

# Hierarchical control of goal-directed action in the cortical–basal ganglia network

Bernard W Balleine<sup>1</sup>, Amir Dezfouli<sup>1</sup>, Makato Ito<sup>2</sup> and Kenji Doya<sup>2</sup>



Goal-directed control depends on constructing a model of the world that maps actions onto specific outcomes, allowing choice to remain adaptive when the values of outcomes change. In complex environments, however, such models can become computationally unwieldy. One solution to this problem is to develop a hierarchical control structure within which more complex, or abstract, actions are built from simpler ones. Here we review findings suggesting that the acquisition, evaluation and execution of goal-directed actions accords well with predictions from hierarchical models. We describe recent evidence that hierarchical action control is implemented in a series of feedback loops integrating secondary motor areas with the basal ganglia and describe how such a structure not only overcomes issues of dimensionality, but also helps to explain the formation of actions sequences, action chunking and the relationship between goal-directed actions and habits.

## Addresses

<sup>1</sup> Brain & Mind Research Institute, University of Sydney, Australia

<sup>2</sup> Okinawa Institute of Science and Technology Graduate University, Japan

Corresponding author: Balleine, Bernard W  
([Bernard.balleine@sydney.edu.au](mailto:Bernard.balleine@sydney.edu.au))

Current Opinion in Behavioral Sciences 2015, 5:1–7

This review comes from a themed issue on **Neuroeconomics**

Edited by **John P O'Doherty** and **Colin C Camerer**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 17th June 2015

<http://dx.doi.org/10.1016/j.cobeha.2015.06.001>

2352-1546/© 2015 Elsevier Ltd. All rights reserved.

## Introduction

### Goal-directed decision-making

A decision-making process is called goal-directed if the actions with which it is concerned are evaluated with respect to their causal relationship to, and the value of, the goals that the decision-maker desires to attain. The capacity for goal-directed action has been demonstrated in both human and non-human animals [1]. This evidence has, therefore, also established that the brain is able to build an internal model of the environment that

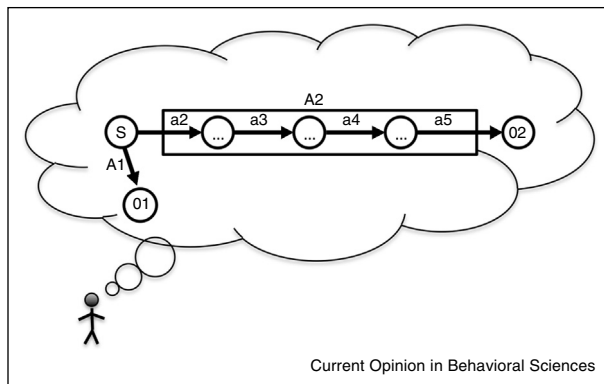
represents the relationship between actions and their specific outcomes [2–4].

Despite being widely conserved, the capacity to build such internal models is not in itself sufficient to explain successful decision-making. In complex environments the increasing number of alternative actions can result in computational challenges — the, so-called, curse of dimensionality. A simple solution to this problem is for the decision-maker to construct macroscopic actions by combining simpler actions so as to avoid having to make all decisions at the level of the most primitive action [5]. The formation of new actions can be as simple as concatenating simpler actions together to form action sequences [6,7], or it can involve a more complex configuration of simple actions into what have been called abstract actions [8]. No matter how they are structured, however, these new actions have to be integrated within the goal-directed decision-making process and one way in which they can do so is by furnishing a hierarchical goal-directed decision-making system [9,10,11•] (see [Figure 1](#)). In such a system, abstract action representations at the upper level facilitate the flexible search for goal-directed actions whereas primitive actions at the lower level can be flexibly reused for different goals as needed ([Figure 1](#)).

## Hierarchical reinforcement learning

Reinforcement learning (RL) is a computational framework for learning actions by exploration and reward feedback [12], which has served as the theoretical basis for the neuroscience of decision-making [13]. Although there have been theorems assuring learning of optimal behaviors after sufficient exploration, in the application of RL to large degree-of-freedom problems, such as the control of humanoid robots, the need for a huge number of explorative trials makes the use of simple RL impractical. This problem motivated the development of hierarchical reinforcement learning (HRL) frameworks. One class of HRL algorithm, for example, ‘Options’ [14] and MAXQ learning [15], assumes a single reward function and derives efficient ways of estimating the values of both the abstract actions and the subgoals achieved by them. Another class of HRL assumes different reward functions for higher (abstract) and lower level (simple) actions; the higher level for achievement of an extrinsic goal and the lower level for achievement of the subgoals specified by the higher level [16,17]. In either case, HRL has two

Figure 1



An example illustrating the representation of hierarchical goal-directed decision-making. The decision-maker is in state 'S' and there are two available actions 'A1' and 'A2'. The outcome of taking the simple action 'A1' is O1, and of taking the more complex action sequence 'A2' is O2. 'A2' is an abstract action, which is composed of simpler actions a2 to a5. Decision-making evaluates both O1 and O2 and selects one based on their relative value. If outcome O1 is chosen then the agent will take A1. If O2 is chosen then the agent will take abstract action A2 after which the control of actions will transfer to a lower level process to implement A2. Such an implementation can involve simply executing actions a2 to a5 in a sequence.

major advantages: first, it allows rapid learning of the entire task by searching with abstract states and actions, that is, **the decision-maker can skip over the intermediate effects of individual actions and directly evaluate the overall outcome of abstract actions**, and, second, **it provides the ability to re-use the same component actions for multiple goals**. The effectiveness of HRL, however, depends on how accurately the hierarchy used by the learner matches the actual structure of the environment. How the appropriate hierarchy of actions is established is a field of ongoing work [18].

### Behavioral elements of hierarchical actions

The operation of a hierarchical goal-directed decision-making system at a behavioral level relies on four elements corresponding to the (goal-directed) *acquisition*, *evaluation*, *execution*, and *interruption* of abstract actions. All actions, whether simple or abstract, are evaluated based on the value of their consequences (evaluation); simple actions based on their individual effects, abstract actions based on their composite effects — that is, the outcome of an abstract action is represented based on the action as a whole rather than the outcome of the actions of which it is composed. Studies evaluating this property of abstract actions are scarce; however, recent evidence for this evaluative process has been reported in studies of spatial navigation [19,20] and from experiments assessing goal-directed action sequences in humans and rodents [11,21].

With regard to execution, after choosing an abstract action, lower-level processes come into play in order to

### Box 1 Habits, model-free and model-based hierarchical RL.

In his description of habits, William James wrote: 'If an act requires for its execution a chain, A, B, C, D, E, F, G, etc., of successive nervous events, then in the first performances of the action the conscious will must choose each of these events from a number of wrong alternatives that tend to present themselves; but habit soon brings it about that each event calls up its own appropriate successor, without any alternative offering itself and without any reference to the conscious will, until at last the whole chain, A, B, C, D, E, F, G rattles itself off as soon as A occurs, just as if A and the rest of the chain were fused into a continuous stream.' [67]. James' conception of habit readily maps onto contemporary theories of hierarchical reinforcement learning in which the chain (A, B, C, ...) is considered an abstract action and acquiring habits is equivalent to forming abstract actions from simpler actions. However, in the broader scheme, abstract actions themselves also need to pass through an evaluation and selection process. From William James' point of view, this selection process is cue-triggered; for example, 'The marksman sees the bird, and, before he knows it, he has aimed and shot'. Such a selection process is equivalent to the *model-free* evaluation of actions, which implies that the decision-maker already knows which actions should be taken when encountering a certain cue without needing to evaluate the consequences of those actions. But he also advances an alternative view in which abstract actions are part of the goal-directed process in the service of reaching goals observing that: 'habit simplifies the movements required to achieve a given result'. This resembles *model-based* evaluation in which all actions including abstract actions are evaluated with respect to their outcomes. On this later view, goal-directed actions and habits are parts of the same global machinery to reach goals, whereas in the former view habits constitute a separate system working in parallel to the goal-directed process. A third possibility is a hybrid system in which abstract actions are evaluated in *both* a model-free and model-based manner. Of these three possibilities, the model-based evaluation of abstract actions appears to be both necessary and sufficient to explain the available behavioral data [9], however some fMRI studies in humans have reported the footprint of model-free evaluation of abstract actions [68,69], lending support to the hybrid scheme.

execute and expand the selected action into simple actions, implying that these lower level processes have learned to perform the abstract action (acquisition). Such an acquisition process could be especially important, particularly at finer (sensorimotor) levels, for the rapid formation of some forms of habitual action (such as those induced by interval schedules of reinforcement [9,22]) — see Box 1. Here, there is a body of evidence showing that humans [23,24], non-human primates [25,26], and rodents [27] are able to form and execute sequences of actions as a single response unit. The formation of these response units is marked by fast reaction times between the elements of the action sequences as well as the feedback-free operation of the sequences, meaning that, once started, a sequence will run to its end without requiring feedback from the performance of each individual action [28,29].

Nevertheless, the performance of abstract actions is not unregulated; once initiated, a sequence of simple actions can be interrupted [30] both in situations in which there is a better alternative than completing the current abstract

action, resulting in a new action inserted in its place [11<sup>•</sup>,31], or where ongoing sequences need rapidly to be inhibited without replacement, for example, in the stop-signal task [32<sup>••</sup>].

## Neural substrates of hierarchical goal-directed decision-making

At a neural level, the parallel organization of the cortico-basal ganglia circuit has been suggested as a possible substrate for hierarchical action control and the implementation of hierarchical reinforcement learning [26,33,34,36] – see Figure 2. The network linking the basal ganglia and the cerebral cortex is known to be topographically organized and to be subdivided into three major loops: first, a motor loop; second, a cognitive loop; and third, an affective (limbic) loop [32<sup>••</sup>,37–40]. The motor loop takes care of detailed motor commands like stepping forward or reaching an arm to a certain position. The cognitive loop plans actions at an abstract level, such as choosing the left button over the right, irrespective of which hand or finger is used. The affective loop may decide which of these actions to implement given current motivational demands [10,41<sup>•</sup>,42<sup>•</sup>]. At an early stage of

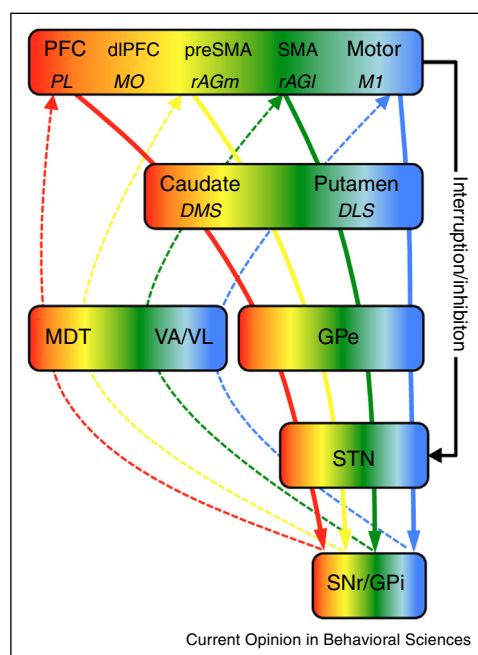
learning, choice behavior is guided by the cognitive loop using more abstract, extrinsic spatial coordinates and each selected sub goal is converted to detailed motor actions. After repeated experience, however, the motor loop can learn the appropriate sequence of detailed motor actions [35]. The algorithms used in cognitive and motor loops can be based on either model-based or model-free RL. However, when a task is designed as a simple rule referring to extrinsic objects, such as the positions or colors of buttons, building an internal model of the task is made much easier using abstract states and actions than detailed motor variables, which favors the use of a model-based strategy by the cognitive loop. The motor loop can benefit from the focused search for relevant motor states and actions guided by the cognitive loop, which makes model-free learning feasible despite its huge search space [43]. The operation of the motor loop can be further facilitated by internal models of sensorimotor dynamics, which are thought to be encoded by the cerebellum by allowing it to work with the predicted rather than the actual sensory state, which is subject to feedback delay [3].

A recent study that compared the neural representations in the dorsolateral, dorsomedial, and ventral striatum showed that the representation of task-level action values (for example, goodness of choosing left) was most prominently found in dorsomedial striatum, a part of the cognitive loop [41<sup>•</sup>]. Although the authors used a model-free reinforcement learning algorithm to estimate the action values from animals' behavior, it is difficult to specify whether the neural activity was due to model-free or model-based computation as their differences are minor in their task. However, neural activity changed swiftly in accordance with the animals' behavior following changes in reward setting, which is consistent with control of this activity by the goal-directed decision system.

## Encoding and retrieving abstract goal-directed actions

There have been considerable advances in our understanding of the neural processes that mediate hierarchical action control and its involvement in the acquisition, evaluation, execution and interruption of abstract actions. There is, for example, broad recognition that a circuit involving medial prefrontal cortical inputs to the dorsomedial striatum is critically involved in the acquisition of goal-directed actions as part of a corticostriatal loop, that feedbacks to the medial prefrontal cortex via the basal ganglia and mediodorsal thalamus [1,44]. However, the role of this medial prefrontal cortex-basal ganglia loop in ongoing goal-directed decision-making is far less certain; evidence suggests it is only engaged during new learning [45] and, therefore, inputs to striatum from other control structures must be involved subsequently in the evaluation and execution of goal-directed actions. Chief amongst these are the secondary motor areas particularly the pre-supplementary motor area (pre-SMA) in humans and primates and homologous rostral medial agranular

Figure 2



The topographically organized cortical-basal ganglia loops proposed to mediate hierarchical action control. Whereas a more medial loop integrating PFC (PL in the rat) with caudate (DMS) is engaged in goal-directed learning, we propose loops integrating dlPFC (MO) and preSMA (rAGm) provide support for cognitive control of hierarchical action selection and involving SMA (rAGl) and motor areas with putamen (DLS) in a motor loop in the implementation of abstract actions by unpacking sequences of simple actions. Loops feedback to cortex topographically, as coded by color, though basal ganglia (GPe, STN and SNr/GPi) and motor thalamus (MDT/VA/VL). See text for abbreviations.

cortex (rAGm) in rodents [46] (Figure 2). Unlike the supplementary motor area, the pre-SMA has no direct connections to spinal motor pools or motor cortex. It projects both directly and indirectly to the dorsomedial striatum via key frontal cortical areas, for example, the dorsolateral prefrontal cortex (in humans) and medial orbital cortex (in rodents) and receives direct feedback from the basal ganglia via the mediodorsal thalamus [46–48].

In line with this involvement, activity in the pre-SMA occurs prior to self-initiated movements when subjects are free to choose rather than when actions are instructed by environmental stimuli [49,50]. Neurons in the pre-SMA also increase their activity prior to the performance of action sequences [25,51] implying that the pre-SMA is involved in the evaluation and selection of both abstract and simple goal-directed actions. Where the values of these simple and abstract actions are compared is an open question, although some recent evidence suggests that the dlPFC is involved in the comparison of action values for decision-making in humans and exerts control over the motor system in a response specific manner [52\*]. Furthermore, whereas, at least in mice, lesions to rAGm abolish goal-directed control of simple actions [53,54], these lesions spare choice between such actions whilst abolishing the ability of rats to choose between goal-directed action sequences [21]. Hence, this loop appears to play a key role both in maintaining the performance of individual actions and in comparing abstract action values for decision-making, suggesting that the hierarchical control of goal-directed actions involves this pre-SMA (rAGm)-basal ganglia feedback circuit.

More recently, the distinction between simple and abstract actions has been proposed to underlie the difference in goal-directed and habitual action control [9\*,11\*\*,55\*] (see Box 1). On this account, the performance of individual goal-directed actions becomes habitual through their incorporation into action sequences, which are open to selection by the hierarchical controller. Another way of describing hierarchical control is, therefore, in terms of interactions in selection between individual goal-directed actions and goal-directed sequences of habitual actions. Interestingly, habit learning have been found to involve dorsolateral striatum (putamen in humans) [56,57], which has also been implicated in the acquisition and chunking of action sequences in rodents [58,59\*\*] and humans [24\*\*] and lesions of this structure appear to cause a deterioration of sequence learning [60] and to enhance the control of individual goal-directed actions when these processes are put into competition [61]. This account is, therefore, broadly consistent with prior findings suggesting that the neural bases of goal-directed actions and habits can be dissociated across the medial and lateral regions of the dorsal striatum (see Box 2).

#### Box 2 Interpreting the effect of striatal manipulations within the current framework.

Prior research has typically reported that manipulations of dorsomedial striatum (DMS), using lesions or pharmacological inactivation, render actions less goal-directed (i.e., insensitive to changes in outcome value or the action-outcome contingency) [1,44] whereas similar manipulations of the dorsolateral striatum (DLS) render actions less habitual [57]. According to the current framework this means to say that DMS manipulations detach actions from their consequences in a manner that makes the selection of simple actions appear more nearly random in a two choice situation. DMS manipulations may affect the selection of abstract actions similarly; however, it is possible that they affect the selection of single actions without affecting the goal-directed control of action sequences and the available data are consistent with both of these alternatives. In contrast, manipulations of DLS should be predicted to reduce the selection of abstract actions and so favor the selection of individual actions either because the former have been abolished or because the hierarchical controller is unable to identify the consequences of abstract actions or determine the value of those consequences.

It is not currently known what effect is produced by direct neural manipulations of the hierarchical controller and predictions from this hierarchical framework are equivocal; it is not clear whether the current account necessarily assumes that all goal-directed actions are under hierarchical control. Being model based, it does, however, assume that hierarchical control is goal-directed. If all goal-directed actions are hierarchically controlled then the loss of hierarchical control would be predicted to impinge on action selection generally; however, some prior data are inconsistent with this conclusion [21]. If goal-directed actions aren't all hierarchically controlled then only the selection of those that are will be affected; for example, loss of the hierarchical controller would, by necessity, impair the goal-directed evaluation of abstract actions meaning that one might be able to 'do A then do B then do C' to reach a specified goal without being able to evaluate or execute the chunked sequence 'ABC' with respect to that goal.

It appears likely, therefore, that there are two components to the neural bases of hierarchical control; one mediating the selection of both individual goal-directed actions and abstract actions involving a pre-SMA basal ganglia loop and a second involving the execution of those actions by unpacking the performance of abstract actions into component sequences of simple actions within the motor loop. This latter appears to involve a parallel circuit linking supplementary motor area (rostro-lateral agranular cortex in rodents), its projections to the putamen and feedback via the basal ganglia and ventrolateral thalamus; a suggestion that has much in common with prior models of sequence learning such as that described by Hikosaka and colleagues [26]. How the interaction of these loops is achieved is an open question, however there are grounds for supposing integration could occur within the thalamo-cortical feedback circuit, suggesting similar integration of pre-SMA and SMA controlled loops to that suggested to underlie cognitive–emotional integration during goal-directed learning [42\*]. It remains possible, however, that multiple forms of sequence learning coexist; that there are multiple hierarchical processes within the CNS operating at different levels (cf. Box 1).



### Interrupting action sequences

The regulation of hierarchical control is critical if actions at different levels are to remain adaptive. Importantly, both the pre-SMA and SMA maintain direct connections with the subthalamic nucleus through the hyperdirect pathway and evidence suggests this pathway can rapidly inhibit ongoing motor activity controlled by cortical–basal ganglia circuits [62,63]. How more subtle shaping of ongoing sequences of actions is achieved, for example, the replacement of elements in a sequence or the deflection of a partially completed sequence into another sequence, is, however, not currently known. There are numerous alternative mechanisms that could achieve this; for example, one quite attractive recent suggestion is that it emerges through inhibitory feedback from the external globus pallidus directly to the striatum to inhibit ongoing actions and disinhibit alternatives [64,65].

### Open questions

The evidence for hierarchical control of goal-directed action described above suggests that this solution to the challenges posed by a changing environment has been implemented in biological systems. Nevertheless, it does not demonstrate that this is the only solution or even the most commonly used solution to these issues. Furthermore, there are many open questions relating to the operation of such a controller in the context of goal-directed action. Perhaps the most difficult to resolve is whether, computationally, such a controller is solely model based, selecting from abstract and simple actions at a single level, or whether it involves mixed model-based and model-free processes with higher level control processes supervising the activity of lower level ones.

Whatever structure is favored, however, researchers working on these forms of hierarchical controller have yet to stipulate exactly how values are calculated and compared to guide the selection and execution of higher versus lower level actions within the respective action control hierarchies. One advantage of the single level architecture is its parsimony on this front. Although the brain maintains an impressive array of evaluative processes, it is possible that this variety reflects different levels of processing rather than fundamentally different processors. Alternatively, the recent discovery of ventromedial to dorsolateral differences in the response properties of dopamine neurons [66] and the spiraling connections between dopamine neurons and the striatum [38] may suggest different valuation mechanisms for upper and lower level actions. A standard interpretation of the activation of dorsolateral SNc dopamine neurons by both reward and punishment, as well as stimuli associated with these events, is that they encode motivational salience [66]. However, in a hierarchical reinforcement learning system, a lower-level action that minimizes an aversive outcome needs to be rewarded, even when the overall outcome is negative. A possible role for the aversive

responses of dorsolateral dopamine neurons in the SNc, which project mainly to the dorsolateral striatum, might be to promote avoidance or coping motor responses. Nevertheless, to date dopamine-related reward prediction errors have largely been documented in Pavlovian conditioning; what different RPE's contribute to goal-directed action is currently one of the more important open questions.

Finally, although the integration of goal-directed and habitual actions within a single model-based hierarchical controller is, we believe, a significant advance, it is clear that the relationship between actions and habits is not always so easily integrated. Indeed, this relationship can be a highly competitive one with intrusive inhibition, disinhibition and failures of inhibition characterizing this interaction and resulting in suboptimal outcomes. Understanding not just the inhibition of habits but their regulation within a hierarchical goal-direct control process will be a critical test for any unified theory of adaptive behavior.

### Conflict of interest

Nothing declared.

### Acknowledgements

This work was supported by grants from the Australian Research Council, grants FL0992409 and DP150104878, and by a Senior Principal Research Fellowship, APP1079561, from the National Health & Medical Research Council of Australia to BWB and by KAKENHI Grants 23120007 to KD and 26120729, and 25430017 to MI.

### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Balleine B, O'Doherty J: **Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action.** *Neuropsychopharmacology* 2010, **35**:48–69.
  2. Doya K: **Complementary roles of basal ganglia and cerebellum in learning and motor control.** *Curr Opin Neurobiol* 2000, **10**:732–739.
  3. Wolpert DM, Miall RC, Kawato M: **Internal models in the cerebellum.** *Trends Cogn Sci* 1998, **2**:338–347.
  4. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control.** *Nat Neurosci* 2005, **8**:1704–1711.
  5. Botvinick MM, Niv Y, Barto AC: **Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective.** *Cognition* 2009, **113**:262–280.
  6. Rhodes BJ, Bullock D, Verwey WB, Averbeck BB, Page MP: **Learning and production of movement sequences: behavioral, neurophysiological, and modeling perspectives.** *Hum Mov Sci* 2004, **23**:699–746.
  7. Lashley KS: **The problem of serial order in behavior.** In *Cerebral Mechanisms in Behavior*. Edited by Jeffress LA. Wiley; 1951:112–136.
  8. Hengst B: **Hierarchical approaches.** In *Reinforcement Learning: State-of-the-Art*. Edited by Wiering M, Otterlo MV. 2012:293–323. Provides a useful introduction to hierarchical reinforcement learning and the algorithms that it uses to encode the hierarchical structure of the environment from state interactions.

9. Dezfouli A, Balleine BW: **Habits, action sequences and reinforcement learning.** *Eur J Neurosci* 2012, **35**:1036-1051.  
Describes a model of habit learning proposing that habits are macroactions composed of sequences of microactions under the control of a hierarchical controller that can select simple goal-directed actions or macroactions to reach goals. Simulations of outcome devaluation and contingency degradation effects are also described.
10. Ito M, Doya K: **Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit.** *Curr Opin Neurobiol* 2011, **21**:368-373.
11. Dezfouli A, Balleine BW: **Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized.** *PLoS Comput Biol* 2013, **9**:e1003364.  
Provides a test of hierarchical model of habit learning as habit sequences using a two-stage discrimination task to examine choices made after reward and non-reward both at the first stage of the discrimination and over the two stages as a sequence of actions.
12. Sutton RS, Barto AG: *Reinforcement Learning*. Cambridge, MA: MIT Press; 1998.
13. Doya K: **Reinforcement learning: computational theory and biological mechanisms.** *HFSP J* 2007, **1**:30-40.
14. Sutton RS, Singh S, Precup D, Ravindran B: **Improved switching among temporally abstract actions.** *Advances in Neural Information Processing Systems 11*. 1999.
15. Dietterich TG: **Hierarchical reinforcement learning with the MAXQ value function decomposition.** *J Artif Intell Res* 2000, **13**:227-303.
16. Dayan P, Hinton GE: **Feudal reinforcement learning.** In *Advances in Neural Information Processing Systems 5*. Edited by Giles CL, Hanson SJ, Cowan JD. 1993:271-278.
17. Morimoto J, Doya K: **Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning.** *Robot Auton Syst* 2001, **36**:37-51.
18. Elfving S, Doya K, Christensen HI: **Evolutionary development of hierarchical learning structures.** *IEEE Trans Evol Comput* 2007, **11**:249-264.
19. Solway A, Diuk C, Cordova N, Yee D, Barto AG, Niv Y, Botvinick MM: **Optimal behavioral hierarchy.** *PLoS Comput Biol* 2014, **10**:e1003779.  
Describes a mathematical account and experiments in humans comparing the different ways of decomposing task structure to build action hierarchies. The model describes how optimal hierarchies of actions form given a specific task set and the experiments in humans demonstrate that these optimal hierarchies are spontaneously discovered.
20. Wiener JM, Mallot HA: **'Fine-to-coarse' route planning and navigation in regionalized environments.** *Spat Cogn Comput* 2003, **34**:331-358.
21. Ostlund SB, Winterbauer NE, Balleine BW: **Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex.** *J Neurosci* 2009, **29**:8280-8287.
22. Dickinson A: **Instrumental conditioning.** In *Animal Cognition and Learning*. Edited by Mackintosh NJ. Academic Press; 1994:4-79.
23. Rosenbaum DA, Cohen RG, Jax SA, Weiss DJ, van der Wel R: **The problem of serial order in behavior: Lashley's legacy.** *Hum Mov Sci* 2007, **26**:525-554.
24. Wymbs NF, Bassett DS, Mucha PJ, Porter MA, Grafton ST: **Differential recruitment of the sensorimotor putamen and frontoparietal cortex during motor chunking in humans.** *Neuron* 2012, **74**:936-946.  
Describes experiments comparing the neural loci of action concatenation and action segmentation during the chunking of a sequence of action elements using fMRI in humans. The authors describe activity in bilateral putamen associated with the concatenation of the sequence and a frontoparietal network associated with segmentation of the sequence.
25. Tanji J: **Sequential organization of multiple movements: involvement of cortical motor areas.** *Annu Rev Neurosci* 2001, **24**:631-651.
26. Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, Doya K: **Parallel neural networks for learning sequential procedures.** *Trends Neurosci* 1999, **22**:464-471.
27. Jin X, Costa RM: **Start/stop signals emerge in nigrostriatal circuits during sequence learning.** *Nature* 2010, **466**:457-462.
28. Pew RW: **Acquisition of hierarchical control over the temporal organization of a skill.** *J Exp Psychol* 1966, **71**:764-771.
29. Keele SW: **Movement control in skilled motor performance.** *Psychol Bull* 1968, **70**:387-403.
30. Sutton RS, Precup D, Singh S: **Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning.** *Artif Intell* 1999, **112**:181-211.
31. Matsumoto N, Hanakawa T, Maki S, Graybiel AM, Kimura M: **Role of [corrected] nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner.** *J Neurophysiol* 1999, **82**:978-998.
32. Logan GD, Van Zandt T, Verbruggen F, Wagenmakers EJ: **On the ability to inhibit thought and action: general and special theories of an act of control.** *Psychol Rev* 2014, **121**:66-95.  
This review paper updates and significantly expands the original race model of Logan and Cowan (1984) used to estimate stopping time in the stop-signal task and revealing a subject's capacity to inhibit specific actions. This model incorporates decision-making variables into models of stop and go processes in action control.
33. Haber SN, Knutson B: **The reward circuit: linking primate anatomy and human imaging.** *Neuropsychopharmacology* 2010, **35**:4-26.
34. Haruno M, Kawato M: **Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning.** *Neural Netw* 2006, **19**:1242-1254.
35. Nakahara H, Doya K, Hikosaka O: **Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuo-motor sequences – a computational approach.** *J Cogn Neurosci* 2001, **13**:626-647.
36. Samejima K, Doya K: **Multiple representations of belief states and action values in corticobasal ganglia loops.** *Ann N Y Acad Sci* 2007, **1104**:213-228.
37. Alexander GE, Crutcher MD: **Functional architecture of basal ganglia circuits: neural substrates of parallel processing.** *Trends Neurosci* 1990, **13**:266-271.
38. Haber SN, Fudge JL, McFarland NR: **Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum.** *J Neurosci* 2000, **20**:2369-2382.
39. Middleton FA, Strick PL: **Basal ganglia and cerebellar loops: motor and cognitive circuits.** *Brain Res Rev* 2000, **31**:236-250.
40. Voorn P, Vanderschuren LJ, Groenewegen HJ, Robbins TW, Pennartz CM: **Putting a spin on the dorsal-ventral divide of the striatum.** *Trends Neurosci* 2004, **27**:468-474.
41. Ito M, Doya K: **Distinct neural representation in the dorsolateral, dorsomedial, and ventral parts of the striatum during fixed- and free-choice tasks.** *J Neurosci* 2015, **35**:3499-3514.  
Compares the distinct functions of phasically active neurons in dorsomedial, dorsolateral and ventral striata during decision-making, finding state information in the ventral striatum, action and action value information in dorsomedial striatum and action command information in the dorsolateral striatum consistent with a hierarchical reinforcement learning model of the interaction of these striatal areas.
42. Balleine BW, Morris RW, Leung BK: **Thalamocortical integration of instrumental learning and performance and their disintegration in addiction.** *Brain Res* 2014 <http://dx.doi.org/10.1016/j.brainres.2014.12.023>.  
Advances a systems integration hypothesis of the interaction of the processes engaged during instrumental learning and performance within the thalamocortical circuit; particularly integration of action-outcome encoding with the value based processes that mediate changes in experienced value and predicted value as assessed by outcome devaluation and Pavlovian-instrumental transfer, respectively.

43. Uchibe E, Doya K: **Competitive-cooperative-concurrent reinforcement learning with importance sampling.** *International Conference on Simulation of Adaptive Behavior: From Animals to Animats (SAB2004)*; Los Angeles, CA, USA, 13–17 July: 2004:287–296.
44. Balleine BW, Delgado MR, Hikosaka O: **The role of the dorsal striatum in reward and decision-making.** *J Neurosci* 2007, **27**:8161–8165.
45. Ostlund SB, Balleine BW: **Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning.** *J Neurosci* 2005, **25**:7763–7770.
46. Reep RL, Corwin JV, Hashimoto A, Watson RT: **Efferent connections of the rostral portion of medial agranular cortex in rats.** *Brain Res Bull* 1987, **19**:203–221.
47. Nachev P, Kennard C, Husain M: **Functional role of the supplementary and pre-supplementary motor areas.** *Nat Rev Neurosci* 2008, **9**:856–869.
48. Reep RL, Corwin JV, King V: **Neuronal connections of orbital cortex in rats: topography of cortical and thalamic afferents.** *Exp Brain Res* 1996, **111**:215–232.
49. Nachev P, Rees G, Parton A, Kennard C, Husain M: **Volition and conflict in human medial frontal cortex.** *Curr Biol* 2005, **15**:122–128.
50. Jahanshahi M, Jenkins IH, Brown RG, Marsden CD, Passingham RE, Brooks DJ: **Self-initiated versus externally triggered movements: I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects.** *Brain* 1995, **118**:913–933.
51. Nakamura K, Sakai K, Hikosaka O: **Neuronal activity in medial frontal cortex during learning of sequential procedures.** *J Neurophysiol* 1998, **80**:2671–2687.
52. Morris RW, Dezfouli A, Griffiths KR, Balleine BW: **Action-value comparisons in the dorsolateral prefrontal cortex control choice between goal-directed actions.** *Nat Commun* 2014, **5**:4390.

Describes an experiment assessing the locus of pre-choice action value comparisons in humans using fMRI and dynamic causal modeling establishing that action values are compared using an advantage signal and finding that the dlPFC calculates the advantage of a specific action and modulates the motor cortex in an action specific manner.
53. Yin HH: **The role of the murine motor cortex in action duration and order.** *Front Integr Neurosci* 2009, **3**:23.
54. Gremel CM, Costa RM: **Premotor cortex is critical for goal-directed actions.** *Front Comput Neurosci* 2013, **7**:110.
55. Dezfouli A, Lingawi NW, Balleine BW: **Habits as action sequences: hierarchical action control and changes in outcome value.** *Philos Trans R Soc Lond B Biol Sci* 2014, **369**.

Examines the influence of outcome devaluation on hierarchical action control finding that changes in value influence the performance of goal-directed actions but not action sequences in a two stage discrimination task in humans. The paper also examines the influence of extinction on the breakdown of action sequences and the recovery of individual goal-directed actions in rats. Both predictions accord with a hierarchical theory of action control.
56. Tricomi E, Balleine BW, O'Doherty JP: **A specific role for posterior dorsolateral striatum in human habit learning.** *Eur J Neurosci* 2009, **29**:2225–2232.
57. Yin HH, Knowlton BJ, Balleine BW: **Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning.** *Eur J Neurosci* 2004, **19**:181–189.
58. Graybiel AM: **The basal ganglia and chunking of action repertoires.** *Neurobiol Learn Mem* 1998, **70**:119–136.
59. Jin X, Tecuapetla F, Costa RM: **Basal ganglia subcircuits •• distinctively encode the parsing and concatenation of action sequences.** *Nat Neurosci* 2014, **17**:423–430.

Assesses chunking of actions in mice using rapid homogeneous action sequences finding neurons sensitive to the starting and stopping point of the sequence as well as neurons that were inhibited or that were excited for the entire action sequence. Importantly, D1 and D2 pathways were found to be synchronised during sequence initiation but not during sequence performance.
60. Yin HH: **The sensorimotor striatum is necessary for serial order learning.** *J Neurosci* 2010, **30**:14719–14723.
61. Bradfield LA, Balleine BW: **Hierarchical and binary associations compete for behavioral control during instrumental biconditional discrimination.** *J Exp Psychol Anim Behav Process* 2013, **39**:2–13.
62. Aron AR, Poldrack RA: **Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus.** *J Neurosci* 2006, **26**:2424–2433.
63. Nambu A, Tokuno H, Takada M: **Functional significance of the cortico-subthalamo-pallidal 'hyperdirect' pathway.** *Neurosci Res* 2002, **43**:111–117.
64. Mallet N, Micklem BR, Henny P, Brown MT, Williams C, Bolam JP, Nakamura KC, Magill PJ: **Dichotomous organization of the external globus pallidus.** *Neuron* 2012, **74**:1075–1086.
65. Abdi A, Mallet N, Mohamed FY, Sharott A, Dodson PD, Nakamura KC, Suri S, Avery SV, Larvin JT, Garas FN et al.: **Prototypic and arkypallidal neurons in the dopamine-intact external globus pallidus.** *J Neurosci* 2015, **35**:6667–6688.
66. Matsumoto M, Hikosaka O: **Two types of dopamine neuron distinctly convey positive and negative motivational signals.** *Nature* 2009, **459**:837–841.
67. James W: **The Principles of Psychology.** New York: Holt; 1890.
68. Diuk C, Tsai K, Wallis J, Botvinick MM, Niv Y: **Hierarchical • learning induces two simultaneous, but separable, prediction errors in human basal ganglia.** *J Neurosci* 2013, **33**:5797–5805.

Examined prediction errors associated with different levels of a hierarchical task structure finding distinct error signals associated with the levels of the tasks processed in parallel thereby satisfying a key prediction of hierarchical reinforcement learning; that is, that distinct error signals can be calculated that temporally span over several states and actions.
69. Ribas-Fernandes JJF, Solway A, Diuk C, McGuire JT, Barto AG, Niv Y et al.: **A neural signature of hierarchical reinforcement learning.** *Neuron* 2011, **71**:370–379.