ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Current Opinion in Neurobiology

Hierarchical reinforcement learning and decision making

Matthew Michael Botvinick

The hierarchical structure of human and animal behavior has been of critical interest in neuroscience for many years. Yet understanding the neural processes that give rise to such structure remains an open challenge. In recent research, a new perspective on hierarchical behavior has begun to take shape, inspired by ideas from machine learning, and in particular the framework of hierarchical reinforcement learning. Hierarchical reinforcement learning builds on traditional reinforcement learning mechanisms, extending them to accommodate temporally extended behaviors or subroutines. The resulting computational paradigm has begun to influence both theoretical and empirical work in neuroscience, conceptually aligning the study of hierarchical behavior with research on other aspects of learning and decision making, and giving rise to some thought-provoking new findings.

Address
Princeton Neuroscience Institute and Department of Psychology,
Princeton University, United States

Corresponding author: Botvinick, Matthew Michael
(matthewb@princeton.edu)

Current Opinion in Neurobiology 2012, **22**:956–962

This review comes from a themed issue on **Decision making**

Edited by **Kenji Doya** and **Michael N Shadlen**

For a complete overview see the Issue and the Editorial

Available online 11th June 2012

0959-4388/$ – see front matter, © 2012 Elsevier Ltd. All rights reserved.

http://dx.doi.org/10.1016/j.conb.2012.05.008

Over recent years, decision making research has gradually embraced the issue of learning, coming to inquire not only how decisions are reached but also how the relevant processes are shaped by past experience. As detailed in several recent reviews (e.g. [1–3]), work on this problem has been increasingly dominated by ideas imported from computational reinforcement learning (RL). Beginning with a proposed link between dopaminergic function and the reward-prediction error signal in temporal-difference (TD) learning [4,5] (Figure 1a), ideas from RL have been used to interpret the functions of numerous cortical and subcortical structures, and to explain a wide range of behavioral phenomena (see e.g. [6–9]).

These applications of RL within neuroscience represent a genuine scientific success story, and there is good reason to celebrate them. However, at the same time, a look back at the computational literature suggests that there may be difficulties lying in wait. For even as RL algorithms have permeated neuroscientific research, computational work has become increasingly focused on the limitations of such algorithms, limitations that may have direct implications for neuroscience.

The curse of dimensionality and the blessing of abstraction

The basic problem is that RL algorithms, like many computational procedures, suffer from a 'curse of dimensionality.' Their effectiveness deteriorates as the size of the learning problem grows [10••]. This scaling problem has obvious relevance for applications of RL within neuroscience, since the learning problems humans and other animals routinely face are notoriously large, involving a wide range of relevant environmental states and action–outcome relationships. With this in mind, it becomes germane for neuroscience to ask how work in computational RL has sought to cope with the scaling problem.

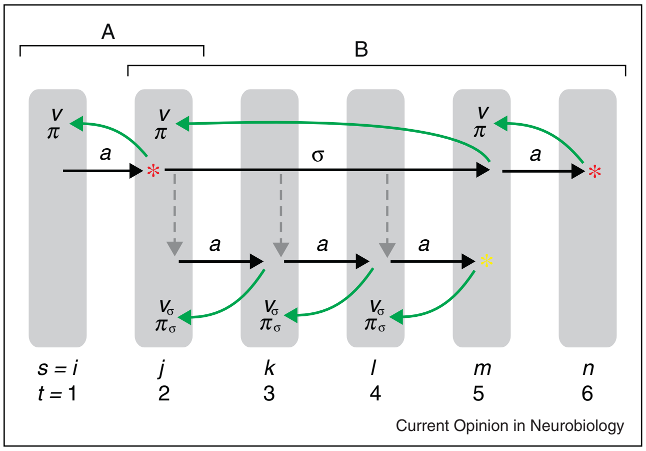
One key method, in this respect, has been to employ some form of abstraction. In *state abstraction*, for example, the learning agent collapses across a number of interrelated environmental states (e.g. specific locations within a single room), treating them as equivalent or interchangeable [11]. By reducing the number of states the agent must learn about, state abstraction effectively shrinks the learning problem, making it potentially easier to conquer.

A second approach, of key interest in the present review, employs *temporal abstraction*. Here, multiple actions (e.g. *open laptop*, *mouse to browser icon*, *double-click*, *enter URL*, *enter password*, among others) are packaged together into discrete subroutines (*check email*). Selecting among such high-level action representations can reduce the number of decisions required to solve a problem, once again making large problems tractable [12•,13].

Hierarchical reinforcement learning

The term *hierarchical reinforcement learning* (HRL) refers to a set of computational techniques that extend standard RL procedures to accommodate temporally abstract actions [10••,12•,13,14•]. Whereas a standard RL agent selects among primitive actions, the HRL agent can also select among subroutines, each associated with its own behavioral policy and its own designated subgoals. TD mechanisms, centering on reward-prediction error signals, allow the agent to learn which subroutines to select in particular situations, as well as to learn subroutine

Figure 1



Schematic representation of basic operations in TD and HRL. At each time-step (gray boxes) the agent arrives in a new situation or state s, and selects an action based on its policy (π) for that state. Each state is also associated with a value (v), which represents a prediction concerning cumulative future reward. Area A: at $t = 1$, the agent selects a primitive action a. Upon arriving in the resulting state (s_j) and receiving primary reward (red asterisk), the agent generates a reward-prediction error (leftward-most green arrow), computed as the difference between the expectation v from s_i and the outcome (the immediate reward received plus the new prediction v for s_j). This PE is used to update both v and π for s_i. If it is positive, v is increased, as is the tendency to select the same action again from that state in future. A negative PE causes the opposite changes. Area B: under HRL, the agent can also select subroutines. In the diagram, at $t = 2$, a subroutine σ is selected, and its policy drives selection of a sequence of primitive actions (lower tier), resulting in receipt of pseudo-reward (yellow asterisk) when the subroutine's subgoal is attained. A pseudo-reward prediction error is computed after each action (lower arrows), and used to update the subroutine's policy (π_σ) and subroutine-specific state values (v_σ). When the subroutine terminates, a PE is generated (extended arrow), and used to update the value and policy for the state in which the subroutine was initiated. A new action is then selected at the top level, yielding primary reward.

policies that strive toward each subroutine's subgoals (see Figure 1b and [10••,12•,13,14•]).

Adding temporal abstraction to RL can have dramatic payoffs for learning, significantly easing the scaling problem. In view of this fact, and given the relevance of the scaling problem to neuroscience, it seems natural to ask whether HRL might shed light on the neural mechanisms underlying decision making. This inquiry is further encouraged by the fact, often noted in behavioral research, that human action is hierarchically organized: Actions cohere into subtask sequences, which fit together to achieve overall task goals [15•,16,17•]. Neuroscientific data point in the same direction, with multiple lines of evidence suggesting that the brain parses ongoing behavior into discrete, bounded segments [18,19,20,21•,22,23]. The question is whether such observations fit into a larger story that can be understood in terms provided by computational HRL.

Potential neural correlates of HRL

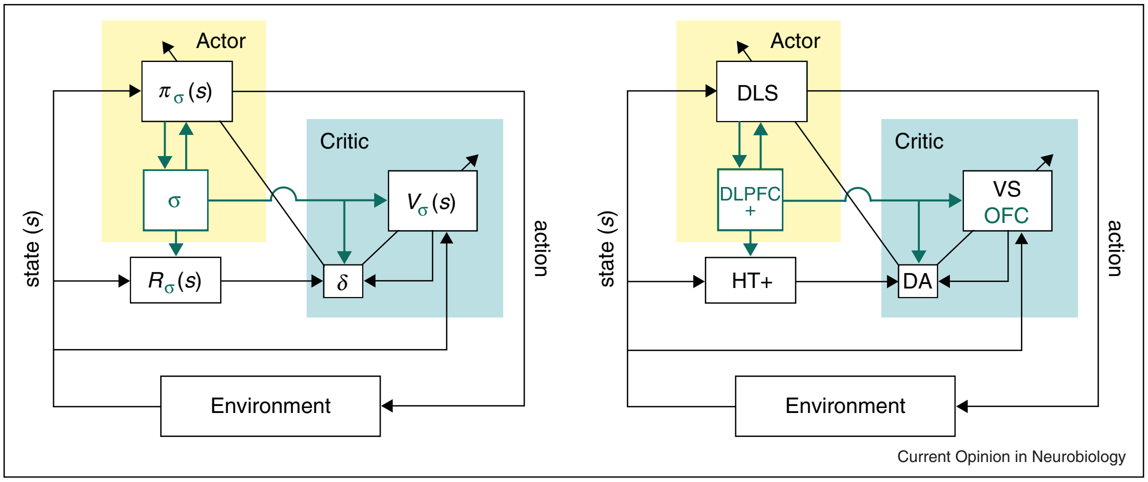
If HRL is relevant to brain function, one should expect to find that each of its 'moving parts' has a clear neural correlate. To evaluate this, Botvinick *et al.* [10••] revisited an influential theory that maps neural structures onto elements within the actor-critic implementation of standard RL [24] (Figure 2). The goal was to determine, based on a computational analysis, what minimum set of structural modifications would be required to allow the actor-critic architecture to support HRL, and then to inquire whether each newly added component might have a plausible neural counterpart.

Botvinick *et al.* [10••] concluded that, in fact, only four basic extensions would be required to implement HRL, each of which matched with credible neural correlates. The most obvious extension, perhaps, involved the addition of a mechanism for representing the currently active subroutine (see Figure 2). Here, a compelling neural counterpart is provided by dorsolateral prefrontal cortex, a structure that has been clearly shown to carry representations of task set [25], marking out subtask sequences and boundaries [19,26•,27] and regulating stimulus–response mappings [28]. Botvinick *et al.* [10••] drew on further neuroscientific findings to map other essential components of HRL to potential correlates in orbitofrontal cortex, ventral striatum and dorsal striatum (another structure that signals the boundaries of temporally extended action events [21•,22,29]). At the same time, they identified several novel predictions arising from HRL, a subset of which have since been put to experimental test.

One prediction arising from HRL is that reward-prediction errors should arise from events at each level of hierarchical task structure (see Figure 1). To test this, Diuk *et al.* [30••] employed a hierarchical gambling task. Here, participants chose repeatedly between two casinos, and then chose, within each casino, among a set of slot machines. Behavioral analyses indicated that participants learned at both levels of the task, discovering both which casino and which slots were most lucrative. fMRI indicated that outcomes at both levels generated independent reward prediction-error signals, detectable within the ventral striatum. Strikingly, prediction errors at the slot-machine and casino levels could be separately identified even when task events triggered them concurrently.

A second prediction from HRL is that TD prediction errors should occur in response to progress toward task subgoals, even when these are not themselves directly associated with primary reward (see Figure 1). Computational HRL attaches a special reward-like signal to the attainment of subgoals, referred to as *pseudo-reward.* Unexpected changes in progress toward a subgoal generate a pseudo-reward prediction error (PPE), a signal that drives learning of subroutine action policies [10••,14•].

Figure 2



Left: schematic of the actor-critic architecture, with extensions as proposed by Botvinick, Niv and Barto [10••]. An 'actor' component houses the agent's action policy π while a 'critic' component houses its state-value function (*V*; see Figure 1). Prediction errors (δ) based on action outcomes, including reward (*R*), drive updating of both the policy and state values. Green elements indicate modifications required by HRL, including mechanisms for representing the currently selected subroutine (σ), and subroutine-specific policies and state-values (highlighted subscripts). Right: putative neural correlates. *Abbreviations*: DA: dopamine; DLS, dorsolateral striatum; HT+: hypothalamus and other structures, potentially including the habenula, the pedunculopontine nucleus, and the superior colliculus; VS, ventral striatum. Structures proposed by Botvinick *et al*. [10••] to participate in implementation of HRL include dorsolateral prefrontal cortex, along with related frontal structures (DLPFC+), and orbitofrontal cortex (OFC).

The PPE does not occur in ordinary, non-hierarchical RL, and thus stands as a distinguishing feature of HRL. Noting this, Ribas-Fernandes *et al.* [31••] used fMRI and EEG, in conjunction with a hierarchical navigation task, to test for a neural correlate of the PPE. Task events that increased the distance to a subgoal location, but left the distance to a final goal location unchanged, triggered focal activation in dorsal anterior cingulate cortex, a structure previously implicated in prediction-error signaling [32].

Related proposals

Taken together with previous findings, the prediction-error phenomena reported by Diuk *et al.* [30••] and Ribas-Fernandes [31••] provide encouraging initial support for the account proposed by Botvinick *et al.* [10••], mapping the functional components of HRL onto a specific network of neural structures. Subsequent to the latter work, several parallel proposals have been put forth for how the brain may implement HRL-like mechanisms.

Building directly on the framework in Botvinick *et al.* [10••], Holroyd and Yeung [33••] have proposed an account centering on dorsal anterior cingulate cortex. Motivated in part by evidence that the cingulate is involved in regulating overall levels of task engagement, they hypothesize that the cingulate in fact serves to select among temporally abstract actions, triggering activity in dorsolateral prefrontal cortex that guides execution of those actions. In the proposed account, as in Botvinick *et al.* [10••], subroutine-specific policies are implemented within the dorsal striatum; state values, computed in relation to current subgoals, are represented in ventral striatum and orbitofrontal cortex; and dopamine drives learning at multiple levels of task structure.

Ito and Doya [34,35] propose a different theory, according to which HRL mechanisms play out across a ventromedial-to-dorsolateral axis within the striatum. Building on anatomical findings indicating a cascading pattern of connectivity along this axis [36], they hypothesize that the ventral striatum (VS), the dorsomedial striatum (DMS), and the dorsolateral striatum (DLS) constitute parallel but hierarchically interrelated learning modules, guiding action selection at progressive levels of temporal granularity: The VS selects top-level goals (e.g. *run the maze*), the DMS intermediate-level actions or subroutines (*turn left*); and DLS fine grained movements (*advance left leg*).

Another HRL-related proposal comes from Reynolds and O'Reilly [37]. This work builds on a computational framework introduced by O'Reilly and Frank [38], within which the dorsal striatum learns, through dopamine-driven TD-like mechanisms (see Figure 1), to regulate or 'gate' activity in prefrontal cortex (see also [39]). Reynolds and O'Reilly [37] extended this model by introducing a hierarchical pattern of connectivity within the prefrontal cortex, an innovation motivated by convergent neuroscientific data (see [40–42]). They applied the resulting model to a task involving cues that must be held in working memory, a task that is hierarchical in the sense that the occurrence of one type of cue sets the context for interpreting subsequent cues of another type. The model's connectivity led to the emergence of a hierarchical distribution of function within its prefrontal sector, with higher-level segments maintaining information about superordinate task cues (see also [43]).

One appealing aspect of the Reynolds and O'Reilly [37] model, shared by the O'Reilly and Frank [38] model from which it derives, is its resonance with empirical findings suggesting that dopaminergic and striatal function play a critical role in the 'chunking' of action routines [23,29,44]. At the same time, because the task employed by Reynolds and O'Reilly [37] demands that context information be maintained over time, it is not clear whether the proposed learning mechanisms would give rise to temporal abstraction in tasks that do not involve this memory requirement (e.g. the Markov decision problems typically considered in work on computational HRL). It is also worth noting that the Reynolds and O'Reilly [37] model forgoes any distinction between primary reward and pseudo-reward, a distinction that is critical to adaptive behavior in computational HRL systems [31••].

In quite recent work, Frank and Badre [45••,46] proposed an HRL-based model motivated by experimental findings suggesting hierarchical processing in prefrontal cortex. The relevant data arise from a task with two conditions: a 'flat' condition where the correct response depends on a full three-way conjunction of stimulus color, shape and orientation; and a 'hierarchical' condition, where the color of the stimulus indicates whether only shape or only orientation is relevant. Studying this task with fMRI, Badre and colleagues [46,47] found that two frontal regions, dorsal premotor cortex (PMd) and an area anterior to it (prePMd) were both initially active. As learning progressed, however, PMd remained active in both conditions, while prePMD activity persisted only in the hierarchical condition. Frank and Badre [45••] interpret this as showing that prePMd is involved in discovering and implementing 'abstract' stimulus–response rules, exerting a hierarchical influence on PMd, which in turn directly regulates response selection. To evaluate this account, they implemented it in a neural network model. Here, through dopamine-based, TD-like mechanisms, the striatum learns to gate color information into prePMd, and to use the resulting pattern of prePMd activity to filter the output of PMd, allowing only information about the task-relevant feature dimension to pass from PMd to motor cortex. In the flat version of the task, prePMd eventually falls inactive, since there is no 'abstract rule' (color-to-feature-dimension mapping) to be learned.

One can view the role of prePMd in the Frank and Badre [47] model as involving temporal abstraction, since it effectively selects between two subroutines (a 'color task' and a 'shape task'). However, note that the prePMd can also be viewed as performing state abstraction: By focusing attention exclusively on task-relevant feature dimensions, it causes disparate stimuli to be treated as equivalent. The close correspondence here between temporal and state abstraction is no coincidence. Both in the neuroscience laboratory [48] and in daily life, different tasks can call for attention to different perceptual feature dimensions: When sorting laundry, color may be most relevant, but when sorting silverware, shape is likely to be more relevant. In response to this point, computational HRL models often link individual subroutines to specific state abstractions, a move that can lead to significant payoffs in learning speed and transfer of knowledge between tasks [49].

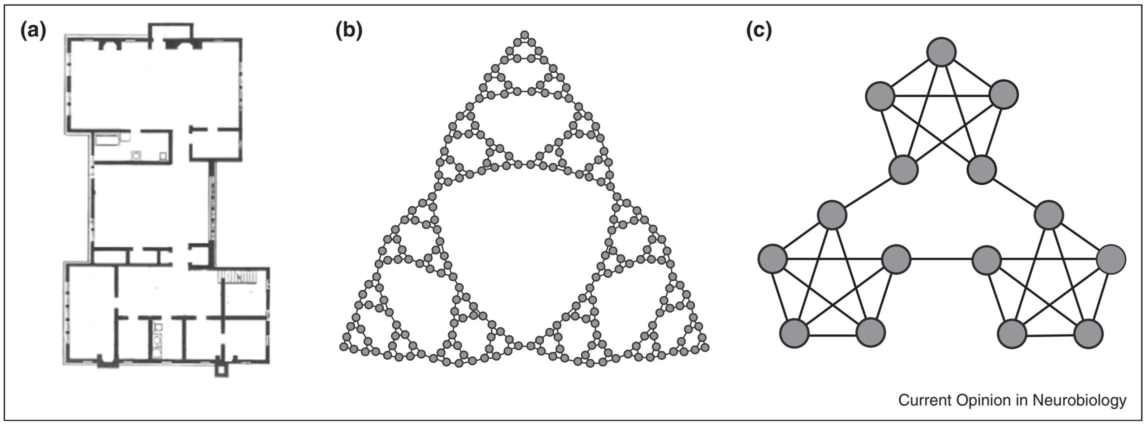
One possibility, consistent with the results of Frank and Badre [47], is that prePMd function is fundamentally tied to state abstraction, and not more generically to temporal abstraction. That possibility strains a bit against the idea that prePMd forms part of a corticotopic hierarchy, where different levels of task structure are represented within different prefrontal regions [40]. That anatomical theory is put under further pressure by recent studies that have failed to confirm some of its predictions [50], as well as by recent neurophysiologic data indicating that multiple levels of task structure may be represented concurrently within the same region of cortex [26•,27,51]. Uithol *et al.* [52•] have recently made the point that, even if prefrontal regions do interconnect hierarchically, the resulting organization might not align with the task-subtask-action hierarchies that characterize sequential behavior. That insight may perhaps prove useful in resolving apparent contradictions in the data currently surrounding the prefrontal-hierarchy hypothesis.

Subgoal discovery and the burden of abstraction

As noted earlier, adding temporally abstract actions to RL can help overcome the curse of dimensionality. However, the question inevitably arises: How are the relevant subroutines initially acquired or learned? HRL furnishes procedures by which the agent can learn to accomplish the subgoals linked to any subroutine. Thus, in the final analysis, the key question is that of how the agent can identify useful *subgoals*. Of course, not just any set of subgoals will do. The agent must identify subgoals that carve tasks 'at their joints,' yielding useful building blocks for the construction of novel behaviors [10••,18,30••,53]. Importantly, the relevant subgoals, unlike top-level goals, will not typically be marked by an association with primary reward. This makes subgoal discovery a vexing problem, arguably the most difficult in HRL.

How do humans and other animals solve the subgoal discovery problem? Putting the HRL literature together with ideas from developmental and cognitive psychology, Botvinick *et al.* [10••] enumerated a set of available

Figure 3



Behavioral bottlenecks in three example domains. **(a)** In a building, doorways, stairways and hallways operate as bottlenecks, affording access from one large set of spatial locations to another. Using such bottlenecks as a basis for temporal and spatial abstraction allows the building to be represented simply as a set of rooms, allowing for a compact specification of navigation plans ('To get from bedroom to kitchen, go through the living room to the dining room, and from there enter the kitchen'). **(b)** A graphical representation of the state-space for the Tower of Hanoi puzzle, in which a set of five disks must be moved among three posts, in order to transform an initial disk arrangement into a goal arrangement. Each node in the graph corresponds to a particular disk configuration, with edges marking feasible moves. The graph displays three salient bottlenecks (edges connecting the three largest clusters). Simsek and Barto [56] showed that an HRL agent that uses these as a basis for temporal abstraction can solve the problem much faster than a standard RL agent. **(c)** The graph employed in experiments by Schapiro and colleagues (reported in [30••]). Bottlenecks correspond to the edges running between star-like clusters.

hypotheses, ranging from evolutionary selection to 'intrinsically motivated' learning [54]. One particularly interesting possibility, for which some neuroscientific evidence can be adduced, centers on the notion of behavioral *bottlenecks*. In many problem domains, it is possible to find states that stand as passages between one large set of states and another (Figure 3). A number of studies in computational HRL have shown that such bottleneck states make natural subgoals [55,56].

Inspired by this observation, a set of recent experiments has investigated whether human learners may attend to behavioral bottlenecks in order to identify useful subgoals. Behavioral experiments show that human learners are rather good at identifying bottleneck states in novel domains [30••], and that such states can indeed provide the basis for spontaneous temporal abstraction. In the latter connection, Schapiro *et al.* (reported in [30••]) presented sequences of visual stimuli, with stimulus order determined by the graph in Figure 3c. Each node in the graph was associated with a specific fractal image, and stimulus sequences were generated based on a random walk through the graph. After some experience with the resulting sequences, participants were asked to parse them wherever event boundaries were perceived. Confirming predictions, participants showed a tendency to parse at moments just following the traversal of a graph bottleneck (see Figure 3c). The notion of temporal abstraction predicts, further, that stimuli lying between such parse-points — that is, stimuli within the same cluster in the underlying graph — should be represented as being similar to one another, by virtue of their joint membership in coherent, temporally extended events. To test this, Schapiro *et al.* [30••] used the same stimulus paradigm with fMRI. Multi-voxel pattern analysis [57] revealed a set of cortical regions displaying the predicted effect. Within these regions, stimuli belonging to the same graph cluster induced activation patterns that were more similar than those induced by stimuli belonging to different clusters. Extrapolating from this finding, Shapiro *et al.* [30••] proposed a view of task decomposition that relates it to computational procedures used to detect local 'communities' in complex networks [58].

Conclusions

While ideas from computational RL have been ubiquitous in neuroscience for some time, attention has expanded to HRL only very recently. Time will tell whether HRL provides a genuinely valid blueprint for understanding neural processes. Clearly, however, the framework has already proven useful in a heuristic role, stimulating novel interpretations of existing data and generating interesting new hypotheses. In several cases, experimental work has yielded new findings that appear to support HRL-based models of learning and decision making, encouraging further investigation.

A number of questions stand out as important targets for the next stage of research. First, what are the more detailed neural mechanisms that might underpin HRL-like learning and decision making? In particular, what role does dopaminergic signaling play? How do the prefrontal cortex and its subsectors contribute, and in what way does this depend on their interactions with striatal mechanisms? Second, in line with the foregoing discussion, how are subroutines initially established and their subgoals selected? Do the apparent 'chunking' phenomena observed in some neuroscientific studies [18–20,21•,22,23] fit into a larger story involving HRL-like mechanisms?

One further set of questions arises from research distinguishing between habit learning and goal-directed planning, tying these to dissociable neural systems [59]. Recent computational analyses draw a parallel between the operations of these two systems and two broad classes of RL: respectively, model-free RL (based on situation-response associations) and model-based RL (involving an internal causal model of the world, used for planning) [8,60]. Given the growing importance of this opposition in neuroscience, it is worth noting that HRL also comes in model-free and model-based varieties [10••,14•,61], with the latter supporting planning or reasoning entirely at the level of subroutines ('First I'll catch the bus to the airport, then I'll check in, and then board my flight…'). Most HRL-based neuroscientific research has so far focused, either explicitly or implicitly, on the model-free case. However, a handful of studies hint at the relevance of model-based HRL to both neuroscience and behavior [30••,62], providing an additional point of entry for the next stage of research.

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. Shah A: **Psychological and neuroscientific connections with reinforcement learning**. In *Reinforcement Learning: State of the Art.* Edited by Wiering M, Van Otterlo M. Springer-Verlag; 2012:507-537.

2. Niv Y: **Reinforcement learning and the brain**. *J Math Psychol* 2009, **53**:139-154.

3. Lee D, Seo H, Jung MW: **Neural basis of reinforcement learning and decision making**. *Annu Rev Neurosci* 2012, **35**:287-308.

4. Montague PR, Hyman SE, Cohen JD: **Computational roles for dopamine in behavioral control**. *Nature* 2004, **411**:760-767.

5. Bromberg-Martin ES, Matsumoto M, Hikosaka O: **Dopamine in motivational control: rewarding, aversive and alerting**. *Neuron* 2010, **68**:815-834.

6. Klucharev V, Hytonen K, Rijpkema M, Smidts A, Fernandez G: **Reinforcement learning signal predicts social conformity**. *Neuron* 2009, **61**:140-151.

7. Maia TV, Frank MJ: **From reinforcement learning models to psychiatric and neurological disorders**. *Nat Neurosci* 2011, **14**:154-162.

8. Solway A, Botvinick MM: **Goal directed decision making as probabilistic inference: a computational framework and potential neural correlates**. *Psychol Rev* 2012, **119**:120-154.

9. Niv Y, Edlund JA, Dayan P, O'Doherty JP: **Neural prediction errors reveal a risk-sensitive reinforcement learning process in the human brain**. *J Neurosci* 2012, **32**:551-562.

10. •• Botvinick MM, Niv Y, Barto AC: **Hierarchically organized behavior and its neural foundations: a reinforcement-learning perspective**. *Cognition* 2009, **113**:262-280.

Surveys potential links between HRL and neuroscience, leveraging a computational analysis to identify potential neural correlates for key elements of HRL.

11. Ponsen M, Taylor ME, Tuyls K: **Abstraction and generalization in reinforcement learning: a summary and framework**. *Adaptive Learn. Agents* 2010, **5924**:1-32.

12. • Hengst B: **Hierarchical approaches**. In *Reinforcement Learning: State of the Art.* Edited by Wiering M, Van Otterlo M. Springer-Verlag; 2012:293-323.

A recent survey of computational HRL.

13. Barto A, Mahadevan S: **Recent advances in hierarchical reinforcement learning**. *Discrete Event Dyn Syst* 2003, **13**:341-379.

14. • Sutton RS, Precup D, Singh S: **Between MDPs and semi-MDPs: a framework for temporal abstraction in reinforcement learning**. *Artificial Intelligence* 1999, **112**:181-211.

Paper originally proposing one of the most influential HRL implementations, known as the 'options' framework.

15. • Logan GD, Crump MJC: **Hierarchical control of cognitive processes: the case for skilled typewriting**. In *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol 54. Edited by Ross BH. Academic Press; 2011:2-19.

Reviews work from one of the few lines of research that have focused on real-time decision making in hierarchical task settings.

16. Shallice T, Cooper RP: *The Organization of Mind*. Oxford, UK: Oxford University Press; 2011.

17. • Botvinick M: **Hierarchical models of behavior and prefrontal function**. *Trends Cogn Sci* 2008.

A brief survey of computational approaches to hierarchical action production, including but not limited to HRL.

18. Zacks JM, Kurby CA, Eisenberg ML, Haroutunian N: **Prediction error associated with the perceptual segmentation of naturalistic events**. *J Cogn Neurosci* 2011, **23**:4057-4066.

19. Fujii N, Graybiel AM: **Representation of action sequence boundaries by macaque prefrontal cortical neurons**. *Science* 2003, **301**:1246-1249.

20. Graybiel AM: **The basal ganglia and chunking of action repertoires**. *Neurobiol Learn Mem* 1998, **70**:119-136.

21. • Barnes TD, Mao J-B, Hu D, Kubota Y, Dreyer AA, Stamoulis C, Brown EN, Graybiel AM: **Advance cueing produces enhanced action-boundary patterns of spike activity in the sensorimotor striatum**. *J Neurophysiol* 2011, **105**:1861-1878.

Latest installment in a series of studies revealing phasic neural activation, across a number of neural structures, coinciding with the onset and termination of temporally extended behaviors.

22. Thorn C, Atallah H, Howe M, Graybiel AM: **Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning**. *Neuron* 2010, **66**:781-795.

23. Jin X, Costa RM: **Start/stop signals emerge in nigrostriatal circuits during sequence learning**. *Nature* 2010, **466**:457-461.

24. O'Doherty J, Dayan P, Schulz L, Diechmann R, Friston KJ, Dolan RJ: **Dissociable roles of ventral and dorsal striatum in instrumental conditioning**. *Science* 2004, **304**:452-454.

25. Sakai K: **Task set and prefrontal cortex**. *Annu Rev Neurosci* 2008, **31**:219-245.

26. • Sigala N, Kusunoki M, Nimmo-Smith I, Gaffan D, Duncan J: **Hierarchical coding for sequential task events in the monkey prefrontal cortex**. *Proc Natl Acad Sci U S A* 2008, **105**:11969-11974.

One of several papers documenting distributed, overlapping representation of multiple levels of task structure within a single local cortical region.

27. Saga Y, Iba M, Tanji J, Hoshi E: **Development of multidimensional representations of task phases in the lateral prefrontal cortex**. *J Neurosci* 2011, **31**:10648-10665.

28. Miller EK, Cohen JD: **An integrative theory of prefrontal cortex function**. *Annu Rev Neurosci* 2001, **24**:167-202.

29. Boyd LA, Edwards JD, Siengsukon CS, Vidoni ED, Wessel BD, Linsdell MA: **Motor sequence chunking is impaired by basal ganglia stroke**. *Neurobiol Learn Mem* 2009, **92**:35-44.

30. •• Diuk C, Schapiro A, Cordova N, Ribas-Fernandes JJF, Niv Y, Botvinick M: **Divide and conquer: task decomposition and hierarchical reinforcement learning in humans**. In *Computational and Robotic Models of the Hierarchical Organization of Behavior*. Edited by Baldassare G, Mirolli M: Springer Verlag; 2012, in press.

Presents a set of recent behavioral and neuroscientific studies evaluating the relevance of HRL to human action selection.

31. •• Ribas-Fernandes JJF, Solway A, Diuk C, Barto AG, NIv Y, Botvinick M: **A neural signature of hierarchical reinforcement learning**. *Neuron* 2011, **71**:370-379.

Presents results from EEG and fMRI experiments, providing evidence for subgoal-related prediction error signaling, as predicted by HRL.

32. Holroyd CB, Coles MGH, Nieuwenhuis S: **Medial prefrontal cortex and error potentials**. *Science* 2002, **296**:1610-1611.

33. •• Holroyd CB, Yeung N: **Motivation of extended behaviors by anterior cingulate cortex**. *Trends Cogn Sci* 2012, **16**:122-128.

Leverages constructs from HRL to propose a novel theory of anterior cingulate cortex function.

34. Doya K: **What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?** *Neural Networks* 1999, **12**:961-974.

35. Ito M, Doya K: **Hierarchical information coding in the striatum during decision making tasks**. *Neurosci Res* 2010, **68**:e187.

36. Haber SN: **Neural circuits of reward and decision making: integrative networks across corticobasal ganglia loops**. In *Neural Basis of Motivational and Cognitive Control.* Edited by Mars RB, Sallet J, Rushworth MFS, Yeung N. MIT Press; 2011:21-35.

37. Reynolds JR, O'Reilly RC: **Developing PFC representations using reinforcement learning**. *Cognition* 2009, **113**:281-292.

Extends an influential computational model of reward-based learning and action selection to incorporate hierarchical structure.

38. O'Reilly RC, Frank MJ: **Making working memory work: a computational model of learning in prefrontal cortex and basal ganglia**. *Neural Comput* 2006, **18**:283-328.

39. Todd MT, Niv Y, Cohen JD: **Learning to use working memory in partially observable environments through dopaminergic reinforcement**. In *Advances in Neural Information Processing Systems (NIPS)*, vol 21. Edited by Koller D. Curran Associates; 2009.

40. Badre D: **Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes**. *Trends Cogn Sci* 2008, **12**:193-200.

41. Koechlin E, Jubault T: **Broca's area and the hierarchical organization of behavior**. *Neuron* 2006, **50**:963-974.

42. Verstynen T, Badre D, Jarbo K, Schneider W, Microstructural organizational patterns in the human corticostriatal system. *J Neurophysiol* in press.

43. Botvinick MM: **Multilevel structure in behaviour and the brain: a model of Fuster's hierarchy**. *Philos Trans R Soc Lond, Series B* 2007, **362**:1615-1626.

44. Tremblay PL, Bedard MA, Langlois D: **Movement chunking during sequence learning is a dopamine-dependent process: a study conducted in Parkinson's disease**. *Exp Brain Res* 2010, **205**:375-385.

45. •• Frank MJ, Badre D: **Mechanisms of hierarchical reinforcement learning in corticostraital circuits 1: computational analysis**. *Cereb Cortex* 2012, **22**:509-526.

Presents neural network and hierarchical Bayesian models of learning in a task with interesting hierarchical structure.

46. Badre D, Frank MJ: **Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: evidence from fMRI**. *Cereb Cortex* 2012, **22**:527-536.

Leverages the models from the previous reference to test predictions concerning neural activity during hierarchical task acquisition and performance, as measured with fMRI.

47. Badre D, Kayser AS, D'Esposito M: **Frontal cortex and the discovery of abstract action rules**. *Neuron* 2010, **66**:315-326.

48. Roy JE MR, Poggio T, Miller EK: **Prefrontal cortex activity during flexible categorization**. *J Neurosci* 2010, **30**:8519-8528.

49. Dietterich TG: **Hierarchical reinforcement learning with the maxq value function decomposition**. *J Artif Intell Res* 2000, **13**:227-303.

50. Reynolds JR, O'Reilly RC, Cohen JD, Braver TS: **The functional organization of lateral prefrontal cortex: a test of competing hypotheses**. *PLoS One* 2012, **7**:e30284.

51. Bonini L, Serventi FU, Simone L, Rozzi S, Ferrari PF, Fogassi L: **Grasping neurons of monkey parietal and premotor cortices encode action goals at distinct levels of abstraction during complex action sequences**. *J Neurosci* 2011, **31**:5876-5886.

52. • Uithol S, Van Rooij I, Bekkering H, Haselager P: **Hierarchies in action and motor control**. *J Cogn Neurosci* 2012, **24**:1077-1086.

Considers alternative interpretations of what form of hierarchy may be pertinent to understanding prefrontal cortical function.

53. Botvinick M, Cohen JD, Computational models of executive control: charted territory and new frontiers. *Cogn Sci*, in press.

54. Vigorito CM, Barto AG: **Intrinsically motivated hieararchical skill learning in structured environments**. *IEEE Trans Autonomous Mental Develop (T-AMD)* 2010:2.

55. Şimşek Ö, Barto AG: In *Skill characterization based on betweenness.* Edited by Koller D, Schuurmans D, Bengio Y, Bottou L..2009, **21**:1497-1504.

56. Moradi P, Shiri ME, Rad AA, Khadivi A, Hasler M: **Automatic skill acquisition in reinforcement learning using graph centrality measures**. *Intell Data Anal* 2012, **16**:113-135.

57. Pereira F, Mitchell T, Botvinick M: **Machine learning classifiers and fMRI: a tutorial overview**. *Neuroimage* 2010, **45**:199-209.

58. Rosvall M, Bergstrom CT: **Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems**. *PLoS One* 2011:6.

59. Balleine BW, O'Doherty JP: **Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action**. *Neuropsychopharmacol Rev* 2010, **35**:48-69.

60. Daw N, Gershman SJ, Seymour B, Dayan P, Dolan RJ: **Model-based influences on humans' choices and striatal prediction errors**. *Neuron* 2011, **69**:1204-1215.

61. Bornstein AM, Daw ND: **Multiplicity of control in the basal ganglia: computational roles of striatal subregions**. *Curr Opin Neurobiol* 2011, **21**:374-380.

62. 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.