Reward-driven adaptation of movements requires strong recurrent basal ganglia-cortical loops

Arthur Leblois[1], Thomas Boraud[1], David Hansel[2]

1 UMR 5293, IMN, Univ. Bordeaux, CNRS, Bordeaux, France. 2 Cerebral Dynamics, Plasticity and Learning Laboratory, CNRS, Univ. Paris Cité, Paris, France.

*Arthur Leblois

Email : arthur.leblois@u-bordeaux.fr

Abstract

The basal ganglia (BG), a set of subcortical nuclei involved in motor control, sensorimotor integration, and procedural learning, modulate movement through tonic inhibition of thalamo-cortical networks. While essential for sensorimotor integration and learning, the BG are not necessary for executing well-learned movements. During skill learning, they guide behavioral corrections via dopamine-dependent cortico-striatal plasticity, but these corrections become BG-independent as modifications occur in the motor cortex. Existing models of BG function often overlook the feedback dynamics of cortico-BG-thalamo-cortical loops, and do not address the relative role of cortex and the BG in the generation and adaptation of movement. In this work, we develop a theoretical model of this multiregional network, integrating anatomical, physiological, and behavioral evidence to explore how its dynamics shape movement execution and reward-based adaptation. We show that the BG-thalamo-cortical network influences motor output through three key factors: (i) the rich dynamics of its closed-loop architecture, (ii) attractor dynamics from recurrent cortical connections, and (iii) classical reinforcement learning via dopamine-dependent cortico-striatal plasticity. Our study highlights that efficient visuomotor adaptation requires strong feedback from the BG to the cortex. Finally, we propose a mechanism for initial movement learning through motor babbling. This model suggests the BG-cortical network shapes motor output through its intricate closed-loop dynamics and cortico-striatal dopamine-dependent plasticity.

Introduction

The primary motor cortex (M1) plays a key role in the control of voluntary limb movement(1–5). The dynamics of this cortical area depend on its recurrent connectivity as well as on other players like the basal ganglia network (BG). The BG can shape movement through their tonic inhibition of thalamo-cortical networks(6, 7). They are involved in sensorimotor integration and procedural learning REF.

Transient inactivation of BG output does not degrade the performance of well-learned skills(8), suggesting that BG are not involved in the production of already-learned movement(8, 9). Pallidotomy, in which the BG output nucleus is lesioned, is an effective therapy for Parkinsonian and dystonic motor symptoms but has only few deleterious side-effects on movement(10, 11). By contrast, the BG are essential in guiding reinforcement-driven behavioral corrections during skill learning(12, 13). During such learning, neuronal activity changes in the striatum but only later in the motor cortex (16, 17). Moreover, optogenetic modulation of the dopaminergic input to the BG can guide sensorimotor adaptation(14, 15). However, as learning progresses, behavioral corrections become BG independent(16) while slow synaptic modification(17) and macroscale reorganization(18, 19) occur in the motor cortex.

Cognitive, motor and premotor cortical areas provide a strong input to the BG through their striatal and subthalamic projections(20, 21). In return, the activity in the cortex is modulated by the BG output(20, 22) giving rise to loops(23) which are closed at cellular spatial scale(22, 24, 25). In previous theoretical studies, we developed a dynamical model of the BG-thalamo-cortical network incorporating these recurrent feedback loops(26–28).

Current mechanistic circuit models of RL(29) rely on dopamine-dependent long-term plasticity at cortico-striatal synapses(30, 31). They assume that during reward-based learning, these synapses undergo changes which modify the cortical input to BG and thus their output in such a way that the actions which are produced maximize the rewarding outcome of the learned behaviors(23, 29, 30). However, these models do not consider the feedback dynamics of cortico-BG-thalamo-cortical loops. In this work, we develop a theoretical model of the cortico-BG-thalamo-cortical multiregional network to investigate how its dynamics can shape the execution of reaching movements as well as its role in reward-based movement adaptation.

Results

The cortico-BG-thalamo-cortical circuit model. Our model consists of seven populations of neurons representing the striatum, the internal and external parts of the globus pallidus (GPi, GPe), the subthalamic nucleus (STN) and the motor cortex (M_{ctx}) (Fig 1A). Its architecture reflects the anatomy(20, 32) and the connectivity(22, 24) of the cortico-BG-thalamo-cortical network. We assume that the behavioral output of the are reaching movements whose direction and vigor depend on the activity of the neurons in M_{ctx} ((6, 41) see Material and Methods). Specifically, we consider reaching tasks in which a subject as to performed in the direction of a transient visual cue lasting for 250 ms. Anatomy and physiology show that motor and sensory cortical inputs related to the same movement converge in the striatum(35, 36). We therefore include in the model an additional population of cortical neurons, V_{ctx} where neurons are selective to the direction of the visual cue. It corresponds for instance the parietal or temporal cortex(37, 38). This additional population projects to M_{ctx} as well to the striatum (Material and Methods). The visual related input from V_{ctx} into M_{ctx} and striatal neurons depends on the difference of their preferred direction (PD) and the direction of the cue. Finally, during a movement, M_{ctx} neurons receive a homogeneous external time dependent input, $I_{\text{mov}}(t)$, reflecting the motor drive from premotor areas.

Activity in the circuit during a reaching movement. Unless specified otherwise, the parameters of the model are as in Table 1. At rest $I_{\text{mov}}(t) = 0$ and the activity is homogeneous in all the

populations in the network (Figs 1C-D, Fig Sup.) In agreement with experimental data, the population average spontaneous activity in the GPi, GPe, Thalamus and STN is much larger than in M_{ctx} whereas the activity in the striatum is much smaller than in the latter.

When the cue appears in direction θ_{cue},, the striatum receives cue-related input encoding this direction. This input is maximized for striatal neurons whose preferred direction is $\theta = \theta_{\text{cue}}$. At the onset of movement, the input $I_{\text{mov}}(t)$ increases over 500 ms, followed by a decrease lasting 50 ms, resulting in a bell-shaped temporal profile that lasts for 1 second after movement initiation. During movement, the motor cortex activity displays a bumpy spatial profile, where only a subset of neurons increases their firing rate, while others maintain activity near their spontaneous levels (Fig. 1C). This occurs despite the homogeneity of $I_{\text{mov}}(t)$, emerging from the interplay between the recurrent dynamics within M_{ctx} and the strong feedback from the direct and hyperdirect loops. The location of the activity bump in M_{ctx} is biased by the striatum via the direct feedback loop. As a result, the activity in M_{ctx} encodes the direction, θ_{cue}, of the visual cue.

During the movement, the population vector of M_{ctx} neuronal activity varies. It encodes (Material and Methods) the trajectory of the hand (Fig. 1G). The instantaneous direction of the vector determines the direction of the movement. It points the direction of the cue up to small fluctuations which result from the noise in the dynamics. The amplitude of the movement is determined by the length of the population vector at its maximum. When this occurs, the hand starts to move back to its original location.

Movement-related activity reverberates throughout the entire basal ganglia (BG)-thalamo-cortical loop. Consequently, all populations in the basal ganglia, as well as the thalamus, exhibit heterogeneous activity patterns that are directionally tuned to movement. These patterns are inherited from the cortex. Neurons in the internal segment of the globus pallidus (GPi) are strongly inhibited by the striatum (Fig 1D). This inhibition tends to reduce the activity of GPi neurons whose preferred direction (PD) aligns with the movement direction. In contrast, GPi neurons with a PD opposite to the movement direction are strongly excited due to the combined effect of divergent excitatory projections from the subthalamic nucleus (STN) and reduced striatal activity. This differential activation leads to a "bump" of activity in the GPi, which is inverted compared to the activity seen in the cortex, striatum, and STN. As a result, the tuning curves of GPi neurons are inverted (Fig 1H), consistent with experimental findings (39, 40).

Synaptic plasticity between sensory cortex and striatum drives movement adaptation. Next, we investigated in the model the conditions under which reaching movements can adapt through reward-driven learning. To this end we consider a hand-reaching task in absence of visual feedback about the actual hand location(41–43). Originally the subject is instructed to move in the direction of the cue, θ_{cue} . It performs well in this condition in most of the trials (Fig. 1). Later the condition changes. The subject must perform the movement to target a location which is rotated by $\Delta\theta$ with respect to θ_{cue} (Fig 2A). Importantly, the subject is not informed of this new condition. The only information on performance is a reward provided by the experimentalist at the end of a trial, according to the angular distance of the hand with respect to the desired target. Specifically, the rewarded interval is $[\theta_{\text{cue}} + \Delta - \epsilon, \theta_{cue} + \Delta\theta + \epsilon]$. At the end of the nth trial, the weight of the cortico-striatal projection from a sensory neuron with PD, $\theta_{V_{\text{ctx}}}$, and a striatal neuron with PD, θ_{str} , changes according to the reward-modulated tri-partite learning rule (Williams, 1992; Loewenstein&Seung, 2006; see Material and Methods for details)

$$\begin{aligned} W_n(\theta_{\text{str}}, \theta_{V_{\text{ctx}}}) &= W_{n-1}(\theta_{\text{str}}, \theta_{V_{\text{ctx}}}) \\ &\quad + \eta R(n)(m_{\text{str}}(t_{\text{cue}}, \theta_{\text{str}}) - \prec m_{\text{str}}(\theta_{\text{str}}) \succ)\, m_{V_{\text{ctx}}}(t_{\text{cue}}, \theta_{V_{\text{ctx}}} - \theta_{\text{cue}}) \end{aligned}$$

where η is the learning rate, R(n) the value (0 or 1) of the reward at the end of the trial. Here, $m_{\text{str}}(t_{\text{cue}}, \theta_{\text{str}})$ and $m_{V_{\text{ctx}}}(t_{\text{cue}}, \theta_{V_{\text{ctx}}} - \theta_{\text{cue}})$ denote the activity of the striatal and sensory neuron

when the cue appears in the nth trial in the direction θ_{cue}. Finally, $< \mathrm{m}_{\mathrm{str}}(\theta_{\mathrm{str}}) >$ is the time average of the striatal neuron activity over the 500 msec preceding the onset of the cue.

Before adaptation, the aggregate excitatory input received by the striatal neurons from V_{ctx} is biased toward the cue direction. This is because the projections from V_{ctx} to the striatum are direction specific (red line in Fig 2B). During adaptation, these projections change according to the equation above. As a result, their functional specificity is modified and the direction bias of the aggregate input from V_{ctx} into the striatum rotates toward the rewarded desired direction (Fig. 2C).

Figure 2F shows how the reward-driven adaptation is driven by the BG-cortical network. Because the production of movement is noisy, the hand explores many final reaching, particularly those within the rewarded interval. The projections from neurons in V_{ctx} to striatum which are selective to the rewarded direction strengthen. Because the projections along the direct pathway are functionally specific, GPi, Thalamic and M_{ctx} neurons which are representing the rewarded direction become inhibited, disinhibited and activated, respectively. This is shown for GPi and M_{ctx} neurons in Fig 2D&E (green line). As a result, the aggregate input from the BG-Thalamic circuit into the M_{ctx} becomes more biased as learning progresses. Thus, the probability that the hand movement ends in the rewarded interval increases and as a consequence the fraction of rewarded trials becomes larger (Fig 2G).

The learning curve averaged over twenty repetitions of the protocol (Fig. 2G&H) indicates that at the end of the learning the movement has adapted to the new condition. This depends crucially on the fact that after learning the input from BG-thalamic circuit to M_{ctx} is biased toward the desired movement direction. Inhibition of the GPi will suppress this bias and thus lead to a loss of the adaptation (Fig 2F-H).

Strong cortico-BG motor loops support efficient adaptation. As explained above, during adaptation, changes in the input from V_{ctx} into striatal neurons modifies the pattern of movement-related activity in the GPi. As a result, the feedback input from the direct pathway into M_{ctx}, biases the population vector of the cortical neurons toward the desired movement direction. How does this depend on the strength of the projections in the direct loop? How the efficiency of the learning process depends on these parameters?

To answer these questions, we simulated the model for different strength of the cortico-striatal projections. The learning dynamics slow down as this strength decreases (Fig 3C). Moreover, on average, the rotation angle of the movement direction becomes smaller (Fig 3D). For too weak cortico-striatal projections the probability of success during learning is extremely small and thus the direction of the movement has hardly changed at the end of a block of trials. Therefore, the feedback loop dynamics of the BG-thalamo-cortical loop appears essential for correct adaptation of the movement.

To investigate further the role of the BG-cortical loop in the adaptation process we studied the dynamics of the network when there are no recurrent interactions in M_{ctx} . We fixed all parameters but the M_{ctx}-striatal strength (see Fig. 3E-F). In this condition, the BG-cortical loop displays three different dynamical regimes. For intermediate values of the cortico-striatal gain the homogeneous activity state of the network is stable. In particular, the activity in M_{ctx} is homogeneous. When this parameter is increased, this state loses stability, and a symmetry breaking bifurcation occurs (Fig. 3F, bottom). As a result, the activity in M_{ctx}, is heterogeneous: It has a 'bump' shape. Interestingly, this bifurcation does not depend on the negative feedback mediated by the hyperdirect and indirect pathways (Fig. 3E). Finally, for sufficiently small cortico-striatal gain, the network undergoes a

bifurcation toward a state in which the activity oscillates in all the populations. These oscillations are driven by the hyperdirect feedback loop whose global gain is negative(26).

We then explored how recurrent dynamics in the M_{ctx} network cooperate with the BG-thalamo-cortical dynamics to shape the movement-related activity in M_{ctx}. Figure 3A shows how this profile is affected as the strength of the M_{ctx}-striatal projections varies. The activity bump in M_{ctx} is broad when these projections are weak, and it sharpens with their strength. For strong enough M_{ctx}-striatal projections the intrinsic dynamics of the BG-cortical loop bifurcate to the bump state. This results to even more sharpening of the bump in M_{ctx} and a strong amplification of the activity. This effect of the M_{ctx}-striatal projections is robust to changes in other parameters of the model. This is shown in Fig 3B for the cortico-subthalamic synaptic weights.

The role of the basal ganglia (BG) in motor adaptation can be understood as follows. Reaching movements are generated by the attractor dynamics in M_{ctx}without requiring direct involvement of the BG. Before adaptation, input from V_{ctx} to M_{ctx} influences the attractor dynamics in M_{ctx}, causing its population vector to be biased toward the cue. This bias is also reflected in the inputs from M_{ctx} and V_{ctx} into the striatum, which then propagate through the BG and thalamus back to M_{ctx}. During adaptation, the direct input from V_{ctx} to M_{ctx} remains biased toward the cue. However, reward-driven plasticity in the projections from V_{ctx} to the striatum results in inputs to the BG that gradually shift toward the target. The actual direction of movement is thus determined by the interaction between these two conflicting inputs. Specifically, the activity bump in M_{ctx} is shaped by the balance between the bias from V_{ctx} and the bias from the BG-thalamo-cortical pathway. When the gain of the BG-thalamo-cortical pathway is weak, its influence on M_{ctx} is minimal, and the activity remains biased toward the cue (Fig. 3D). Conversely, when the gain is strong, the BG input significantly shifts the bump in M_{ctx}. As a result, with learning, the population vector in M_{ctx} becomes more aligned with the target direction (Fig. 3C-D).

Shaping adaptation in the low-noise regime. In our model, neural noise is essential for driving exploration toward the rewarded direction. The exploration and thus the learning rate are highly dependent on the noise level. Specifically, the initial learning phase—measured as the time taken to reach the first reward—can be prolonged if noise is too weak, making rewarded trials rare (Fig. 4A-D). When noise is very low (ocher), learning progresses at an extremely slow pace, with fewer than 1% of movements reaching the reward region after 1,000 trials. As the noise level increases (red), the duration of the initial learning phase shortens, allowing the first reward to be achieved more quickly. Following this, the adaptation process accelerates, achieving an average performance of 80% after 700 trials. However, further increases in noise (blue) speed up learning at the cost of reducing the final performance. Excessive noise leads to imprecise movements, both before and after adaptation, with the noise amplitude remaining stable throughout the reward-driven learning process(44, 45). The balance between sufficient exploration and movement precision places a constraint on adaptation efficiency. This limitation can be addressed by employing a "shaping" protocol to enhance the adaptation process(43, 46, 47) (Fig. 4).

Behavioral experiments demonstrate that reaching movements are generally accurate at baseline(48, 49), implying low noise levels and slow adaptation. To speed up learning, one approach would be to progressively increase noise levels, though this is challenging to achieve experimentally. An alternative is to shape the behavior by gradual increase of the task difficulty(47).

In a shaping protocol, the target direction can be adjusted based on current performance (Fig. 4E-H). Initially, the rewarded area is set near the cue direction (15° ± 10°), leading to frequent rewards. Reward-dependent cortico-striatal plasticity then shifts movements closer to the desired target. As the target direction is gradually updated, movements continue to adjust toward the final goal (Fig. 4E-F). This method allows for faster adaptation, even with low noise, without compromising the accuracy of baseline reaching movements (Fig. 4G-H).

Basal ganglia drive *de-novo* visuo-motor association. The association between visual cues and movement direction is not innate(50). Instead, visuo-motor associations are learned during early childhood. More broadly, sensorimotor associations are developed through spontaneous exploration of possible movements, commonly known as "babbling" (30, 54). Here, we investigate the hypothesis that the basal ganglia (BG) are crucial for the initial acquisition of visuo-motor adaptations(51).

To test this hypothesis, we consider a "tabula-rasa" BG-thalamo-cortical network in which the projections from V_{ctx} to the striatum are initially non-specific (Fig. 5B, red), and projections from V_{ctx} to $M_{\mathrm{ctx}}$are absent. We assume the neurons in the sensory cortex are tuned to the location of visual cues and that projections from $M_{\mathrm{ctx}}$to the striatum are topographically organized. Since our primary focus is on how BG-thalamo-cortical loops facilitate the alignment between sensory representations in V_{ctx} and motor representations in the striatum, we do not include recurrent connectivity within M_{ctx}. As shown in Fig. 5, this alignment between sensory and motor representations emerges through the interaction of reward-driven cortico-striatal synaptic plasticity and the dynamics of these loops (See Material&Methods).

We begin by considering a scenario in which the same cue is presented repeatedly in a fixed direction. Initially, the projections from V_{ctx} to the striatum are non-specific, meaning that the sensory input received by striatal neurons during movement (as described in Fig. 1) does not favor any direction. Meanwhile, the dynamics of BG-thalamo-cortical loops generate a bump of activity in the $M_{\mathrm{ctx}}$with the location of this bump determined by ongoing synaptic noise. As a result, the movement fluctuates randomly across all possible directions over many trials (Figure 5A, top).

When movements occur in the direction of the presented cue (θ_{cue}) and fall within a rewarded range θ_{cue} ± 20°, grey area in Fig. 5C), the projections from V_{ctx} neurons selective to the cue location to striatal neurons tuned to that movement direction become progressively stronger. This leads to the emergence of direction-specificity in the V_{ctx}-striatal projections (Fig. 5B, top). Due to the topography of striato-pallidal projections, neurons in GPi that are preferentially active during movement in direction θ_{cue} are inhibited. Since the cortico-BG-thalamo-cortical loops are closed, cortical neurons tuned to θ_{cue} increase their activity, while the activity of less-tuned neurons is suppressed. Consequently, the variability in movement direction decreases (Fig. 5C&D), increasing the likelihood that the movement falls within the rewarded region (Fig. 5E). Repeating this process over many trials in the cue direction progressively strengthens the Vctx-striatal projections in a manner specific to that direction (Fig. 5B).

We now consider the scenario in which multiple cues are presented randomly, each corresponding to one of 16 possible directions uniformly distributed between 0 and 360 degrees (Fig. 5F-H). Initially, the projections from V_{ctx} to the striatum are non-specific. In this early phase, movement directions are random and independent of cue direction. During the learning process, movements directed toward one of the 16 specified directions are rewarded. Consequently, the projections from V_{ctx} to the striatum are selectively strengthened according to Eq. (1). As learning progresses, the M_{ctx} -striatal connection matrix becomes more structured, promoting movements that align with the 16 cue directions (Fig. 5F). Gradually, movement directions converge with the cue directions for all 16 possibilities. After 2000 trials, the movements consistently fall within θ_{cue}± 5° for most attempts (Fig. 5F), leading to an average reward rate of over 80% across 20 runs (Fig. 5G-H). The resulting synaptic weight matrix between M_{ctx} and the striatum reveals a clear peak, indicating alignment between pre-synaptic cue direction in M_{ctx} and post-synaptic movement direction in the striatum (Fig. 5I).

Discussion

Inhibiting the output of the basal ganglia (BG) impacts the vigor of reaching movements without affecting their direction(8). However, the BG play a vital role in the reward-based adaptation of reaching movement direction(41). To investigate the underlying mechanisms, we created a simplified dynamical model of the BG-thalamo-cortical network that incorporates these aspects. Our key conclusion is that the efficiency of adaptation is significantly influenced by the dynamical regime in which the BG network operates.

Sensorimotor adaptation occurs through two primary modes. Error-based learning and reward-driven learning(48, 52). Error-based adaptation involves utilizing sensory feedback to correct predictable perturbations on a trial-by-trial basis(53–55) and is likely mediated by cerebellar circuits(56, 57). When sensory feedback is available, error-based corrections tend to dominate the adaptation of reaching movements(58–60).In contrast, reward-driven adaptation relies on a binary evaluation of action outcomes. This form of adaptation has been studied in humans performing motor tasks where outcomes are rewarded based on rules known only to the experimenter. Reward-driven adaptation can pertain to various movement features, such as curvature(44) or direction56, 67, 68). Notably, reward-based sensorimotor adaptation remains intact in individuals with cerebellar damage(43) but requires functioning basal ganglia (BG) circuitry(12, 13, 16, 61). Furthermore, dopaminergic input to the BG is both necessary and sufficient for facilitating rapid reinforcement-driven behavioral corrections(15, 62, 63).

The mechanisms by which BG circuitry and its dynamics contribute to reward-based adaptation are still largely unclear. Our work seeks to address this issue using an abstract model of the cortex-BG network.

Comparison with other models. The dorsal basal ganglia (BG) have been proposed to implement reinforcement learning (RL) mechanisms(29, 64). Standard RL algorithms, which are effective for learning discrete tasks (e.g., choosing between two options), have been examined in models where actions are represented by discrete output populations of the BG(29, 65). However, developing a mechanistic model for learning simple continuous movements remains a significant challenge(66–68).

Previous theories regarding song reinforcement learning(69, 70) suggested that reward-modulated plasticity occurs at cortico-cortical synapses, with an external experimenter inducing perturbations in the motor output network. In contrast, it has been found that dopamine-dependent plasticity primarily takes place in cortico-striatal connections, which subsequently drive BG-dependent motor corrections within cortical networks(16, 71). Tesileanu et al.(72) proposed a mechanism whereby a cortico-cortical network learns from another cortical area acting as a "tutor," which has been trained through reinforcement learning.

The mechanism we propose for the role of the basal ganglia (BG)-cortical network in reward-based sensorimotor adaptation differs fundamentally from previous theories(29, 71). It is based on the emergence of attractors within the recurrent dynamics of the BG-cortex circuit. In our model, these attractors arise from feedback loops that topographically connect the cortex to itself via the basal ganglia and thalamus, as shown in anatomical studies(22, 24, 25) and more recent physiological experiments(24). These experiments revealed that specific subsets of cortical information are transmitted through the cortico-basal ganglia-thalamic network back to the original cortico-striatal neurons of each subnetwork, thereby forming a bona fide closed loop(24). A similar integration of reward-driven learning and attractor dynamics has been proposed in the context of operant conditioning, where subjects choose from discrete options(73). This combination has also been modeled in the hippocampal system for spatial navigation(74), where the activation patterns of place cells are influenced by local attractor dynamics, and rewards can facilitate the learning of preferred locations within an environment.

In our model, reaching movements are driven by the recurrent cortical attractor(75–77) (Ruti et al., 1995;. These movements are represented by the population vector of neuronal activity in the motor cortex, where the vigor and direction of the movement correspond to the norm and direction of this vector. Alternative models that do not assume functional organization have been proposed to explain the changes in the direction of the vector representing motor cortex activity in neural space from preparation to execution of a movement(78). However, recent work by Brunel and Hatsopoulos(33) demonstrated that such changes can also be accounted for in models with functional organization.

Our model intentionally omits recurrent connectivity in the BG nuclei for the sake of simplicity. While recurrent connectivity is relatively weak in the subthalamic nucleus (STN) and thalamus, it is more pronounced in other BG nuclei, such as the external globus pallidus (GPe)(20). Understanding the functional implications of this dense recurrent connectivity would necessitate further theoretical investigation.

Generality of the results. While our model is grounded in the anatomy of the basal ganglia (BG) and reaching-related activity in primates(34), the principles it reveals may broadly apply to BG-cortical circuits in vertebrates. This generalizability stems from two key features: (i) topographically organized closed-loop BG-cortical circuits, which have been observed in rodents, songbirds, and non-human primates (NHPs)(22, 24, 25); and (ii) the competition between pathways of opposing polarity within the BG, a feature present in species as diverse as lamprey, songbirds, rodents, and NHPs(79–81), with a remarkably well-conserved functional connectivity across species(21, 82, 83). This suggests that similar network dynamics may underpin movement production and adaptation across a variety of vertebrate species.

In our model, the topographic projections to the motor cortex from a population of neurons selective to the direction of the visual cue play a crucial role in aligning the sensory and motor representations of movement direction. This population could correspond, for example, to a region in the parietal or temporal cortex(37, 38). We demonstrate that this alignment can emerge de novo through the same plasticity rules when paired with large-scale exploratory behavior, such as motor babbling (Fig. 5), similar to vocal babbling observed in infants and songbirds(50, 84).

Predictions and perspectives. Our theory explains experimental findings and generates specific predictions in two areas: reaching movements in primates and vocal production in birds. In non-human primates (NHP), inactivating the globus pallidus internus (GPi) reduces the vigor of reaching movements without impairing movement execution(8). In birds, while song production is not dependent on basal ganglia (BG) integrity, sensorimotor learning and adaptation are(85, 86). Our model accounts for both phenomena.

In our framework, reward-driven adaptation induces synaptic changes in cortico-striatal projections but does not alter recurrent connectivity in the motor cortex. Therefore, we predict that if the output of the BG to the motor cortex is suppressed shortly after learning, performance will revert to baseline. This could be tested in NHPs trained without visual feedback, guided only by external rewards (*e.g.*, as in (87)), by pharmacologically inhibiting the GPi (as shown in Fig 2C).

All models of reward-based sensorimotor adaptation involving the BG predict that inhibiting the GPi will disrupt adaptation. A key prediction of our model is that adaptation will also be impaired by weakening the projections from the motor cortex to the striatum. This hypothesis could be tested in NHPs by expressing chemogenetic receptors in motor cortex projection neurons(88), and locally activating these receptors in the dorsal striatum.This would reduce synaptic release by cortico-striatal axons in the striatum without affecting other movement control centers. We predict that such a manipulation would impair adaptation. Our theory can also be tested in the context of pitch adaptation in songbirds(16). Here, chemogenetic methods could be used to reduce the strength of

projections from the Lateral Magnocellular Anterior Nucleus, a cortical region, to area X, the songbird equivalent of the BG (striatum, globus pallidus). Our model predicts that this reduction would impair pitch adaptation.

Materials and Methods

Architecture of the model. The model is grounded in the anatomy of the basal ganglia (BG)-thalamo-cortical network, particularly emphasizing the topographic organization of feedback loops with opposing polarities (as demonstrated by Foster, Sabatini, Luo, and Strick). It includes the divergent projections from the subthalamic nucleus (STN) to the globus pallidus (GPi, GPe), as shown in studies by many anatomical studies(89, 90).

We assume that in the motor cortex, nearby neurons typically share similar preferred directions for arm movements(77, 91, 92). This functional organization is thought to extend to the BG nuclei and the thalamus through topographic connectivity. Our model accounts for directional tuning across the BG-thalamo-cortical network(39, 40, 93).

In terms of learning, the model incorporates a reward-modulated tri-partite learning rule(94, 95) that aligns with dopamine's role in signaling reward(64) and modulating cortico-striatal plasticity(62). Moreover, the convergence of motor and sensory cortical inputs related to the same task or movement in the striatum, supported by anatomical and physiological studies(35, 36, 96), is central to our theory.

Anatomical structure of the BG-thalamo-cortical network model. Our model of the basal ganglia (BG) network consists of six neuron populations: three excitatory and three inhibitory. The excitatory populations include the motor cortex (Ctx), the ventral anterior and ventral lateral nuclei of the thalamus (Thal), and the subthalamic nucleus (STN). The inhibitory populations correspond to the sensory–motor regions of the striatum (Str) and both the external (GPe) and internal (GPi) segments of the globus pallidus.

The network consists of parallel, partially segregated feedback loops, a configuration that is supported by recent experimental studies in non-human primates (NHP), birds, and rodents (references 22, 24, 25). Anatomical and electrophysiological studies have shown that the direct pathway is organized topographically, with information flow maintaining functional and somatotopic segregation throughout the pathway (references 22, 24). In line with these findings, the connectivity patterns in our model reflect this topographic structure, as demonstrated in Figure 1.

The architecture of the circuit is depicted in Figure 1A. The motor cortex is represented by a population of neurons. Each neuron is assigned a preferred direction (PD) between 0° and 360° (Fig 1B). The strength of the connection between two neurons depends on the difference between their PD.

The various BG nuclei and the BG-receiving motor thalamus are also modeled as populations of as rate units (Fig 1B). The architecture of the network connecting these various populations reflects the major anatomical connections of the BG-thalamo-cortical loop(6, 20). As a result, neurons in each population of the network can be associated with a PD corresponding to the PD of the cortical neuron they are in relation with. Indeed, cortical neurons with 0° preferred direction preferentially project to striatal and STN neurons also located at 0°, which themselves project to pallidal neurons located at 0°. This organization results in a reflection of the motor cortex population code, with movement-related activity tuned to movement direction in all populations of the network (Fig 1C) consistent with activity patterns observed experimentally(39, 40). While mainly segregated along the direct/indirect pathways through the BG and in their thalamo-cortical feedback, the parallel and

partially segregated feedback loops of the BG interact through the divergent projection from the STN to the GPe and GPi(20, 89). This is reflected in our model by a broader connection profile from the STN to the GPe and GPi compared to other connections of the network (Fig 1A).

In a visuomotor association task, such as the one modeled in our study, the convergence of sensory information—specifically the visual cue's position—into the motor BG-thalamo-cortical network is essential for instructing movement direction. Anatomical and physiological evidence supports that motor and sensory cortical inputs related to the same task or movement converge in the striatum (46, 47). In our model, we incorporate this feature by considering the output of a sensory cortical population that encodes the position of the visual stimulus and projects to the striatal population within the motor BG-thalamo-cortical network (Fig 1A). This sensory-to-motor projection ensures that the visual information is effectively integrated to guide the movement direction.

Single neuron dynamics. Neurons are modeled as rate units (97, 98). The dynamics of neuron is defined as follows. We denote by $I_\alpha(\theta, t)$ the total input to the neuron in population $\alpha \in$ {Ctx, Thalamus, Striatum, STN, GPe, GPi} and PD, θ at time t. The activity of the neuron at time t , $A_\alpha(\theta, t)$, is related to this input by

$$A_\alpha(\theta, t) = S_\alpha\big(I_\alpha(\theta, t)\big)$$

where is the non-linear transfer function neurons in population α the rectified linear function

$$S_\alpha(I) = [I - T_\alpha]_+$$

with $[x]_+ = x$ for $x > 0$ and 0 , T_α is a threshold which depends only of the population.

We denote by $m_{\alpha,\beta}(\theta, \theta', t)$ the input received by neuron α, θ from neuron β, θ' whose dynamics is (Ermentrout, 1996; Shriki et al, 2003)

$$\tau_{\alpha\beta} \frac{dm_{\alpha\beta}(\theta)}{dt} = -m_{\alpha\beta}(\theta) + A_\beta(\theta)$$

Here $\tau_{\beta\alpha}$ is the time constant of the dynamics. We will assume that the time constant is the same for all the populations, $\tau_{\beta\alpha} = \tau$, except for $(\alpha, \beta) =$ (Ctx, STN) to account account for the presence of NMDA receptors on STN neurons.

Connectivity between neuronal populations. The connectivity between two connected populations in the circuit is topographically organized, connecting preferentially neurons associated to the same preferred direction in each population.

The strength of the connection between two neurons of these two populations depends on the "angular" distance (|θ - θ'| [2π] between these two neurons, θ and θ' being their respective preferred directions. We consider the following profile for the connection strengths between neurons (i, j) of populations α and β, whose preferred directions are and θ and θ':

$$G_{ij}^{\alpha\beta} = G^{\alpha\beta}(\theta - \theta') = \pm \frac{G_{\alpha\beta}^0}{\sqrt{\pi}\sigma_{\alpha\beta}} \exp\left(-\frac{(|\theta - \theta'|[2\pi])^2}{\sigma_{\alpha\beta}^2}\right)$$

Where σαβ denotes the width (or footprint) of the divergent connections from one neuron to the post-synaptic population and G0αβ the global gain of the connection. For the sake of simplicity, we do not introduce any lateral interactions in the various populations of the model.

Total inputs to the neurons. The total synaptic inputs received by neuron in population $\alpha \in$ {Ctx, Thalamus, Striatum, STN, GPe, GPi} with PD θ at time t, are given by the following:

$$I_{Ctx}(\theta,t) = \int_{\theta'=-\pi}^{\pi} G_{CtxTh}(\theta-\theta')m_{CtxTh}(\theta',t-\Delta_{CtxTh})d\theta + H_{Ctx}(\theta,t) + \eta_{Ctx}(\theta,t)$$

$$I_{Str}(\theta,t) = \int_{\theta'=-\pi}^{\pi} G_{StrCtx}(\theta-\theta')m_{StrCtx}(\theta',t-\Delta_{StrCtx})d\theta + H_{Str}(\theta,t) + \eta_{Str}(\theta,t)$$

$$I_{NST}(\theta,t) = \int_{\theta'=-\pi}^{\pi} G_{NSTCtx}(\theta-\theta')m_{NSTCtx}(\theta',t-\Delta_{NSTCtx})d\theta + \eta_{NST}(\theta,t)$$

$$I_{GPi}(\theta,t) = \int_{\theta'=-\pi}^{\pi} G_{GPiNST}(\theta-\theta')m_{GPiNST}(\theta',t-\Delta_{GPiNST})d\theta - \int_{\theta'=-\pi}^{\pi} G_{GPiStr}(\theta-\theta')m_{GPiStr}(\theta',t-\Delta_{GPiStr})d\theta + \eta_{GPi}(\theta,t)$$

$$I_{Th}(\theta,t) = -\int_{\theta'=-\pi}^{\pi} G_{ThGPi}(\theta-\theta')m_{ThGPi}(\theta',t-\Delta_{ThGPi})d\theta + \eta_{Th}(\theta,t)$$

where Δαβ denotes the synaptic delay from population α to population β. The last term in these equations, represents additional Gaussian white noisy inputs with a zero mean and SD as in Table 1.

Movement execution. Neuron activation in the motor cortex directly drives movement. In our model, the instantaneous movement direction and speed are determined by a population vector. This vector is calculated as the sum of individual neurons' vectors, each pointing toward its preferred direction, with the amplitude proportional to its activity above 10 spikes per second (sp/s). The 10 sp/s threshold ensures no movement occurs during typical cortical rest activity, which averages around 5 sp/s. A movement is triggered as soon as any neuron exceeds the 10 sp/s activity level.

We assume that during the planning and execution of movements, neurons in both cortical and striatal populations receive additional external inputs. These inputs represent synaptic signals arriving at cortical and striatal neurons from other cortical areas and brain regions that are not explicitly included in the model. For simplicity, the inputs to the cortical populations are considered homogeneous across the entire population. In both the text and figures, external inputs are expressed in arbitrary units.

The inputs to cortical neurons. The two cortical populations in our model represent neurons in the motor cortex that encode two different motor programs. Since other brain regions that send inputs to the motor cortex are not explicitly modeled, we account for these by introducing an external input, $I_{\text{mov}}(t)$ to the system dynamics. We assume that the temporal profile of this external input takes the following form: it is active during $t_{\text{m}} - \frac{D_{\text{mvt}}}{2} \leq t \leq t_{\text{m}} + \frac{D_{\text{mvt}}}{2}\frac{D_{\text{mvt}}}{2}$ and zero otherwise. Here, D_{mvt} is the total duration of the input associated with the movement, and t_{m} is the time when the input reaches its peak value.

Inputs to Striatal Neurons. During motor planning, information from the sensorimotor cortical area, related to the motor program to be executed, is sent to the sensory–motor striatum. This sensory information is represented in the model by adding a transient, weakly selective external input to the striatum. This input begins at the same time as the cortical input, but its duration (D_{Str})

is much shorter than the duration of the cortical input (D_{mvt}). The input follows a specific profile defined within the model.

Integration method. The model's dynamics were simulated using a standard first-order Euler algorithm with a fixed time step of 0.5 ms. To ensure the accuracy of the simulations, additional runs were performed using a smaller time step (0.05 ms), confirming that the dynamics remained consistent and unchanged.

Acknowledgments
AL thanks Alexis Dubreuil, Nicolas Mallet and Marc Deffains for thoughtful discussions. DH expresses gratitude to Carole Levenes, Yonatan Loewenstein, and the late Carl van Vreeswijk for their insightful discussions. He also thanks the Edmond and Lily Safra Brain Research Center, and the Loewenstein's lab at The Hebrew University of Jerusalem, Israe,l for their warm hospitality. This work was supported by funding from the Agence National pour la Recherche (SleepBrainDyn, 20-CE37-0023-02), Fonds pour la Recherche Médicale (EQU202203014688) to AL, and the France Israel Center for Neural Computation (FICNC, HUJI/CNRS) to DH.

References

1. E. V. Evarts, Relation of pyramidal tract activity to force exerted during voluntary movement. *J Neurophysiol* **31**, 14–27 (1968).

2. E. E. Fetz, Cortical mechanisms controlling limb movement. *Curr Opin Neurobiol* **3**, 932–939 (1993).

3. T. C. Harrison, O. G. S. Ayling, T. H. Murphy, Distinct cortical circuit mechanisms for complex forelimb movement and motor map topography. *Neuron* **74**, 397–409 (2012).

4. M. S. A. Graziano, C. S. R. Taylor, T. Moore, D. F. Cooke, The cortical control of movement revisited. *Neuron* **36**, 349–362 (2002).

5. I. Q. Whishaw, Loss of the innate cortical engram for action patterns used in skilled reaching and the development of behavioral compensation following motor cortex lesions in the rat. *Neuropharmacology* **39**, 788–805 (2000).

6. J. T. Dudman, J. W. Krakauer, The basal ganglia: from motor commands to the control of vigor. *Current Opinion in Neurobiology* **37**, 158–166 (2016).

7. J. M. Deniau, G. Chevalier, Disinhibition as a basic process in the expression of striatal functions. II. The striato-nigral influence on thalamocortical cells of the ventromedial thalamic nucleus. *Brain Res* **334**, 227–233 (1985).

8. M. Desmurget, R. S. Turner, Motor Sequences and the Basal Ganglia: Kinematics, Not Habits. *J. Neurosci.* **30**, 7685–7690 (2010).

9. C. Piron, *et al.*, The globus pallidus pars interna in goal-oriented and routine behaviors: Resolving a long-standing paradox. *Mov Disord* **31**, 1146–1154 (2016).

10. R. Hayashi, T. Hashimoto, T. Tada, S. Ikeda, Effects of unilateral pallidotomy on voluntary movement, and simple and choice reaction times in Parkinson's disease. *Mov Disord* **18**, 515–523 (2003).

11. J. A. Obeso, *et al.*, What can man do without basal ganglia motor output? The effect of combined unilateral subthalamotomy and pallidotomy in a patient with Parkinson's disease. *Exp Neurol* **220**, 283–292 (2009).

12. O. Hikosaka, K. Nakamura, K. Sakai, H. Nakahara, Central mechanisms of motor skill learning. *Current Opinion in Neurobiology* **12**, 217–222 (2002).

13. A. Leblois, D. J. Perkel, "The Song Circuit as a Model of Basal Ganglia Function" in *The Neuroethology of Birdsong*, Springer Handbook of Auditory Research., J. T. Sakata, S. C. Woolley, R. R. Fay, A. N. Popper, Eds. (Springer International Publishing, 2020), pp. 93–125.

14. R. J. Beninger, The role of dopamine in locomotor activity and learning. *Brain Res* **287**, 173–196 (1983).

15. L. Xiao, *et al.*, A Basal Ganglia Circuit Sufficient to Guide Birdsong Learning. *Neuron* **98**, 208-221.e5 (2018).

16. A. S. Andalman, M. S. Fee, A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12518–12523 (2009).

17. J. A. Kleim, *et al.*, Cortical synaptogenesis and motor map reorganization occur during late, but not early, phase of motor skill learning. *J Neurosci* **24**, 628–633 (2004).

18. Y. Matsuzaka, N. Picard, P. L. Strick, Skill representation in the primary motor cortex after long-term practice. *J Neurophysiol* **97**, 1819–1832 (2007).

19. R. J. Nudo, Mechanisms for recovery of motor function following cortical damage. *Curr Opin Neurobiol* **16**, 638–644 (2006).

20. A. Parent, L.-N. Hazrati, Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews* **20**, 91–127 (1995).

21. A. Nambu, H. Tokuno, M. Takada, Functional significance of the cortico–subthalamo–pallidal 'hyperdirect' pathway. *Neuroscience Research* **43**, 111–117 (2002).

22. R. M. Kelly, P. L. Strick, "Macro-architecture of basal ganglia loops with the cerebral cortex: use of rabies virus to reveal multisynaptic circuits" in *Progress in Brain Research*, (Elsevier, 2004), pp. 447–459.

23. The functional anatomy of basal ganglia disorders: Trends in Neurosciences. Available at: https://www.cell.com/trends/neurosciences/abstract/0166-2236(89)90074-X [Accessed 20 September 2024].

24. N. N. Foster, *et al.*, The mouse cortico–basal ganglia–thalamic network. *Nature* **598**, 188–194 (2021).

25. M. Luo, L. Ding, D. J. Perkel, An avian basal ganglia pathway essential for vocal learning forms a closed topographic loop. *J Neurosci* **21**, 6836–6845 (2001).

26. A. Leblois, T. Boraud, W. Meissner, H. Bergman, D. Hansel, Competition between Feedback Loops Underlies Normal and Pathological Dynamics in the Basal Ganglia. *J. Neurosci.* **26**, 3567–3583 (2006).

27. M. Guthrie, A. Leblois, A. Garenne, T. Boraud, Interaction between cognitive and motor cortico-basal ganglia loops during decision making: a computational study. *Journal of Neurophysiology* **109**, 3025–3040 (2013).

28. M. Topalidou, A. Leblois, T. Boraud, N. P. Rougier, A long journey into reproducible computational neuroscience. *Front. Comput. Neurosci.* **9** (2015).

29. K. Doya, Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology* **10**, 732–739 (2000).

30. J. R. Wickens, Synaptic plasticity in the basal ganglia. *Behavioural Brain Research* **199**, 119–128 (2009).

31. Fino, Spike-timing dependent plasticity in the striatum. *Front.Syna.Neurosci.* (2010). https://doi.org/10.3389/fnsyn.2010.00006.

32. A. B. Nelson, A. C. Kreitzer, Reassessing Models of Basal Ganglia Function and Dysfunction. *Annu. Rev. Neurosci.* **37**, 117–135 (2014).

33. L. Bachschmid-Romano, N. G. Hatsopoulos, N. Brunel, Interplay between external inputs and recurrent dynamics during movement preparation and execution in a network model of motor cortex. *eLife* **12**, e77690 (2023).

34. A. Georgopoulos, M. DeLong, M. Crutcher, Relations between parameters of step-tracking movements and single cell discharge in the globus pallidus and subthalamic nucleus of the behaving monkey. *J. Neurosci.* **3**, 1586–1598 (1983).

35. A. W. Flaherty, A. M. Graybiel, Two input systems for body representations in the primate striatal matrix: experimental evidence in the squirrel monkey. *J Neurosci* **13**, 1120–1137 (1993).

36. S. Charpier, M. Pidoux, S. Mahon, Converging sensory and motor cortical inputs onto the same striatal neurons: An in vivo intracellular investigation. *PLoS One* **15**, e0228260 (2020).

37. J. P. Gottlieb, M. Kusunoki, M. E. Goldberg, The representation of visual salience in monkey parietal cortex. *Nature* **391**, 481–484 (1998).

38. X. Y, Parietal-driven visual working memory representation in occipito-temporal cortex. *Current biology : CB* **33** (2023).

39. R. S. Turner, M. E. Anderson, Pallidal Discharge Related to the Kinematics of Reaching Movements in Two Dimensions. *Journal of Neurophysiology* **77**, 1051–1074 (1997).

40. B. Pasquereau, *et al.*, Shaping of Motor Responses by Incentive Values through the Basal Ganglia. *J. Neurosci.* **27**, 1176–1183 (2007).

41. J. Izawa, R. Shadmehr, Learning from Sensory and Reward Prediction Errors during Motor Adaptation. *PLoS Comput Biol* **7**, e1002012 (2011).

42. O. Codol, P. J. Holland, J. M. Galea, The relationship between reinforcement and explicit control during visuomotor adaptation. *Sci Rep* **8**, 9121 (2018).

43. A. S. Therrien, D. M. Wolpert, A. J. Bastian, Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain* **139**, 101–114 (2016).

44. H. G. Wu, Y. R. Miyamoto, L. N. G. Castro, B. P. Ölveczky, M. A. Smith, Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat Neurosci* **17**, 312–321 (2014).

45. S. J. Sober, M. J. Wohlgemuth, M. S. Brainard, Central Contributions to Acoustic Variation in Birdsong. *J. Neurosci.* **28**, 10370–10379 (2008).

46. E. C. Tumer, M. S. Brainard, Performance variability enables adaptive plasticity of “crystallized” adult birdsong. *Nature* **450**, 1240–1244 (2007).

47. J. S. Skinner, Commentaries. *Can Med Assoc J* **96**, 794 (1967).

48. A. M. Haith, J. W. Krakauer, Motor learning: the great rate debate. *Curr Biol* **24**, R386-388 (2014).

49. Y. Mandelblat-Cerf, *et al.*, The neuronal basis of long-term sensorimotor learning. *J Neurosci* **31**, 300–313 (2011).

50. M. Hadders-Algra, Early human motor development: From variation to the ability to vary and adapt. *Neuroscience & Biobehavioral Reviews* **90**, 411–427 (2018).

51. F. Hadj-Bouziane, M. Meunier, D. Boussaoud, Conditional visuo-motor learning in primates: a key role for the basal ganglia. *J Physiol Paris* **97**, 567–579 (2003).

52. J. A. Taylor, J. W. Krakauer, R. B. Ivry, Explicit and implicit contributions to learning in a sensorimotor adaptation task. *J Neurosci* **34**, 3023–3032 (2014).

53. H. V. Helmholtz, *Treatise on physiological optics*, JPC Southall, Trans. (1962).

54. R. Shadmehr, F. A. Mussa-Ivaldi, Adaptive representation of dynamics during learning of a motor task. *J Neurosci* **14**, 3208–3224 (1994).

55. J. W. Krakauer, Z. M. Pine, M.-F. Ghilardi, C. Ghez, Learning of Visuomotor Transformations for Vectorial Planning of Reaching Trajectories. *J. Neurosci.* **20**, 8916–8924 (2000).

56. M. J. Weiner, M. Hallett, H. H. Funkenstein, Adaptation to lateral displacement of vision in patients with lesions of the central nervous system. *Neurology* **33**, 766–772 (1983).

57. M. A. Smith, R. Shadmehr, Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *J Neurophysiol* **93**, 2809–2821 (2005).

58. J. R. Morehead, J. A. Taylor, D. E. Parvin, R. B. Ivry, Characteristics of Implicit Sensorimotor Adaptation Revealed by Task-irrelevant Clamped Feedback. *J Cogn Neurosci* **29**, 1061–1074 (2017).

59. J. W. Krakauer, A. M. Hadjiosif, J. Xu, A. L. Wong, A. M. Haith, Motor Learning. *Compr Physiol* **9**, 613–663 (2019).

60. K. M. Bond, J. A. Taylor, Flexible explicit but rigid implicit learning in a visuomotor adaptation task. *J Neurophysiol* **113**, 3836–3849 (2015).

61. T. L. Warren, E. C. Tumer, J. D. Charlesworth, M. S. Brainard, Mechanisms and time course of vocal learning and consolidation in the adult songbird. *Journal of Neurophysiology* **106**, 1806–1821 (2011).

62. J. N. J. Reynolds, B. I. Hyland, J. R. Wickens, A cellular mechanism of reward-related learning. *Nature* **413**, 67–70 (2001).

63. L. A. Hoffmann, V. Saravanan, A. N. Wood, L. He, S. J. Sober, Dopaminergic Contributions to Vocal Learning. *J. Neurosci.* **36**, 2176–2189 (2016).

64. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).

65. N. D. Daw, K. Doya, The computational neurobiology of learning and reward. *Current Opinion in Neurobiology* **16**, 199–204 (2006).

66. N. Bernstein - The Coordination And Regulation Of Movements : Free Download, Borrow, and Streaming : Internet Archive. Available at: https://archive.org/details/bernsteinthecoordinationandregulationofmovements [Accessed 20 September 2024].

67. A. K. Dhawale, S. B. E. Wolff, R. Ko, B. P. Ölveczky, The basal ganglia control the detailed kinematics of learned motor skills. *Nat Neurosci* **24**, 1256–1269 (2021).

68. R. Parr, S. Russell, Reinforcement Learning with Hierarchies of Machines in *Advances in Neural Information Processing Systems*, (MIT Press, 1997).

69. I. R. Fiete, M. S. Fee, H. S. Seung, Model of Birdsong Learning Based on Gradient Estimation by Dynamic Perturbation of Neural Conductances. *Journal of Neurophysiology* **98**, 2038–2057 (2007).

70. M. A. Farries, A. L. Fairhall, Reinforcement Learning With Modulated Spike Timing–Dependent Synaptic Plasticity. *Journal of Neurophysiology* **98**, 3648–3665 (2007).

71. M. S. Fee, J. H. Goldberg, A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* **198**, 152–170 (2011).

72. T. Teşileanu, B. Ölveczky, V. Balasubramanian, Rules and mechanisms for efficient two-stage learning in neural circuits. *eLife* **6**, e20944 (2017).

73. Covariance-based synaptic plasticity in an attractor network model accounts for fast adaptation in free operant learning - PubMed. Available at: https://pubmed-ncbi-nlm-nih-gov.insb.bib.cnrs.fr/23345226/ [Accessed 20 September 2024].

74. R. Cazé, M. Khamassi, L. Aubin, B. Girard, Hippocampal replays under the scrutiny of reinforcement learning models. *J Neurophysiol* **120**, 2877–2896 (2018).

75. D. Hansel, H. Sompolinsky, 13 Modeling Feature Selectivity in Local Cortical Circuits. 69.

76. R. Darshan, A. Rivkind, Learning to represent continuous variables in heterogeneous neural networks. *Cell Reports* **39**, 110612 (2022).

77. B. Amirikian, A. P. Georgopoulos, Modular organization of directionally tuned cells in the motor cortex: Is there a short-range order? *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12474–12479 (2003).

78. M. M. Churchland, B. M. Yu, S. I. Ryu, G. Santhanam, K. V. Shenoy, Neural Variability in Premotor Cortex Provides a Signature of Motor Preparation. *J. Neurosci.* **26**, 3697–3712 (2006).

79. S. Grillner, J. Hellgren, A. Menard, K. Saitoh, M. Wikstrom, Mechanisms for selection of basic motor programs – roles for the striatum and pallidum. *Trends in Neurosciences* **28**, 364–370 (2005).

80. A. Doupe, D. Perkel, A. Reiner, E. Stern, Birdbrains could teach basal ganglia research a new song. *Trends in Neurosciences* **28**, 353–363 (2005).

81. T. Boraud, A. Leblois, N. P. Rougier, A natural history of skills. *Progress in Neurobiology* **171**, 114–124 (2018).

82. A. Leblois, Á. L. Bodor, A. L. Person, D. J. Perkel, Millisecond Timescale Disinhibition Mediates Fast Information Transmission through an Avian Basal Ganglia Loop. *J. Neurosci.* **29**, 15420–15433 (2009).

83. N. Maurice, J.-M. Deniau, J. Glowinski, A.-M. Thierry, Relationships between the Prefrontal Cortex and the Basal Ganglia in the Rat: Physiology of the Cortico-Nigral Circuits. *J. Neurosci.* **19**, 4674–4681 (1999).

84. R. Darshan, W. E. Wood, S. Peters, A. Leblois, D. Hansel, A canonical neural mechanism for behavioral variability. *Nat Commun* **8**, 15415 (2017).

85. S. W. Bottjer, E. A. Miesner, A. P. Arnold, Forebrain Lesions Disrupt Development But Not Maintenance of Song in Passerine Birds. *Science* **224**, 901–903 (1984).

86. F. Ali, *et al.*, The Basal Ganglia Is Necessary for Learning Spectral, but Not Temporal, Features of Birdsong. *Neuron* **80**, 494–506 (2013).

87. S. E. Pekny, J. Izawa, R. Shadmehr, Reward-dependent modulation of movement variability. *J Neurosci* **35**, 4015–4024 (2015).

88. M. Deffains, *et al.*, In vivo electrophysiological validation of DREADD-based modulation of pallidal neurons in the non-human primate. *Eur J Neurosci* **53**, 2192–2204 (2021).

89. B. P. Kolomiets, *et al.*, Segregation and Convergence of Information Flow through the Cortico-Subthalamic Pathways. *J. Neurosci.* **21**, 5764–5772 (2001).

90. J. W. Mink, The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology* **50**, 381–425 (1996).

91. E. Stark, I. Asher, M. Abeles, Encoding of Reach and Grasp by Single Neurons in Premotor Cortex Is Independent of Recording Site. *Journal of Neurophysiology* **97**, 3351–3364 (2007).

92. Y. Ben-Shaul, *et al.*, Dynamical Organization of Directional Tuning in the Primate Premotor and Primary Motor Cortex. *Journal of Neurophysiology* **89**, 1136–1142 (2003).

93. A. Georgopoulos, J. Kalaska, R. Caminiti, J. Massey, On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).

94. R. J. Williams, Simple statistical gradient-following algorithms for connectionist reinforcement learning.

95. Y. Loewenstein, H. S. Seung, Operant matching is a generic outcome of synaptic plasticity based on the covariance between reward and neural activity. *Proceedings of the National Academy of Sciences* **103**, 15224–15229 (2006).

96. S. Ramanathan, J. J. Hanley, J.-M. Deniau, J. P. Bolam, Synaptic Convergence of Motor and Somatosensory Cortical Afferents onto GABAergic Interneurons in the Rat Striatum. *J. Neurosci.* **22**, 8158–8169 (2002).

97. W. Hr, C. Jd, Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal* **12** (1972).

98. O. Shriki, D. Hansel, H. Sompolinsky, Rate Models for Conductance-Based Cortical Neuronal Networks. *Neural Computation* **15**, 1809–1841 (2003).

99. D. J. Pinto, J. C. Brumberg, D. J. Simons, G. B. Ermentrout, A quantitative population model of whisker barrels: re-examining the Wilson-Cowan equations. *J Comput Neurosci* **3**, 247–264 (1996).

100. T. W. Gardiner, R. J. Nelson, Striatal neuronal activity during the initiation and execution of hand movements made in response to visual and vibratory cues. *Exp Brain Res* **92**, 15–26 (1992).

101. I. H. Lee, J. A. Assad, Putaminal Activity for Simple Reactions or Self-Timed Movements. *Journal of Neurophysiology* **89**, 2528–2537 (2003).

Figures and Tables

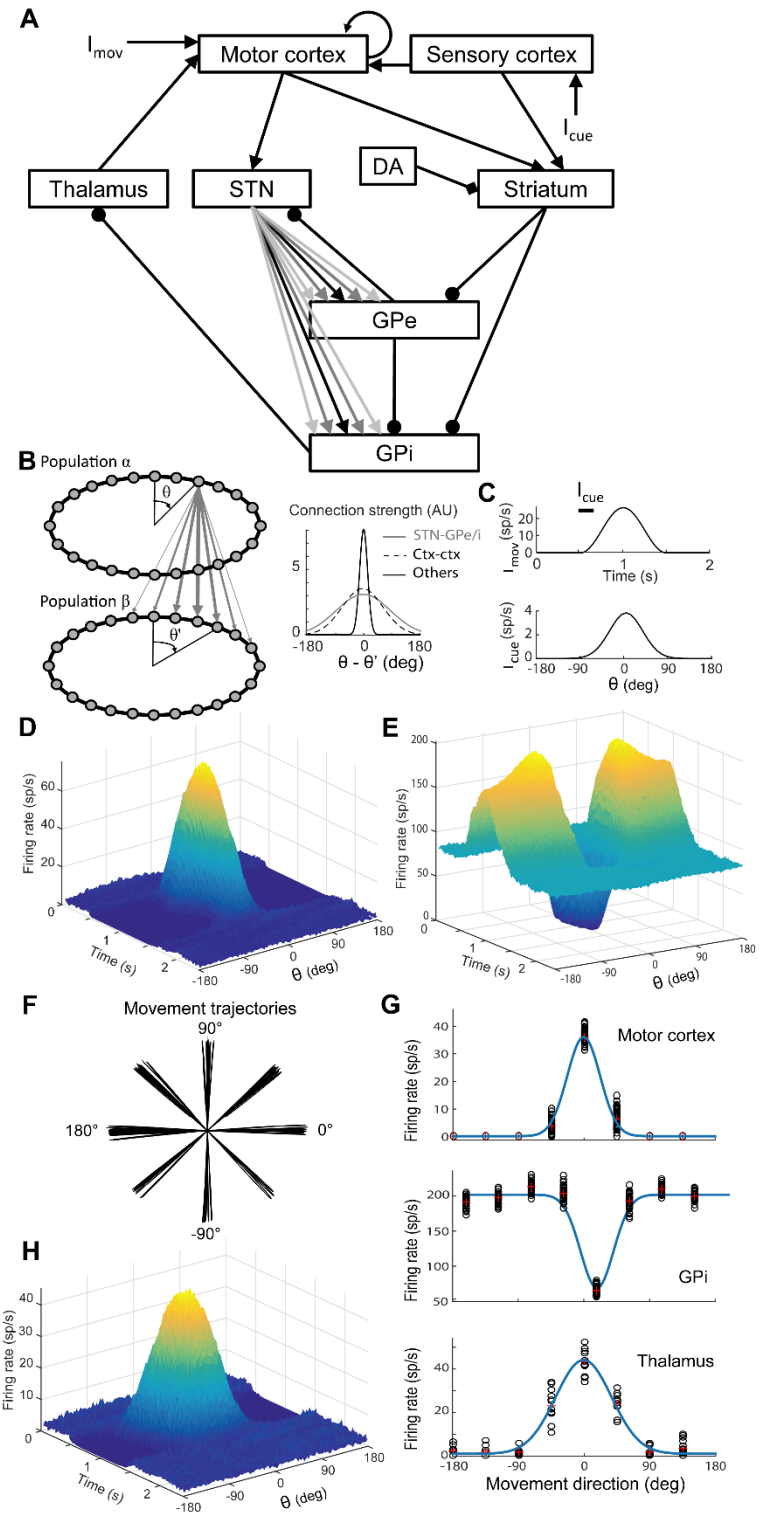

Figure 1: The BG-thalamo-cortical network model, tuning of synaptic weights, movement-related activity in the network and tuning curves in the cortex and GPi. A: Architecture of the BG model network. Each neural population is represented by a ring where neurons are evenly distributed and associated with a PD between -180° and 180°. B: The connectivity between the different populations of the network is focal and organized in parallel beams (bottom right inset). In two connected populations, the strength of the connection between two neurons from these two populations is a gaussian function of the angular distance between the PDs of the two neurons (see Methods). The width of the gaussian connectivity function is tight (σ = 9°) in all connections of the network but the divergent projection from STN to GPe and GPi (σ = 360°). C: Movement-related input in the model. Top: time course of the input $\mathbf{I_{mov}}$ added to the cortical population to induce movement. The thick black line indicates the duration of the weak and transient movement-related input from the sensory cortex to the striatum $\mathbf{I_{cue}}$ (pointing the direction of movement). Bottom: Profile of the weak and transient movement-related input to the striatum. D: Movement-related activity induced in the cortex in the presence of the full closed-loop BG-thalamo-cortical network. The cortical population displays a strong activity bump peaking at the angle pointed by the input from the sensory cortex ($\mathbf{I_{cue}}$). E: the GPi population displays a sharp profile with decreased activity around the movement direction and increased activity away from the movement direction. F: Example trajectories generated in the model using a population vector computed

as the sum of the PD of motor cortical neurons weighted by their firing rate at each time step. G: Tuning curve of the movement-related activity for neurons in the cortex (Top) and GPi (Bottom). The angle associated with each neuron in the population corresponds to the preferred direction, as defined by the movement direction evoking maximal activity increase (in the cortex) or decrease (in the GPi). H: Movement-related activity induced in the cortex following the inhibition of the GPi (50% of neurons silenced in the GPi). Note that the cortical population still displays a strong activity bump peaking at the angle pointed by the input from the sensory cortex (**I_{cue}**), although with lower amplitude compared to the situation with intact BG-cortical loop in D. Parameters are as in Table 1.

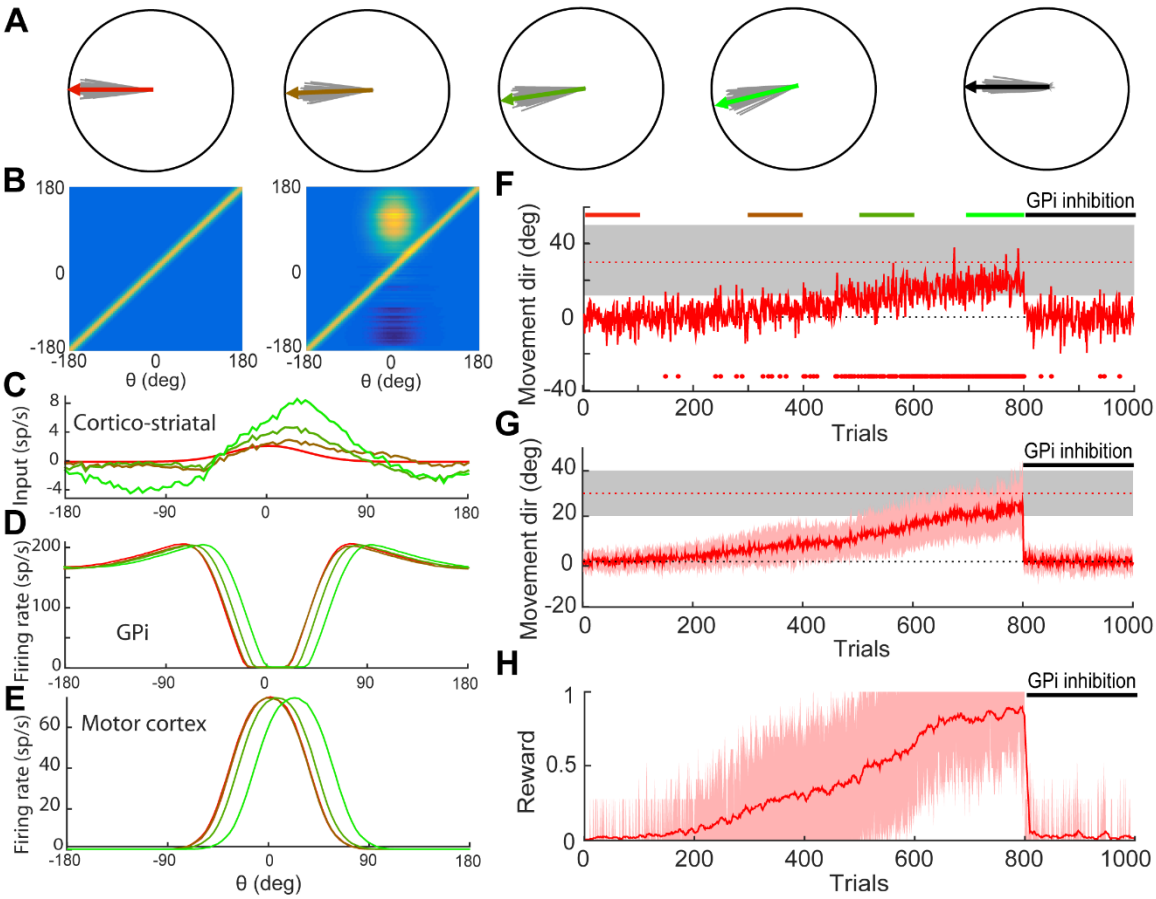

Figure 2: Reward-modulated cortico-striatal plasticity allows adaptation of movement direction in a reward-driven rotation protocol with 30° shift. GPi inhibition after learning cancels the learned adaptation bias. A: direction of movements produced along the reward-driven adaptation protocol. Example trajectories of 20 movements toward the target at different stages of the adaptation protocol (from left to right: trials 1-20, 201-220, 401-420, 781-800). B: Matrix of synaptic weights between Mctx and the Striatum before (left) and after(right) adaptation. C-E: Profil of the input from sensory cortex to the striatal neurons (C) and activity of the GPi neurons (D) and cortical neurons (E) across trials along the adaptation protocol. Example trials are coloured according to their position in the protocol (red: early, green: late). Arrows along the bottom X-axis indicate the movement direction for the corresponding example trials. F: Movement direction along the adaptation protocol in an example run. The shaded area indicates rewarded movement directions. Points along the x axis indicate rewarded trials. In trials 800-1000 the GPi is inhibited (100% inhibition). G-H: Average movement direction (G) and reward rate (H) over 20 runs of the

adaptation protocol. Shaded areas indicate the mean +/- SD of the movement direction (in G) or reward rate (in H).

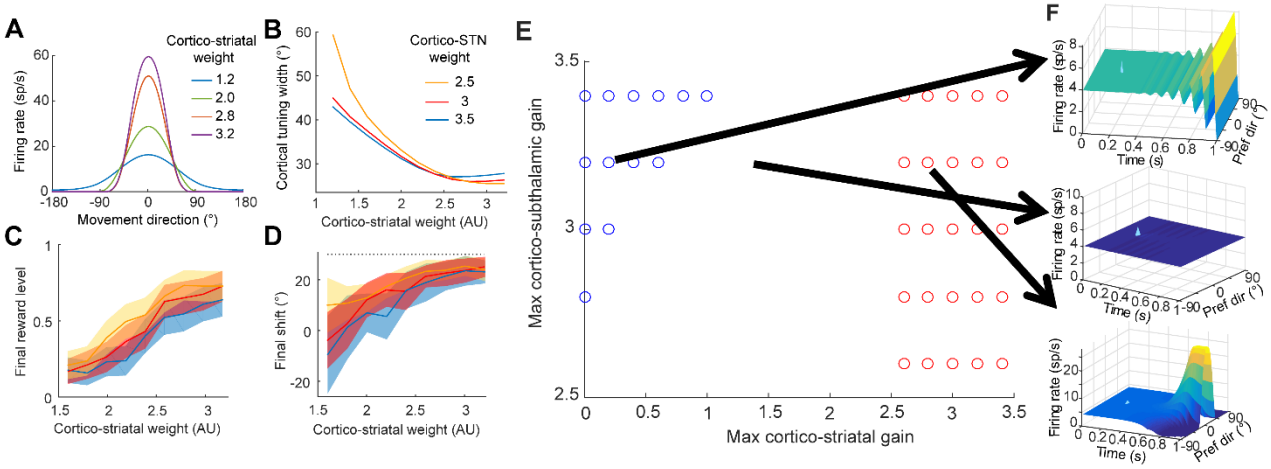

Figure 3: Cortical movement-related activity profile and the network ability to adapt movement direction strongly depends on the average cortico-striatal synaptic weight. A: Cortical tuning curve is narrow for strong cortico-striatal weights and broad when cortico-striatal connection is weakened. B: The decrease in cortical tuning with the average cortico-striatal weight is true for a large range of cortico-STN synaptic weights. C-D: The success of movement adaptation (target at 30° +/- 20°) is compromised when cortico-striatal synapses are weakened, as evidenced by final reward level (C) or final movement direction after 1000 trials (D). E: Phase diagram showing the three dynamical regimes of the BG-cortical network as a function of the maximal cortico-striatal (Gctxstr) and cortico-subthalamic (Gctxstn) synaptic gains. For each value of (Gctxstr,Gctxstn), the network is set at the steady state of its dynamics, where constant external inputs lead to constant activity levels in all populations. A perturbation is applied to this steady state to test its stability, highlighting the properties of the possible instabilities of the network. The white area corresponds to the domain where the steady state is stable and the network remains at constant activity levels after perturbation. Red dots indicate a symmetry breaking instability while blue dots denote an oscillatory instability. F: Cortical activity profile following the perturbation of the network from its steady state. Top: oscillatory instability leads to cortical oscillations that propagate through the

network. Middle: Stable steady state. Bottom: symmetry breaking instability leads to a sharp activity profile in the cortical population due to the network dynamics.

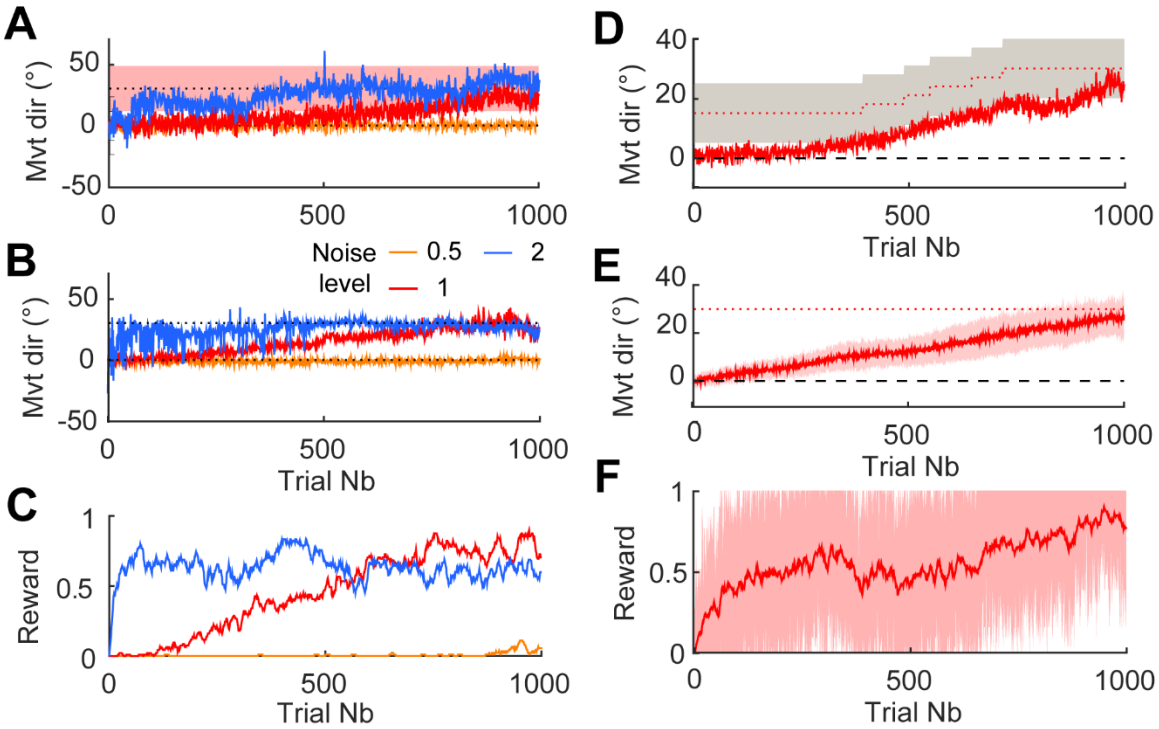

Figure 4. Adaptation is very slow for low movement variability, and adaptation in realistic conditions requires shaping. A-B: Example learning curves (A) and reward ratios (B) for adaptation of movement direction in the model with various levels of striatal synaptic noise (0.5, 1 and 4). The shaded area represents rewarded movement directions (30° +/- 20°). While low noise level lead to low movement variability and thus prevent adaptation, high noise leads to fast convergence. C-D: Mean learning curves (C) and reward ratios (D) for striatal synaptic noise levels of 0.5-1-2-4. E-F: Example learning curves (E) and reward ratios (F) during an adaptation protocol including shaping. The direction of the target (and thus rewarded movements) is slowly increased as the movement direction is modified. G-H: Mean learning curves (C) and reward ratios (D) over 20 runs in the shaping protocol.

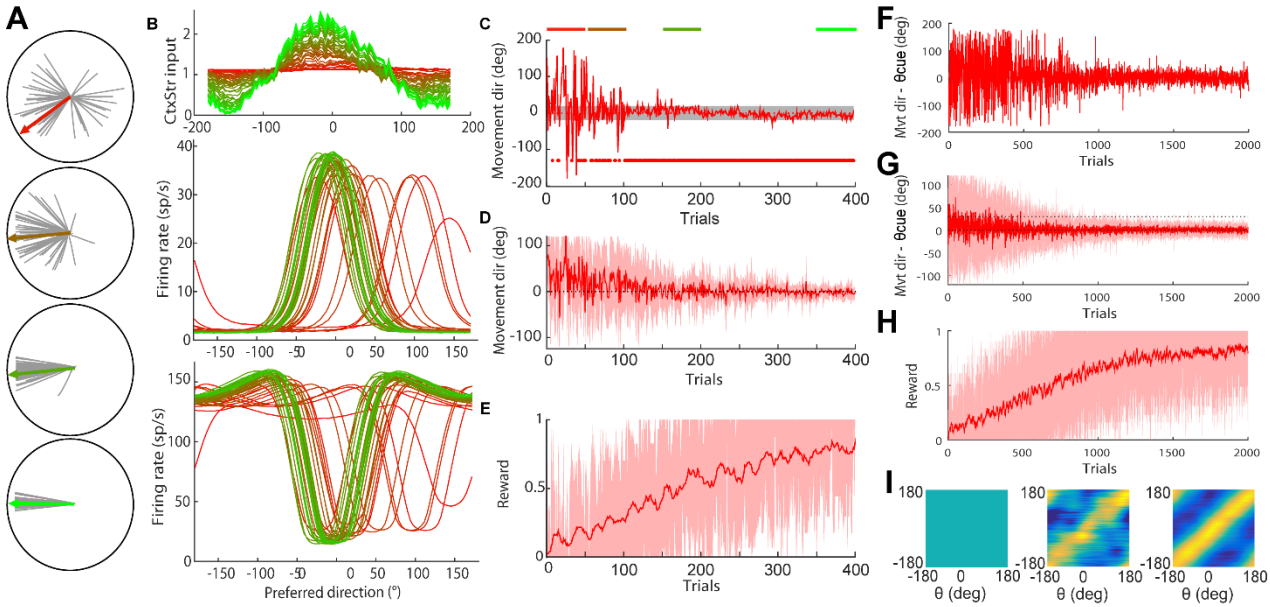

Figure 5. De novo learning: aligning visual cue direction with movement direction of spontaneous reaching movements in a system with no cortico-cortical connectivity. A: direction of movements produced along the reward-driven adaptation protocol. Example trajectories of 20 movements toward the target at different stages of the initial learning protocol (from left to right: trials 1-20, 101-120, 201-220, 381-400). B: Profil of the input from sensory cortex to the striatal neurons (top panel) and activity of the GPi neurons (middle panel) and cortical neurons (bottom panel) across trials along the initial learning protocol. Example trials are coloured according to their position in the protocol (red: early, green: late). C: Movement direction along the adaptation protocol in an example run. The shaded area indicates rewarded movement directions. Points along the x axis indicate rewarded trials. In trials 400-500 the GPi is inhibited (100% inhibition). D-E: Average movement direction (D) and reward rate (E) over 20 runs of the initial learning protocol. Shaded areas indicate the mean +/- SD of the movement direction (in D) or reward rate (in E). F-I: De novo learning of 16 directions, with multiple cues presented in a random order and uniformly distributed between 0 and 360 degrees. F: During the learning process, movements directed toward the presented cue are rewarded, and movements progressively align to cue directions. G-H: For 20 runs of the learning of 16 direction, the learning is efficient in around 1000 trials, as shown with movement direction with respect to cue direction (G) and reward rate (H). I: During the process, the synaptic weights from Vctx to striatum, initially non-specific (left) become structured (final matrix of Vctx-striatum synaptic weights after 200 trials: example run, middle; average over 20 runs, right).

m0ctx	m0str	m0stn	m0gpe	m0gpi	m0thal	Gstrctx Max	Ggpistr	Gthagpi	Gctxtha	Gstnctx	Ggpistn	Ggpigpe	Ggpestr	Ggpestn	Gstngpe
4Hz	-2Hz	16	60	80	20	2	2.6	0.3	1	3	3.3	0.3	0.8	0.8	0.6
dt	η	σconnect	$\sigma_{stn\text{-}connect}$	Δstrctx	Δgpistr	Δthagpi	Δctxtha	Δstnctx	Δgpistn	Δgpigpe	Δgpestr	Δgpestn	Δstngpe	β	
0.5 ms	$2*10^{-4}$	9°	360°	6ms	10ms	5ms	5ms	5ms	5ms	2ms	4ms	3ms	3ms	4	
σ^{noise}_{ctx}	σ^{noise}_{str}	σ^{noise}_{gpi}	σ^{noise}_{gpe}	σ^{noise}_{tha}	σ^{noise}_{stn}	$\tau^{correlnoise}$	σ^{correl}_{noise}	A^{correl}_{noise}							
2Hz	4Hz	10Hz	10Hz	4Hz	4Hz	50ms	45°	1.5							

`**Table 1.** Model parameters