Article

Neuron

Volitional Modulation of Primary Visual Cortex Activity Requires the Basal Ganglia

Highlights

- Rodents learn to produce activity patterns in V1 in order to receive a reward
- Visual cortex neurons are volitionally modulated in the absence of visual input
- Inhibition of DMS impairs learning, but not production of learned patterns
- Basal ganglia circuits play a general role in shaping cortical activity

Authors

Ryan M. Neely, Aaron C. Koralek, Vivek R. Athalye, Rui M. Costa, Jose M. Carmena

Correspondence

rc3031@columbia.edu (R.M.C.), jcarmena@berkeley.edu (J.M.C.)

In Brief

Neely et al. use brain-machine interface training in rodents to demonstrate that neurons in the primary visual cortex can acquire learned modulations in the absence of visual input, and that this form of learning requires participation of the dorsomedial striatum.

Neely et al., 2018, Neuron *97*, 1–13
March 21, 2018
https://doi.org/10.1016/j.neuron.2018.01.051

CellPress

Please cite this article in press as: Neely et al., Volitional Modulation of Primary Visual Cortex Activity Requires the Basal Ganglia, Neuron (2018), https://doi.org/10.1016/j.neuron.2018.01.051

Neuron
Article

Volitional Modulation of Primary Visual Cortex Activity Requires the Basal Ganglia

Ryan M. Neely,[1,7] Aaron C. Koralek,[2,7] Vivek R. Athalye,[2,3] Rui M. Costa,[2,4,6,*] and Jose M. Carmena[1,3,5,6,8,*]
[1]Helen Wills Neuroscience Institute, University of California-Berkeley, Berkeley, CA 94720, USA
[2]Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon 1400-038, Portugal
[3]Department of Electrical Engineering and Computer Sciences, University of California-Berkeley, Berkeley, CA, 94720, USA
[4]Department of Neuroscience and Neurology, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA
[5]Joint Graduate Group in Bioengineering UCB/UCSF, Berkeley, CA 94720, USA
[6]Senior author
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: rc3031@columbia.edu (R.M.C.), jcarmena@berkeley.edu (J.M.C.)
https://doi.org/10.1016/j.neuron.2018.01.051

SUMMARY

Animals acquire behaviors through instrumental conditioning. Brain-machine interfaces have used instrumental conditioning to reinforce patterns of neural activity directly, especially in frontal and motor cortices, which are a rich source of signals for voluntary action. However, evidence suggests that activity in primary sensory cortices may also reflect internally driven processes, instead of purely encoding antecedent stimuli. Here, we show that rats and mice can learn to produce arbitrary patterns of neural activity in their primary visual cortex to control an auditory cursor and obtain reward. Furthermore, learning was prevented when neurons in the dorsomedial striatum (DMS), which receives input from visual cortex, were optogenetically inhibited, but not during inhibition of nearby neurons in the dorsolateral striatum. After learning, DMS inhibition did not affect production of the rewarded patterns. These data demonstrate that cortico-basal ganglia circuits play a general role in learning to produce cortical activity that leads to desirable outcomes.

INTRODUCTION

Behavioral flexibility is essential for survival in changing and uncertain environments. Task-relevant modification or enhancement of sensory representations can be important to improve behavioral outcomes: for example, attentional resources can be used to amplify activity related to salient stimuli while ignoring distractors. Many sensory areas of the cortex, including primary sensory areas, display activity that reflects task parameters, changing behavioral contexts, and shifts of attention, suggesting that computations in these regions are influenced by ongoing internal processes (Keller et al., 2012; Martínez et al., 1999; Niell and Stryker, 2010; Shuler and Bear, 2006; Steinmetz et al., 2000; Zhang et al., 2014). These task-relevant modulations of ongoing sensory representations can emerge and evolve following repeated training or association with a salient stimulus (Makino and Komiyama, 2015). An important question is how modulatory inputs to functionally diverse cortical circuits are tuned such that their outputs contribute positively to the behavioral outcomes of an individual. The basal ganglia, through the striatum, receives input from most cortical areas (Hintiryan et al., 2016; Kemp and Powell, 1970; McGeorge and Faull, 1989; Webster, 1965), feeds back to the cortex via multiple recurrent pathways (Redgrave et al., 2010), and dynamically encodes action-outcome contingencies (Samejima et al., 2005; Tricomi et al., 2004), making this structure a likely candidate to shape cortical activity based on behavioral experience (Barnes et al., 2005; Graybiel, 2008; Hinterberger et al., 2005; Swanson, 2000). Previous work has demonstrated the importance of the striatum for voluntary behavior and instrumental learning (Hikosaka et al., 1999; Yin et al., 2005, 2006, 2009). Similarly, we have shown that a brain-machine interface controlled by neurons in the primary motor cortex also requires cortico-striatal plasticity in order for animals to learn a novel neuroprosthetic action (Koralek et al., 2012, 2013). However, in addition to overt motor behaviors driven by motor cortices, cortico-striatal circuits have been theorized to also support abstract forms of learning, such as cognitive associations (Graybiel, 1997; Middleton and Strick, 1994). Furthermore, damage to basal ganglia structures in human patients, either through stroke or in diseases like Parkinson's, have been associated with deficits in sensory perception and the control of visual attention (Brown et al., 1997; Husain et al., 1997; Mercuri et al., 1997; Wright et al., 1990; Yamaguchi and Kobayashi, 1998). These data suggest that basal ganglia circuits may be involved in learning modulatory signals that influence many forms of cortical processing based on experience. However, observing and measuring these influences can be difficult, especially when their contributions to overt behavior may not be immediately apparent.

One strategy to overcome this difficulty is to use brain-machine interfaces (BMIs) that directly map a subject's internally generated neural activity to the movement of an artificial effector. By explicitly defining the behavioral relevance of observable

patterns of neural activity, BMI can be an important tool for studying how these patterns are generated. In clinical applications, the activity of cortical neurons in humans and non-human primates has been decoded as a proof-of-principle control signal to replace lost motor function by controlling prosthetic devices (Aflalo et al., 2015; Bouton et al., 2016; Collinger et al., 2013; Gilja et al., 2015; Hochberg et al., 2012). However, an important observation is that populations of cortical neurons whose activity is remapped to the movement of an artificial effector can undergo marked learning-related changes, and observing this learning process provides a unique window into how learning proceeds in the cortex (Arduin et al., 2013; Ganguly and Carmena, 2009; Ganguly et al., 2011; Hwang et al., 2013; Jarosiewicz et al., 2008; Prsa et al., 2017; Sadtler et al., 2014). In order to better facilitate such observations, BMI studies can thus be designed to observe the acquisition and evolution of volitional control signals, rather than to optimize the performance and control of a complex effector.

Here, we asked whether neurons in the primary visual cortex, an area involved in processing low-level visual features, could be instrumentally conditioned to produce arbitrary modulations of ongoing spike activity, and whether this abstract form of learning was dependent on the basal ganglia. To address this question, we trained rats and mice to perform a neuroprosthetic task that virtually re-routed spike activity from the primary visual cortex (V1) into the frequency of an auditory cursor. This allowed us to facilitate and observe learned modulations of V1 activity with a known relationship to behavior. Animals trained on the task successfully learned to produce this novel action by voluntarily modulating spike activity in V1. Then, using the red-shifted inhibitory opsin Jaws (Chuong et al., 2014) to inactivate striatal neurons on a trial-by-trial basis, we then investigated to what degree this instrumental learning process in V1 was also dependent on activity in dorsomedial and dorsolateral striatum.

RESULTS

A V1-Driven Closed-Loop Brain-Machine Interface

We adapted our previously developed neuroprosthetic task for rodents (Koralek et al., 2012) in order to directly study abstract reinforcement learning in V1 (Figure 1A). Briefly, 10 rats (and 12 mice, see later) were implanted chronically with microwire electrode arrays positioned in V1 layer 5 (L5), allowing us to isolate and record individual units (Figure 1A; Figures S1A and S1B). In addition to V1, 8 rats were also implanted with electrode arrays in the dorsomedial striatum (DMS) near the projection target of V1 (Hintiryan et al., 2016; Khibnik et al., 2014; McGeorge and Faull, 1989) (Figures S4A and S4B). During the course of the experiment, animals were placed in a totally dark or lighted behavioral chamber and allowed to move freely while listening to auditory feedback that reported their neural state in real time. Each day, two neural ensembles, consisting of two well-isolated units each, were randomly chosen to directly control the continuous auditory cursor (direct units), while the remaining units recorded in V1 had no defined relationship to cursor control (indirect units, Figure S1B). Activity of the two direct-unit ensembles had an opposing relationship, such that spikes produced by ensemble 1 (E1) moved the cursor closer to the rewarded frequency, while spikes in ensemble 2 (E2) moved the cursor away from the rewarded frequency and toward the unrewarded frequency (Figures 1A–1C; Movie S1). The highest and lowest possible tones were randomly assigned to be rewarded or unrewarded for each animal, and this association remained constant for the duration of training. Prior the start of every session, a baseline distribution of neural states (binned E1-E2 spike counts) was used to initialize the target values such that the chance rate of attaining each target (rewarded and unrewarded) within a trial period was approximately 30% (Figures S1C and S1D; see STAR Methods). Animals were given 30 s to reach either target; otherwise the trial was considered a miss and animals received a white noise burst followed by a time out.

Rodents Learn to Control V1 Activity Patterns

Over the course of 7–12 training sessions (average 9.11 sessions), rats learned to perform the task well above chance level (Figures 1F and 1G). Animals in the late phase of learning, considered here the final 3 training sessions, exhibited significant improvements in the percentage of rewarded targets compared to their performance in the early phase (during the first three sessions of training) (Figure 1G). Additionally, over this same interval, latencies to rewarded targets decreased significantly, while changes in latencies to unrewarded targets were non-significant (Figure S2A). We observed that simply pairing particular auditory tones with reward was not sufficient to drive V1 activity. After 5 days of performance above chance level, we decoupled auditory tones from neural activity mid-session by playing back the sequence of tones generated in the first part of the session. Although reward was still delivered in tandem with the rewarded tone, modulation of the direct unit ensembles was markedly decreased when animals' neural activity was not driving the cursor (Figures 1D and 1E). This suggests that the learning that we observed was not an effect of classical conditioning and instead resulted from an intentional modulation of V1 activity. Because the chance rate for each target was reset at the start of each session to approximately 30%, increases in performance seen as animals progressed from early to late phases reflected greater improvements within single training sessions across days. We quantified this by comparing performance in the first 10 min of a session (first trials) and the last 10 min (last trials) (Figure 2A). This suggests that the learned ability to control V1 activity was retained between training sessions, even though animals needed to perform some *de novo* learning to adjust to the initial parameters of the transform in any given day. We observed that in late learning, over the course of single training sessions, animals acquired a strong preference for the rewarded target relative to the unrewarded target (Figures 2C and 2D). This was reflected in the shift of the distribution of auditory cursor values in the direction corresponding to the rewarded target tone (Figure 2B). Additionally, we observed that auditory feedback was necessary for learning: sessions in which the feedback tones were muted resulted in no significant difference between the number of rewarded and unrewarded targets ($p = 0.738$, Figures S2B and 2C), even though these no-feedback sessions were conducted after several days of successful normal training. These data demonstrate that closed-loop

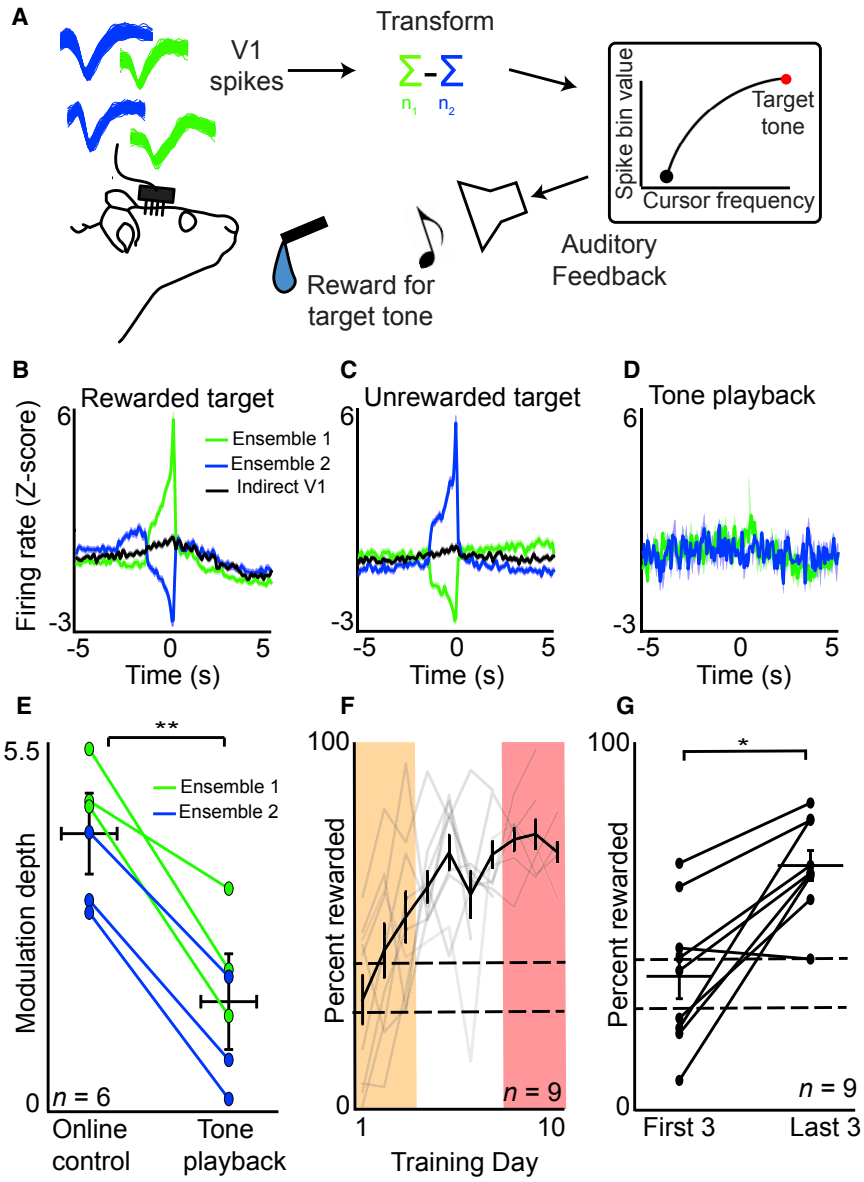

Figure 1. Learning to Operantly Control Activity in Primary Visual Cortex Using a Closed-Loop BMI Paradigm

(A) Schematic of V1-BMI paradigm. Activity of well-isolated V1 units (top left) were used to generate auditory tones using a differential transform (top right). Animals were rewarded for producing a target tone (red). A second tone (black) at the opposite end of the frequency range terminated the trial but was not rewarded. For more detail on the calculation of the transform, see Figures S1B–S1D.

(B) Average *Z* scored firing rates of V1 neurons arbitrarily assigned to ensemble 1 (green), ensemble 2 (blue), or unassigned (indirect; black), time locked to rewarded targets. Shaded areas show SEM.

(C) Same as in (B) but time locked to the unrewarded target. Increased activity in ensemble 1 units moved the tone frequency in the opposite direction as increased activity in ensemble 2 units.

(D) Ensemble 1 (green) and ensemble 2 (blue) modulations during passive playback of tones decoupled from ongoing neural activity, time locked to the rewarded target tone. Shaded areas represent SEM.

(E) Modulation depth of ensemble 1 and ensemble 2 during online control compared to tone playback. Mean modulation depth online control = 4.341; mean tone playback = 1.739; p = 0.000348 (paired t test). Black bars show mean and SEM. Lines show data from individual animals.

(F) Time course of learning across training days. Bold line shows the mean and SEM across 9 rats; gray lines show learning curves for animals individually. One animal only completed 4 sessions; data for this animal have been excluded from this plot. Dashed lines bound the range of chance performance levels. Orange highlighted region shows data range classified as the early learning phase for all animals; red region shows range for late learning phase.

(G) Comparison of performance between early learning phases and late learning phases (N = 9 animals, mean for the first 3 days = 39.6% rewarded; mean for the last 3 days 67.2%; p = 0.00162 [paired t test]). Black bars show mean and SEM, lines show data from individual animals.

neurofeedback-based reinforcement training can be used to condition the activity of neurons in the primary visual cortex.

V1 Modulation Is Sensitive to Task Contingency

We next investigated the sensitivity of performance to changing task conditions. In late learning, animals were able to quickly shift their neural activity to produce the rewarded tone frequency once the auditory feedback began (Figures 2C and 2D). We asked whether animals' behavior was habitual, and therefore insensitive to changes in action-reward contingencies, or goal-directed, in which performance remains sensitive to changing task contingencies (Dias-Ferreira et al., 2009). To test whether performance of the V1-controlled task fit either of these regimes, we degraded the task contingency so that animals received randomly timed rewards irrespective of target hits but at a similar rate. Similar to our observations of abstract skill learning in M1 (Koralek et al., 2012), task performance during degradation dropped significantly (Figure 2E) but returned to pre-manipulation levels once the reward contingency was reinstated. These results suggest that modulation of V1 activity was intentional and goal directed. To test whether the association between neural activity and rewarded cursor movement could be flexibly adapted to a new task contingency, we reversed the transform after animals had achieved saturating performance. This manipulation caused the cursor to move in the opposite direction for a given spike rate modulation than what animals had initially learned. Although this manipulation caused an initial decrease in performance, animals were able to learn the reversed behavior when given sufficient training time (Figure 2F; Figure S2E), showing that animals could dynamically adapt to changes in the relationship between neural patterns and reward.

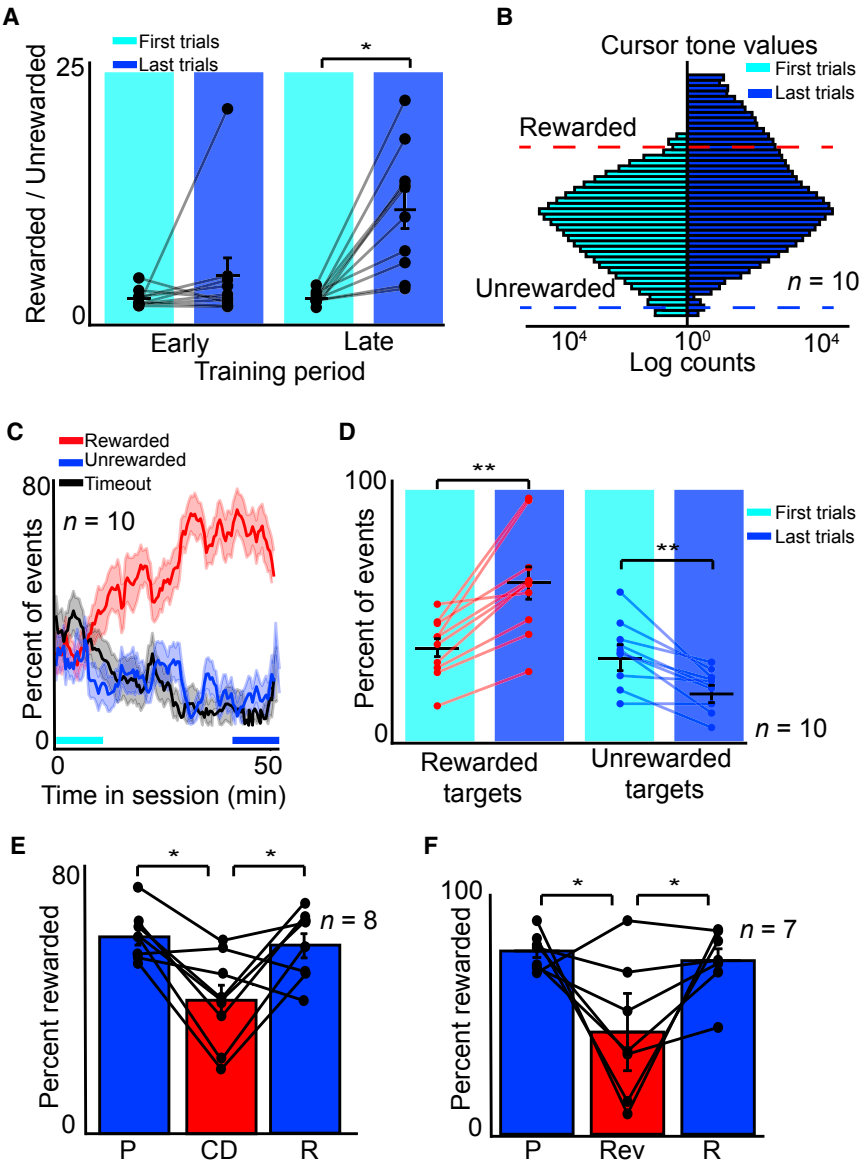

Figure 2. Operant Modulation of V1 Activity Is Sensitive to Changes in Contingency

(A) Comparison of within-session improvements during the early learning period (first 3 days) of training relative to the late learning period (last 3 days) for each animal (N = 9 animals), expressed as the ratio of rewarded to unrewarded targets. Data to compute the ratio for first trials (shaded in cyan) were averaged over the first 10 min of each session, while data for the last trials (shaded in blue) were averaged over the last 10 min. During the early learning period, the mean rewarded/unrewarded ratio for first trials was 1.12, last trials session = 3.848; p = 0.275 (paired t test). For the late period, mean ratio early in session = 1.195, mean late in session = 11.667; p = 0.00130 (paired t test). Black bars show mean and SEM, lines show data from individual animals.

(B) Mean distribution of cursor values for all animals for the first trials (first 10 min) compared to the last trials (last 10 min) of sessions during the late period. Cyan bars show the initial distribution, based on baseline activity, used to set the task parameters, while blue bars show the distribution at the end of the training session for the last trials. Dashed lines show the thresholds corresponding to the rewarded and unrewarded targets.

(C) Time course of mean within-session task learning during the late period (last 3 sessions) of training for all animals. Shaded areas represent SEM. Chance rates for rewarded and unrewarded targets were set at approximately 30% at the start of each training session. Comparisons in performance were made during over first 10 min of each session (first trials, cyan bar) and over the last 10 min (last trials, blue bar) of each session. N = 10 animals, mean of 3 sessions per animal.

(D) Quantification of rewarded and unrewarded target hits for the first trials compared to the last trials; data same as (C). N = 10 animals, mean of the late period (last 3 sessions) for each animal. Paired t test between first trials and last trials for rewarded targets: p = 0.00055; mean early = 36.9%, mean late = 59.1%. Paired t test between first trials and last trials for unrewarded targets: p = 0.00145; mean early = 33.0%, mean late = 20.6%. Red and blue indicate the rewarded and unrewarded targets, respectively (**p < 0.001). Black lines show mean and SEM. Lines show data from individual animals.

(E) Percentage of rewarded trials for contingency degradation sessions. Bars show means, lines show individual animals. N = 8 animals, mean of 2 sessions per animal. P, pre-degradation, mean = 64.5%. CD, peri-degradation (reward decoupled from cursor), mean = 43.1%. R, reinstatement of reward, mean = 61.4%. Paired t test between pre- (P) and peri- (CD) degradation, p = 0.0038. Paired t test between CD and reinstatement (R), p = 0.0283. Paired t test between pre-degradation (P) and reinstatement (R), p = 0.289. Error bars show SEM.

(F) Quantification of performance in contingency reversal sessions; N = 7 animals; average of 1.57 sessions per animal (range 1–2). P, pre-reversal (mean = 85.8%); Rev, peri-reversal (mean = 47.8%); R, recovery of performance with the decoder still reversed (mean = 81.2%). Paired t test between P and Rev, p = 0.0435; paired t test between Rev and R, p = 0.447; paired t test between P and R, p = 0.689. For a time course of reversal, see Figure S2E. Error bars show SEM, lines show data for individual animals.

Changes in Visual Context Affect Performance of a V1-Driven Task

Neurons in the primary visual cortex are known to respond to visual stimuli. The observation that animals can learn to successfully modulate V1 neurons in total darkness indicates that this activity is at least partially independent of visual input. One possibility is that over the course of learning, E1 and E2 units become decoupled from bottom-up sources of visual input, for example, visually driven activity from the lateral geniculate nucleus of the thalamus. If this were true, then we can expect trained animals to be able to perform the task under any light condition. To test this, we trained animals both in light and total dark conditions. Interestingly, no significant difference was observed in performance at the end of a training session (last

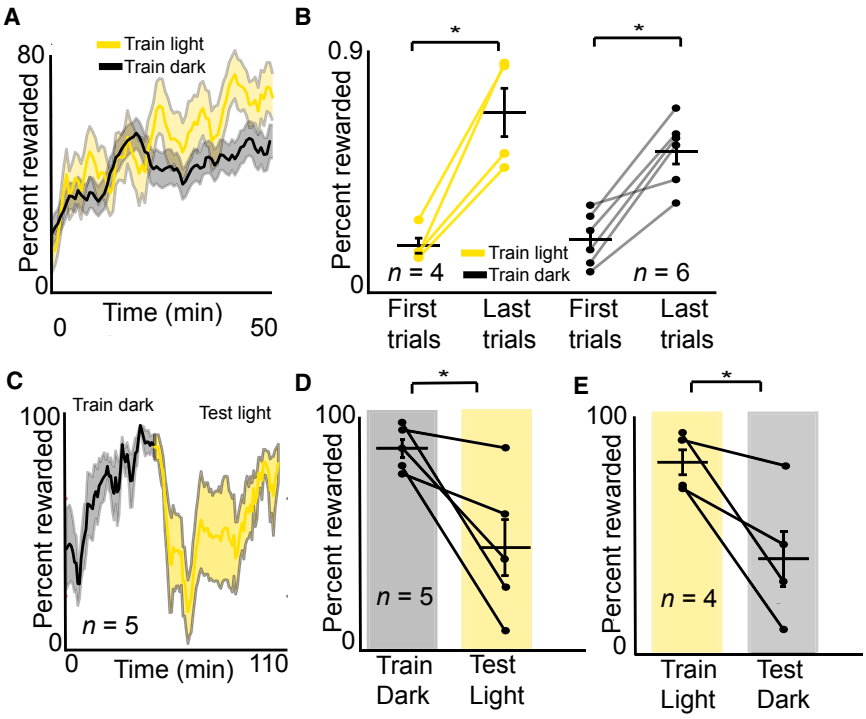

Figure 3. Operant Control of V1 Activity Is Sensitive to Changes in Visual Context

(A) Time course showing the mean percentage of rewarded trials within all sessions under lighted conditions (yellow) compared to dark conditions (black). Shaded areas represent SEM.

(B) Quantification of data in (A), using the first trials and last trials in a session. Train light: N = 4 animals; average of 7.5 sessions per animal (range = 5 to 10 sessions). Paired t test between first trials and last trials: p = 0.0113; mean first trials = 37.8%; mean last trials = 74.0%. Train dark: N = 6 animals; 8 sessions per animal. Paired t test between first trials and last trials: p = 0.00167; mean early = 0.396; mean late = 0.635. Unpaired t test between last trials for light sessions and last trials for dark sessions: p = 0.205 (*p < 0.05; black crosses show mean and SEM). Colored lines show data from individual animals.

(C) Time course showing the mean percentage of rewarded trials when animals learned a decoder under dark conditions and were switched to a lighted condition mid-session ("train dark, test light"). Shaded areas represent SEM.

(D) Mean percentage of rewarded trials when animals were trained in dark, and then tested in the light (same data as C). N = 5 animals, mean of 1.8 sessions per animal (range 1–2). Data taken from last 10 min of dark training and first 10 min of light testing. Mean train dark: 87.4%, mean test light 42.6%; p = 0.0309 (paired t test). Error bars show SEM; horizontal lines show mean (*p < 0.05).

(E) Mean percentage of rewarded trials when animals were trained in the light and tested in the dark. N = 4 animals, mean of 1.5 sessions per animal (range 1–2). Mean train light: 77.9%. Mean test dark: 35.4%. p = 0.043 (paired t test). For a time course, see Figure S2H. Error bars show SEM.

trials) for sessions performed entirely in the dark compared to in the light (train light = 74.0%, train dark = 63.5%, p = 0.205, Figures 3A and 3B), suggesting that learning can occur both in the presence and absence of visual stimuli. However, changing the context within a training session, i.e., from dark to light after the animals had learned the task in darkness, or vice versa, had a significant negative impact on performance (Figures 3C–3E; Figures S2H–S2J). Changing the light conditions affected the mean spike rates of all V1 neurons (Figure S2G; mean rate in dark = 4.13 Hz; mean rate in light 6.18 Hz). These results suggest that a modulatory input can learn to generate rewarded patterns of activity in direct units under stable network conditions but that changing the state of the network (for example, by adding or removing visually evoked activity) is disruptive and requires compensatory re-learning.

Learning to Modulate V1 Neurons Is Associated with Changes in Neural Dynamics

Next, we chose to examine the neural dynamics associated with learning goal-directed modulations in V1. Correlations between E1 units, whose combined positive activity modulations moved the cursor in the rewarded direction, significantly increased over the course of the session, suggesting that training resulted in increased coordination between these units. No such change was observed between E2 units or between E1 and E2 units (Figures 4A and 4B). In 72 out of 102 sessions, the relationship between performance and the E1 unit correlation was positive (mean Pearson correlation coefficient = 0.187). Of these sessions, 55.72% exhibited a significant ($p < 0.05$) correlation. An example session is shown in Figure 4C. We also observed that the correlation between E1 units was significantly greater in a 1 s window prior to rewarded target hits, relative to unrewarded targets or timeouts (Figure 4D). These data suggest that coordination between E1 units was important for success.

Interestingly, in the late learning phase, we observed an increase in LFP power in V1 prior to rewarded target hits (Figure 4E). Similar changes in ongoing oscillatory activity have previously been associated with top-down processing in visual cortices (Engel et al., 2001), which is one potential mechanism by which animals may be performing the task. To further explore this possibility, we then calculated the coherence between spikes in direct (combined E1 and E2) units and local field potentials (LFP) in V1, time locked to rewarded targets. Previous reports have found that attention alters alpha-band (approximately 5–15 Hz) coherence in the deep layers of visual cortex (Buffalo et al., 2011). We found that the alpha-band spike-field coherence (SFC) of direct units increased from early to late phases of learning (Figures 4G and 4H). This effect was stronger for E1 than for E2 units (Figures S3G–S3I). This increase was only observed during task performance but not when animals were engaged in passive behavior (Figure 4F). Indirect neurons did not show this effect (Figure 4G), suggesting that these learning-related dynamics were specific to units directly involved in cursor control. However, a relatively constant fraction of indirect neurons in each training session did show predictive power for target choice (Figures S3E and S3F), suggesting that there is

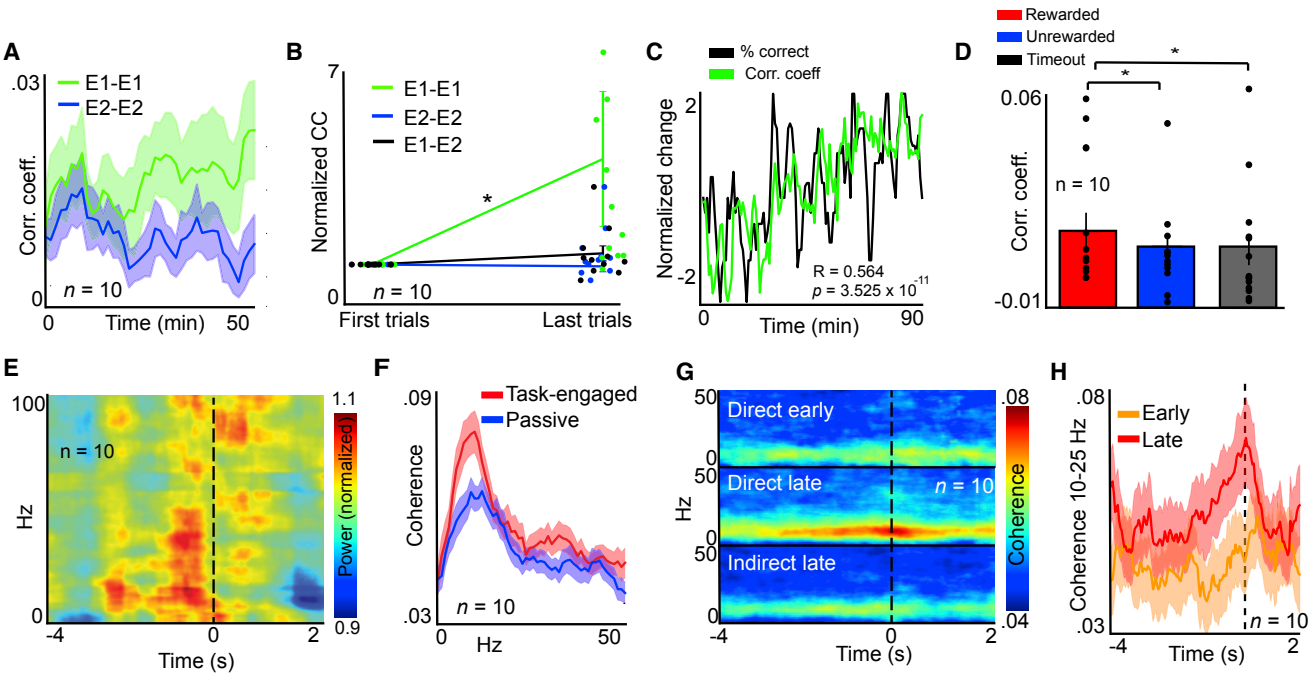

Figure 4. Evolving Neural Dynamics during V1-BMI Learning

(A) Mean pairwise correlations between units within ensemble 1 (green) or within ensemble 2 (blue) during training. Correlations were calculated using 1 ms bins. Shaded areas show SEM.

(B) Change in normalized correlation coefficients (cc) from the first trials to the last trials within sessions. CC calculated between units within ensemble 1 (mean change = 4.418; p = 0.0417; paired t test), within ensemble 2 (mean change = 0.942; p = 0.775; paired t test), or between units in ensembles 1 and 2 (mean change = 1.361; p = 0.491; paired t test). N = 10 rats; mean of 4.4 sessions per rat (range 3–6). Error bars show SEM.

(C) Example data from one session showing the relationship between changes in correlation of ensemble 1 units (green) and task performance (percent of correct trials, black).

(D) Mean correlation between ensemble 1 units during a 2 s window prior to target hits or timeouts. Mean cc prior to rewarded targets = 0.0285; mean cc prior to unrewarded targets = 0.0241 (p = 0.00167; paired t test). Mean cc prior to timeouts = 0.0240. Comparison between rewarded targets and timeouts: p = 0.0362; comparison between unrewarded targets and timeouts: p = 0.549. Error bars show SEM.

(E) Mean spectrogram of V1 LFP time-locked to rewarded target hits.

(F) Spike-field coherence between direct units and V1 LFP for late learning periods during task performance (red) compared to non-engaged passive behavior (blue). Shaded area represents SEM.

(G) Spike-field coherograms showing the evolution of coherence over time for LFPs in V1 and spikes from direct (combined ensemble 1 and ensemble 2) units in during early training periods (days 1–3; top plot), late training periods (last 3 days, middle plot), and indirect units (no direct relationship to cursor control) for late periods (bottom plot) time locked to rewarded targets. See Figures S3G–S3I for a separate analysis of E1 and E2 units.

(H) Coherence between direct units and V1-LFP in the 10–25 Hz band for early training compared to late training. Solid lines show the mean for 10 animals and shaded areas represent SEM.

a broader network of neurons in V1 that may have supported learning in the direct units.

Striatal Activity Becomes Task Relevant with Learning

To address whether the dorsomedial striatum (DMS) plays a role in V1-based reinforcement learning, we next examined whether activity in an area of DMS that receives input from V1 changed with learning. Electrical stimulation of V1 produced a reliable spike response in both putative DMS output neurons and interneurons (see STAR Methods) with a delay of approximately 6 ms, suggesting a direct projection as shown in previous reports (Allen Institute for Brain Science, 2015; Faull et al., 1986; Hintiryan et al., 2016; Khibnik et al., 2014; McGeorge and Faull, 1989) (Figures 5A and 5B; Figures S4D and S4E). Conversely, stimulation of DMS while recording in V1 did not produce an observable response in most (but not all) units (Figure S4C). In late learning, DMS output neurons exhibited a strong modulation time locked to the rewarded target that was not present in the early phase. This response was characterized by an increase in activity that emerged approximately 4 s prior to target hit, followed by a sharp decline during the reward period (Figures 5C and 5D). Units classified as interneurons exhibited the opposite profile (Figures S4G and S4H). This was accompanied by increases in beta and gamma LFP power within the same time interval (Figure 5F). We next asked whether learning was accompanied by changes in the dependent relationship between direct unit activity and activity of the recorded population in DMS. A linear regression analysis using data from DMS units (including both classified output neurons and interneurons) revealed that over the course of training days, population activity of recorded DMS units increasingly co-varied with V1 direct unit activity in a 500 ms window prior to target hits, such that a greater proportion

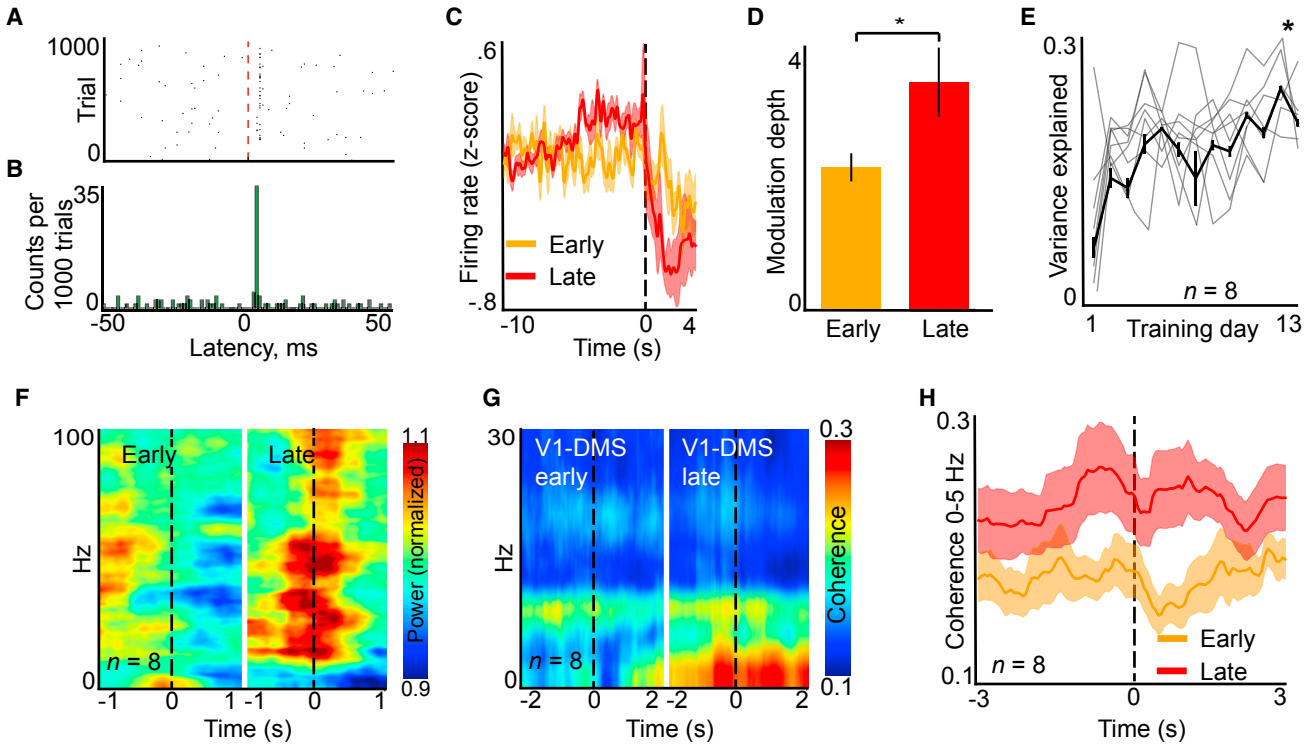

Figure 5. Dorsomedial Striatum Activity Becomes Engaged with V1 Ensembles during Learning

(A) Raster plot of an example putative DMS output neuron time-locked to V1 ICMS. For data from interneurons, see Figures S4D and S4E.

(B) Histogram of evoked spike counts from unit shown in (A) bin counts are total counts over 1,000 trials of ICMS.

(C) Mean *Z* scored firing rate of classified output units recorded in the dorsal medial striatum (DMS) time locked to rewarded targets for early compared to late learning. For data from interneurons, see Figures S4G and S4H. Shaded areas show SEM.

(D) Modulation depth of output DMS units in a 1 s window surrounding rewarded target hits. Mean for early periods = 2.132; mean for late periods = 3.711; p = 0.049. Error bars show SEM, dots show data for individual animals.

(E) Proportion of variance (R2) of direct unit spikes in V1 explained by DMS unit spikes (all units) in a 500 ms window prior to rewarded target hits, computed using a linear regression analysis on individual training days. Solid black lines show mean and SEM; gray lines show data for individual animals. Mean of first 2 days = 0.0788; mean of last 2 days = 0.276; p = 0.0196 (paired t test).

(F) Mean spectrogram of DMS LFP time locked to rewarded target hits for all animals during the early period (sessions 1–3, left), compared to the late period (last 3 sessions, right).

(G) Field-field coherograms between V1 LFP and DMS LFP for early (left) compared to late (right) training sessions time locked to rewarded targets.

(H) Field-field coherence between V1 and DMS in the 0–5 Hz band in for early and late training, time locked to rewarded targets. Solid lines show mean for all animals; shaded areas represent SEM.

of variance of direct unit activity could be explained by activity in DMS (Figure 5E, Figure S4F). This result suggests that extended training of V1 activity increasingly recruited the striatum. In accordance with this observation, field-field coherence between V1 LFP and DMS LFP was also increased in late learning around the time of target achievement (Figures 5G and 5H), suggesting a possible role for the striatum in learning to produce rewarded modulations in V1.

Dorsomedial, but Not Dorsolateral Striatal Activity Is Critical for Learning to Modulate V1 Activity

Next, we asked whether DMS activity was necessary for learning to volitionally modulate V1 neural activity. Mice were injected bilaterally with the red-shifted inhibitory opsin Jaws (rAAV8/CamKII-Jaws-KGC-GFP-ER2; Chuong et al., 2014) into the area of DMS that receives input from V1 and implanted chronically with optical fibers targeting DMS and with recording electrodes in L5 of V1 (Figures 6A and 6B). Red light stimulation through the optical fiber decreased activity in optogentically identified DMS units (Figure 6B; Figures S5C, S5D, S5G, and S5H) but had no effect on spike rates in V1 (Figure S5I). Mice were trained on the same task as rats as described above and in Figure 1. During the first 7 days of training, Jaws-injected mice and GFP-injected controls randomly received red light stimulation on 50% of trials for the whole duration of the trial (see STAR Methods; LED 50, Figure 6A, bottom). Control GFP animals were able to learn the task and improved significantly with several days of training, while Jaws animals did not (Figures 6C and 6D). This appeared to be a deficit in acquisition of the task and not just poor performance in LED on trials, as no difference in success rate was observed between performance during LED on trials compared to LED off trials in any of these sessions (Figure S5A). However, Jaws animals were able to learn the task and improve over the course of 5 subsequent training days if no

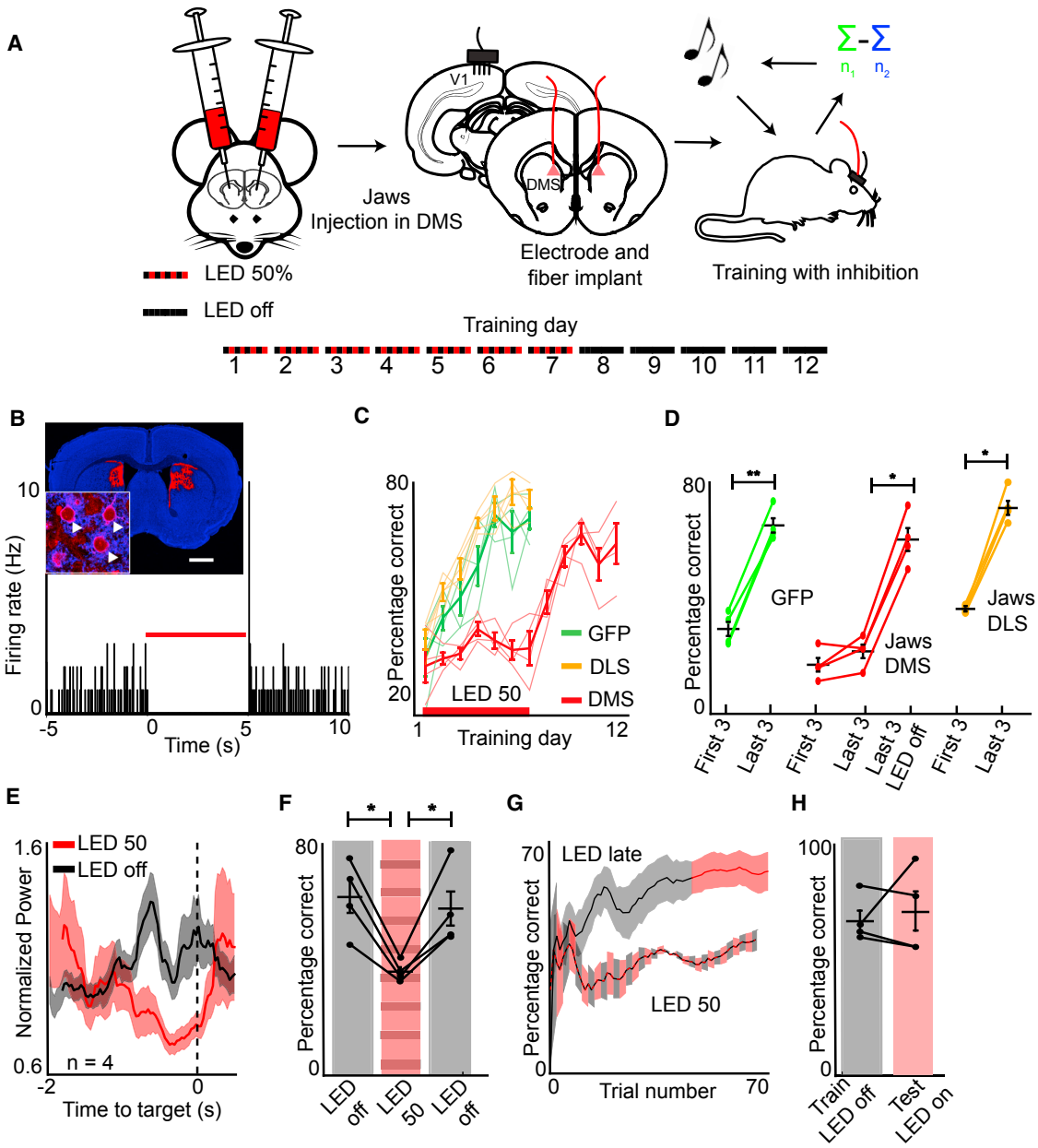

Figure 6. Optogenetic Inhibition of DMS in Mice Impairs Learning, but Not Performance, of V1-BMI

(A) Schematic showing the replication of rat experiments in Jaws-injected mice. Left: mice were injected bilaterally with rAAV8/CamKII-Jaws-KGC-GFP-ER2. Middle: after Jaws expression stabilized, mice were implanted unilaterally with an electrode array in V1 L5 and bilaterally in DMS or DLS. Right: animals were trained on the same task as rats (see Figure 1A) with the addition of optogenetic inhibition. Bottom: time course of experiments. LED 50% indicates that Jaws was activated via red LED light in 50% of trials.

(B) Jaws-mediated inhibition of DMS neurons. Inset shows a coronal section stained for neuronal cell bodies (blue) and Jaws expression (red). Scale bar, 1 mm. Zoomed inset shows a magnification with labeling apparent on single neurons (white arrows). Arrow edge, 10 μm. Histogram plot shows suppression of spike activity in one example putative DMS output neuron during Jaws-mediated inhibition (red bar). For an example interneuron, see Figure S5D.

(C) Percentage of rewarded trials for animals expressing Jaws in DMS (red) or DLS (orange) compared to control GFP (green) mice over the course of several days of training. Red bar indicates sessions where the LED was active on 50% of trials for both groups (LED 50). Error bars show SEM, and thin lines show data from individual animals.

(D) Quantification of performance across days for animals expressing Jaws in DMS (red; N = 4); animals expressing GFP in DMS (green, N = 4); and animals expressing Jaws in DLS (orange; n = 4). Each time period is a mean over 3 sessions. Mean GFP, first 3 sessions = 39.7%, mean GFP, last 3 sessions = 69.7%; p = 0.00072 (paired t test). Mean Jaws DMS, first 3 sessions = 29.3%, mean Jaws DMS, last 3 sessions LED 50 = 33.3%; p = 0.183. Mean Jaws DMS, last 3 sessions LED off = 62.1%. Paired t test between Jaws DMS last 3 sessions LED 50 and Jaws DMS LED off: p = 0.00131. Mean Jaws DLS, first 3 sessions = 52.3%,

(legend continued on next page)

LED activation occurred and activity DMS was unimpaired (Figures 6C and 6D). Interestingly, V1 LFP power in the gamma band (25–60 Hz) during sessions with Jaws inhibition was markedly reduced around the time of target hits, while gamma power during LED-off sessions in which animals successfully learned the task was similar to that observed in rats after learning (Figures 4E and 6E; Figures S5E and S5F). One possible explanation for these results may be that DMS activity is necessary for learning to modulate activity in V1. However, a second possibility may be that Jaws-mediated inhibition of DMS reduced animals' motivation to perform the task or that inactivation of these neurons created a distraction that prevented learning from taking place. To further control for these other possibilities, we performed the same experiments in a third cohort of mice in which Jaws expression was targeted to the dorsolateral striatum (DLS). The DLS has an established role in motor learning (McHaffie et al., 2005; Redgrave et al., 2010, 2011) but does not receive a substantial projection from V1 (Allen Institute for Brain Science, 2015; Faull et al., 1986; Hintiryan et al., 2016; Khibnik et al., 2014; McGeorge and Faull, 1989). Once expression was stable, DLS-Jaws mice were trained on the same task as DMS-Jaws mice, including LED activation to inhibit DLS activity on a randomized 50% of trials. Interestingly, animals were still able to learn the task in a similar manner to control animals with ongoing DLS inactivation (Figures 6C and 6D). Together, these results suggest that inhibition of DMS activity prevents animals from learning to generate arbitrary patterns of V1 activity in order to obtain a desirable outcome.

Optogenetic Inhibition of DMS Does Not Impair Learned Performance

In our task, the parameters used to translate neural activity to auditory tones were re-calculated each day such that at the start of every session animals were required to undergo some *de novo* learning in order to adapt to the new calibration. Despite this, animals were able to retain some memory of training from previous days to perform better over the course of each session as training progressed (Figure 2A; comparison between early and late phases). Interestingly, we observed that returning Jaws animals to an LED on 50% condition after several successful days of LED off training impaired their ability to learn during that session, even though they performed well above chance during the previous session and during following session when the LED was again inactive (Figure 6F). This could suggest that striatal activity is required for animals to learn or adapt to the initial parameters set at the beginning of each session; however, it could also indicate that striatal inhibition was interfering with task performance rather than acquisition. To disambiguate these two possibilities, we allowed animals to achieve saturating performance on the task within a single session (Train LED off, Figure 6H) and then continued to train animals with the LED turned on in every trial for the remainder of the session (Test LED on, Figure 6H). Interestingly, inhibition of the striatum after within-session learning had taken place did not impair animals' ability to perform the task, and they continued to perform well above chance level (Figures 6G and 6H) and in an indistinguishable manner from the LED off trials. These data suggest that striatal inhibition prevents animals from learning to modulate neural patterns in V1 through instrumental conditioning, but not from executing these patterns after learning has taken place.

DISCUSSION

Here we demonstrate that animals can learn to modulate spike activity in the primary visual cortex in a goal-directed manner using an abstract virtual task. These data demonstrate that feedback-based reinforcement learning can modulate activity on the scale of a few neurons, even in a primary sensory region that is strongly driven by external sensory input. Because we observed that successful performance occurred in both the presence and absence of light, it would appear that the learned modulation of V1 units in our task is an internally driven process. Such internally generated modulatory activity may arise from cognitive and attention-related control systems in the frontal cortices, for example, by leveraging circuits that modify sensory processing in V1 to fit task requirements (Makino and Komiyama, 2015; Zhang et al., 2014). Alternatively, movement and gaze-related activity is also known to change the activity of V1 neurons (Duhamel et al., 1992; Keller et al., 2012; Niell and Stryker, 2010), and it is possible that these mechanisms may still function in the absence of sensory input. However, the differential nature of the transform used to drive cursor activity was designed to discourage a movement-based strategy, and previous work demonstrated that movements were not used to solve a similar task, even when control was generated from the primary motor cortex (Koralek et al., 2012). Additionally, we observed that changing the light context, which would not be expected to affect movements, had a negative effect on task performance.

mean Jaws DLS, last 3 sessions = 77.0%; p = 0.00149. Colored lines show data for individual animals, horizontal bars show mean across animals, and error bars show SEM.

(E) Mean gamma power (25–60 Hz, solid lines) in V1 time locked to rewarded targets during LED-50 versus LED-off sessions in DMS-injected animals. Shaded areas show SEM. For color plots, see Figures S5E and S5F.

(F) Mean performance of trained DMS Jaws-injected animals after several days of training. Data are plotted in the order that the training sessions occurred. Black bars indicate sessions without LED activation. Striped red bars indicate sessions where the red LED was active on 50% of trials. N = 4 animals. Mean LED off, first session = 63.3%; mean LED 50% (second session) = 38.9%; mean LED off, last session = 59.6%. Paired t test between LED off, first session and LED 50%: p = 0.0122. Paired t test between LED 50% and LED off, last session: p = 0.0259. Paired t test between LED off, first and last sessions: p = 0.502. Error bars show SEM.

(G) Mean performance within a session for trained Jaws-DMS animals with late-session LED only activation compared to full-session LED 50% sessions. Shaded areas show SEM.

(H) Quantification of performance when Jaws-DMS animals were trained with LED off and tested with LED on in the same session; data same as in (F). N = 4 animals; mean of train LED off = 59.6%; mean of test LED on = 64.4%; p = 0.657 (paired t test). Lines show data from individual animals, horizontal bars show mean across animals, and error bars show SEM.

Finally, the lack of any observable learning deficit during DLS inhibition, a structure crucial for motor learning (Redgrave et al., 2010; Yin et al., 2006), would argue against a motor-based strategy. Identifying the source of the learned modulatory input in V1 is of great interest for future investigations. Taken along with a body of previous work describing brain-machine interface learning in other diverse cortical areas (Carmena et al., 2003; Cerf et al., 2010; Clancy et al., 2014; Fetz, 2007; Musallam et al., 2004; Prsa et al., 2017; Schafer and Moore, 2011; Shibata et al., 2011), these results suggest that this type of instrumental learning ability may be a common feature that tunes the activity of cortical circuits more generally.

In the realm of motor control, the cortico-basal ganglia circuit has been hypothesized to perform a selection function in which competing cortical motor programs are either maintained or released from inhibitory control (Costa, 2011; Redgrave et al., 2011). A similar function has also been postulated to operate in the realm of abstract cognition, by which various cognitive patterns may be selected that are appropriate for the current behavioral context, and have previously led to positive outcomes (Graybiel, 1997). These models propose an inhibitory feed-forward projection from basal ganglia output nuclei (globus pallidus internal (GPi) and substantia nigra pars reticulata (SNr) that can activate cortical programs when inhibition is transiently released from the thalamus. Interestingly, although basal ganglia outputs are known to target frontal cortical areas and even higher-order visual areas like area TE in the primate (Middleton and Strick, 1994, 1996), we are not aware of any direct projections from the basal ganglia that target V1-projecting thalamic nuclei.

Despite this, we observed that activity in the striatum was necessary for instrumental learning of neural patterns in the primary visual cortex. This result may be due to an induced learning deficit in a cortical region other than V1 whose input modulates ensemble 1 and ensemble 2 activity in the absence of visual stimulation. Frontal cortical areas, such as the cingulate cortex (Cg) in the rodent, are known to powerfully and directly influence processing in V1 to select and amplify representations of behaviorally relevant stimuli (Zhang et al., 2014). Thus, disrupting cortico-basal ganglia circuit function might impair learning of a top-down modulatory signal that generates rewarded activity in V1. Perhaps analogously, Parkinson's patients with abnormal basal ganglia function have been shown to be impaired in voluntary and sustained control of visual attention in the absence of eye movements (Wright et al., 1990; Yamaguchi and Kobayashi, 1998). Alternatively, the learning process we observed could have recruited structures upstream of V1, such as the superior colliculus, that also form connections with the basal ganglia (McHaffie et al., 2005). Our results do not rule out these or other possibilities; rather, they simply demonstrate that the basal ganglia are necessary to learn to modulate activity in V1 and that the input for this circuit is the striatum.

In our experiments, we observed that animals performed in a goal-directed manner: performance remained sensitive to changing task contingencies, even after many days of training (Figure 2E). The projection of the primary visual cortex to the striatum lies along the most medial-dorsal aspect, adjacent to the ventricle wall (Khibnik et al., 2014). This lies well within the dorsomedial division of the striatum, which is known to be necessary for and to facilitate goal-directed behaviors (Yin et al., 2005, 2009), as opposed to the dorsolateral division that is required for habitual action (Redgrave et al., 2010; Wickens et al., 2007; Yin et al., 2006). One possibility is that segregation of V1 projections in the dorsomedial division of the striatum favors goal-directed learning and behavior in V1. However, it is also possible that the daily recalibration of task parameters or simply insufficient training time prevented behavior from becoming habitual. Previous work utilizing a similar task design but controlled by neurons in M1 also observed that animals behaved in a goal-directed manner (Clancy et al., 2014; Koralek et al., 2012).

From our analyses, we observed that learning-related changes in neural dynamics, such as changes in correlations and spike-field coherence (Figure 4), were largely restricted to the direct population consisting of units from ensemble 1 and ensemble 2. Absolute modulation depth of indirect (non-E1 or E2) neurons in V1 remained low relative to direct units (Figures S3A–S3C), suggesting that the learning we observed operated primarily on the small scale of a few neurons. Furthermore, the modulation depth of task-irrelevant indirect neurons declined over the course of training (Figure S3C), echoing similar results observed across mice and monkeys using calcium imaging and electrophysiology techniques (Clancy et al., 2014; Ganguly et al., 2011; Prsa et al., 2017). It has been reported that the activity of single cells in sensory cortex is sufficient to drive a percept (Houweling and Brecht, 2008), suggesting that cortical circuits may be optimized to operate on these microscales. However, a closer analysis of indirect unit activity showed that many single units as well as the full population of indirect cells contained activity that was predictive of target choice (Figures S3E and S3F). These results suggest a subtle role for this population in the learning and execution of the task.

It has been reported previously that plasticity in the cortico-striatal circuit is required for animals to operantly learn to control patterns of activity in primary motor cortex, as assessed by selective deletion of striatal of NMDA receptors (Koralek et al., 2012). A complete dissection of the circuit that begins with changes at the cortico-striatal synapse and returns back to the task-relevant cortical neurons is a project that we find compelling and may warrant a study using a different set of methods. However, we can speculate on the mechanism using existing data. Although early models of cortico-basal ganglia-thalamo-cortical loops proposed a segregated, parallel model (Alexander et al., 1986), more recent data suggest that there is a substantial degree of interconnection and integration in the circuit, both within the striatum as well as the thalamus (Haber and Calzavara, 2009; Joel and Weiner, 1994). Regions of DMS that are recipient of V1 activity may increasingly engage frontal cortices through divergent and non-reciprocal projections to SNr and GPi, or act by changing synaptic weights in "hot spots" of convergence within the thalamus. The recruitment of frontal cortices may in turn increase the influence of top-down inputs to V1, which has been demonstrated to occur with experience (Makino and Komiyama, 2015). Precise tuning of the circuit that includes direct units may then happen locally, through recruitment of neuromodulatory systems that are known to influence V1 plasticity

(Gu, 2002, 2003), or through Hebbian learning mechanisms that can adjust the tuning of the relevant units to better suit the task at hand (Legenstein et al., 2009). Future projects will no doubt involve closer study of how nearby and distal neural populations support learning in a particular subset of cells.

An important goal of BMI research is often to decode movement parameters with high accuracy in order to translate a subjects' existing motor control repertoire into the movement of a complex artificial effector. In these cases, using high channel count recordings in motor cortices is an effective strategy due to the rich encoding of movement parameters in areas such as M1 and PMd. Here, our goal was not to optimize performance or control of an effector, but rather to develop a task that would enable us to study learning. It is important to note our task is different in several respects from BMI paradigms designed with the goal of maximizing performance and control in a therapeutic context—our goal was to use a BMI paradigm as a method of operantly conditioning neural activity in V1 directly in order to study the learning process and the neural dynamics associated with it. This enabled us to define the final output layer of neurons directly responsible for controlling a virtual action as well as their relationship to task performance and allowed us to observe their activity relative to each other, other V1 neurons, and activity in the dorsomedial striatum.

Although neurons in the primary visual cortex are thought to represent low-level visual features early in the visual processing stream, we observed that V1 neurons could learn to produce rewarded activity patterns that were independent of visual stimulation when spike activity was used as a control signal for a closed-loop brain-machine interface task. While here we focus on learning in the primary visual cortex, the dynamics of striatal activation, cortico-striatal dynamics over the course of learning, and the necessity of the striatum in the learning process is similar to what has been observed in a variety of tasks that engage diverse cortical regions (Barnes et al., 2005; Corbit and Janak, 2010; Koralek et al., 2012; Pasupathy and Miller, 2005; Shohamy et al., 2004; Yin et al., 2009). These results suggest that cortico-striatal projections, as part of larger cortico-basal ganglia circuits, play a generalizable role in shaping cortical activity based on ongoing experience and behavioral outcomes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Rats
  - Mice
- METHOD DETAILS
  - Electrophysiology
  - Behavioral Task
  - Behavioral Manipulations
  - Optical Inhibition
  - Data Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and one movie and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.01.051.

ACKNOWLEDGMENTS

We thank M. Correia and R. Oliveira for technical assistance and helpful discussions. This work was supported by the National Defense Science and Engineering fellowship to R.M.N., award #348880 from the Simons foundation, and award #413/14 from the Bial Foundation to A.C.K., European Research Council Consolidator Grant (COG 617142), HHMI International Early Career Scientist Grant (IEC 55007415), and ERA-Net NEURON grants to R.M.C., and Office of Naval Research grant N00014-15-1-2312 to J.M.C.

AUTHOR CONTRIBUTIONS

R.M.N., A.C.K., V.R.A., R.M.C., and J.M.C. designed experiments. R.M.N., A.C.K., and V.R.A. conducted experiments. R.M.N., A.C.K., V.R.A., R.M.C., and J.M.C. analyzed data and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 18, 2017
Revised: December 20, 2017
Accepted: January 29, 2018
Published: March 1, 2018

REFERENCES

Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., Shanfield, K., Hayes-Jackson, S., Aisen, M., Heck, C., et al. (2015). Neurophysiology. Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. Science *348*, 906–910.

Alexander, G.E., DeLong, M.R., and Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu. Rev. Neurosci. *9*, 357–381.

Allen Institute for Brain Science (2015). Allen Mouse Brain Atlas (Allen Mouse Brain Atlas).

Arduin, P.J., Frégnac, Y., Shulz, D.E., and Ego-Stengel, V. (2013). "Master" neurons induced by operant conditioning in rat motor cortex during a brain-machine interface task. J. Neurosci. *33*, 8308–8320.

Barnes, T.D., Kubota, Y., Hu, D., Jin, D.Z., and Graybiel, A.M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature *437*, 1158–1161.

Bouton, C.E., Shaikhouni, A., Annetta, N.V., Bockbrader, M.A., Friedenberg, D.A., Nielson, D.M., Sharma, G., Sederberg, P.B., Glenn, B.C., Mysiw, W.J., et al. (2016). Restoring cortical control of functional movement in a human with quadriplegia. Nature *533*, 247–250.

Brown, L.L., Schneider, J.S., and Lidsky, T.I. (1997). Sensory and cognitive functions of the basal ganglia. Curr. Opin. Neurobiol. *7*, 157–163.

Buffalo, E.A., Fries, P., Landman, R., Buschman, T.J., and Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. Proc. Natl. Acad. Sci. USA *108*, 11262–11267.

Carmena, J.M., Lebedev, M.A., Crist, R.E., O'Doherty, J.E., Santucci, D.M., Dimitrov, D.F., Patil, P.G., Henriquez, C.S., and Nicolelis, M.A. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. PLoS Biol. *1*, E42.

Cerf, M., Thiruvengadam, N., Mormann, F., Kraskov, A., Quiroga, R.Q., Koch, C., and Fried, I. (2010). On-line, voluntary control of human temporal lobe neurons. Nature *467*, 1104–1108.

Chuong, A.S., Miri, M.L., Busskamp, V., Matthews, G.A.C., Acker, L.C., Sørensen, A.T., Young, A., Klapoetke, N.C., Henninger, M.A., Kodandaramaiah, S.B., et al.

(2014). Noninvasive optical inhibition with a red-shifted microbial rhodopsin. Nat. Neurosci. *17*, 1123–1129.

Clancy, K.B., Koralek, A.C., Costa, R.M., Feldman, D.E., and Carmena, J.M. (2014). Volitional modulation of optically recorded calcium signals during neuroprosthetic learning. Nat. Neurosci. *17*, 807–809.

Collinger, J.L., Wodlinger, B., Downey, J.E., Wang, W., Tyler-Kabara, E.C., Weber, D.J., McMorland, A.J., Velliste, M., Boninger, M.L., and Schwartz, A.B. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. Lancet *381*, 557–564.

Corbit, L.H., and Janak, P.H. (2010). Posterior dorsomedial striatum is critical for both selective instrumental and Pavlovian reward learning. Eur. J. Neurosci. *31*, 1312–1321.

Costa, R.M. (2011). A selectionist account of de novo action learning. Curr. Opin. Neurobiol. *21*, 579–586.

Dias-Ferreira, E., Sousa, J.C., Melo, I., Morgado, P., Mesquita, A.R., Cerqueira, J.J., Costa, R.M., and Sousa, N. (2009). Chronic stress causes frontostriatal reorganization and affects decision-making. Science *325*, 621–625.

Duhamel, J.R., Colby, C.L., and Goldberg, M.E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. Science *255*, 90–92.

Engel, A.K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. Nat. Rev. Neurosci. *2*, 704–716.

Faull, R.L.M., Nauta, W.J.H., and Domesick, V.B. (1986). The visual corticostriato-nigral pathway in the rat. Neuroscience *19*, 1119–1132.

Fetz, E.E. (2007). Volitional control of neural activity: implications for brain-computer interfaces. J. Physiol. *579*, 571–579.

Ganguly, K., and Carmena, J.M. (2009). Emergence of a stable cortical map for neuroprosthetic control. PLoS Biol. *7*, e1000153.

Ganguly, K., Dimitrov, D.F., Wallis, J.D., and Carmena, J.M. (2011). Reversible large-scale modification of cortical networks during neuroprosthetic control. Nat. Neurosci. *14*, 662–667.

Gilja, V., Pandarinath, C., Blabe, C.H., Nuyujukian, P., Simeral, J.D., Sarma, A.A., Sorice, B.L., Perge, J.A., Jarosiewicz, B., Hochberg, L.R., et al. (2015). Clinical translation of a high-performance neural prosthesis. Nat. Med. *21*, 1142–1145.

Graybiel, A.M. (1997). The basal ganglia and cognitive pattern generators. Schizophr. Bull. *23*, 459–469.

Graybiel, A.M. (2008). Habits, rituals, and the evaluative brain. Annu. Rev. Neurosci. *31*, 359–387.

Gregoriou, G.G., Gotts, S.J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science *324*, 1207–1210.

Gu, Q. (2002). Neuromodulatory transmitter systems in the cortex and their role in cortical plasticity. Neuroscience *111*, 815–835.

Gu, Q. (2003). Contribution of acetylcholine to visual cortex plasticity. Neurobiol. Learn. Mem. *80*, 291–301.

Haber, S.N., and Calzavara, R. (2009). The cortico-basal ganglia integrative network: the role of the thalamus. Brain Res. Bull. *78*, 69–74.

Hikosaka, O., Nakahara, H., Rand, M.K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., and Doya, K. (1999). Parallel neural networks for learning sequential procedures. Trends Neurosci. *22*, 464–471.

Hinterberger, T., Veit, R., Wilhelm, B., Weiskopf, N., Vatine, J.J., and Birbaumer, N. (2005). Neuronal mechanisms underlying control of a brain-computer interface. Eur. J. Neurosci. *21*, 3169–3181.

Hintiryan, H., Foster, N.N., Bowman, I., Bay, M., Song, M.Y., Gou, L., Yamashita, S., Bienkowski, M.S., Zingg, B., Zhu, M., et al. (2016). The mouse cortico-striatal projectome. Nat. Neurosci. *19*, 1100–1114.

Hochberg, L.R., Bacher, D., Jarosiewicz, B., Masse, N.Y., Simeral, J.D., Vogel, J., Haddadin, S., Liu, J., Cash, S.S., van der Smagt, P., and Donoghue, J.P. (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature *485*, 372–375.

Houweling, A.R., and Brecht, M. (2008). Behavioural report of single neuron stimulation in somatosensory cortex. Nature *451*, 65–68.

Husain, M., Shapiro, K., Martin, J., and Kennard, C. (1997). Abnormal temporal dynamics of visual attention in spatial neglect patients. Nature *385*, 154–156.

Hwang, E.J., Bailey, P.M., and Andersen, R.A. (2013). Volitional control of neural activity relies on the natural motor repertoire. Curr. Biol. *23*, 353–361.

Jarosiewicz, B., Chase, S.M., Fraser, G.W., Velliste, M., Kass, R.E., and Schwartz, A.B. (2008). Functional network reorganization during learning in a brain-computer interface paradigm. Proc. Natl. Acad. Sci. USA *105*, 19486–19491.

Jarvis, M.R., and Mitra, P.P. (2001). Sampling properties of the spectrum and coherency of sequences of action potentials. Neural Comput. *13*, 717–749.

Jin, X., and Costa, R.M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature *466*, 457–462.

Joel, D., and Weiner, I. (1994). The organization of the basal ganglia-thalamocortical circuits: open interconnected rather than closed segregated. Neuroscience *63*, 363–379.

Keller, G.B., Bonhoeffer, T., and Hübener, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. Neuron *74*, 809–815.

Kemp, J.M., and Powell, T.P. (1970). The cortico-striate projection in the monkey. Brain *93*, 525–546.

Khibnik, L.A., Tritsch, N.X., and Sabatini, B.L. (2014). A direct projection from mouse primary visual cortex to dorsomedial striatum. PLoS ONE *9*, e104501.

Koralek, A.C., Jin, X., Long, J.D., 2nd, Costa, R.M., and Carmena, J.M. (2012). Corticostriatal plasticity is necessary for learning intentional neuroprosthetic skills. Nature *483*, 331–335.

Koralek, A.C., Costa, R.M., and Carmena, J.M. (2013). Temporally precise cell-specific coherence develops in corticostriatal networks during learning. Neuron *79*, 865–872.

Legenstein, R., Chase, S.M., Schwartz, A.B., and Maass, W. (2009). Functional network reorganization in motor cortex can be explained by reward-modulated Hebbian learning. Adv. Neural Inf. Process. Syst. *2009*, 1105–1113.

Lepage, K.Q., Kramer, M.A., and Eden, U.T. (2011). The dependence of spike field coherence on expected intensity. Neural Comput. *23*, 2209–2241.

Makino, H., and Komiyama, T. (2015). Learning enhances the relative impact of top-down processing in the visual cortex. Nat. Neurosci. *18*, 1116–1122.

Martínez, A., Anllo-Vento, L., Sereno, M.I., Frank, L.R., Buxton, R.B., Dubowitz, D.J., Wong, E.C., Hinrichs, H., Heinze, H.J., and Hillyard, S.A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat. Neurosci. *2*, 364–369.

McGeorge, A.J., and Faull, R.L.M. (1989). The organization of the projection from the cerebral cortex to the striatum in the rat. Neuroscience *29*, 503–537.

McHaffie, J.G., Stanford, T.R., Stein, B.E., Coizet, V., and Redgrave, P. (2005). Subcortical loops through the basal ganglia. Trends Neurosci. *28*, 401–407.

Mercuri, E., Atkinson, J., Braddick, O., Anker, S., Cowan, F., Rutherford, M., Pennock, J., and Dubowitz, L. (1997). Basal ganglia damage and impaired visual function in the newborn infant. Arch. Dis. Child. Fetal Neonatal *77*, F111–F114.

Middleton, F.A., and Strick, P.L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. Science *266*, 458–461.

Middleton, F.A., and Strick, P.L. (1996). The temporal lobe is a target of output from the basal ganglia. Proc. Natl. Acad. Sci. USA *93*, 8683–8687.

Musallam, S., Corneil, B.D., Greger, B., Scherberger, H., and Andersen, R.A. (2004). Cognitive control signals for neural prosthetics. Science *305*, 258–262.

Niell, C.M.C.C.M., and Stryker, M.P.M. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. Neuron *65*, 472–479.

Pasupathy, A., and Miller, E.K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature *433*, 873–876.

Prsa, M., Galiñanes, G.L., and Huber, D. (2017). Rapid integration of artificial sensory feedback during operant conditioning of motor cortex neurons. Neuron *93*, 929–939.e6.

Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M.C., Lehericy, S., Bergman, H., Agid, Y., DeLong, M.R., and Obeso, J.A. (2010). Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. Nat. Rev. Neurosci. *11*, 760–772.

Redgrave, P., Vautrelle, N., and Reynolds, J.N.J. (2011). Functional properties of the basal ganglia's re-entrant loop architecture: selection and reinforcement. Neuroscience *198*, 138–151.

Sadtler, P.T., Quick, K.M., Golub, M.D., Chase, S.M., Ryu, S.I., Tyler-Kabara, E.C., Yu, B.M., and Batista, A.P. (2014). Neural constraints on learning. Nature *512*, 423–426.

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. Science *310*, 1337–1340.

Schafer, R.J., and Moore, T. (2011). Selective attention from voluntary control of neurons in prefrontal cortex. Science *332*, 1568–1571.

Shibata, K., Watanabe, T., Sasaki, Y., and Kawato, M. (2011). Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. Science *334*, 1413–1415.

Shohamy, D., Myers, C.E., Grossman, S., Sage, J., Gluck, M.A., and Poldrack, R.A. (2004). Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain *127*, 851–859.

Shuler, M.G., and Bear, M.F. (2006). Reward timing in the primary visual cortex. Science *311*, 1606–1609.

Steinmetz, P.N., Roy, A., Fitzgerald, P.J., Hsiao, S.S., Johnson, K.O., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. Nature *404*, 187–190.

Suner, S., Fellows, M.R., Vargas-Irwin, C., Nakata, G.K., and Donoghue, J.P. (2005). Reliability of signals from a chronically implanted, silicon-based electrode array in non-human primate primary motor cortex. IEEE Trans. Neural Syst. Rehabil. Eng. *13*, 524–541.

Swanson, L.W. (2000). Cerebral hemisphere regulation of motivated behavior. Brain Res. *886*, 113–164.

Thomson, D.J. (1982). Spectrum estimation and harmonic analysis. Proc. IEEE *70*, 1055–1096.

Tricomi, E.M., Delgado, M.R., and Fiez, J.A. (2004). Modulation of caudate activity by action contingency. Neuron *41*, 281–292.

Webster, K.E. (1965). The cortico-striatal projection in the cat. J. Anat. *99*, 329–337.

Wickens, J.R., Horvitz, J.C., Costa, R.M., and Killcross, S. (2007). Dopaminergic mechanisms in actions and habits. J. Neurosci. *27*, 8181–8183.

Wright, M.J., Burns, R.J., Geffen, G.M., and Geffen, L.B. (1990). Covert orientation of visual attention in Parkinson's disease: an impairment in the maintenance of attention. Neuropsychologia *28*, 151–159.

Yamaguchi, S., and Kobayashi, S. (1998). Contributions of the dopaminergic system to voluntary and automatic orienting of visuospatial attention. J. Neurosci. *18*, 1869–1878.

Yin, H.H., Ostlund, S.B., Knowlton, B.J., and Balleine, B.W. (2005). The role of the dorsomedial striatum in instrumental conditioning. Eur. J. Neurosci. *22*, 513–523.

Yin, H.H., Knowlton, B.J., and Balleine, B.W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav. Brain Res. *166*, 189–196.

Yin, H.H., Mulcare, S.P., Hilário, M.R.F., Clouse, E., Holloway, T., Davis, M.I., Hansson, A.C., Lovinger, D.M., and Costa, R.M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat. Neurosci. *12*, 333–341.

Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., and Dan, Y. (2014). Selective attention. Long-range and local circuits for top-down modulation of visual cortex processing. Science *345*, 660–665.

Zhong, Y., and Bellamkonda, R.V. (2007). Dexamethasone-coated neural probes elicit attenuated inflammatory response and neuronal loss compared to uncoated neural probes. Brain Res. *1148*, 15–27.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
rAAV8/CamKII-Jaws-KGC-GFP-ER2	University of North Carolina	Addgene plasmid #65015
rAAV8/CamKII-GFP	University of North Carolina	Addgene plasmid #64545; GenBank: KM000926.1
Experimental Models: Organisms/Strains		
Long-Evans rats (Male, between 200 g and 300 g)	Charles River	Crl:LE Strain Code: 006
Mice, C57BL, (Male, between 2.5-3.5 months old)	Jackson Labs	000664
Software and Algorithms		
Python 2.7; Anaconda distribution	Anaconda	https://www.anaconda.com/download/
Chronux toolbox	Chronux	http://chronux.org; RRID: SCR_005547
Scikit-learn toolbox	Scikit-learn	http://www.scikit-learn.org; RRID: SCR_002577
Plexon SortClient	Plexon	http://plexon.com/products/software; RRID: SCR_003170
Other		
Plexon MAP System	Plexon	http://plexon.com/products/multichannel-acquisition-processor-map-data-acquisition-system
Microwire arrays- fixed	Innovative Neurophysiology	http://www.inphysiology.com/fixed-arrays/
Microwire arrays- moveable	Innovative Neurophysiology	http://www.inphysiology.com/movable-arrays/
Fiber-coupled LED system	Prizmatix	http://www.prizmatix.com/optogenetics/Prizmatix-in-vivo-Optogenetics-Toolbox.htm
OmniPlex system	Plexon	https://plexon.com/products/omniplex-d-neural-data-acquisition-system-1/
Operant test chamber	Lafayette Neuroscience	Model 80004NS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Jose M. Carmena (jcarmena@berkeley.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Rats

All rat experiments were performed in compliance with the regulations of the Animal Care and Use Committee at the University of California, Berkeley. A total of 13 singly housed, male Long-Evans rats on a 12h light/dark cycle weighing roughly 250 g were used for the experiments. All rats were chronically implanted with microwire arrays in V1, with 8 rats also receiving implants in the dorsomedial striatum. Each array contained 16 or 32 tungsten microelectrodes (35 μm diameter, 250 μm electrode spacing, 8x2 or 8x4 configuration; Innovative Neurophysiology, Durham, NC). Stereotactic coordinates relative to bregma and lambda were used to center the arrays (1 mm anterior of lambda, 3.5 mm lateral from the midline, and 1.25 mm ventral from the cortical surface for V1; 1.8 mm anterior of bregma, 2 mm lateral of the midline, and 5.5 mm ventral from the cortical surface for DMS). V1 implants were targeted for layer 5 neurons based on insertion depth, and this was verified histologically at the end of experiments (Figure S1A,b). Rodents were anesthetized with Ketamine (50 mg/kg) and Xylazine (5 mg/kg) with supplemental isoflurance gas as needed. Craniectomies were sealed with cyanoacrylate and rats were allowed to recover for ten days after implantation before behavioral training. Rats were given dexamethasone treatment (0.5 mg/kg) for one week following surgery to minimize tissue damage around the implant (Zhong and Bellamkonda, 2007).

Mice

Mouse experiments were performed in accordance with the Champalimaud Centre for the Unknown Ethics Committee guidelines and approved by the Portuguese Veterinary General Board (Direcção Geral de Veterinária, approval 0421/000/000/2014). A total of 12 singly housed male C57BL mice of 2.5-3.5 months of age were used (four experimental mice with DMS-targeted Jaws, four with DLS-targeted Jaws, and four DMS-targeted GFP control mice). Surgeries were performed under isoflurane anesthesia (1%–3%). All mice with DMS-targeted injections were injected bilaterally with 1 μL of viral solution in the dorsomedial striatum using coordinates relative to Bregma (0.9 mm AP, ± 1.5 mm ML, 2 mm below brain surface). For DLS injections, 0.5 μL was used and the coordinates used were +0.5 mm AP, +/−2.3 ML, and −2.3 DV. Viral solution was injected through a glass pipette by pressure (Nanojet II from Drummond Scientific, 4.6 nL pulses at a rate of 0.2 Hz). For experimental animals, the virus injected was rAAV8/CamKII-Jaws-KGC-GFP-ER2 (University of North Carolina, titer $5.9x10^{12}$). For control animals, the virus injected was rAAV8/CamKII-GFP (University of North Carolina, titer $2.8x10^{12}$). For delivery of red light, mice were implanted bilaterally with optical fibers (250 μm diameter, NA 0.63). All mice were also implanted with 16-channel movable electrode arrays (electrode diameter 23 μm; Innovative Neurophysiology, Durham, NC) in the right primary visual cortex using coordinates relative to Lambda (0 mm AP, 0.3 mm ML, 0.6 mm below brain surface).

METHOD DETAILS

Electrophysiology

Single unit activity and local field potentials were simultaneously recorded with a Multichannel Acquisition Processor (MAP in rats, OmniPlex in mice; Plexon, Dallas, TX). Activity was sorted using an online sorting application (Plexon, Dallas, TX) prior to each daily recording session. Only units with a clearly identified waveform and signal-to-noise ratio greater than 2 were used. Sorting templates were further refined using an offline sorting application (Plexon, Dallas, TX). Behavioral timestamps were sent to the MAP recording system through custom Python and C++ programs and synchronized to the neural data for later analyses. Recording arrays were grounded to a screw in the occipital bone, and both arrays were also referenced locally using the online program Ref2 (Plexon, Dallas, TX) to eliminate effects of volume conduction. For referencing, an electrode on each array was chosen to be subtracted from all other electrodes on that array. This was done independently for both V1 and DMS.

Behavioral Task

After recovering from surgery, rats were trained on the task in a dark behavioral chamber (Lafayette Instrument Company, Lafayette, IL) unless otherwise specified. During training, rats only received access to water during the task unless supplemental water was needed to maintain their body weight at a healthy level. At the start of each session, two ensembles of 2 well-isolated V1 units each were chosen for inclusion in the direct population based on SNR, interspike- interval histograms, and refractory periods. No other selection criteria were used to partition the recorded cells into each ensemble. Although these direct units were consistently well-isolated, we also ensured that many well-isolated units were included in the indirect population to enable a proper comparison. The units assigned to the direct population remained relatively constant throughout training using the stability of spike waveforms, sorting templates, and interspike intervals across sessions as a guide. After sorting and partitioning of direct and indirect units, a 15-minute baseline period was recorded in which animals received a sucrose water reward on a variable-interval schedule. During this time, spike counts were recorded for each ensemble binned into 100 ms bins, and a distribution of state values was calculated by subtracting the counts from E1 from E2 in each time bin. From here, the distribution was fit by a Gaussian mixture model (GMM) comprising between 1 and 10 Gaussian components (the exact number was chosen by finding the Akaike Information Criterion (AIC) value for each possible number of components, and choosing the number with the lowest AIC value). The probability density function (PDF) was then computed for the chosen model. By default, the neural state values (E1 – E2 counts for a 100 ms bin) corresponding to the low and high targets were set at the points on the computed PDF where the area under the curve was equivalent to 1.5% and 98.5% of the total area, respectively (Figure S1A). However, these target values were iteratively updated by running a simulation using the data recorded in the baseline period until the probability of hitting each target was approximately 30%. Finally, using the lowest and highest target values as well as the mean of the GMM, in combination with the lowest (1000 Hz), highest (15000 Hz) and midpoint (7000 Hz) frequencies, a 2-degree polynomial function was fit to these values in order to map neural state to a frequency value. During online performance, the state value used to compute the instantaneous feedback frequency was smoothed over the most recent 10 bins, updated every 100 ms. The rodents had to then precisely modulate these neuronal ensembles to move the cursor to one of two target frequencies, one which was randomly chosen using a coin flip on a per-animal basis to be associated with a 20% sucrose solution reward (and kept consistent across training sessions). Rodents were free to reach either target, although the cursor value had to return to the middle value for a new trial to begin. A trial was marked as a miss if neither of these target states were achieved within 30 s of trial initiation. Recorded neural data were entered in real time to custom routines in Python and C++ that then translated those activity levels into the appropriate feedback frequency as described above and played the frequency on speakers mounted above the behavioral chamber. When a target was hit, a Data Acquisition board (National Instruments, Austin, TX) controlled by a Python script triggered the operant box to supply the appropriate reward to rodents.

Behavioral Manipulations

After initial training of at least 8 days, a contingency degradation was performed. This took place over two sessions: during the first session, animals were allowed to achieve stable performance on the task, which took approximately 30 min (pre-degradation). Then, reward delivery was decoupled from task performance and delivered on a variable-interval schedule that approximated the rate of reward during training conditions (CD). During this time, animals still received auditory feedback linked to their neural state. The contingency degradation continued for the remainder of the session. The next day, animals were again trained on the task under normal conditions (reinstatement). Similarly, for contingency reversal sessions, we reversed the contingency approximately 30 min into a training session. During sessions that involved light manipulation, we again waited for animals to achieve saturating performance in a single session, and then placed a light into the behavioral chamber in an overhead configuration.

Optical Inhibition

For optical inhibition experiments, red light was applied to the striatum of both experimental and control groups on 50% of all trials in a session. Light was applied through a fiber-coupled LED system (Prizmatix, Givat-Shmuel, Israel). Power levels tested through the system at the optical fiber tip ranged from 4-6 mW. Red light application consisted of a square pulse that lasted the full duration of the current trial, from trial initiation until either a target or timeout was reached. Both groups were trained in this manner for 7 days. Next, the experimental animals were trained for 5 additional days in the absence of red light. After this initial training, both groups were tested to determine the role of striatal circuits in learning versus performance of the task. On day LED 50, red light was applied on 50% of all trials. On day LED late, no light was applied for the first 45 trials, after which red light was applied on 50% of all remaining trials.

Data Analysis

Analyses were performed in Python with custom- written routines utilizing publicly available software packages. Unit data were first binned in 1 ms time bins and digitized. To classify recorded striatal units as either output neurons or interneurons, we used the method of (Jin and Costa, 2010). Briefly, units with a waveform trough half-width of less than 100 μs and a baseline firing rate of more than 10 Hz were considered to be fast-spiking interneurons, while units that did not meet these criteria were classified as output neurons. Approximately 91% of the units we recorded were classified as output neurons. Firing rate analyses were performed in relation to target achievement as indicated in figures. Unless otherwise specified, firing rates were binned into 50 ms bins for all analyses. Only two-sided statistical tests were used to determine significance. The term "early" indicates that analyses were performed using data taken from animals during the first 3 days of training (inclusive), while "late" specifies data taken from animals during the final 3 days of training. The "first trials" of a session indicates trials in the first 10 min, while the "last trials" of a session are defined as occurring in the last 10 min, unless otherwise specified in the text. Modulation depths were computed as the difference between the maximum and minimum firing rate values in a 2 s window centered around target achievement. Coherence analyses were performed using algorithms translated to Python from the Chronux toolbox (http://chronux.org) in conjunction with custom routines in Python. A multi-taper method was used to compute spectral estimates of spiking and LFP activity (Jarvis and Mitra, 2001; Thomson, 1982). A total of 5 tapers were used with a time-bandwidth product of 3, and estimates were computed every 50 ms with a window size of 500 ms. Coherence between spiking in LFP activity was calculated and defined as:

$$C_{xy} = \frac{|R_{xy}|}{\sqrt{|R_{xx}|}\sqrt{|R_{yy}|}}$$

where R_{xx} and R_{yy} are the power spectra and R_{xy} is the cross-spectrum. Spectral analyses were calculated relative to the delivery of reward and averaged across trials and animals.

Coherence estimates can be affected by firing rate(Lepage et al., 2011) and we therefore performed a thinning procedure to equate firing rates between conditions in which rates differed (Gregoriou et al., 2009). Trial-averaged spike trains in the neuronal populations were smoothed with a moving average of 10 ms. The difference in firing rate between the populations normalized by the maximum firing rate at a given time point determined the probability that a spike would need to be removed from the population with a higher firing rate. Spikes were then removed from the population with a higher firing rate based on this probability in order to eliminate any possible influence of firing rate on coherence estimates.

The signal-to-noise ratio for each recorded waveform was quantified as:

$$SNR = \frac{A}{2 * SD_{noise}}$$

where A is the peak-to-peak voltage of the mean waveform and SD_{noise} is the standard deviation of the residuals from each waveform after the mean waveform has been subtracted (Suner et al., 2005). Units included in the analysis had a minimum SNR of 2.

For logistic regression analyses, we used functions from the publicly available python package scikit-learn (http://.scikit-learn.org). Regression was performed using a window of spike activity 500 ms prior to target hits, binned into 50 ms bins. L2 Regularization was done using cross-validation to determine the optimum regularization parameter. 3-fold cross validation was performed 5 times using left out data to compute accuracies, and the average of all 5 results was taken to be the final accuracy value. Chance rates were taken

as the accuracy of the analyses using shuffled data. To determine significance values, a permutation test was used that compared the accuracy of the prediction using the original dataset compared to dataset in which target identities for all trials were shuffled. Neural activity was considered to be significantly predictive of target choice if the accuracy of the prediction using the original dataset outperformed the accuracy of the shuffled version on 95% of 500 iterations.

The linear regression model was fit for each session as follows. First, a data window was defined for each trial as the 500 ms prior to target hit. Summed spike counts over this interval were computed for all DMS units at every trial in the session, and arranged to create a matrix, X, of dimensions (trials * units), where each entry is a spike count. Then for each direct (V1) unit, a similar process was used to create a vector, y, for that particular unit, with length trials, and each entry the spike count over the interval for that trial. These matrices were used to fit an ordinary least-squares regression and estimate the coefficients of the regression model, which was done using the LinearRegression method of the python package scikit-learn (sklearn.org). R-squared (variance explained) was computed using three-fold cross validation. In order to compute the significance of the model prediction, we used a permutation test with 10,000 iterations. On each iteration, the vector y was shuffled and the R-squared was computed for the shuffled data. The prediction was considered significant if the intact dataset resulted in a greater quantity of explained variance than shuffled datasets on 95% of iterations. This process was repeated for each direct unit using the same X matrix of DMS unit data. The resulting R-squared value for each direct unit was then averaged across all direct units to compute the mean R-squared value for that session.

QUANTIFICATION AND STATISTICAL ANALYSIS

In all figures, unless otherwise specified, the bold lines represent the mean of the data, and error bars denote standard error of the mean (SEM). Translucent lines or dots are used to show data from individual animals, as indicated in the Figure legends. Means, standard error, and statistical tests were computed across animals. Number of subjects and sessions analyzed for each figure are indicated in the figure legends or in the figure panel. ''Percent correct'' was calculated by dividing the total number of correct trials in a given interval by the sum of the total trials completed in that interval, including trials that resulted in a rewarded target hit, an unrewarded target hit, or a time-out. Many comparisons of learning were done by analyzing performance differences between different epochs- as indicated in the Results, ''early'' and ''late'' refer to the first 3 and last 3 training sessions for each animal, respectively. Sometimes, we analyzed ''first trials'' and ''last trials,'' which refers to the trials completed in the first 10 min of a session and last 10 min of a session, respectively. The epoch that is being analyzed in a given figure is clearly specified in the Results and Figure legend. Determination of chance performance rates is described above in Method Details, under the subheading Behavioral task. The statistical test used in each comparison is specified in the Figure legend, along with the resulting p value. Parametric tests were used to compare performance across training epochs, while permutation tests were used to derive significance levels for regression models. Significance was considered as $p < 0.05$. A single asterisk was used to denote significance of at least $p < 0.05$, while a double asterisk was used to denote $p < 0.001$. All statistical tests were performed using custom-written scripts in Python.