

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

Nathan F Parker¹, Courtney M Cameron¹, Joshua P Taliaferro¹, Junuk Lee¹, Jung Yoon Choi¹, Thomas J Davidson², Nathaniel D Daw¹ & Ilana B Witten¹

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 whether and how DA neurons are specialized on the basis of 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.

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^{1–4}, 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 because of its role in movement disorders in humans (for example, Parkinson's disease) and the dramatic motor phenotypes following DA perturbations in animals^{3,5,6}. However, seminal studies 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^{7,8}, which in theory could be used to learn which stimuli or actions are likely to lead to reward². Consistent with this idea, phasic activation of DA neurons is sufficient to support learning^{9,10}. 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^{11,12} or affecting the balance between activity in the output pathways of the striatum^{13,14}.

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 and SNc receive different inputs and project topographically to different striatal subregions^{15–19}. These striatal subregions themselves have specialized anatomical and functional organizations, with dorsal regions, such as the dorsomedial striatum (DMS), implicated in evaluating and generating actions^{20–22} and ventral regions, such as the nucleus accumbens (NAc), implicated in processing reward^{23,24}.

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 neurons might contribute to movement not merely as an indirect consequence of learning or motivation; 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^{25,26} 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

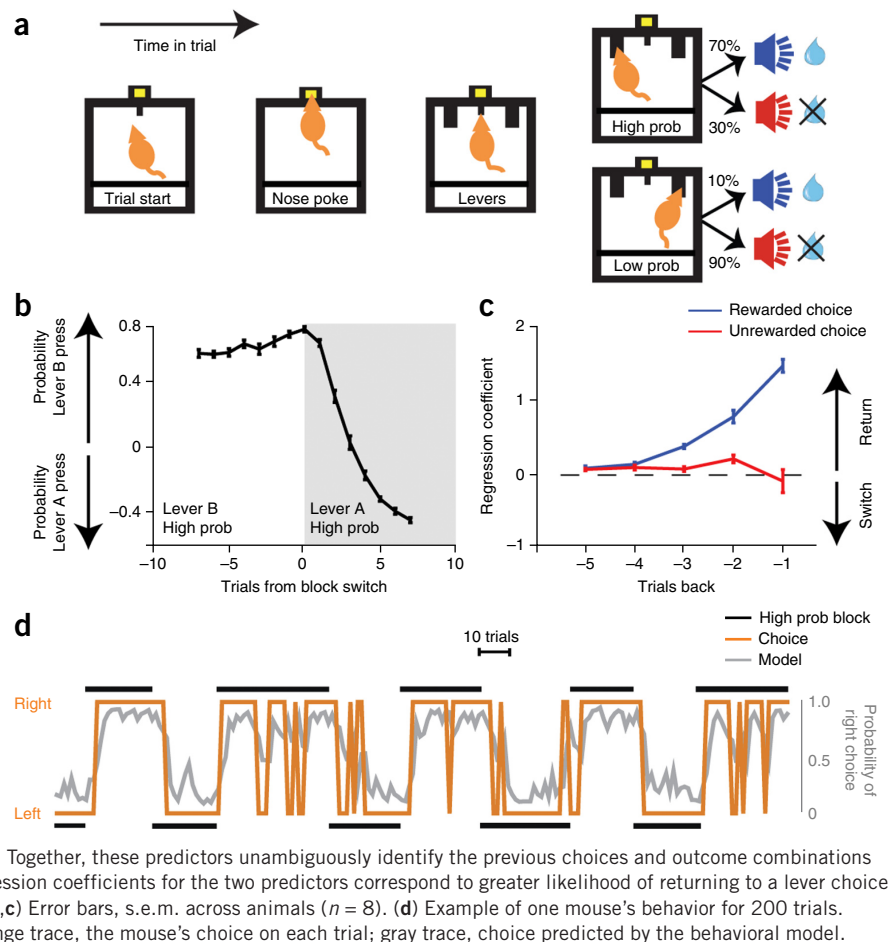
Mice were trained to perform an instrumental reversal learning task (**Fig. 1a**). The start of each trial was signified by the illumination of a central nose port. After the mice entered the nose port (“nose poke”), two levers were presented (“lever presentation”). One lever corresponded to a high probability 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 ten rewarded trials. Lever presses that resulted in a reward were followed by one auditory stimulus (positive conditioned stimulus, or CS+) and lever

¹Princeton Neuroscience Institute and Department of Psychology, Princeton University, Princeton, New Jersey, USA. ²Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, California, USA. Correspondence should be addressed to I.B.W. (iwitten@princeton.edu).

Received 8 December 2015; accepted 14 March 2016; published online 25 April 2016; doi:10.1038/nn.4287

Figure 1 Mice continually learn which choice to make on the basis of recent experience.

(a) A trial starts with the illumination of a central nose port ("Trial start"). As a consequence of entering the central nose port ("Nose poke"), the mouse is presented with two 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. The blue and red icons represent an auditory tone used as a positive and negative conditioned stimulus (CS+ and CS−), respectively. (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 predicting each mouse's choice on a given trial on the basis of previous trial choice and outcome. Two sets of predictors were used: 'rewarded choice' on previous trials (blue 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 (red 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 the predicted choice and the previous choice/outcome that was used to predict it). Together, these predictors unambiguously identify the previous choices and outcome combinations for the five trials preceding any choice. Positive regression coefficients for the two predictors correspond to greater likelihood of returning to a lever choice that resulted in reward, or no reward, respectively. (b,c) Error bars, s.e.m. across animals ($n = 8$). (d) Example of one mouse's behavior for 200 trials. Black bars, identity of the high probability lever; orange trace, the mouse's choice on each trial; gray trace, choice predicted by the behavioral model.



presses that were not associated with reward were followed by a different auditory stimulus (CS−). A temporal jitter between 0 and 1 s 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 on the basis of 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) on the basis of previous trial choices and outcomes^{20,27,28}. 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 action's values, which ultimately determines the subsequent choices²⁸. 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 (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 whether 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 (Supplementary Fig. 1a). Mice expressing Cre recombinase under the control of the tyrosine hydroxylase (*Th*) promoter (TH::IRES-Cre mice) received injections in the VTA of an adeno-associated virus (AAV) expressing either Cre-dependent NpHR-YFP (halorhodopsin from *Natronomonas* with yellow fluorescent protein) or YFP only (control virus), as well as bilateral optical fiber implants above each VTA (Fig. 2a–c). To confirm the efficacy of NpHR-mediated inhibition in tyrosine hydroxylase-expressing (TH⁺) neurons, in a separate group of mice, we performed whole cell recordings in brain slices and observed large photocurrents (689 ± 16 pA) and effective elimination of current-evoked spiking (Supplementary Fig. 1b–d). After mice were trained on the behavioral task (Fig. 1a), we bilaterally inhibited DA neurons on a randomly selected 10% of trials throughout the duration of a trial (from the initial nose poke until either the end of 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 Figure 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, $n = 8$ 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 that the effect (like that of rewards) was mediated by error-driven learning. In contrast, light

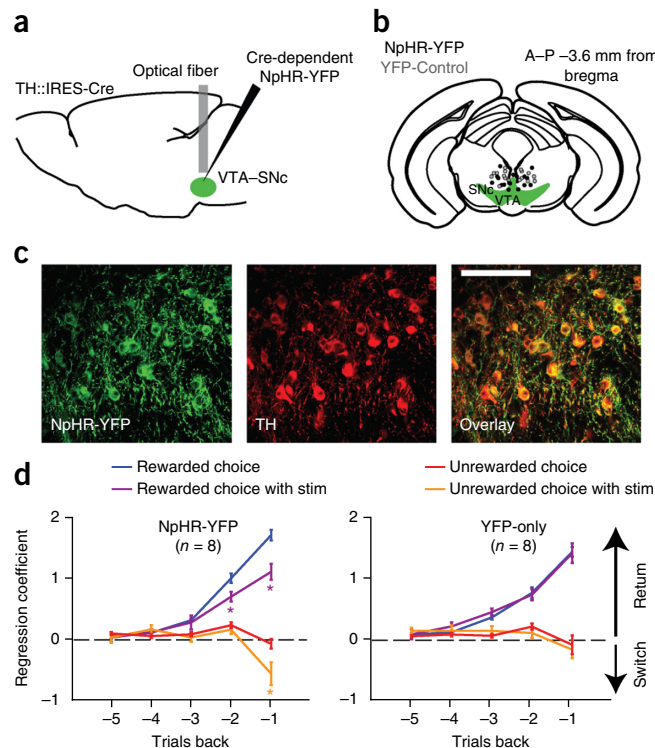
Figure 2 Inhibiting DA neurons in the VTA–SNc alters an animal's choice on future trials. (a) Surgical schematic showing the optical fiber implant (gray rectangle) and injection site (black needle) in the VTA–SNc (green). (b) Coronal section with anterior–posterior (A–P) optical fiber tip locations for mice injected with AAV5 DIO-EF1 α -eNpHR3.0-YFP (black points) or DIO-EF1 α -YFP (gray points). (c) Confocal images of the VTA, showing expression of NpHR-YFP (left), anti-tyrosine hydroxylase (TH) staining (middle panel) and an overlay of the two images (right) demonstrating co-localization. Scale bar, 100 μ 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 probability of returning to the lever chosen on the previous trial; a positive coefficient indicates an increase. Rewarded choices with stimulation (stim) decreased the probability of returning to the chosen lever in comparison to rewarded choices without stimulation in NpHR-YFP mice (left; $P = 0.01$, $t_{(7)} = 3.39$ for one trial back, $P = 0.04$, $t_{(7)} = 2.45$ for two trials back; two-tailed t -test comparing coefficients across animals 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; $P = 0.01$, $t_{(7)} = 3.352$ for one 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; $P = 0.85$, $t_{(7)} = 0.202$, effect of stimulation on rewarded choice one trial back; $P = 0.41$, $t_{(7)} = 0.876$, effect of stimulation on unrewarded choice one trial back; two-tailed t -tests). Error bars represent s.e.m. across animals ($n = 8$ for each panel, $*P < 0.05$).

had no discernible effect on behavior in YFP-control littermates (Fig. 2d, $n = 8$ YFP-only mice; choice $\times$ light $\times$ group were 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 on the basis of previous rewarded and unrewarded choice, light and group, and all interactions thereof). Given recent reports of Cre expression in neurons that do not express TH near the VTA of TH::IRES-Cre mice²⁹, we replicated the same experiment in mice expressing Cre recombinase under control of the dopamine transporter (*Dat*) promoter (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 help determine an animal's choice on future trials in this task and are 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 ± 0.17 s after stimulation versus 9.93 ± 0.13 s after control trials). This indicates that our manipulation is relatively specific for choice. To determine whether 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 2 s 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

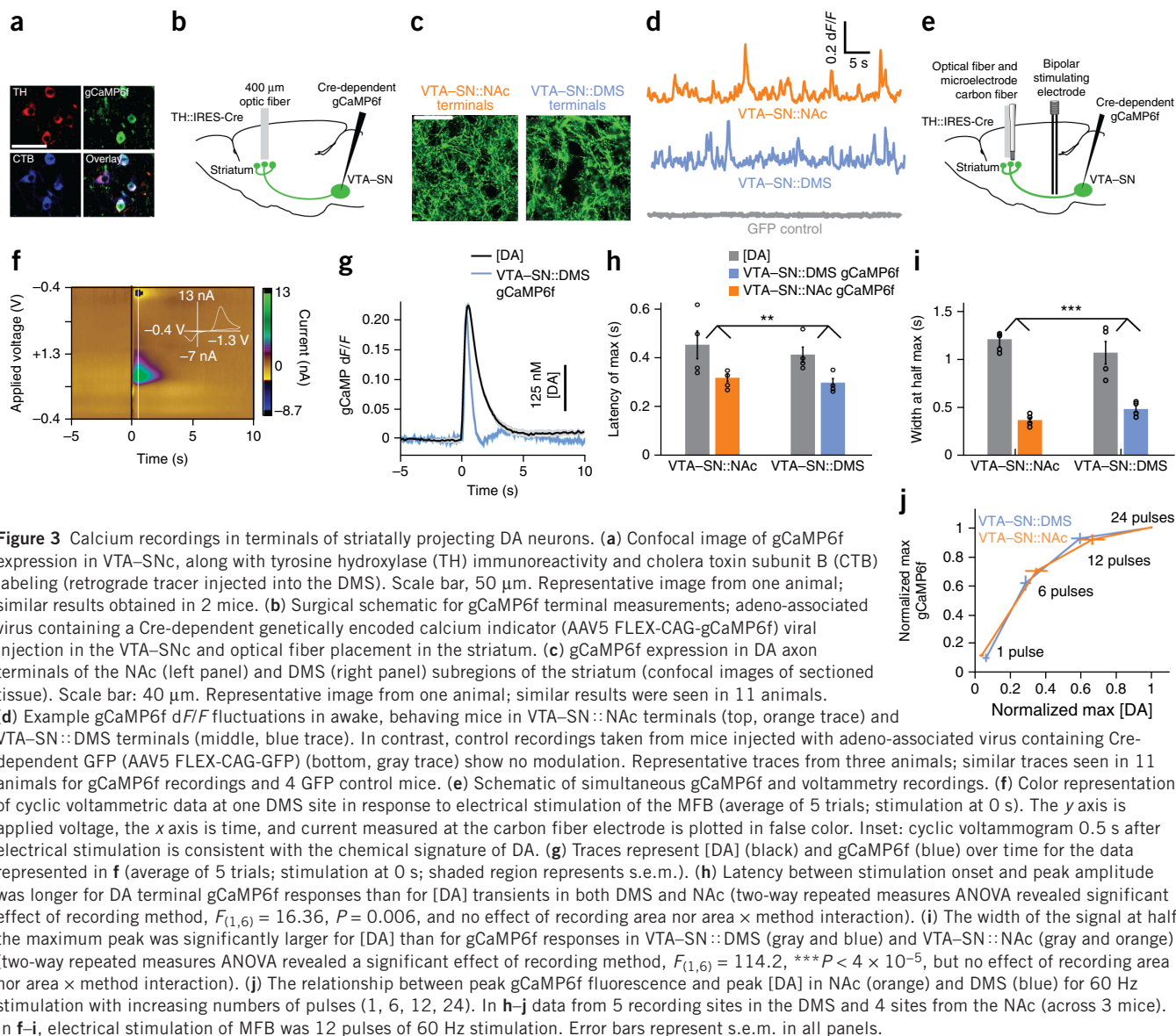
After confirming that DA neuron activity in the VTA–SNc affected task performance, we employed a genetically encoded calcium indicator to measure and compare neural activity in the terminals of DA neurons originating in the VTA–SNc and projecting to either the dorsomedial striatum ("VTA–SN::DMS") or the nucleus



accumbens core ("VTA–SN::Nac") during the reversal learning task. We analyzed retrograde tracer injections in the two striatal regions and found 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).

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. 3; 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 (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).

We measured fluctuations in fluorescence in the VTA–SN::Nac and VTA–SN::DMS terminals in the striatum through fiber photometry^{26,30} (Fig. 3c,d). In VTA–SN::Nac and VTA–SN::DMS terminals expressing gCaMP6f, but not terminals expressing GFP (control virus), we found large fluctuations in the fluorescence signal, reflecting underlying fluctuations in neural activity (Fig. 3d, change in fluorescence relative to baseline (dF/F) of 0.60 ± 0.17 for VTA–SN::Nac gCaMP6f terminals, $n = 11$ sites; peak dF/F of 0.40 ± 0.05



for VTA-SN::NAc gCaMP6f terminals, $n = 11$ sites; and peak dF/F of 0.06 ± 0.001 for GFP terminals, $n = 4$ striatal sites).

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 by affixing the 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; 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, we found a significantly shorter latency to the peak gCaMP6f response across the population of recording sites compared to latency to the peak [DA] for both the DMS and NAc recording sites (Fig. 3h, $n = 5$ and $n = 4$ recording sites for DMS and NAc, respectively; two-way repeated measures ANOVA showed a significant effect of recording method, $F_{(1,6)} = 16.36$, $P = 0.0067$ but no effect of recording area nor of area–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; two-way repeated measures ANOVA revealed a significant effect of recording method, $F_{(1,6)} = 114.2$, $P = 3.96 \times 10^{-5}$, but no effect of recording area nor of area–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 calcium concentration in the terminals or saturation in the sensor itself (Fig. 3j, $n = 4$ recording sites per condition in NAc and $n = 5$ recording sites per condition in DMS; peak was normalized relative to the largest response for each recording site). Given that a burst of phasic dopamine is typically 2–8 spikes^{8,31}, 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

To isolate the neural response associated with individual behavioral events (nose poke, lever presentation, lever press, CS+, CS−, reward consumption) while avoiding contamination from neighboring events, we fitted a linear regression model in which we modeled the measured gCaMP6f trace as the sum of the convolution of each behavioral event with a response kernel that corresponded to that event (Fig. 4a; details in Online Methods). We chose the coefficients of the kernels to minimize the difference between the modeled and actual gCaMP6f signal. The kernels for a particular behavioral event can be interpreted as the isolated gCaMP6f response to that event without contamination from other nearby events, assuming linearity in the responses to each event. This statistical disambiguation of the responses to neighboring events was enabled by 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 (Fig. 4b). Aligning the z-scored gCaMP6f responses 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), showing that these responses are associated with the anticipatory component of reward prediction error, as previously reported in DA neuron cell body recordings⁸.

The average CS+ response for VTA-SN::Nac terminals was larger than for VTA-SN::DMS terminals, which was evident in the single-trial heat maps and the average time-locked response (Fig. 4c,d and Supplementary Fig. 3a; time-locked VTA-SN::Nac responses were 360% higher than VTA-SN::DMS). 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 and

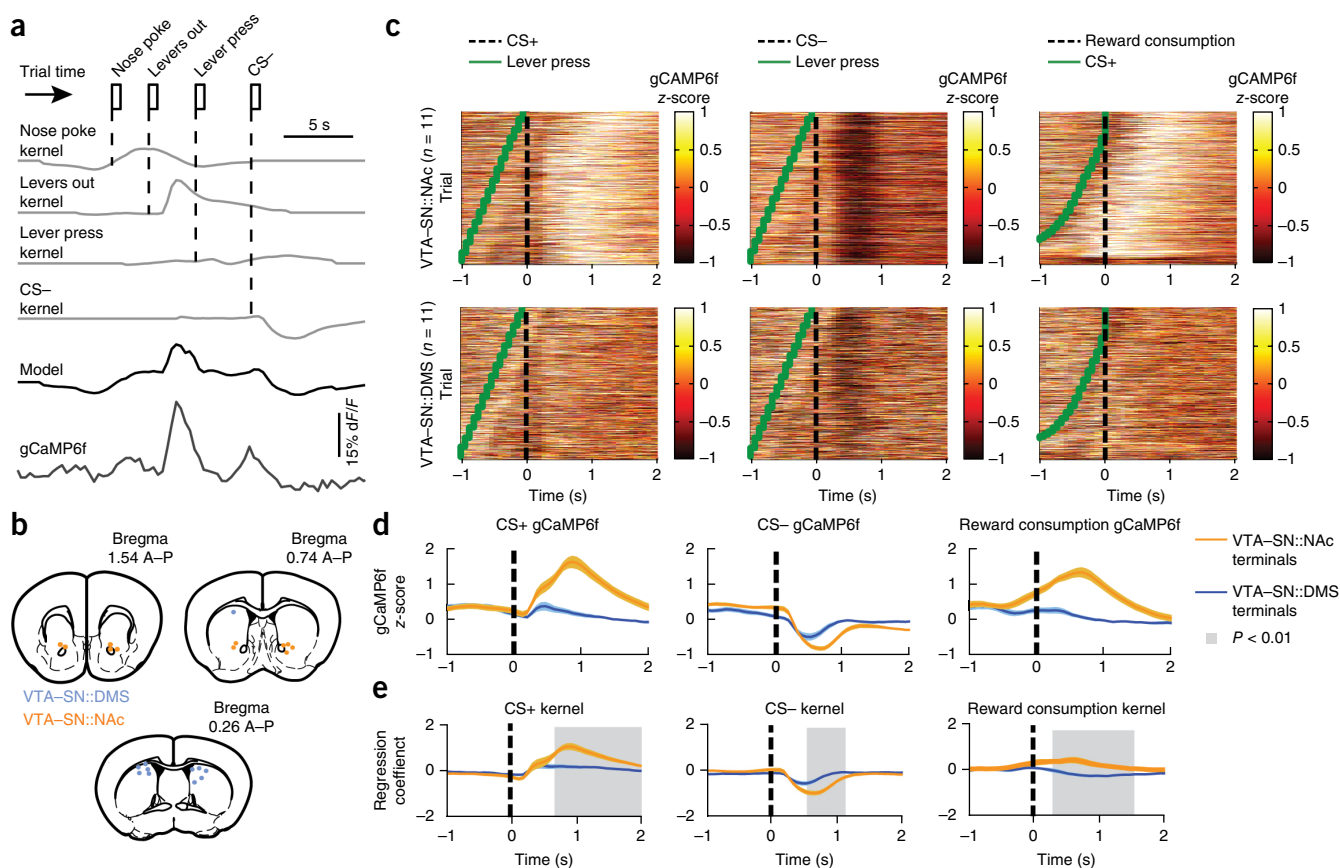
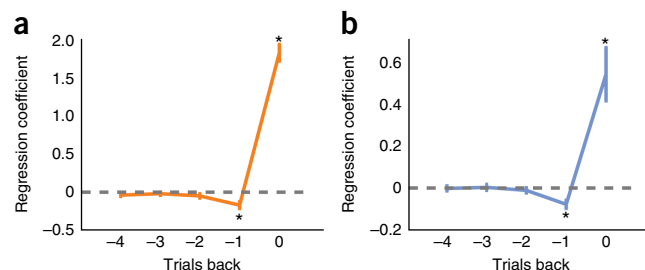


Figure 4 Response to reward consumption and to reward-predictive cues dominates in VTA-SN::Nac relative to VTA-SN::DMS terminals.

(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 time-shifted according to the time of the corresponding events (light gray traces) and then summed to generate the modeled gCaMP6f trace (black trace), which correlates with the recorded gCaMP6f trace (dark gray trace). (b) Coronal sections of the mouse brain showing location of optic fibers used for gCaMP6f terminal recordings. Orange and blue indicate VTA-SN::Nac and VTA-SN::DMS recordings, respectively. (c) Heat maps represent the 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 on the basis of the time of lever press (for left and middle heat maps) or the time of the CS+ (for the rightmost heat map).

(d) Average gCaMP6f z-score across all recording sites time-locked to CS+ (left), CS− (middle) and reward consumption (right; $n = 11$ recording sites from 11 animals for both VTA-SN::Nac and VTA-SN::DMS). Orange traces are from VTA-SN::Nac recording sites; blue, VTA-SN::DMS. (e) Average kernel across all recording sites corresponding to CS+ (left), CS− (middle), and reward consumption (right; $n = 11$ recording sites for 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 a. Two-tailed t -test comparing VTA-SN::Nac and VTA-SN::DMS kernels in 0.1 s bins reveals statistically significant difference between 0.7–2.0 s post-CS+ onset for CS+ kernel (left), 0.6–1.1 s post-CS− onset for CS− kernel, and 0.3–1.6 s post-reward consumption onset for reward consumption kernel ($P < 0.01$, significance represented by gray region). In d,e error bars represent s.e.m. 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 (0.2–1.0 s post CS onset, time range selected on the basis of the CS+ and CS– kernels in **Fig. 4e**) is predicted with a multiple linear regression with the reward outcome of the current trial (labeled trial 0) and the previous trial(s) (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 indicate the encoding of a reward prediction error. **(b)** Same as **a**, but for VTA-SN::DMS terminal recordings. All error bars represent s.e.m. across mice; $n = 11$ recording sites from 11 animals in each panel; t -test, $*P < 0.01$.



Supplementary Fig. 3b). The VTA-SN::Nac and VTA-SN::DMS kernels were statistically different from 0.7 to 2.0 s post-CS+ onset (**Fig. 4e**; peak kernel in VTA-SN::Nac was 290% larger than in VTA-SN::DMS; $n = 11$ 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 heat maps and time-locked average response for VTA-SN::Nac relative to VTA-SN::DMS terminals (**Fig. 4c,d** and **Supplementary Fig. 3a**; minimum time-locked VTA-SN::Nac response was 78% more negative on average compared with VTA-SN::DMS). A statistical comparison between the two populations revealed a statistically significant difference in the kernel from 0.6–1.1 s post-CS– onset across the two areas (**Fig. 4e**; minimum kernel was 85% more negative on average in VTA-SN::Nac than in VTA-SN::DMS; significance defined as $P < 0.01$ for two-tailed t -test comparing 0.1-s time bins; $n = 11$ recording sites for both VTA-SN::DMS and 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**; VTA-SN::Nac response was 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 small consumption responses (**Fig. 4e**; peak kernel in VTA-SN::Nac was 400% larger than in VTA-SN::DMS). Indeed, statistically comparing the magnitude of reward consumption kernels revealed a significant difference between 0.3 and 1.6 s post-consumption (**Fig. 4e**; significance defined as $P < 0.01$, two-tailed t -test comparing 0.1 s time bins, $n = 11$ recording sites for both VTA-SN::DMS and VTA-SN::Nac). Previous studies using electrophysiological recordings of putative DA neurons in rodents have also reported a residual response to fully predicted rewards^{32–34}.

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 may serve as a reinforcement signal^{2,35,36}. Specifically, the strength of the response to a reward

Figure 6 Contralateral response preference in VTA-SN::DMS terminals. **(a)** Heat maps represent the z-score of gCaMP6f recordings time-locked to the nose poke (left) or lever presentation (right), for VTA-SN::Nac terminals. Heat maps show 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 on the basis of the time of illumination of the trial start light in the nose port (for left column), or the time of the nose poke (for right column). **(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, $n = 11$ recording sites). **(c)** Nose poke and lever presentation kernels are derived from the same gCaMP6f statistical model as described in **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 versus contralateral VTA-SN::Nac kernels in 100 ms 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 0.1 s bins revealed a significant difference (threshold of $P < 0.01$ for shading) between 0.4–0.9 s for contralateral versus ipsilateral nose poke kernels (left) and 0.3–0.5 s for contralateral versus ipsilateral lever presentation kernel. Error bars in **b,c,e,f** represent s.e.m. across sites.

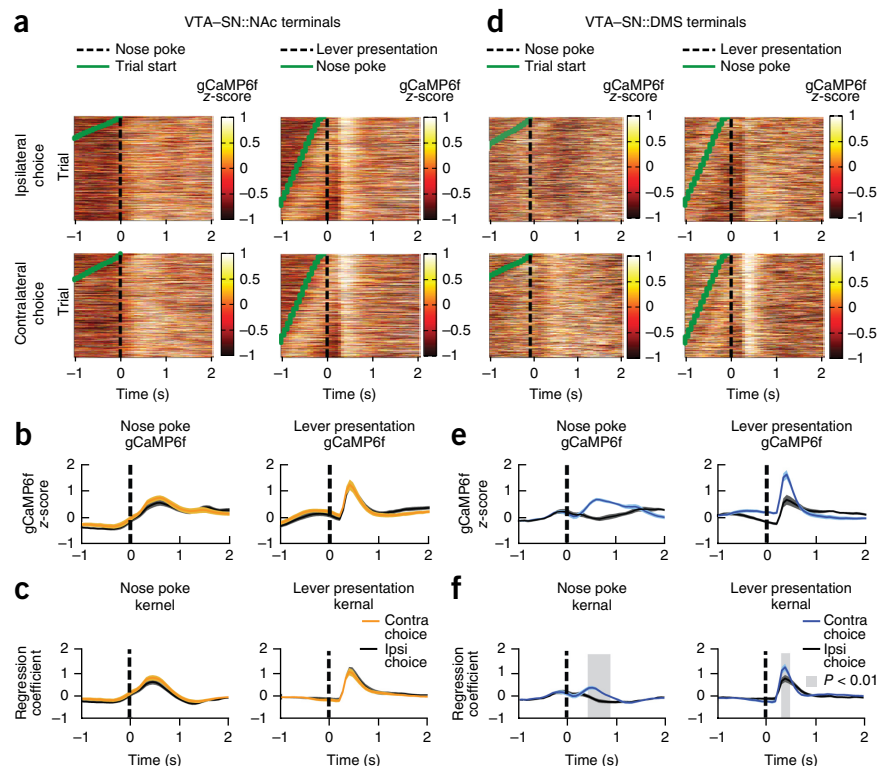
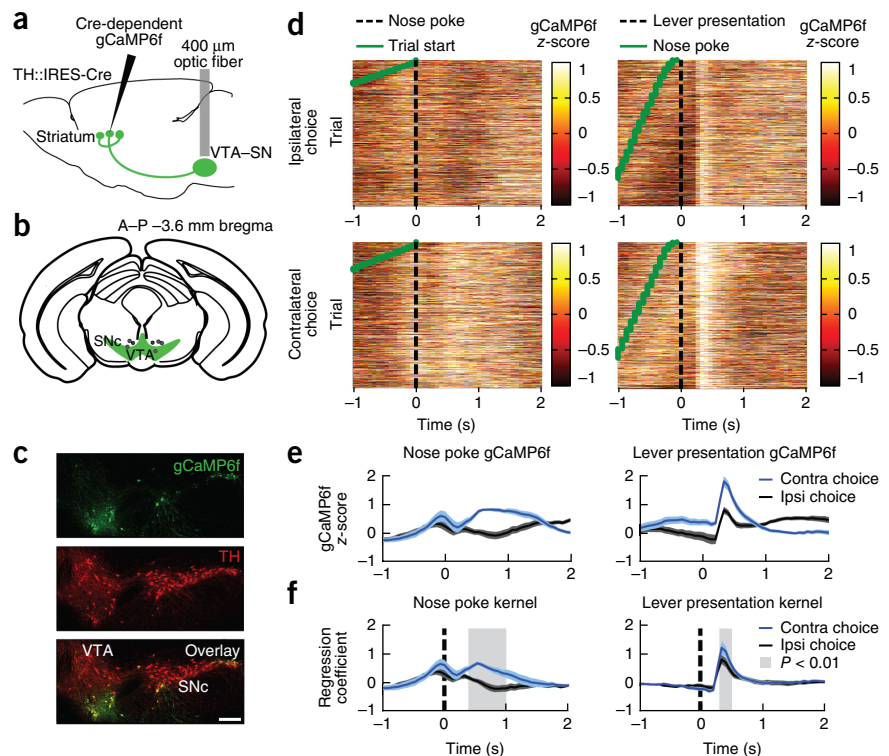


Figure 7 Contralateral response preference in VTA–SN::DMS cell bodies. (a) Surgical schematic. Cre-dependent gCaMP6f was injected into terminals of the DMS (black needle) and expression in VTA–SN::DMS cell bodies (green) was recorded through an optical fiber (gray rectangle). (b) Coronal sections of the mouse brain showing location of optical fiber recording locations (gray circles). (c) Confocal images of the VTA and 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 μ m. Representative image from one animal; similar results were seen across 4 animals. (d) Heat maps represent the z-score of gCaMP6f recordings time-locked to the nose poke (left) or lever presentation (right), for VTA–SN::DMS cell bodies. Heat maps show 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 on the basis of the time of illumination of the trial start light in the nose port (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. Gray shading shows a significant difference between ipsilateral and contralateral kernels from 0.4–0.9 s after time of nose poke and 0.4–0.6 s after lever presentation (two-tailed t -test, 100-ms bins, threshold of $P < 0.01$). Error bars in e,f represent s.e.m. across sites.



is modulated by reward history such that an unexpected reward leads to a stronger response than expected rewards. To determine whether 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 Glimcher³⁵ and created a multiple linear regression model in which we used the presence or absence of reward on the current and previous trials to predict the gCaMP6f response to the conditioned stimulus (CS+ or CS–) on the current trial (response time window of 0.2–1.0 s after CS onset, chosen on the basis of 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: response to the CS+ was positively modulated by reward on that trial, but negatively modulated by reward on previous trials (Fig. 5, $P = 1.84 \times 10^{-8}$, $t_{(10)} = 4.11$ and $P = 0.0034$, $t_{(10)} = -3.89$ for current and one-trial-back modulation, 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 terminal activity in VTA–SN::NAc or VTA–SN::DMS neurons 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 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 versus contralateral choice trials for either nose poke or lever presentation (Fig. 6c; paired t -test, 0.1 s 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; 1,200% larger peak response for the nose poke preceding contralateral versus ipsilateral choices, 144% larger response for lever presentation preceding contralateral versus ipsilateral choices). 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 and 0.9 s for the nose poke kernel and between 0.3 and 0.5 s for the lever presentation kernel (Fig. 6f; 150% larger response for contralateral versus ipsilateral nose poke kernel, 57% larger response for lever presentation; significance defined as $P < 0.01$ for two-tailed t -test comparing 0.1 s time bins; $n = 11$ recordings sites for both VTA–SN::DMS and VTA–SN::NAc; 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 versus 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 the nose poke and lever presentation kernels obtained from striatal terminals of animals expressing gCaMP6f, the nose poke and lever presentation kernels from striatal terminals of the GFP control mice did not show significant modulation (**Supplementary Fig. 5**, $P > 0.01$ for all timepoints, t -test comparing contralateral and ipsilateral kernels in 0.1-s time bins; $n = 4$ recordings sites).

Given the task structure, an animal's choice on one trial is correlated with its choice on the previous trial. To determine whether neural responses were better explained by current trial versus previous trial choice, we created a new variant of the regression model that 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 its 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 into the DMS (**Fig. 7a–c**). As in our findings in VTA-SN::DMS terminals (**Fig. 6d–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**; the average z-scored gCaMP6f response was 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 versus ipsilateral choices, 42% larger for lever presentation; significance defined as $P < 0.01$ for two-tailed t -test comparing 0.1 s time bins; $n = 7$, individual traces shown in **Supplementary Fig. 7b**). These results demonstrate that the choice encoding seen in dopaminergic 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 task encoding 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 on the basis of 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 optogenetically inhibiting activity throughout a trial on a randomly selected subset of trials (**Fig. 2**). As expected, we found that 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 relative to control trials, only with a shifted baseline, such 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 continually on the basis of 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 transient dopamine neuron inhibition cancelling 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 considering that optogenetic inhibition throughout the VTA-SNc was likely incomplete as a result of light scattering in neural 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–e**). In both areas, the stimulus that predicted reward (CS+) led to an increase in the calcium signal and the stimulus that predicted absence of reward (CS−) led to a decrease in the calcium signal. The magnitudes of responses to both stimuli were 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 the magnitude of such responses was also larger in VTA-SN::Nac, consistent with our finding of more reward-related responses in that population (**Fig. 5a,b**). Reward prediction error was quantified on the basis of a multiple linear regression model in which the response to the reward-predicting cue (CS+ or CS−) on the current trial was predicted on the basis of the reward outcomes of current and previous trials³⁵ (**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^{27,35}. 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 dopaminergic 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 learning³⁷.

Preferential contralateral responses in VTA–SN::DMS terminals

We found that contralateral choices drive consistently stronger responses than ipsilateral choices in VTA–SN::DMS cell bodies and terminals (Figs. 6 and 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 a greater 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 because previous recordings have not isolated responses in VTA–SN::DMS neurons. In fact, electrophysiology experiments have often identified DA neurons in part by response to unexpected reward, which may have selected against recording from these neurons, given this subpopulation's relatively weak responses to reward (Fig. 4). Our finding of contralateral response preference is consistent with the extensive literature demonstrating lateralized motor deficits as a result of unilateral manipulation of DA neurons⁶.

Given that the dorsal striatum is involved in action selection and generation, the contralateral choice preference we found in VTA–SN::DMS terminal responses suggests that DA neurons projecting to that region are specialized to support its function. One possibility is that DA not only affects movement indirectly through an effect on learning or motivation, but is also directly involved in signaling movement choice and even controlling movement execution (for example, via effects on medium spiny neurons in the contralateral striatum). Alternatively, it may be that contralateral choice selectivity arises as part of a 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^{38–40}.

Prior evidence of spatial heterogeneity in striatal dopamine

Previous electrophysiological 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 and ventral striatum, the medial–lateral axis of the VTA–SNc should roughly distinguish between DA neurons projecting to ventral versus dorsal striatum. Using this approach, one study reported more reward encoding in putative DA neurons in the VTA compared to SNc⁴¹, while other studies have reported a more homogeneous distribution^{8,42,43}. 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 VTA–SNc projections to the striatum.

In addition, neurochemical approaches have been employed to measure [DA] in specific subregions of the striatum during behavior^{36,44–46}. Although most voltammetry studies have focused on the NAc^{36,46–48}, other studies have recorded in the dorsal striatum, and their 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^{49,50}. Similarly, ramping patterns of [DA] as a rat approaches a reward location are more prominent in the ventromedial compared to the dorsolateral striatum⁴⁵. Our finding of 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, but 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 suggest that the DA innervation of striatal subregions is specialized to contribute to the specific function of the target region. This new concept, that the DA innervation of these striatal subregions contributes to their specialized function, complements the major current view of striatal DA as encoding a reward prediction error signal to support learning.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

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. I.B.W. was supported by the Pew, McKnight, NARSAD and Sloan Foundations, NIH DP2 New Innovator Award and an R01 MH106689-02; N.E.P. was supported by an NSF Graduate Research Fellowship; and J.P.T. and I.B.W. were supported by the Essig and Enright '82 Fund.

AUTHOR CONTRIBUTIONS

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

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Morris, G., Nevet, A., Arkadir, D., Vaadia, E. & Bergman, H. Midbrain dopamine neurons encode decisions for future action. *Nat. Neurosci.* **9**, 1057–1063 (2006).
- Schultz, W., Dayan, P. & Montague, P.R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Starr, B.S. & Starr, M.S. 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* **25**, 455–463 (1986).
- Wise, R.A. Dopamine, learning and motivation. *Nat. Rev. Neurosci.* **5**, 483–494 (2004).
- Marshall, J.F. & Berrios, N. Movement disorders of aged rats: reversal by dopamine receptor stimulation. *Science* **206**, 477–479 (1979).
- Arbuthnott, G.W. & Crow, T.J. Relation of contraversive turning to unilateral release of dopamine from the nigrostriatal pathway in rats. *Exp. Neurol.* **30**, 484–491 (1971).

7. Hollerman, J.R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1**, 304–309 (1998).
8. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
9. Witten, I.B. *et al.* Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* **72**, 721–733 (2011).
10. Tsai, H.-C. *et al.* Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* **324**, 1080–1084 (2009).
11. Nestler, E.J. & Carlezon, W.A. Jr. The mesolimbic dopamine reward circuit in depression. *Biol. Psychiatry* **59**, 1151–1159 (2006).
12. Niv, Y., Daw, N.D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191**, 507–520 (2007).
13. Nicola, S.M., Surmeier, J. & Malenka, R.C. Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annu. Rev. Neurosci.* **23**, 185–215 (2000).
14. Graybiel, A.M. The basal ganglia. *Curr. Biol.* **10**, R509–R511 (2000).
15. Domesick, V.B. Neuroanatomical organization of dopamine neurons in the ventral tegmental area. *Ann. NY Acad. Sci.* **537**, 10–26 (1988).
16. Lammel, S., Ion, D.I., Roeper, J. & Malenka, R.C. Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron* **70**, 855–862 (2011).
17. Lynd-Balta, E. & Haber, S.N. The organization of midbrain projections to the striatum in the primate: sensorimotor-related striatum versus ventral striatum. *Neuroscience* **59**, 625–640 (1994).
18. Lammel, S. *et al.* Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. *Neuron* **57**, 760–773 (2008).
19. Lerner, T.N. *et al.* Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. *Cell* **162**, 635–647 (2015).
20. Tai, L.-H., Lee, A.M., Benavidez, N., Bonci, A. & Wilbrecht, L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat. Neurosci.* **15**, 1281–1289 (2012).
21. Balleine, B.W., Delgado, M.R. & Hikosaka, O. The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* **27**, 8161–8165 (2007).
22. Graybiel, A.M., Aosaki, T., Flaherty, A.W. & Kimura, M. The basal ganglia and adaptive motor control. *Science* **265**, 1826–1831 (1994).
23. Roitman, M.F., Wheeler, R.A. & Carelli, R.M. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron* **45**, 587–597 (2005).
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. Brain Res. Rev.* **31**, 6–41 (1999).
25. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
26. Gunaydin, L.A. *et al.* Natural neural projection dynamics underlying social behavior. *Cell* **157**, 1535–1551 (2014).
27. Lau, B. & Glimcher, P.W. Value representations in the primate striatum during matching behavior. *Neuron* **58**, 451–463 (2008).
28. Lau, B. & Glimcher, P.W. Dynamic response-by-response models of matching behavior in rhesus monkeys. *J. Exp. Anal. Behav.* **84**, 555–579 (2005).
29. Lammel, S. *et al.* Diversity of transgenic mouse models for selective targeting of midbrain dopamine neurons. *Neuron* **85**, 429–438 (2015).
30. Cui, G. *et al.* Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature* **494**, 238–242 (2013).
31. Hyland, B.I., Reynolds, J.N.J., Hay, J., Perk, C.G. & Miller, R. Firing modes of midbrain dopamine cells in the freely moving rat. *Neuroscience* **114**, 475–492 (2002).
32. Pan, W.-X., Schmidt, R., Wickens, J.R. & Hyland, B.I. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J. Neurosci.* **25**, 6235–6242 (2005).
33. Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
34. Roesch, M.R., Calu, D.J. & Schoenbaum, G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.* **10**, 1615–1624 (2007).
35. Bayer, H.M. & Glimcher, P.W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
36. Day, J.J., Roitman, M.F., Wightman, R.M. & Carelli, R.M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* **10**, 1020–1028 (2007).
37. Ilango, A. *et al.* Similar roles of substantia nigra and ventral tegmental dopamine neurons in reward and aversion. *J. Neurosci.* **34**, 817–822 (2014).
38. Gershman, S.J., Pesaran, B. & Daw, N.D. Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *J. Neurosci.* **29**, 13524–13531 (2009).
39. O'Reilly, R.C. & Frank, M.J. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.* **18**, 283–328 (2006).
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.* **33**, 5797–5805 (2013).
41. Matsumoto, M. & Hikosaka, O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459**, 837–841 (2009).
42. Mirenowicz, J. & Schultz, W. Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* **72**, 1024–1027 (1994).
43. Mirenowicz, J. & Schultz, W. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* **379**, 449–451 (1996).
44. Stefani, M.R. & 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.* **26**, 8810–8818 (2006).
45. Howe, M.W., Tierney, P.L., Sandberg, S.G., Phillips, P.E.M. & Graybiel, A.M. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* **500**, 575–579 (2013).
46. Phillips, P.E.M., Stuber, G.D., Heien, M.L.A.V., Wightman, R.M. & Carelli, R.M. Subsecond dopamine release promotes cocaine seeking. *Nature* **422**, 614–618 (2003).
47. Hart, A.S., Rutledge, R.B., Glimcher, P.W. & Phillips, P.E.M. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J. Neurosci.* **34**, 698–704 (2014).
48. Hamid, A.A. *et al.* Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).
49. Willuhn, I., Burgeno, L.M., Everitt, B.J. & Phillips, P.E.M. Hierarchical recruitment of phasic dopamine signaling in the striatum during the progression of cocaine use. *Proc. Natl. Acad. Sci. USA* **109**, 20703–20708 (2012).
50. Brown, H.D., McCutcheon, J.E., Cone, J.J., Ragozzino, M.E. & Roitman, M.F. Primary food reward and reward-predictive stimuli evoke different patterns of phasic dopamine signaling throughout the striatum. *Eur. J. Neurosci.* **34**, 1997–2006 (2011).

ONLINE METHODS

Mice. 58 male TH::IRES-Cre C57BL/6J mice expressing Cre recombinase under the control of the tyrosine hydroxylase promoter (TH::IRES-Cre), 14 female TH::IRES-Cre C57BL/6J mice and 6 male DAT::IRES-Cre C57BL/6J mice, all of ages 12–20 weeks, were used for these experiments. Mice were group-housed with 3–5 mice/cage on a 12 h on, 12 h offlight schedule. All behavioral testing was performed during the light off time. All experimental procedures and animal care was performed in accordance with the guidelines set by the National Institutes of Health and were approved by the Princeton University Institutional Animal Care and Use Committee.

Instrumental reversal learning task. Mice undergoing behavioral testing were water-restricted to 80–85% of their *ad-libitum* weight. All behavioral experiments were performed in a 21 × 18 cm modular operant chamber (MED Associates, ENV-307W). The nose port, levers and rewards were confined to one wall of the chamber. The beginning of each trial was signaled by the illumination of a central nose poke. Upon entering the nose poke, animals initiated the release of two levers spaced equidistantly right and left of the center nose poke after a random jittered delay between 0–1 s. Animals were given 10 s to press a lever; failure to do so resulted in an abandoned trial (<2% of trials for all mice) that was excluded from analysis. A lever press resulted in either the CS– (0.5 s of white noise), and no reward delivery, or the CS+ (0.5 s of 5 kHz pure tone) and the delivery of 4 μ l reward (10% sucrose in H₂O) in a custom reward port located central to the two levers and directly below the central nose poke. The time from lever press to CS also had a delay between 0–1 s in 0.1 s intervals sampled from a uniform distribution. All trials were separated by a constant 3 s intertrial interval. Mice were trained daily and each behavioral session lasted one hour. Prior to data collection, mice were trained for 1–3 d to complete a nose poke and then a lever press in order to receive a reward (at this stage reward was not contingent on which lever was pressed). After 50 correct trials, the mice were then trained for 3–5 d on a paradigm where one lever was rewarded with 100% probability and the other lever was not rewarded, and which lever was rewarded was switched after at least 10 rewarded trials. Mice were then trained on the final paradigm for an additional 2 d before data was collected. Each block consisted of 10 correct trials plus a random number of trials drawn from a geometric distribution with a mean of 11 trials.

NpHR photocurrent measurements in TH⁺ neurons. Coronal slices containing the VTA–SNc were prepared from 3–4 month-old female TH::Cre mice (~4 weeks after DIO-NpHR-EYFP virus injection). Mice were deeply anesthetized with isoflurane and decapitated to remove the brain. The isolated whole brains were immersed in ice-cold carbogenated NMDG ACSF solution (92 mM NMDG, 2.5 mM KCl, 1.25 mM NaH₂PO₄, 30 mM NaHCO₃, 20 mM HEPES, 25 mM glucose, 2 mM thiourea, 5 mM Na-ascorbate, 3 mM Na-pyruvate, 0.5 mM CaCl₂, 10 mM MgSO₄ and 12 mM N-acetyl-L-cysteine, pH adjusted to 7.3–7.4). Afterwards brain slices (300 μ m) were sectioned using a vibratome (VT1200s, Leica, Germany) and then incubated in NMDG ACSF at 34 °C for 15 min. Slices were then transferred into a holding solution of HEPES ACSF (92 mM NaCl, 2.5 mM KCl, 1.25 mM NaH₂PO₄, 30 mM NaHCO₃, 20 mM HEPES, 25 mM glucose, 2 mM thiourea, 5 mM Na-ascorbate, 3 mM Na-pyruvate, 2 mM CaCl₂, 2 mM MgSO₄ and 12 mM N-acetyl-L-cysteine, bubbled at room temperature with 95% O₂ and 5% CO₂). During recordings, slices were perfused (flow rate of 4–5 ml/min) with a recording ACSF solution (120 mM NaCl, 3.5 mM KCl, 1.25 mM NaH₂PO₄, 26 mM NaHCO₃, 1.3 mM MgCl₂, 2 mM CaCl₂ and 11 mM D-(+)-glucose), and continuously bubbled with 95% O₂ and 5% CO₂. Whole-cell recordings were performed using a Multiclamp 700B (Molecular Devices, Sunnyvale, CA) using pipettes with a resistance of 3–5 M Ω filled with a potassium-based internal solution containing 120 mM potassium gluconate, 0.2 mM EGTA, 10 mM HEPES, 5 mM NaCl, 1 mM MgCl₂, 2 mM Mg-ATP and 0.3 mM Na-GTP, with the pH adjusted to 7.2 with KOH. TH⁺ neurons were identified for recordings on the basis of YFP expression. Photostimulation parameters were 560 nm and 5 mW per mm². Neurons were held at –70 mV during photocurrent measurements. To confirm the ability of photocurrents to eliminate action potentials (Supplementary Fig. 1c), action potentials were induced by positive current injection (120 pA, 100 ms pulse duration, 5 Hz).

Optogenetic inhibition of putative dopamine neurons. 16 male TH::IRES-Cre mice were bilaterally injected in the VTA (M–L $\pm$ 0.5, A–P –3.1 and D–V

–4.75 mm) with either AAV5 Ef1a-DIO-eNpHR3.0-EYFP (n = 8 NpHR mice, UNC virus core, injected 0.5 μ l per hemisphere of titer of 4×10^{12} pp per ml) or AAV5-Ef1a-DIO-EYFP (n = 8 control mice, UNC virus core, final titer of 4×10^{12} virus, with littermates randomly assigned to each group. 6 male DAT::Cre mice were bilaterally injected with AAV5 Ef1a-DIO-eNpHR3.0-EYFP. Optical fibers (300 μ m core diameter, 0.37 NA) delivering 3–4 mW of 532 nm laser light (measured at the fiber tip) were implanted bilaterally above the VTA (M–L $\pm$ 0.4, A–P –3.3 and D–V –4.0 mm). Animals were anesthetized for implant surgeries with isoflurane (3–4% induction and 1–2% maintenance). Animals were given 5 d of recovery after the surgical procedure before behavioral testing. During behavioral sessions, light was administered randomly on 10% of trials beginning at the illumination of the central nose poke and remaining until either the duration of the CS– or the end of reward consumption. Days with optical stimulation alternated with days without stimulation; days without stimulation were excluded from analysis. All mice were tested without knowledge of the group in which the mouse belonged to known (in other words, the experimenter was blind to conditions). All segregation into groups performed after data collection. Anatomical targeting was confirmed as successful in all mice through histology after the experiment (Fig. 2b), and therefore no mice were excluded from this data set.

Model of choice behavior. The mouse's choice on trial i was modeled on the basis of the outcomes and choices on the previous n trials with a logistic regression model:

$$\log\left(\frac{C(i)}{1-C(i)}\right) = \beta_0 + \sum_{j=1}^n \beta_j^R R(i-j) + \sum_{j=1}^n \beta_j^N N(i-j) + \text{error}$$

where $C(i)$ is the probability of choosing the right lever on trial i . The rewarded press variable, $R(i)$, was defined as +1 when the i^{th} trial was both rewarded and a right press, –1 when the i^{th} trial was rewarded and a left press and 0 when it was unrewarded. Similarly the unrewarded press variable, $N(i)$, was defined as +1 when the i^{th} trial was both unrewarded and a right press, –1 when the i^{th} trial was unrewarded and a left press and 0 when it was rewarded. Together, these two variables define reward and choice history without redundancy. We chose these variables because of the ease of interpretation of the corresponding coefficients (β_j^R and β_j^N), which reflect the strength of the relationship of rewarded (or unrewarded) outcomes in previous trials with the animal's current choice. For each mouse, the regression coefficients were fit using a logistic regression function in MATLAB (function *glmfit*). For Figures 1c and 2d, coefficients represent the average across mice⁵¹.

To examine the effect of optical inhibition of DA neurons on the animal's choice, we expanded the model to include the time of optical stimulation, as well as the interaction between choice and stimulation:

$$\log\left(\frac{C(i)}{1-C(i)}\right) = \beta_0 + \sum_{j=1}^n \beta_j^R R(i-j) + \sum_{j=1}^n \beta_j^N N(i-j) + \sum_{j=1}^n \beta_j^{LR} L(i-j)R(i-j) + \sum_{j=1}^n \beta_j^{LN} L(i-j)N(i-j) + \sum_{j=1}^n \beta_j^L L(i-j) + \text{error}$$

where $L(i)$ represents optical stimulation on the i^{th} trial (1 for optical stim, 0 for control trials).

Finally, to statistically test the interaction between an animal's choice, optical stimulation and the presence or absence of a light-activated opsin (i.e., NpHR versus YFP-control groups), we further augmented the model to fully cross all effects with an indicator for opsin. We estimated this model as a mixed effects model using the lme4 package⁵² in the R statistical language (R Core Team, 2014) where the factors with repeated measures (intercept, stimulation, reward, and their interactions) were allowed to vary across animals.

Calcium recordings in dopaminergic terminals. 11 mice were injected bilaterally with AAV5-CAG-Flex-GCamp6f-WPRE-SV40 (UPenn virus core, injected titer of 3.53×10^{12} pp per ml) in VTA using the coordinates described above. Optical fibers (400 μ m core diameter, 0.48 NA) were implanted in either the DMS (M–L $\pm$ 1.5, A–P 0.74 and D–V –2.4 mm) or the NAc (M–L $\pm$ 1.25, A–P

1.68 and D–V –4.15 mm). Each animal was implanted with one optical fiber in each brain region on opposite hemispheres, alternating which fiber was placed on which hemisphere between mice.

To record neural activity with gCaMP6f in terminals of dopaminergic neurons, light from the excitation laser (488 nm) was chopped at 400 Hz before entering the implanted optical fiber (square wave with 50% duty cycle) and emission light was collected through the same fiber using a photodetector (New Focus, Femtowatt Photoreceiver model 2151), similarly to ref. 26. The excitation light was first filtered through a 488 nm UV/VIS bandpass filter (Thor Labs, FL488-10) to reduce background noise in the system. Emission light was filtered through a 525 nm GFP clean up filter (Thor Labs, MF525-39) before reaching the photodetector. The signal from the photodetector was then processed with a lock-in amplifier (Signal Recovery, Model 7265) to measure signal time-locked to the 400 Hz chopped excitation laser. The fluorescence signal from the lock-in amplifier was digitized at 15 Hz on a data acquisition board (USB-201, Measurement Computing) and stored.

Targeting was confirmed as successful in all but one fiber through histology after the experiment, and therefore that recording site was excluded from analysis. Another recording site was excluded because there were no discernible gCaMP6f transients. No other data was excluded.

Animals expressing gCaMP6f were run on the probabilistic reversal learning task. For the data in **Figures 4** and **5**, 3,755 rewarded trials and 4,003 unrewarded trials were obtained across 11 VTA–SN::Nac animals, and for VTA–SN::DMS recordings there were 4,861 rewarded trials and 5,247 unrewarded trials. For the data in **Figure 6**, VTA–SN::Nac recordings contained 3,526 ipsilateral and 4,232 contralateral trials from 11 recording sites, and VTA–SN::DMS recordings contained 4,822 ipsilateral and 5,286 contralateral trials from 11 recording sites. For the cell body VTA–SN::DMS recordings from **Figure 7** there were 2,312 ipsilateral and 2,626 contralateral trials from 7 recordings sites.

Calcium recordings in VTA–SN cell bodies. 44 male TH::IRES-Cre mice were injected bilaterally with AAV5-CAG-Flex-GCamp6f-WPRE-SV40 (UPenn virus core, final titer of 3.53×10^{12} pp per ml) in DMS (M–L ± 1.4 , A–P 0.74, D–V –2.6 mm). Optical fibers (400 μ m core diameter, 0.48 NA) were implanted in the VTA–SN (M–L ± 0.9 , A–P –3.3 and D–V –4.2 mm). gCaMP6f recording were performed as described above and targeting was confirmed and reported in **Figure 7b**. One data point was excluded because there were no discernible gCaMP6f transients.

Analysis of calcium data. Data were post-processed in MATLAB using a high-pass FIR filter with a passband of 0.4 Hz, stopband of 0.1 Hz and a stopband attenuation of 10 dB to remove the baseline fluorescence and correct for drift in the baseline. dF/F was calculated by dividing the high-pass filtered signal by the mean of the signal before high-pass filtering. To compare signal across recording sites, the z -score of dF/F was calculated by dividing each recording site's dF/F trace by its s.d.

In order to calculate kernels that correspond to the isolated response to each behavioral event (**Figs. 4e, 6c,f** and **7f**), the time-dependent gCaMP6f signal was modeled as the sum of the response to each behavioral event. The response to each behavioral event was considered as the convolution of a time series representing the time of the event as a series of 0s and 1s, where the 1s correspond to the time of the event and to the kernel corresponding to the response profile to that event. The model can be written as follows, where $g(t)$ is the gCaMP6f signal, a , b , etc., are example behavioral event time series (e.g., nose poke, lever presentation) and k^a and k^b are the kernels for the corresponding event:

$$g(t) = g_0 + \sum_{t'=-1s}^{2s} a(t-t')k^a(t') + \sum_{t'=-1s}^{2s} b(t-t')k^b(t') + \dots + \text{error}$$

For each recording site, the coefficients of the kernels were solved using the method of least-squares in MATLAB (with the function *regress*). **Figures 4e, 6c,f** and **7f** display the average (and SEM) kernel across recording sites⁵¹, with each kernel based on data from 1 h recording sessions across 4 d from one site. The calculation described in this paragraph is depicted schematically in **Figure 4a**. This linear regression model successfully captured a significant amount of variance in the gCaMP6f response (R^2 of 0.24 ± 0.020 for VTA–SN::Nac terminals and 0.11 ± 0.016 for VTA–SN::DMS terminals).

Fast scan cyclic voltammetry. 3 TH::IRES-Cre mice previously injected bilaterally with AAV5 CAG-Flex-GCamp6f-WPRE-SV40 in VTA (as described above) were used for these experiments ($n = 9$ recording sites). Mice were anesthetized with urethane (1.5 g per kg, i.p.) and placed in a stereotaxic frame. A bipolar stimulating electrode (Plastics 1 Inc., Roanoke, VA) was placed in the medial forebrain bundle (MFB; M–L ± 1.1 , A–P –2.4 and D–V –5 mm from bregma). A Ag–AgCl reference electrode was placed contralateral to the stimulating electrode in the left or right forebrain and held in place with a stainless steel skull screw. During the experiment, carbon-fiber microelectrodes (90–110 μ m length) that had been coupled to 400 micron core optical fibers were placed into the DMS and Nac core (M–L ± 1.7 , A–P 1.68 mm from bregma and D–V –3.0 to –4.3 mm from brain surface).

Changes in DA concentration during electrical stimulation of the MFB were assessed using fast scan cyclic voltammetry (FSCV). The carbon-fiber and Ag–AgCl reference electrodes were connected to an external headstage. Voltammetric recordings were made every 100 ms by applying a triangular waveform (–0.4 to +1.3 V, 400 V per s). Data were digitized and stored to a computer using software written in LabVIEW (National Instruments, Austin, TX).

For each animal, a fresh carbon-fiber microelectrode was initially inserted 2.5 mm into the brain above the region of interest, and the triangular waveform was applied at a frequency of 60 Hz for 20 min. Waveform application was then changed to 10 Hz, and the electrode was lowered to the recording site. Electrical stimulation was applied to the MFB at a frequency of 60 Hz and pulse numbers varying between 24, 12, 6, and 1 biphasic pulse (300 μ A, 2 ms each phase). These stimulation parameters were repeated 5 times at each recording site.

DA signals from FSCV were identified as previously described²³. For analyte identification, current during a voltammetric scan is plotted against applied potential to yield a cyclic voltammogram (the chemical signature of the analyte). Cyclic voltammetric data were analyzed on stimulation trials. A background signal from 1 voltammetric scan (100 ms time bin) before a stimulation was subtracted from the remainder of the scans to reveal changes in DA concentration (rather than absolute values).

Current was examined at the oxidation potential of DA to reveal changes in DA current versus time during stimulation trials (I versus t). After the experiment, changes in current were converted to changes in DA concentration by post calibration. Carbon-fiber microelectrodes were calibrated by recording the *in vitro* response to injections of known DA concentrations ranging from 300 nM to 3 μ M in ACSF. The calibration curve was generated using 10 different electrodes and gave a linear response across dopamine concentrations with an R^2 value of 0.84. The equation generated from the linear response was used to determine an electrode response of 32.2 nA per μ M.

Histology. After completion of behavioral experiments, mice were perfused with 4% PFA in PBS, and then brains were removed and postfixed in 4% PFA for 24 additional hours before transferring to 30% sucrose in PBS. After post-fixing, 40 micron sections were made with an American Optical 860 microtome. For immunohistochemistry, slices were blocked in 3% normal donkey serum (NDS) in PBS with 0.25% Triton X-100 for 30 min. Sections were then incubated at 4 °C overnight in polyclonal chicken anti-TH primary antibody (1:1,000, Aves Labs, No. TH) and in monoclonal rabbit anti-GFP primary antibody (1:1,000, Life Technologies, No. G10362). PBS washes were performed to remove primary antibody, and slices were then incubated overnight in AlexaFluor488-conjugated donkey anti-rabbit IgG (1:1,000 dilution, Jackson ImmunoResearch, No. 711-545-152) and AlexaFluor647-conjugated donkey anti-chicken IgG (1:1,000 dilution, Jackson ImmunoResearch, No. 703-545-155). Following PBS washes, slices were mounted in 1:2,500 DAPI in Fluoromount-G. Whole slices were imaged with a Nikon Ti2000E epifluorescence microscope to determine optical fiber targeting while confocal images of virus expression in cell bodies and terminals were taken using a Leica TCS SP8.

For the retrograde tracing experiments, 200–400 nl of cholera toxin subunit B conjugated to either the Alexa-555 or -647 fluorophore (Life Technologies C-34776 & C-34778) was injected into the Nac or DMS, respectively, and 5–7 d post-injection, animals were perfused as described above.

Statistics. All paired and unpaired *t*-tests mentioned in the Results were performed using MATLAB, using either the *ttest* or *ttest2* function. In all cases *t*-tests

were two-tailed. The two-way repeated measures ANOVAs reported in results paragraph 7 were done in R. The logistic regressions used to generate **Figures 1c** and **2d** and the linear regression used to generate the response kernels (**Figs. 4e, 6c,f** and **7f**) were done in MATLAB using the *glmfit* function. To estimate the s.d. for the penetrance and specificity of gCaMP6f in TH⁺ cells (Results paragraph 6) we used the s.d. for a binomial distribution: $\sqrt{np(1-p)}$.

For all *t*-tests in this paper, data distributions were assumed to be normal, but this was not formally tested. No statistical methods were used to predetermine sample sizes, but our sample sizes were similar to those generally employed in the field.

We performed additional analyses on data of **Figures 6** and **7** that were not included in the main text, but that supported the same conclusions as the statistics that were included in the main text. The tests and results are described in detail below. Given the potentially longer time course of VTA-SN::Nac versus VTA-SN::DMS terminal responses (compare **Fig. 6c** to **Fig. 6f**), we assessed whether re-analyzing the VTA-SN::Nac terminal responses using longer analysis time windows would reveal significant differences in ipsilateral and contralateral responses. First, we expanded the size of the time bins used in the original *t*-test (see shaded region in **Fig. 6f**) from 100 to 200 ms and still found no significant difference between ipsilateral and contralateral response kernels for either nose poke or lever presentation (paired *t*-test; all 200-ms time bins from −1 s to 2 s; $P > 0.01$). Second, to determine whether a significant difference existed outside of the range of the initial time window of −1 to 2 s (kernels in **Figs. 4, 6** and **7**), we expanded the bin-wise analysis of VTA-SN::Nac terminal kernels from −2 to 3 s and again found no significant difference between ipsilateral and contralateral responses (paired *t*-test; all time bins; $P > 0.01$).

We also performed an additional statistical test to compare contralateral versus ipsilateral responses in VTA-SN::Nac terminals that did not involve bin-wise comparisons (**Fig. 6b,c**). Toward this end, we directly compared the peak response of ipsilateral and contralateral time-locked responses or kernels (**Fig. 6b,c**) across animals. This analysis revealed no significant difference in the time-locked responses between ipsilateral and contralateral trials (paired *t*-test; nose poke, $P = 0.24$; lever presentation, $P = 0.78$) nor the nose poke or lever presentation kernels (paired *t*-test; nose poke, $P = 0.063$; lever presentation, $P = 0.76$).

For the data in **Figure 6f**, the bin-wise *t*-test comparing ipsilateral and contralateral response kernels (indicated as significant by the shaded gray area)

for VTA-SN::DMS terminal recordings produced a minimum *P*-value of 2.51×10^{-4} for the nose poke kernels and 8.84×10^{-4} for the lever presentation kernel. Similarly, the same bin-wise analysis on the gCaMP6f time-locked responses in **Figure 6e** results in a minimum *P*-value of 1.31×10^{-7} for the nose poke and 1.55×10^{-6} for the lever presentation. For the data in **Figure 7f**, the bin-wise *t*-test comparing ipsilateral and contralateral response kernels (indicated as significant by the shaded gray area) for VTA-SN::DMS terminal recordings produced *P*-values as low as 3.22×10^{-4} for the nose poke kernels and 0.005 for the lever presentation kernel. Similarly, the same bin-wise analysis on the gCaMP6f time-locked responses in **Figure 7e** results in *P*-values as low as 4.54×10^{-5} for nose poke and 6.44×10^{-5} for lever presentation.

We performed an additional statistical test to compare contralateral versus ipsilateral responses in VTA-SN::DMS cell bodies that did not involve bin-wise comparisons. Directly comparing the peak of the contralateral kernel with the peak of the ipsilateral kernel (from −1 s to +2 s event onset) results in $P = 7.88 \times 10^{-4}$ for nose poke and $P = 0.012$ for lever presentation (two-tailed paired *t*-test).

Finally, in the Results section, we compare the maximum difference in the contralateral and ipsilateral responses between VTA-SN::Nac and VTA-SN::DMS terminals (unpaired *t*-test; nose poke, $P = 0.01$, lever presentation, $P = 0.024$). Given the similarity of the VTA-SN::DMS terminal and cell body data (**Figs. 6f** and **7f**), we also merged the two VTA-SN::DMS groups to make a larger combined group, and again performed the same statistical comparison between the combined VTA-SN::DMS and the VTA-SN::Nac terminal data. As expected, the *P*-values were smaller in this case due to the larger sample size ($P = 9.202 \times 10^{-4}$ for nose poke and 0.011 for lever presentation).

A **Supplementary Methods Checklist** is available.

Code accessibility. Custom MATLAB code used in this study is available upon request.

51. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. Preprint at <http://arxiv.org/abs/1406.5823> (2014).
52. Friston, K.J., Stephan, K.E., Lund, T.E., Morcom, A. & Kiebel, S. Mixed-effects and fMRI studies. *Neuroimage* 24, 244–252 (2005).