA terminals in dorsomedial striatum responded more strongly to contralateral choices. In both cases the terminals encoded a reward prediction error. Our results suggest that the DA modulation of the striatum is spatially organized to support the specialized function of the targeted subregion.

### Introduction

Two essential aspects of behavior are learning to perform actions that lead to reward, and generating those actions. The neuromodulator DA is implicated in both processes<sup>1–4</sup>, but how DA supports these related but distinct functions remains the subject of debate. Historically, dopamine was thought to be involved primarily in generating actions, due to its role in movement disorders in humans (e.g. Parkinson's disease) and the dramatic motor phenotypes following DA perturbations in animals<sup>3,5,6</sup>. However, seminal work demonstrated that rather than encoding actions, phasic activity in putative DA neurons in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) encodes a reward prediction error, or reinforcement signal<sup>7,8</sup>, which in theory could be used to learn which

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>3</sup>Correspondence to: iwitten@princeton.edu.

**Author Contributions** N.F.P., C.M.C., J.P.T. & J.L. performed the experiments; N.F.P., C.M.C., J.P.T., J.L. & J.Y.C. analyzed the data; T.J.D. provided advice on rig design; N.D.D. & I.B.W. provided advice on statistical analysis; N.F.P., N.D.D. & I.B.W. designed experiments and interpreted the results; N.F.P. & I.B.W. wrote the manuscript.

**Competing Financial Interests** The authors declare no competing financial interests.

stimuli or actions are likely to lead to reward<sup>2</sup>. Consistent with this idea, phasic activation of DA neurons is sufficient to support learning<sup>9,10</sup>. Thus, the idea emerged that phasic DA activity may only have an indirect effect on movement, either by driving learning, or by other mechanisms, such as altering motivational state<sup>11,12</sup> or by affecting the balance between activity in the output pathways of the striatum<sup>13,14</sup>.

However, most of our understanding of what DA neurons do and do not encode has emerged from recordings from putative DA neurons without knowledge of the striatal subregion targeted by the neurons. Notably, different DA neurons in the VTA/SNc receive different inputs, and project in a topographic manner to different striatal subregions<sup>15–19</sup>. These striatal subregions themselves have specialized anatomical and functional organization, with dorsal regions, such as the dorsomedial striatum (DMS), implicated in evaluating and generating actions<sup>20–22</sup>, and ventral regions, such as the nucleus accumbens (NAc), implicated in processing reward<sup>23,24</sup>.

Indeed, these considerations lead to the intriguing hypothesis that DA projections to striatal subregions could support the specialized function of the target area. For example, subpopulations of DA neuron might contribute to movement not merely as an indirect consequence of learning or motivation, but instead they could directly encode actions. To test the idea that subpopulations of DA neurons that project to specific striatal regions encode specialized information, we sought to compare the representation of rewards and actions in DA axon terminals in the DMS and the NAc as mice performed an instrumental reversal learning task. Toward this end, we combined recent advances in measuring neural activity using calcium indicators in axons terminals of genetically specified neurons deep in the brain<sup>25,26</sup>, along with other complementary approaches, including a statistical model that allowed isolation of the calcium responses to individual behavioral events, optogenetic perturbations, and fast scan cyclic voltammetry.

## Results

### Task and behavior

Mice were trained to perform an instrumental reversal learning task (task schematic, Fig. 1a). The start of each trial was signified by the illumination of a central noseport. After the mice entered the noseport (“nose poke”), two levers were presented (“lever presentation”). One lever corresponded to a high probability of reward (70%) while the other lever corresponded to a low probability (10%); which lever (right or left) corresponded to the high versus low probability of reward was reversed in a probabilistic manner after at least 10 rewarded trials. Lever presses which resulted in a reward were followed by one auditory stimulus (positive conditioned stimulus, or CS+) and lever presses which were not associated with reward were followed by a different auditory stimulus (CS–). A temporal jitter between 0 and 1s was introduced between the nose poke and the lever presentation, as well as the lever press and the CS, to enable separation of the neural responses of temporally neighboring behavioral events.

Mice continually learned which lever to press based on recent experience, as they were more likely to press the lever with the higher probability of reward, and their choices closely

followed the reversal of the lever probabilities (Fig. 1b). To quantify how mice were using previous trial outcomes to inform their choice, we used a logistic regression model to predict the animal's choice (right or left lever) based on previous trial choices and outcomes<sup>20,27,28</sup>. In this model, a positive regression coefficient indicates that an animal was more likely to return to a previously chosen lever, while a negative regression coefficient indicates that an animal was more likely to switch to the other lever. The model revealed that previously rewarded choices increase the likelihood of returning to the same lever in comparison to unrewarded choices (Fig. 1c). The effects of previous trials decayed roughly exponentially with increasing number of trials back (Fig. 1c). This pattern is consistent with error-driven learning models, in which reward (relative to lack of reward) drives learning about an actions' values and ultimately choices<sup>28</sup>. The model provided a good fit to each mouse's behavior, indicating that mice indeed learned from recent experience to guide their choice of action (example model fit in Fig. 1d;  $R^2$  ranged from 0.29 to 0.49 with a median of 0.40,  $n=16$  mice).

### Timed inhibition of dopamine neurons disrupts learning

To determine if and how DA neuron activity affected the animal's choice on future trials, we transiently inhibited DA neurons in the VTA/SNc on a subset of trials as mice performed this task (example behavior in Supplementary Fig. 1a). TH: IRES-Cre mice received injections in the VTA of an AAV expressing either Cre-dependent NpHR-YFP or YFP-only (control virus), as well as bilateral optical fiber implants above each VTA (Fig. 2A). To confirm the efficacy of NpHR-mediated inhibition in TH+ neurons, in a separate group of mice, we performed whole cell recordings in brain slices and observed large photocurrents (689 $\pm$ 16 pA) and effective elimination of current-evoked spiking (Supplementary Fig. 1b,d). After mice were trained on the behavioral task (Fig. 1a), DA neurons were inhibited bilaterally on a randomly selected 10% of trials throughout the duration of a trial (from the initial nose poke until either the CS- or the end of reward consumption). We quantitatively described the effect of DA inhibition on behavior by incorporating the optical inhibition into the logistic regression model of choice introduced in Fig. 1c. This model revealed a significant interaction between the animals' choice and the optical inhibition of DA neurons, as optical inhibition on previous trials decreased the probability of returning to the previously chosen lever in comparison to previous trials without inhibition (Fig. 2d, left panel,  $n=8$  NpHR-YFP mice). The effect of optical inhibition was similar whether the previous trial was rewarded or unrewarded (Fig. 2d). This effect was greatest in the case of inhibition on the previous trial, and declined with the number of trials separating the inhibition and the choice, suggesting the effect (like that of rewards) was mediated by error-driven learning. In contrast, light had no discernible effect on behavior in YFP-control littermates (Fig. 2d, right panel,  $n=8$  YFP-only mice; Choice X Light X Opsin was statistically significant in the case of both rewarded and unrewarded choices one trial back;  $p=0.0043$  for rewarded choice and  $p=0.0134$  for unrewarded choices;  $p$ -values from a mixed-effect logistic regression that predicted choice based on previous rewarded and unrewarded choice, light and opsin, and all interactions thereof). Given recent reports of Cre expression in neurons that do not express TH near the VTA of TH:IRES-Cre mice<sup>29</sup>, we replicated the same experiment in DAT: Cre mice and observed comparable effects of DA neuron inhibition on behavior (Supplementary Fig. 1e,f). Together, these results indicate that

activity in DA neurons plays a role in determining an animal's choice on future trials in this task, and is consistent with dopamine neuron inhibition functioning as a negative reward prediction error.

We also compared the latency to initiate the next trial after stimulation versus control trials, and found no significant differences (average latency was  $9.89 \pm 0.17$ s after stimulation versus  $9.93 \pm 0.13$ s after control trials). This indicates that our manipulation is relatively specific for choice. In order to determine if there is a time window within the trial that contributes preferentially to the effect of inhibition on choice, we inhibited DA neurons either from trial initiation to lever press, or else from lever press until 2s after reward consumption. Neither of these 'subtrial' inhibitions had an observable effect on choice (Supplementary Fig. 1g).

### Recordings and analysis of calcium signals from VTA/SN::NAc and VTA/SN::DMS terminals

After confirming that DA neuron activity in the VTA/SNc affected task performance, we employed a genetically encoded calcium indicator to measure and compare neural activity in the terminals of DA neurons originating in the VTA/SNc and projecting to either the dorsomedial striatum ("VTA/SN: :DMS") or the nucleus accumbens core ("VTA/SN: :NAc") during the reversal learning task. Retrograde tracer injections in the two striatal regions demonstrated that both regions received inputs from the lateral portion of the VTA and medial portion of the SNc. Despite the fact that these two striatal regions received inputs from a partially overlapping part of the VTA/SNc, quantification of the labeling profile of the retrograde tracers revealed that the subpopulation of TH+ neurons projecting to each region was largely non-overlapping (only  $8 \pm 4\%$  of TH+ neurons that projected to either area projected to both areas,  $n=487$  neurons from  $n=6$  mice; Supplementary Fig. 2).

In order to record from the terminals of these DA neuron subpopulations, a Cre-dependent AAV virus expressing the genetically encoded calcium indicator, gCaMP6f, was injected in the VTA of TH: IRES-Cre mice, and optical fibers were implanted above the DMS and NAc (Fig. 3C; each mouse had one recording site in DMS and one in NAc). We first confirmed colocalization of gCaMP6f and TH+ in VTA/SN: :NAc and VTA/SN: :DMS neurons by quantifying overlap in VTA/SNc between a retrograde tracer injected into NAc or DMS, TH expression, and Cre-mediated gCaMP6f expression (example image in Fig. 3a). This revealed a high degree of specificity between gCaMP6f and TH expression in VTA/SN: :NAc and VTA/SN: :DMS neurons ( $99 \pm 0.5\%$  of VTA/SN: :NAc and  $95 \pm 2\%$  of VTA/SN: :DMS neurons that expressed gCaMP6f also expressed TH) as well as penetrance of gCaMP6f expression in TH+ neurons ( $90 \pm 2\%$  of NAc-projecting TH+ cells and  $88 \pm 3\%$  of DMS-projecting TH+ cells expressed gCaMP6f;  $n=229$  VTA/SN: :NAc neurons and  $n=211$  VTA/SN: :DMS neurons;  $n=2$  mice).

Fluctuations in fluorescence in the VTA/SN: :NAc and VTA/SN: :DMS terminals in the striatum were measured through fiber photometry<sup>26,30</sup> (Fig. 3c,d). In VTA/SN: :NAc and VTA/SN: :DMS terminals expressing gCaMP6f, but not terminals expressing GFP (control virus), there were large fluctuations in the fluorescence signal, reflecting underlying fluctuations in neural activity (Fig. 3d, peak  $dF/F$  of  $0.60 \pm 0.17$  for VTA/SN: :NAc

gCaMP6f terminals,  $n=11$  sites; peak  $dF/F$  of  $0.40 \pm 0.05$  for VTA/SN: NAc gCaMP6f terminals,  $n=11$  sites; and peak  $dF/F$  of  $0.06 \pm 0.001$  for GFP terminals,  $n=4$  sites for DMS and NAc).

To aid in quantitatively interpreting these recordings of calcium dynamics in DA terminals, we compared the dynamics of the gCaMP6f terminal signal with the dopamine concentration ([DA]) in the same striatal location. This was achieved by affixing an optical fiber for measuring DA terminal calcium dynamics (gCaMP6f fluorescence) to a voltammetry electrode for measuring extracellular [DA] (in anesthetized mice), and measuring both signals simultaneously while stimulating DA fibers of passage in the medial forebrain bundle (MFB; experimental schematic in Fig. 3e). Comparison of calcium dynamics and [DA] at an example recording site revealed faster dynamics of gCaMP6f fluorescence relative to [DA] (Fig. 3g; stimulation of 12 pulses at 60 Hz). Similarly, across the population of recording sites, there was a significantly shorter latency to the peak gCaMP6f response compared to [DA] for both the DMS and NAc recording sites (Fig. 3h,  $n=5$  and 4 recording sites for DMS and NAc, respectively; 2-way repeated measures ANOVA, significant effect of recording method,  $F(1,6)=16.36$ ,  $p=0.0067$ , no effect of recording area nor area X method interaction). In addition, the width of the response at half the peak value was significantly shorter in the gCaMP6f recording relative to the [DA] (Fig. 3i; 2-way repeated measures ANOVA revealed a significant effect of recording method,  $F(1,6) = 114.2$ ,  $p=3.96e-5$ , but no effect of recording area nor area X method interaction).

Finally, to assess how the magnitude of the gCaMP6f signal and the [DA] related to the strength of activity in DA neurons, we varied the number of pulses of the MFB phasic electrical stimulation. A comparison of the peak response of the two measurements across the population of recording sites revealed a monotonically increasing relationship between gCaMP6f and [DA] with increasing pulse number, a relationship that was remarkably similar for DMS and NAc recording sites (Fig. 3j). At high pulse numbers, the gCaMP6f response began to saturate relative to [DA], suggesting a saturation in either  $[Ca^{++}]$  in the terminals or saturation in the sensor itself (Fig. 3j,  $n = 4$  and 5 recording sites per condition in NAc and DMS, respectively; peak is normalized relative to largest response for each recording site). Given that a burst of phasic dopamine is typically 2–8 spikes<sup>8,31</sup>, the data suggests a relatively linear relationship between gCaMP6f and [DA] in the relevant dynamic range (Fig. 3j).

### Reward and prediction error encoding in striatal DA terminals

In order to isolate the neural response associated with individual behavioral events (nose poke, lever presentation, lever press, CS+, CS-, reward consumption) while avoiding contamination of neighboring events, we fit a linear regression model in which the measured gCaMP6f trace was modeled as the sum of the convolution of each behavioral event with a response kernel that corresponded to that event (schematic in Fig. 4a, details in Methods). The coefficients of the kernels were chosen to minimize the difference between the modeled and the actual gCaMP6f signal. The kernels for a particular behavioral event can be interpreted as the isolated gCaMP6f response to this event without contamination of other nearby events, under the assumption of linearity in the responses to each event. This

statistical disambiguation of the responses to neighboring events was possible due to the temporal jitter that we had introduced between events in the task, as well as by the variability in each mouse's behavioral latencies.

We first examined the encoding of reward predicting stimuli and reward consumption in VTA/SN: :NAc and VTA/SN: :DMS terminals (recording locations in Fig. 4b). Aligning the Z-scored gCaMP6f response for all trials of all mice to the CS+ or the CS- revealed a positive response to the CS+ and a negative response to the CS- in both the VTA/SN: :DMS and VTA/SN: :NAc terminals (Fig. 4c, left and middle columns) as previously reported in DA neuron cell body recordings, and associated with the anticipatory component of reward prediction error<sup>8</sup>.

The average CS+ response for VTA/SN: :NAc terminals was larger than for VTA/SN: :DMS terminals, which was evident in the single trial heatmaps and the average time-locked response (Fig. 4c,d, left columns; time-locked VTA/SN: :NAc responses 360% higher than VTA/SN: :DMS; all individual recording sites plotted in Supplementary Fig. 3a, left panel). This was further quantified by comparing the VTA/SN: :NAc and VTA/SN: :DMS CS+ kernels derived from the linear model of the gCaMP6f signal (Fig. 4e; individual traces in Supplementary Fig. 3b, left panel). The VTA/SN: :NAc and VTA/SN: :DMS kernels were statistically different from 0.7–2.0s post-CS+ onset (Fig. 4e, left panel; 290% larger peak kernel in VTA/SN: :NAc than VTA/SN: :DMS; shaded region defined as  $p < 0.01$  for two-tailed t-test comparing kernels in 0.1s time bins;  $n=11$  recording sites for both VTA/SN: :DMS and VTA/SN: :NAc).

Similarly, the negative response to the CS- was more pronounced in the single trial heatmaps and time-locked average response for VTA/SN: :NAc relative to VTA/SN: :DMS terminals (Fig. 4c,d, middle column; time-locked VTA/SN: :NAc response 78% more negative on average compared with VTA/SN: :DMS; all individual traces plotted in Supplementary Fig. 3a, middle panel). A statistical comparison between the two populations revealed a statistically significant difference in the kernel from 0.6–1.1s post-CS- onset across the two areas (Fig. 4e, middle panel; minimum kernel is 85% more negative on average in VTA/SN: :NAc than VTA/SN: :DMS; significance defined as  $p < 0.01$  for two-tailed t-test comparing 0.1s time bins;  $n=11$  recordings sites for both VTA/SN: :DMS VTA/SN: :NAc).

In addition, the response to reward consumption was also stronger in the VTA/SN: :NAc terminals relative to the VTA/SN: :DMS terminals. This was evident in the single trial heat maps as well as the time-locked average response (Fig. 4c,d, right panel; VTA/SN: :NAc response 390% larger). In fact, the VTA/SN: :NAc terminals displayed a positive response before and throughout the consumption of reward whereas the VTA/SN: :DMS terminals showed very little consumption responses (Fig. 4e, right panel; 400% larger peak kernel in VTA/SN: :NAc than VTA/SN: :DMS). Indeed, statistically comparing the magnitude of reward consumption kernels revealed a significant difference between 0.3 and 1.6s post consumption (Fig. 4e; significance defined as  $p < 0.01$ , two-tailed t-test comparing 0.1s time bins,  $n=11$  recordings sites for both VTA/SN: :DMS and VTA/SN: :NAc). A residual

response to fully predicted rewards has previously been reported from electrophysiological recordings of putative DA neurons in rodents<sup>32–34</sup>.

A well characterized feature of the electrophysiological recordings of activity in DA neurons in the VTA/SNc is the encoding of a reward prediction error to outcomes, which is thought to serve as a reinforcement signal<sup>2,35,36</sup>. Specifically, the strength of the response to a reward is modulated by reward history such that an unexpected reward leads to a stronger response than expected rewards. To determine if reward prediction error was a feature of VTA/SN: :NAc and VTA/SN: :DMS terminals during this task, we adapted the analysis from Bayer and Glimcher<sup>35</sup>, and created a multiple linear regression model in which the presence or absence of reward on the current and previous trials was used to predict the gCaMP6f response to the conditioned stimulus (CS+ or CS–) on the current trial (response time window of 0.2–1.0s after CS onset, chosen based on the time course of the CS+ and CS– kernels (Fig. 4e). A reward prediction error would involve a positive modulation of the response by the current trial's reward, together with a negative modulation by the rewards on previous trials. (This is because the prediction error takes the form of obtained minus predicted rewards, and reward predictions are expected to arise from a recency-weighted average over rewards obtained on previous trials, as also suggested by behavior, Fig. 1c) This is indeed what we saw, both in VTA/SN: :NAc and VTA/SN: :DMS terminals (Fig. 5), as response to the CS+ was indeed positively modulated by reward on that trial, but negatively modulated by reward on previous trials (Fig. 5,  $p=1.84e-8$ ,  $t(10)=4.11$  and  $p=0.0034$ ,  $t(10)=-3.89$  for current and one-trial back, respectively, for VTA/SN: :NAc;  $p=0.0017$ ,  $t(10)=16.04$  and  $p=0.0025$ ,  $t(10)=-3.81$  for current and one-trial back modulation, respectively, for VTA/SN: :DMS; two-tailed t-test).

### Choice encoding in striatal DA terminals

We next examined how VTA/SN: :NAc or VTA/SN: :DMS neurons terminal activity was modulated by the mouse's choice of action. In the VTA/SN: :NAc recordings, Z-scored gCaMP6f recordings aligned to both the nose poke and lever presentation revealed no consistent response preference across the population on trials with an ipsilateral or contralateral lever choice, as expected for dopamine neurons (Fig. 6a,b; ipsilateral and contralateral are defined relative to the side of the recording). Similarly, a quantitative comparison of the VTA/SN: :NAc response kernels calculated for upcoming ipsilateral and contralateral choices showed no significant difference between responses of ipsilateral vs contralateral choice trials for either nose poke or lever presentation (Fig. 6c; paired t-test, 0.1s time bins,  $p>0.01$ ,  $n=11$  recording sites).

In contrast, for the VTA/SN: :DMS recordings, the Z-scored gCaMP6f response of trials time-locked with the nose poke and lever presentation showed a qualitatively larger response when the lever chosen was contralateral (rather than ipsilateral) to the recording location (Fig. 6d,e; 1200% larger for the nose poke response preceding contralateral vs ipsilateral choices, 144% for lever presentation). Indeed, a quantitative comparison of the response kernels associated with the nose poke and lever presentation preceding contralateral versus ipsilateral choices showed a significantly larger response on contralateral trials between 0.4–0.9s for the nose poke kernel and 0.3–0.5s for the lever presentation kernel (Fig. 6f;

150% larger for contralateral vs ipsilateral nose poke kernel, 57% for lever presentation; significance defined as  $p < 0.01$  for two-tailed t-test comparing 0.1s time bins;  $n = 11$  recordings sites for both VTA/SN: :DMS and VTA/SN: :NAc; individual traces shown in Supplementary Fig 4). In addition, a direct comparison of the extent of contralateral response preference between VTA/SN: :DMS versus VTA/SN: :NAc terminals revealed a significant difference (unpaired two tailed t-test; maximum difference in ipsilateral vs contralateral responses between VTA/SN: :DMS and VTA/SN: :NAc sites;  $p = 0.01$ ,  $t(10) = 2.406$  for nose poke response,  $p = 0.024$ ,  $t(10) = 1.08$  for lever presentation responses)

To control for the possibility that this stronger response for contralateral choices could be related to a movement artifact rather than a true measure of neural activity, we examined the difference in ipsilateral and contralateral encoding in mice that expressed a calcium-insensitive GFP rather than gCaMP6f in DA neurons. Unlike animals expressing gCaMP6f, the nose poke and lever presentation kernels obtained from striatal terminals of the GFP control mice did not show significant modulation (Supplementary Fig. 5,  $n = 4$  recordings sites).

Given the task structure, an animal's choice on one trial is correlated with its choice on the previous trial. To determine if neural responses were better explained by current trial versus previous trial choice, we created a new variant of the regression model which contained interactions between the behavioral event kernels and both current and previous choice. The animal's current choice had a much larger influence on predicting VTA/SN: :DMS gCaMP6f signal than previous choice, indicating that the choice selective activity was reflective of current, and not past, choices (Supplementary Fig. 6).

### Choice encoding in VTA/SN::DMS cell bodies

The preferential response to contralateral choices in VTA/SN: :DMS terminals may be generated by axo-axonal interactions within the striatum, or it may already be present in the VTA/SN: :DMS cell bodies. To distinguish between these two possibilities, we recorded from VTA/SN: :DMS cell bodies in TH: :IRES-Cre mice, selectively targeted by injecting Cre-dependent gCaMP6f virus in the DMS (Fig. 7a,b,c). Similar to the findings in VTA/SN: :DMS terminals (Fig. 6d,e,f), VTA/SN: :DMS cell bodies showed qualitatively larger gCaMP6f responses on contralateral trials to both the nose poke and lever presentation (Fig. 7d,e; average Z-scored gCaMP6f response 900% larger on contralateral trials for nose poke and 64% larger for lever presentation; averages across recording sites;  $n = 7$ ; individual difference traces shown in Supplementary Fig. 7a).

Similarly, the regression kernels for the nose poke and lever presentation were significantly larger for contralateral relative to ipsilateral trials (Fig. 7f; 490% larger for the nose poke kernel preceding contralateral vs ipsilateral choices, 42% for lever presentation; significance defined as  $p < 0.01$  for two-tailed t-test comparing 0.1s time bins;  $n = 7$ , individual traces show in in Supplementary Fig. 7b). These results demonstrate that the choice encoding seen in DAergic terminals in the DMS is also present in the cell bodies of these projecting neurons.



## Discussion

Here, we recorded neural activity from striatal terminal regions of midbrain DA neurons. We found striking differences in the encoding of reward and choice in VTA/SN: :NAC and VTA/SN: :DMS DA terminals, with VTA/SN: :NAC terminals preferentially responding to reward and VTA/SN: :DMS terminals preferentially responding to contralateral (versus ipsilateral) choice. We also found an important commonality in the encoding of the task in the two populations, as both populations represented a reward prediction error, or reinforcement signal.

### Transient dopamine neuron inhibition affects future choice

Our instrumental reversal learning task enabled mice to continually learn which lever to press based on trial and error. The large number of learning trials afforded the statistical power to quantitatively model the effect of transient dopamine neuron inhibition on choice, as well as to determine how neural activity related to reward, choice, and prediction error. Before examining neural correlates of the task in DA subpopulations, we first confirmed that task performance depended on DA activity by performing optogenetic inhibition of activity throughout a trial on a randomly selected subset of trials (Fig. 2). As expected, future choice was affected in a manner consistent with dopamine neuron inhibition functioning as a negative reward prediction error. Our quantitative modeling of behavior revealed a preserved differential effect of rewarded and unrewarded choices on future choice after dopamine neuron inhibition, but a shifted baseline, in that the mice had a tendency to avoid previous choices associated with dopamine neuron inhibition. Note that this effect appeared to be specific to choice, as there was no change in latency to initiate the next trial after dopamine neuron inhibition. This specificity may be due to the fact that choice is easier to perturb than initiation latency in our task, given that mice are trained to adjust their choices perpetually based on reward feedback. We did not observe an effect of inhibiting dopamine neurons for only a subset of time within a trial (inhibiting either before or after the lever press; Supplementary Fig. 1g). This negative result may be caused by rebound excitation after dopamine transient neuron inhibition canceled the effect of inhibition. Alternatively, it may be that inhibition for only a subset of a trial is not a strong enough manipulation to perturb behavior, especially in light of the likely incomplete optogenetic inhibition throughout the VTA/SNc occurring in highly scattering brain tissue.

### Preferential responses to reward in VTA/SN::NAC terminals

A striking result in this study is the greater strength of responses to reward-predictive cues and reward consumption in VTA/SN: :NAC terminals relative to VTA/SN: :DMS terminals (Fig. 4c,d,e). In both areas, the stimulus that predicted reward (CS+) lead to an increase in the calcium signal, and the stimulus that predicted absence of reward (CS-) lead to a decrease in the calcium signal. The magnitude of the response to both stimuli was much larger in the VTA/SN: :NAC terminals. In addition, a positive reward consumption response was evident in the VTA/SN: :NAC terminals, while if anything the VTA/SN: :DMS terminals exhibited a depression during reward consumption. These results support the notion that the VTA/SN: :NAC terminals are more involved in reward processing than are VTA/SN: :DMS terminals.

Similarly, reward prediction errors were evident in both populations, and consistent with the finding that there was more reward-related responses in VTA/SN: :NAc, the magnitude was larger in that population (Fig. 5). Reward prediction error was quantified based on a multiple linear regression model in which the response to the reward predicting cue (CS+ or CS-) on the current trial was predicted based on the reward outcome of current and previous trials<sup>35</sup> (Fig. 5). In this analysis, the signature of a reward prediction error is a positive coefficient in the regression model for the current trial outcome along with negative coefficients for the previous trial outcomes, with more recent previous trials weighted more heavily. This represents a reward prediction error in that reward on previous trials decreases the strength of the response to reward on the current trial. In particular, if reward predictions are learned via the Rescorla-Wagner rule, then the regression coefficients for earlier trials' outcomes should decline exponentially<sup>27,35</sup>. The multiple linear regression model revealed a profile consistent with these expectations in both VTA/SN: :NAc and VTA/SN: :DMS terminal recordings (Fig. 5). Thus, these data support the view that both DAergic projections encode a reward prediction error, or reinforcement signal, and are consistent with recent evidence that DA terminals in both dorsal and ventral striatum are sufficient to drive learning<sup>37</sup>.

### **Preferential contralateral responses in VTA/SN::DMS terminals**

Perhaps the most novel aspect of this dataset is the discovery that contralateral choices drive consistently stronger responses than ipsilateral choices in VTA/SN: :DMS cell bodies and terminals (Fig. 6,7). In other words, neural activity is greater preceding a contralateral versus an ipsilateral movement (lever press), supporting a role for this projection not only in learning and motivation, but also in generating or expressing the movement that the mouse has chosen. There was significantly more contralateral response preference in the VTA/SN: :DMS relative to the VTA/SN: :NAc terminals, although there appeared to be individual VTA/SN: :NAc recording sites that preferentially encode one choice or the other (Supplementary Fig. 4).

To our knowledge, preferential responses for contralateral choices have not been previously reported in DA neurons. This could be due to the fact that previous recordings have not isolated responses in VTA/SN: :DMS neurons. In fact, electrophysiology experiments have often identified DA neurons in part based on a response to unexpected reward, which may have biased against recording from these neurons, given the relatively weak responses to reward in this subpopulation (Fig. 4). Despite the fact that contralateral response preference has not been previously reported in DA neurons, the finding is consistent with the extensive literature demonstrating lateralized motor deficits as a result of unilateral manipulation of DA neurons<sup>6</sup>.

Given that the dorsal striatum is involved in action selection and generation, the finding of contralateral choice preference of the VTA/SN: :DMS terminal responses suggests that DA neurons that project to that region are specialized to support its function. One possibility is that DA does not only affect movement indirectly through an effect on learning or motivation, but is involved directly in signaling movement choice and even controlling movement execution (e.g. via effects on medium spiny neurons in the contralateral striatum). Alternatively, it may be that contralateral choice selectivity arises as part of some specialized

prediction error signal appropriate to the dorsal striatum. For instance, while the standard prediction error is scalar and unitary, more sophisticated computational models have sometimes invoked vector-valued prediction errors with separate components specific to different actions, effectors, or situations<sup>38–40</sup>.

### **Prior evidence of spatial heterogeneity in striatal dopamine**

Previous electrophysiology recordings of putative midbrain DA neurons in monkeys have related neural responses to the location of the recording electrode. Given the topography between the VTA/SNc and the dorsal/ventral striatum, the medial/lateral axis of the VTA/SNc should roughly distinguish between DA neurons projection to ventral versus dorsal striatum. Using this approach, one study reported more reward encoding in putative DA neurons in the VTA compared to SNc<sup>41</sup>, while other studies have reported a more homogeneous distribution<sup>8,42,43</sup>. This lack of consensus may arise from the challenge of determining the precise anatomical location of the recording electrode, coupled with the imperfect topography in the VTA/SNc projection to the striatum.

In addition, neurochemical approaches have been employed to measure [DA] in specific subregions of the striatum during behavior<sup>36,44–46</sup>. Although most voltammetry studies have focused on the NAc<sup>36,46–48</sup>, there have also been recordings in the dorsal striatum, and the results are consistent with our finding that reward responses are stronger in the ventral than in the dorsal striatum. For example, reward-related [DA] transients appear after more extensive behavioral training in the dorsal striatum relative to the ventral striatum<sup>49,50</sup>. Similarly, ramping patterns of [DA] as a rat approaches a reward location are more prominent in the ventromedial compared to the dorsolateral striatum<sup>45</sup>. Our result that there is differential encoding of reward and choice in the two terminal regions both confirms and extends these results. In particular, the presence of features that are encoded more prominently in VTA/SN: :DMS terminals (contralateral choice preference) suggests that the observation of weaker reward encoding in VTA/SN: :DMS terminals is not related to technical difficulties in obtaining DA signals in the more dorsal striatal regions, and is indeed a fundamental organizational feature of striatal DA.

### **Conclusions**

In summary, given that the NAc is thought to be preferentially involved in reward processing while the DMS is thought to be preferentially involved in action selection and generation, the relatively strong reward-related responses in VTA/SN: :NAc terminals and the presence of contralateral choice selectivity in VTA/SN: :DMS terminals suggests that the DA innervation of striatal subregions is specialized to contribute the specific function of the target region. The concept that the DA innervation of these striatal subregions contributes to their specialized function is novel, and compliments the major current view of striatal DA as encoding a reward prediction error signal to support learning.

### **Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

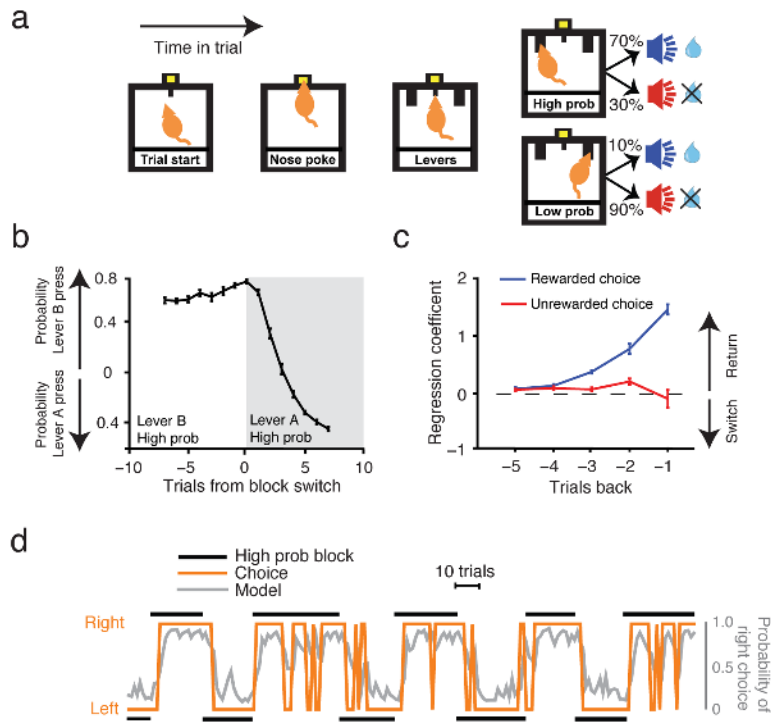
We thank C. Gregory, M. Applegate and J. Finkelstein for assistance in data collection; J. Pillow and A. Conway for advice with data analysis; M. Murugan and B. Engelhard for comments on the manuscript; and D. Tindall and P. Wallace for administrative support. This work was supported by the Pew, McKnight, NARSAD and Sloan Foundations and an NIH DP2 New Innovator Award and an R01 MH106689-02 (IBW), an NSF Graduate Research Fellowship (NFP), and the Essig and Enright `82 Fund (JFT, IBW).

## References

1. Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H. Midbrain dopamine neurons encode decisions for future action. *Nat. Neurosci.* 2006; 9:1057–1063. [PubMed: 16862149]
2. Schultz W, Dayan P, Montague PR. A Neural Substrate of Prediction and Reward. *Science.* 1997; 275:1593–1599. [PubMed: 9054347]
3. Starr BS, Starr MS. Differential effects of dopamine D1 and D2 agonists and antagonists on velocity of movement, rearing and grooming in the mouse: Implications for the roles of D1 and D2 receptors. *Neuropharmacology.* 1986; 25:455–463. [PubMed: 3488514]
4. Wise RA. Dopamine, learning and motivation. *Nat. Rev. Neurosci.* 2004; 5:483–494. [PubMed: 15152198]
5. Marshall JF, Berrios N. Movement disorders of aged rats: reversal by dopamine receptor stimulation. *Science.* 1979; 206:477–479. [PubMed: 504992]
6. Arbuthnott GW, Crow TJ. Relation of contraversive turning to unilateral release of dopamine from the nigrostriatal pathway in rats. *Exp. Neurol.* 1971; 30:484–491. [PubMed: 5554236]
7. Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* 1998; 1:304–309. [PubMed: 10195164]
8. Schultz W. Predictive Reward Signal of Dopamine Neurons. *J. Neurophysiol.* 1998; 80:1–27. [PubMed: 9658025]
9. Witten IB, et al. Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron.* 2011; 72:721–733. [PubMed: 22153370]
10. Tsai H-C, et al. Phasic Firing in Dopaminergic Neurons Is Sufficient for Behavioral Conditioning. *Science.* 2009; 324:1080–1084. [PubMed: 19389999]
11. Nestler EJ, Carlezon WA Jr. The Mesolimbic Dopamine Reward Circuit in Depression. *Biol. Psychiatry.* 2006; 59:1151–1159. [PubMed: 16566899]
12. Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.).* 2006; 191:507–520. [PubMed: 17031711]
13. Nicola SM, Surmeier DJ, Malenka RC. Dopaminergic Modulation of Neuronal Excitability in the Striatum and Nucleus Accumbens. *Annu. Rev. Neurosci.* 2000; 23:185–215. [PubMed: 10845063]
14. Graybiel AM. The basal ganglia. *Curr. Biol.* 2000; 10:R509–R511. [PubMed: 10899013]
15. Domesick VB. Neuroanatomical Organization of Dopamine Neurons in the Ventral Tegmental Area. *Ann. N. Y. Acad. Sci.* 1988; 537:10–26. [PubMed: 3059917]
16. Lammel S, Ion DI, Roeser J, Malenka RC. Projection-Specific Modulation of Dopamine Neuron Synapses by Aversive and Rewarding Stimuli. *Neuron.* 2011; 70:855–862. [PubMed: 21658580]
17. Lynd-Balta E, Haber SN. The organization of midbrain projections to the striatum in the primate: Sensorimotor-related striatum versus ventral striatum. *Neuroscience.* 1994; 59:625–640. [PubMed: 7516506]
18. Lammel S, et al. Unique Properties of Mesoprefrontal Neurons within a Dual Mesocorticolimbic Dopamine System. *Neuron.* 2008; 57:760–773. [PubMed: 18341995]
19. Lerner TN, et al. Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell.* 2015; 162:635–647. [PubMed: 26232229]
20. Tai L-H, Lee AM, Benavidez N, Bonci A, Wilbrecht L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat. Neurosci.* 2012; 15:1281–1289. [PubMed: 22902719]
21. Balleine BW, Delgado MR, Hikosaka O. The Role of the Dorsal Striatum in Reward and Decision-Making. *J. Neurosci.* 2007; 27:8161–8165. [PubMed: 17670959]

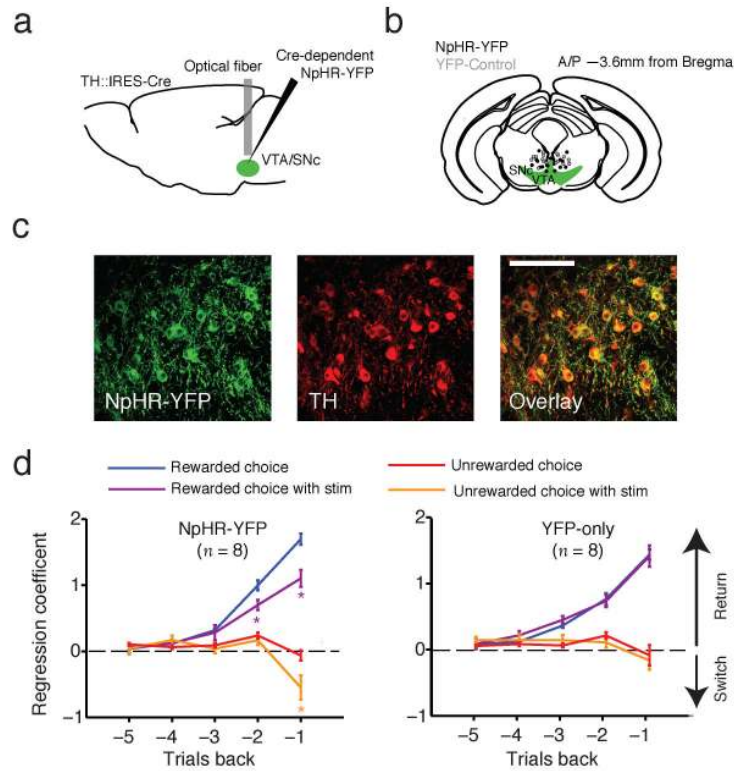
22. Graybiel AM, Aosaki T, Flaherty AW, Kimura M. The basal ganglia and adaptive motor control. *Science*. 1994; 265:1826–1831. [PubMed: 8091209]
23. Roitman MF, Wheeler RA, Carelli RM. Nucleus Accumbens Neurons Are Innately Tuned for Rewarding and Aversive Taste Stimuli, Encode Their Predictors, and Are Linked to Motor Output. *Neuron*. 2005; 45:587–597. [PubMed: 15721244]
24. Ikemoto S, Panksepp J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Res. Rev.* 1999; 31:6–41. [PubMed: 10611493]
25. Chen T-W, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*. 2013; 499:295–300. [PubMed: 23868258]
26. Gunaydin LA, et al. Natural Neural Projection Dynamics Underlying Social Behavior. *Cell*. 2014; 157:1535–1551. [PubMed: 24949967]
27. Lau B, Glimcher PW. Value Representations in the Primate Striatum during Matching Behavior. *Neuron*. 2008; 58:451–463. [PubMed: 18466754]
28. Lau B, Glimcher PW. Dynamic Response-by-Response Models of Matching Behavior in Rhesus Monkeys. *J. Exp. Anal. Behav.* 2005; 84:555–579. [PubMed: 16596980]
29. Lammel S, et al. Diversity of Transgenic Mouse Models for Selective Targeting of Midbrain Dopamine Neurons. *Neuron*. 2015; 85:429–438. [PubMed: 25611513]
30. Cui G, et al. Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature*. 2013; 494:238–242. [PubMed: 23354054]
31. Hyland BI, Reynolds JNJ, Hay J, Perk CG, Miller R. Firing modes of midbrain dopamine cells in the freely moving rat. *Neuroscience*. 2002; 114:475–492. [PubMed: 12204216]
32. Pan W-X, Schmidt R, Wickens JR, Hyland BI. Dopamine Cells Respond to Predicted Events during Classical Conditioning: Evidence for Eligibility Traces in the Reward-Learning Network. *J. Neurosci.* 2005; 25:6235–6242. [PubMed: 15987953]
33. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*. 2012; 482:85–88. [PubMed: 22258508]
34. Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.* 2007; 10:1615–1624. [PubMed: 18026098]
35. Bayer HM, Glimcher PW. Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron*. 2005; 47:129–141. [PubMed: 15996553]
36. Day JJ, Roitman MF, Wightman RM, Carelli RM. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* 2007; 10:1020–1028. [PubMed: 17603481]
37. Ilango A, et al. Similar Roles of Substantia Nigra and Ventral Tegmental Dopamine Neurons in Reward and Aversion. *J. Neurosci.* 2014; 34:817–822. [PubMed: 24431440]
38. Gershman SJ, Pesaran B, Daw ND. Human Reinforcement Learning Subdivides Structured Action Spaces by Learning Effector-Specific Values. *J. Neurosci.* 2009; 29:13524–13531. [PubMed: 19864565]
39. O'Reilly R, Frank M. Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Comput.* 2006; 18:283–328. [PubMed: 16378516]
40. Diuk C, Tsai K, Wallis J, Botvinick M, Niv Y. Hierarchical Learning Induces Two Simultaneous, But Separable, Prediction Errors in Human Basal Ganglia. *J. Neurosci.* 2013; 33:5797–5805. [PubMed: 23536092]
41. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*. 2009; 459:837–841. [PubMed: 19448610]
42. Mirenowicz J, Schultz W. Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* 1994; 72:1024–1027. [PubMed: 7983508]
43. Mirenowicz J, Schultz W. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*. 1996; 379:449–451. [PubMed: 8559249]

44. Stefani MR, Moghaddam B. Rule learning and reward contingency are associated with dissociable patterns of dopamine activation in the rat prefrontal cortex, nucleus accumbens, and dorsal striatum. *J. Neurosci. Off. J. Soc. Neurosci.* 2006; 26:8810–8818.
45. Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature.* 2013; 500:575–579. [PubMed: 23913271]
46. Phillips PEM, Stuber GD, Heien MLAV, Wightman RM, Carelli RM. Subsecond dopamine release promotes cocaine seeking. *Nature.* 2003; 422:614–618. [PubMed: 12687000]
47. Hart AS, Rutledge RB, Glimcher PW, Phillips PEM. Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. *J. Neurosci.* 2014; 34:698–704. [PubMed: 24431428]
48. Hamid AA, et al. Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* 2016; 19:117–126. [PubMed: 26595651]
49. Willuhn I, Burgeno LM, Everitt BJ, Phillips PEM. Hierarchical recruitment of phasic dopamine signaling in the striatum during the progression of cocaine use. *Proc. Natl. Acad. Sci.* 2012; 109:20703–20708. [PubMed: 23184975]
50. Brown HD, McCutcheon JE, Cone JJ, Ragozzino ME, Roitman MF. Primary food reward and reward-predictive stimuli evoke different patterns of phasic dopamine signaling throughout the striatum. *Eur. J. Neurosci.* 2011; 34:1997–2006. [PubMed: 22122410]



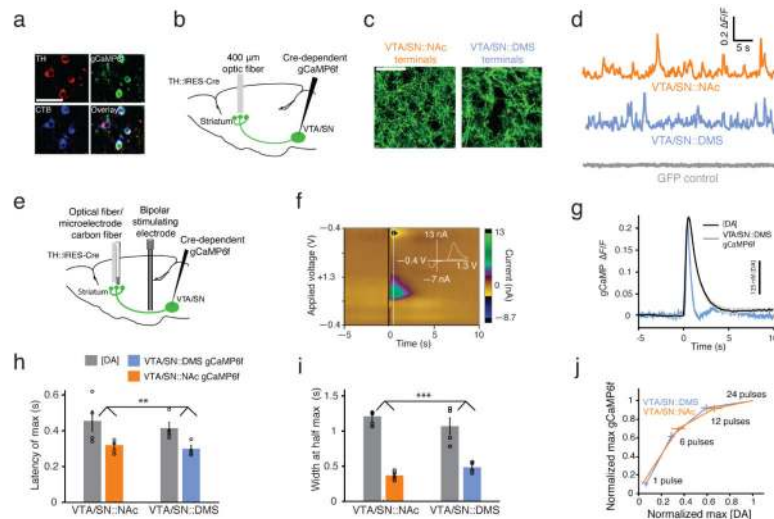
**Figure 1. Mice continually learn which choice to make based on recent experience**

(a) A trial starts with the illumination of a central noseport (“Trial Start”). As a consequence of entering the central nose port (“Nose poke”), the mouse is presented with two levers (“Levers”). Pressing one lever results in reward with high probability (70%; “High prob”) and the other lever results in reward with low probability (10%; “Low prob”). The identity of the high reward lever reverses on a pseudorandom schedule. (b) The average probability of pressing the high or low probability lever relative to the trial in which the identity of the high probability lever is switched. (c) Regression coefficients from a logistic regression model to predict each mouse’s choice on a given trial based on previous trial choice and outcome. Two sets of predictors were used: “rewarded choice” on previous trials (red line), which identifies a choice as a rewarded right lever choice (+1), rewarded left lever choice (–1), or unrewarded (0), and “unrewarded choice” on previous trials (blue line), which identifies a choice as unrewarded right lever choice (+1), unrewarded left lever choice (–1), or rewarded (0). Each predictor was included in the model, with a shift of 1–5 trials preceding the choice being predicted (x-axis represents trials between predicted choice and previous choice/outcome that was used to predict it). Together, these predictors unambiguously identify the previous choices and outcome combinations for the 5 trials preceding any choice. Positive regression coefficients for the two predictors correspond to greater likelihood to returning to a lever choice that resulted in reward, or no reward, respectively. (d) Example of one mouse’s behavior for 200 trials. Black bar signifies the identity of the high probability lever. The solid orange trace is the mouse’s choice on each trial, and the grey trace is the choice predicted by the behavioral model. **b, c** Error bars represent SEM across animals ( $n=8$ ).



**Figure 2. Inhibiting DA neurons in the VTA/SNc alters an animal's choice on future trials**  
**(a)** Surgical schematic. The optical fiber implant (grey rectangle) and injection site (black needle) in the VTA/SNc (green) are shown. **(b)** Coronal section with optical fiber tip locations for mice injected with AAV5 DIO-EF1a-eNpHR3.0-YFP (black points) or DIO-EF1a-YFP (grey points). **(c)** Confocal images of the VTA. Expression of NpHR-YFP (left panel), anti-*tyrosine hydroxylase* (TH) staining (middle panel) and an overlay of the two images (right panel) demonstrate co-localization. Scale bar: 100  $\mu$ m. Representative image from one animal; similar results were seen across 16 animals **(d)** Coefficients from a logistic regression model demonstrating the influence of VTA/SNc cell body inhibition on lever choice in subsequent trials in NpHR-YFP and YFP-control mice. A negative coefficient indicates a reduction in the return probability to the lever chosen on the previous trial. Conversely, a positive coefficient indicates that the animal is more likely to return to the previously chosen lever. Rewarded choices with stimulation decreased the probability of returning to the chosen lever in comparison to rewarded choices without stimulation in NpHR-YFP mice (left panel;  $p = 0.01$ ,  $t(7)=3.39$  for 1 trial back,  $p=0.04$ ,  $t(7)=2.45$  for 2 trials back; two-tailed t-test comparing coefficients of “rewarded choice” in blue with “rewarded choice+rewarded choice  $\times$  stim” in purple). Likewise, unrewarded choice with stimulation significantly decreased the probability of returning to the chosen lever compared to unrewarded choice alone (left panel;  $p=0.01$ ,  $t(7)=3.352$  for 1 trial back, two-tailed t-test comparing “unrewarded choice” in red with “unrewarded choice+unrewarded choice  $\times$  stim” in orange). In contrast, no effect of stimulation on future choice was observed in the YFP-control animals (right panel;  $p>0.1$ , two-tailed t-tests). Error bars represent SEM across animals ( $n=8$  for each panel).





### Figure 3. Calcium recordings in terminals of striatally-projecting DA neurons

(a) Confocal image of gCaMP6f expression in VTA/SNc, along with *tyrosine hydroxylase* (TH) immunoreactivity and CTB labeling (retrograde tracer injected into the DMS). Scale bar: 50  $\mu$ m. Representative image from one animal; similar results obtained in 2 mice. (b) Surgical schematic for gCaMP6f terminal measurements; AAV5 FLEX-CAG-gCaMP6f viral injection in the VTA/SNc and optical fiber placement in the striatum. (c) gCaMP6f expression in DA axon terminals of the NAc (left panel) and DMS (right panel) subregions of the striatum (confocal images of sectioned tissue). Scale bar: 40  $\mu$ m. Representative image from one animal; similar results were seen in 11 animals (d) Example gCaMP6f dF/F fluctuations in awake, behaving mice in VTA/SN: NAc terminals (top, orange trace) and VTA/SN: DMS terminals (middle, blue trace). In contrast, control recordings taken from mice injected with a AAV5 FLEX-CAG-GFP (bottom, grey trace) show no modulation. Representative traces from three animals, similar traces seen in 11 animals for gCaMP6f recordings, 4 from GFP controls (e) Schematic of simultaneous gCaMP6f and voltammetry recordings. (f) Color representation of cyclic voltammetric data at one DMS site in response to electrical stimulation of the MFB (average of 5 trials; stimulation at 0s). The y-axis is applied voltage, the x-axis is time, and current measured at the carbon fiber electrode is plotted in false color. Inset: cyclic voltammogram at 0.5s following electrical stimulation is consistent with the chemical signature for DA. (g) Traces represent [DA] (black) and gCaMP6f (blue) over time for the data represented in B (average of 5 trials; stimulation at 0s). (h) Latency between stimulation onset and peak amplitude was longer for DA terminal gCaMP6f responses compared to [DA] transients in both DMS and NAc (2-way repeated measures ANOVA revealed significant effect of recording method,  $F(1,6)=16.36$ ,  $p=0.006$ , and not recording site nor site X method interaction). (i) The width of the signal at half the maximum peak was significantly larger for [DA] than for gCaMP6f responses in VTA/SN: DMS (grey and blue) and VTA/SN: NAc (grey and orange) (2-way repeated measures ANOVA revealed a significant effect of recording method,  $F(1,6) = 114.2$ ,  $p < 4 \times 10^{-5}$ , but no effect of recording area nor area X method interaction). (j) The relationship between peak gCaMP6f fluorescence and peak [DA] in NAc and DMS for 60Hz stimulation with increasing numbers of pulses (1, 6, 12, 24). In **h**, **i** and **j** data from 5 recording sites in the

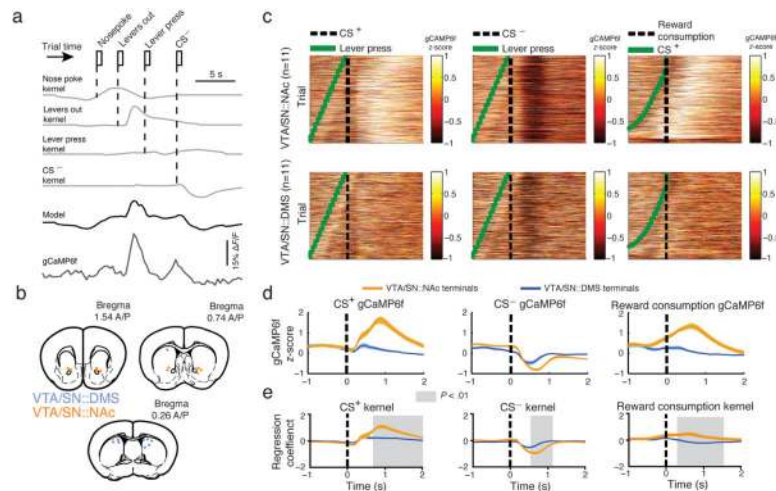
DMS and 4 sites from the NAc (across 3 mice). In **f, g, h & i**, electrical stimulation of MFB was 12 pulses of 60Hz stimulation. Error bars represent SEM in all panels.

Author Manuscript

Author Manuscript

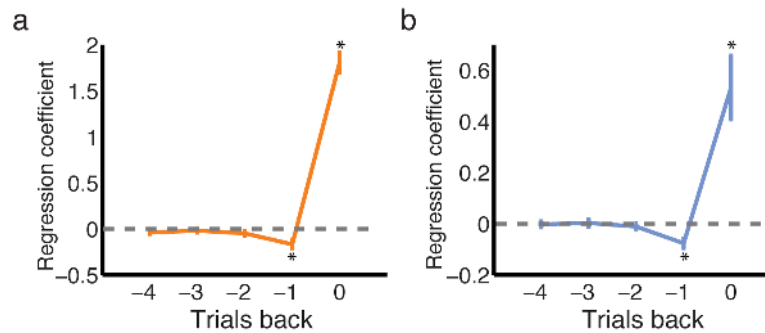
Author Manuscript

Author Manuscript



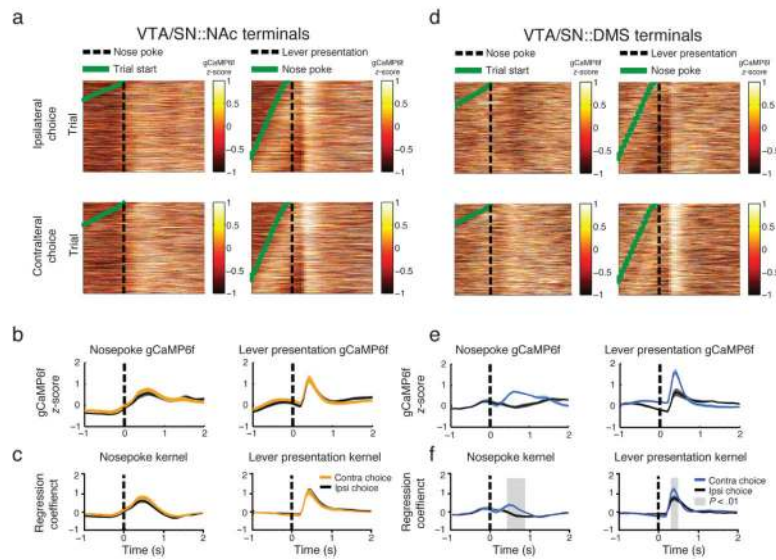
**Figure 4. Response to reward consumption and to reward-predictive cues dominates in VTA/SN:NAc relative to VTA/SN:DMS terminals**

(a) Schematic of linear model used to predict gCaMP6f signal from task events. Each behavioral event is depicted as rectangular tick in the top row. Each kernel is timeshifted based on the time of the corresponding events (grey) and then summed to generate the modeled gCaMP6f trace (black trace), which correlates with the recorded gCaMP6f trace (bottom grey trace). (b) Coronal sections of the mouse brain showing location of optic fibers used for gCaMP6f terminal recordings. VTA/SN:NAc and VTA/SN:DMS recordings indicated by orange and blue, respectively. (c) Heatmaps represent Z-score of all single trial gCaMP6f recordings time locked to the CS+ (left), CS- (middle), or onset of reward consumption (right), for recordings in VTA/SN:NAc (top row) and VTA/SN:DMS (bottom row) terminals. All recorded trials from all recording sites are represented, and the trials are sorted based on the time of lever press (for left and middle heatmaps), or the time of the CS+ (for the rightmost heatmap). (d) Average gCaMP6f Z-score across all recording sites time-locked to CS+ (left), CS- (middle) and Reward Consumption (right,  $n=11$  recording sites taken from 11 animals in both VTA/SN:NAc and VTA/SN:DMS). Orange traces are from VTA/SN:NAc recording sites; blue from VTA/SN:DMS. (e) Average kernel across all recording sites corresponding to CS+ (left), CS- (middle), and Reward Consumption (right;  $n=11$  recording sites in both VTA/SN:NAc and VTA/SN:DMS). Kernels are coefficients from a linear regression model where each behavioral event is convolved with a corresponding kernel and all convolved traces are summed to predict the measured gCaMP6f signal, as schematized in b. Two tailed t-test comparing VTA/SN:NAc and VTA/SN:DMS kernels in 0.1s bins reveals statistically significant difference between 0.7–2s post-CS+ onset for CS+ kernel (left), 0.6–1.1s post-CS- onset for CS- kernel, and 0.3–1.6s post-reward consumption onset for reward consumption kernel ( $p < .01$ , significance represented by grey region). In d, e error bars represent SEM across sites.



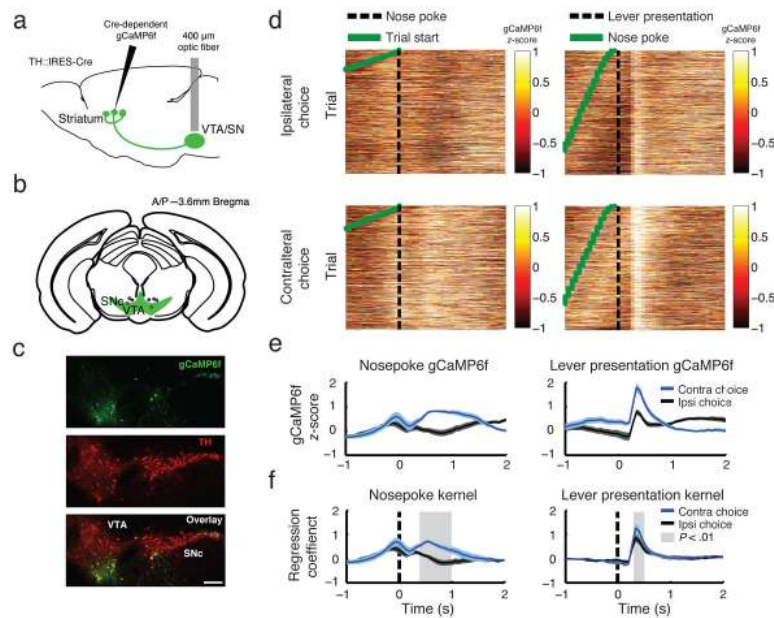
**Figure 5. Evidence of reward prediction error encoding in both VTA/SN: :NAc and VTA/SN: :DMS terminals**

(a) The CS response in VTA/SN: :NAc terminal recordings (0.2-1.0s post CS onset, time range selected based on the CS+ and CS- kernels in Figure 4e) is predicted with a multiple linear regression with the reward outcome of the current trial (labeled trial 0) and previous trial (labeled with negative numbers) used as predictors. The average regression coefficient across mice is plotted for current and previous trials. The positive coefficient for the current trial and negative coefficients for previous trials indicates the encoding of a reward prediction error. (b) Same as a, but for VTA/SN: :DMS terminal recordings. All error bars represent SEM across mice;  $n=11$  recording sites from 11 animals in each panel; t-test, \*  $p<0.01$ .



### Figure 6. Contralateral response preference in VTA/SN: :DMS terminals

(a) Heatmaps represent Z-score of gCaMP6f recordings time-locked to the nose poke (left) or lever presentation (right), for VTA/SN: :NAc terminals. Heat maps shows all trials preceding either an ipsilateral lever press choice (top) or a contralateral lever press choice (bottom). (Ipsilateral and contralateral lever presses are defined relative to the recording site in the brain.) All recorded trials from all recording sites are represented ( $n=11$  recording sites), and the trials are sorted based on the time of illumination of the trial start light in the noseport (for left two columns), or the time of the nosepoke (for right two columns). (b) Average Z-Scored gCaMP6f signal time-locked to the nose poke (left panel) and lever presentation (right panel) for ipsilateral (black) or contralateral (orange) trials in VTA/SN: :NAc recording sites (averages taken across mice,  $n=11$ ). (c) Nose poke and lever presentation kernels are derived from the same gCaMP6f statistical model as described previously (Figure 4a), but in this case the events preceding an ipsilateral lever choice versus a contralateral lever choice are considered separately. No significant difference in the ipsilateral and the contralateral kernels ( $p>0.01$  for all time bins, two-tailed t-test comparing ipsilateral vs contralateral VTA/SN: :NAc kernels in 100ms bins). (d) Same as a but for VTA/SN: :DMS terminal recordings ( $n=11$  recording sites). (e) Same as b but for VTA/SN: :DMS recording sites ( $n=11$ ). (f) Same as c but for VTA/SN: :DMS recording sites ( $n=11$  recording sites). Two tailed t-test comparing ipsilateral and contralateral kernels in 0.1s bins revealed a significant difference between 0.4-0.9s for contralateral vs ipsilateral nose poke kernels (left) and 0.3-0.5s for contralateral vs ipsilateral lever presentation kernel (threshold of  $p<0.01$  for shading). (Error bars in b, c, e, f represent SEM across sites).



**Figure 7. Contralateral response preference in VTA/SN: *DMS* cell bodies**

(a) Surgical schematic. Cre-dependent gCaMP6f was injected into terminals of the DMS (black needle) and expression in VTA/SN: *DMS* cell bodies (green) was recorded through an optical fiber (grey rectangle). (b) Coronal sections of the mouse brain showing location of optical fiber recording locations (grey circles) (c) Confocal images of the VTA/SN. Expression of gCaMP6f (top panel), anti- *tyrosine hydroxylase* (TH) staining (middle panel) and an overlay of the two images (bottom panel) demonstrates retrograde travel following injection in the DMS. Scale bar: 200  $\mu$ m. Representative image from one animal; similar results were seen across 4 animals (d) Heatmaps represent Z-score of gCaMP6f recordings time-locked to the nose poke (left) or lever presentation (right), for VTA/SN: *DMS* cell bodies. Heat maps shows all trials preceding either an ipsilateral lever press choice (top) or a contralateral lever press choice (bottom). All recorded trials from all recording sites are represented ( $n=7$  recording sites), and the trials are sorted based on the time of illumination of the trial start light in the noseport (for left column), or the time of the nose poke (for right column). (e) Average gCaMP6f Z-score time-locked to the nose poke (left panel) and lever presentation (right panel) for ipsilateral or contralateral trials in VTA/SN: *DMS* neurons (average across  $n=7$  recording sites). (f) Nose poke (left) and lever presentation (right) kernels for ipsilateral and contralateral trials derived from the statistical model described in Figure 4a. Significant difference between ipsilateral and contralateral kernels from 0.4-0.9s after time of nose poke and 0.4–0.6s after lever presentation (two-tailed t-test, 100ms bins, threshold of  $p<0.01$  for shading). (Error bars in e, f represent SEM across sites).