

Thalamocortical contribution to solving credit assignment in neural systems

Mien Brabeeba Wang^{*†} and Michael M. Halassa[†]

April 6, 2021

Abstract

Animal brains evolved to optimize behavior in dynamically changing environments, selecting actions that maximize future rewards. A large body of experimental work indicates that such optimization changes the wiring of neural circuits, appropriately mapping environmental input onto behavioral outputs. A major unsolved scientific question is how optimal wiring adjustments, which must target the connections responsible for rewards, can be accomplished when the relation between sensory inputs, action taken, environmental context with rewards is ambiguous. The computational problem of properly targeting cues, contexts and actions that lead to reward is known as *structural*, *contextual* and *temporal credit assignment* respectively. In this review, we survey prior approaches to these three types of problems and advance the notion that the brain’s specialized neural architectures provide efficient solutions. Within this framework, the thalamus with its cortical and basal ganglia interactions serve as a systems-level solution to credit assignment. Specifically, we propose that thalamocortical interaction is the locus of meta-learning where the thalamus provides cortical control functions that parametrize the cortical activity association space. By selecting among these control functions, the basal ganglia hierarchically guide thalamocortical plasticity across two timescales to enable meta-learning. The faster timescale establishes contextual associations to enable rapid behavioral flexibility while the slower one enables generalization to new contexts. Incorporating different thalamic control functions under this framework clarifies how thalamocortical-basal ganglia interactions may simultaneously solve the three credit assignment problems.

Introduction Learning which action to choose in an uncertain environment is a hallmark of intelligence [1–3]. When animals explore unfamiliar environments, they tend to reinforce actions that lead to unexpected rewards. A common notion in contemporary neuroscience is that such behavioral reinforcement emerges from changes in synaptic connectivity, where synapses that contribute to the unexpected reward are strengthened [4–8]. A prominent model for connecting synaptic to behavioral reinforcement is dopaminergic innervation of

¹Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

²Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge MA.

basal ganglia (BG), where dopamine (DA) carries the reward prediction error (RPE) signals to guide synaptic learning [9–12]. This circuit motif is thought to implement a basic form of the reinforcement learning algorithm [13–19], which has had much success in explaining simple Pavlovian and instrumental conditioning [3, 15, 19, 20]. However, what allows this circuit to reinforce the appropriate connections in complex natural environments where animals are presented with multiple cues in multiple contexts and make multiple actions before they receive the reward, is unknown. If one naively credits all synapses with the RPE signals, the learning will be highly inefficient since different cues, contexts and actions contribute to the RPE signals differently. To properly credit the cues, context and actions that lead to unexpected reward is a challenging problem, known as the *credit assignment* problem [8, 21–23].

One can roughly categorize the credit assignment into *structural credit assignment*, *contextual credit assignment* and *temporal credit assignment* (Figure 1). In structural credit assignment, animals may make decisions in a multi-cue environment and should be able to credit those cues that contribute to the rewarding outcome. Similarly, if actions are being chosen based on internal decision variables, then the underlying activity states must also be reinforced. In such cases, neurons that are selective to external cues or internal latent variables need to adjust their downstream connectivity based on its contribution of their downstream targets to the RPE. This is a challenging computation to implement because, for upstream neurons, the RPE will be dependent on downstream neurons that are several connections away. For example, a sensory neuron needs to know the action chosen in the motor cortex to selectively credit the sensory synapses that contribute to the action. In contextual credit assignment, animals not only need to appropriately credit the sensory cues and actions that lead to the reward but also need to credit the sensorimotor combination in the right context. For example, when one is in the United States, one learns to first look left before crossing the street, whereas, in the United Kingdom, one learns to look right instead. However, after spending time in the UK, someone from the US should not unlearn the behavior of looking left first when they return home because their brain ought to properly assign the credit to a different context. In the temporal credit assignment problem, animals make decisions in an environment with distant rewards and need to figure out which past sensory cues and actions lead to the current reward. For example, in a game of Go, even though the result of the game is only revealed after hundreds of hands, professional players can recognize which moves in the past are good and reinforce such moves.

In this review, we will first go over common approaches from machine learning to tackle these three credit assignment problems. In doing so, we highlight the challenge in their efficient implementation within biological neural circuits. We also highlight some recent proposals that advance the notion of specialized neural hardware that approximate more general solutions for credit assignment [24–37]. Along these lines, we propose an efficient systems-level solution involving the thalamus and its interaction with the cortex and BG for these three credit assignment problems.

Common machine learning approaches to credit assignment One solution to structural credit assignment in machine learning is backpropagation [22]. Backpropagation recursively computes the vector-valued error signal for synapses based on their contribution

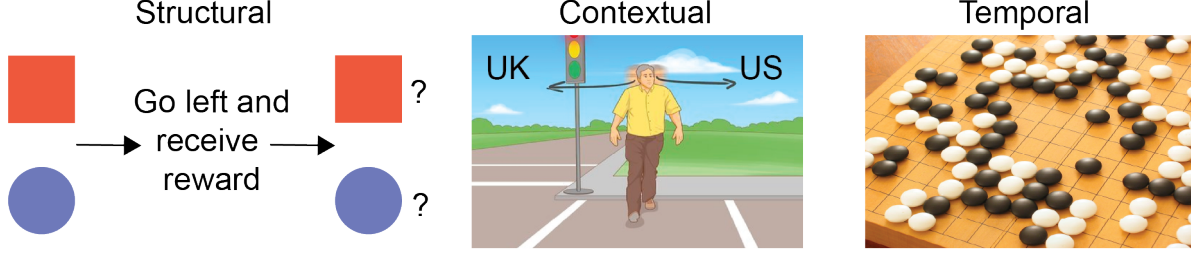


Figure 1: **Different types of credit assignment** When animals face a decision with multiple cues, contexts and actions before the reward, animals need to assign proper credit to the cues, contexts and actions that lead to behavioral outcomes. In structural credit assignment, animals need to credit the cues that lead to the reward but not the distractor cues. In contextual credit assignment, animals need to credit sensory cues and actions in the right context. In temporal credit assignment, animals need to figure out which past actions lead to the current reward.

to the error signal. There is much empirical success of backpropagation in surpassing human performance in supervised learning such as image recognition [38, 39] and reinforcement learning such as playing the game of Go and Atari [40–43]. Additionally, comparing artificial networks trained with backpropagation to neural responses from the ventral visual stream of non-human primates shows comparable internal representations [44, 45]. Despite its empirical success in superhuman level performance and matching the internal representation of actual brains, backpropagation may not be straightforward to implement in biological neural circuits as we explain below.

In its most basic form, backpropagation requires symmetric connections between neurons (forward and backward connections). Mathematically, we can write down the backpropagation in Equation 1:

$$\delta W_i \propto \frac{\partial E}{\partial W_i} = e_i f(a_{i-1})^\top \quad (1)$$

where

$$e_i = W_{i+1}^\top e_{i+1} \circ f'(a_i),$$

E is the total error, e_i is the vector error at layer i , W_i is the synaptic weight connecting layer $i - 1$ to layer i and f is the nonlinearity. Intuitively, this is saying that the change of synaptic weight W_i is computed by a Hebbian learning rule between backpropagation error e_i and activity from last layer $f(a_{i-1})$ while the backpropagation error is computed by backpropagating the error in the next layer through symmetric feedback weights $W_{i+1}^\top$. Importantly, in this algorithm, error signals do not alter the activity of neurons in the preceding layers and instead operate independently from the feedforward activity. However, such arrangement is not observed in the brain; symmetric connections across neurons are not a universal feature of circuit organization, and biological neurons may encode both feedforward inputs and errors through changes in spike output (changes in activity) [32, 46]. Therefore, it is hard to imagine how the basic form of backpropagation (symmetry and error/activity separation) is physically implemented in the brain.

Furthermore, while an animal can continually learn to behave across different contexts, artificial neural networks trained by backpropagation struggle to learn and remember different

tasks in different contexts: a problem known as catastrophic forgetting [47–51]. Specifically, this problem occurs when the tasks are trained sequentially because the weights optimized for former tasks will be modified to fit the later tasks. One of the common solutions is to interleave the tasks from different contexts to jointly optimize performance across contexts by using an episodic memory system and replay mechanism [49, 52]. This approach has received empirical success in artificial neural networks including learning to play many Atari game [40, 43]. However, since one needs to store past training data in memory to replay during learning, this approach demands a high computational overhead and can be inefficient as the number of the contexts increases. On the other hand, humans and animals acquire diverse sensorimotor skills in different contexts throughout their life span: a feat that cannot be solely explained by memory replay [51, 53–55]. Therefore, biological neural circuits are likely to employ other solutions to contextual credit assignment in addition to memory replay.

Lastly, one of the most common solutions to temporal credit assignment in reinforcement learning is temporal difference (TD) learning [15, 19]. Instead of directly learning from distant rewards, animals use more proximal inputs (which form a current state) to learn the expected future rewards. Specifically, the expected future reward can be expressed as

$$V(s_t) = \mathbb{E}[r_t + r_{t+1} + r_{t+2} + \dots | s_t] = \mathbb{E}\left[\sum_{i=0}^{\infty} r_{t+i} \middle| s_t\right], \quad (2)$$

where s_t is the state at time t and r_t is the reward at time t . The learning is guided by the temporal prediction error

$$e_t = r_t + V(s_{t+1}) - V(s_t). \quad (3)$$

TD learning brings the distant reward to the present by estimating the future rewards given the current state, and thus solve the temporal credit assignment problem by reinforcing the action that leads to positive TD RPE. However, for TD learning to be valid, state transition and rewards need to depend on only the current state and the current action. In natural settings, both rewards and sensory inputs can depend on past sensory information and past actions. For example, when people communicate with each other, their intentions may only be clear after a whole sentence, not the individual words spoken. Thus, the naive implementation of TD learning on sensory inputs can fail on such tasks with sequential dependency. To overcome this shortcoming, a working memory buffer may be required, and in machine learning can be implemented via Long Short-Term Memory networks (LSTMs) [56–58]. An LSTM can selectively maintain the past sensory inputs in hidden memory states. However, one needs to use backpropagation to train LSTM which is unlikely to be implemented in the brain as such. The brain probably has an alternative mechanism to selectively maintain the past sensory inputs and actions in conjunction with TD RPE signals from DA to solve temporal credit assignment.

Therefore, to solve the three credit assignment problems in the brain, one needs to seek different solutions. One of the pitfalls of backpropagation is that it is a general algorithm that works on any architecture. However, actual brains are collections of specialized hardware put together in a specialized way. It can be conceived that through clever coordination between different cell types and different circuits, the brains can solve the credit assignment problem by leveraging its specialized architectures. Along this line of ideas, many investigators have

proposed cellular [26, 29, 30, 32, 34–36] and circuit level mechanisms [24, 25, 28, 33] to assign credit appropriately. In this review, we would like to advance the notion that the specialized hardware arrangement also happens at the system level and propose that the thalamus and its interaction with basal ganglia (BG) and the cortex serve as a system-level solution for these three types of credit assignment.

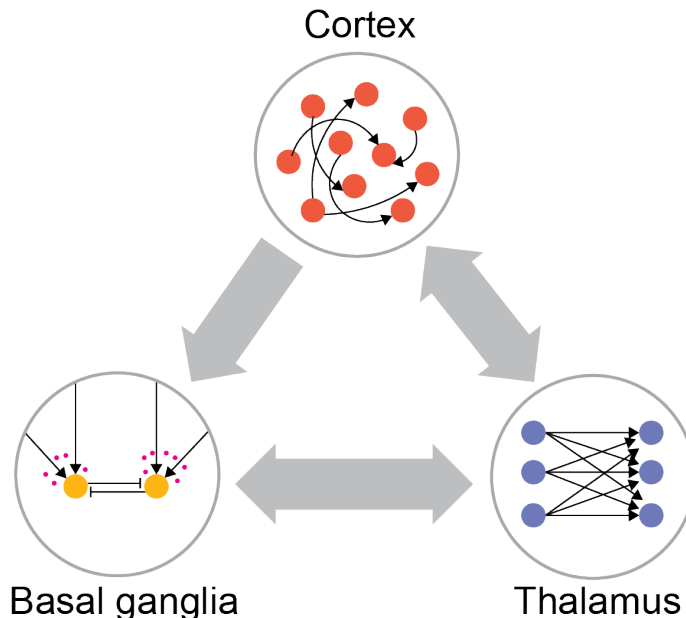


Figure 2: **Distinct architectures of cortex, thalamus and basal ganglia** Cortex is largely composed of excitatory neurons with extensive recurrent connectivity. Thalamus consists of mostly excitatory neurons without lateral connections. Basal ganglia consist of mostly inhibitory neurons driven by cortical and thalamic inputs, and the corticostriatal plasticity is modulated by dopamine.

A proposal: thalamocortical-basal ganglia interactions enable meta-learning to solve credit assignment. To motivate the notion of thalamocortical-basal ganglia interactions being a potential solution for credit assignment, we will start with a brief introduction. The cortex, thalamus and basal ganglia are the three major components of the mammalian forebrain – the part of the brain to which high level cognitive capacities are attributed to [2, 3, 59–65]. Each of these components has its specialized internal architectures; the cortex is dominated by excitatory neurons with extensive lateral connectivity profiles [66–68], the thalamus is grossly divided into different nuclei harboring mostly excitatory neurons devoid of lateral connections [69–71], and the basal ganglia are a series of inhibitory structures driven by excitatory inputs from the cortex and thalamus [72–74] (Figure 2). A popular view within system neuroscience stipulates that BG and the cortex underwent different learning paradigms where BG is involved in reinforcement learning while the cortex is involved in unsupervised learning [75, 76]. Specifically, the input structure of the basal ganglia known as the striatum is thought to be where reward gated plasticity takes place to implement reinforcement learning [3, 12, 35, 64, 77, 78]. One such evidence is the high temporal precision of DA

activity in the striatum. To accurately attribute the action that leads to positive RPE, DA is released into the relevant corticostriatal synapses. However, DA needs to disappear quickly to prevent the next stimulus-response combination from being reinforced. In the striatum, this elimination process is carried out by dopamine active transporter (DAT) to maintain a high temporal resolution of DA activity to support reinforcement learning [79–81]. In contrast, although the cortex also has dopaminergic innervation, cortical DAT expression is low and therefore DA levels may change at a timescale that is too slow to support reinforcement learning [79, 81–83] but instead support other processes related to learning [2, 2, 61]. In fact, ample evidence indicates that cortical structures undergo Hebbian-like long term potentiation (LTP) and long term depression (LTD) [84–86]. However, despite the unsupervised nature of these processes, cortical representations are task-relevant and include appropriate sensorimotor mappings that lead to rewards [87–92]. How could this arise from an unsupervised process? One possible explanation is that basal ganglia activate the appropriate cortical neurons during behaviors and the cortical network collectively consolidates high reward sensorimotor mappings via Hebbian-like learning [93–97]. Previous computational accounts of this process have emphasized a consolidation function for the cortex in this process, which naively would beg the question of why duplicate a process that seems to function well in the basal ganglia and perhaps include a lot of details of the associated experience?

The answer to this question is the core of our proposal. We propose that the learning process is not a duplication, but instead that the reinforcement process in the basal ganglia selects thalamic control functions that subsequently activate cortical associations to allow flexible mappings across different contexts (Figure 3).

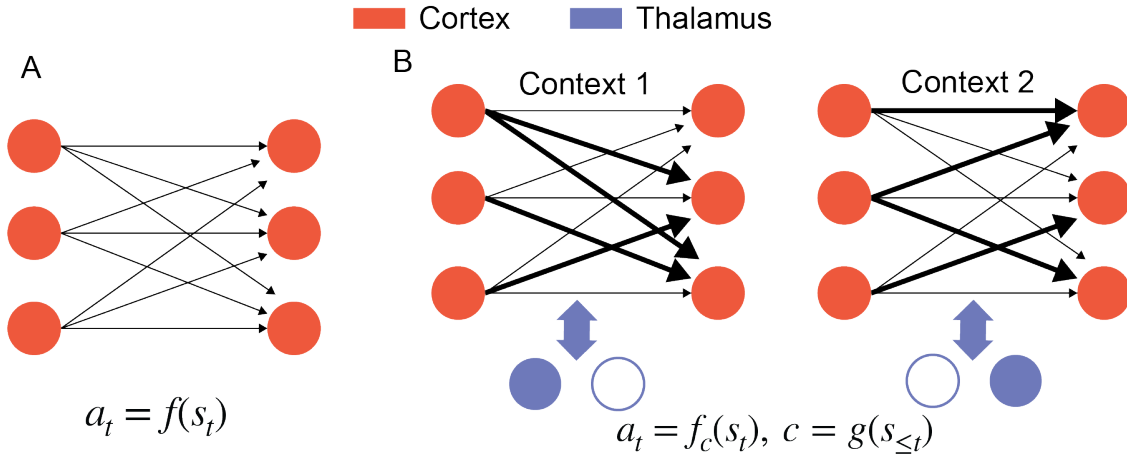


Figure 3: **Two views of learning in the cortex** A. One possible view is that the Hebbian cortical plasticity consolidates the sensorimotor mapping from BG to learn a stimulus-action mapping $a_t = f(s_t)$. B. We propose that thalamocortical systems do meta-learning by consolidating the teaching signals from BG to learn a context-dependent mapping $a_t = f_c(s_t)$ where the context c is computed by past stimulus history and represented by different thalamic activities.

To understand this proposition, we need to take a closer look at the involvement of these distinct network elements in task learning. Learning in basal ganglia happens in corticostriatal synapses where the basic form of reinforcement learning is implemented. Specifically,

the coactivation of sensory and motor cortical inputs generates eligibility traces in corticostriatal synapses that get captured by the presence or absence of DA [35, 98, 99]. This RL algorithm is fast at acquiring simple associations but slow at generalization to other behaviors. On the other hand, the cortical plasticity operates in a much slower timescale but seems to allow flexible behaviors and fast generalization [2, 60, 100, 101]. How does the cortex exhibit slow synaptic plasticity and flexible behaviors at the same time? An explanatory framework is meta-learning [102, 103], where the flexibility arises from network dynamics and the generalization emerges from slow synaptic plasticity across different contexts. In other words, synaptic plasticity stores a higher-order association between contexts and sensorimotor associations while the network dynamics switches between different sensorimotor associations based on this higher order association. However, properly arbitrating between synaptic plasticity and network dynamics to store such higher order association is a non-trivial task [104]. We propose that the thalamocortical system learns these dynamics, where the thalamus provides control nodes that parametrize the cortical activity association space. Basal ganglia inputs to the thalamus learn to select between these different control nodes directly implementing the interface between weight adjustment and dynamical controls. Our proposal rests on the following three specific points.

First, building on a line of the literature that shows diverse thalamocortical interaction in sensory, cognitive and motor cortex, we propose that thalamic output may be described as control functions over cortical computations. These control functions can be purely in the sensory domain like attentional filtering, in the cognitive domain like manipulating working memory or in the motor domain like preparation for movement [105–114]. These functions directly relate thalamic activity patterns to different cortical dynamical regimes and thus offer a way to establish higher order association between context and sensorimotor mapping within the thalamocortical pathways. Second, based on previous studies on direct and indirect BG pathways that influence most cortical regions [115–118], we propose that BG hierarchically selects these thalamic control functions to influence activities of the cortex toward rewarding behavioral outcomes. Lastly, we propose that thalamocortical structure consolidate the selection of BG through a two timescales Hebbian learning process to enable meta-learning. Specifically, the faster corticothalamic plasticity learns the higher order association that enables flexible contextual switching with different thalamic patterns [113, 119] while the slower cortical plasticity learns the shared representations that allow generalization to new behaviors. Below, we will go over the supporting literature that leads us to this proposal.

More general roles of thalamocortical interaction and basal ganglia Classical literature has emphasized the role of the thalamus in transmitting sensory inputs to the cortex. This is because some of the better studied thalamic pathways are those connected to sensors on one end and primary cortical areas on another [120–124]. From that perspective, thalamic neurons being devoid of lateral connection transmit their inputs (e.g. from the retina in the case of the lateral geniculate nucleus (LGN)) to the primary sensory cortex (V1 in this same example case) and the input transformation (center-surround to oriented edges) occurs within the cortex [123–126]. In many cases, these formulations of thalamic “relay” have generalized to how motor and cognitive thalamocortical interactions may be operating. How-

ever, in contrast to the classical relay view of the thalamus, more recent studies have shown diverse thalamic functions in sensory, cognitive and motor processing [105–113]. For example in mice, sensory thalamocortical transmission can be adjusted based on PFC-dependent, top-down biasing signals transmitted through non-classical basal ganglia pathways involving the thalamic reticular nucleus (TRN) [107, 117, 127]. Interestingly, these task-relevant PFC signals themselves require long range interactions with the associative mediodorsal (MD) thalamus to be initiated, maintained and flexibly switched [107, 110, 113]. One can also observe nontrivial control functions in the motor thalamus. Motor preparatory activities in the anterior motor cortex (ALM) show persistent activities that predicted future actions. Interestingly, the motor thalamus also shows similar preparatory activities that predict future actions and by optogenetically manipulate the motor thalamus activities, the persistent activities in ALM quickly diminished [111]. Based on the above studies, we propose that the thalamus provides a set of control functions to the cortex. Specifically, cortical computations may be flexibly switched to different dynamical modes by activating a particular thalamic output that corresponds to that mode.

On the other hand, the selective role of BG in motor and cognitive control also has dominated the literature because thalamocortical-basal ganglia interaction is the most well studied in frontal systems [62–64, 128, 129]. However, classical and contemporary studies have recognized that all cortical areas, including primary sensory areas project to the striatum [115, 116, 118]. Similarly, the basal ganglia can project to the more sensory parts of the thalamus through lesser-studied pathways to influence the sensory cortex [115, 117, 118]. Specifically, a non-classical BG pathway projects to TRN which in turn modulates the activities of LGN to influence sensory thalamocortical transmission [117]. On the other hand, it has also been argued that BG are involved in gating working memory [129, 130]. This shows that BG has a much more general role than classical action and action strategy selection. Therefore, combining with our proposals on thalamic control functions, we propose that BG hierarchically selects different thalamic control functions to influence all cortical areas in different contexts through reinforcement learning.

Furthermore, there are series of the work that indicates the role of BG to guide plasticity in thalamocortical structures [94, 96–98, 131]. In particular, there is evidence that BG is critical for the initial learning and less involved in the automatic behaviors once the behaviors are learned across different species. In zebra finches, the lesion of BG in adult zebra finch has little effects on song production, but the lesion of BG in juvenile zebra finch prevents the bird from learning the song [99, 132, 133]. Similar patterns can be observed in people with Parkinson’s disease. Parkinson’s patients who have a reduction of DA and striatal defects have troubles in solving procedural learning tasks but can produce automatic behaviors normally [134–136]. This behavioral evidence suggests that thalamocortical structures consolidate the learning from BG as the behaviors become more automatic. Furthermore, on the synaptic level, a songbird learning circuit also demonstrates this cortical consolidation motif [97, 131]. In a zebra finch, the premotor nucleus HVC (a proper name) projects to the motor nucleus robust nucleus of the arcopallium (RA) to produce the song. On the other hand, RA also receives BG nucleus Area X mediated inputs from the lateral nucleus of the medial nidopallium (LMAN). The latter pathway is believed to be a locus of reinforcement learning in the songbird circuit. By burst stimulating both input pathways in different time lags, one can discover that HVC-RA and LMAN-RA underwent opposite plasticity [131].

This suggests that the learning is gradually transferred from LMAN-RA to HVC-RA pathway [97, 99, 131]. This indicates a general role of BG as the trainer for cortical plasticity. We further propose that BG is the trainer in two different timescales for thalamocortical structures to enable meta-learning. The faster timescale trainer trains the corticothalamic connections to select the appropriate thalamic control functions in different contexts while the slower timescale trainer trains the cortical connections to form a task-relevant and generalizable representation.

The thalamocortical structure consolidates the BG selections on thalamic control functions in different timescales to enable meta-learning. In this section, we will first propose a general network consolidation motif inspired by a songbird circuit mentioned above and then show how one can apply this motif in two different timescales in thalamocortical-basal ganglia interactions to enable meta-learning. We propose the following general network motif on learning consolidation (Figure 4). In this motif, there are three groups of neurons, input, output and teacher neurons. This is an abstract depiction that could map on multiple circuit scenarios (i.e. what the actual input, output and teacher circuits are). However, as a way of example, we can imagine an implementation of the input and output being intracortical and the teacher being the basal ganglia. As a general rule, the output receives both input and teacher connections, with the former undergoing unsupervised plasticity and the latter undergoing reinforcement learning. A key idea that we propose, is that unsupervised plasticity requires coactivation of both the direct and indirect pathways depicted in Figure 4. Specifically, the direct connections strengthen when it coincides with the teacher inputs (Hebbian-like) while the teacher connections weaken under the same conditions (anti-Hebbian-like). Therefore, at the initial stage of learning, the indirect pathway will have a strong influence on the output, which will lessen as the direct pathway is strengthened and thus the behaviors become less teacher dependent. We describe a basic mathematical setup of the network motif in Box 1.

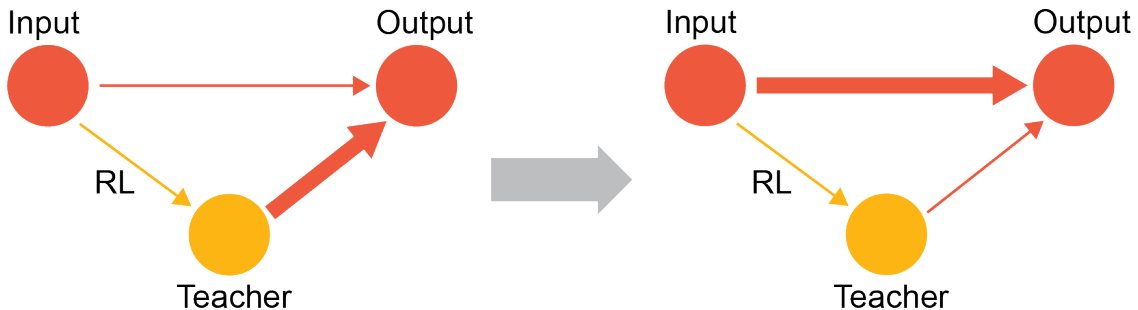


Figure 4: **A circuit motif for learning consolidation.** The outputs receive both student inputs and teacher inputs. Teacher inputs are learned from reinforcement learning. At the initial stage of learning, teacher inputs dominate the output. However, whenever the student inputs and the teacher inputs coincide, the student inputs are strengthened while the teacher inputs are weakened. Therefore, when the students learn the task, the learning is consolidated into the student connections and the teacher has little influence on the outputs.

Box 1: A network motif for structural credit assignment with RL

In this box, we will introduce a basic setup of this motif and the intuition on how it works. Given the activity of the input neurons $x(t) \in \mathbb{R}^n$, the teacher neurons $y(t) \in \mathbb{R}^m$, the output neurons $o(t) \in \mathbb{R}^m$ and the input-teacher synapses $W \in \mathbb{R}^{m \times n}$, input- output synapses $M \in \mathbb{R}^{m \times n}$ and teacher-output synapses $w \in \mathbb{R}^m$, we have the following relationship

$$o(t) = Mx(t) + w \cdot y(t), \quad (4)$$

$$g(t) = Wx(t), \quad y_i(t) = g_i(t) \text{ if } i = a_t^*, \quad y_i(t) = 0 \text{ otherwise} \quad (5)$$

where $\cdot$ is the entrywise product and a_t^* is the action selected by the teacher. And the animals select the action a_s^* as the behavior output based on the output activity. One possible action selection mechanism could be

$$\mathbb{P}(a_t^* = i) = \frac{e^{g_i(t)}}{\sum_{j=1}^m e^{g_j(t)}}, \quad \mathbb{P}(a_s^* = i) = \frac{e^{s_i(t)}}{\sum_{j=1}^m e^{s_j(t)}}. \quad (6)$$

The plasticity rule of the input-teacher synapse can be described by reinforcement learning

$$\Delta W_{a_s^*} \propto (r_t - s_{a_s^*}(t))x(t), \quad \Delta W_i = 0 \text{ if } i \neq a_s^*. \quad (7)$$

The input-output synapses got strengthen if the input-output and teacher-output synapses coactivate the output. Below is one implementation of this idea

$$\Delta M_{a_t^*} \propto w_{a_t^*} t_{a_t^*} M_{a_t^*} \cdot x(t). \quad (8)$$

And furthermore, the teacher-output synapses undergo the opposite plasticity

$$\Delta w_{a_t^*} \propto -w_{a_t^*} t_{a_t^*} M_{a_t^*}^\top x(t), \quad (9)$$

so the net contribution to the output is balanced. The decision of the output will rely on the teacher less as the input-output synapses learn the right sensorimotor association from the teacher.

From the songbird example, we see how thalamocortical structures can consolidate simple associations learned through the basal ganglia. To enable meta-learning, we propose that this general direct/indirect connectivity idea operates over two different timescales within thalamocortical-basal ganglia interactions (Figure 5). First, combining the idea of thalamic outputs as control functions over cortical network activity patterns and the basal ganglia selecting such functions, we frame learning in basal ganglia as a process that connects contextual associations (higher order) with the appropriate dynamical control that maximizes reward at the sensorimotor level (lower order). Under this framing, corticothalamic plasticity consolidates the higher order association within a fast timescale. This allows flexible switching between different thalamic control functions in different contexts. On the other hand, the cortical plasticity consolidates the sensorimotor association over a slow timescale to allow

shared representation that can generalize across different contexts. As the thalamocortical structures learn the higher order association, the behaviors become less BG-dependent and the network is able to switch between different thalamic control functions to induce different sensorimotor mappings in different contexts. By having two learning timescales, animals can conceivably both adapt quickly in changing environments with fast learning of corticothalamic connections while maintaining the important information across the environment in the cortical connections.

Some anatomical observations support this idea. Classically, thalamocortical neurons are classified into matrix and core [70, 137]. Matrix has a more modulatory role to the cortical dynamics in a diffusive projection while core has a driver role to the cortical dynamic in a topographically restricted dense projection. Most importantly, matrix has innervation to striatum while core does not. This indicates that matrix might serve as the role of control functions in the faster consolidation loop with the feedback to striatum to conduct credit assignment. On the other hand, core might be more involved in the slower consolidation loop with the feedback to striatum coming from the cortex to train the common cortical representation across contexts.

In summary, this two timescales network consolidation scheme provides a general way for BG to guide plasticity in the thalamocortical architecture to enable meta-learning and thus solves structural credit assignment as a special case. Along these lines, experimental evidence supports the notion that when faced with multi-sensory inputs, the BG can selectively disinhibit a modality-specific subnetwork of the thalamic reticular nucleus (TRN) to filter out the sensory inputs that are not relevant to the behavior outcomes and thus solve the structural credit assignment problem.

In the discussion above, we discuss our proposal under a general formulation of thalamic control functions. In the next sections, we will specify other thalamic control functions suggested by recent studies and observe how they can solve contextual and temporal credit assignments under this framework as well.

The Thalamus selectively amplifies functional cortical connectivity as a solution to contextual credit assignment and catastrophic forgetting. One of the pitfalls of the artificial neural network is catastrophic forgetting. If one trains an artificial neural network on a sequence of tasks, the performance on the older task will quickly deteriorate as the network learns the new task [47–51]. On the other hand, the brain can achieve *continual learning*, the ability to learn different tasks in different contexts without catastrophic forgetting, and thus solves contextual credit assignment [53–55, 138]. There are three main approaches in machine learning in dealing with catastrophic forgetting. First, one can use the regularization method to mostly update the weights that are less important to the prior tasks [139–144]. This idea is inspired by experimental and theoretical studies on how synaptic information is selectively protected in the brain [145–149]. However, it is unclear how to biologically compute the importance of each synapse to prior tasks nor how to do global regularization locally. Second, one can also use a dynamic architecture in which the network expands the architecture by allocating a subnetwork to train with the new information while preserving old information [150–153]. However, this type of method is not scalable since the number of neurons needs to scale linearly with the number of the task. Lastly, one can

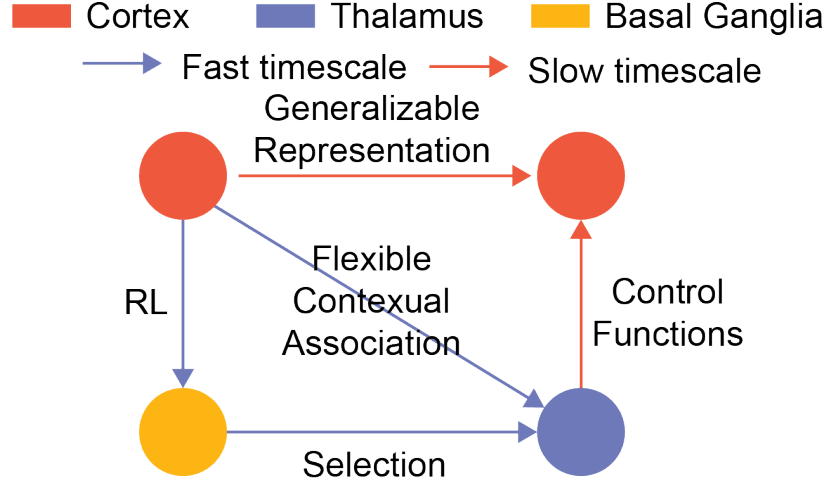


Figure 5: **Two timescales learning in thalamocortical structures** We propose that one can learn the thalamocortical structure to enable meta-learning by applying the general network motif in two different timescales. First, one can learn the corticothalamic connections by applying the motif on the blue loop with a faster timescale. This allows the network to consolidate flexible switching behaviors. Second, one can learn the cortical connections by applying the motif on the orange loop in a slower timescale. This allows cortical neurons to develop a task-relevant shared representation that can generalize across contexts.

use a memory buffer to replay past tasks to avoid catastrophic forgetting by interleaving the experience of the past tasks with the experience of the present task [49, 52, 154, 155]. However, this type of method cannot be the sole solution as the memory buffer needs to scale linearly with the number of the tasks and potentially the number of the trials.

We propose that the thalamus provides another way to solve contextual credit assignment and catastrophic forgetting via selectively amplifying parts of the cortical connections in different contexts (Figure 6). Specifically, we propose that a population of thalamic neurons topographically amplify the connectivity of cortical subnetworks as their control functions. During a behavioral task, BG selects subsets of the thalamus which selectively amplify the connectivity of cortical subnetworks. Because of the reinforcement learning in BG, the subnetwork that is the most relevant to the current task will be more preferentially activated and updated. By selecting only the relevant subnetwork to activate in one context, the thalamus protects other subnetworks which can have useful information in another context from being overwritten. The corticothalamic structures can then consolidate these BG-guided flexible switching behaviors via our proposed network motif and the switching becomes less BG-dependent. Furthermore, our proposed solution has implications on generalization as well. Different tasks can have principles in common that can be transferred. For example, although the rules of chess and Go are very different, players in both games all need to predict what the other players are going to do and counterattack based on the prediction. Since BG selects the subnetwork at each hierarchy that is most relevant to the current tasks, in addition to selecting different subnetworks to prevent catastrophic forgetting, BG can also select subnetworks that are beneficial to both tasks as well to achieve generalization. Therefore, the cortex can develop a modular hierarchical representation of the world that

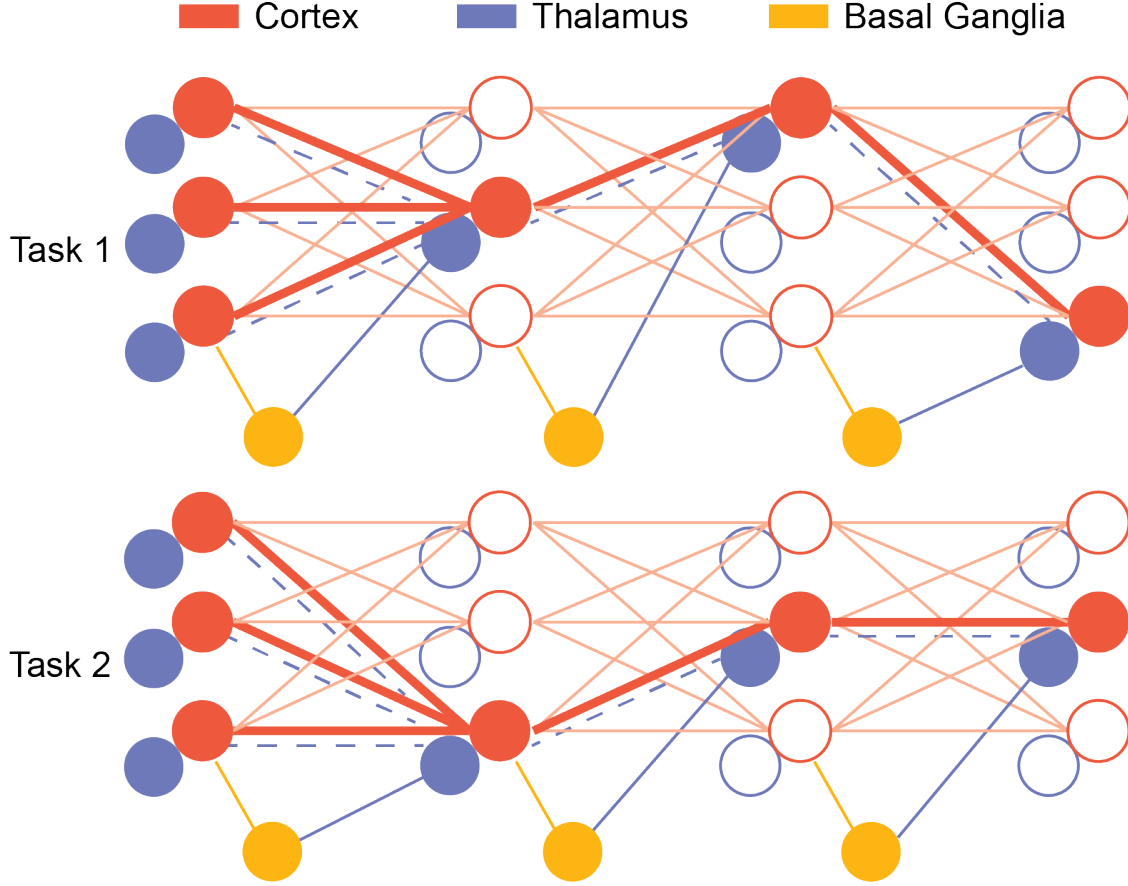


Figure 6: **A thalamocortical architecture with interaction with BG for contextual credit assignment** During task execution, BG selects thalamic neurons that amplify the relevant cortical subnetwork. This protects other parts of the network that are important for another context from being overwritten. When the other task comes, BG selects other thalamic neurons and since the synapses are protected from the last task, animals can freely switch from different tasks without forgetting the previous tasks. Furthermore, as the corticothalamic synapses learn how to select the right thalamic neurons in a different context (blue dash line), task execution can become less BG dependent.

can be easily generalized.

The idea of protecting relevant information from the past tasks to be overwritten has been applied before computationally and has decent success in combating catastrophic forgetting in deep learning [139]. Experimentally, we also have found thalamic neurons selectively amplify the cortical connectivity to solve the contextual credit assignment problem. In a task where the mice need to switch between different sets of task cues that guided the attention to the visual or auditory target, the performance of the mice does not deteriorate much after switching to the original context which is an indication of contextual credit assignment [113]. Electrophysiological recording of PFC and mediodorsal thalamic nucleus (MD) neurons, we discovered that PFC neurons preferentially code for the rule of the attention while MD neurons preferentially code for the contexts of different sets of the cues. Thalamic neurons that encode the task-relevant context translate this neural representation into the

amplification of cortical activity patterns associated with that context (despite the fact that cortical neurons themselves only encode the context implicitly). These experimental observations are consistent with our proposed solution: by incorporating the thalamic population that can selectively amplify connectivity of cortical subnetworks, the thalamus and its interaction with cortex and BG solve the contextual credit assignment problem and prevent catastrophic forgetting.

Thalamus’s role as a flexible gate for working memory makes temporal credit assignment more efficient. Finally, in addition to structural credit assignment and contextual credit assignment, an animal also needs to solve temporal credit assignment, the ability to attribute the distant reward to the relevant past actions. One well-known mechanism to solve this problem is through temporal difference reward prediction error in the ventral tegmental area (VTA) of BG. Since the TD RPE captures the value function of discounted future reward, an organism does not need to wait until reward delivery to be able to credit the actions that lead to positive TD RPE. However, TD learning assumes that the state transition and the reward depend only on the current state and the current action while in a natural environment, many rewards depend on not only the current sensory inputs but also the past sensory inputs. One possible solution is to store all past sensory inputs and actions in the current state. However, this is highly inefficient since not all past information is relevant to the task. Therefore, the ability to store the relevant sensory inputs and actions in the working memory is important to solve temporal credit assignments in a natural environment. One such solution is LSTM network [56–58] which selectively maintains past sensory inputs in the hidden memory. However, since one needs to use backpropagation to train the network, based on everything we have discussed so far, the brain ought to have an alternative solution.

We propose that the thalamus is also an integral component of the brain’s solution to temporal credit assignment. By selectively changing the effective time constant of the working memory in cortical recurrent networks, thalamic outputs can precisely accomplish this computational objective. Specifically, when the sensory inputs arrive at prefrontal cortex (PFC), BG uses reinforcement learning to set thalamic activity patterns thereby selectively amplifying the appropriate cortical associations relevant to the task’s working memory component (Figure 7). Mathematically, we can consider the following simplified model of PFC

$$\tau \frac{dr_{PFC}}{dt} = -r_{PFC} + f(inp_{PFC}), \quad inp_{PFC} = (1 + r_{MD})w_{PFC}r_{PFC} + r_s \quad (10)$$

where τ is the time constant, r_{PFC} is the PFC activity, $f(x) = (1 + e^{-x})^{-1}$ is the sigmoid function, r_{MD} is the input from MD to PFC, w_{PFC} is the recurrent weight in PFC and r_s is the input from sensory cortex. By linearizing the dynamic, we can derive the effective time constant of the working memory as

$$\tau_{eff} = \frac{\tau}{1 - (1 + r_{MD})w_{PFC}f(inp_{PFC})(1 - f(inp_{PFC}))} \quad (11)$$

where larger r_{MD} creates longer effective time constant in memory. Because of the reinforcement learning in BG to set the activity of r_{MD} , PFC is able to selectively store the task

relevant sensory inputs into working memory. Furthermore, as the learning consolidates to corticothalamic connections, one then needs less BG dependency on adjusting the time constant of the working memory. By selectively maintaining the task-relevant information in working memory, one is able to credit the past information at the time when the reward is delivered.

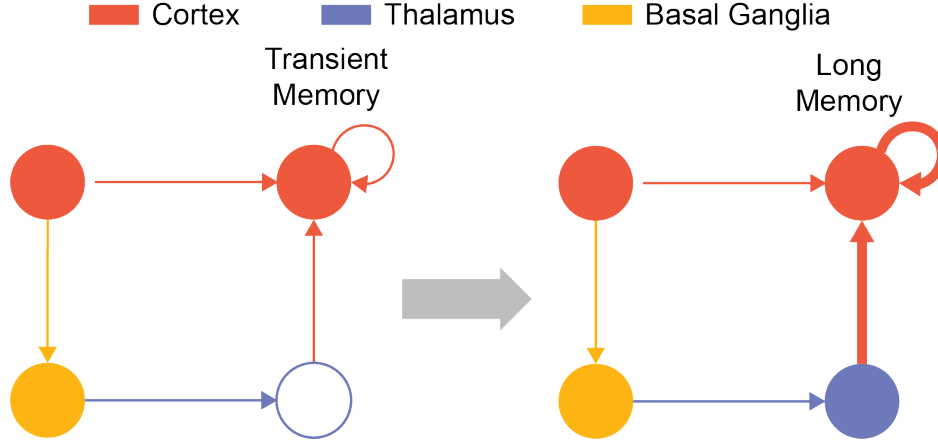


Figure 7: **BG uses reinforcement learning to adjust the time constant of the working memory** By inhibiting the specific thalamus that amplifies the connectivity of the working memory, BG gates the information from storing in the working memory. On the other hand, by disinhibiting the thalamus, BG selectively adjusts the time constant of the working memory.

The idea that the coordination between PFC, thalamus and BG selectively gates working memory has been explored computationally before and shows success in simple working memory tasks [156]. Experimentally, it is also well known that both thalamus [109–111] and basal ganglia [129, 130] are involved in cortical working memory. In our proposal, we further refine the idea by allowing thalamic control functions to fine-tune the time constant of the working memory through amplifying the cortical connectivity [110, 113, 114]. By incorporating these thalamic populations that selectively adjust the effective time constant of cortical working memory into our framework, the thalamocortical-basal ganglia system can flexibly maintain the relevant sensory inputs to solve temporal credit assignment efficiently.

Summary In summary, in contrast to the traditional relay view of the thalamus, we propose that thalamocortical interaction is the locus of meta-learning where the thalamus provides cortical control functions, such as sensory filtering, working memory gating or motor preparation, that parametrize the cortical activity association space. Furthermore, we propose a two timescale learning consolidation framework where BG hierarchically selects these thalamic control functions to enable meta-learning, solving the credit assignment problem. The faster plasticity learns contextual associations to enable rapid behavioral flexibility while the slower plasticity establishes cortical representation that generalizes. Additionally, by considering the ability of the thalamus to selectively amplify functional cortical connectivity, the thalamocortical-basal ganglia network is able to flexibly learn context-dependent association without catastrophic forgetting while generalizing to the new contexts. Lastly,

by considering thalamic control functions in adjusting the time constant of cortical working memory, our framework provides an efficient solution to temporal credit assignment as well.

References

- [1] E. Thorndike, *Animal Intelligence: Experimental Studies*. Taylor & Francis, 2017.
- [2] E. K. Miller and J. D. Cohen, “An integrative theory of prefrontal cortex function,” *Annu Rev Neurosci*, vol. 24, pp. 167–202, 2001.
- [3] Y. Niv, “Reinforcement learning in the brain,” *Journal of Mathematical Psychology*, vol. 53, pp. 139–154, June 2009.
- [4] D. Hebb, *The Organization of Behavior: A Neuropsychological Theory*. Taylor & Francis, 2002.
- [5] T. V. Bliss and T. Lomo, “Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path,” *J Physiol*, vol. 232, pp. 331–356, Jul 1973.
- [6] L. F. Abbott and S. B. Nelson, “Synaptic plasticity: taming the beast,” *Nat Neurosci*, vol. 3 Suppl, pp. 1178–1183, Nov 2000.
- [7] P. Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, 2005.
- [8] J. C. R. Whittington and R. Bogacz, “Theories of Error Back-Propagation in the Brain,” *Trends Cogn Sci*, vol. 23, pp. 235–250, Mar 2019.
- [9] P. R. Montague, P. Dayan, and T. J. Sejnowski, “A framework for mesencephalic dopamine systems based on predictive Hebbian learning,” *J Neurosci*, vol. 16, pp. 1936–1947, Mar 1996.
- [10] W. Schultz, P. Dayan, and P. R. Montague, “A neural substrate of prediction and reward,” *Science*, vol. 275, pp. 1593–1599, Mar 1997.
- [11] H. M. Bayer and P. W. Glimcher, “Midbrain dopamine neurons encode a quantitative reward prediction error signal,” *Neuron*, vol. 47, pp. 129–141, Jul 2005.
- [12] N. S. Bamford, R. M. Wightman, and D. Sulzer, “Dopamine’s Effects on Corticostriatal Synapses during Reward-Based Behaviors,” *Neuron*, vol. 97, pp. 494–510, Feb 2018.
- [13] J. C. Houk, J. L. Davis, and D. G. Beiser, *Adaptive Critics and the Basal Ganglia*, pp. 215–232. 1994.
- [14] R. E. Suri and W. Schultz, “A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task,” *Neuroscience*, vol. 91, no. 3, pp. 871–890, 1999.

- [15] R. S. Sutton and A. G. Barto, “Time-derivative models of pavlovian reinforcement,” in *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, pp. 497–537, MIT Press, 1990.
- [16] J. R. Wickens and R. Kotter, “Cellular Models of Reinforcement,” in *Models of Information Processing in the Basal Ganglia*, The MIT Press, Nov 1994.
- [17] G. Morris, A. Nevet, D. Arkadir, E. Vaadia, and H. Bergman, “Midbrain dopamine neurons encode decisions for future action,” *Nat Neurosci*, vol. 9, pp. 1057–1063, Aug 2006.
- [18] M. R. Roesch, D. J. Calu, and G. Schoenbaum, “Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards,” *Nat Neurosci*, vol. 10, pp. 1615–1624, Dec 2007.
- [19] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning series, MIT Press, 2018.
- [20] S. Ikemoto and J. Panksepp, “The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking,” *Brain Res Brain Res Rev*, vol. 31, pp. 6–41, Dec 1999.
- [21] M. Minsky, “Steps toward artificial intelligence,” *Proceedings of the IRE*, vol. 49, no. 1, pp. 8–30, 1961.
- [22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, pp. 533–536, Oct. 1986.
- [23] T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, and G. Hinton, “Backpropagation and the brain,” *Nat Rev Neurosci*, vol. 21, pp. 335–346, Jun 2020.
- [24] R. C. O’Reilly, “Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm,” *Neural Computation*, vol. 8, no. 5, pp. 895–938, 1996.
- [25] P. R. Roelfsema and A. van Ooyen, “Attention-gated reinforcement learning of internal representations for classification,” *Neural Comput*, vol. 17, pp. 2176–2214, Oct 2005.
- [26] I. R. Fiete and H. S. Seung, “Gradient learning in spiking neural networks by dynamic perturbation of conductances,” *Phys. Rev. Lett.*, vol. 97, p. 048104, Jul 2006.
- [27] N. Ketz, S. G. Morkonda, and R. C. O’Reilly, “Theta coordinated error-driven learning in the hippocampus,” *PLOS Computational Biology*, vol. 9, pp. 1–9, Jun 2013.
- [28] T. P. Lillicrap, D. Cownden, D. B. Tweed, and C. J. Akerman, “Random synaptic feedback weights support error backpropagation for deep learning,” *Nat Commun*, vol. 7, p. 13276, Nov 2016.

- [29] M. Schiess, R. Urbanczik, and W. Senn, “Somato-dendritic synaptic plasticity and error-backpropagation in active dendrites,” *PLOS Computational Biology*, vol. 12, pp. 1–18, Feb 2016.
- [30] L. Kusmierz, T. Isomura, and T. Toyoizumi, “Learning with three factors: modulating Hebbian plasticity with errors,” *Curr Opin Neurobiol*, vol. 46, pp. 170–177, Oct 2017.
- [31] F. Zenke and S. Ganguli, “SuperSpike: Supervised Learning in Multilayer Spiking Neural Networks,” *Neural Comput*, vol. 30, pp. 1514–1541, Jun 2018.
- [32] B. A. Richards and T. P. Lillicrap, “Dendritic solutions to the credit assignment problem,” *Curr Opin Neurobiol*, vol. 54, pp. 28–36, Feb 2019.
- [33] P. R. Roelfsema and A. Holtmaat, “Control of synaptic plasticity in deep cortical networks,” *Nat Rev Neurosci*, vol. 19, pp. 166–180, Feb 2018.
- [34] J. Sacramento, R. Ponte Costa, Y. Bengio, and W. Senn, “Dendritic cortical microcircuits approximate the backpropagation algorithm,” in *Advances in Neural Information Processing Systems*, vol. 31, pp. 8735–8746, Curran Associates, Inc., 2018.
- [35] J. Kornfeld, M. Januszewski, P. Schubert, V. Jain, W. Denk, and M. Fee, “An anatomical substrate of credit assignment in reinforcement learning,” *bioRxiv*, 2020.
- [36] Y. H. Liu, S. Smith, S. Mihalas, E. Shea-Brown, and U. Smbl, “A solution to temporal credit assignment using cell-type-specific modulatory signals,” *bioRxiv*, 2020.
- [37] R. C. O’Reilly, J. L. Russin, M. Zolfaghar, and J. Rohrlich, “Deep predictive learning in neocortex and pulvinar,” 2021.
- [38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., 2012.
- [39] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [40] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, pp. 529–533, Feb 2015.
- [41] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the game of Go with deep neural networks and tree search,” *Nature*, vol. 529, pp. 484–489, Jan 2016.

- [42] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, “Mastering the game of Go without human knowledge,” *Nature*, vol. 550, pp. 354–359, Oct 2017.
- [43] J. Schrittwieser, I. Antonoglou, T. Hubert, K. Simonyan, L. Sifre, S. Schmitt, A. Guez, E. Lockhart, D. Hassabis, T. Graepel, T. Lillicrap, and D. Silver, “Mastering Atari, Go, chess and shogi by planning with a learned model,” *Nature*, vol. 588, pp. 604–609, Dec 2020.
- [44] C. F. Cadieu, H. Hong, D. L. K. Yamins, N. Pinto, D. Ardila, E. A. Solomon, N. J. Majaj, and J. J. DiCarlo, “Deep neural networks rival the representation of primate it cortex for core visual object recognition,” *PLOS Computational Biology*, vol. 10, pp. 1–18, Dec 2014.
- [45] D. L. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, “Performance-optimized hierarchical models predict neural responses in higher visual cortex,” *Proc Natl Acad Sci U S A*, vol. 111, pp. 8619–8624, Jun 2014.
- [46] F. Crick, “The recent excitement about neural networks,” *Nature*, vol. 337, pp. 129–132, Jan 1989.
- [47] M. McCloskey and N. J. Cohen, “Catastrophic interference in connectionist networks: The sequential learning problem,” vol. 24 of *Psychology of Learning and Motivation*, pp. 109–165, Academic Press, 1989.
- [48] R. M. French, “Catastrophic forgetting in connectionist networks,” *Trends Cogn Sci*, vol. 3, pp. 128–135, Apr 1999.
- [49] D. Kumaran, D. Hassabis, and J. L. McClelland, “What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated,” *Trends Cogn Sci*, vol. 20, pp. 512–534, 07 2016.
- [50] R. Kemker, M. McClure, A. Abitino, T. Hayes, and C. Kanan, “Measuring catastrophic forgetting in neural networks,” in *AAAI Conference on Artificial Intelligence*, 2018.
- [51] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, “Continual lifelong learning with neural networks: A review,” *Neural Netw*, vol. 113, pp. 54–71, May 2019.
- [52] J. L. McClelland, B. L. McNaughton, and R. C. O’Reilly, “Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory,” *Psychol Rev*, vol. 102, pp. 419–457, Jul 1995.
- [53] M. M. Murray, D. J. Lewkowicz, A. Amedi, and M. T. Wallace, “Multisensory Processes: A Balancing Act across the Lifespan,” *Trends Neurosci*, vol. 39, pp. 567–579, Aug 2016.

- [54] J. D. Power and B. L. Schlaggar, “Neural plasticity across the lifespan,” *Wiley Interdiscip Rev Dev Biol*, vol. 6, Jan 2017.
- [55] F. Zenke, W. Gerstner, and S. Ganguli, “The temporal paradox of Hebbian learning and homeostatic plasticity,” *Curr Opin Neurobiol*, vol. 43, pp. 166–176, Apr 2017.
- [56] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, Nov. 1997.
- [57] Y. Yu, X. Si, C. Hu, and J. Zhang, “A review of recurrent neural networks: Lstm cells and network architectures,” *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
- [58] G. V. Houdt, C. Mosquera, and G. Nápoles, “A review on the long short-term memory model,” *Artificial Intelligence Review*, vol. 53, pp. 5929–5955, May 2020.
- [59] G. E. Alexander, M. R. DeLong, and P. L. Strick, “Parallel organization of functionally segregated circuits linking basal ganglia and cortex,” *Annu Rev Neurosci*, vol. 9, pp. 357–381, 1986.
- [60] E. K. Miller, “The prefrontal cortex and cognitive control,” *Nature Reviews Neuroscience*, vol. 1, pp. 59–65, Oct. 2000.
- [61] D. Badre, A. S. Kayser, and M. D’Esposito, “Frontal cortex and the discovery of abstract action rules,” *Neuron*, vol. 66, pp. 315–326, Apr 2010.
- [62] M. Seo, E. Lee, and B. B. Averbeck, “Action selection and action value in frontostriatal circuits,” *Neuron*, vol. 74, pp. 947–960, Jun 2012.
- [63] H. Makino, E. J. Hwang, N. G. Hedrick, and T. Komiyama, “Circuit Mechanisms of Sensorimotor Learning,” *Neuron*, vol. 92, pp. 705–721, Nov 2016.
- [64] J. Cox and I. B. Witten, “Striatal circuits for reward learning and decision-making,” *Nat Rev Neurosci*, vol. 20, pp. 482–494, Aug 2019.
- [65] M. Wolff and S. D. Vann, “The Cognitive Thalamus as a Gateway to Mental Representations,” *J Neurosci*, vol. 39, pp. 3–14, Jan 2019.
- [66] P. Rakic, “Evolution of the neocortex: a perspective from developmental biology,” *Nat Rev Neurosci*, vol. 10, pp. 724–735, Oct 2009.
- [67] J. Fuster, *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. Lippincott-Raven, 1997.
- [68] W. Singer, T. Sejnowski, and P. Rakic, *The Neocortex*. Strüngmann Forum Reports, MIT Press, 2019.
- [69] E. G. Jones, ed., *The Thalamus*. Springer US, 1985.
- [70] S. M. Sherman and R. W. Guillery, *Exploring the Thalamus and Its Role in Cortical Function, Second Edition*. The MIT Press, hardcover ed., Dec 2005.

- [71] J. A. Harris, S. Mihalas, K. E. Hirokawa, J. D. Whitesell, H. Choi, A. Bernard, P. Bohn, S. Caldejon, L. Casal, A. Cho, A. Feiner, D. Feng, N. Gaudreault, C. R. Gerfen, N. Graddis, P. A. Groblewski, A. M. Henry, A. Ho, R. Howard, J. E. Knox, L. Kuan, X. Kuang, J. Lecoq, P. Lesnar, Y. Li, J. Luviano, S. McConoughey, M. T. Mortrud, M. Naeemi, L. Ng, S. W. Oh, B. Ouellette, E. Shen, S. A. Sorensen, W. Wakeman, Q. Wang, Y. Wang, A. Williford, J. W. Phillips, A. R. Jones, C. Koch, and H. Zeng, “Hierarchical organization of cortical and thalamic connectivity,” *Nature*, vol. 575, pp. 195–202, Nov 2019.
- [72] C. Gerfen and J. Bolam, “The neuroanatomical organization of the basal ganglia,” *Handbook of Behavioral Neuroscience*, vol. 20, pp. 3–28, Dec 2010.
- [73] A. Nambu, “Somatotopic organization of the primate Basal Ganglia,” *Front Neuroanat*, vol. 5, p. 26, 2011.
- [74] J. L. Lanciego, N. Luquin, and J. A. Obeso, “Functional neuroanatomy of the basal ganglia,” *Cold Spring Harb Perspect Med*, vol. 2, p. a009621, Dec 2012.
- [75] K. Doya, “What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?,” *Neural Netw*, vol. 12, pp. 961–974, Oct 1999.
- [76] K. Doya, “Complementary roles of basal ganglia and cerebellum in learning and motor control,” *Curr Opin Neurobiol*, vol. 10, pp. 732–739, Dec 2000.
- [77] O. Hikosaka, H. F. Kim, M. Yasuda, and S. Yamamoto, “Basal ganglia circuits for reward value-guided behavior,” *Annu Rev Neurosci*, vol. 37, pp. 289–306, 2014.
- [78] E. Perrin and L. Venance, “Bridging the gap between striatal plasticity and learning,” *Curr Opin Neurobiol*, vol. 54, pp. 104–112, Feb 2019.
- [79] P. A. Garriss and R. M. Wightman, “Different kinetics govern dopaminergic transmission in the amygdala, prefrontal cortex, and striatum: an in vivo voltammetric study,” *J Neurosci*, vol. 14, pp. 442–450, Jan 1994.
- [80] B. J. Ciliax, C. Heilman, L. L. Demchyshyn, Z. B. Pristupa, E. Ince, S. M. Hersch, H. B. Niznik, and A. I. Levey, “The dopamine transporter: immunochemical characterization and localization in brain,” *J Neurosci*, vol. 15, pp. 1714–1723, Mar 1995.
- [81] W. A. Cass and G. A. Gerhardt, “In vivo assessment of dopamine uptake in rat medial prefrontal cortex: comparison with dorsal striatum and nucleus accumbens,” *J Neurochem*, vol. 65, pp. 201–207, Jul 1995.
- [82] C. C. Lapish, S. Kroener, D. Durstewitz, A. Lavin, and J. K. Seamans, “The ability of the mesocortical dopamine system to operate in distinct temporal modes,” *Psychopharmacology (Berl)*, vol. 191, pp. 609–625, Apr 2007.
- [83] J. K. Seamans and T. W. Robbins, *Dopamine Modulation of the Prefrontal Cortex and Cognitive Function*, pp. 373–398. Totowa, NJ: Humana Press, 2010.

- [84] A. Kirkwood, M. C. Rioult, and M. F. Bear, “Experience-dependent modification of synaptic plasticity in visual cortex,” *Nature*, vol. 381, pp. 526–528, Jun 1996.
- [85] D. E. Feldman, “Synaptic mechanisms for plasticity in neocortex,” *Annu Rev Neurosci*, vol. 32, pp. 33–55, 2009.
- [86] S. F. Cooke and M. F. Bear, “Visual experience induces long-term potentiation in the primary visual cortex,” *J Neurosci*, vol. 30, pp. 16304–16313, Dec 2010.
- [87] C. H. Donahue and D. Lee, “Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex,” *Nat Neurosci*, vol. 18, pp. 295–301, Feb 2015.
- [88] K. Tsutsui, T. Hosokawa, M. Yamada, and T. Iijima, “Representation of Functional Category in the Monkey Prefrontal Cortex and Its Rule-Dependent Use for Behavioral Selection,” *J Neurosci*, vol. 36, pp. 3038–3048, Mar 2016.
- [89] W. E. Allen, I. V. Kauvar, M. Z. Chen, E. B. Richman, S. J. Yang, K. Chan, V. Gradinaru, B. E. Deverman, L. Luo, and K. Deisseroth, “Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex,” *Neuron*, vol. 94, pp. 891–907, May 2017.
- [90] C. C. H. Petersen, “Sensorimotor processing in the rodent barrel cortex,” *Nat Rev Neurosci*, vol. 20, pp. 533–546, Sep 2019.
- [91] P. Enel, J. D. Wallis, and E. L. Rich, “Stable and dynamic representations of value in the prefrontal cortex,” *Elife*, vol. 9, Jul 2020.
- [92] D. S. Jacobs and B. Moghaddam, “Prefrontal Cortex Representation of Learning of Punishment Probability During Reward-Motivated Actions,” *J Neurosci*, vol. 40, pp. 5063–5077, Jun 2020.
- [93] F. G. Ashby, J. M. Ennis, and B. J. Spiering, “A neurobiological theory of automaticity in perceptual categorization,” *Psychol Rev*, vol. 114, pp. 632–656, Jul 2007.
- [94] A. S. Andalman and 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*, vol. 106, pp. 12518–12523, Jul 2009.
- [95] T. L. Warren, E. C. Tumer, J. D. Charlesworth, and M. S. Brainard, “Mechanisms and time course of vocal learning and consolidation in the adult songbird,” *J Neurophysiol*, vol. 106, pp. 1806–1821, Oct 2011.
- [96] S. Hélie, S. W. Ell, and F. G. Ashby, “Learning robust cortico-cortical associations with the basal ganglia: an integrative review,” *Cortex*, vol. 64, pp. 123–135, Mar 2015.
- [97] T. Tesileanu, B. Olveczky, and V. Balasubramanian, “Rules and mechanisms for efficient two-stage learning in neural circuits,” *Elife*, vol. 6, Apr 2017.

- [98] I. R. Fiete, M. S. Fee, and H. S. Seung, “Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances,” *J Neurophysiol*, vol. 98, pp. 2038–2057, Oct 2007.
- [99] M. S. Fee and J. H. Goldberg, “A hypothesis for basal ganglia-dependent reinforcement learning in the songbird,” *Neuroscience*, vol. 198, pp. 152–170, Dec 2011.
- [100] C. Kim, N. F. Johnson, S. E. Cilles, and B. T. Gold, “Common and distinct mechanisms of cognitive flexibility in prefrontal cortex,” *J Neurosci*, vol. 31, pp. 4771–4779, Mar 2011.
- [101] V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome, “Context-dependent computation by recurrent dynamics in prefrontal cortex,” *Nature*, vol. 503, pp. 78–84, Nov 2013.
- [102] J. X. Wang, Z. Kurth-Nelson, D. Kumaran, D. Tirumala, H. Soyer, J. Z. Leibo, D. Hassabis, and M. Botvinick, “Prefrontal cortex as a meta-reinforcement learning system,” *Nat Neurosci*, vol. 21, pp. 860–868, Jun 2018.
- [103] M. Botvinick, S. Ritter, J. X. Wang, Z. Kurth-Nelson, C. Blundell, and D. Hassabis, “Reinforcement Learning, Fast and Slow,” *Trends Cogn Sci*, vol. 23, pp. 408–422, May 2019.
- [104] H. Sohn, N. Meirhaeghe, R. Rajalingham, and M. Jazayeri, “A Network Perspective on Sensorimotor Learning,” *Trends Neurosci*, vol. 44, pp. 170–181, 03 2021.
- [105] M. Tanaka, “Cognitive signals in the primate motor thalamus predict saccade timing,” *J Neurosci*, vol. 27, pp. 12109–12118, Oct 2007.
- [106] Y. B. Saalman and S. Kastner, “The cognitive thalamus,” *Front Syst Neurosci*, vol. 9, p. 39, 2015.
- [107] R. D. Wimmer, L. I. Schmitt, T. J. Davidson, M. Nakajima, K. Deisseroth, and M. M. Halassa, “Thalamic control of sensory selection in divided attention,” *Nature*, vol. 526, pp. 705–709, Oct 2015.
- [108] H. Zhou, R. J. Schafer, and R. Desimone, “Pulvinar-Cortex Interactions in Vision and Attention,” *Neuron*, vol. 89, pp. 209–220, Jan 2016.
- [109] S. S. Bolkan, J. M. Stujenske, S. Parnaudeau, T. J. Spellman, C. Rauffenbart, A. I. Abbas, A. Z. Harris, J. A. Gordon, and C. Kellendonk, “Thalamic projections sustain prefrontal activity during working memory maintenance,” *Nat Neurosci*, vol. 20, pp. 987–996, Jul 2017.
- [110] L. I. Schmitt, R. D. Wimmer, M. Nakajima, M. Happ, S. Mofakham, and M. M. Halassa, “Thalamic amplification of cortical connectivity sustains attentional control,” *Nature*, vol. 545, pp. 219–223, May 2017.

- [111] Z. V. Guo, H. K. Inagaki, K. Daie, S. Druckmann, C. R. Gerfen, and K. Svoboda, “Maintenance of persistent activity in a frontal thalamocortical loop,” *Nature*, vol. 545, no. 7653, pp. 181–186, 2017.
- [112] W. Guo, A. R. Clause, A. Barth-Maron, and D. B. Polley, “A Corticothalamic Circuit for Dynamic Switching between Feature Detection and Discrimination,” *Neuron*, vol. 95, pp. 180–194, Jul 2017.
- [113] R. V. Rikhye, A. Gilra, and M. M. Halassa, “Thalamic regulation of switching between cortical representations enables cognitive flexibility,” *Nat Neurosci*, vol. 21, pp. 1753–1763, Dec 2018.
- [114] A. Mukherjee, N. Bajwa, N. H. Lam, C. Porrero, F. Clasca, and M. M. Halassa, “Variation of connectivity across exemplar sensory and associative thalamocortical loops in the mouse,” *Elife*, vol. 9, Oct 2020.
- [115] B. J. Hunnicutt, B. C. Jongbloets, W. T. Birdsong, K. J. Gertz, H. Zhong, and T. Mao, “A comprehensive excitatory input map of the striatum reveals novel functional organization,” *Elife*, vol. 5, Nov 2016.
- [116] H. Jiang and H. F. Kim, “Anatomical Inputs From the Sensory and Value Structures to the Tail of the Rat Striatum,” *Front Neuroanat*, vol. 12, p. 30, 2018.
- [117] M. Nakajima, L. I. Schmitt, and M. M. Halassa, “Prefrontal Cortex Regulates Sensory Filtering through a Basal Ganglia-to-Thalamus Pathway,” *Neuron*, vol. 103, pp. 445–458, Aug 2019.
- [118] A. J. Peters, J. M. J. Fabre, N. A. Steinmetz, K. D. Harris, and M. Carandini, “Striatal activity topographically reflects cortical activity,” *Nature*, Jan 2021.
- [119] T. F. Marton, H. Seifkar, F. J. Luongo, A. T. Lee, and V. S. Sohal, “Roles of Prefrontal Cortex and Mediodorsal Thalamus in Task Engagement and Behavioral Flexibility,” *J Neurosci*, vol. 38, pp. 2569–2578, Mar 2018.
- [120] D. H. Hubel and T. N. Wiesel, “Integrative action in the cat’s lateral geniculate body,” *J Physiol*, vol. 155, pp. 385–398, Feb 1961.
- [121] S. M. Sherman and P. D. Spear, “Organization of visual pathways in normal and visually deprived cats,” *Physiol Rev*, vol. 62, pp. 738–855, Apr 1982.
- [122] P. Reinagel, D. Godwin, S. M. Sherman, and C. Koch, “Encoding of visual information by LGN bursts,” *J Neurophysiol*, vol. 81, pp. 2558–2569, May 1999.
- [123] W. M. Usrey, J. M. Alonso, and R. C. Reid, “Synaptic interactions between thalamic inputs to simple cells in cat visual cortex,” *J Neurosci*, vol. 20, pp. 5461–5467, Jul 2000.
- [124] A. D. Lien and M. Scanziani, “Cortical direction selectivity emerges at convergence of thalamic synapses,” *Nature*, vol. 558, pp. 80–86, Jun 2018.

- [125] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *J Physiol*, vol. 160, pp. 106–154, Jan 1962.
- [126] K. P. Hoffmann, J. Stone, and S. M. Sherman, “Relay of receptive-field properties in dorsal lateral geniculate nucleus of the cat,” *J Neurophysiol*, vol. 35, pp. 518–531, Jul 1972.
- [127] J. M. Phillips, N. A. Kambi, and Y. B. Saalman, “A Subcortical Pathway for Rapid, Goal-Driven, Attentional Filtering,” *Trends Neurosci*, vol. 39, pp. 49–51, Feb 2016.
- [128] O. Monchi, M. Petrides, A. P. Strafella, K. J. Worsley, and J. Doyon, “Functional role of the basal ganglia in the planning and execution of actions,” *Ann Neurol*, vol. 59, pp. 257–264, Feb 2006.
- [129] F. McNab and T. Klingberg, “Prefrontal cortex and basal ganglia control access to working memory,” *Nat Neurosci*, vol. 11, pp. 103–107, Jan 2008.
- [130] B. Voytek and R. T. Knight, “Prefrontal cortex and basal ganglia contributions to visual working memory,” *Proc Natl Acad Sci U S A*, vol. 107, pp. 18167–18172, Oct 2010.
- [131] W. H. Mehafeey and A. J. Doupe, “Naturalistic stimulation drives opposing heterosynaptic plasticity at two inputs to songbird cortex,” *Nat Neurosci*, vol. 18, pp. 1272–1280, Sep 2015.
- [132] F. Sohrabji, E. J. Nordeen, and K. W. Nordeen, “Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch,” *Behav Neural Biol*, vol. 53, pp. 51–63, Jan 1990.
- [133] C. Scharff and F. Nottebohm, “A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning,” *J Neurosci*, vol. 11, pp. 2896–2913, Sep 1991.
- [134] P. Soliveri, R. G. Brown, M. Jahanshahi, T. Caraceni, and C. D. Marsden, “Learning manual pursuit tracking skills in patients with Parkinson’s disease,” *Brain*, vol. 120 (Pt 8), pp. 1325–1337, Aug 1997.
- [135] V. Thomas-Ollivier, J. M. Reymann, S. Le Moal, S. Schück, A. Lieury, and H. Alain, “Procedural memory in recent-onset Parkinson’s disease,” *Dement Geriatr Cogn Disord*, vol. 10, no. 2, pp. 172–180, 1999.
- [136] F. Asmus, H. Huber, T. Gasser, and L. Schöls, “Kick and rush: paradoxical kinesia in Parkinson disease,” *Neurology*, vol. 71, p. 695, Aug 2008.
- [137] E. G. Jones, “The thalamic matrix and thalamocortical synchrony,” *Trends Neurosci*, vol. 24, pp. 595–601, Oct 2001.
- [138] D. J. Lewkowicz, “Early experience and multisensory perceptual narrowing,” *Dev Psychobiol*, vol. 56, pp. 292–315, Feb 2014.

- [139] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, “Overcoming catastrophic forgetting in neural networks,” *Proc Natl Acad Sci U S A*, vol. 114, pp. 3521–3526, 03 2017.
- [140] F. Zenke, B. Poole, and S. Ganguli, “Continual learning through synaptic intelligence,” in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70 of *Proceedings of Machine Learning Research*, pp. 3987–3995, Aug 2017.
- [141] C. Fernando, D. Banarse, C. Blundell, Y. Zwols, D. Ha, A. A. Rusu, A. Pritzel, and D. Wierstra, “Pathnet: Evolution channels gradient descent in super neural networks,” *CoRR*, vol. abs/1701.08734, 2017.
- [142] Z. Li and D. Hoiem, “Learning without forgetting,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2935–2947, 2018.
- [143] H. Jung, J. Ju, M. Jung, and J. Kim, “Less-forgetful learning for domain expansion in deep neural networks,” in *AAAI Conference on Artificial Intelligence*, 2018.
- [144] D. Maltoni and V. Lomonaco, “Continuous learning in single-incremental-task scenarios,” *Neural Netw*, vol. 116, pp. 56–73, Aug 2019.
- [145] S. Fusi, P. J. Drew, and L. F. Abbott, “Cascade models of synaptically stored memories,” *Neuron*, vol. 45, pp. 599–611, Feb 2005.
- [146] G. Yang, F. Pan, and W. B. Gan, “Stably maintained dendritic spines are associated with lifelong memories,” *Nature*, vol. 462, pp. 920–924, Dec 2009.
- [147] J. Cichon and W. B. Gan, “Branch-specific dendritic Ca(2+) spikes cause persistent synaptic plasticity,” *Nature*, vol. 520, pp. 180–185, Apr 2015.
- [148] A. Hayashi-Takagi, S. Yagishita, M. Nakamura, F. Shirai, Y. I. Wu, A. L. Loshbaugh, B. Kuhlman, K. M. Hahn, and H. Kasai, “Labelling and optical erasure of synaptic memory traces in the motor cortex,” *Nature*, vol. 525, pp. 333–338, Sep 2015.
- [149] M. K. Benna and S. Fusi, “Computational principles of synaptic memory consolidation,” *Nat Neurosci*, vol. 19, pp. 1697–1706, Dec 2016.
- [150] T. Xiao, J. Zhang, K. Yang, Y. Peng, and Z. Zhang, “Error-driven incremental learning in deep convolutional neural network for large-scale image classification,” in *ACM Multimedia*, Nov 2014.
- [151] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell, “Progressive neural networks,” *CoRR*, vol. abs/1606.04671, 2016.
- [152] T. J. Draelos, N. E. Miner, C. C. Lamb, J. A. Cox, C. M. Vineyard, K. D. Carlson, W. M. Severa, C. D. James, and J. B. Aimone, “Neurogenesis deep learning: Extending deep networks to accommodate new classes,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 526–533, 2017.

- [153] C. Cortes, X. Gonzalvo, V. Kuznetsov, M. Mohri, and S. Yang, “AdaNet: Adaptive structural learning of artificial neural networks,” in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70 of *Proceedings of Machine Learning Research*, pp. 874–883, Aug 2017.
- [154] H. Shin, J. K. Lee, J. Kim, and J. Kim, “Continual learning with deep generative replay,” in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.
- [155] R. Kemker and C. Kanan, “Fearnnet: Brain-inspired model for incremental learning,” in *International Conference on Learning Representations*, 2018.
- [156] R. C. O'Reilly and M. J. Frank, “Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia,” *Neural Computation*, vol. 18, pp. 283–328, Feb. 2006.