Covert skill learning in a cortical-basal ganglia circuit

Jonathan D. Charlesworth[†], **Timothy L. Warren**[†], and **Michael S. Brainard**[†]

[†]W. M. Keck Center for Integrative Neuroscience, Department of Physiology, and the Neuroscience Graduate Program, University of California, San Francisco, California 94143, USA

We learn complex skills like speech and dance through a gradual process of trial-and-error. Cortical-basal ganglia circuits play an important yet unresolved role in such trial-and-error skill learning[1]; influential 'actor-critic' models propose that basal ganglia circuits generate a variety of behaviors during training and learn to implement the successful behaviors in their repertoire[2–3]. Here we show that the anterior forebrain pathway (AFP), a cortical-basal ganglia circuit[4], contributes to skill learning even when it does not contribute to such 'exploratory' variation in behavioral performance during training. Blocking the output of the AFP while training Bengalese finches to modify their songs prevented the gradual improvement that normally occurs in this complex skill during training. Surprisingly, however, unblocking the output of the AFP after training caused an immediate transition from naïve performance to excellent performance, indicating that the AFP covertly gained the ability to implement learned skill performance without contributing to skill practice. In contrast, inactivating the AFP nucleus LMAN during training completely prevented learning, indicating that learning requires activity within the AFP during training. Our results suggest a revised model of skill learning: basal ganglia circuits can monitor the consequences of behavioral variation produced by other brain regions and then direct those brain regions to implement more successful behaviors. The ability of the AFP to identify successful performances generated by other brain regions indicates that basal ganglia circuits receive a remarkably detailed efference copy of premotor activity in those regions. The capacity of the AFP to implement successful performances that were initially produced by other brain regions indicates precise functional connections between basal ganglia circuits and the motor regions that directly control performance.

We assessed the contributions of basal ganglia circuitry to learned modification of adult Bengalese finch song, a complex behavior consisting of a sequence of 30–100ms long 'syllables,' each with a highly stereotyped acoustic structure. The song-specific motor control system consists of a motor pathway, which is analogous to mammalian premotor and primary motor cortex and is sufficient to produce well-learned elements of song, and the AFP, a cortical-basal ganglia circuit that is necessary for juvenile song learning and adult song modification[4]. We elicited learning by training birds with aversive reinforcement contingent on the fundamental frequency of individually targeted syllables (Figure 1a–b).

Aversive reinforcement consisted of loud, 50–80ms bursts of white noise[5–6]. Training with aversive reinforcement elicited changes to fundamental frequency that adaptively reduced white noise exposure; delivering white noise to performances of a syllable with fundamental frequency below a threshold elicited an increase in mean fundamental frequency of that syllable (Figure 1b) whereas delivery of white noise to performances with fundamental frequency above that threshold elicited a decrease in mean fundamental frequency. These adaptive changes developed within hours and were specific to fundamental frequency of the targeted syllable.

Influential actor-critic models[2–3], inspired by reinforcement learning theory[7] and supported by empirical evidence[8–9], propose that basal ganglia circuits such as the AFP are a crucial substrate for trial-and-error learning, generating a variety of behavioral performances and ultimately implementing only the performances that have led to successful outcomes. In the context of fundamental frequency modification (Figure 1a–b), the actor-critic model proposes that on each trial the AFP (the actor) generates distinct fundamental frequency values (exploratory behavioral variation, Figure 1c), receives reinforcement signals about the consequences of that variation from dopaminergic neurons (the critic, Figure 1d), and changes the probability of generating that fundamental frequency value in the future based on its consequences[4,10–12]. Over time, the AFP gradually adjusts its output to implement (i.e. cause the execution of) behaviors with better consequences, leading to adaptive changes in fundamental frequency and thus improved skill performance (Figure 1e). Consistent with this model, blocking AFP output through lesions or reversible inactivations reduces song variation, indicating that the AFP generates variation in song performance that might serve as motor exploration[4–5] (Figure 1c,f). Moreover, blocking AFP output after learning reduces the expression of recently learned song changes, suggesting that the AFP can contribute to learning by biasing the motor pathway to implement more successful behaviors[13–14] (as suggested in Figure 1e). A critical yet untested proposition of this model is that learning requires reinforcement of exploratory behavioral variation generated by the AFP, and thus preventing the AFP from contributing to behavioral variation during training should prevent trial-and-error learning (Figure 1f–g).

We tested this prediction by pharmacologically blocking the output of the AFP, training birds with aversive reinforcement, and then unblocking the output of the AFP. To block contributions of the AFP to exploratory variation in song during training, while leaving intrinsic AFP circuitry intact, we exploited a pharmacological distinction between inputs that song motor nucleus RA receives from premotor nucleus HVC and from AFP nucleus LMAN. Inputs from LMAN are mediated almost exclusively by NMDA receptors whereas inputs from HVC are mediated by both NMDA and AMPA receptors[4] (Figure 2a). Thus, to reversibly disrupt AFP output, we inserted microdialysis probes into RA and used retrodialysis to switch between a control solution (ACSF) and a solution containing 1–5mM of the NMDA receptor antagonist APV (Figure 2a). Consistent with previous reports[14–15], this manipulation affected song in the same manner as pharmacological inactivations or lesions of LMAN[14,16], reducing the coefficient of variation (CV) of fundamental frequency by 31.7 +/− 5.6% (n=12 syllables in 9 birds) without causing systematic changes in song structure (Figure 2b-c, Supplementary Figure 2). The APV-dependent reduction in song variation was reversible; switching the infusion solution back to ACSF restored the CV of

fundamental frequency to 96.5 +/− 4.6% of baseline (Figure 2c, Supplementary Figure 2c). These data indicate that infusing APV into RA effectively and reversibly prevents the AFP from contributing to song variation (as schematized in Figure 1c,f).

As predicted by an actor-critic model of AFP function, there was no expression of learning while AFP output was blocked during training. We compared learning in control experiments (e.g. Figure 3a) to learning in experiments with APV in RA throughout training (e.g. Figure 3c). Training consisted of administering aversive reinforcement contingent on the fundamental frequency of a targeted syllable (Figure 1a–b). To ensure that a similar proportion of syllable renditions received aversive reinforcement across experiments despite the reduced range of variation following APV infusion, we set the threshold for avoiding white noise at approximately the baseline median fundamental frequency for each targeted syllable (see Online Methods). To simplify presentation, we have plotted data so that the direction of learning (that reduces white noise exposure) is always upwards. For control experiments (n=14 experiments for 9 syllables in 7 birds), there was significant expression of learning during the training period; the mean shift of fundamental frequency in the adaptive direction was 33.5Hz, corresponding to a 1.1 +/− 0.35% change in fundamental frequency (Figure 3b, left bar, P<0.01, signed-rank test). In contrast, for experiments with APV in RA (n=21 experiments for 12 syllables in 9 birds), there was no expression of learning during the training period (Figure 3d, left bar); the mean shift in fundamental frequency was 5.3Hz (a 0.20 +/− 0.15% change) which was significantly less than in control conditions (P=0.02, rank-sum test) and not significantly different from zero (P=0.15, signed-rank test). These results indicate that infusing APV into RA eliminates any expression of learning during training and thus provide further support that this manipulation blocks AFP output.

Surprisingly, learned changes to song appeared immediately when AFP output was unblocked after training. If learning required the AFP to transmit song variation during training, as predicted by an actor-critic model of AFP function, then blocking AFP output during training should have prevented learning and thus unblocking AFP output after training should not have revealed any learned changes to fundamental frequency (Figure 1f–g). Contrary to this prediction, we observed learned changes to fundamental frequency after unblocking AFP output (Figure 3c–d). These learned changes could not be predicted by any subtle changes in fundamental frequency during training (Supplementary Figure 3) and were specific to the fundamental frequency of the targeted syllable (Figure 3e, Supplementary Figure 4). The average learned change across experiments was 27.6Hz, corresponding to a 0.99 +/− 0.17% change in fundamental frequency (n=21 experiments in 9 birds, P<0.001, signed-rank test, Figure 3d, right bar). The magnitude of learning expressed after training was statistically indistinguishable from the magnitude of learning in control experiments (Figure 3b,d, right bars, P>0.9, rank-sum test). In contrast to the gradual progression of learning in control experiments, maximal learning was expressed immediately after unblocking AFP output and did not require further practice with AFP output unblocked (Figure 3f). Thus, during training with AFP output blocked, the AFP had not only encoded a 'policy' specifying the change in song that would improve outcomes (e.g. fundamental frequency of the targeted syllable should be increased), but had already altered its activity to implement that change.

The acquisition of learning during training with APV in RA is consistent with three classes of mechanisms. First, learning could require activity in the AFP during training. Second, learning could require plasticity upstream of the AFP, possibly in the ventral tegmental area (VTA), and the AFP could merely serve as a conduit between the site of plasticity and behavioral output. Third, learning could require plasticity downstream of the AFP, in RA, but the expression of that learning could be gated by AFP output[14]. To discriminate between these possible mechanisms, we inactivated LMAN during training, by infusing muscimol (n=12 experiments in 3 birds) or lidocaine (n=2 experiments in 1 bird) into LMAN (Figure 4a). Whereas infusing APV into RA blocks AFP output while leaving activity in the AFP intact, inactivating LMAN not only blocks AFP output but also disrupts activity within the AFP.

We found that activity in LMAN during training is crucial for learning. Inactivating LMAN reversibly reduced variation in fundamental frequency by the same amount as lesions of LMAN or infusion of APV into RA (CV reduction of 31.2 +/− 6.5%, n=14, Supplementary Figure 2b). Importantly for the interpretation of these experiments, we ensured in each case that the threshold for reinforcement continued to provide a directed instructive signal during the training period despite the reduced range of fundamental frequency variation (as in APV experiments, see Online Methods)[6]. As with infusing APV into RA, inactivating LMAN prevented any expression of learning during training; expression of learning during training with LMAN inactivated was -0.19 +/− 0.37% (n=14, P=0.9 signed-rank test) compared to 0.90 +/− 0.09% (n=14, P=1.2e-4 signed-rank test) in control experiments (Figure 4b–d). However, in contrast to experiments with APV in RA, inactivation of LMAN during training prevented any acquisition of learning as assessed following the washout of drug (-0.07 +/− 0.21%, n=14, P=0.95 signed-rank test, Figure 4b–d). These results demonstrate that inactivating AFP nucleus LMAN during training prevents the acquisition of learning and thus activity within the AFP during training is essential for learning.

Together, our results indicate that the capacity to adaptively modify a complex motor skill developed within the AFP during training with AFP output blocked. The prevention of learning by inactivating LMAN during training indicates that activity in the AFP is required for learning (Figure 4). The immediate transition from naïve performance to learned performance when we unblocked AFP output after training (Figure 3) demonstrates that, during training, the AFP had gained the ability to improve behavior even though that improvement was not yet expressed. For simpler forms of conditioning[17–18], such covert learning, indicating learning-related plasticity in the brain that is not accompanied by behavioral improvement, would only require that the brain region involved in learning received coarse signals about actions and stimuli[19]. In contrast, our results indicate that the brain region involved in learning, the AFP, receives detailed information (an efference copy[20]) about the precise dynamics and timing of behavioral performance from the other brain regions controlling that performance.

Our results motivate a revision to models of song plasticity[10–12] and influential actor-critic models of skill learning[2–3], which propose that essential learning-related signals develop only in brain regions that are "acting" (i.e. controlling behavior). In contrast, our results indicate that the essential learning-related signals necessary to adaptively bias behavior

develop in a basal ganglia circuit, the AFP, while it is prevented from contributing to behavioral performance and motor exploration. This indicates that motor exploration (i.e. variation) generated by the AFP is not necessary for learning and thus a source of variation independent of the AFP can be exploited for reinforcement learning. Presumably, this variation arises in the motor pathway, possibly in RA[21–22], and is transmitted to the AFP. Under normal circumstances with AFP output intact, variation contributed by the AFP itself may also be used for reinforcement learning. Thus, the AFP may be a specialized hub where information about behavioral variation from multiple sources converges and is associated with reinforcement signals to guide learning.

The specificity of learning with AFP output blocked (Figure 3e, Supplementary Figure 4) implies that the AFP associates reinforcement signals with detailed information about ongoing song performance, including both the identity of the syllable being produced and the rendition-by-rendition variation in the fundamental frequency of that syllable. Reinforcement signals, indicating the presence or absence of white noise, could be conveyed to the AFP via known projections from neuromodulatory nuclei such as the ventral tegmental area (VTA)[4,10]. Signals encoding syllable identity are conveyed to the AFP via projections from nucleus HVC in the motor pathway to Area X[4]. In principle, auditory feedback could provide information about variation in fundamental frequency, but such auditory signals appear to be absent in the AFP during singing[23]. Thus we favor the alternative possibility that information about fundamental frequency variation is transmitted to the AFP via an efference copy of activity in premotor regions, by way of projections from HVC to Area X and/or projections from RA to thalamic nucleus DLM[24–25] (Supplementary Figure 1). This is consistent with a recent proposal that transmission of efference copy signals from motor cortex (HVC and/or RA) to basal ganglia circuitry (AFP) plays a fundamental role in mammalian skill learning[26].

Our results also indicate remarkably precise functional coordination between the AFP and the motor pathway. Immediately after unblocking AFP output, we observed learning that was specific to the reinforced features of song, indicating that the AFP had modified its output to direct production of those specific features by the motor pathway. This implies that the AFP not only receives detailed information about the song performances produced by the motor pathway during training, but that it also changes its output to specifically implement the features of those performances that were reinforced. Such a capacity of the AFP to precisely monitor and modify the activity of the motor pathway indicates fine-scale functional coordination both in the projections from the motor pathway to the AFP and in the projections from the AFP back to the motor pathway. Such bi-directional coordination might be mediated by segregated functional loops between the AFP and motor pathway, each encoding a particular feature of song, such as high fundamental frequency in a particular syllable (Supplementary Figure 1). Under normal conditions, with AFP output intact, such functional loops could enable the AFP to amplify and bias specific behavioral features, functions that have been attributed to mammalian basal ganglia circuits[27–28]. More generally, our results suggest that precise functional coordination between motor cortex and basal ganglia circuitry is important for enabling motor skill learning.

Methods Summary

All experiments were performed on adult (> 120 day old) Bengalese finches (*Lonchura striata domestica*) singing undirected song. Song recording and feedback delivery were performed using software[5] that recognized a targeted syllable and delivered a 50–80ms burst of white noise unless the FF met an escape criterion. For experiments with APV in RA and associated controls, the threshold for escaping white noise was set near median FF of the targeted syllable; thus approximately 50% of syllable performances initially avoided white noise. We used reverse microdialysis[14] to deliver the NMDA-receptor antagonist DL-APV (1–5 mM in ACSF) to RA and the GABA(A) agonist muscimol (100–500 μM) or the sodium channel blocker lidocaine (2%) to LMAN. To ensure complete wash-in of drug, we delayed 1–2 hours between drug infusion and the beginning of the training period. Immediately after training, the solution was switched back to ACSF. To ensure complete wash-out of drug, we delayed at least 1 hour between switching the solution to ACSF and measuring FF performance after training.

Online Methods

Animal Care

Adult (> 120 day old) Bengalese finches (*Lonchura striata domestica*) were bred in our colony and housed with their parents until at least 60 days of age. During experiments, birds were housed individually in sound-attenuating chambers (Acoustic Systems) with food and water provided *ad libitum.* All song recordings were from undirected song (i.e. no female was present). All procedures were performed in accordance with established protocols approved by the University of California, San Francisco Institutional Animal Care and Use Committee.

Training

The same training parameters were used for control experiments and experiments with pharmacological manipulations. Song acquisition and feedback delivery were accomplished using previously described LabView software (EvTaf [5]), which recognized a specific time (contingency time) in a targeted syllable of song based on its spectral profile. Upon recognition, EvTaf recorded the time and calculated the fundamental frequency (FF) during the previous 8ms of song. If the FF met the escape criterion (i.e. above or below a threshold), then no disruptive feedback was delivered. Otherwise, a 50–80ms burst of white noise was delivered starting <1ms after the contingency time. The duration of white noise was constant for a given experiment. To allow quantification of FF during training, a randomly interleaved 10% of songs were allocated as catch trials and did not receive white noise.

Experiments with reversible disruption of LMAN transmission to RA via reverse microdialysis

We interfered with LMAN transmission to RA using a previously described reverse microdialysis technique[14], in which solution diffuses into targeted brain areas across the dialysis membranes of implanted probes. RA was mapped electrophysiologically during

cannula implantation in order to direct probes to the center of RA. Between probe insertion and white noise training, there was a >48h period in which control solution (ACSF) was dialyzed at a flow rate of 1 μL/min. The dialysis solution was switched from ACSF to the NMDA-receptor antagonist DL-APV (2–5 mM in ACSF; Ascent) at least 1.5 hours prior to the onset of white noise training so that the threshold for escaping white noise could be determined based on song performance with APV in RA. During this period, we evaluated the efficacy of APV by assessing the rendition-to-rendition variability of FF for individual syllables. FF variability reduced and stabilized at an asymptotic level within the first 30 minutes of APV dialysis, indicating rapid onset and equilibrium of drug effect. We observed a reduction in variability similar to that reported after lesions or inactivations of LMAN[14,16]. For clarity of presentation in Figure 3, running averages of FF performance for experiments with APV in RA omit the period of time during APV wash-in before white noise onset. For experiments with APV in RA and the accompanying control experiments, white noise was delivered for 4–14 waking hours. Blocking AFP output reduced variation in FF by an average of 31.7%, meaning that setting the threshold for avoiding white noise at a certain level above mean FF (e.g. +30Hz) in control experiments and experiments with AFP output blocked would result in a greater proportion of syllable performances escaping aversive reinforcement in control experiments. To avoid this confound and ensure that a similar proportion of syllable renditions received aversive reinforcement in control experiments and experiments with AFP output blocked, we set the threshold for avoiding white noise at approximately the baseline median FF performance (between the 40th and 60th percentile in all experiments). To ensure that our assessment of learning during the training period evaluated the effects of white noise training as opposed to the acute effects of APV, FF change at the end of the training period was quantified by subtracting FF immediately prior to training (during the time period with APV in RA prior to the onset of WN) from FF at the end of the training period. Immediately after the conclusion of white noise training, the dialysis solution was switched back to ACSF. Learning after the training period was quantified by measuring the difference between FF performance after white noise training (with ACSF in RA) and FF performance before white noise training and prior to infusing APV into RA (i.e. with ACSF in RA). Although the latency between switching the solution remotely at the pumping apparatus and changing the solution at the probe tips is only six minutes in our experimental setup[14], the APV-dependent reduction in FF variability typically remained for hours after switching back to ACSF, presumably reflecting the combined kinetics of passive diffusion, active clearance and degradation mechanisms. In all experiments, birds were prevented from singing for at least 1.5 hours after switching from APV to ACSF to provide time for APV washout. For quantification of learning expressed immediately after training (Figure 3f), we analyzed the first songs performed after this period. To further ensure that persisting effects of APV would not cause an underestimation of learning in our primary representations of the data (Figure 3), expression of learning was assessed the morning after the training period. This allowed sufficient time for the APV-dependent block of AFP output to subside while providing limited opportunity for the birds to sing in the absence of white noise, which could lead to extinction. In a subset of experiments (8 of 24) white noise training was terminated (and APV was switched to ACSF) at least three hours before sleep. In these experiments we found that the expression of learning before sleep was significantly greater than zero (0.95+/− 0.25% change in FF,

P<0.02, signed-rank test) and only slightly less than learning the next morning (1.3% +/− 0.18% change in FF). This indicates that washout of APV, independently of a period of sleep, is sufficient to enable the expression of learning. Probe position in RA was established using electrophysiological mapping of RA during implantation and confirmed post mortem by identifying cannula tracts in brain sections stained for Nissl bodies. Additionally, in three birds, biotinylated muscimol (EZ-link biotin kit; Pierce; diluted to 500 µm) was dialyzed across the diffusion membrane in order to estimate the path of diffusion from the membrane[14]. In these birds, probe position was determined post mortem by histological staining for biotin and by comparing interleaved sections stained for Nissl bodies. Spread of drug outside RA tended to be in regions dorsal to RA, along the cannula, but not into the lateral areas where nucleus Ad is located.

Experiments with reversible inactivation of LMAN via reverse microdialysis

We examined the progression of learning for data from experiments in which we transiently inactivated LMAN using the same reverse dialysis technique that we used for infusing APV into RA[14]. To inactivate LMAN, we switched the dialysis solution from ACSF to the GABA$_A$ agonist muscimol (100–500 µM; Sigma; 3 birds, 12 experiments) or the Na$^+$ channel blocker lidocaine (2%; Hospira; 1 bird, 2 experiments) at a flow rate of 1 µl/min. Inactivations lasted for 3–4 h, during which a 1 µl/min flow rate was maintained. At the conclusion of inactivation, the dialyzing solution was switched back to ACSF. We applied white noise contingent on FF over a total period of two or more days, during both control and LMAN inactivation periods. The threshold for escaping white noise was incrementally raised to drive progressive changes in FF. In each experiment, FF eventually reached a stable value because we stopped raising the threshold. We only considered LMAN inactivations on days before FF reached this stable value, to ensure that the bird retained the capacity for further learning. For each LMAN inactivation, learning after training was quantified as the difference in FF between the last 50 renditions of the syllable before infusion of drug and the first 50 renditions of the syllable after drug washout, normalized as for experiments with APV in RA. We excluded the first hour after switching the infusion solution to ACSF to allow for washout. During the period with LMAN inactivated, which lasted a minimum of 3 hours, the threshold for escaping white noise was set so that greater than 50% but less than 90% of syllables escaped and thus a learning signal of differential reinforcement was present in each experiment. This is crucial for interpretation of the lack of learning in these experiments since learning in this paradigm does not proceed without such differential reinforcement[6]. Learning during training with LMAN inactivated was quantified using a linear regression of FF on the renditions of the targeted syllable during training with LMAN inactivated. For each inactivation, matched learning in control conditions was quantified by calculating the average rate of change in FF (per hour) during ACSF infusion on the day of that inactivation and multiplying that rate by the number of hours that LMAN was inactivated. Probe positioning and the path of drug diffusion were evaluated post mortem by histological staining of sectioned tissue as described previously[14]. Tissue damage caused by cannulae enabled confirmation that probes were accurately targeted to LMAN. In addition, biotinylated muscimol or ibotenic acid were used to estimate the spread of diffusion as described previously[14].

Author Manuscript

Analysis

All analyses were performed with custom software written in MATLAB (Mathworks). For a given syllable, FF was measured over a consistent time window aligned to syllable onset; for syllables targeted with WN feedback, the measurement time window was centered at the median point at which feedback was delivered. FF was calculated as described previously[6] for both targeted syllables and non-targeted syllables of the same song. Spectral entropy, volume and duration were calculated as described previously[5]. Statistical significance was tested using non-parametric statistical tests; Wilcoxon signed-rank tests and Wilcoxon rank-sum tests were used where appropriate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Hikosaka O, Nakamura K, Sakai K, Nakahara H. Central mechanisms of motor skill learning. Curr Opin Neurobiol. 2002; 12:217–22. [PubMed: 12015240]

2. Houk, JC.; Adams, JL.; Barto, AG. Models of Information Processing in the Basal Ganglia. Houk, JC.; Davis, JL.; Beiser, DG., editors. MIT Press; Cambridge, Massachusetts: 1995. p. 249-270.

3. Suri RE, Schultz W. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. Neuroscience. 1999; 91:871–890. [PubMed: 10391468]

4. Mooney R. Neural mechanisms for learned birdsong. Learn Mem. 2009; 16:655–669. [PubMed: 19850665]

5. Tumer EC, Brainard MS. Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature. 2007; 450:1240–1244. [PubMed: 18097411]

6. Charlesworth JD, Tumer EC, Warren TL, Brainard MS. Learning the microstructure of successful behavior. Nat Neurosci. 2011; 14:373–380. [PubMed: 21278732]

7. Sutton, RS.; Barto, AG. Reinforcement Learning: An Introduction. MIT Press; Cambridge, Massachusetts: 1998.

8. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275:1593–1599. [PubMed: 9054347]

9. Reynolds JN, Hyland BI, Wickens JR. A cellular mechanism of reward-related learning. Nature. 2001; 413:67–70. [PubMed: 11544526]

10. Fee MS, Goldberg JH. A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. Neuroscience. 2011; 198:152–70. [PubMed: 22015923]

11. Fiete IR, Fee MS, Seung HS. Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. J Neurophys. 2007; 98:2038–2057.

12. Doya, K.; Sejnowski, T. The New Cognitive Neurosciences. Gazzaniga, M., editor. MIT Press; Cambridge, Massachusetts: 2000. p. 469-482.

13. Andalman AS, Fee MS. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. Proc Natl Acad Sci USA. 2009; 106:12518–12523. [PubMed: 19597157]

14. Warren TL, Tumer EC, Charlesworth JD, Brainard MS. Mechanisms and time course of vocal learning and consolidation in the adult songbird. J Neurophysiol. 2011; 106:1806–1821. [PubMed: 21734110]

15. Olveczky BP, Andalman AS, Fee MS. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol. 2005; 3:e153. [PubMed: 15826219]

16. Hampton CM, Sakata JT, Brainard MS. An avian basal ganglia-forebrain circuit contributes differentially to syllable versus sequence variability of adult Bengalese finch song. J Neurophysiol. 2009; 101:3235–3245. [PubMed: 19357331]

17. Krupa DJ, Thompson JK, Thompson RF. Localization of a memory trace in the mammalian brain. Science. 1993; 260:989–991. [PubMed: 8493536]

18. Atallah HE, Lopez-Paniagua D, Rudy JW, O'Reilly RC. Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. Nat Neurosci. 2007; 10:126–131. [PubMed: 17187065]

19. Balleine BW, Ostlund SB. Still at the choice-point: action selection and initiation in instrumental conditioning. Ann N Y Acad Sci. 2007; 1104:147–171. [PubMed: 17360797]

20. Crapse TB, Sommer MA. Corollary discharge across the animal kingdom. Nat Rev Neurosci. 2008; 9:587–600. [PubMed: 18641666]

21. Olveczky BP, Otchy TM, Goldberg JH, Aronov D, Fee MS. Changes in the neural control of a complex motor sequence during learning. J Neurophysiol. 2011; 106:386–397. [PubMed: 21543758]

22. Sober SJ, Wohlgemuth MJ, Brainard MS. Central contributions to acoustic variation in birdsong. J Neurosci. 2008; 28:10370–10379. [PubMed: 18842896]

23. Leonardo A. Experimental test of the birdsong error-correction model. Proc Natl Acad Sci USA. 2004; 101:16935–16940. [PubMed: 15557558]

24. Vates GE, Vicario DS, Nottebohm F. Reafferent thalamo- "cortical" loops in the song system of oscine songbirds. J Comp Neurol. 1997; 380:275–290. [PubMed: 9100137]

25. Goldberg JH, Fee MS. A cortical motor nucleus drives the basal ganglia-recipient thalamus in singing birds. Nat Neurosci. (Epub February 12, 2012).

26. Redgrave P, Gurney K. The short-latency dopamine signal: a role in discovering novel actions? Nat Rev Neurosci. 2006; 7:967–975. [PubMed: 17115078]

27. Turner RS, Desmurget M. Basal ganglia contributions to motor control: a vigorous tutor. Curr Opin Neurobiol. 2010; 20:704–716. [PubMed: 20850966]

28. Frank MJ. Computational models of motivated action selection in corticostriatal circuits. Curr Opin Neurobiol. 2011; 21:381–386. [PubMed: 21498067]
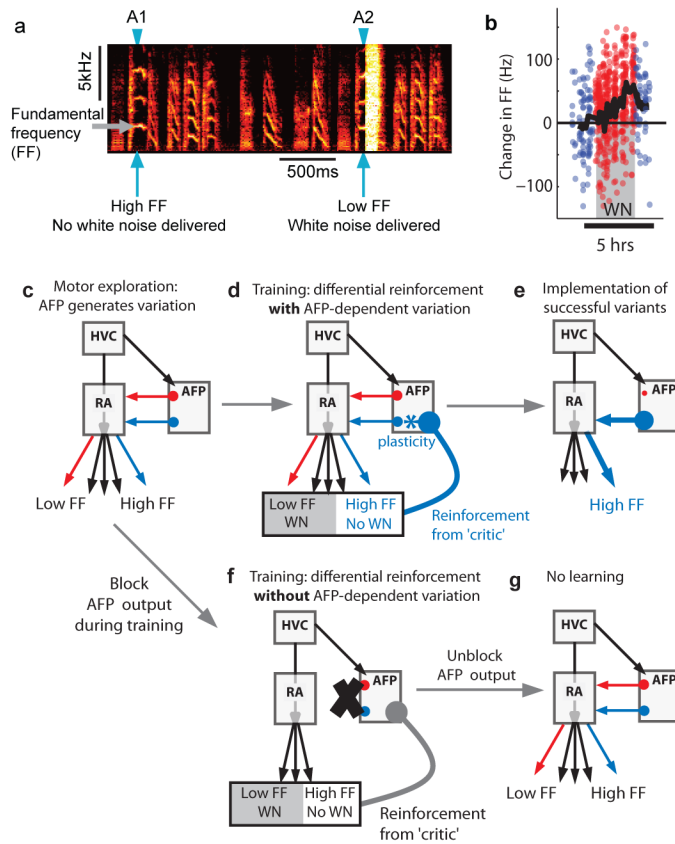
Figure 1. Trial-and-error learning in adult birdsong

a. Spectrogram of song during an experiment in which white noise (WN) was delivered to targeted syllable (A) renditions with low fundamental frequency (FF) but not high FF. **b.** Delivering WN to syllables with low FF (shaded region) elicited increases in FF. Each point corresponds to one syllable rendition; black line indicates running average. **c.** The song circuit includes a motor pathway, containing HVC and RA, and the anterior forebrain pathway (AFP), important for learning. The AFP generates variation in performance (motor exploration); red and light blue indicate distinct activity patterns in the AFP that lead to distinct FF values on different renditions of the same syllable. **d.** Actor-critic models propose that the AFP receives feedback about the behavioral variants that it generates, and this feedback strengthens patterns of AFP activity yielding better outcomes (light blue, feedback shown) and weakens patterns of AFP activity yielding worse outcomes (red). **e.** This changes the output of the AFP so that it selectively implements more successful behaviors. **f.** We tested this model by blocking the output of the AFP during training, thus preventing the AFP from generating variation in FF. **g.** The model predicts that this will prevent learning-related plasticity in the AFP, and thus there will be no change in FF, even when AFP output is unblocked after training.
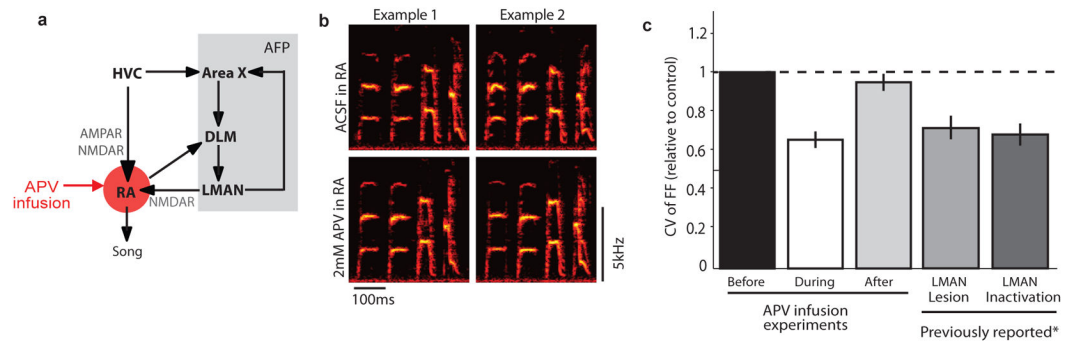
Figure 2. Infusing APV into RA reversibly reduced song variability without distorting song structure

a. The AFP contains the striatopallidal nucleus Area X, the thalamic nucleus DLM, and the cortical nucleus LMAN, which projects to RA. We blocked AFP output to the motor pathway by infusing the NMDA receptor antagonist APV into RA. **b.** Infusing APV into RA did not markedly change song. **c.** Infusions of APV into RA reduced the coefficient of variation (CV) of FF, which recovered after switching back to ACSF (n=12 syllables in 9 birds). The CV reduction with APV in RA (31.7% +/− 5.6%) was not significantly different from previously reported effects of lesions (34.1 +/− 4.5%) and inactivations (28.4 +/− 6.0%) of LMAN in adult Bengalese finches. Error bars indicate +/− s.e.m. *Previously reported values from Hampton et al.[16] and Warren et al.[14].
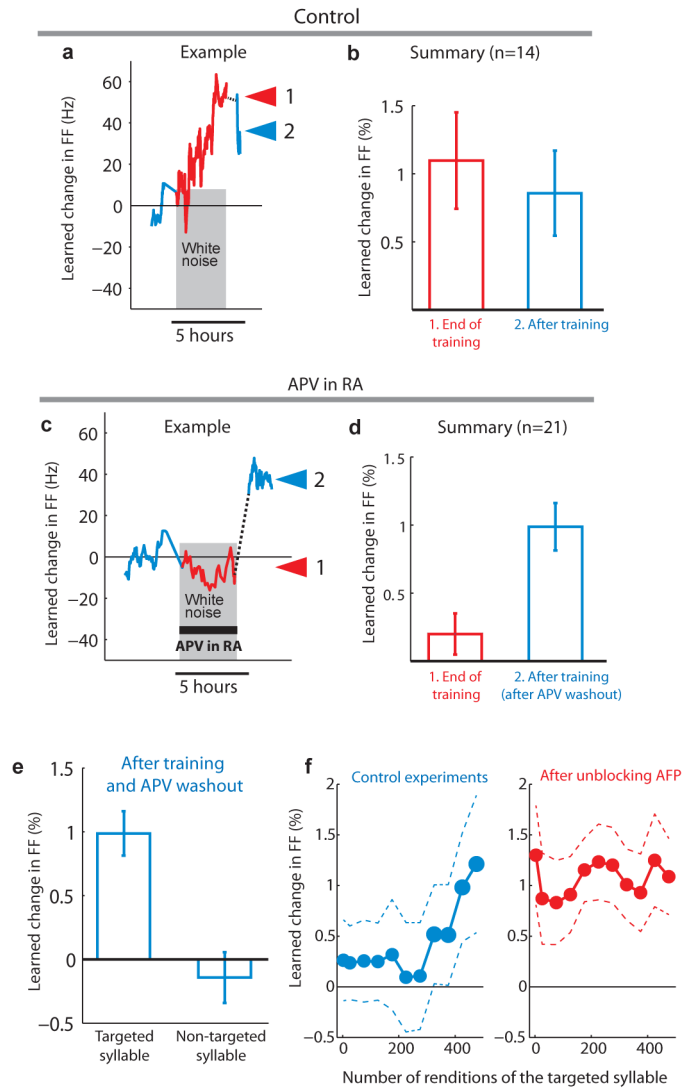
Figure 3. Infusing APV into RA prevents expression but not acquisition of learning
a. Control experiment (ACSF in RA) in which white noise was delivered to targeted syllables with low FF. Arrowheads indicate FF at end of training (1) and after training (2). Dashed line indicates delay between measurements at the end of training and after training. **b.** For control experiments (n=14 experiments in 7 birds), learning was expressed at a similar magnitude at the end of training (left) and after training (right). Learning was normalized as a percentage of baseline FF. Error bars indicate +/− s.e.m. **c.** Example of experiment with AFP output blocked (APV infused into RA) throughout the training period. Arrowheads indicate FF at end of training (1) and after training and APV washout (2). **d.** For experiments with APV in RA (n=21 experiments in 9 birds), learning at end of training (left) was not significantly greater than zero and was significantly less than in control experiments. Learning after training and APV washout (right) was significantly greater than zero and was the same magnitude as in control experiments. **e.** After training and APV washout, learning was evident in syllables targeted with reinforcement (left) but not in other syllables of the same songs that were not targeted with reinforcement (right). This analysis

was performed for each experiment in which FF of a non-targeted syllable could be reliably quantified (n=17 of 21 total experiments). **f.** Mean progression of learning for control experiments (left) and after unblocking AFP output for experiments with APV in RA (right). Points correspond to syllable renditions 1–5, 1–50, 51–100,…451–500. Dashed lines indicate +/− s.e.m.

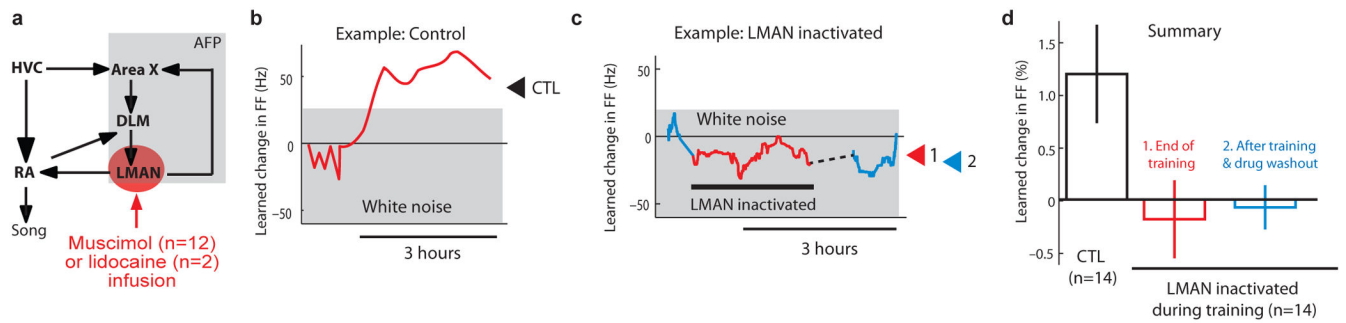**Figure 4. Inactivating LMAN during training prevents both expression and acquisition of learning**

a. We inactivated LMAN by infusing the $GABA_A$ antagonist muscimol (n=12 experiments in 3 birds) or the sodium channel blocker lidocaine (n=2 experiments in 1 bird) into LMAN. **b.** Control experiment in which white noise was delivered to renditions of a targeted syllable with low FF. **c.** Same as panel b, but with LMAN inactivated during training. Arrowheads indicate FF at the end of training with LMAN inactivated (1) and following training and muscimol washout (2). **d.** For experiments with LMAN inactivated (n=14), there was neither evidence for learning at the end of training (red) nor after training and drug washout (light blue). Error bars indicate +/− s.e.m.