Neural Networks 144 (2021) 507–521

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2021 Special Issue on AI and Brain Science: AI-powered Brain Science

Parallel and hierarchical neural mechanisms for adaptive and predictive behavioral control

Tom Macpherson [a,1], Masayuki Matsumoto [b,1], Hiroaki Gomi [c,1], Jun Morimoto [d,e,1], Eiji Uchibe [d,1], Takatoshi Hikida [a,*,1]

[a] Laboratory for Advanced Brain Functions, Institute for Protein Research, Osaka University, Osaka, Japan
[b] Division of Biomedical Science, Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan
[c] NTT Communication Science Laboratories, Nippon Telegraph and Telephone Co., Kanagawa, Japan
[d] Department of Brain Robot Interface, ATR Computational Neuroscience Laboratories, Kyoto, Japan
[e] Graduate School of Informatics, Kyoto University, Kyoto, Japan

ARTICLE INFO

Article history:
Available online 17 September 2021

Keywords:
Parallel processing
Hierarchical processing
Behavioral flexibility
Movement control
Artificial intelligence
Humanoid robotics

ABSTRACT

Our brain can be recognized as a network of largely hierarchically organized neural circuits that operate to control specific functions, but when acting in parallel, enable the performance of complex and simultaneous behaviors. Indeed, many of our daily actions require concurrent information processing in sensorimotor, associative, and limbic circuits that are dynamically and hierarchically modulated by sensory information and previous learning. This organization of information processing in biological organisms has served as a major inspiration for artificial intelligence and has helped to create *in silico* systems capable of matching or even outperforming humans in several specific tasks, including visual recognition and strategy-based games. However, the development of human-like robots that are able to move as quickly as humans and respond flexibly in various situations remains a major challenge and indicates an area where further use of parallel and hierarchical architectures may hold promise. In this article we review several important neural and behavioral mechanisms organizing hierarchical and predictive processing for the acquisition and realization of flexible behavioral control. Then, inspired by the organizational features of brain circuits, we introduce a multi-timescale parallel and hierarchical learning framework for the realization of versatile and agile movement in humanoid robots.

1. Introduction

Many of our daily actions require the efficient and concurrent performance of multiple tasks. Imagine that you watch a movie while snacking on popcorn. Your brain receives and analyzes visual and auditory information, for example, to recognize who the actor is and what they are saying, while simultaneously generating the motor command to grab a piece of popcorn and take it to your mouth. The information processing underlying such tasks is thought to be facilitated by hierarchical mechanisms that occur within largely parallel neural circuits. Within the brain, 'hierarchical organization' is often used to describe sequential processing that occurs within a network of 'modules' (Meunier et al., 2010, 2009). These modules are often discrete brain regions or subregions that perform a specific sub-role within a larger functional circuit. For instance, separate layers of the visual cortex are known to hierarchically process the various perceptual features of a stimulus in a feedforward manner that integrates the receptive fields of the lower layers to eventually allow us to 'perceive' an image. Similar such hierarchical circuits are found throughout the brain and wider central nervous system, including in sensorimotor, associative, and limbic circuits, and when acting in parallel allow for complex and adaptive behavioral control.

The ubiquity of hierarchical and parallel organizational properties in biological organisms has led researchers to suggest that these features are likely evolutionarily conserved due to their efficiency for information processing (Alcocer-Cuarón et al., 2013; Mengistu et al., 2016). As such, similar such organizational properties have been advocated for artificial intelligence (AI) systems. Indeed, modern computers now commonly employ two or more central processing units (CPU) that simultaneously handle separate parts of a larger task (so-called parallel computing). Similarly, *in silico* artificial neural networks (ANN) incorporating hierarchical processing layers have been developed as an analogy of complex and flexible brain processing, in order to realize complex nonlinear mapping, classification, and optimization, by

* Correspondence to: Laboratory for Advanced Brain Functions, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan.
E-mail address: hikida@protein.osaka-u.ac.jp (T. Hikida).

[1] All authors contributed equally to this work.

https://doi.org/10.1016/j.neunet.2021.09.009

learning with massive parallel computation of simple calculation elements. For example, in the field of artificial visual intelligence, ANNs modeled on the hierarchical processing architecture of the brain's visual ventral stream are now capable of human-level object recognition (Nonaka et al., 2020; Yamins et al., 2014). Similar such integration of brain-like processing architectures may help to provide solutions in other areas of AI that have proved challenging, such as behavioral flexibility and smooth motor control in robots.

In this review article, we discuss two neural circuits that demonstrate the parallel and hierarchical features of the brain. The first is the cortico-basal ganglia-thalamo-cortical loop circuit in which frontal cortex projections to the basal ganglia are relayed back to the frontal cortex via the thalamus. More specifically, this circuit consists of several loops that mediate distinct functions, including sensorimotor, associative, and limbic processing. Here we summarize the anatomical structure of cortico-basal ganglia-thalamo-cortical loop circuits and discuss how the separation of functions to parallel loops facilitates economical behavioral control. Then we describe recent computational models of reinforcement learning incorporating parallel and hierarchical control of model-based and model-free systems for adaptive behavior in artificial systems.

The second type of circuit that we discuss are sensorimotor circuits that execute information processing at different levels of the central nervous system, including the spinal cord, midbrain, cerebellum, and cortex. As with cortico-basal ganglia-thalamo-cortical loop circuits, there are in fact several sensorimotor loop circuits that allow for various sensory inputs to be processed and translated to motor output. Here we describe two types of these circuits, somatosensory–motor and visuomotor circuits, and examine their hierarchical organization. In both circuits, low level loops are essential for stably achieving action targets in dynamic environments due to the slow speed of information processing in high level loops. We also introduce a new type of flexible regulation mechanism in low level somatosensory–motor circuits, as well as a distinctive visual processing mechanism specialized for low level visuomotor circuits, both of which are essential to realize human-like dynamic movements in robots.

Finally, we introduce model-free and model-based learning approaches utilizing hierarchical and parallel architectures, inspired by those found in brain circuits, to derive control strategies for multi-joint robotic systems. Specifically, we describe a multi-timescale learning framework that enables flexible and agile humanoid body movements by utilizing hierarchical processing layers similar to those found in the sensorimotor circuits and parallel information processing across these layers similar to that found in cortico-basal ganglia-thalamo-cortical loops. These findings demonstrate the utility of using biologically-inspired architectures for the design of artificial movement systems, and highlight the potential value of further use of hierarchical and parallel mechanisms for behavioral control in robots.

2. The cortico-basal ganglia-thalamo-cortical loop circuit

The ability to select, evaluate, and execute purposeful actions in response to external and internal cues is thought to be managed primarily by a series of hierarchical parallel processing loops known collectively as the cortico-basal ganglia-thalamo-cortical loop circuit (Balleine, 2019; Macpherson & Hikida, 2019; Peak et al., 2019). These loops are largely topographically organized and classically have been broadly categorized according to the proposed roles of the domains they incorporate; the sensorimotor, associative, and limbic loops (Fig. 1) (Alexander et al., 1986; Foster et al., 2020; Haber, 2003; Parent & Hazrati, 1995; Wichmann & DeLong, 2006).

At the highest level of the circuit, cortices associated with the sensorimotor (somatosensory and motor cortices), associative (lateral and ventromedial orbitofrontal, prelimbic, and infralimbic cortices), and limbic (agranular insular, ventromedial orbitofrontal, and prelimbic cortices) loops, project to striatal subregions approximately corresponding to the dorsolateral (DLS), dorsomedial (DMS), and ventral striatum, respectively (Hintiryan et al., 2016; Hooks et al., 2018; Hunnicutt et al., 2016; Li et al., 2018; Oh et al., 2014). In the striatum, corticostriatal inputs converge with intralaminar thalamostriatal inputs, predominantly from the centrolateral/centromedian nucleus and parafascicular nucleus in the dorsal striatum and the paraventricular thalamic nucleus in the ventral striatum (Hunnicutt et al., 2016; Li et al., 2018; Wall et al., 2013). From the dorsal striatum, dopamine D1 receptor-expressing striatal projection neurons (D1-SPNs) form a "direct" basal ganglia projection pathway to the substantia nigra pars reticulata (SNr) and globus pallidus internal segment (GPi), while dopamine D2 receptor-expressing striatal projection neurons (D2-SPNs) project to the SNr and GPi via an "indirect" pathway that includes the globus pallidus external segment (GPe) and subthalamic nucleus (STN) (Fig. 1) (Alexander & Crutcher, 1990; Haber, 2003). The influence of STN projections to nigral and pallidal regions is also regulated by a direct input from the cortex, primarily somatosensory and motor cortices, that forms a "hyperdirect" basal ganglia pathway (Aristieta et al., 2021; Nambu, 2004; Nambu et al., 2002). In contrast to the dorsal striatum, D1- and D2-SPNs of the nucleus accumbens (NAc) of the ventral striatum are not as clearly delineated to "direct" and "indirect" pathways, with D1-SPNs projecting to the SNr, ventral tegmental nucleus (VTA) and ventral pallidum (VP), and D2-SPNs projecting only to the VP (Kupchik et al., 2015). From the GPi/SNr, information is then sent to the motor (ventromedial, ventral anterior, and ventrolateral thalamic nuclei), associative (submedial, intralaminar, mediodorsal, laterodorsal, and lateral posterior thalamic nuclei), and limbic (anterior, mediodorsal, central medial, and midline thalamic nuclei) regions of the thalamus, which in turn project back to their respective functional areas in the striatum and cortex (Antal et al., 2014; Deniau et al., 2007; Foster et al., 2020; Haber & Calzavara, 2009; Mandelbaum et al., 2019; Wall et al., 2013).

Here we have focused on the broad categorization of cortico-basal ganglia-thalamo-cortical loops into sensorimotor, associative, and limbic loops; however, there is now a growing appreciation of the complex organizational properties of the striatum (Hintiryan et al., 2016; Hunnicutt et al., 2016; Stanley et al., 2020), as well as of the cortico-basal ganglia-thalamo-cortical loop circuit as a whole (Foster et al., 2020). In the striatum, as many as 29 discrete subregions (domains) have been identified, each contained within intermediate-scale networks (communities) that themselves combine to form large-scale networks (divisions) (Hintiryan et al., 2016). While intermediate-scale communities broadly corresponded to the sensorimotor, associative, and limbic loops described above, each striatal domain receives a unique pattern of cortical input, suggesting potentially heterogeneous functionality (Hintiryan et al., 2016). Similarly, a recent tracing study has suggested the existence of at least 6 parallel cortico-basal ganglia-thalamo-cortical loops, each with complex patterns of convergence and divergence in anatomically distinct domains of the various cortical, basal ganglia, and thalamic nuclei they incorporate (Foster et al., 2020).

2.1. The functional roles of cortico-basal ganglia-thalamo-cortical loops

Economical instrumental behavior requires the ability to encode associations between actions and their resulting outcomes,

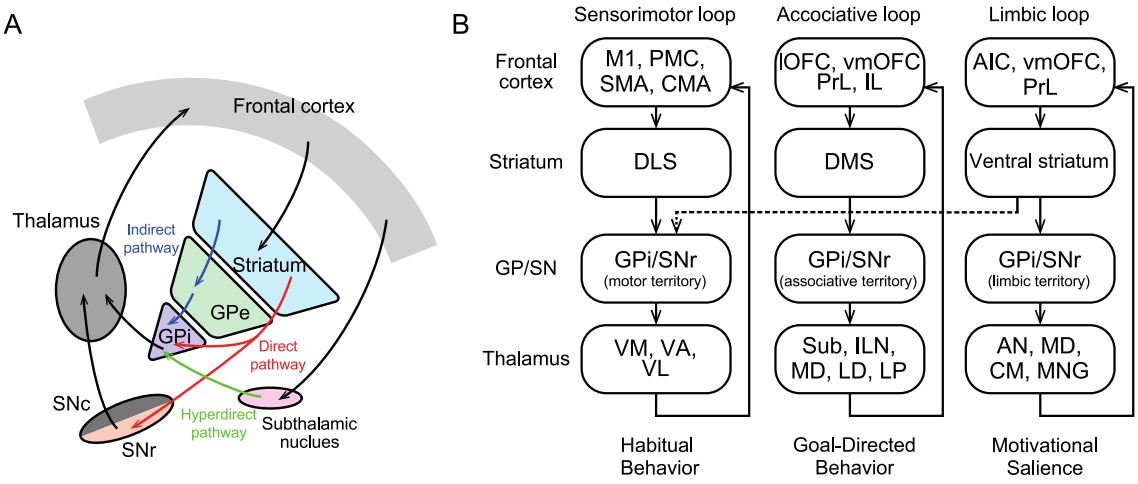

Fig. 1. Hierarchical and parallel processing in cortico-basal ganglia-thalamo-cortical loops. A. Connectivity in the cortico-basal ganglia-thalamo-cortical loop circuit. Projections from the frontal cortex are transmitted to the thalamus and eventually back up to the cortex largely via three routes through the basal ganglia; the direct, indirect, and hyperdirect pathways (indicated by red, blue, and green arrows, respectively). **B.** The cortico-basal ganglia-thalamo-cortical loop circuit can be functionally divided into separate sensorimotor, associative, and limbic loops, that are arranged hierarchically and in parallel to each other. AIC, agranular insular cortex; AN, anterior nucleus of thalamus; CM, central medial nucleus of thalamus; CMA, cingulate motor area; DMS, dorsomedial striatum; DLS, dorsolateral striatum; GPe, globus pallidus external segment; GPi, globus pallidus internal segment; IL, infralimbic cortex; ILN, intralaminar nucleus of thalamus; LD, lateral dorsal nucleus of thalamus; lOFC, lateral orbitofrontal cortex; LP, lateral posterior nucleus of thalamus; M1, primary motor cortex; MD, mediodorsal nucleus of thalamus; MNG, midline nuclear group; PMC, premotor cortex; PrL, prelimbic cortex; SMA, supplementary motor area; SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata; Sub, submedial nucleus of thalamus; VA, ventral anterior nucleus of thalamus; VL, ventrolateral nucleus of thalamus; VM, ventromedial nucleus of thalamus; vmOFC, ventromedial orbitofrontal cortex. The dotted arrow indicates a putative striatonigral connection between the limbic and sensorimotor loops.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

initiate and automate stable desired actions while suppressing unproductive actions, and dynamically adapt behaviors in response to changes in outcome values. The cortico-basal ganglia-thalamo-cortical loop circuit is thought to facilitate these operations and allow them to be performed concurrently by dividing the various processes and types of learning required across its constituent parallel loops. While the roles of specific cell types will not be discussed here, these processes are also enabled by separate or concurrent activation of striatal D1- and D2-SPNs (reviewed in Balleine, 2019; Macpherson et al., 2014; Peak et al., 2019).

2.1.1. Goal-directed learning and behavioral flexibility

In the early stages of instrumental learning, behavior is guided by learned associations between motor actions and their expected outcomes (action-outcome (A-O) associations). These goal-directed responses are flexible and highly sensitive to changes in the A-O contingency. In a series of lesion and chemical inactivation studies, various regions of the associative loop have been demonstrated to play important roles in the acquisition and performance of goal-directed responses. Indeed, while the prelimbic cortex (particularly the medial prefrontal cortex) (Corbit & Balleine, 2003; Coutureau et al., 2009; Hart, Bradfield, & Balleine, 2018; Hart, Bradfield, Fok et al., 2018; Tran-Tu-Yen et al., 2009) and mediodorsal thalamus (Corbit et al., 2003; Ostlund & Balleine, 2008) are reported to be necessary for the acquisition of A-O associations, the posterior DMS, which receives input from the prelimbic cortex, has been reported to be necessary for the performance of A-O-guided actions (Stalnaker et al., 2010; Yin, Knowlton et al., 2005; Yin, Ostlund et al., 2005).

Effective goal-directed behavior also requires actions to be dynamically altered or abandoned in response to changes in the value of the associated outcome, potentially requiring inhibition of an outdated response and acquisition of a novel response. Both the associative and limbic loops are suggested to play important roles supporting behavioral flexibility, and a variety of tasks have been devised to measure the ability to switch to a response requiring a novel strategy (strategy/set-shifting) or reverse a previously learnt strategy (reversal learning). Interestingly, these two abilities are dissociated within the cortex, with lesions and inactivation of the medial prefrontal cortex (mPFC) impairing the acquisition of a novel strategy in attentional set-shifting but not reversal learning (Birrell & Brown, 2000; Bissonette et al., 2008; Floresco et al., 2008), and disruption of the orbitofrontal cortex (OFC) resulting in the opposite phenotype (Bissonette et al., 2008; Bohn et al., 2003; Ghods-Sharifi et al., 2008; Graybeal et al., 2011). In the DMS, a target of both the mPFC and the OFC, lesions have been reported to impair both strategy-shifting and reversal learning (Braun & Hauber, 2011; Castañé et al., 2010; Ragozzino, 2007; Ragozzino et al., 2002). Similarly, in the NAc, which also receives inputs from both the mPFC and OFC, intra-NAc infusions of GABA or dopamine D2 receptor agonists, as well as genetic inactivation of neurotransmission from NAc D2-SPNs, have been reported to impair the ability for strategy switching requiring either set-shifting or reversal learning by increasing perseveration of previously correct strategies (Floresco et al., 2006; Haluk & Floresco, 2009; Macpherson et al., 2016; Yawata et al., 2012). Accordingly, intra-NAc D2 receptor antagonism, and optogenetic stimulation of prelimbic cortex inputs onto NAc D2-SPNs, have been demonstrated to facilitate set-shifting in a task-switching paradigm and reduce perseverative errors during early reversal learning in a visual discrimination task (Cui et al., 2018; Sala-Bayo et al., 2020).

While both associative and limbic loops appear to be necessary for behavioral flexibility, the precise functional roles that each loop plays are still unclear. NAc dysfunction is reported to increase perseverative errors, suggesting that it supports the inhibition of outdated responses (Cui et al., 2018; Macpherson et al., 2016); whereas, some (Ragozzino, 2007; Ragozzino et al., 2002), but not all (Castañé et al., 2010), studies of DMS inactivation have reported errors to be unrelated to perseveration, suggesting the associative loop to potentially play a greater role in the maintenance of novel strategies. Inconsistencies in the findings of studies investigating the role of the DMS in behavioral flexibility may be explained by differences in the sites of inactivation, with

strategy maintenance-related errors typically occurring following lesions of the anterior DMS, an area known to receive different cortical inputs to the posterior DMS (Hunnicutt et al., 2016; Ragozzino, 2007; Ragozzino et al., 2002).

2.1.2. Attribution of motivational salience

Goal-directed behavior is also known to be influenced by various motivational processes (Balleine & Killcross, 2006; Berridge, 2004; Dickinson & Balleine, 1994). Specifically, motivational salience attributed to the goal, or a goal-associated stimulus, is able to increase the efficiency and vigor of reward-seeking actions (Berridge, 2007; Berridge & Robinson, 1998; Ikemoto & Panksepp, 1999; Salamone et al., 2007). Within the limbic loop, the NAc is proposed to integrate inputs from cortical, midbrain dopaminergic, and limbic regions, including the hippocampus and amygdala, that supply information about the motivational salience of a goal or stimulus (Floresco, 2015; Mannella et al., 2013; Mogenson et al., 1980). This integrated information is then suggested to be conveyed back to the prefrontal cortex, via the limbic loop, where it can act to facilitate value-based action selection (Mannella et al., 2013; Nicola, 2007). Accordingly, NAc functioning appears to be particularly necessary when goal-directed actions require significant effort, as well as in ambiguous or complex situations, such as in the 5-choice serial-reaction time task, a visuospatial attention and impulsivity test requiring animals to accurately respond at one of five illuminated windows (Christakou et al., 2004; Pezze et al., 2007), and radial arm tasks where mice have to navigate through up to eight possible spatial locations to find a reward (Floresco et al., 1997; Gal et al., 1997). Whereas, tasks with low effort requirements or complexity, such as the acquisition of a task requiring animals to discriminate between two-to-four spatial locations (Castañé et al., 2010; Macpherson et al., 2016) or discrete cues (texture or smell) (Floresco et al., 2006), are reported to be insensitive to NAc inactivation.

Within the NAc, core and shell subregions are proposed to play different functional roles in reward-related behaviors, likely by action of their differing inputs and output projections (Kupchik et al., 2015; Li et al., 2018). The NAc core is innervated primarily by the prefrontal, prelimbic, and insular cortices, the basolateral amygdala, and the VTA, and is implicated in the ability of conditioned stimuli to instigate Pavlovian approach/avoidance behavior and to invigorate instrumental responding (Corbit & Balleine, 2011; Floresco, 2015; Parkinson et al., 2000; Saunders & Robinson, 2012). Whereas, the NAc shell is innervated primarily by the prefrontal cortex, the subiculum and CA1 regions of the hippocampus, the paraventricular thalamus, and the lateral hypothalamus, and is implicated in evaluating the hedonic value of natural (including food and novelty) and drug reinforcers (Bassareo et al., 2002; Bossert et al., 2007; Castro & Berridge, 2014; Fuchs et al., 2008; Hooks & Kalivas, 1995; Peciña & Berridge, 2005), as well as in drug reinstatement (Bossert et al., 2007; Fuchs et al., 2008) and outcome-specific invigoration of instrumental behaviors by conditioned stimuli (Corbit & Balleine, 2011). These studies have led to the suggestion that the NAc core is critical for the attribution of motivational salience to a goal or stimuli (based on both appetitive/aversive value and novelty value) (Ambroggi et al., 2011; Floresco, 2015; Mannella et al., 2013). Whereas, the NAc shell, is suggested to be critical for weighting the relative importance of goals and using this information to guide and modulate behavior, including the suppression of goal-irrelevant behaviors (Floresco, 2015; Mannella et al., 2013).

While the precise mechanisms by which activity in the limbic loop is able to modulate the performance of actions by the motor loop are still unclear, recent evidence has indicated that limbic information may be integrated with the motor loop at the level of the SNr (Foster et al., 2020). This is further supported by evidence that optogenetic stimulation of either NAc core or DLS D1-SPN axon terminals in the SNr resulted activation of the same areas in the motor cortex (M1) (Aoki et al., 2019).

2.1.3. Habitual behavior

Following repetition, behavior can become habitual and guided by learned associations between environmental stimuli and the motor responses they co-occur with (stimulus–response (S-R) associations). These stimulus-driven responses no longer rely on the representation of the expected outcome, and are thus relatively impervious to changes in the outcome value, such as outcome devaluation, where the value of a reward is reduced by allowing open access to the same reward (pre-feeding of a food reward) or pairing it with an aversive stimulus (taste aversion) (Dickinson, 1985; Packard & Knowlton, 2002; Smith & Graybiel, 2014). In contrast to goal-directed learning, the acquisition and performance of habitual behaviors triggered by S-R associations is thought to be mediated within the sensorimotor loop. Lesion, chemical inactivation, genetic deletion, and optogenetic inhibition studies have revealed that disruption of the infralimbic cortex (Barker et al., 2017; Coutureau & Killcross, 2003; Killcross & Coutureau, 2003; Smith et al., 2012) and DLS (Lingawi & Balleine, 2012; Tricomi et al., 2009; Yin et al., 2004, 2006; Yu et al., 2009) renders previously habitual responses sensitive to changes in the outcome value, likely by allowing the associative loop to regain response control.

Habitual motor behaviors are thought to be supported by the "chunking" of action sequences, whereby modular motor components (individual actions) are concatenated to form sequences of movements (Graybiel, 1998; Hikosaka et al., 1995; Sakai et al., 2003; Smith & Graybiel, 2014, 2016). These action sequences strengthen with repeated training, resulting in increased performance speed and accuracy, until sequences are automatically performed in their entirety upon being triggered, independent of the outcome of individual actions (Kubota et al., 2009; Smith & Graybiel, 2013; Thorn et al., 2010). Thus, through the development of chunked habitual responses, action sequences frequently resulting in rewards can not only be accelerated and refined, but also automated, conserving valuable cognitive resources, albeit at the cost of behavioral flexibility.

Interestingly, following repeated training, activity in DLS and infralimbic cortex neurons gradually becomes more prominent at the initiation and termination of action sequences, so-called "task-bracketing" activity (Jin & Costa, 2010; Jin et al., 2014; Jog et al., 1999; Smith & Graybiel, 2013). This task-bracketing activity has been shown to be directly correlated with the degree automation of action sequences in a trial-by-trial manner, with trials beginning and ending in strong DLS activation associated with less investigation of possible choices (less cognitive deliberation) during performance (Desrochers et al., 2015; Smith & Graybiel, 2013, 2016). While the precise function of these start and end activities is still unclear, it is suggested that they may facilitate chunking by signaling the boundaries of action sequences, allowing them to be represented as a single unit that can more readily be initiated by S-R associations (Graybiel & Grafton, 2015). Additionally, it is important to note that this task-bracketing activity is observed only in a subset of neurons, or not at all in some tasks (Sales-Carbonell et al., 2018), with the majority of DLS neurons reported to be engaged throughout the entirety of action sequences and responsible for the integration of contextual and kinematic information relevant to the task (Rueda-Orozco & Robbe, 2015; Vandaele et al., 2019).

2.2. Animal studies of hierarchical and parallel control of adaptive behavior in cortico-basal ganglia-thalamo-cortical loops

While goal-directed and habitual behaviors generally dominate the early and late stages of training, respectively, there is now considerable evidence that these two forms of behavioral

control develop in parallel (Bergstrom et al., 2018; Bradfield & Balleine, 2013; Smith & Graybiel, 2013, 2016; Thorn et al., 2010; Vandaele et al., 2019). Thus, it would appear that goal-directed and habitual behaviors either compete or cooperate to control adaptive behavior.

Support for competition between goal-directed and habitual control largely comes from studies demonstrating that attenuation of either type of behavior is able to bias control towards the other. While inactivation of the DMS renders behavior stimulus-bound and insensitive to outcome devaluation (Yin, Ostlund et al., 2005), inactivation of the DLS is reported to facilitate goal-directed behavior in early learning (Bergstrom et al., 2018; Bradfield & Balleine, 2013), and can block the transition to habitual behavior following extended training (Yin et al., 2004). The dominance of goal-directed and habitual strategies during early and late stages of training, respectively, may therefore represent hierarchical biasing of one of these two competing strategies during the appropriate time. Indeed, in mice it has been reported that DMS-projecting OFC neurons act to attenuate DMS activity following extended training, allowing the competing DLS-associated habitual strategy to gain control of behavior (Gremel et al., 2016; Gremel & Costa, 2013; Yin et al., 2009).

Interestingly, in contrast to studies demonstrating a decrease in DMS activity following the transition to habitual behavior (Gremel et al., 2016; Gremel & Costa, 2013; Yin et al., 2009), other studies have reported that both the DMS and DLS continue to be engaged following repeated training and both demonstrate task-bracketing activity (Stalnaker et al., 2010; Vandaele et al., 2019). Accordingly, pharmacological inactivation of either the DMS or the DLS is reported to impair sequence performance in overtrained mice (Vandaele et al., 2019). These findings indicate that the DMS may not be disengaged following the transition to habitual behavioral control, and may instead cooperate with the DLS to control the performance of habitual action sequences. Further support for collaboration between goal-directed and habitual strategies is provided by studies in mice demonstrating that the outcome of an instrumental action can itself act as a stimulus to guide the selection of subsequent actions (Balleine & Dezfouli, 2019; Balleine & Ostlund, 2007; Ostlund & Balleine, 2007). These findings suggest that S-R associations may proceed and act to trigger R-O associations resulting in the execution of instrumental responses, and have led to the development of a model proposing hierarchical integration of A-O and S-R associations, allowing collaboration between goal-directed and habitual strategies in order to achieve adaptive instrumental behavior (Balleine & Dezfouli, 2019).

Finally, while the importance of the DLS in habit formation has been established in tasks largely utilizing natural rewards, studies investigating the development of addiction to drugs of abuse, a phenomenon that has been likened to an extreme form of habitual behavior, have highlighted the importance of drug-induced synaptic plasticity in the NAc and VTA of the limbic loop (Everitt & Robbins, 2005; Francis et al., 2019; Lipton et al., 2019; Lüscher, 2013; Lüscher & Malenka, 2011; Scofield et al., 2016; Wolf, 2016). While the precise role of limbic loops in controlling habitual behaviors is still unclear, it has been reported that the ventral striatum also shows task-bracketing activity following the transition to habitual behavior, suggesting it potentially contributes to the performance of chunked action sequences (Atallah et al., 2014).

2.3. Computational modeling of hierarchical and parallel control of behavior

The processes governing action control in the brain have been computationally modeled through reinforcement learning (RL), a machine learning framework that describes how an optimal policy and corresponding value function are updated through experience (state–action-reward sequences). Specifically, it has been argued that goal-directed and habitual strategies can be algorithmically equated to model-based and model-free RL approaches, respectively (Daw et al., 2011, 2005; Dolan & Dayan, 2013; Doya et al., 2002; Huang et al., 2020). Here, model-based RL utilizes estimates of the reward function (an evaluation of how effective a learning agent is) and the state transition probability (an evaluation of how the environment changes according to the agent's action) to calculate the optimal policy for the given state. Whereas, model-free RL selects the optimal policy based on a value function that is directly learned from experiences obtained through interaction with the environment (i.e., actions that have been rewarded in the past). While both types of RL are driven by prediction errors, model-based RL acquires the state transition probability from a state prediction error (SPE), whereas model-free RL acquires the value function from a reward prediction error (RPE). These differences place the two systems at opposite ends of a trade-off, with the sophistication of the model-based system resulting in more flexible, but computationally demanding and slower, decision-making than the statistically efficient, but inflexible, model-free RL, which relies on "cached" action values (Daw et al., 2005; Wood & Rünger, 2015). While these RL learning approaches appear to account for many of the features of goal-directed and habitual learning in biological organisms, it should be noted that the extent to which model-based and model-free RL are able to fully capture the properties of goal-directed and habit learning has been questioned (Collins & Cockburn, 2020; Friedel et al., 2014; Gillan et al., 2015).

Among studies suggesting competition between parallelly processed action plans of model-based (goal-directed) and model-free (habitual) systems, several methods of arbitration have been hypothesized. An influential proposal is that arbitration may occur on the basis of the relative uncertainty of the estimates of each controller (Daw et al., 2005). Indeed, Lee et al. (2014) proposed that the reliability (a measure of certainty) of controllers could be calculated using their prediction errors, with the reliability of the model-based system (approximated from the aggregated SPE) needing to outweigh that of the model-free system (approximated from the aggregated RPE) in order to be selected. Others have similarly suggested model-free RL to be the default controller, but propose that model-based RL is selected when a cost-benefit-based arbitrator evaluates the benefits of using the model-based system (calculated based upon the difference in uncertainty between the two controllers) to outweigh the computational costs (cognitive effort and time) (Pezzulo et al., 2013). A major strength of these uncertainty-based arbitration theories is that they are able to convincingly explain why habitual behaviors tend to dominate following repeated training, as the uncertainty of model-free estimates would decrease following accumulation of knowledge about the environment (Dolan & Dayan, 2013). Interestingly, recent evidence has indicated that task complexity may interact with uncertainty in the arbitration process. Kim et al. (2019) revealed that human participants performing a two-stage Markov decision task, a paradigm designed to dissociate model-based and model-free choices, increasingly used model-based RL as the task complexity grew, but switched to model-free RL when both uncertainty and task complexity were high.

Beyond arbitration mechanisms, collaboration between model-based and model-free RL controllers has been suggested for control of action selection. Drummond and Niv (2020) recently described how model-based and model-free RL could collaborate to facilitate decision-making using the example of a chess game. They suggest that cached model-free values could be used by model-based planning to compute the action value of simulated

possible future moves, allowing the elimination of model-based simulated moves with low cached values. Oppositely, it has been suggested that the model-based system may act to train the model-free system by simulating or replaying experiences (Gershman et al., 2014; Sutton, 1991). This proposal is similar to the hierarchical RL model recently introduced by Balleine and Dezfouli (2019) (also described briefly in Section 2.2), in which learnt S-R associations (potentially represented by model-based RL) may proceed, and themselves act to trigger, R-O associations (potentially represented by model-free RL).

With regard to using model-based and model-free systems to aid robots in RL tasks, the performances of model-based or model-free RL alone, random or entropy-based (a measure similar to uncertainty evaluation) arbitration of the two controllers, or fusion of the two controllers using various ensemble RL methods (majority vote, rank vote, Boltzmann multiplication, Boltzmann addition), were recently compared in a conveyor belt block-pushing task using a robotic arm (Renaudo et al., 2015). Interestingly, it was revealed that under stable environmental conditions (constant speed of the conveyor belt), random arbitration of model-based and model-free action plans outperformed (lead to a greater cumulative reward) all other methods. Whereas, under unstable environmental conditions (changing speed of conveyor belt), fusion-based methods were found, in general, to result in greater or equivalent performance to random arbitration, and to outperform model-based or model-free propositions alone, as well as criterion-based arbitration.

2.4. Summary

To summarize, the organization of the cortico-basal ganglia-thalamo-cortical loop circuit into parallel loops allows for concurrent sensorimotor, associative, and limbic processing. These parallel processing loops enable the initiation and flexible alteration of goal-directed behaviors during early learning, and allow behavior to become automated when repeated under stable conditions, such as in overtraining. Biological and computational evidence suggest the ability of the cortico-basal ganglia-thalamo-cortical loop circuit to adaptively control behavior may be realized via hierarchical control of goal-directed and habitual behavioral strategies. In the next section, we will further describe how hierarchical organization of brain circuits facilitates behavioral control, specifically focusing on dynamic motor control via sensorimotor loop circuits.

3. Hierarchical sensorimotor loops

Parallel signaling mechanisms in different levels of the central nervous system (CNS), including the spinal cord, midbrain, cerebellum, and cortex, form an important hierarchy for sensorimotor control (Merel et al., 2019; Pearson & Gordon, 2000). For example, somatosensory signals detected at sensory receptors enter into the dorsal horn of the spinal cord where some of them monosynaptically and polysynaptically connect to motor neurons in the spinal cord, forming the lowest level sensorimotor loop, while others continue up to higher CNS structures and contribute to cortically descending motor commands. In the cortex, somatosensory signals are distributed to the secondary somatosensory cortex, parietal cortex, basal ganglia, and thalamus (Künzle, 1977; Lewis & Essen, 2000) via the primary sensory cortex, as well as being sent directly to the motor cortex via the ventrolateral nucleus of the thalamus (Iriki et al., 1991). On the other hand, visual motion signals detected by the retina, which are important for motor control, are sent to the extrastriate motion-sensitive area (MT, MST, V5) via cortical (Dubner & Zeki, 1971; Komatsu & Wurtz, 1988), subcortical (Berman & Wurtz, 2011), and direct (Schmid et al., 2010) projection pathways, from where they are then sent on towards various areas of the parietal cortex. With integrated multimodal information represented in the parietal cortex, action and motor planning emerges in the frontal, parietal, and basal ganglia networks (described in Section 2) so as to generate motor commands driving limb, eye, and body movements. In addition, the side-pathways of cortico-cerebellar and spinocerebellar networks play essential roles in automatic sensorimotor coordination (Ito, 2012).

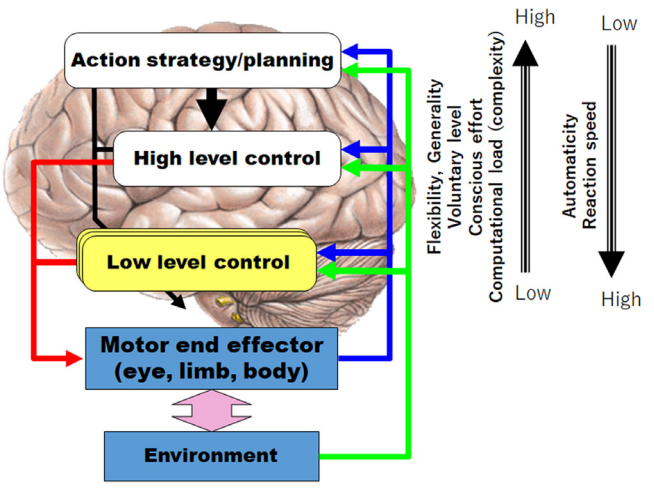

Fig. 2. Schematic diagram of the hierarchical structure of sensorimotor control. Low and high level controllers both receive proprioceptive (blue) and exteroceptive (green) sensory signals and generate motor commands. Low level controllers also receive modification signals from high level controllers to increase flexibility and maintain high reaction speeds.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Although sensorimotor networks, from sensory signals to motor commands, are complex, vary depending on each effector, and have not yet been fully elucidated, this section will focus on a simplified hierarchical structure of fundamental loops controlling implicit sensorimotor processing. The complex circuits mentioned above can be functionally categorized into low and high level controller and planning computation layers, as depicted in Fig. 2, each of which receives proprioceptive and exteroceptive (e.g., vision) sensory signals. High level computation can provide greater control flexibility for interacting within various environments (as in Fig. 2), but requires a greater amount of time for computation. In contrast, low level computation can realize faster reactions, but is not as flexible. To improve the flexibility of low level control, a smart regulation mechanism by the high level controller and planning computation layers is necessary (as illustrated by the arrow obliquely entering into the low level control box in Fig. 2). Here, we will discuss previous investigation of hierarchical mechanisms in the somatosensory–motor system and introduce novel frameworks of hierarchical interactions and parallel processing

3.1. Hierarchical mechanisms in the somatosensory–motor system

Somatosensory–motor loops have been extensively investigated for more than a century. The muscle sensory system (muscle spindle and Golgi tendon organ) conveys information about the muscle state (length, velocity, and force) through Ia, II, and Ib fibers, some of which are regulated actively by the fusimotor system at the sensory periphery using gamma motor signals. These sensory signals dominantly control the production of various movements (Gandevia & McCloskey, 1977), although cutaneous

and joint receptor information are also involved in motor control (Edin & Johansson, 1995). Sensory signals are sent to interneurons in the spinal cord and directly utilized for motor control, in addition to the midbrain and sensory areas of the cortex.

Since the muscles themselves and these sensorimotor loops combine to create spring-like properties in limbs, a servo control mechanism has been proposed as a computational model of brain motor control (Hammond, 1960; Marsden et al., 1976). Although this mechanism simplifies postural and movement control (Bizzi et al., 1984; Feldman, 1986), it cannot fully account for nonlinear motor dynamics during movements (Gomi & Kawato, 1996), resulting in the requirement for internal model control by the CNS (Kawato, 1999). Internal model control in a feedforward manner is, however, insufficient to deal with disturbances caused by neural noise (Harris & Wolpert, 1998) and various interactions with environments (Gomi & Kawato, 1993). Quick feedback control is, therefore, essential to reduce movement errors and stabilize posture.

To understand how feedback control in somatosensory–motor loops is regulated in various environments, the stretch reflex, which can be experimentally induced by mechanical perturbation, has been investigated (Prochazka et al., 2000). The stretch reflex response is driven by several sensorimotor loops from spinal to transcortical levels. An early study (Tanji & Evarts, 1976) showed modulation of the long-latency stretch reflex and accompanying activity changes of cortical neurons, suggesting task-dependent sensorimotor regulation in the cortex. In the last few decades, it has been shown that the long-latency stretch reflex is flexibly modulated by various contexts: task instruction (Hammond, 1956; Shemmell et al., 2009), upcoming force field (Kimura & Gomi, 2009; Kimura et al., 2006), stability change (Gomi et al., 2003; Shemmell et al., 2009), and the spatial properties of visual targets (Nashed et al., 2012; Pruszynski et al., 2008; Yang et al., 2011). These studies, alongside evidence that a shorter reflex response can also be modulated (Tanji & Evarts, 1976), indicate that different sensorimotor loops are regulated according to the task requirements.

In addition to these task/environmental contexts, recent studies (Ito & Gomi, 2020; Izawa & Shadmehr, 2008) have shown that reflexes are also modulated by state uncertainty. Fig. 3 shows a schematic diagram of the modulation of the long-latency stretch reflex by a directional mismatch (90 and 180 deg rotation) between the actual hand movement and visual hand cursor. A series of visual manipulation experiments suggest that the stretch reflex amplitude is regulated according to the uncertainty of body states estimated by a combination of visual and somatosensory information. In contrast to the uncertainty in estimated states, environmental uncertainty increases the reflex gain to reduce errors in goal-directed tasks (Franklin et al., 2012). These ideas would be, to some extent, in favor of the optimal feedback control framework (Scott, 2004; Todorov & Jordan, 2002), in which limb states are estimated by the integration of multimodal sensory inflow and internal prediction, considering temporal delay in each modality (Crevecoeur et al., 2016). As such, through interaction between high and low level controllers in a motor hierarchy, humans can realize dexterous tasks in various environments.

3.2. Parallel processing in the visuomotor system

In addition to somatosensory–motor loops, visuomotor loops are well studied, especially for smooth eye movement control. Moving targets can be smoothly and accurately tracked by the eyes (Krauzlis & Lisberger, 1994) and surrounding large-field visual motion induces a short-latency eye movement, known as the ocular following response (OFR) (Miles et al., 1986). To generate these eye movements, visual motion analysis is required

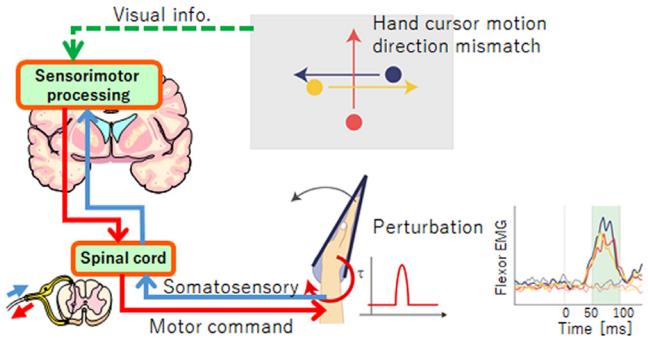

Fig. 3. Implicit somatosensory–motor coordination. A stretch reflex response observed in an electromyogram (EMG) following mechanical perturbation applied during a single joint reaching movement. The long latency component (50–100 ms) of the EMG is deemed to be generated by a cortical loop, and was reduced by the addition of a discrepancy of visual feedback motion and by eliminating the cursor feedback.

in these loops. Interestingly, the spatiotemporal frequency tuning of visual motion perception is similar to that of smooth pursuit (Matsumiya & Shioiri, 2015), but is greatly different from that of OFR (Burr & Ross, 1982; Gomi et al., 2006; Miles et al., 1986), suggesting a difference between visual motion analyses in perception and OFR generation processes (Boström & Warzecha, 2010; Glasser & Tadin, 2014).

Dissociation of visual analysis can be also found in perception and hand movements. Patients showing impaired card-slot orientation discrimination but the ability for card insertion to the slot suggest a dissociation between vision-for-action and vision-for-perception (Goodale & Milner, 1992). Further, visual motion analysis for hand control is different from that for perception (Gomi et al., 2006). A short-latency hand response induced by large-field visual motion, known as the manual following response (MFR), has a tuning peak at a low-spatial and high-temporal frequency (0.04 c/deg, 16.9 Hz), whereas motion perception sensitivity peaks at a higher-spatial and lower-temporal frequency (0.79 c/deg, 4.4 Hz) (Fig. 4).

One interesting question in information coding of a moving visual target is which of 'motion' or 'position' signals is employed in the brain during hand reaching movements towards the target object. In vision science, it is well known that the surrounding visual motion signal affects the perceived position (MIPS: Motion Induced Position Shift). Therefore, one theory (Whitney & Cavanagh, 2000; Whitney et al., 2003) has proposed that visually guided reaching is driven by a target position representation influenced by various motion signals (e.g., target texture and surroundings). In a recent study, Ueda et al. (2019) examined the temporal dynamics of visual motion effects on reaching adjustments and revealed that the onset of the indirect effect is significantly slower than the adjustment onset itself. This evidence indicates multi-stream processing in visuomotor control: a fast and direct contribution of visual motion for quick action initiation, and a relatively slow contribution of position representation updated by relevant motion signals for continuous action regulation. This distinctive visuomotor mechanism is crucial for successfully interacting with time-varying environments in the real world.

3.3. Summary

Decades of extensive physiological and psychophysical studies into sensorimotor processing have led to the development of several important computational standpoints: (1) fast reaction loops that functionally generate responses with short latencies

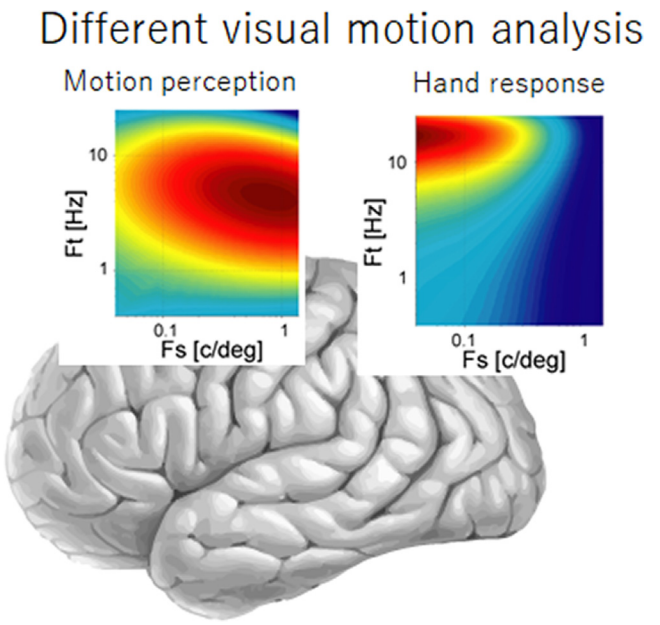

Fig. 4. Parallel visual processing for perception and hand control. The top two panels depict the spatiotemporal frequency tunings of perception sensitivity (left) and MFR (right). Most sensitive spatiotemporal frequencies greatly differ between perception and MFR.

are needed for dynamic interaction with the external world; (2) adaptability is needed even in fast loops in order to respond to various changes in the external world; and (3) adaptability is also needed in response to the uncertainty of one's own state representation. To achieve flexible and sophisticated interactions with the environment, complex and time-consuming computations are insufficient. Information processing mechanisms of the CNS indicate that in order to for robots to effectively interact with dynamic environments, like humans and animals they will likely require hierarchical and parallel loops for adaptive sensorimotor control.

4. A hierarchical learning approach for humanoid motor control

In the previous sections we described how hierarchical and parallel mechanisms within the cortico-basal ganglia-thalamo-cortical loop circuit and sensorimotor loop circuits allow adaptive behavioral control in biological organisms. Now, in this section, we discuss how AI and robotics communities can take inspiration from the architectural principles of such brain circuits to create artificial learning systems capable of flexible and agile body movements in humanoid robots.

Back in 2015, DARPA organized a robotics challenge consisting of eight tasks (such as stair climbing and valve opening) to evaluate how humanoid-type robots coped with a disaster situation in outdoor environments. Of the 23 teams that entered the challenge, only three were able to accomplish all eight tasks and required an average of around 50 min to complete them, compared to an average of five minutes for humans (Kajita et al., 2016). Thus, while the neuroscience and AI communities have made great progress in identifying and simulating the brain mechanisms responsible for other brain functions, including image and speech recognition, our understanding of motor control systems, and the motor control performance of humanoid robots, remains vastly inadequate. One of the biggest differences between human and humanoid motor control is that humans can generate a wide variety of dynamic movements to cope with various situations in real time. However, humanoid robot studies have typically focused on accomplishing a single task with careful, sometimes quasi-static, movements. The hierarchical and parallel architecture of the human brain may provide inspiration for the creation of robotic systems capable of human-level motor control to generate agile and versatile movements.

4.1. Hierarchical and parallel model-free learning approaches

In order for the acquisition of policies to generate a variety of movements, end-to-end deep learning methods have recently been applied to humanoid simulation models (Heess et al., 2017). While locomotion policies for complicated virtual environments can be acquired from reward functions as simple as moving forward, huge amounts of learning trials and carefully designed learning schedules are necessary. In contrast, the human brain does not seem to require such large-scale data or intricate scheduling to learn daily motions, but instead utilizes modularized and hierarchically organized networks for sample-efficient learning (Merel et al., 2019). Such use of multiple compact networks to represent behavioral modules, rather than using one big end-to-end network to learn a policy, not only requires fewer learning trials, but also allows efficient exploration with multiple resolution of control output to cope with novel situations (Sutton et al., 1999).

Based upon these principles, our group has proposed a hierarchical RL method in which upper and lower layers process information in parallel to output the desired postures and joint torque commands, respectively, to control a multi-link robot in a stand-up task from a prone state (Morimoto & Doya, 2001). Specifically, the upper-layer system learns to output a target posture as the next sub-goal for the lower-layer system in a Q-learning framework, whereas the lower-layer system tries to achieve the target posture provided by the upper-layer system in an actor–critic framework. Although the upper layer does not explicitly consider physical plausibility when the system outputs target postures as a sub-goal for the lower layer, only physically meaningful target postures are eventually selected since the sub-goals, which are unreachable by the lower layer, cannot have high action values. As a result, the multi-link robot was able to acquire a stand-up policy using our proposed hierarchical RL method in a simulated environment. We then applied the acquired policy to multi-link robot control in a real environment. Due to the modeling error between the simulated and real environments, the real robot failed to stand at the first trial; however, many fewer trials were required for the robot to adapt to the real environment and generate successful stand-up movements when compared to learning the stand-up policy from scratch. This is because the upper-layer policy, which outputs a more abstract control command (i.e., target posture) than the actual joint torque output of the lower-layer modules, could easily be generalized to the real environment. These findings demonstrate that hierarchical RL can be an efficient sim-to-real method.

Finally, more recently, deep RL methods that utilize hierarchical architectures have been proposed (Kulkarni et al., 2016). In one such study of hierarchical deep RL, it was shown that a simulated biped model could learn to effectively generate walking movements over uneven terrain, suggesting the utility of such methods for controlling the movement of humanoid robots (Peng et al., 2017).

4.2. Hierarchical and parallel model-based learning approaches

For high-dimensional systems such as humanoid robots a huge number of trials is required to acquire a global optimal policy that covers the entire state space thorough model-free policy learning methods. Therefore, for real-world applications, finding a local optimal trajectory-based policy based on model-based approaches is favorable. To find a locally optimal control sequence, differential dynamic programming (DDP) and its simplified version, iterative linear quadratic Gaussian (iLQG), have become standard methodologies. In DDP, the value function is propagated backward in time around the trajectory to derive a feedforward control sequence and a local linear feedback controller. To conduct this backward propagation, dynamic models of the robot and the surrounding environment are required. However, accurately identifying these models is not always easy. For such cases, policy-updating methods have recently attracted attention because they use multiple movement trajectories sampled forward in time rather than backward propagation of the value function (Theodorou et al., 2010).

Although trajectory-based policies are useful for high-dimensional systems, they struggle to deal with large external disturbances that take a robot away from its planned trajectory. To cope with such disturbances, model predictive control (MPC) is now widely used in the robotics community. One popular MPC approach in robotics studies is the use of DDP as a trajectory optimizer (Tassa et al., 2012). Using this implementation, an optimal control sequence is first derived by DDP and only the first control command is used, without another consecutive command sequence. The optimal control sequence is then calculated again based on the initial state at the next time step. Consequently, MPC enables the robot to cope with a sudden state change caused by an external disturbance. However, the application of MPC to real-time humanoid control for generating agile movements is cumbersome due to the heavy computational burden of performing the DDP calculations at each time step. On the other hand, the human brain can control its own body in real time, which is thought to be achieved by hierarchically connected control modules.

In order to address these issues, our group have proposed a computationally efficient hierarchical MPC for real-time humanoid control based on the idea of a singular perturbation method (Kokotovic et al., 1999), and inspired by the hierarchical control architecture used in sensorimotor circuits in the brain (described in Section 3.2.) (Fig. 5; Ishihara et al., 2019). Specifically, similar to sensorimotor circuits where environmental information is received into all hierarchical layers, here also information from the sensors is sent to all constituent layers where it can be processed in parallel. The top layer of the hierarchy uses a long-term horizon and a large time-step size to optimize entire body movements whereas the middle layer uses a short-time horizon and a small time-step size to optimize the motion of each limb using MPC calculations. Specifically, we extract fast dynamics from the humanoid robot system by introducing two different time scales. When compared with the system with smaller time scale, the larger-time-scale system can be considered to be a static environment. We then focus on optimizing the movements that belongs to the smaller-time-scale dynamics for the short-time horizon with the small time-step size. Therefore, the middle layer can quickly re-plan movement to cope with rapid changes in the environment. At the bottom layer, a reflex-based controller maintains the robot's posture with a very short control period. This controller is not model-based, but rather is inspired by and similar to the reflex-based controller found in biological systems (introduced in Section 3.1). We evaluated our framework in skating tasks with simulated and real lower-body humanoids that have rollers on their feet. Our simulated robot was able to generate agile motions in real time, including jumping over an obstacle and flipping down a cliff. In a real lower-body humanoid, our model was also able to successfully generate walking movement down a slope, indicating its effectiveness for controlling agile movement in humanoid robots.

5. Discussion

Adaptive and predictive behavioral control are facilitated by the organizational properties of information processing circuits in the brain. Here we described how parallel and hierarchical mechanisms in cortico-basal ganglia-thalamo-cortical and sensorimotor circuits enable flexible action selection and dynamic motor control. We then evidenced how hierarchical organization, as well as parallel processing within hierarchical layers, can be utilized in AI systems to facilitate motor control in humanoid robots.

Complex behaviors often require the concurrent performance of several tasks, such as processing visual stimuli and making motor responses when driving a car. The separation of these functions into largely anatomically independent circuits in the brain allows for them to be performed in parallel rather than creating a bottleneck by processing them serially. In this review we discussed how parallel brain circuits allow for simultaneous sensorimotor, associative, and limbic processing in cortico-basal ganglia-thalamo-cortical loop circuits, and concurrent somatosensory–motor and visuomotor processing in sensorimotor circuits. Similar such task parallelism is now ubiquitous in modern computers systems where multiple tasks can be split between several processors of a single machine (parallel computing) or be divided across multiple networked machines (distributed computing). This technological progress has been tightly coupled with advances in the capabilities of artificial intelligence and has dramatically reduced the time taken to train deep neural networks (Ben-Nun & Hoefler, 2019).

Parallelism in artificial systems can also be utilized through the use of different RL approaches that can be used separately or combined to facilitate adaptive behavior. In Section 2.3, we discussed how model-based and model-free RL approaches have been equated to goal-directed and habit learning in the brain, how arbitration and hierarchical control models have been created as explanations for how behavior is adaptively controlled in biological organisms, and how fusion of model-based and model-free systems has been demonstrated to improve RL in robotic systems under unstable environments. Interestingly, integration of model-based RL algorithms into a model-free RL system has also been reported to facilitate humanoid robot movement in complex trajectory-centric tasks by combining the sample efficiency of model-based systems with the generality of model-free systems, offering advantages to using each system separately (Chebotar et al., 2017). These findings indicate that arbitration or integration of model-based and model-free RL systems may provide a useful framework for behavioral control in robots. However, it is important to note that it has been questioned whether the use of simple RL dichotomies, such as model-based and model-free systems, is able to capture the rich and complex learning and decision-making processes of biological organisms (Collins & Cockburn, 2020). Thus, further elucidation of how parallel processing architectures are implemented in the brain may help to inspire the next generation of artificial behavioral control systems.

In addition to parallelism, we also discussed the hierarchical properties of the brain in this review. A major advantage of modular hierarchical organization is that it reduces the connection costs of the network, increasing energy efficiency (Laughlin & Sejnowski, 2003). By reducing the number and length of

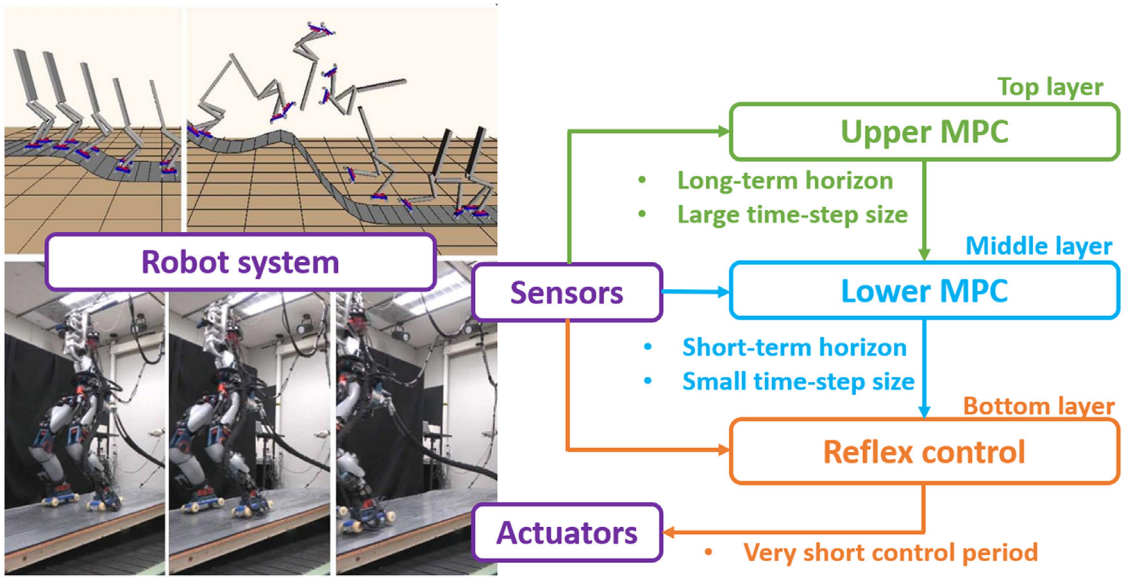

Fig. 5. Multi-timescale hierarchical and parallel control framework for generating versatile humanoid movements. Three hierarchical processing layers work in parallel to control humanoid robot movement. The top layer of the hierarchy uses a long-term horizon and a large time-step size to optimize entire body movements, while the middle layer uses a short-time horizon and a small time-step size to optimize each limb motion using MPC calculations. At the bottom layer, a reflex-based controller maintains the robot's posture with a very short control period. Details are explained in Ishihara et al. (2019).

connections in a network, energy expenditure for the creation, maintenance, and transmission of information across connections can be minimized. Indeed, computational evolution experiments designed to maximize network performance and minimize connection costs have shown that networks organized into modular hierarchies are evolutionarily favored (Clune et al., 2013; Mengistu et al., 2016). In addition to processing efficiency, there is strong evidence that hierarchical organization may also be advantageous for effectively adapting to changing environments (Kashtan & Alon, 2005; Sun & Deem, 2007). In hierarchical systems, modules can be trained to perform specific sub-tasks and can be added, altered, or replaced, when necessary, without the need for replacement of the entire system and the risk of loss of function in modules that are already performing effectively (Meunier et al., 2010). For example, in the visual system, modules specializing in processing of specific perceptual properties of visual stimuli, such as contrast and motion detection, could be improved over time without requiring large-scale changes in the entire visual system. Similarly, we have described in this review how the hierarchical organization of a modular motor system means that specific movements controlled by lower-level motor areas can be adaptively controlled by higher-level areas involved in motor sequence planning and posture. This hierarchical control allows quick adaptation to novel environments where motor sequences might have to be dynamically altered or newly learnt. Indeed, we have demonstrated how the incorporation of such hierarchical control into artificial motor systems can reduce the amount of time needed to train humanoid robots to perform specific agile movements in virtual and real environments.

Here we have mainly discussed hierarchical learning frameworks that contain different time-scale learning systems at each layer; however, hierarchical organization of learning parameters can also significantly improve policy performance. For example, hyperparameters such as learning rates or discount factors in RL algorithms need to be carefully selected for successful policy acquisition. In our brain, these hyperparameters may be tuned through evolutionary processes or learned through long-term learning trials (Doya, 2002). These kinds of meta-learning algorithms are recently gaining much attention and will likely play key roles in the development of a life-long learning system (Xu et al., 2020).

While hierarchical and parallel organization of control systems in the brain provide useful clues for the creation of artificial systems capable of adaptive and predictive behavioral control, we are still far from a sufficiently deep understanding of the required control mechanisms to develop an artificial intelligence system capable of human-like diverse and dynamic limbic, cognitive, and motor control. This is an area where significant progress is expected from collaboration among the neuroscience, AI, and robotics fields.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This is supported by MEXT, Japan Grant-in-Aid Scientific Research on Innovative Areas "Correspondence and Fusion of Artificial Intelligence and Brain Science" (JP16H06568 to TH and TM, JP16H06567 to MM, JP16H06566 to HG, JP19H05001 to EU, JP16H06565 to JM.)

References

Alcocer-Cuarón, C., Rivera, A. L., & Castaño, V. M. (2013). Hierarchical structure of biological systems. *Bioengineered*, *5*(2), 73–79. http://dx.doi.org/10.4161/bioe.26570.

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends in Neurosciences*, *13*(7), 266–271. http://dx.doi.org/10.1016/0166-2236(90)90107-l.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*, 357–381. http://dx.doi.org/10.1146/annurev.ne.09.030186.002041.

Ambroggi, F., Ghazizadeh, A., Nicola, S. M., & Fields, H. L. (2011). Roles of nucleus accumbens core and shell in incentive-cue responding and behavioral inhibition. *The Journal of Neuroscience*, *31*(18), 6820–6830. http://dx.doi.org/10.1523/jneurosci.6491-10.2011.

Antal, M., Beneduce, B. M., & Regehr, W. G. (2014). The substantia nigra conveys target-dependent excitatory and inhibitory outputs from the basal ganglia to the thalamus. *The Journal of Neuroscience*, *34*(23), 8032–8042. http://dx.doi.org/10.1523/jneurosci.0236-14.2014.

Aoki, S., Smith, J. B., Li, H., Yan, X., Igarashi, M., Coulon, P., Wickens, J. R., Ruigrok, T. J., & Jin, X. (2019). An open cortico-basal ganglia loop allows limbic control over motor output via the nigrothalamic pathway. *ELife*, *8*, Article e49995. http://dx.doi.org/10.7554/elife.49995.

Aristieta, A., Barresi, M., Lindi, S. A., Barrière, G., Courtand, G., Crompe, B. de la, Guilhemsang, L., Gauthier, S., Fioramonti, S., Baufreton, J., & Mallet, N. P. (2021). A disynaptic circuit in the globus pallidus controls locomotion inhibition. *Current Biology*, *31*(4), 707–721.e7. http://dx.doi.org/10.1016/j.cub.2020.11.019.

Atallah, H. E., McCool, A. D., Howe, M. W., & Graybiel, A. M. (2014). Neurons in the ventral striatum exhibit cell-type-specific representations of outcome during learning. *Neuron*, *82*(5), 1145–1156. http://dx.doi.org/10.1016/j.neuron.2014.04.021.

Balleine, B. W. (2019). The meaning of behavior: Discriminating reflex and volition in the brain. *Neuron*, *104*(1), 47–62. http://dx.doi.org/10.1016/j.neuron.2019.09.024.

Balleine, B. W., & Dezfouli, A. (2019). Hierarchical action control: Adaptive collaboration between actions and habits. *Frontiers in Psychology*, *10*(2735), http://dx.doi.org/10.3389/fpsyg.2019.02735.

Balleine, B. W., & Killcross, S. (2006). Parallel incentive processing: an integrated view of amygdala function. *Trends in Neurosciences*, *29*(5), 272–279. http://dx.doi.org/10.1016/j.tins.2006.03.002.

Balleine, B. W., & Ostlund, S. B. (2007). Still at the choice-point. *Annals of the New York Academy of Sciences*, *1104*(1), 147–171. http://dx.doi.org/10.1196/annals.1390.006.

Barker, J. M., Glen, W. B., Linsenbardt, D. N., Lapish, C. C., & Chandler, L. J. (2017). Habitual behavior is mediated by a shift in response-outcome encoding by infralimbic cortex. *ENeuro*, *4*(6), http://dx.doi.org/10.1523/eneuro.0337-17.2017, ENEURO.0337-17.2017.

Bassareo, V., Luca, M. A. D., & Chiara, G. D. (2002). Differential expression of motivational stimulus properties by dopamine in nucleus accumbens shell versus core and prefrontal cortex. *Journal of Neuroscience*, *22*(11), 4709–4719. http://dx.doi.org/10.1523/jneurosci.22-11-04709.2002.

Ben-Nun, T., & Hoefler, T. (2019). Demystifying parallel and distributed deep learning. *ACM Computing Surveys*, *52*(4), 1–43. http://dx.doi.org/10.1145/3320060.

Bergstrom, H. C., Lipkin, A. M., Lieberman, A. G., Pinard, C. R., Gunduz-Cinar, O., Brockway, E. T., Taylor, W. W., Nonaka, M., Bukalo, O., Wills, T. A., Rubio, F. J., Li, X., Pickens, C. L., Winder, D. G., & Holmes, A. (2018). Dorsolateral striatum engagement interferes with early discrimination learning. *Cell Reports*, *23*(8), 2264–2272. http://dx.doi.org/10.1016/j.celrep.2018.04.081.

Berman, R. A., & Wurtz, R. H. (2011). Signals conveyed in the pulvinar pathway from superior colliculus to cortical area MT. *The Journal of Neuroscience*, *31*(2), 373–384. http://dx.doi.org/10.1523/jneurosci.4738-10.2011.

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & Behavior*, *81*(2), 179–209. http://dx.doi.org/10.1016/j.physbeh.2004.02.004.

Berridge, K. C. (2007). The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology*, *191*(3), 391–431. http://dx.doi.org/10.1007/s00213-006-0578-x.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, *28*(3), 309–369. http://dx.doi.org/10.1016/s0165-0173(98)00019-8.

Birrell, J. M., & Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *Journal of Neuroscience*, *20*(11), 4320–4324. http://dx.doi.org/10.1523/jneurosci.20-11-04320.2000.

Bissonette, G. B., Martins, G. J., Franz, T. M., Harper, E. S., Schoenbaum, G., & Powell, E. M. (2008). Double dissociation of the effects of medial and orbital prefrontal cortical lesions on attentional and affective shifts in mice. *The Journal of Neuroscience*, *28*(44), 11124–11130. http://dx.doi.org/10.1523/jneurosci.2820-08.2008.

Bizzi, E., Accornero, N., Chapple, W., & Hogan, N. (1984). Posture control and trajectory formation during arm movement. *Journal of Neuroscience*, *4*(11), 2738–2744. http://dx.doi.org/10.1523/jneurosci.04-11-02738.1984.

Bohn, I., Giertler, C., & Hauber, W. (2003). Orbital prefrontal cortex and guidance of instrumental behaviour in rats under reversal conditions. *Behavioural Brain Research*, *143*(1), 49–56. http://dx.doi.org/10.1016/s0166-4328(03)00008-1.

Bossert, J. M., Poles, G. C., Wihbey, K. A., Koya, E., & Shaham, Y. (2007). Differential effects of blockade of dopamine D1-family receptors in nucleus accumbens core or shell on reinstatement of heroin seeking induced by contextual and discrete cues. *The Journal of Neuroscience*, *27*(46), 12655–12663. http://dx.doi.org/10.1523/jneurosci.3926-07.2007.

Boström, K. J., & Warzecha, A.-K. (2010). Open-loop speed discrimination performance of ocular following response and perception. *Vision Research*, *50*(9), 870–882. http://dx.doi.org/10.1016/j.visres.2010.02.010.

Bradfield, L. A., & Balleine, B. W. (2013). Hierarchical and binary associations compete for behavioral control during instrumental biconditional discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *39*(1), 2–13. http://dx.doi.org/10.1037/a0030941.

Braun, S., & Hauber, W. (2011). The dorsomedial striatum mediates flexible choice behavior in spatial tasks. *Behavioural Brain Research*, *220*(2), 288–293. http://dx.doi.org/10.1016/j.bbr.2011.02.008.

Burr, D. C., & Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Research*, *22*(4), 479–484. http://dx.doi.org/10.1016/0042-6989(82)90196-1.

Castañé, A., Theobald, D. E. H., & Robbins, T. W. (2010). Selective lesions of the dorsomedial striatum impair serial spatial reversal learning in rats. *Behavioural Brain Research*, *210*(1), 74–83. http://dx.doi.org/10.1016/j.bbr.2010.02.017.

Castro, D. C., & Berridge, K. C. (2014). Opioid hedonic hotspot in nucleus accumbens shell: Mu, delta, and kappa maps for enhancement of sweetness "Liking" and "Wanting". *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(12), 4239–4250. http://dx.doi.org/10.1523/jneurosci.4458-13.2014.

Chebotar, Y., Hausman, K., Zhang, M., Sukhatme, G., Schaal, S., & Levine, S. (2017). Combining model-based and model-free updates for trajectory-centric reinforcement learning. ArXiv.

Christakou, A., Robbins, T. W., & Everitt, B. J. (2004). Prefrontal cortical–ventral striatal interactions involved in affective modulation of attentional performance: Implications for corticostriatal circuit function. *The Journal of Neuroscience*, *24*(4), 773–780. http://dx.doi.org/10.1523/jneurosci.0949-03.2004.

Clune, J., Mouret, J.-B., & Lipson, H. (2013). The evolutionary origins of modularity. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1755), 20122863. http://dx.doi.org/10.1098/rspb.2012.2863.

Collins, A. G. E., & Cockburn, J. (2020). Beyond dichotomies in reinforcement learning. *Nature Reviews Neuroscience*, *21*(10), 576–586. http://dx.doi.org/10.1038/s41583-020-0355-6.

Corbit, L. H., & Balleine, B. W. (2003). The role of prelimbic cortex in instrumental conditioning. *Behavioural Brain Research*, *146*(1–2), 145–157. http://dx.doi.org/10.1016/j.bbr.2003.09.023.

Corbit, L. H., & Balleine, B. W. (2011). The general and outcome-specific forms of pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *The Journal of Neuroscience*, *31*(33), 11786–11794. http://dx.doi.org/10.1523/jneurosci.2711-11.2011.

Corbit, L. H., Muir, J. L., & Balleine, B. W. (2003). Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. *European Journal of Neuroscience*, *18*(5), 1286–1294. http://dx.doi.org/10.1046/j.1460-9568.2003.02833.x.

Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioural Brain Research*, *146*(1–2), 167–174. http://dx.doi.org/10.1016/j.bbr.2003.09.025.

Coutureau, E., Marchand, A. R., & Scala, G. D. (2009). Goal-directed responding is sensitive to lesions to the prelimbic cortex or basolateral nucleus of the amygdala but not to their disconnection. *Behavioral Neuroscience*, *123*(2), 443–448. http://dx.doi.org/10.1037/a0014818.

Crevecoeur, F., Munoz, D. P., & Scott, S. H. (2016). Dynamic multisensory integration: Somatosensory speed trumps visual accuracy during feedback control. *The Journal of Neuroscience*, *36*(33), 8598–8611. http://dx.doi.org/10.1523/jneurosci.0184-16.2016.

Cui, Q., Li, Q., Geng, H., Chen, L., Ip, N. Y., Ke, Y., & Yung, W.-H. (2018). Dopamine receptors mediate strategy abandoning via modulation of a specific prelimbic cortex-nucleus accumbens pathway in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(21), E4890–E4899. http://dx.doi.org/10.1073/pnas.1717106115.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. http://dx.doi.org/10.1016/j.neuron.2011.02.027.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711. http://dx.doi.org/10.1038/nn1560.

Deniau, J. M., Mailly, P., Maurice, N., & Charpier, S. (2007). The pars reticulata of the substantia nigra: a window to basal ganglia output. *Progress in Brain Research*, *160*, 151–172. http://dx.doi.org/10.1016/s0079-6123(06)60009-5.

Desrochers, T. M., Amemori, K., & Graybiel, A. M. (2015). Habit learning by naive macaques is marked by response sharpening of striatal neurons representing the cost and outcome of acquired action sequences. *Neuron*, *87*(4), 853–868. http://dx.doi.org/10.1016/j.neuron.2015.07.019.

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*(1135), 67–78. http://dx.doi.org/10.1098/rstb.1985.0010.

Dickinson, A., & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1–18. http://dx.doi.org/10.3758/bf03199951.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325. http://dx.doi.org/10.1016/j.neuron.2013.09.007.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*(4–6), 495–506. http://dx.doi.org/10.1016/s0893-6080(02)00044-8.

Doya, K., Samejima, K., Katagiri, K., & Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Computation*, *14*(6), 1347–1369. http://dx.doi.org/10.1162/089976602753712972.

Drummond, N., & Niv, Y. (2020). Model-based decision making and model-free learning. *Current Biology*, *30*(15), R860–R865. http://dx.doi.org/10.1016/j.cub.2020.06.051.

Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, *35*(2), 528–532. http://dx.doi.org/10.1016/0006-8993(71)90494-x.

Edin, B. B., & Johansson, N. (1995). Skin strain patterns provide kinaesthetic information to the human central nervous system. *The Journal of Physiology*, *487*(1), 243–251. http://dx.doi.org/10.1113/jphysiol.1995.sp020875.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, *8*(11), 1481–1489. http://dx.doi.org/10.1038/nn1579.

Feldman, A. G. (1986). Once more on the equilibrium-point hypothesis (λ model) for motor control. *Journal of Motor Behavior*, *18*(1), 17–54. http://dx.doi.org/10.1080/00222895.1986.10735369.

Floresco, S. B. (2015). The nucleus accumbens: an interface between cognition, emotion, and action. *Annual Review of Psychology*, *66*(1), 25–52. http://dx.doi.org/10.1146/annurev-psych-010213-115159.

Floresco, S. B., Block, A. E., & Tse, M. T. L. (2008). Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated procedure. *Behavioural Brain Research*, *190*(1), 85–96. http://dx.doi.org/10.1016/j.bbr.2008.02.008.

Floresco, S. B., Ghods-Sharifi, S., Vexelman, C., & Magyar, O. (2006). Dissociable roles for the nucleus accumbens core and shell in regulating set shifting. *The Journal of Neuroscience*, *26*(9), 2449–2457. http://dx.doi.org/10.1523/jneurosci.4431-05.2006.

Floresco, S. B., Seamans, J. K., & Phillips, A. G. (1997). Selective roles for hippocampal, prefrontal cortical, and ventral striatal circuits in radial-arm maze tasks with or without a delay. *Journal of Neuroscience*, *17*(5), 1880–1890. http://dx.doi.org/10.1523/jneurosci.17-05-01880.1997.

Foster, N. N., Korobkova, L., Garcia, L., Gao, L., Becerra, M., Sherafat, Y., Peng, B., Li, X., Choi, J.-H., Gou, L., Zingg, B., Azam, S., Lo, D., Khanjani, N., Zhang, B., Stanis, J., Bowman, I., Cotter, K., Cao, C., ... Dong, H. (2020). The mouse cortico-basal ganglia-thalamic network. http://dx.doi.org/10.1101/2020.10.06.326876, BioRxiv, 2020.10.06.326876.

Francis, T. C., Gantz, S. C., Moussawi, K., & Bonci, A. (2019). Synaptic and intrinsic plasticity in the ventral tegmental area after chronic cocaine. *Current Opinion in Neurobiology*, *54*, 66–72. http://dx.doi.org/10.1016/j.conb.2018.08.013.

Franklin, S., Wolpert, D. M., & Franklin, D. W. (2012). Visuomotor feedback gains upregulate during the learning of novel dynamics. *Journal of Neurophysiology*, *108*(2), 467–478. http://dx.doi.org/10.1152/jn.01123.2011.

Friedel, E., Koch, S. P., Wendt, J., Heinz, A., Deserno, L., & Schlagenhauf, F. (2014). Devaluation and sequential decisions: linking goal-directed and model-based behavior. *Frontiers in Human Neuroscience*, *8*(587), http://dx.doi.org/10.3389/fnhum.2014.00587.

Fuchs, R. A., Ramirez, D. R., & Bell, G. H. (2008). Nucleus accumbens shell and core involvement in drug context-induced reinstatement of cocaine seeking in rats. *Psychopharmacology*, *200*(4), 545–556. http://dx.doi.org/10.1007/s00213-008-1234-4.

Gal, G., Joel, D., Gusak, O., Feldon, J., & Weiner, I. (1997). The effects of electrolytic lesion to the shell subterritory of the nucleus accumbens on delayed non–matching-to-sample and four-arm baited eight-arm radial-maze tasks. *Behavioral Neuroscience*, *111*(1), 92–103. http://dx.doi.org/10.1037/0735-7044.111.1.92.

Gandevia, S. C., & McCloskey, D. I. (1977). Changes in motor commands, as shown by changes in perceived heaviness, during partial curarization and peripheral anaesthesia in man. *The Journal of Physiology*, *272*(3), 673–689. http://dx.doi.org/10.1113/jphysiol.1977.sp012066.

Gershman, S. J., Markman, A. B., & Otto, A. R. (2014). Retrospective revaluation in sequential decision making: A tale of two systems. *Journal of Experimental Psychology: General*, *143*(1), 182–194. http://dx.doi.org/10.1037/a0030844.

Ghods-Sharifi, S., Haluk, D. M., & Floresco, S. B. (2008). Differential effects of inactivation of the orbitofrontal cortex on strategy set-shifting and reversal learning. *Neurobiology of Learning and Memory*, *89*(4), 567–573. http://dx.doi.org/10.1016/j.nlm.2007.10.007.

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 523–536. http://dx.doi.org/10.3758/s13415-015-0347-6.

Glasser, D. M., & Tadin, D. (2014). Modularity in the motion system: Independent oculomotor and perceptual processing of brief moving stimuli. *Journal of Vision*, *14*(3), 28. http://dx.doi.org/10.1167/14.3.28.

Gomi, H., Abekawa, N., & Nishida, S. (2006). Spatiotemporal tuning of rapid interactions between visual-motion analysis and reaching movement. *The Journal of Neuroscience*, *26*(20), 5301–5308. http://dx.doi.org/10.1523/jneurosci.0340-06.2006.

Gomi, H., & Kawato, M. (1993). Recognition of manipulated objects by motor learning with modular architecture networks. *Neural Networks*, *6*(4), 485–497. http://dx.doi.org/10.1016/s0893-6080(05)80053-x.

Gomi, H., & Kawato, M. (1996). Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement. *Science*, *272*(5258), 117–120. http://dx.doi.org/10.1126/science.272.5258.117.

Gomi, H., Saijo, N., & Haggard, P. (2003). Flexible sensorimotor transformation during arm movements for interacting with environments. In *33rd annual meeting of society for neuroscience*, Program No. 492.11.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. http://dx.doi.org/10.1016/0166-2236(92)90344-8.

Graybeal, C., Feyder, M., Schulman, E., Saksida, L. M., Bussey, T. J., Brigman, J. L., & Holmes, A. (2011). Paradoxical reversal learning enhancement by stress or prefrontal cortical damage: rescue with BDNF. *Nature Neuroscience*, *14*(12), 1507–1509. http://dx.doi.org/10.1038/nn.2954.

Graybiel, A. M. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory*, *70*(1–2), 119–136. http://dx.doi.org/10.1006/nlme.1998.3843.

Graybiel, A. M., & Grafton, S. T. (2015). The striatum: Where skills and habits meet. *Cold Spring Harbor Perspectives in Biology*, *7*(8), Article a021691. http://dx.doi.org/10.1101/cshperspect.a021691.

Gremel, C. M., Chancey, J. H., Atwood, B. K., Luo, G., Neve, R., Ramakrishnan, C., Deisseroth, K., Lovinger, D. M., & Costa, R. M. (2016). Endocannabinoid modulation of orbitostriatal circuits gates habit formation. *Neuron*, *90*(6), 1312–1324. http://dx.doi.org/10.1016/j.neuron.2016.04.043.

Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, *4*(2264), http://dx.doi.org/10.1038/ncomms3264.

Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy*, *26*(4), 317–330. http://dx.doi.org/10.1016/j.jchemneu.2003.10.003.

Haber, S. N., & Calzavara, R. (2009). The cortico-basal ganglia integrative network: The role of the thalamus. *Brain Research Bulletin*, *78*(2–3), 69–74. http://dx.doi.org/10.1016/j.brainresbull.2008.09.013.

Haluk, D. M., & Floresco, S. B. (2009). Ventral striatal dopamine modulation of different forms of behavioral flexibility. *Neuropsychopharmacology*, *34*(8), 2041–2052. http://dx.doi.org/10.1038/npp.2009.21.

Hammond, P. H. (1956). The influence of prior instruction to the subject on an apparently involuntary neuro-muscular response. *The Journal of Physiology*, *132*(1), 17-8P.

Hammond, P. H. (1960). An experimental study of servo-action in the human muscular control. In *Prod. 3rd int. conf. med. electron.* (pp. 190–199).

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784. http://dx.doi.org/10.1038/29528.

Hart, G., Bradfield, L. A., & Balleine, B. W. (2018). Prefrontal corticostriatal disconnection blocks the acquisition of goal-directed action. *Journal of Neuroscience*, *38*(5), 1311–1322. http://dx.doi.org/10.1523/jneurosci.2850-17.2017.

Hart, G., Bradfield, L. A., Fok, S. Y., Chieng, B., & Balleine, B. W. (2018). The bilateral prefronto-striatal pathway is necessary for learning new goal-directed actions. *Current Biology*, *28*(14), 2218–2229. http://dx.doi.org/10.1016/j.cub.2018.05.028.e7.

Heess, N., TB, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., Eslami, S. M. A., Riedmiller, M., & Silver, D. (2017). Emergence of locomotion behaviours in rich environments. ArXiv.

Hikosaka, O., Rand, M. K., Miyachi, S., & Miyashita, K. (1995). Learning of sequential movements in the monkey: process of learning and retention of memory. *Journal of Neurophysiology*, *74*(4), 1652–1661. http://dx.doi.org/10.1152/jn.1995.74.4.1652.

Hintiryan, H., Foster, N. N., Bowman, I., Bay, M., Song, M. Y., Gou, L., Yamashita, S., Bienkowski, M. S., Zingg, B., Zhu, M., Yang, X. W., Shih, J. C., Toga, A. W., & Dong, H.-W. (2016). The mouse cortico-striatal projectome. *Nature Neuroscience*, *19*(8), 1100–1114. http://dx.doi.org/10.1038/nn.4332.

Hooks, M. S., & Kalivas, P. W. (1995). The role of mesoaccumbens-pallidal circuitry in novelty-induced behavioral activation. *Neuroscience*, *64*(3), 587–597. http://dx.doi.org/10.1016/0306-4522(94)00409-x.

Hooks, B. M., Papale, A. E., Paletzki, R. F., Feroze, M. W., Eastwood, B. S., Couey, J. J., Winnubst, J., Chandrashekar, J., & Gerfen, C. R. (2018). Topographic precision in sensory and motor corticostriatal projections varies across cell type and cortical area. *Nature Communications*, *9*(1), 3549. http://dx.doi.org/10.1038/s41467-018-05780-7.

Huang, Y., Yaple, Z. A., & Yu, R. (2020). Goal-oriented and habitual decisions: Neural signatures of model-based and model-free learning. *NeuroImage*, *215*, Article 116834. http://dx.doi.org/10.1016/j.neuroimage.2020.116834.

Hunnicutt, B. J., Jongbloets, B. C., Birdsong, W. T., Gertz, K. J., Zhong, H., & Mao, T. (2016). A comprehensive excitatory input map of the striatum reveals novel functional organization. *ELife*, *5*, Article e19103. http://dx.doi.org/10.7554/elife.19103.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*, *31*(1), 6–41. http://dx.doi.org/10.1016/s0165-0173(99)00023-5.

Iriki, A., Pavlides, C., Keller, A., & Asanuma, H. (1991). Long-term potentiation of thalamic input to the motor cortex induced by coactivation of thalamocortical and corticocortical afferents. *Journal of Neurophysiology*, *65*(6), 1435–1441. http://dx.doi.org/10.1152/jn.1991.65.6.1435.

Ishihara, K., Itoh, T. D., & Morimoto, J. (2019). Full-body optimal control toward versatile and agile behaviors in a humanoid robot. *IEEE Robotics and Automation Letters*, *5*(1), 119–126. http://dx.doi.org/10.1109/lra.2019.2947001.

Ito, M. (2012). *The cerebellum: Brain for an implicit self*. FT Press.

Ito, S., & Gomi, H. (2020). Visually-updated hand state estimates modulate the proprioceptive reflex independently of motor task requirements. *ELife*, *9*, Article e52380. http://dx.doi.org/10.7554/elife.52380.

Izawa, J., & Shadmehr, R. (2008). On-line processing of uncertain information in visuomotor control. *The Journal of Neuroscience*, *28*(44), 11360–11368. http://dx.doi.org/10.1523/jneurosci.3063-08.2008.

Jin, X., & Costa, R. M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature*, *466*(7305), 457–462. http://dx.doi.org/10.1038/nature09263.

Jin, X., Tecuapetla, F., & Costa, R. M. (2014). Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nature Neuroscience*, *17*(3), 423–430. http://dx.doi.org/10.1038/nn.3632.

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., & Graybiel, A. M. (1999). Building neural representations of habits. *Science*, *286*(5445), 1745–1749. http://dx.doi.org/10.1126/science.286.5445.1745.

Kajita, S., Morisawa, M., Nakaoka, S., Cisneros, R., Sakaguchi, T., Kaneko, K., & Kanehiro, F. (2016). Development and lessons learned in DARPA robotics challenge FinalsDevelopment and lessons learned in DARPA robotics challenge finals. *Journal of the Robotics Society of Japan*, *34*(6), 360–365. http://dx.doi.org/10.7210/jrsj.34.360.

Kashtan, N., & Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 13773–13778. http://dx.doi.org/10.1073/pnas.0503610102.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*(6), 718–727. http://dx.doi.org/10.1016/s0959-4388(99)00028-8.

Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex*, *13*(4), 400–408. http://dx.doi.org/10.1093/cercor/13.4.400.

Kim, D., Park, G. Y., O'Doherty, J. P., & Lee, S. W. (2019). Task complexity interacts with state-space uncertainty in the arbitration between model-based and model-free learning. *Nature Communications*, *10*(1), 5738. http://dx.doi.org/10.1038/s41467-019-13632-1.

Kimura, T., & Gomi, H. (2009). Temporal development of anticipatory reflex modulation to dynamical interactions during arm movement. *Journal of Neurophysiology*, *102*(4), 2220–2231. http://dx.doi.org/10.1152/jn.90907.2008.

Kimura, T., Haggard, P., & Gomi, H. (2006). Transcranial magnetic stimulation over sensorimotor cortex disrupts anticipatory reflex gain modulation for skilled action. *The Journal of Neuroscience*, *26*(36), 9272–9281. http://dx.doi.org/10.1523/jneurosci.3886-05.2006.

Kokotovic, P., Khalil, H. K., & O'Reilly, J. (1999). *Singular perturbation methods in control: Analysis and design*. Society for Industrial and Applied Mathematics.

Komatsu, H., & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology*, *60*(2), 580–603. http://dx.doi.org/10.1152/jn.1988.60.2.580.

Krauzlis, R. J., & Lisberger, S. G. (1994). Temporal properties of visual motion signals for the initiation of smooth pursuit eye movements in monkeys. *Journal of Neurophysiology*, *72*(1), 150–162. http://dx.doi.org/10.1152/jn.1994.72.1.150.

Kubota, Y., Liu, J., Hu, D., DeCoteau, W. E., Eden, U. T., Smith, A. C., & Graybiel, A. M. (2009). Stable encoding of task structure coexists with flexible coding of task events in sensorimotor striatum. *Journal of Neurophysiology*, *102*(4), 2142–2160. http://dx.doi.org/10.1152/jn.00522.2009.

Kulkarni, T. D., Narasimhan, K. R., Saeedi, A., & Tenenbaum, J. B. (2016). Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. ArXiv.

Künzle, H. (1977). Projections from the primary somatosensory cortex to basal ganglia and thalamus in the monkey. *Experimental Brain Research*, *30*(4), 481–492. http://dx.doi.org/10.1007/bf00237639.

Kupchik, Y. M., Brown, R. M., Heinsbroek, J. A., Lobo, M. K., Schwartz, D. J., & Kalivas, P. W. (2015). Coding the direct/indirect pathways by D1 and D2 receptors is not valid for accumbens projections. *Nature Neuroscience*, *18*(9), 1230–1232. http://dx.doi.org/10.1038/nn.4068.

Laughlin, S. B., & Sejnowski, T. J. (2003). Communication in neuronal networks. *Science*, *301*(5641), 1870–1874. http://dx.doi.org/10.1126/science.1089662.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, *81*(3), 687–699. http://dx.doi.org/10.1016/j.neuron.2013.11.028.

Lewis, J. W., & Essen, D. C. V. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *Journal of Comparative Neurology*, *428*(1), 112–137. http://dx.doi.org/10.1002/1096-9861(20001204)428:1<112::aid-cne8>3.0.co;2-9.

Li, Z., Chen, Z., Fan, G., Li, A., Yuan, J., & Xu, T. (2018). Cell-type-specific afferent innervation of the nucleus accumbens core and shell. *Frontiers in Neuroanatomy*, *12*(84), http://dx.doi.org/10.3389/fnana.2018.00084.

Lingawi, N. W., & Balleine, B. W. (2012). Amygdala central nucleus interacts with dorsolateral striatum to regulate the acquisition of habits. *The Journal of Neuroscience*, *32*(3), 1073–1081. http://dx.doi.org/10.1523/jneurosci.4806-11.2012.

Lipton, D. M., Gonzales, B. J., & Citri, A. (2019). Dorsal striatal circuits for habits, compulsions and addictions. *Frontiers in Systems Neuroscience*, *13*(28), http://dx.doi.org/10.3389/fnsys.2019.00028.

Lüscher, C. (2013). Drug-evoked synaptic plasticity causing addictive behavior. *The Journal of Neuroscience*, *33*(45), 17641–17646. http://dx.doi.org/10.1523/jneurosci.3406-13.2013.

Lüscher, C., & Malenka, R. C. (2011). Drug-evoked synaptic plasticity in addiction: From molecular changes to circuit remodeling. *Neuron*, *69*(4), 650–663. http://dx.doi.org/10.1016/j.neuron.2011.01.017.

Macpherson, T., & Hikida, T. (2019). Role of basal ganglia neurocircuitry in the pathology of psychiatric disorders. *Psychiatry and Clinical Neurosciences*, *13*(266), http://dx.doi.org/10.1111/pcn.12830.

Macpherson, T., Morita, M., & Hikida, T. (2014). Striatal direct and indirect pathways control decision-making behavior. *Frontiers in Psychology*, *5*(1301), http://dx.doi.org/10.3389/fpsyg.2014.01301.

Macpherson, T., Morita, M., Wang, Y., Sasaoka, T., Sawa, A., & Hikida, T. (2016). Nucleus accumbens dopamine D2-receptor expressing neurons control behavioral flexibility in a place discrimination task in the IntelliCage. *Learning & Memory (Cold Spring Harbor, NY)*, *23*(7), 359–364. http://dx.doi.org/10.1101/lm.042507.116.

Mandelbaum, G., Taranda, J., Haynes, T. M., Hochbaum, D. R., Huang, K. W., Hyun, M., Venkataraju, K. U., Straub, C., Wang, W., Robertson, K., Osten, P., & Sabatini, B. L. (2019). Distinct cortical-thalamic-striatal circuits through the parafascicular nucleus. *Neuron*, *102*(3), 636–652. http://dx.doi.org/10.1016/j.neuron.2019.02.035.e7.

Mannella, F., Gurney, K., & Baldassarre, G. (2013). The nucleus accumbens as a nexus between values and goals in goal-directed behavior: a review and a new hypothesis. *Frontiers in Behavioral Neuroscience*, *7*(135), http://dx.doi.org/10.3389/fnbeh.2013.00135.

Marsden, C. D., Merton, P. A., & Morton, H. B. (1976). Servo action in the human thumb. *The Journal of Physiology*, *257*(1), 1–44. http://dx.doi.org/10.1113/jphysiol.1976.sp011354.

Matsumiya, K., & Shioiri, S. (2015). Smooth pursuit eye movements and motion perception share motion signals in slow and fast motion mechanisms. *Journal of Vision*, *15*(11), 12. http://dx.doi.org/10.1167/15.11.12.

Mengistu, H., Huizinga, J., Mouret, J.-B., & Clune, J. (2016). The evolutionary origins of hierarchy. *PLoS Computational Biology*, *12*(6), Article e1004829. http://dx.doi.org/10.1371/journal.pcbi.1004829.

Merel, J., Botvinick, M., & Wayne, G. (2019). Hierarchical motor control in mammals and machines. *Nature Communications*, *10*(1), 5489. http://dx.doi.org/10.1038/s41467-019-13239-6.

Meunier, D., Lambiotte, R., & Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, *4*(200), http://dx.doi.org/10.3389/fnins.2010.00200.

Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D., & Bullmore, E. T. (2009). Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, *3*(37), http://dx.doi.org/10.3389/neuro.11.037.2009.

Miles, F. A., Kawano, K., & Optican, L. M. (1986). Short-latency ocular following responses of monkey. I. Dependence on temporospatial properties of visual input. *Journal of Neurophysiology*, *56*(5), 1321–1354. http://dx.doi.org/10.1152/jn.1986.56.5.1321.

Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. *Progress in Neurobiology*, *14*(2–3), 69–97.

Morimoto, J., & Doya, K. (2001). Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning. *Robotics and Autonomous Systems*, *36*(1), 37–51. http://dx.doi.org/10.1016/s0921-8890(01)00113-0.

Nambu, A. (2004). A new dynamic model of the cortico-basal ganglia loop. *Progress in Brain Research*, *143*, 461–466. http://dx.doi.org/10.1016/s0079-6123(03)43043-4.

Nambu, A., Tokuno, H., & Takada, M. (2002). Functional significance of the cortico–subthalamo–pallidal 'hyperdirect' pathway. *Neuroscience Research*, *43*(2), 111–117. http://dx.doi.org/10.1016/s0168-0102(02)00027-5.

Nashed, J. Y., Crevecoeur, F., & Scott, S. H. (2012). Influence of the behavioral goal and environmental obstacles on rapid feedback responses. *Journal of Neurophysiology*, *108*(4), 999–1009. http://dx.doi.org/10.1152/jn.01089.2011.

Nicola, S. M. (2007). The nucleus accumbens as part of a basal ganglia action selection circuit. *Psychopharmacology*, *191*(3), 521–550. http://dx.doi.org/10.1007/s00213-006-0510-4.

Nonaka, S., Majima, K., Aoki, S. C., & Kamitani, Y. (2020). Brain hierarchy score: Which deep neural networks are hierarchically brain-like? http://dx.doi.org/10.1101/2020.07.22.216713, BioRxiv, 2020.07.22.216713.

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A. M., Mortrud, M. T., Ouellette, B., Nguyen, T. N., Sorensen, S. A., Slaughterbeck, C. R., Wakeman, W., Li, Y., Feng, D., Ho, A., ... Zeng, H. (2014). A mesoscale connectome of the mouse brain. *Nature*, *508*(7495), 207–214. http://dx.doi.org/10.1038/nature13186.

Ostlund, S. B., & Balleine, B. W. (2007). Selective reinstatement of instrumental performance depends on the discriminative stimulus properties of the mediating outcome. *Animal Learning & Behavior*, *35*(1), 43–52. http://dx.doi.org/10.3758/bf03196073.

Ostlund, S. B., & Balleine, B. W. (2008). Differential involvement of the basolateral amygdala and mediodorsal thalamus in instrumental action selection. *The Journal of Neuroscience*, *28*(17), 4398–4405. http://dx.doi.org/10.1523/jneurosci.5472-07.2008.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*(1), 563–593. http://dx.doi.org/10.1146/annurev.neuro.25.112701.142937.

Parent, A., & Hazrati, L.-N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews*, *20*(1), 91–127. http://dx.doi.org/10.1016/0165-0173(94)00007-c.

Parkinson, J. A., Willoughby, P. J., Robbins, T. W., & Everitt, B. J. (2000). Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs pavlovian approach behavior: Further evidence for limbic cortical-ventral striatopallidal systems. *Behavioral Neuroscience*, *114*(1), 42–63. http://dx.doi.org/10.1037//0735-7044.114.1.42.

Peak, J., Hart, G., & Balleine, B. W. (2019). From learning to action: the integration of dorsal striatal input and output pathways in instrumental conditioning. *European Journal of Neuroscience*, *49*(5), 658–671. http://dx.doi.org/10.1111/ejn.13964.

Pearson, K., & Gordon, J. (2000). Spinal reflexes. In E. Kandel, T. Schwatz, & T. Jessell (Eds.), *Principles of neural science* (4th edition). McGraw-Hill.

Peciña, S., & Berridge, K. C. (2005). Hedonic hot spot in nucleus accumbens shell: where do mu-opioids cause increased hedonic impact of sweetness? *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *25*(50), 11777–11786. http://dx.doi.org/10.1523/jneurosci.2329-05.2005.

Peng, X. B., Berseth, G., Yin, K., & Panne, M. V. D. (2017). Deeploco. *ACM Transactions on Graphics*, *36*(4), 1–13. http://dx.doi.org/10.1145/3072959.3073602.

Pezze, M.-A., Dalley, J. W., & Robbins, T. W. (2007). Differential roles of dopamine D1 and D2 receptors in the nucleus accumbens in attentional performance on the five-choice serial reaction time task. *Neuropsychopharmacology*, *32*(2), 273–283. http://dx.doi.org/10.1038/sj.npp.1301073.

Pezzulo, G., Rigoli, F., & Chersi, F. (2013). The mixed instrumental controller: Using value of information to combine habitual choice and mental simulation. *Frontiers in Psychology*, *4*(92), http://dx.doi.org/10.3389/fpsyg.2013.00092.

Prochazka, A., Clarac, F., Loeb, G. E., Rothwell, J. C., & Wolpaw, J. R. (2000). What do reflex and voluntary mean? Modern views on an ancient debate. *Experimental Brain Research*, *130*(4), 417–432. http://dx.doi.org/10.1007/s002219900250.

Pruszynski, J. A., Kurtzer, I., & Scott, S. H. (2008). Rapid motor responses are appropriately tuned to the metrics of a visuospatial task. *Journal of Neurophysiology*, *100*(1), 224–238. http://dx.doi.org/10.1152/jn.90262.2008.

Ragozzino, M. E. (2007). The contribution of the medial prefrontal cortex, orbitofrontal cortex, and dorsomedial striatum to behavioral flexibility. *Annals of the New York Academy of Sciences*, *1121*(1), 355–375. http://dx.doi.org/10.1196/annals.1401.013.

Ragozzino, M. E., Ragozzino, K. E., Mizumori, S. J. Y., & Kesner, R. P. (2002). Role of the dorsomedial striatum in behavioral flexibility for response and visual cue discrimination learning. *Behavioral Neuroscience*, *116*(1), 105–115. http://dx.doi.org/10.1037/0735-7044.116.1.105.

Renaudo, E., Girard, B., Chatila, R., & Khamassi, M. (2015). Which criteria for autonomously shifting between goal-directed and habitual behaviors in robots? *2015 joint IEEE international conference on development and learning and epigenetic robotics*, *25*, 4–260. http://dx.doi.org/10.1109/devlrn.2015.7346152.

Rueda-Orozco, P. E., & Robbe, D. (2015). The striatum multiplexes contextual and kinematic information to constrain motor habits execution. *Nature Neuroscience*, *18*(3), 453–460. http://dx.doi.org/10.1038/nn.3924.

Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during human visuomotor sequence learning. *Experimental Brain Research*, *152*(2), 229–242. http://dx.doi.org/10.1007/s00221-003-1548-8.

Sala-Bayo, J., Fiddian, L., Nilsson, S. R. O., Hervig, M. E., McKenzie, C., Mareschi, A., Boulos, M., Zhukovsky, P., Nicholson, J., Dalley, J. W., Alsiö, J., & Robbins, T. W. (2020). Dorsal and ventral striatal dopamine D1 and D2 receptors differentially modulate distinct phases of serial visual reversal learning. *Neuropsychopharmacology*, *45*(5), 736–744. http://dx.doi.org/10.1038/s41386-020-0612-4.

Salamone, J. D., Correa, M., Farrar, A., & Mingote, S. M. (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology*, *191*(3), 461–482. http://dx.doi.org/10.1007/s00213-006-0668-9.

Sales-Carbonell, C., Taouali, W., Khalki, L., Pasquet, M. O., Petit, L. F., Moreau, T., Rueda-Orozco, P. E., & Robbe, D. (2018). No discrete start/stop signals in the dorsal striatum of mice performing a learned action. *Current Biology*, *28*(19), 3044–3055. http://dx.doi.org/10.1016/j.cub.2018.07.038.e5.

Saunders, B. T., & Robinson, T. E. (2012). The role of dopamine in the accumbens core in the expression of Pavlovian-conditioned responses. *The European Journal of Neuroscience*, *36*(4), 2521–2532. http://dx.doi.org/10.1111/j.1460-9568.2012.08217.x.

Schmid, M. C., Mrowka, S. W., Turchi, J., Saunders, R. C., Wilke, M., Peters, A. J., Ye, F. Q., & Leopold, D. A. (2010). Blindsight depends on the lateral geniculate nucleus. *Nature*, *466*(7304), 373–377. http://dx.doi.org/10.1038/nature09179.

Scofield, M. D., Heinsbroek, J. A., Gipson, C. D., Kupchik, Y. M., Spencer, S., Smith, A. C. W., Roberts-Wolfe, D., & Kalivas, P. W. (2016). The nucleus accumbens: Mechanisms of addiction across drug classes reflect the importance of glutamate homeostasis. *Pharmacological Reviews*, *68*(3), 816–871. http://dx.doi.org/10.1124/pr.116.012484.

Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, *5*(7), 532–545. http://dx.doi.org/10.1038/nrn1427.

Shemmell, J., An, J. H., & Perreault, E. J. (2009). The differential role of motor cortex in stretch reflex modulation induced by changes in environmental mechanics and verbal instruction. *The Journal of Neuroscience*, *29*(42), 13255–13263. http://dx.doi.org/10.1523/jneurosci.0892-09.2009.

Smith, K. S., & Graybiel, A. M. (2013). A dual operator view of habitual behavior reflecting cortical and striatal dynamics. *Neuron*, *79*(2), 361–374. http://dx.doi.org/10.1016/j.neuron.2013.05.038.

Smith, K. S., & Graybiel, A. M. (2014). Investigating habits: strategies, technologies and models. *Frontiers in Behavioral Neuroscience*, *8*(39), http://dx.doi.org/10.3389/fnbeh.2014.00039.

Smith, K. S., & Graybiel, A. M. (2016). Habit formation. *Dialogues in Clinical Neuroscience*, *18*(1), 33–43. http://dx.doi.org/10.31887/dcns.2016.18.1/ksmith.

Smith, K. S., Virkud, A., Deisseroth, K., & Graybiel, A. M. (2012). Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *109*(46), 18932–18937. http://dx.doi.org/10.1073/pnas.1216264109.

Stalnaker, T. A., Calhoon, G. G., Ogawa, M., Roesch, M. R., & Schoenbaum, G. (2010). Neural correlates of stimulus–response and response–outcome associations in dorsolateral versus dorsomedial striatum. *Frontiers in Integrative Neuroscience*, *4*(12), http://dx.doi.org/10.3389/fnint.2010.00012.

Stanley, G., Gokce, O., Malenka, R. C., Südhof, T. C., & Quake, S. R. (2020). Continuous and discrete neuron types of the adult murine striatum. *Neuron*, *105*(4), 688–699. http://dx.doi.org/10.1016/j.neuron.2019.11.004.e8.

Sun, J., & Deem, M. W. (2007). Spontaneous emergence of modularity in a model of evolving individuals. *Physical Review Letters*, *99*(22), Article 228107. http://dx.doi.org/10.1103/physrevlett.99.228107.

Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM SIGART Bulletin*, *2*(4), 160–163. http://dx.doi.org/10.1145/122344.122377.

Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*(1–2), 181–211. http://dx.doi.org/10.1016/s0004-3702(99)00052-1.

Tanji, J., & Evarts, E. V. (1976). Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. *Journal of Neurophysiology*, *39*(5), 1062–1068. http://dx.doi.org/10.1152/jn.1976.39.5.1062.

Tassa, Y., Erez, T., & Todorov, E. (2012). Synthesis and stabilization of complex behaviors through online trajectory optimization. In *2012 IEEE/RSJ international conference on intelligent robots and systems* (pp. 4906–4913). http://dx.doi.org/10.1109/iros.2012.6386025.

Theodorou, E. A., Buchli, J., & Schaal, S. (2010). A generalized path integral control approach to reinforcement learning. *Journal of Machine Learning Research*, *11*, 3137–3181.

Thorn, C. A., Atallah, H., Howe, M., & Graybiel, A. M. (2010). Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron*, *66*(5), 781–795. http://dx.doi.org/10.1016/j.neuron.2010.04.036.

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, *5*(11), 1226–1235. http://dx.doi.org/10.1038/nn963.

Tran-Tu-Yen, D. A. S., Marchand, A. R., Pape, J., Scala, G. D., & Coutureau, E. (2009). Transient role of the rat prelimbic cortex in goal-directed behaviour. *European Journal of Neuroscience*, *30*(3), 464–471. http://dx.doi.org/10.1111/j.1460-9568.2009.06834.x.

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, *29*(11), 2225–2232. http://dx.doi.org/10.1111/j.1460-9568.2009.06796.x.

Ueda, H., Abekawa, N., Ito, S., & Gomi, H. (2019). Distinct temporal developments of visual motion and position representations for multi-stream visuomotor coordination. *Scientific Reports*, *9*(1), 12104. http://dx.doi.org/10.1038/s41598-019-48535-0.

Vandaele, Y., Mahajan, N. R., Ottenheimer, D. J., Richard, J. M., Mysore, S. P., & Janak, P. H. (2019). Distinct recruitment of dorsomedial and dorsolateral striatum erodes with extended training. *ELife*, *8*, Article e49536. http://dx.doi.org/10.7554/elife.49536.

Wall, N. R., De La Parra, M., Callaway, E. M., & Kreitzer, A. C. (2013). Differential innervation of direct- and indirect-pathway striatal projection neurons. *Neuron*, *79*(2), 347–360. http://dx.doi.org/10.1016/j.neuron.2013.05.014.

Whitney, D., & Cavanagh, P. (2000). Motion distorts visual space: shifting the perceived position of remote stationary objects. *Nature Neuroscience*, *3*(9), 954–959. http://dx.doi.org/10.1038/78878.

Whitney, D., Westwood, D. A., & Goodale, M. A. (2003). The influence of visual motion on fast reaching movements to a stationary object. *Nature*, *423*(6942), 869–873. http://dx.doi.org/10.1038/nature01693.

Wichmann, T., & DeLong, M. R. (2006). Deep brain stimulation for neurologic and neuropsychiatric disorders. *Neuron*, *52*(1), 197–204. http://dx.doi.org/10.1016/j.neuron.2006.09.022.

Wolf, M. E. (2016). Synaptic mechanisms underlying persistent cocaine craving. *Nature Reviews Neuroscience*, *17*(6), 351–365. http://dx.doi.org/10.1038/nrn.2016.39.

Wood, W., & Rünger, D. (2015). Psychology of habit. *Annual Review of Psychology*, *67*(1), 1–26. http://dx.doi.org/10.1146/annurev-psych-122414-033417.

Xu, Z., Hasselt, H. van, Hessel, M., Oh, J., Singh, S., & Silver, D. (2020). Gradient reinforcement learning with an objective discovered online. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems (Vol. 33)* (pp. 15254–15264). Curran Associates, Inc.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. http://dx.doi.org/10.1073/pnas.1403112111.

Yang, L., Michaels, J. A., Pruszynski, J. A., & Scott, S. H. (2011). Rapid motor responses quickly integrate visuospatial task constraints. *Experimental Brain Research*, *211*(2), 231–242. http://dx.doi.org/10.1007/s00221-011-2674-3.

Yawata, S., Yamaguchi, T., Danjo, T., Hikida, T., & Nakanishi, S. (2012). Pathway-specific control of reward learning and its flexibility via selective dopamine receptors in the nucleus accumbens. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(31), 12764–12769. http://dx.doi.org/10.1073/pnas.1210797109.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, *19*(1), 181–189. http://dx.doi.org/10.1111/j.1460-9568.2004.03095.x.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2005). Blockade of NMDA receptors in the dorsomedial striatum prevents action–outcome learning in instrumental conditioning. *European Journal of Neuroscience*, *22*(2), 505–512. http://dx.doi.org/10.1111/j.1460-9568.2005.04219.x.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action–outcome contingency in instrumental conditioning. *Behavioural Brain Research*, *166*(2), 189–196. http://dx.doi.org/10.1016/j.bbr.2005.07.012.

Yin, H. H., Mulcare, S. P., Hilário, M. R. F., Clouse, E., Holloway, T., Davis, M. I., Hansson, A. C., Lovinger, D. M., & Costa, R. M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nature Neuroscience*, *12*(3), 333–341. http://dx.doi.org/10.1038/nn.2261.

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, *22*(2), 513–523. http://dx.doi.org/10.1111/j.1460-9568.2005.04218.x.

Yu, C., Gupta, J., Chen, J.-F., & Yin, H. H. (2009). Genetic deletion of A2A adenosine receptors in the striatum selectively impairs habit formation. *The Journal of Neuroscience*, *29*(48), 15100–15103. http://dx.doi.org/10.1523/jneurosci.4215-09.2009.