

Reviewed Preprint

v1 • September 1, 2026

Not revised

✉ For correspondence:

nicolas.rougier@inria.fr

arthur.leblois@u-bordeaux.fr

† These authors contributed equally to this work

Competing interests: No competing interests declared

Funding: See page 26

Reviewing editor: Long Ding, University of Pennsylvania, United States

© 2026, Sankar et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Dual pathway architecture in songbirds enables robust sensorimotor learning

Remya Sankar¹, Atharv Suryawanshi², Nicolas P Rougier^{1,†} ✉, Arthur Leblois^{3,†} ✉

¹Univ. Bordeaux, Inria, IMN, UMR 5293, Bordeaux, France • ²LMU Munich, Planegg, Germany • ³Univ. Bordeaux, CNRS, IMN, UMR 5293, Bordeaux, France

eLife Assessment

This study modeled vocal learning in zebra finches with a network of three components: a pathway with delayed/slow Hebbian learning that mimics the HVC-RA projection, a pathway with reinforcement learning that mimics the HVC-BG-RA projection, and a motor unit that mimics the syrinx and produces song output. The model **convincingly** reproduces key features of song learning and is a **valuable** step towards understanding how vocal learning is substantiated in the song system. The generic and simplified nature of the model limits its generalization to other forms of motor learning. Further examination of model assumptions and presentation of testable predictions would increase the impact of the study.

<https://doi.org/10.7554/eLife.112141.1.sa4>

Abstract

The acquisition of sensorimotor skills critically depends on basal ganglia (BG)–thalamo–cortical circuits. Prevailing theories propose that the BG optimize motor output through reinforcement learning (RL), using internal performance evaluations to approximate stochastic gradient ascent. However, this framework struggles in non-convex performance landscapes, where local optima hinder efficient learning. Songbirds provide a compelling biological example of robust sensorimotor learning, mastering complex vocalizations through trial-and-error within a specialized BG–thalamo–cortical architecture. Here, we present a computational model constrained by the anatomy, physiology, and developmental trajectory of the zebra finch song system. The model combines a BG-driven RL pathway with a parallel cortical motor pathway that progressively consolidates successful motor patterns via Hebbian plasticity. In addition, we incorporate synaptic volatility within the BG pathway, introducing structured variability across learning. Through simulations of vocal learning using both a biophysical syrinx model and synthetic performance landscapes, we demonstrate that this dual-pathway architecture reliably converges to global optima and outperforms standard and noise-annealed RL approaches. The model reproduces key experimental features of song learning, including non-monotonic learning trajectories, a gradual reduction in motor variability, and the developmental transfer of motor control from subcortical to cortical circuits. Mechanistically, delayed maturation of the cortical pathway provides an implicit regulation of the exploration–exploitation trade-off, while synaptic volatility enables escape from local optima. These results highlight the importance of neural circuit architecture and dynamics in efficient learning, and suggest biologically inspired design principles for improving the robustness and sample efficiency of artificial RL systems in complex sensorimotor domains.

Introduction

Acquisition and maintenance of sensorimotor skills are based on optimizing motor output based on internal evaluations of the behaviors performed and their results (Hikosaka et al., 2002). In the brain, this process is heavily dependent on basal ganglia (BG)–thalamo–cortical loops (Hikosaka et al., 2002; Graybiel and Grafton, 2015). A prominent view frames BG–cortical interactions as a form of reinforcement learning (RL), in which dopamine influx in the BG signals performance (Schultz et al., 1997; Kasdin et al., 2025) and dopamine-dependent cortico-striatal plasticity strengthens motor programs that previously yielded better-than-expected outcomes (Wickens et al., 2003; Doya, 2000). Over many trials, this mechanism approximates stochastic gradient ascent on performance (Williams, 1992). However, when applied to continuous embodied sensorimotor tasks, the gradient-based RL framework faces major limitations. High-dimensional state–action spaces (Bellman, 1957), nonlinear body dynamics (Bernstein, 1967), and complex performance landscapes make gradient ascent slow, sample-inefficient, and vulnerable to local optima (Peters and Schaal, 2008; Dhawale et al., 2017). Artificial RL agents typically rely on heuristic exploration strategies (e.g. *ε*-greedy or uncertainty bonuses), (Sutton and Barto, 2018; George and Powell, 2006) that lack biological grounding and do not resemble the richer exploratory behaviors observed in humans or animals (Gershman and Daw, 2017; Shen and Dayan, 2025).

Juvenile songbirds provide a model system that exposes these shortcomings. Learning to imitate a tutor’s song requires exploring a vast sensorimotor space to approach an optimal performance (Mooney, 2009; Fee and Goldberg, 2011). Yet, the trajectory of learning is highly irregular and non-monotonic. Performance often deteriorates after sleep before improving with practice, and the magnitude of this post-sleep deterioration predicts later imitation quality (Derégnaucourt et al., 2005). Moreover, changes within the day, overnight fluctuations, and slower consolidation on a week scale display a partial misalignment (Kollmorgen et al., 2020), making song learning an erratic process that defies simple gradient-ascent interpretations. These saltatory transitions suggest that the brain employs mechanisms beyond straightforward local optimization, allowing discontinuous traversals across rugged performance landscapes.

Songbirds rely on a dual pathway brain circuitry dedicated to vocal learning for song acquisition and maintenance. A cortical-like motor pathway and a BG-thalamo-cortical circuit interact during learning, with structured variability, pathway-specific plasticity, and delayed maturation providing a substrate for guided exploration and consolidation (Mooney, 2009). The motor pathway comprising HVC (used as a proper name) and the robust nucleus of the arcopallium (RA) executes vocal output, generating the precise timing, sequencing, and motor patterns driving the song (Hahnloser et al., 2002; Sober and Brainard, 2009). During development, RA firing evolves from variable juvenile patterns to highly stereotyped adult burst sequences aligned to the song with 10–20 ms precision (Ölveczky and Gardner, 2011). In parallel, a BG–thalamo–cortical loop (Area X → DLM → LMAN → Area X) receives dense dopaminergic input, evaluates performance (Kasdin et al., 2025), and injects variability and corrective bias into RA through LMAN (Warren et al., 2011; Andalman and Fee, 2009; Kao et al., 2005). This BG-derived bias is gradually consolidated into the increasingly influential motor pathway through activity-dependent plasticity (Aronov et al., 2008; Mehaffey and Doupe, 2015). Toward the beginning of learning, the BG pathway dominates exploration, while direct cortical motor projections strengthen over time, eventually making BG input unnecessary to produce the crystallized song (Bottjer et al., 1984; Scharff and Nottebohm, 1991), though its removal reduces residual variability (Dhawale et al., 2017; Kao et al., 2005).

To explore how biological systems solve complex sensorimotor problems, we develop a computational model that captures vocal learning in a way consistent with and constrained by the behavioral, physiological, and anatomical characteristics of song learning in zebra finches (Figure 1a,b). In this model, we propose a neural mechanism inspired by the song system that addresses key limitations of standard RL. Through simulations of the neuronal network driving vocal learning across diverse performance landscapes, we show that the interaction between circuit

architecture and synaptic plasticity enables the system to escape local optima and more consistently reach global solutions compared to standard RL algorithms. We further propose that synaptic volatility and delayed maturation of the cortical pathway could produce the observed non-monotonic learning features such as post-sleep deterioration while ensuring optimal final performance. Altogether, this work highlights architectural principles that could enhance the robustness and sample efficiency of RL for real-world sensorimotor control.

Results

The dual pathway model

We investigate the neural mechanisms of song learning using a computational model of the song-related network constrained by the anatomy and physiology of the song system described in the Introduction (Figure 1a-c). The model includes three populations, each representing the contribution of a song system nucleus or pathway to song acquisition: one population representing nucleus HVC, one representing nucleus RA, and a BG population representing the contribution of the basal ganglia-thalamo-cortical circuit. HVC and RA are connected via two pathways: a direct cortical (HVC-RA) pathway and a polysynaptic pathway through BG.

Song learning is simulated over a 60-day sensorimotor period mimicking the sensorimotor period of song learning, with 1,000 motif renditions per day (Figure 1a, Tchernichovski et al. (2000)). HVC provides a fixed, time-locked representation of the song motif sequence (Figure 1d, Hahnloser et al. (2002)), assumed for simplicity to be already in place by the beginning of the sensorimotor learning (Mackevicius et al., 2023). Initially, HVC-RA synapses are absent and gradually strengthen during learning (Mooney and Rao, 1994), while HVC-BG connectivity is random. RA activity drives a 2D motor output controlling vocal production, which is evaluated at each time step (see Methods) to generate a dopaminergic performance prediction error signal (Gadagkar et al., 2016; Kasdin et al., 2025). Performance is assessed either through a biophysical syrinx model, comparing produced vocalizations to tutor syllables (Eq 17, see Methods, Figure S1a, Figure 1e, Titze and Martin (1998); Amador and Margoliash (2013)), or via synthetic landscapes with controlled numbers of local optima (Figure S1c, d, Figure 1f). In both cases, a performance prediction error signal calculated from current and past performance as Eq 10 is fed into a tri-partite reinforcement learning plasticity rule at HVC-BG synapses following Eq 9. This plasticity rule is known to drive gradient ascent in the performance landscape (Williams, 1992). It will allow BG output, and thereby RA output, to gradually approach the closest local maximum in the performance landscape. This gradient ascent strategy is clearly insufficient and bound to fail in landscapes with multiple sub-optimal maxima such as the ones derived from the syrinx model (Figure 1e) or designed artificially to test our model (Figure 1f).

To overcome the limitations of gradient-based learning in multi-peaked landscapes, the model incorporates two key biological features. First, Hebbian plasticity in the HVC-RA pathway progressively consolidates BG-driven motor patterns (Eq 12, Mehaffey and Doupe (2015); Teşileanu et al. (2017)). Second, synaptic volatility in HVC-BG connections introduces overnight perturbations (Eq 13), mimicking the continuous remodeling of the cortico-striatal synapses with volatility over hours to days (Kuo and Liu, 2019; Cepeda et al., 2003; Holtmaat and Svoboda, 2009). This volatility induces post-sleep deterioration (Derégnaucourt et al., 2005), promoting non-monotonic learning and exploration. Together, these mechanisms enable learning in non-convex performance landscapes. In the following sections, we compare the model's behavior to songbird data, analyze the interaction between pathways, and benchmark its performance against alternative learning strategies.

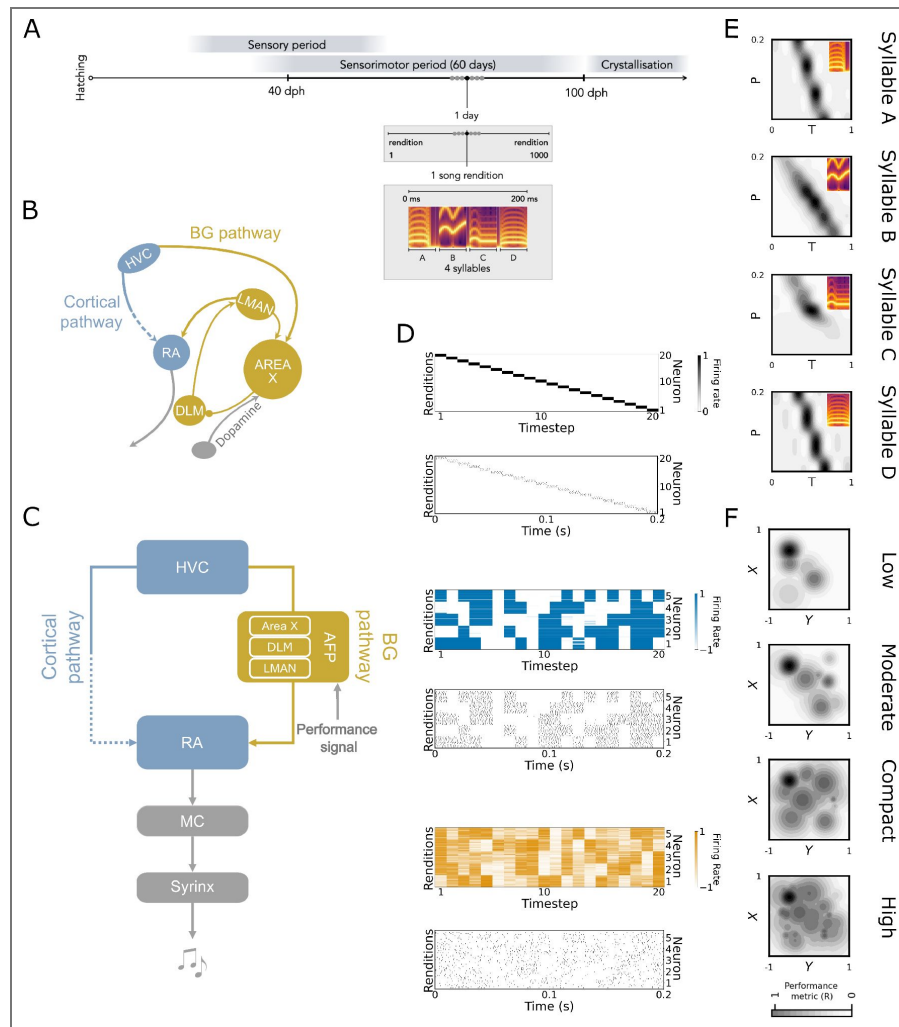


Figure 1. The dual pathway model for vocal learning.

A. Timeline of song learning. The model learns to produce 4 syllables (50 ms each) per motif, and practices the motif 1000 times per day for 60 days (dph: days post hatch), similar to the sensorimotor phase of zebra finches. **B.** The neural substrates involved in vocal learning in zebra finches, the song system. The cortical pathway driving adult song production, in blue, with dotted lines representing its delayed maturation and gradual strengthening over the sensorimotor phase. The subcortical pathway, in yellow, necessary for song learning and maintenance, receiving information about performance quality through dopaminergic input. **C.** The architecture of the dual pathway model is derived from the song system. The cortical pathway, in blue, consists of two nuclei, HVC (used as a proper name) and the robust nucleus of the arcopallium (RA). The pathway strengthens during learning (dotted). The subcortical pathway, in yellow, consists of one basal ganglia (BG) neuronal population, representing the BG-thalamo-cortical loop (comprising Area X, medial part of the dorsolateral nucleus (DLM) and lateral magnocellular nucleus of the anterior neostriatum (LMAN)). The motor control (MC) layer transforms the RA neural commands to the appropriate vocalisation parameters (see Methods). **D.** Representative neural activity in each neuronal population of the model at the end of the sensorimotor phase. For each population, the top panel shows the firing rate generated for each 10 ms time bin during song, and the bottom panel shows illustrative spiking activity (see Methods), where each vertical black line denotes a spike. HVC neurons (black) fire sequentially, providing a timing scaffold for the song. By the end of the sensorimotor period, RA neuronal populations (blue) display sparse, stereotyped bursts across renditions. BG neuronal populations (yellow) exhibit high rendition-to-rendition variability in firing rate. **E.** Examples of performance landscapes generated using a biophysical model of the syrinx (see Methods) with four target syllables (inset). The performance metric at each point (determined by pressure, P, on y-axis, tension, T, on x-axis) on the landscape is indicated by the grey heatmap. The landscapes have a few global optima (performance metric ≈ 1) and several local optima (performance metric $< .7$). **F.** Examples of Gaussian-based landscapes from each density category. The landscape is generated by superimposing several Gaussian functions (see Methods). The superimposition results in several local (performance metric $< .7$) and one globally optimal solution (performance metric = 1). The landscapes are divided into four classes according to the number of local optima present: low (2-5), moderate (3-9), compact (8-18), high (11-25).

Song imitation with an artificial syrinx or in arbitrary performance landscapes

In Figure 2 [↗](#), we evaluate the dual-pathway model on a motor learning task involving vocal production through a biophysical syrinx model (see Figure 1 [↗](#)). The network learns to generate RA activity patterns that reproduce four tutor syllables (each 50 ms duration) forming a song motif (Figure 2a [↗](#)) by maximizing the similarity between produced and target vocalizations. Initially, outputs lie far from optimal regions of the performance landscape, but through exploration the model converges to accurate imitations (Figure 1b-e [↗](#)). This improvement is accompanied by systematic exploration of motor parameters controlling syrinx dynamics. Across 100 simulations, individual syllables were successfully imitated (terminal performance > 0.7) in over 85% of cases, with high median performance (median terminal performance [All summary statistics are reported as median [interquartile range (IQR)], unless otherwise stated.] of A: 0.93 [0.92-0.97], B: 0.98 [0.95-0.99], C: 0.98 [0.98-0.99], D: 0.98 [0.97-0.99]) (Figure 2f [↗](#)), and full motifs were successfully learned in more than 95% of simulations. Failures (~5%) corresponded to convergence to suboptimal local maxima.

We then assessed performance on synthetic landscapes with a single global optimum and varying densities of local optima (Figure 1f [↗](#)). As distractor density increased, success rates declined modestly and plateaued, reflecting increased trapping in local optima (Figure 2g [↗](#)). At high densities, performance improved again, as closer spacing between optima facilitated transitions toward the global solution. Overall, these results demonstrate that the dual-pathway model achieves robust learning across a wide range of landscape complexities.

Subcortical circuitry is required for vocal learning by the dual pathway model

To disentangle the roles of the cortical (HVC–RA) and subcortical (HVC–BG–RA) pathways across learning, we performed *in silico* lesions mimicking experimental manipulations in songbirds. Consistent with prior studies, we selectively inactivated BG inputs to RA at different developmental stages to assess their contribution to learning and production. Early BG inactivation (before or at 40 dph) resulted in low imitation quality and limited improvement over time (Figure 3a [↗](#), left panel), with reduced motor variability, indicating that the cortical pathway alone cannot support learning at this stage. Mid-stage inactivation (70 dph) led to moderate performance impairments (success rate drops from 91% to 80%) and a sharp drop in variability (from 0.11 [0.09-0.14] to 0.03 [0.02-0.04]), with learning plateauing thereafter (Figure 3a,b,c [↗](#), middle panels). In contrast, late inactivation (100 dph) had minimal impact on vocal output and performance (0.98 [0.97-0.99] vs 0.98 [0.96-0.99], Figure 3a-c [↗](#)), with only a slight reduction in variability (from 0.04 [0.03-0.06] to 0.01 [0.002-0.01]), showing that BG input becomes largely dispensable once learning has converged in the model, as shown experimentally (Bottjer et al., 1984 [↗](#); Scharff and Nottebohm, 1991 [↗](#); Aronov et al., 2008 [↗](#), 2011 [↗](#)).

Conversely, inactivating the cortical pathway (HVC–RA) at any stage prevented stable performance despite ongoing exploration (Figure 3d-e [↗](#)). At mid and late stages, this caused a dramatic drop in performance (dph 70: from 0.92 [0.89 - 0.94] to 0.28 [0.14 - 0.39] and dph 100: from 0.98 [0.96 - 0.99] to 0.24 [0.15 - 0.31], success rate falling to 0%), and a strong increase in variability (dph 70: from 0.11 [0.09 - 0.15] to 0.44 [0.36 - 0.49] and dph 100: from 0.04 [0.03 - 0.06] to 0.53 [0.46 - 0.57]), reflecting unopposed BG-driven fluctuations. The BG pathway is thus essential for early exploration and learning but becomes unnecessary for stable production after crystallization, while the cortical pathway progressively assumes control and suppresses variability. BG lesions also abolish overnight performance deterioration in the model, as it is driven by BG-mediated synaptic volatility. Together, these results support a dual-pathway mechanism in which BG-driven exploration enables learning, while cortical circuits consolidate stable motor patterns.

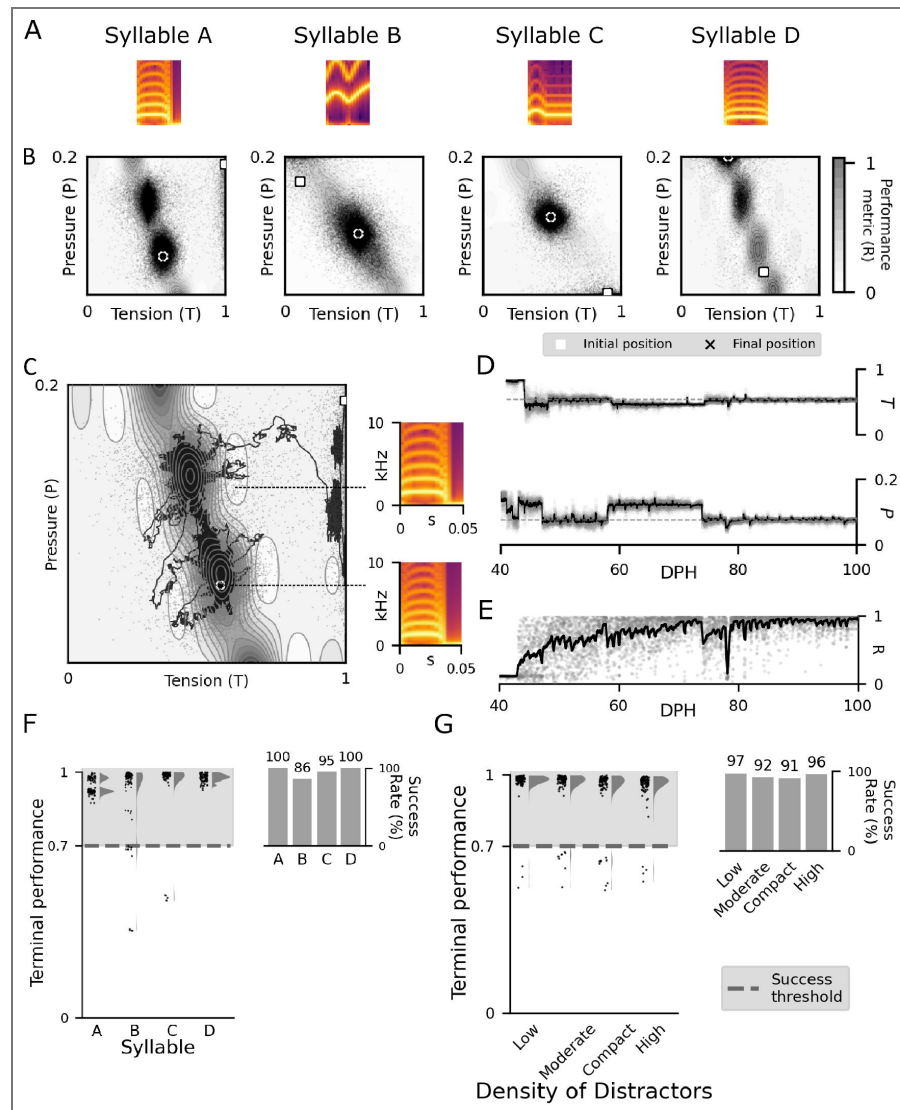


Figure 2. Song acquisition in the dual pathway model.

A. Target song composed of 4 syllables. **B.** Model simulation of the learning process for each syllable over the sensorimotor period. The model starts at a random initial position (square) and is able to find the optimal solution (cross) over the sensorimotor period of 60 days with 1000 iterations each day. The output at each iteration is shown as a black dot. Each panel shows the motor outputs produced for the corresponding syllable. **C.** Example trajectory of the model output. The two-dimensional MC output is fed as pressure and tension in the syrinx model, driving the production of each syllable. The model starts at a random initial position (square), produces several sub-optimal vocalisations, over the course of learning (smoothed trajectory in black) and eventually finds the globally optimal solution (cross). Two example output vocalisations are shown on the right, corresponding to a locally (top) and globally (bottom) optimal performance, with differing fundamental frequencies. **D.** Smoothed trajectories for the control parameters representing the syrinx pressure and tension for the example syllable simulation in C are shown in black. The target output for each control parameter is shown in grey horizontal dotted lines. The model explores different outputs (grey dots) in the two dimensions in the beginning of sensorimotor learning and eventually converges at the target output. **E.** Smoothed trajectory for the performance metric obtained by the model at every rendition of the example syllable over the sensorimotor period (in black). Grey dots represent performance metric at each rendition. **F.** Performance of the model over 100 simulations on the syrinx-based landscapes. Left: Terminal performance for all simulations (black dots), calculated as the average performance over the last 100 trials on the last day (see Methods). Performance metrics above the threshold (shaded portion above the black dotted line), lie close to the global optimum. The violin plot denotes the distribution of the terminal performance (black dots) across the 100 simulations. Right: Success rate for each syllable, i.e. the proportion of simulations that learned the optimal performance (achieved terminal performance > .7). **G.** Performance of the model over 100 simulations on the gaussian-based landscapes, with increasing density of local optima. Same conventions as in F.

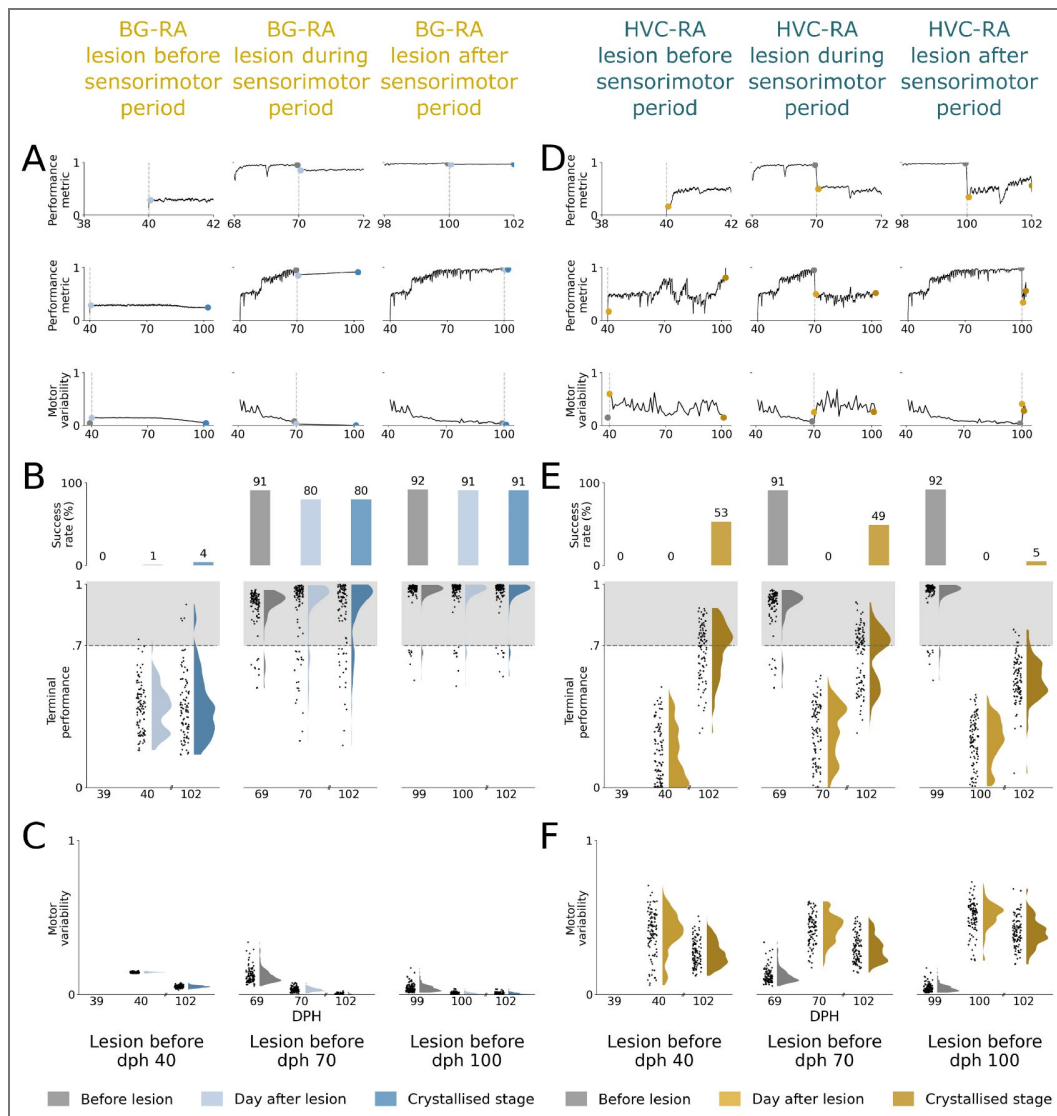


Figure 3. Effects of inactivating the BG and cortical inputs to RA, respectively, at different stages of the sensorimotor learning period.

A. Performance before and after inactivation of BG input to RA at different stages of sensorimotor learning in representative simulations. The three columns show the effect of BG inputs being inactivated before (left), during (middle) and after (right) sensorimotor learning. Top panel: Performance (black) two days before and after lesion. Vertical dotted line denotes the lesion time. Middle panel: Performance (black) throughout the sensorimotor period. Bottom panel: Motor variability over the sensorimotor period. In all three panels, the grey dots denote the performance metric value at the end of the day before inactivation. The light blue dots denote the metric at the start of day after inactivation. The blue/yellow dots denote the metric at the end of the sensorimotor period. **B.** Distribution of success rate (top) and terminal performance (bottom) across 100 simulations before and after inactivation of BG input to RA performed at different stages of sensorimotor learning. Bottom: Terminal performance for all simulations (black dots, see Methods). Performance metrics above the threshold (shaded portion above the black dotted line), lie close to the global optimum. The violin plot denotes the distribution of the terminal performance (black dots) across the 100 simulations. Top: Success rate for each condition (proportion of simulation achieving terminal performance > 0.7). **C.** Motor variability across 100 simulations before and after inactivation of BG input to RA performed at different stages of sensorimotor learning. Black dots show the motor variability for each simulation (see Methods) while violin plot denotes the distribution of motor variability across the 100 simulations. **D-F.** Corresponds to A-C for the suppression of HVC input into RA (in yellow).

Neural activity patterns in the model parallels song system neurophysiology

The model reproduces key neurophysiological features of the song system. HVC exhibits sparse, time-locked burst firing, with each time point corresponding to a distinct target output and performance landscape (Figure 4a). HVC neurons project to both RA and the basal ganglia (BG), forming parallel pathways. Throughout learning, BG activity remains variable across renditions (dph 40: 0.19 ± 0.09 ; dph 99: 0.17 ± 0.05 ; mean $\pm$ std, Figure 4c), while RA dynamics evolve markedly (Figure 4b). Early in learning, weak HVC–RA synapses make RA activity largely driven by the BG pathway, resulting in high variability (dph 40: 0.36 ± 0.29 ; mean $\pm$ std). As HVC–RA synapses strengthen, RA becomes increasingly driven by stable HVC input, reducing sensitivity to BG fluctuations and leading to lower variability (dph 100: 0.05 ± 0.13 ; mean $\pm$ std). This transition from variable to stereotyped RA activity mirrors experimental observations in juvenile zebra finches (Ölveczky et al., 2011) and reflects the progressive stabilization of motor output. Overall, the model captures both the structured temporal drive from HVC and the gradual refinement of activity underlying vocal learning, with behavior shifting from exploratory to stereotyped output.

Overnight performance deterioration aids sensorimotor learning

Juvenile zebra finches exhibit non-monotonic learning, with performance often deteriorating after sleep (Derégnaucourt et al., 2005). In our model, we introduce random synaptic perturbations at HVC–BG synapses during the night. The magnitude of this volatility is stochastic and modulated inversely by the extent of synaptic change accumulated during the preceding day (see section Methods). It produces daily shifts in motor output. Without this volatility, the model follows local gradient ascent, converging monotonically to nearby suboptimal solutions and becoming trapped (Figure 5a–c). In contrast, with overnight volatility, synaptic changes are minimal during the day but substantially larger across nights (0.11 [0.05 – 0.21] vs 0.27 [0.09 – 0.51], Mann-Whitney $U = 4.38e7$, $p < 0.001$, $r = 0.38$), leading to systematic displacements in BG and RA activity. This induces broader, cross-day exploration of the performance landscape (Figure 5d–f), enabling the model to escape local optima and discover the global solution.

As a result, learning becomes non-monotonic across days: performance drops after each night but improves within each day, closely matching experimental observations (Figure 5e). This effect is strongest early in learning and diminishes over time. Across 100 simulations, success rates increased markedly with synaptic volatility (93% vs 60%, Mann-Whitney $U = 1.07e9$, $p < 0.001$, $r = 0.408$), along with higher terminal performance (0.98 [0.96 – 0.99] vs 0.90 [0.63 – 0.97], Figure 5h). These findings show that overnight synaptic volatility provides an effective exploration mechanism, facilitating escape from local optima and enabling robust learning in complex, non-convex landscapes.

Delayed maturation of cortical pathway modulates exploration-exploitation transition

The dual-pathway model explores multiple regions of the sensorimotor landscape before converging to a stable solution (Figure 6b–d). Early in learning, the cortical pathway is weak and contributes little to motor output, which is instead dominated by the basal ganglia (BG) pathway. As a result, vocal output is highly variable and reflects BG-driven exploration. Over time, Hebbian plasticity strengthens the cortical pathway, progressively consolidating BG-led vocalizations. Cortical activity increasingly drives motor output, while the influence of the BG pathway diminishes. Consequently, RA activity and vocal output become more stable and less variable across development (Figure 6f).

Importantly, this reduction in motor variability (dph 40: 0.41 ± 0.12 , dph 100: 0.04 ± 0.02 ; mean $\pm$ std) does not stem from changes within the BG itself, which remains consistently variable (dph 40: 0.17 ± 0.01 , dph 100: 0.17 ± 0.01 ; mean $\pm$ std). Rather, the impact of BG-driven fluctuations is attenuated as the cortical pathway strengthens (Figure 6a,e), a process also reflected in the decreasing magnitude of overnight performance shifts (early: 0.4 ± 0.29 , late: 0.05 ± 0.03 ; mean $\pm$ std;

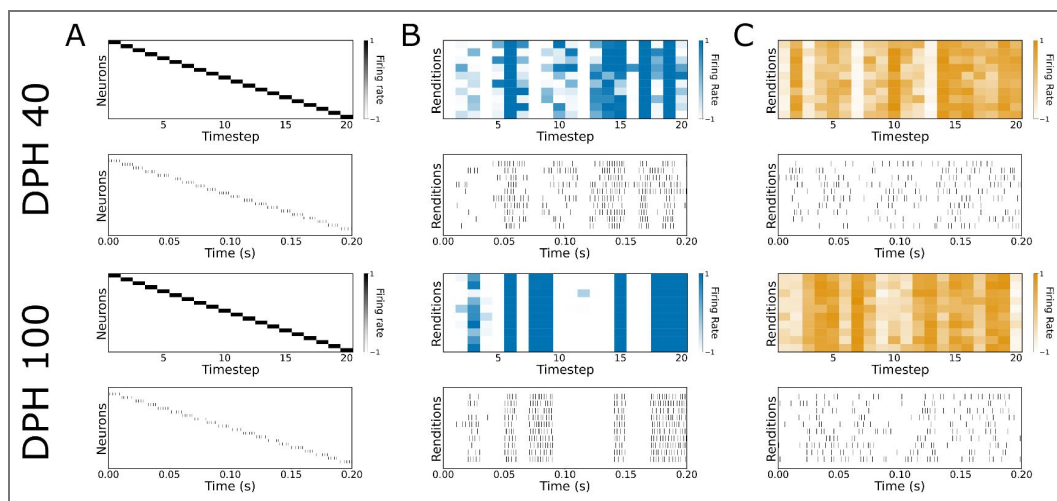


Figure 4. Neural activity patterns across the sensorimotor learning period (40 dph vs 100 dph) in the model.

A. Neurons in the HVC layer encode time during song and provide context information to the model. The activity pattern of HVC neurons (black) across one song rendition. Top panel: Firing rate of multiple neurons across time steps during one song rendition. Bottom panel: Poisson spike trains generated based on the above firing rate profiles (see Methods). Each vertical black line corresponds to one action potential. **B.** Neurons in the RA layer have variable firing rates across song renditions in the beginning of sensorimotor learning and stereotyped patterns towards the end of sensorimotor learning. Top panel: Firing rate of one RA neuron (shades of blue) across ten song rendition. Bottom panel: Poisson spike trains generated based on the above firing rate profiles. **C.** The neurons in the BG layer have variable firing rates across song renditions throughout sensorimotor learning, akin to LMAN neural activity. Top panel: Firing rate of one BG neuron (shades of blue) across 10 song renditions. Bottom panel: Poisson spike trains generated based on the above firing rate profiles.

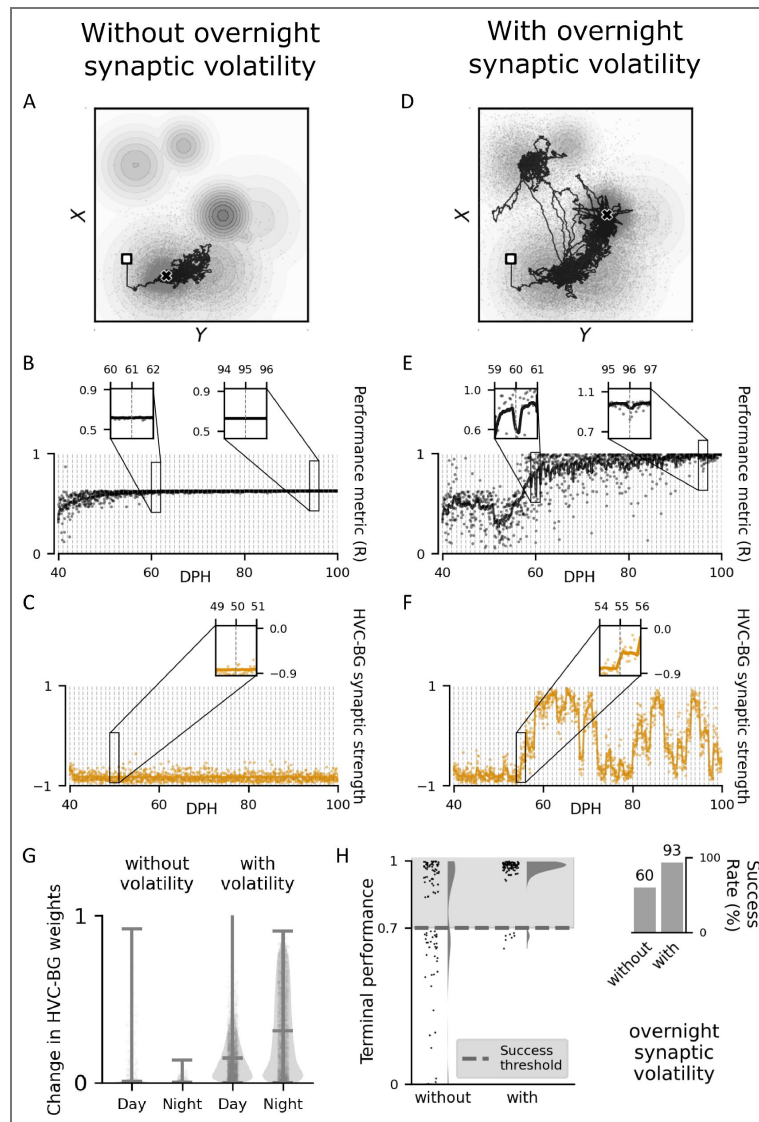


Figure 5. Synaptic volatility induces overnight displacement in performance and aids sensorimotor learning.

The two columns show one representative example simulation of the learning of a single vocal gesture (single time step), with (right) and without (left) overnight synaptic volatility. **A.** Representative learning trajectory in the absence of synaptic volatility. Each black dot corresponds to one rendition of the motor output. The model converges (cross) at the sub-optimal solution closest to the initial position (square). **B.** Smoothened trajectory (black line) of the performance metric (black dots) over a representative simulation. Vertical grey lines denote day change. Inset: Performance metric over two days during the mid and late stages of sensorimotor learning displays a monotonic performance quality. **C.** Smoothened trajectory (yellow line) of the synaptic strength (yellow dots) of one representative HVC-BG weight. Vertical grey lines denote day change. Inset: Synaptic strength over two days shows gradual changes in the HVC-BG weight during the day and absence of overnight fluctuation. **D.** Representative learning trajectory with synaptic volatility, with same convention as in A. The model escapes sub-optimal solutions and converges (cross) at the globally optimal performance. **E.** Smoothened trajectory (black line) of the performance metric (black dots) over a representative simulation. Vertical grey lines denote day change. Inset: Performance metric over two days during the mid and late stages of sensorimotor learning displays gradual changes in performance during the day and sudden overnight fluctuations. The magnitude of overnight performance fluctuation reduces over the sensorimotor period. **F.** Smoothened trajectory (yellow line) of the synaptic strength (yellow dots) of one representative HVC-BG weight. Vertical grey lines denote day change. Inset: Synaptic strength over two days shows gradual changes in the HVC-BG weight during the day and strong fluctuations overnight. **G.** Change in HVC-BG synaptic strengths across the day vs across the night over one simulation with and and one without overnight synaptic volatility in the BG pathway. **H.** Terminal performance (left, black dots) and success rate (right) of the model over 100 simulations with and without overnight synaptic volatility in the BG pathway.

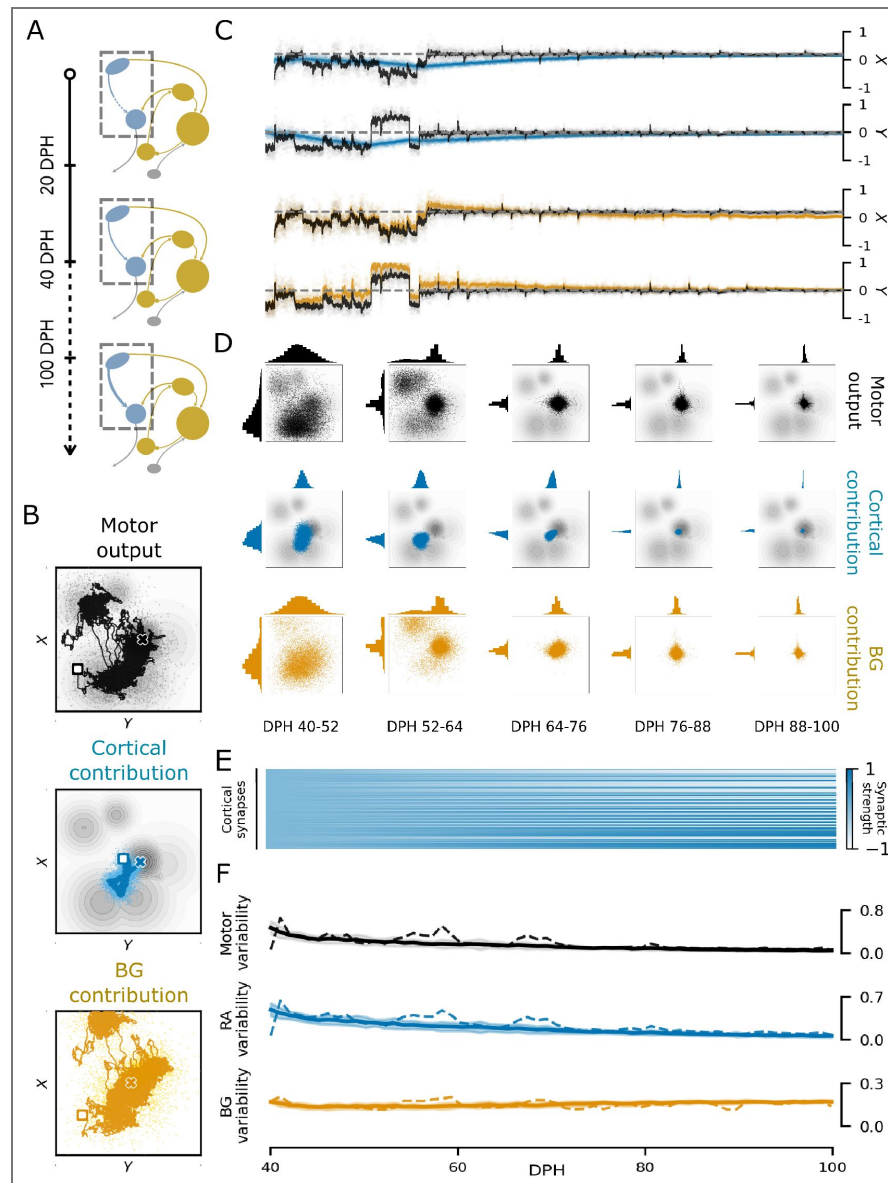


Figure 6. Delayed maturation of cortical pathway modulates the exploration-exploitation trade off.

A. Schema of delayed maturation of the HVC projections to RA and its growing strength over development in male zebra finches. **B.** Representative trajectory of the model output for one time-step (single vocal gesture) in the dual pathway model. Smoothened trajectories of the overall motor output (top, black line), contribution of the cortical pathway to motor output (middle, blue line) and contribution of the BG pathway to motor output (bottom, yellow line). The cortical contribution slowly consolidates a trace of the BG-led exploration of the motor space. Squares denote initial positions and crosses denote final positions. **C.** Learning trajectory along each dimension. The overall motor output (black) overlaps primarily with the BG contribution (yellow) in the beginning of the sensorimotor phase. Over time, it aligns increasingly with the cortical pathway contribution (blue), as the model approaches convergence. As the motor output reaches the target output (grey dotted line), the BG contribution slowly fades to zero. **D.** Change in motor output across different stages of sensorimotor tasks (Top: overall output, middle: cortical pathway contribution, bottom: BG pathway contribution). The histogram show the distribution of motor output (dots) on the landscape in the two dimensions. BG contribution (bottom panel) initially induces a strong displacement in the overall motor output (top panel) as the cortical contribution (middle panel) incorporates a slow trace. However, its effect on motor output fades over time, eventually contributing to moderate exploration around the mainly cortical-driven output. **E.** Growth of synaptic strength (shades of blue) in the cortical pathway over the sensorimotor period (each line represents a single synapse). The HVC-RA synapses, initialized at zero weight, drive a strong net excitatory or net inhibitory effect on the postsynaptic RA populations at the end of learning. **F.** Top panel: Motor variability reduces over the sensorimotor period. Middle panel: Variability of RA activity reduces over the sensorimotor period. Bottom panel: Variability of BG activity remains stable over the sensorimotor period.

Figure 5e [↗](#)). Overall, the model exhibits a gradual transition from BG-driven exploration to cortical-driven exploitation, reflecting a transfer of motor control from subcortical to cortical circuits. This shift emerges naturally from the delayed maturation of the cortical pathway, which implicitly regulates the exploration–exploitation trade-off and supports efficient sensorimotor learning.

Dual pathway model surpasses alternate strategies for sensorimotor learning

We next assessed whether a dual-pathway architecture is necessary for robust learning by comparing it with single-pathway reinforcement learning (RL) models under different strategies. A standard single-pathway RL model (StdRL, see Section Methods) performed poorly on artificial landscapes, succeeding in fewer than one-third of simulations at low noise levels and remaining trapped in suboptimal solutions (Figure 7a [↗](#)). Increasing noise improved exploration and raised success rates, but excessive noise degraded final performance by maintaining high variability. Thus, exploration via noise introduces a limiting trade-off.

To mitigate this, we tested a variant with developmentally decreasing noise (DevRL, see Section Methods). While DevRL improved performance at moderate noise levels by enabling early exploration and later stabilization, success rates remained capped ($\approx 70\%$, Figure 7b [↗](#)). Moreover, the range of noise level that allows good performance (0.3–0.4) is relatively tight and requires finetuning to be successful. At higher noise levels, performance again deteriorated, indicating that this approach does not fully resolve the exploration–exploitation trade-off.

In contrast, the dual-pathway model consistently outperformed both StdRL and DevRL across noise conditions (Figure 7c [↗](#)). Its combination of BG-driven variability, synaptic volatility, and gradual cortical consolidation supports broad early exploration while ensuring stable performance after convergence, without parameter fine-tuning. Finally, we compared the model to simulated annealing (see Section Methods), a standard optimization strategy that explicitly controls exploration via a temperature parameter (Kirkpatrick et al., 1983 [↗](#)). Simulated annealing achieved high success rates at sufficient noise levels while maintaining good final performance (Figure 7d [↗](#)), owing to the temperature parameter that progressively suppresses exploratory variability over time. Despite mechanistic differences, it produces dynamics similar to the dual-pathway model.

Overall, these results show that effective sensorimotor learning requires adaptive regulation of exploration and exploitation. The dual-pathway architecture achieves this intrinsically through its biological design, providing a robust alternative to single-pathway learning strategies.

Robustness

We finally evaluated the robustness of the dual-pathway model to parameter variations and task conditions. Increasing intrinsic BG noise initially improved performance by enhancing exploration, but excessive noise impaired convergence (Figure 8a [↗](#)). Notably, high success rates ($>70\%$) were maintained across a broad intermediate range, indicating tolerance to variability in exploratory drive. Similarly, the model remained stable under substantial noise in RA, with performance degrading only at very high levels (>0.5) where learning became unstable (Figure 8b [↗](#)). This demonstrates robustness to motor variability.

Varying learning rates revealed complementary constraints. BG learning was robust across a range of values but became unstable at high rates (Figure 8c [↗](#)). Cortical learning tolerated a wide range ($1e^{-5}$ to $2e^{-4}$), but very slow learning ($\leq 1e^{-5}$) effectively removed the cortical pathway, while overly fast learning ($\geq 1e^{-3}$) led to premature consolidation and suboptimal convergence (Figure 8d [↗](#)). Thus, efficient learning requires cortical consolidation to operate on a slower timescale than BG-driven exploration.

We also tested robustness to task structure (Figure 8e [↗](#)). The model reliably converged on broad optima and performed well on moderately sharp peaks, though performance declined for extremely narrow optima (62–92% success rate at gaussian peak widths of 0.2–0.3; 16% success rate

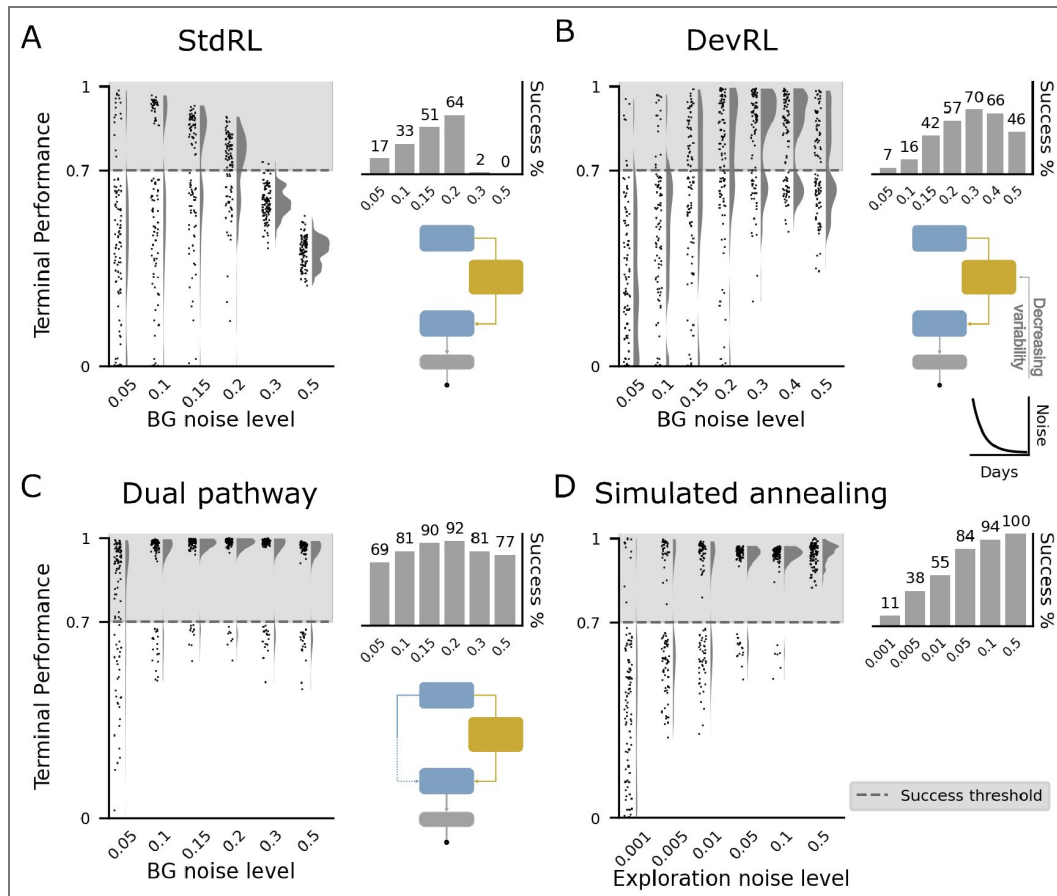


Figure 7. Comparison of different strategies for sensorimotor learning.

In A-D, Terminal performance over 100 simulations (black dots) on artificial landscapes is depicted on the left for various levels of exploratory noise (violin plot: distribution of the terminal performance). Success rate, i.e. the proportion of simulations that learned the optimal performance (achieved terminal performance $> .7$), is depicted on the right for the same levels of exploratory noise. **A.** Performance of a single pathway architecture employing gradient descent based reinforcement learning with no overnight synaptic volatility (StdRL). **B.** Performance of a single pathway architecture employing gradient descent based reinforcement learning with exploratory variability reducing exponentially over development (DevRL). **C.** Performance of the dual pathway model, with overnight synaptic volatility and no explicit reduction in BG variability. **D.** Performance of a simulated annealing algorithm on sensorimotor learning on gaussian-based landscapes.

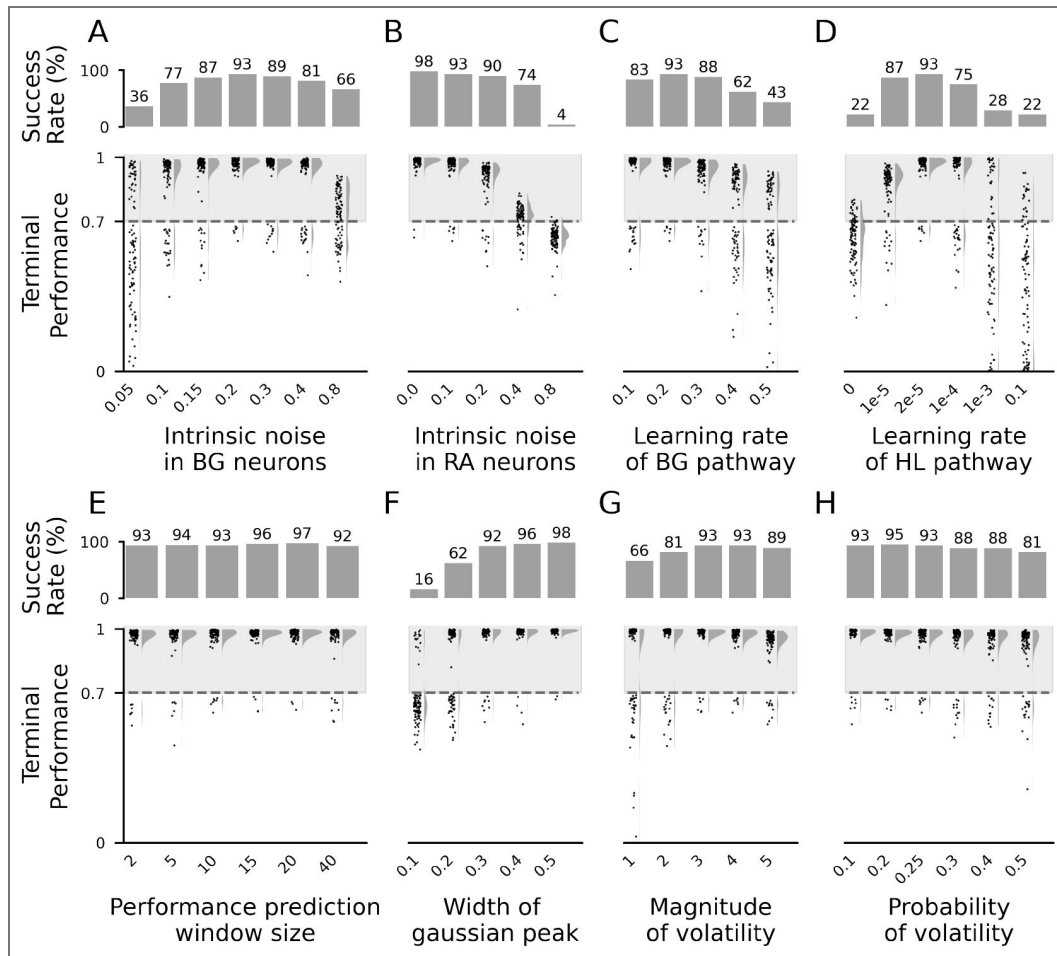


Figure 8. The dual pathway model is robust to fluctuations in key parameters that affect learning.

Here, the performance of the model is evaluated over 100 simulations for different values of each parameter considered. **A.** Synaptic noise into BG neurons. **B.** synaptic noise into RA neurons. **C.** Learning rate for plasticity at HVC-BG synapses. **D.** Learning rate of plasticity at HVC-RA synapses. **E.** Number of renditions over which prediction of performance ($\langle \bar{R} \rangle$) is computed (temporal integration window). **F.** Width of the global optimum gaussian in the gaussian-based performance landscapes. **G.** Scaling factor for overnight synaptic volatility at HVC-BG synapses. **H.** Probability of occurrence of overnight synaptic volatility at HVC-BG synapses.

at gaussian peak widths of 0.1). Learning was largely insensitive to the temporal window used to compute performance prediction error (Figure 8f), with strong performance even for minimal integration (two renditions). Finally, we examined parameters governing overnight synaptic volatility. Low volatility reduced performance, resembling purely local exploration, whereas moderate volatility improved success by enabling escape from local optima (Figure 8g). The model was relatively insensitive to volatility probability, although excessively frequent perturbations disrupted convergence by inducing persistent instability (Figure 8h).

Overall, the dual-pathway framework demonstrates robust performance across a wide range of neural, learning, and task parameters, while highlighting key constraints on the balance between exploration, consolidation, and stability.

Discussion

We developed a computational model of sensorimotor learning grounded in the anatomy, physiology, and developmental trajectory of the zebra finch song system and singing behavior. The model integrates three key features: reinforcement learning in a cortico-BG pathway, Hebbian consolidation in a parallel cortical motor pathway, and overnight synaptic volatility in cortico-BG synapses. Our model reproduces the non-monotonic learning trajectory observed in juvenile zebra finches, the neuronal activity patterns in the network and the reported effect of circuit perturbations. Importantly, the proposed mechanism for sensorimotor learning enables reliable convergence to global optima in complex, multi-peaked performance landscapes, a setting in which standard RL approaches typically fail.

Delayed maturation of cortical pathway implicitly steers the exploration-exploitation trade-off

Motor variability is essential early in learning, enabling exploration of movements that can later be reinforced if successful (Sutton and Barto, 2018; Wu et al., 2014). As learning progresses, however, variability decreases and eventually impairs performance (Cohen et al., 2007; Darshan et al., 2014; Pekny et al., 2015; Dhawale et al., 2017). In zebra finches, increasing similarity to the tutor song is accompanied by reduced vocal variability (Tchernichovski et al., 2001; Ölveczky et al., 2005). Our model reproduces this transition through delayed maturation of the motor pathway (HVC–RA) and explains a key well-established experimental dissociation: the BG pathway is necessary for song learning but largely dispensable for crystallized song production (Bottjer et al., 1984; Scharff and Nottebohm, 1991; Aronov et al., 2008). This emerges from a gradual transfer of motor control from the BG pathway to the cortical pathway. Early in learning, weak HVC–RA connectivity allows BG-driven variability to dominate RA activity. Consistent with this, HVC–RA synapses emerge relatively late (>40 dph) and strengthen progressively through the sensorimotor period until 90 dph (Herrmann and Bischof, 1986; Bottjer et al., 1986; Kittelberger and Mooney, 1999; Garst-Orozco et al., 2014; Sankar et al., 2022). As cortical consolidation proceeds, motor patterns stabilize and dependence on BG input declines. In the model, strengthening cortical input increasingly saturates RA neurons through Hebbian consolidation of BG-guided vocalizations, progressively reducing the impact of BG variability. Consequently, BG lesions have stage-dependent effects: early lesions impair learning, intermediate lesions halt refinement, and late lesions minimally affect performance. This accounts for the decreasing influence of LMAN on RA activity and vocal output, alongside the increasing role of HVC (Brainard and Doupe, 2000; Ölveczky et al., 2005; Aronov et al., 2008; Kao and Brainard, 2006). Notably, LMAN–RA synaptic input remains largely unchanged during learning (Nordeen et al., 1992; Garst-Orozco et al., 2014), and LMAN activity itself remains variable (Achiro et al., 2017). Consistently cutting HVC–RA axons in older juveniles or adults simply reverts song into subsong (Aronov et al., 2008). Similar increases in cortical control during learning have been observed in other species (Hwang et al., 2021). More broadly, this subcortical-to-cortical transfer parallels phenomena such as motor-skill automatization and systems-level memory consolidation. In humans, early learning relies more on prefrontal and striatal circuits, whereas later stages depend increasingly on cortical representations (Doyon and Benali, 2005; Dayan and Cohen,

2011 [↗](#)). Related principles have also been proposed for hippocampal–cortical memory transfer (Benchenane et al., 2011 [↗](#); Maviel et al., 2004 [↗](#)) and cerebellar learning (Mauk, 1997 [↗](#); Bae et al., 2025 [↗](#)). In these systems, Hebbian plasticity in a direct motor or cortical pathway is thought to consolidate behaviors or memories initially acquired through a parallel circuit. The dual-pathway architecture provides a mechanistic account of these transitions grounded in pathway-specific plasticity rather than abstract learning stages.

Role of overnight performance deterioration

Neuromuscular control of the syrinx poses a highly redundant, nonlinear sensorimotor learning problem, in which many motor configurations yield suboptimal vocal outputs (Srivastava et al., 2015 [↗](#)) and transitions between them follow complex, nonlinear and discontinuous trajectories (Tchernichovski et al., 2001 [↗](#)). In such landscapes, gradient-based RL readily becomes trapped in local optima. A central and counterintuitive result of our model is that overnight performance deterioration improves long-term learning. Without synaptic volatility, the system converges reliably but prematurely. Stochastic perturbations in BG synapses displaces vocal output, enabling renewed exploration and eventual convergence to the global optimum.

This provides a functional interpretation of the observation that post-sleep deterioration in juvenile zebra finches predicts better eventual imitation (Derégnaucourt et al., 2005 [↗](#)). In the model, this effect arises not from passive decay but from active synaptic remodeling. Consistent with this idea, rapid spine turnover has been observed over day-long timescales in HVC (Roberts et al., 2010 [↗](#)), as in the neocortex (De Paola et al., 2006 [↗](#); Mongillo et al., 2017 [↗](#)) and cortico-striatal synapses in mammals (Shen and Dayan, 2025 [↗](#); Kuo and Liu, 2019 [↗](#)). Moreover, motor learning is associated with deep structural dynamism of synaptic boutons over days, with diverse changes in neighboring synapses consistent with large stochastic changes in synaptic strength over days (Shen and Dayan, 2025 [↗](#)). We further predict that the magnitude of overnight synaptic change scales inversely with within-day learning, linking daytime consolidation with nighttime exploration. This hypothesis could be tested by tracking synaptic dynamics across development or manipulating sleep. More broadly, these results support the view that sleep actively reorganizes neural circuits to enhance learning (Rasch and Born, 2013 [↗](#); Stickgold, 2005 [↗](#)). Non-monotonic learning trajectories, rather than reflecting instability, may instead be a hallmark of efficient global search in complex sensorimotor landscapes.

Relationship to other computational models

Early RL models of song learning have demonstrated that dopaminergic signals can drive gradient ascent in motor space via BG circuits (Doya and Sejnowski, 1994 [↗](#), 1998 [↗](#)) and have been implemented with realistic plasticity rules and spiking neuronal networks (Fiete et al., 2007 [↗](#); Farries and Fairhall, 2007 [↗](#)). Recent experimental evidence demonstrated that dopamine in the BG encode a performance error signal (Gadagkar et al., 2016 [↗](#); Kasdin et al., 2025 [↗](#)) that can drive song adaptation (Xiao et al., 2018 [↗](#)). Concerning the consolidation of song adaptation initially driven by the BG–cortical circuit in the motor pathway (Andalman and Fee, 2009 [↗](#); Warren et al., 2011 [↗](#)), Teşileanu et al. (2017 [↗](#)) formalized its implementation through experimentally evidenced spike-timing plasticity at HVC–RA synapses (Mehaffey and Doupe, 2015 [↗](#)), showing that the timescale of LMAN input to RA is optimal for such consolidation given the plasticity rule reveal experimentally. However, these models have generally assumed simple or unimodal performance landscapes and have not addressed the challenges posed by multiple local optima. Our model extends this framework by explicitly considering complex, multi-modal landscapes derived from a biophysical model of the syrinx. In this setting, standard RL approaches and noise-based heuristics often fail. By combining BG-driven exploration, cortical consolidation, and synaptic volatility, the model provides a mechanistic solution to escaping local optima, connecting circuit-level plasticity to the non-monotonic behavioral trajectory first described by (Derégnaucourt et al., 2005 [↗](#); Kollmorgen et al., 2020 [↗](#)) and left largely unexplained by prior models.

Our work also situates song learning within the broader landscape of sensorimotor learning theory. During sensorimotor learning, both cerebellar and BG circuits contribute to learning, albeit with learning mechanisms (Doya, 2000 [DOI](#)): supervised learning relying on internal models for the cerebellum (Wolpert and Kawato, 1998 [DOI](#); Ito, 2008 [DOI](#)), RL for the BG (Houk and Wise, 1995 [DOI](#); Fee and Goldberg, 2011 [DOI](#)). While cerebellar circuits may play a role in the temporal patterning of song (Pidoux et al., 2018 [DOI](#); Ursu et al., 2026 [DOI](#)), our model focusses on BG-related learning. Model-free RL approaches applied to continuous motor control (Peters and Schaal, 2008 [DOI](#)) typically assume smooth, convex performance landscapes and do not readily address the acquisition of motor programs from scratch in multi-modal landscapes typical of highly redundant motor systems with high degrees of freedom (Bernstein, 1967 [DOI](#)) such as the zebra finch singing apparatus (Düring et al., 2013 [DOI](#)). Moreover, RL implementation in neural circuits face precisely the exploration–exploitation limitations that motivated our study (Ishii et al., 2002 [DOI](#); Humphries, 2012 [DOI](#)).

Our model also shares conceptual similarities with optimization strategies such as simulated annealing (Kirkpatrick et al., 1983 [DOI](#)). Early high variability and large across-day shifts resemble high-temperature exploration, while gradual stabilization parallels cooling. However, unlike simulated annealing, which requires explicit scheduling and global memory of explored states, the dual-pathway model achieves similar dynamics through distributed, biologically plausible mechanisms. The analogy extends to mini-batch gradient descent (Lecun et al., 1998 [DOI](#); Krizhevsky et al., 2017 [DOI](#)), a variant of stochastic gradient descent (Robbins and Monro, 1951 [DOI](#)) where learning proceeds through successive subsets of data. There exists actually many different strategies (mini-batch sampler) that dictates how to build each mini-batch (Zhao and Zhang, 2014 [DOI](#); Csiba and Richtárik, 2016 [DOI](#); Newling and Fleuret, 2016 [DOI](#)), and reducing the batch size may improve learning (Obando-Ceron et al., 2023 [DOI](#)). In our model, each “day” effectively constitutes a batch, with overnight variability reshaping the sampling distribution. This may contribute to efficient convergence, although the biological implementation differs substantially from engineered algorithms. Overall, the results suggest that structural organization, i.e. parallel pathways with distinct plasticity rules and timescales, can provide a powerful and robust solution to exploration–exploitation challenges in both biological and artificial systems. Other optimization techniques used in machine learning such as simulated annealing may succeed in pathological landscapes that pose difficulties to the model proposed here. However, they are less biologically realistic as they rely on a long-term maintenance of the memory of all explored options.

Limitations

The model incorporates several simplifying assumptions that should be addressed in future work. HVC activity is treated as fixed, whereas experimental data indicate substantial developmental changes (Mackevicius et al., 2023 [DOI](#)), including synaptic remodeling and neurogenesis (Mooney, 2009 [DOI](#)). Incorporating plastic sequence generation could reveal interactions between sensory and motor learning. Many computational models of HVC sequence generation have been proposed (including but not limited to (Jin et al., 2007 [DOI](#); Armstrong and Abarbanel, 2016 [DOI](#); Cannon et al., 2015 [DOI](#))) along with mechanisms for the acquisition of the sequence pattern during sensory learning (Fiete et al., 2010 [DOI](#); Shao et al., 2026 [DOI](#)). The model also assumes a pre-existing dopaminergic performance signal without addressing how this evaluative signal develops. Understanding the emergence and tuning of this “critic” remains an open question.

Motor output is represented in a highly simplified, low-dimensional space. While sufficient to capture key principles, this abstraction does not reflect the full complexity of the syrinx and respiratory system (Düring et al., 2013 [DOI](#)). More detailed biomechanical models could provide a richer test of the framework (Mindlin, 2017 [DOI](#); Amador et al., 2025 [DOI](#)). Given the difficulty of building a realistic vocal production organ, we chose to investigate the capacity of our model for a simple model of the syrinx or for various arbitrary performance landscape. Temporal structure is also simplified: each time step within a motif is treated independently, whereas real song learning involves continuous sequences with temporal dependencies. Extending the model to capture sequence-level learning and timing control is an important direction. Finally, the use of rate-based

neurons omits spike-timing dynamics and detailed synaptic mechanisms. While appropriate for capturing system-level behavior, future implementations could incorporate spiking models (Farries and Fairhall, 2007 [↗](#)) to better align with physiological data.

Predictions and conclusion

Beyond reproducing known phenomena, the model generates several testable predictions. It suggests that synaptic volatility in the BG pathway—potentially mediated by spine turnover—drives overnight changes in behavior. Rapid spine formation and elimination over day-long timescales have been observed in rodent neocortex (Holtmaat and Svoboda, 2009 [↗](#)) and in HVC of zebra finches (Mooney, 2009 [↗](#)), although comparable measurements in Area X are currently lacking. Direct measurements of synaptic dynamics of HVC-X synapses would be required to test this hypothesis. The model also predicts that BG lesions should abolish overnight performance deterioration, and that neural activity patterns in LMAN and RA should show larger changes across nights than within days.

This work bridges behavioral, anatomical and neurophysiological evidence from songbirds and computational modeling to elucidate the mechanisms underlying sensorimotor learning. Our dual-pathway model not only aligns with existing data but also outperforms traditional RL approaches, offering a robust and biologically plausible alternative for real-world sensorimotor control. It achieves functionality analogous to optimization strategies such as simulated annealing with an implicit regulation of the exploration–exploitation trade-off through delayed maturation of the consolidation pathway. Future work could explore the applicability of these principles to other learning systems, both biological and artificial, and further refine our understanding of how the brain optimizes behavior in dynamic environments.

Methods

Model simulations were conducted in Python. Code, parameters, and analysis scripts are publicly available at *Github* [↗](#).

Model Architecture

Inspired by the vocal learning circuitry of songbirds (the song system) (Figure 1b [↗](#)), we designed the model as a multilayered architecture comprising two major parallel pathways, the cortical motor pathway and the subcortical BG pathway (Figure 1c [↗](#)). The model consists of four layers corresponding to HVC, RA, the BG (representing the AFP), and the MC (an abstraction of the downstream motor control structures).

The first layer, HVC, serves as an input layer that provides timing and contextual information, indicating which target syllable is to be produced. The second layer, RA, generates a motor command based on input from HVC and modulatory input from the BG pathway. The third layer, BG, forms a parallel pathway between HVC and RA and receives evaluative feedback signaling performance quality. The fourth layer, MC, mimics the input to an avian syrinx and transforms RA motor commands into a low-dimensional control signal that is subsequently transformed into a syllable vocalisation using a biophysical model of the avian syrinx (Titze and Martin, 1998 [↗](#); Mindlin, 2017 [↗](#)).

In the zebra finch, the RA contains approximately 13,000–15,000 neurons per hemisphere during the early sensorimotor phase (Herrmann and Arnold, 1991 [↗](#)). During this period, approximately 5,000–10,000 neurons from LMAN project to RA (Bottjer et al., 1986 [↗](#)). In addition, the HVC contains roughly 23,000–40,000 neurons per hemisphere during the sensorimotor period, although only a fraction of these neurons project to RA and are active during song production (Hahnloser et al., 2002 [↗](#)). These anatomical proportions and connectivity constraints informed the design of the dual-pathway model architecture.

In the model, the HVC and RA layers each comprise 100 rate-coded neuronal population units, while the BG layer comprises 50 rate-coded units. The relative sizes of these layers were chosen to approximately reflect the proportions of neurons in the corresponding nuclei of the song system,

preserving a 1:2:2 ratio between the number of units in LMAN (BG layer), HVC, and RA, respectively. Both BG and RA units are modeled using sigmoidal activation functions that bound firing rates between -1 and 1 (Eqs 4, 5, 6, 7). Each unit represents the cumulative activity of a local population of excitatory and inhibitory neurons, such that its output corresponds to high (≈ 400 Hz for HVC, RA and ≈ 200 Hz for LMAN) or low (≈ 0 Hz) effective firing rates.

The BG layer receives input from HVC (Eq 1). The RA layer receives two distinct sets of inputs (Figure 1c): a direct projection from HVC via the cortical motor pathway (Eq 2) and a modulatory projection from the BG via the subcortical pathway (Eq 3). RA activity is then transmitted to a downstream motor controller layer (MC), which transforms the RA neural command into a low-dimensional motor control signal (Eq 8).

The HVC and RA layers are fully connected, as are the HVC and BG layers. Synaptic weights in these pathways are plastic, reflecting activity-dependent plasticity at RA and BG synapses. HVC–RA weights are initialized to zero, whereas HVC–BG weights are initialized randomly from a uniform distribution over $[-1, 1]$. Positive weights produce a net excitatory effect on the target population, while negative weights produce a net inhibitory effect. In contrast, the BG–RA and RA–MC pathways are fully connected with fixed positive weights.

The two-dimensional motor output of the MC layer specifies the model's sampling position in the sensorimotor space, which is subsequently transformed into a syllable-level vocalization, as described in subsection Performance Landscapes.

$$\theta_{\text{HVC-BG}} = \frac{J_{\text{HVC}} \cdot W_{\text{HVC-BG}}}{\sum J_{\text{HVC}}} \quad (1)$$

$$\theta_{\text{HVC-RA}} = \frac{J_{\text{HVC}} \cdot W_{\text{HVC-RA}}}{\sum J_{\text{HVC}}} \quad (2)$$

$$\theta_{\text{BG-RA}} = \frac{J_{\text{BG}} \cdot W_{\text{BG-RA}}}{\sum W_{\text{BG-RA}}} \quad (3)$$

$$J_{\text{BG}} = \sigma_{\text{BG}} (\theta_{\text{HVC-BG}} + \xi_{\text{BG}}) \quad (4)$$

$$J_{\text{RA}} = \sigma_{\text{RA}} (\theta_{\text{HVC-RA}} + k\theta_{\text{BG-RA}} + \xi_{\text{RA}}) \quad (5)$$

$$\sigma_{\text{BG}}(x) = 2 \cdot \left(\frac{1}{1 + e^{-m_{\text{BG}}(x - a_{\text{BG}})}} \right) - 1 \quad (6)$$

$$\sigma_{\text{RA}}(x) = 2 \cdot \left(\frac{1}{1 + e^{-m_{\text{RA}}(x - a_{\text{RA}})}} \right) - 1 \quad (7)$$

$$J_{\text{MC}} = \frac{J_{\text{RA}} \cdot W_{\text{RA-MC}}}{\sum W_{\text{RA-MC}}} \quad (8)$$

with

J_x - activity of the layer x ,

W_{x-y} - weights connecting layer x to layer y ,

ξ_x - intrinsic noise in layer x , drawn from a normal distribution

θ_{x-y} - input from layer x to layer y

η - learning rate

k - fixed scaling factor.

Learning

Plastic synaptic weights in the BG pathway ($\mathbf{W}_{\text{HVC-BG}}$) are updated using a reinforcement learning rule based on covariance learning (Eq 9). Exploratory variability is introduced locally within the BG layer as intrinsic noise (ξ_{BG} ; Eq 4). In addition, the BG layer receives a scalar evaluative signal reflecting relative performance quality, quantified as the performance prediction error (PPE). The PPE is defined as the difference between the performance evaluation on a given trial (R) and the expected performance evaluation ($\bar{R}$; Eq 10). In contrast, plastic synaptic weights in the cortical motor pathway ($\mathbf{W}_{\text{HVC-RA}}$) are updated using a Hebbian learning rule, whereby coincident activity in pre- and post-synaptic units leads to potentiation of the corresponding synapse (Eq 12). All plastic synaptic weights are bounded within the range [-1,1].

$$\Delta W_{\text{HVC-BG}} = \eta_{\text{RL}} \cdot J_{\text{HVC}} \cdot J_{\text{BG}} \cdot (\text{PPE}) \quad (9)$$

$$\text{PPE} = R - \bar{R} \quad (10)$$

$$\bar{R}_t = \frac{1}{L} \sum_{l=1}^L R_{t-l} \quad (11)$$

$$\Delta W_{\text{HVC-RA}} = \eta_{\text{HL}} \cdot J_{\text{HVC}} \cdot J_{\text{RA}} \quad (12)$$

Further, the reinforcement learning driven by the BG pathway is supplemented with behavioral features drawn from empirical evidence. Specifically, a daily deterioration in song performance has been shown to occur post-sleep in juvenile birds during the sensorimotor period. This post-sleep deterioration is implemented as a fluctuation of the HVC-BG weights at the end of a day (Eq 13), which results in a displacement of the BG output on the following day. The magnitude of the overnight volatility in the HVC-BG synapses is determined stochastically (ξ_{night}) from a uniform distribution [-1, 1] and is scaled inversely with the change experienced by the HVC-BG weights over the previous day (p) (Eq 14).

Note: For comparisons with the model when lacking overnight jumps (Fig 5), this overnight synaptic volatility was omitted, while all other aspects of the model and learning rules were kept identical.

$$\Delta W_{\text{HVC-BG}} = \eta_{\text{RL}} \cdot \xi_{\text{night}} \cdot p \quad (13)$$

$$p = 1 - f(|W_{\text{HVC-BG}}^{t=T} - W_{\text{HVC-BG}}^{t=0}|) \quad (14)$$

where f is a sigmoid function with an optimum threshold and slope.

Simulation

Within each simulation, the model learns a song motif composed of four syllables in the task with the syrinx-based performance landscapes, and 20 sub-syllabic notes (4 syllables $\times$ 5 sub-syllabic notes) of 10ms each in the task with the artificial performance landscapes, over a 60-day sensorimotor period (Figure 1a). Each day, the model performs 1000 renditions of the song motif, with each rendition consisting of the syllables or sub-syllabic notes generated in a fixed temporal sequence. For each rendition of a syllable or sub-syllabic note, a performance metric,

indicating its similarity to the corresponding target syllable or note, is obtained from the performance landscape corresponding to the target syllable or note, and used to compute the performance prediction error (PPE).

Performance Landscapes

We evaluated the model using two types of performance landscapes: syrinx-based and Gaussian-based landscapes (Figure S1 [\[1\]](#)). For the syrinx-based landscapes, the MC layer generates a motor command that is mapped to a vocal output, which is then evaluated relative to a target tutor song. This transformation and comparison are implemented using a biophysical model of the avian syrinx, allowing motor commands to be translated into vocalizations whose similarity to the tutor template defines performance quality. The two-dimensional scalar output of the MC layer, J_{MC} , is transformed into input signals for the syrinx model according to Eq 15 [\[1\]](#) and Eq 16 [\[1\]](#). Specifically, the MC layer modulates two control signals, J_{MC}^{α} and J_{MC}^{β} , corresponding to air-sac pressure, $\alpha(t)$, and syringeal labial tension, $\beta(t)$, respectively. These pressure and tension waveforms drive oscillations in the trachea, generating a syllable vocalization as a pressure waveform of duration $T = 50$ ms (Supplementary figure S1a [\[1\]](#); Amador and Margoliash (2013 [\[2\]](#))). A spectrogram is constructed from this oscillatory pressure waveform and compared to a target syllable selected from the vocal repertoire of zebra finches. Vocal similarity is quantified using a similarity metric (s) based on the correlation coefficient between the normalized spectrograms of the generated vocalization (S_X) and the target template (S_T) (Eq 17 [\[1\]](#)).

To construct the performance landscape, we simulated vocalizations over a range of MC outputs—corresponding to different combinations of pressure and tension parameters—and computed the similarity metric for each vocalization relative to the target syllable. The resulting performance landscape is represented as a two-dimensional matrix that captures vocal similarity across the pressure–tension input space for a given target syllable. Using four distinct target syllables, we generated four corresponding syrinx-based performance landscapes. As shown in supplementary Figure S1b [\[1\]](#), the resulting performance landscapes exhibit multiple global optima, reflecting the inherent redundancy of the syrinx, whereby distinct combinations of pressure and tension can produce acoustically similar vocal outputs. In addition to these global optima, the landscape contains numerous local optima distributed across the sensorimotor space, arising from the nonlinear biomechanics of sound production.

$$\alpha(t) = J_{MC}^{\alpha} + 0.04 \sin((T/2 + t) * \pi * 10) \quad (15)$$

$$\beta(t) = J_{MC}^{\beta} - 0.2e^{-200t} - 10^{-5}e^{200t} \quad (16)$$

where the output of the MC, J_{MC}^{α} [J^{α} , J^{β}], T : duration of the vocalisation and $t \in [0, T]$.

$$r = \frac{\langle s_X - \bar{s}_X, s_T - \bar{s}_T \rangle}{\|s_X - \bar{s}_X\| \|s_T - \bar{s}_T\|} \quad (17)$$

Here, s_k denotes one-dimensional vector obtained by flattening the two-dimensional matrix S_k , $\bar{s}_k$ denotes its mean and r corresponds to their Pearson correlation coefficient.

Second, to generate performance landscapes with systematically varying complexity, we adopted a direct computational approach. We constructed continuous performance landscapes by mapping the two-dimensional motor space (J_{MC}) onto a one-dimensional scalar performance evaluation. Each performance landscape was generated as the maximum over a set of two-dimensional Gaussian distributions, each centered at a distinct randomly sampled location in motor space and characterized by a specific width. Similar to the syrinx-based performance landscapes, the superposition of these distributions yields landscapes with multiple local optima alongside a single global optimum. Landscape complexity was controlled by varying the number, spatial

Parameter		Value
Reinforcement Learning rate	η_{RL}	0.2
Hebbian Learning Rate	η_{HL}	2×10^{-5}
BG noise	ξ_{BG}	0.2
RA noise	ξ_{RA}	0.1
Performance prediction window size	L	10
Landscape density		Moderate
Global optimum gaussian width		0.3
Magnitude of volatility		0.4
Probability of volatility		0.25

Table 1. Model parameters and their values used in all simulations unless otherwise stated.

arrangement, and widths of the Gaussian components. Local optima were constrained to reach a maximum performance value of 70% of that at the global optimum, and all landscapes were normalized such that the global optimum yielded a performance metric of 1. Based on the density of local optima, landscapes were grouped into four classes of : low (2-5), moderate (3-9), compact (8-18) and high (11-25) number of local optima. All simulations were run on landscapes the moderate density of local optima, unless otherwise stated. [Figure S1d](#) shows representative examples from each class. These Gaussian-based landscapes allow us to systematically assess the versatility and robustness of the model across a wide range of task complexities.

Performance Metrics

Model performance was quantified using two metrics: terminal performance and success rate. The terminal performance of a syllable was defined as the mean performance evaluation obtained from its corresponding performance landscape over the last 100 renditions on the final day of learning. This metric captures the imitation quality of the learned motor output after learning has stabilized.

For the Gaussian-based performance landscapes, the global optimum was assigned a performance value of 1, while all local optima were constrained to reach a maximum performance value of 0.7. Accordingly, a simulation was considered successful if it achieved a terminal performance exceeding 0.7. For the syrinx-based performance landscapes, we observed that the highest peak outside the global optimum had an associated performance value of approximately 0.52, whereas the global optima reached a value of 1. To maintain consistency across landscape types, we apply the same success threshold of 0.7 for syrinx-based simulations. The success rate for a given experimental condition was defined as the proportion of simulations that exceeded the success threshold relative to the total number of simulations.

Motor variability is calculated as standard deviation of the output on each dimension over the day. End-of-day metrics are calculated as the mean of the metric over the last 100 renditions of the day. Start-of-day metrics are calculated as the mean of the metric over the first 100 renditions of the day.

Experiments

To isolate the contribution of the basal ganglia (BG) pathway to sensorimotor learning, we implemented a lesion-like manipulation by silencing BG output to RA. Specifically, the projection from the BG pathway to the RA layer was inactivated, effectively removing LMAN-mediated input to RA while leaving the cortical (HVC–RA) pathway intact. This manipulation allowed us to assess model performance when motor production and learning were governed solely by the cortical pathway. Lesions were applied at different stages of sensorimotor learning to examine developmental effects. We performed the manipulation at three time points corresponding to before learning begins (40 dph), during learning (70 dph), and post learning (100 dph). Following the lesion, simulations were continued without further modification to learning rules or parameters, and performance metrics were evaluated as described above. We performed an analogous manipulation targeting the cortical pathway, inactivating the HVC–RA projection while leaving the BG pathway intact, to assess learning when driven solely by BG input.

Illustrations

To illustrate spiking activity corresponding to the scalar firing rates generated by the model, we create raster plots in [Figure 1](#), [4](#). Spike trains were simulated where the firing rate (r) generated by the model for every 10 ms time bin is linearly mapped onto physiologically observed rate of firing of the particular neuron ($R = 400$ Hz for HVC, RA neurons and $R = 200$ Hz for LMAN neurons). For each time bin of $\Delta = 1$ ms during song, probability of a spike being generated was determined as $p \propto r \cdot R \cdot \Delta t$. A refractory period of 1 ms was enforced after every occurrence of a spike.

Benchmark Algorithms

To compare the performance of the dual-pathway model with established computational approaches, we constructed a single-pathway framework implementing different reinforcement learning (RL) strategies. In this framework, a single pathway injects exploratory noise, receives performance evaluation, updates synaptic weights, and directly governs motor output. Operationally, this corresponds to lesioning the cortical motor pathway and allowing the BG-like pathway to exert sole control over motor output.

First, we evaluated the single-pathway framework using a standard reinforcement learning approach (StdRL). In this condition, learning is governed by the gradient-descent-based update rule applied in the BG pathway of the dual-pathway model (Eq 9 [↗](#)).

Second, we tested a modified reinforcement learning strategy with developmentally decreasing noise (DevRL). In this approach, the magnitude of exploratory noise injected into the single pathway is reduced exponentially over the course of learning (Eq 19 [↗](#)) such that the final noise is around 10% of the initial noise.

$$\xi_{BG} \sim N(0, \sigma_{BG}(t)) \quad (18)$$

$$\sigma_{BG}(t) = \sigma_{BG}(t=0) \exp(-2.2t/T) \quad (19)$$

Third, we compared the dual-pathway model with simulated annealing (SA), a probabilistic optimization technique commonly used to identify global optima in complex search spaces. In this framework, transitions to lower-reward states are permitted with a probability determined by an acceptance function M , which is determined by an exponentially decreasing temperature parameter, scaled by the difference in performance evaluation between successive iterations (Eq 20 [↗](#), 21 [↗](#), 22 [↗](#)). This temperature parameter explicitly controls the balance between exploration and exploitation over learning.

$$T_{tr} = 1 - e^{-c_1/tr} \quad (20)$$

$$\Delta R_{tr} = R_{candidate} - R_{current} \quad (21)$$

$$M(tr) = e^{\frac{\Delta R_{tr}}{T_{tr}}} \quad (22)$$

where ΔR_{tr} denotes the difference in performance evaluation between a randomly chosen candidate motor output, $R_{candidate}$ and the current motor output $R_{current}$. The candidate motor output is sampled within a range corresponding to $\eta_{rl} * \xi$ (Eq 9 [↗](#)), ensuring a step size comparable to that used in the preceding reinforcement learning algorithms.

Supplementary

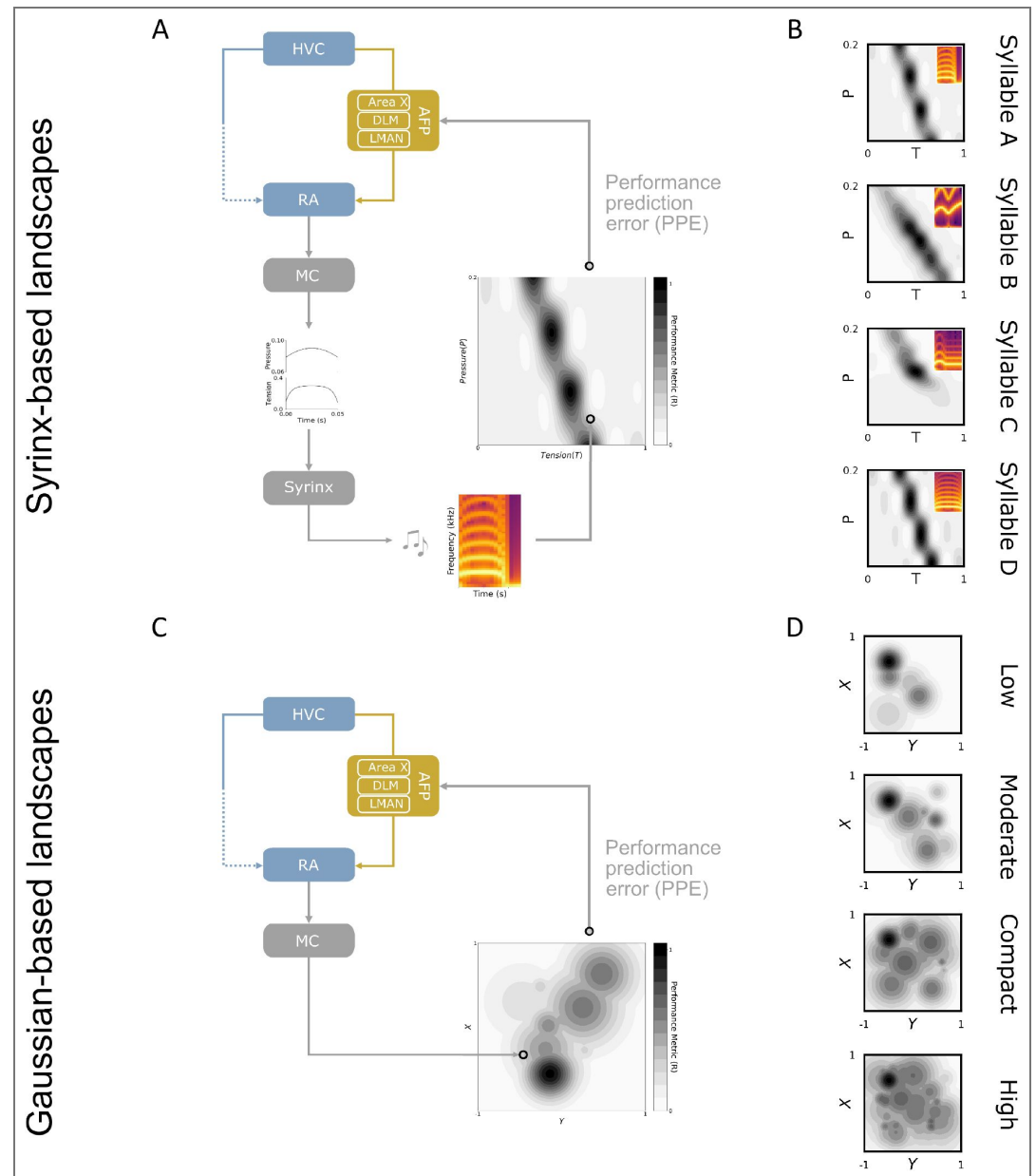


Figure S1. Construction of sensorimotor performance landscapes. **A.** Interface of model with syrinx-based landscapes. The neuronal activity pattern in RA is reduced in dimension at the MC layer. The output from the MC layer modulates 50 ms tension and pressure waves, which is fed as input to a biophysical model of the syrinx to produces zebra finch-like vocalizations. For each output of the model, the produced vocalization is compared with a target syllable to generate a performance metric (grey heatmap). Thus, the performance landscape represents the imitation quality for every model output. **B.** Four sample performance landscapes generated using four target zebra finch syllables (inset). The landscapes have a few global optima (performance metric ≈ 1) and several local optima (performance metric $< .7$). **C.** Interface of model with gaussian-based landscapes. The output from the MC layer is directly used as coordinates on these sensorimotor performance landscapes. **D.** Examples of gaussian-based landscapes from each density category. The landscape is generated by superimposing several gaussian hills. The superimposition results in several local (performance metric $< .7$) and one globally optimal solution (performance metric = 1).

Data availability

Computational article without any dataset associated.

Acknowledgements

We would like to thank Alexis Dubreuil for thoughtful discussions and Manfred Gahr from the Max Planck Institute for Biological Intelligence for technical support and funding. This work has been supported by funding from the French Agence Nationale de la Recherche (grant number ANR-17-CE24-0036) to NR and (20-CE37-0023-02, 22-NEUC-0002-01) to AL, as well as Fonds pour la Recherche Médicale (EQU202203014688) to AL.

Additional information

Funding

Funder	Grant reference number	Author
Agence Nationale de la Recherche (ANR)	ANR-17-CE24-0036	Nicolas P Rougier
Agence Nationale de la Recherche (ANR)	20-CE37-0023-02	Arthur Leblois
Agence Nationale de la Recherche (ANR)	22-NEUC-0002-01	Arthur Leblois
Fondation pour la Recherche Médicale (FRM)	EQU202203014688	Arthur Leblois

Author ORCID iDs

Remya Sankar:  <https://orcid.org/0000-0002-2428-4592>

Nicolas P Rougier:  <https://orcid.org/0000-0002-6972-589X>

Arthur Leblois:  <https://orcid.org/0000-0002-9392-5939>

References

- Achiro JM, Shen J, Bottjer SW** (2017) Neural activity in cortico-basal ganglia circuits of juvenile songbirds encodes performance during goal-directed learning. *eLife* **6**:e26973 <https://doi.org/10.7554/elife.26973> | [PubMed](#)
- Amador A, Margoliash D** (2013) A Mechanism for Frequency Modulation in Songbirds Shared with Humans. *Journal of Neuroscience* **33**:11136-11144 <https://doi.org/10.1523/jneurosci.5906-12.2013> | [PubMed](#)
- Amador A, Mindlin GB, Elemans CPH** (2025) Transitions and tricks: nonlinear phenomena in the avian voice. *Philosophical Transactions of the Royal Society B: Biological Sciences* **4**:380 <https://doi.org/10.1098/rstb.2024.0007> | [PubMed](#)
- Andalman AS, Fee MS** (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proceedings of the National Academy of Sciences* **106**:12518-12523 <https://doi.org/10.1073/pnas.0903214106> | [PubMed](#)
- Armstrong E, Abarbanel HDI** (2016) Model of the songbird nucleus HVC as a network of central pattern generators. *Journal of Neurophysiology* **116**:2405-2419 <https://doi.org/10.1152/jn.00438.2016> | [PubMed](#)
- Aronov D, Andalman AS, Fee MS** (2008) A Specialized Forebrain Circuit for Vocal Babbling in the Juvenile Songbird. *Science* **320**:630-634 <https://doi.org/10.1126/science.1155140> | [PubMed](#)
- Aronov D, Veit L, Goldberg JH, Fee MS** (2011) Two Distinct Modes of Forebrain Circuit Dynamics Underlie Temporal Patterning in the Vocalizations of Young Songbirds. *The Journal of Neuroscience* **31**:16353-16368 <https://doi.org/10.1523/jneurosci.3009-11.2011> | [PubMed](#)
- Bae H, Seo J, Kim CE, Kim SJ** (2025) A normative principle governing memory transfer in cerebellar motor learning. *Nature Communications* **16** <https://doi.org/10.1038/s41467-025-60511-z> | [PubMed](#)

- Bellman R (1957) A Markovian Decision Process. *Journal of Mathematics and Mechanics* **6**:679-684 <https://doi.org/10.1512/iumj.1957.6.56038>
- Benchenane K, Tiesinga PH, Battaglia FP (2011) Oscillations in the prefrontal cortex: a gateway to memory and attention. *Current Opinion in Neurobiology* **21**:475-485 <https://doi.org/10.1016/j.conb.2011.01.004> | PubMed
- Bernstein NA (1967) *The Co-ordination and Regulation of Movements* Oxford, UK: pnPergamon Press.
- Bottjer SW, Miesner EA, Arnold AP (1984) Forebrain Lesions Disrupt Development But Not Maintenance of Song in Passerine Birds. *Science* **224**:901-903 <https://doi.org/10.1126/science.6719123> | PubMed
- Bottjer SW, Miesner EA, Arnold AP (1986) Changes in neuronal number, density and size account for increases in volume of song-control nuclei during song development in zebra finches. *Neuroscience Letters* **67**:263-268 [https://doi.org/10.1016/0304-3940\(86\)90319-8](https://doi.org/10.1016/0304-3940(86)90319-8) | PubMed
- Brainard MS, Doupe AJ (2000) Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* **404**:762-766 <https://doi.org/10.1038/35008083> | PubMed
- Cannon J, Kopell N, Gardner T, Markowitz J (2015) Neural Sequence Generation Using Spatiotemporal Patterns of Inhibition. *PLOS Computational Biology* **11**:e1004581 <https://doi.org/10.1371/journal.pcbi.1004581> | PubMed
- Cepeda C, Hurst RS, Calvert CR, Hernández-Echeagaray E, Nguyen OK, Jocoy E, Christian LJ, Ariano MA, Levine MS (2003) Transient and Progressive Electrophysiological Alterations in the Corticostriatal Pathway in a Mouse Model of Huntington's Disease. *The Journal of Neuroscience* **23**:961-969 <https://doi.org/10.1523/jneurosci.23-03-00961.2003> | PubMed
- Cohen JD, McClure SM, Yu AJ (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**:933-942 <https://doi.org/10.1098/rstb.2007.2098> | PubMed
- Csiba D, Richtárik P (2016) Importance Sampling for Minibatches. *arXiv* <https://doi.org/10.48550/arXiv.1602.02283>
- Darshan R, Leblois A, Hansel D (2014) Interference and Shaping in Sensorimotor Adaptations with Rewards. *PLoS Computational Biology* **10**:e1003377 <https://doi.org/10.1371/journal.pcbi.1003377> | PubMed
- Dayan E, Cohen LG (2011) Neuroplasticity Subservicing Motor Skill Learning. *Neuron* **72**:443-454 <https://doi.org/10.1016/j.neuron.2011.10.008> | PubMed
- De Paola V, Holtmaat A, Knott G, Song S, Wilbrecht L, Caroni P, Svoboda K (2006) Cell Type-Specific Structural Plasticity of Axonal Branches and Boutons in the Adult Neocortex. *Neuron* **49**:861-875 <https://doi.org/10.1016/j.neuron.2006.02.017> | PubMed
- Derégnaucourt S, Mitra PP, Fehér O, Pytte C, Tchernichovski O (2005) How sleep affects the developmental learning of bird song. *Nature* **433**:710-716 <https://doi.org/10.1038/nature03275> | PubMed
- Dhawale AK, Smith MA, Ölveczky BP (2017) The Role of Variability in Motor Learning. *Annual Review of Neuroscience* **40**:479-498 <https://doi.org/10.1146/annurev-neuro-072116-031548> | PubMed
- Doya K (2000) Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology* **10**:732-739 [https://doi.org/10.1016/s0959-4388\(00\)00153-7](https://doi.org/10.1016/s0959-4388(00)00153-7) | PubMed
- Doya K, Sejnowski T (1994) A Novel Reinforcement Model of Birdsong Vocalization Learning. In: *Advances in Neural Information Processing Systems*. **7**
- Doya K, Sejnowski TJ (1998) A Computational Model of Birdsong Learning by Auditory Experience and Auditory Feedback. In: Poon PWF, Brugge JF (Eds). *Central Auditory Processing and Neural Modeling* Springer. pp. 77-88 https://doi.org/10.1007/978-1-4615-5351-9_8

- Doyon J, Benali H (2005) Reorganization and plasticity in the adult brain during learning of motor skills. *Current Opinion in Neurobiology* **15**:161-167 <https://doi.org/10.1016/j.conb.2005.03.004> | PubMed
- Düring DN, Ziegler A, Thompson CK, Ziegler A, Faber C, Müller J, Scharff C, Elemans CP (2013) The songbird syrinx morphome: a three-dimensional, high-resolution, interactive morphological map of the zebra finch vocal organ. *BMC Biology* **11** <https://doi.org/10.1186/1741-7007-11-1> | PubMed
- Farries MA, Fairhall AL (2007) Reinforcement Learning With Modulated Spike Timing-Dependent Synaptic Plasticity. *Journal of Neurophysiology* **98**:3648-3665 <https://doi.org/10.1152/jn.00364.2007> | PubMed
- Fee MS, Goldberg JH (2011) A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* **198**:152-170 <https://doi.org/10.1016/j.neuroscience.2011.09.069> | PubMed
- Fiete IR, Fee MS, Seung HS (2007) Model of Birdsong Learning Based on Gradient Estimation by Dynamic Perturbation of Neural Conductances. *Journal of Neurophysiology* **98**:2038-2057 <https://doi.org/10.1152/jn.01311.2006> | PubMed
- Fiete IR, Senn W, Wang CZH, Hahnloser RHR (2010) Spike-Time-Dependent Plasticity and Heterosynaptic Competition Organize Networks to Produce Long Scale-Free Sequences of Neural Activity. *Neuron* **65**:563-576 <https://doi.org/10.1016/j.neuron.2010.02.003> | PubMed
- Gadagkar V, Puzerey PA, Chen R, Baird-Daniel E, Farhang AR, Goldberg JH (2016) Dopamine neurons encode performance error in singing birds. *Science* **354**:1278-1282 <https://doi.org/10.1126/science.aah6837> | PubMed
- Garst-Orozco J, Babadi B, Ölveczky BP (2014) A neural circuit mechanism for regulating vocal variability during song learning in zebra finches. *eLife* **3**:e03697 <https://doi.org/10.7554/elife.03697> | PubMed
- George AP, Powell WB (2006) Adaptive stepsizes for recursive estimation with applications in approximate dynamic programming. *Machine Learning* **65**:167-198 <https://doi.org/10.1007/s10994-006-8365-9>
- Gershman SJ, Daw ND (2017) Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology* **68**:101-128 <https://doi.org/10.1146/annurev-psych-122414-033625> | PubMed
- Graybiel AM, Grafton ST (2015) The Striatum: Where Skills and Habits Meet. *Cold Spring Harbor Perspectives in Biology* **7**:a021691 <https://doi.org/10.1101/cshperspect.a021691> | PubMed
- Hahnloser RHR, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**:65-70 <https://doi.org/10.1038/nature00974> | PubMed
- Herrmann K, Arnold A (1991) The development of afferent projections to the robust archistriatal nucleus in male zebra finches: a quantitative electron microscopic study. *The Journal of Neuroscience* **11**:2063-2074 <https://doi.org/10.1523/jneurosci.11-07-02063.1991> | PubMed
- Herrmann K, Bischof H (1986) Delayed development of song control nuclei in the zebra finch is related to behavioral development. *Journal of Comparative Neurology* **245**:167-175 <https://doi.org/10.1002/cne.902450204> | PubMed
- Hikosaka O, Nakamura K, Sakai K, Nakahara H (2002) Central mechanisms of motor skill learning. *Current Opinion in Neurobiology* **12**:217-222 [https://doi.org/10.1016/s0959-4388\(02\)00307-0](https://doi.org/10.1016/s0959-4388(02)00307-0) | PubMed
- Holtmaat A, Svoboda K (2009) Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience* **10**:647-658 <https://doi.org/10.1038/nrn2699> | PubMed
- Houk JC, Wise SP (1995) Feature Article: Distributed Modular Architectures Linking Basal Ganglia, Cerebellum, and Cerebral Cortex: Their Role in Planning and Controlling Action. *Cerebral Cortex* **5**:95-110 <https://doi.org/10.1093/cercor/5.2.95> | PubMed
- Humphries M (2012) Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in Neuroscience* **6** <https://doi.org/10.3389/fnins.2012.00009> | PubMed

- Hwang EJ, Dahlen JE, Mukundan M, Komiyama T (2021) Disengagement of Motor Cortex during Long-Term Learning Tracks the Performance Level of Learned Movements. *The Journal of Neuroscience* **41**:7029-7047 <https://doi.org/10.1523/jneurosci.3049-20.2021> | PubMed
- Ishii S, Yoshida W, Yoshimoto J (2002) Control of exploitation–exploration meta-parameter in reinforcement learning. *Neural Networks* **15**:665-687 [https://doi.org/10.1016/s0893-6080\(02\)00056-4](https://doi.org/10.1016/s0893-6080(02)00056-4) | PubMed
- Ito M (2008) Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience* **9**:304-313 <https://doi.org/10.1038/nrn2332> | PubMed
- Jin DZ, Ramazanoğlu FM, Seung HS (2007) Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. *Journal of Computational Neuroscience* **23**:283-299 <https://doi.org/10.1007/s10827-007-0032-z> | PubMed
- Kao MH, Brainard MS (2006) Lesions of an Avian Basal Ganglia Circuit Prevent Context-Dependent Changes to Song Variability. *Journal of Neurophysiology* **96**:1441-1455 <https://doi.org/10.1152/jn.01138.2005> | PubMed
- Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature* **433**:638-643 <https://doi.org/10.1038/nature03127> | PubMed
- Kasdin J, Duffy A, Nadler N, Raha A, Fairhall AL, Stachenfeld KL, Gadagkar V (2025) Natural behaviour is learned through dopamine-mediated reinforcement. *Nature* **641**:699-706 <https://doi.org/10.1038/s41586-025-08729-1> | PubMed
- Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by Simulated Annealing. *Science* **220**:671-680 <https://doi.org/10.1126/science.220.4598.671> | PubMed
- Kittelberger JM, Mooney R (1999) Lesions of an Avian Forebrain Nucleus That Disrupt Song Development Alter Synaptic Connectivity and Transmission in the Vocal Premotor Pathway. *The Journal of Neuroscience* **19**:9385-9398 <https://doi.org/10.1523/jneurosci.19-21-09385.1999> | PubMed
- Kollmorgen S, Hahnloser RHR, Mante V (2020) Nearest neighbours reveal fast and slow components of motor learning. *Nature* **577**:526-530 <https://doi.org/10.1038/s41586-019-1892-x> | PubMed
- Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Communications of the ACM* **60**:84-90 <https://doi.org/10.1145/3065386>
- Kuo HY, Liu FC (2019) Synaptic Wiring of Corticostriatal Circuits in Basal Ganglia: Insights into the Pathogenesis of Neuropsychiatric Disorders. *eneuro* **6**:ENEURO.0076–19.2019 <https://doi.org/10.1523/eneuro.0076-19.2019> | PubMed
- Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**:2278-2324 <https://doi.org/10.1109/5.726791>
- Mackevicius EL, Gu S, Denisenko NI, Fee MS (2023) Self-organization of songbird neural sequences during social isolation. *eLife* **12**:e77262 <https://doi.org/10.7554/elife.77262> | PubMed
- Mauk MD (1997) Roles of Cerebellar Cortex and Nuclei in Motor Learning: Contradictions or Clues?. *Neuron* **18**:343-346 [https://doi.org/10.1016/s0896-6273\(00\)81235-0](https://doi.org/10.1016/s0896-6273(00)81235-0) | PubMed
- Maviel T, Durkin TP, Menzaghi F, Bontempi B (2004) Sites of Neocortical Reorganization Critical for Remote Spatial Memory. *Science* **305**:96-99 <https://doi.org/10.1126/science.1098180> | PubMed
- Mehaffey WH, Doupe AJ (2015) Naturalistic stimulation drives opposing heterosynaptic plasticity at two inputs to songbird cortex. *Nature Neuroscience* **18**:1272-1280 <https://doi.org/10.1038/nn.4078> | PubMed
- Mindlin GB (2017) Avian vocal production beyond low dimensional models. *Journal of Statistical Mechanics: Theory and Experiment* **2017**:024005 <https://doi.org/10.1088/1742-5468/aa54d8>
- Mongillo G, Rumpel S, Loewenstein Y (2017) Intrinsic volatility of synaptic connections — a challenge to the synaptic trace theory of memory. *Current Opinion in Neurobiology* **46**:7-13 <https://doi.org/10.1016/j.conb.2017.06.006> | PubMed

- Mooney R, Rao M** (1994) Waiting periods versus early innervation: the development of axonal connections in the zebra finch song system. *The Journal of Neuroscience* **14**:6532-6543 <https://doi.org/10.1523/jneurosci.14-11-06532.1994> | PubMed
- Mooney R** (2009) Neurobiology of song learning. *Current Opinion in Neurobiology* **19**:654-660 <https://doi.org/10.1016/j.conb.2009.10.004> | PubMed
- Newling J, Fleuret F** (2016) Nested Mini-Batch K-Means. *arXiv* <https://doi.org/10.48550/arXiv.1602.02934>
- Nordeen EJ, Grace A, Burek MJ, Nordeen KW** (1992) Sex-dependent loss of projection neurons involved in avian song learning. *Journal of Neurobiology* **23**:671-679 <https://doi.org/10.1002/neu.480230606> | PubMed
- Obando-Ceron J, Bellemare MG, Castro PS** (2023) Small batch deep reinforcement learning. *arXiv* <https://doi.org/10.48550/arXiv.2310.03882>
- Ölveczky BP, Gardner TJ** (2011) A bird's eye view of neural circuit formation. *Current Opinion in Neurobiology* **21**:124-131 <https://doi.org/10.1016/j.conb.2010.08.001> | PubMed
- Pekny SE, Izawa J, Shadmehr R** (2015) Reward-Dependent Modulation of Movement Variability. *The Journal of Neuroscience* **35**:4015-4024 <https://doi.org/10.1523/jneurosci.3244-14.2015> | PubMed
- Peters J, Schaal S** (2008) Reinforcement learning of motor skills with policy gradients. *Neural Networks* **21**:682-697 <https://doi.org/10.1016/j.neunet.2008.02.003> | PubMed
- Pidoux L, Le Blanc P, Levenes C, Leblois A** (2018) A subcortical circuit linking the cerebellum to the basal ganglia engaged in vocal learning. *eLife* **7**:e32167 <https://doi.org/10.7554/elife.32167> | PubMed
- Rasch B, Born J** (2013) About Sleep's Role in Memory. *Physiological Reviews* **93**:681-766 <https://doi.org/10.1152/physrev.00032.2012> | PubMed
- Robbins H, Monro S** (1951) A Stochastic Approximation Method. *The Annals of Mathematical Statistics* **22**:400-407 <https://doi.org/10.1214/aoms/1177729586>
- Roberts TF, Tschida KA, Klein ME, Mooney R** (2010) Rapid spine stabilization and synaptic enhancement at the onset of behavioural learning. *Nature* **463**:948-952 <https://doi.org/10.1038/nature08759> | PubMed
- Sankar R, Rougier NP, Leblois A** (2022) Computational benefits of structural plasticity, illustrated in songbirds. *Neuroscience and Biobehavioral Reviews* **132**:1183-1196 <https://doi.org/10.1016/j.neubiorev.2021.10.033> | PubMed
- Scharff C, Nottebohm F** (1991) A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *The Journal of Neuroscience* **11**:2896-2913 <https://doi.org/10.1523/jneurosci.11-09-02896.1991> | PubMed
- Schultz W, Dayan P, Montague PR** (1997) A Neural Substrate of Prediction and Reward. *Science* **275**:1593-1599 <https://doi.org/10.1126/science.275.5306.1593> | PubMed
- Shao S, Riquelme JL, Gjorgjieva J** (2026) Spontaneous emergence and evolution of neuronal sequences in recurrent networks. *eLife* **15**:RP110718 <https://doi.org/10.7554/elife.110718>
- Shen T, Dayan P** (2025) Individual differences in tail risk sensitive exploration using Bayes-adaptive Markov decision processes. *eLife* **13**:RP100366 <https://doi.org/10.7554/elife.100366> | PubMed
- Sober SJ, Brainard MS** (2009) Adult birdsong is actively maintained by error correction. *Nature Neuroscience* **12**:927-931 <https://doi.org/10.1038/nn.2336> | PubMed
- Srivastava KH, Elemans CPH, Sober SJ** (2015) Multifunctional and Context-Dependent Control of Vocal Acoustics by Individual Muscles. *The Journal of Neuroscience* **35**:14183-14194 <https://doi.org/10.1523/jneurosci.3610-14.2015> | PubMed
- Stickgold R** (2005) Sleep-dependent memory consolidation. *Nature* **437**:1272-1278 <https://doi.org/10.1038/nature04286> | PubMed
- Sutton RS, Barto AG** (2018) *Reinforcement Learning: An Introduction* (Second) The MIT Press.

- Tchernichovski O, Mitra PP, Lints T, Nottebohm F (2001) Dynamics of the Vocal Imitation Process: How a Zebra Finch Learns Its Song. *Science* **291**:2564-2569 <https://doi.org/10.1126/science.1058522> | PubMed
- Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP (2000) A procedure for an automated measurement of song similarity. *Animal Behaviour* **59**:1167-1176 <https://doi.org/10.1006/anbe.1999.1416> | PubMed
- Teşileanu T, Ölveczky B, Balasubramanian V (2017) Rules and mechanisms for efficient two-stage learning in neural circuits. *eLife* **6**:e20944 <https://doi.org/10.7554/eLife.20944> | PubMed
- Titze IR, Martin DW (1998) Principles of Voice Production. *The Journal of the Acoustical Society of America* **104**:1148-1148 <https://doi.org/10.1121/1.424266>
- Ursu R, Centeno E, Leblois A (2026) A lobule-specific neuronal representation of song temporal structure in the songbird cerebellum. *eLife* **15**:RP109802 <https://doi.org/10.7554/elife.109802>
- Warren TL, Tumer EC, Charlesworth JD, Brainard MS (2011) Mechanisms and time course of vocal learning and consolidation in the adult songbird. *Journal of Neurophysiology* **106**:1806-1821 <https://doi.org/10.1152/jn.00311.2011> | PubMed
- Wickens JR, Reynolds JN, Hyland BI (2003) Neural mechanisms of reward-related motor learning. *Current Opinion in Neurobiology* **13**:685-690 <https://doi.org/10.1016/j.conb.2003.10.013> | PubMed
- Williams RJ (1992) Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning* **8**:229-256 <https://doi.org/10.1007/bf00992696>
- Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. *Neural Networks* **11**:1317-1329 [https://doi.org/10.1016/s0893-6080\(98\)00066-5](https://doi.org/10.1016/s0893-6080(98)00066-5) | PubMed
- Wu HG, Miyamoto YR, Castro LNG, Ölveczky BP, Smith MA (2014) Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature Neuroscience* **17**:312-321 <https://doi.org/10.1038/nn.3616> | PubMed
- Xiao L, Chattree G, Oscos FG, Cao M, Wanat MJ, Roberts TF (2018) A Basal Ganglia Circuit Sufficient to Guide Bird-song Learning. *Neuron* **98**:208-221.e5 <https://doi.org/10.1016/j.neuron.2018.02.020> | PubMed
- Zhao P, Zhang T (2014) Accelerating Minibatch Stochastic Gradient Descent using Stratified Sampling. *arXiv* <https://doi.org/10.48550/arXiv.1405.3080>
- Ölveczky BP, Andalman AS, Fee MS (2005) Vocal Experimentation in the Juvenile Songbird Requires a Basal Ganglia Circuit. *PLoS Biology* **3**:e153 <https://doi.org/10.1371/journal.pbio.0030153> | PubMed
- Ölveczky BP, Otchy TM, Goldberg JH, Aronov D, Fee MS (2011) Changes in the neural control of a complex motor sequence during learning. *Journal of Neurophysiology* **106**:386-397 <https://doi.org/10.1152/jn.00018.2011> | PubMed

Peer reviews

Reviewer #1 (Public review):

The authors sought to devise a model of the song system that captures the essential features of that neural circuitry and couple it to a behavioral model that captures the challenges of motor learning while remaining tractable. They seek to use this model to explain known features of song learning and relate them to the general problems associated with learning via gradient ascent. Their syrinx model uses two control parameters-air sac pressure and syringeal labial tension-to generate birdsong-like spectrograms. Normalized spectrograms generated by the syrinx model are compared to a target spectrogram by computing Pearson's correlation coefficient. This correlation coefficient quantifies the performance of the model, and the goal of learning is to maximize it. The syringeal model, while simplified, is complex enough to generate multiple local maxima in the correlation coefficient within the 2D control space with wide variation in the magnitude of the maxima, making it challenging to find a

good optimum via gradient ascent. They also use a more abstract motor model where local maxima are generated by randomly placing Gaussians within the 2D control space. These two approaches to modeling motor space are a major strength of this work.

The neural model is extremely generic and consists of units (each representing a population of excitatory and inhibitory neurons) with continuous-valued outputs ("firing rate") ranging from -1 to +1 with sigmoidal activation functions. Premotor HVC simply generates a fixed temporal sequence that drives activity in RA (analogous to the primary motor cortex) that constitutes commands to the motor controller that translates RA output into 2D control signals for the syrinx. A second, indirect, pathway from HVC to RA is represented by a single node ("BG"); in songbirds this pathway consists of 3 structures (one of them quite complex and heterogeneous) with recurrent connections (i.e., from LMAN back to Area X). Learning is primarily driven by performance-modulated Hebbian plasticity in HVC-BG connection weights. There is also Hebbian plasticity in HVC-RA weights that are in effect trained by the RA activity patterns driven by BG-RA connections.

I worry that this model is too simplified to capture essential features of the song system (and cortico-basal ganglia circuits more generally). That problem is most acute in the way the authors model (or fail to model) the anterior forebrain pathway (AFP) through X, DLM, and LMAN; their model in effect reduces the AFP to just LMAN. I do not think that invalidates this study, but there is a significant danger that this model will end up missing the mark in some important way relative to a more realistic model. However, there is another flaw that comes close to doing that—the connection weights between the units are allowed to vary from -1 to +1. That means, for example, that the connections between HVC and RA units can be inhibitory and can flip between excitation and inhibition during learning. I can imagine some hand-wavy justifications for this (e.g., the connection becomes "inhibitory" because excitation to inhibitory neurons becomes stronger than that to excitatory neurons within the unit), but I can't easily imagine one that I would find persuasive.

A key feature of this model is "synaptic volatility" in the connections between HVC and BG. In songbirds, performance often improves steadily throughout the day, then deteriorates after sleep, a feature which the authors suggest is an important method for avoiding getting stuck on relatively low-performing local optima. I find this suggestion to be intriguing and reasonably persuasive. To implement this in their model, after a simulated "day" of practice, the HVC-BG connection weights are partially randomized (synaptic volatility). There is nothing intrinsically wrong with this idea, but the authors imply that there is experimental support for this phenomenon, which they relate to "continuous remodeling of the cortico-striatal synapses with volatility over hours to days." None of the papers cited really support this kind of synaptic randomization. However, given the simplicity of the authors' neural model, the HVC-BG weights are probably the only place this randomization can be implemented. The authors' implementation does not just add noise to HVC-BG weights "overnight"; it is also scaled inversely by the magnitude of accumulated weight change through the day. The authors do not appear to provide a justification for this aspect of their synaptic volatility.

If we accept the authors' model as detailed and accurate enough for their purposes, it does explain several features of song learning and relates them to solving general problems of learning through gradient ascent. This is particularly true of the daily deterioration of performance and how that relates to escaping local maxima. They show that their model reproduces some of the known effects of lesioning the motor (HVC-RA) and anterior forebrain (HVC-BG-RA) pathways and how those effects depend on the current stage of song learning. They also explore the implications of delayed maturation of the HVC-RA pathway, exemplified by a gradual increase in HVC-RA weights. However, it is not clear that this actually happens; the one paper they cite in support of this contention (Mooney and Rao, 1994) shows no such thing (it does show that HVC axons enter RA later in development than LMAN axons do).

Moreover, it's not clear how essential this could be given that some songbird species continue to show profound vocal plasticity throughout their lives and presumably long after their HVC-RA pathway has fully matured. The authors compare their dual pathway model to a single pathway model (an AFP-only model, in effect); a very illuminating comparison that demonstrates the advantage of the dual pathway. They close by showing that dual pathway performance is robust under variation of key parameters.

<https://doi.org/10.7554/eLife.112141.1.sa3>

Reviewer #2 (Public review):

Summary:

The authors describe a computational model for the acquisition of sensorimotor skills and explore these dynamics using vocal learning in the zebra finch, a system rich in experimental data, to describe the developmental trajectory of vocal imitation by trial and error. They set up their model as a dual-pathway system, with a cortical pathway that drives the vocal effector and a basal ganglia (BG) pathway that uses dopamine-mediated reinforcement learning (RL) by gradient descent to optimize the vocal imitation process.

Strengths:

A key strength of the model, due in part to the fact that his model was generated by a computational laboratory that has also contributed significantly to the collection of experiment-driven empirical data, is that the model is biologically constrained and incorporates a considerable amount of experimental data, including some of the latest findings in the field. In addition to providing a compelling model for the acquisition of vocal learning, this biologically based RL model outperforms many current models. A key feature of this model, which makes it unique, is the implementation of a synaptic volatility variable within the BG pathway that aims to mimic published work showing that juvenile birds exhibit post-sleep deterioration. The model uses a motor output to drive a biophysical model of the avian vocal organ (syrinx) and explores not only the ability to copy song acoustic units (syllables) but also the underlying neural dynamics in both the cortical and BG pathways, showing that each converges onto the types of neural activity patterns that are observed experimentally. Because of the richness of experimental data in this system, the authors can perform "computational experiments" where they can block sleep-driven synaptic volatility or lesion various pathways to replicate experimental observations.

In addition to providing important computational insights to our understanding of vocal learning in the songbird, this study provides key insights into the general architectures that are optimal for RL by gradient descent. These include the conclusion that effective RL requires adaptive regulation of exploration and exploitation, that cortical consolidation must occur at a slower timescale than BG-driven exploration, and intriguingly that the introduction of synaptic volatility prevents RL models of incomplete learning by getting "stuck" in local minima.

Weaknesses:

In the methods section, the authors state "... HVC and RA layers are fully connected, as are the HVC and BG layers. Synaptic weights in these pathways are plastic, reflecting activity-dependent plasticity at RA and BG synapses." Unless I missed it, it is unclear how much the authors consider the synaptic differences between HVC and BG inputs to RA. This seems like an important feature to highlight, especially given that HVC-RA connections are primarily AMPA-mediated whereas those from LMAN are predominantly NMDA. The authors should be clearer about how they model these synapses and better highlight (and describe) the importance of these synaptic differences in their modeling efforts. Ideally, they should

evaluate whether the differences in synapse type influence the outcome of their model. It would be interesting, for example, to test the effect on learning of synaptic conductance substitution (i.e., replacing NMADA with AMPA) on the LMAN-RA synapse.

The model focuses exclusively on the interaction of two converging pathways, and learning is based purely on acoustic feature properties of what seem like four independent syllables of similar or identical duration. For this model, this is fine. But it would be helpful for the authors to state more clearly that they are not modeling respiratory influences on syllable production, which include amplitude modulation of the syllables and expiratory pulse duration. It should be noted that the authors do not (unless I missed it) mention the existence or role of recurrent loops in song initiation (and possibly syllable sequencing). They should at least mention this in the discussion, perhaps as a limitation and item for future versions of the model.

<https://doi.org/10.7554/eLife.112141.1.sa2>

Reviewer #3 (Public review):

Summary:

This study modeled vocal learning in zebra finches with a network of three components: a pathway with delayed/slow Hebbian learning that mimics the HVC-RA projection, a pathway with reinforcement learning that mimics the HVC-BG-RA projection, and a motor unit that mimics the syrinx and produces song output. The model convincingly reproduces key features of song learning and is a valuable step towards understanding how vocal learning is substantiated in the song system. Further examination of model assumptions and presentation of testable predictions would increase the impact of the study.

Strengths:

- (1) The model reproduces several key features of song learning, including learning in a non-convex performance landscape, decreasing motor variability during learning, the relative importance of the HVC-RA and HVC-BG-RA pathways at different learning stages, and sleep-related deterioration.
- (2) The model incorporates several key physiological properties of the song system, including performance-dependent dopamine signals to the BG, the neural variability in the system, and multiple global/local optima of the motor production landscape.
- (3) The study convincingly demonstrates the advantages of a dual-pathway network over a single-pathway one.
- (4) The model demonstrates the counterintuitive benefit of sleep-related performance deterioration for facilitating the escape from local optima to reach the global optimum.
- (5) The model seems robust to some variations in model parameters and task structure.

Weaknesses:

- (1) The study could be more impactful if the model can generate new, testable predictions. The predictions provided in the Discussion are not well justified. For example, with the HVC spine turnover, it seems unlikely that BG lesions would abolish overnight performance deterioration. Because the volatility term is inversely related to learning during the day, the changes in LMAN and RA during sleep are not necessarily larger than those during the day.

(2) The section "Neural activity patterns in the model parallels song system neurophysiology" seems fully anticipated because the model construction is based on the known neural activity patterns. Are there new testable predictions from the model?

(3) Certain model assumptions lack explanation or justification.

a) The authors treat "the delayed maturation of the cortical pathway" as an important component. However, it is unclear if/how this component was implemented in the model. If it was not included in the model, the authors should remove statements related to the delayed maturation idea.

b) How critical is the inverse relationship between learning and sleep-deterioration? If it is known that sleep-related deterioration is inversely related to learning during the previous day, a citation should be added. Similarly, it should be clarified if the spine turnover observed in HVC depends on previous plastic changes, like the assumed inverse relationship in the model.

c) Spine turnover has been demonstrated in HVC and not yet in Area X, but the model implements volatility only in the HVC-BG pathway. It would be important to compare the effects of sleep-related volatility in the HVC-RA and HVC-BG-RA projections.

d) In Table 1/Figure 8, the learning rate for HVC-BG is 10^4 times bigger than the learning rate for HVC-RA. What is the biological justification for this difference?

(4) Related to #3, it is unclear how changing those assumptions would affect model performance.

(5) It is not explained/shown why cross-day exploration (with sleep, sporadic) is better than continuous, non-sleep-related ones. Figure 7B presents results with different noise levels, which may approximate non-sleep-related, continuous volatility, but only for the single-pathway model. Comparable simulations by adding continuous volatility in the dual-pathway model would be helpful. For example, would just a bigger intrinsic noise in either pathway confer the same benefit in escaping local optima, e.g., epsilon-greedy exploration? If the authors can establish the advantages of sleep-specific volatility in its particular form (based on day learning) and relate it to other sensorimotor learning behaviors, it could increase the general impact of the study.

(6) The model description can be improved.

a) What are mBG and mRA in Eqs 6/7?

b) Where is the learning rate specified in Equations 1-8?

c) It is unexplained why Equations 1-3 are not in the same form: Equations 1 and 2 normalize by activity, but Equation 3 normalizes by weight (Equation 8 also normalizes by weight). The authors should confirm that these equations are correct.

d) How J_{HVC} is generated should be defined.

e) How is R calculated in Equation 10? Does this model maintain a PPE for each syllable or for the overall performance? Would it make any difference in learning?

f) Are the weight updates in Equations 9 and 13 added at different time points (e.g., immediately after each spike versus after a syllable). This should be clarified (perhaps a diagram would help).

g) How are the optimal threshold and slope determined for Equation 14?

h) Parameters in Equation 20 are not defined.

<https://doi.org/10.7554/eLife.112141.1.sa1>

Author response:

We thank all three reviewers for their detailed and constructive reviews, which will help us improve the manuscript. Concerning the simplicity of the model, we will elaborate on the limitations that arise from the modelling choices and better justify these choices. With that in mind, we believe this level of abstraction is well suited to the study's objective. As a rate-coded model, it is designed to capture interactions between multiple circuit pathways, and lays the foundation to delve into neuronal dynamics in the future for further insight, once the network-level questions we address here - regarding the exploration-exploitation tradeoff and evasion of sub-optimal performance convergence - are well examined. For this purpose, we believe the model's simplified network-level perspective is not a bug, but a feature. We will also explore in depth why and how the dual pathway model performs better in non-convex sensorimotor landscapes, building on prior analytical work that speaks to the question (Sankar, Leblois & Rougier, ICDL 2022). We will further substantiate and discuss our choice concerning the potential mechanisms underlying the proposed overnight synaptic volatility. We will include comparable approaches (including relevant machine learning analogues) in the introduction. Finally, we will clarify our interpretation of the sign of synaptic weights. We hope to incorporate all valuable feedback by the reviewers in the revised manuscript.

References

Remya Sankar, Arthur Leblois, Nicolas P. Rougier (2022). Dual pathway architecture underlying vocal learning in songbirds. In IEEE International Conference on Development and Learning, ICDL 2022, London, United Kingdom, September 12-15, 2022. pages 265-271, IEEE, 2022.

<https://doi.org/10.7554/eLife.112141.1.sa0>