Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Neurobiology

Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit

Makoto Ito[1] and Kenji Doya[1,2]

Accumulating evidence shows that the neural network of the cerebral cortex and the basal ganglia is critically involved in reinforcement learning. Recent studies found functional heterogeneity within the cortico-basal ganglia circuit, especially in its ventromedial to dorsolateral axis. Here we review computational issues in reinforcement learning and propose a working hypothesis on how multiple reinforcement learning algorithms are implemented in the cortico-basal ganglia circuit using different representations of states, values, and actions.

Addresses
[1] Neural Computation Unit, Okinawa Institute of Science and Technology, Okinawa 904-0412, Japan
[2] Computational Neuroscience Laboratories, Advanced Telecommunications Research Institute International, Kyoto 619-0288, Japan

Corresponding author: Doya, Kenji (doya@oist.jp)

Current Opinion in Neurobiology 2011, **21**:368–373

This review comes from a themed issue on
Behavioral and cognitive neuroscience
Edited by Ann Graybiel and Richard Morris

Available online 29th April 2011

0959-4388/$ – see front matter

DOI 10.1016/j.conb.2011.04.001

Introduction

The loop network composed by the cerebral cortex and the basal ganglia is now recognized as the major site for decision making and reinforcement learning [1,2]. The theory of reinforcement learning [3] prescribes a number of steps that are required for decision making: 1) recognize the present state of the environment by disambiguating sensory inputs; 2) evaluate the candidate actions in terms of expected future rewards (action values); 3) select an action that is most advantageous; and 4) update the action values based on the discrepancy between the predicted and the actual rewards. Simplistic models of reinforcement learning in the basal ganglia (e.g. [4]) proposed that the cerebral cortex represents the present state and the striatal neurons compute action values [5]. An action is selected in the downstream, the globus pallidus, and the dopamine neurons signal the reward prediction error [6], which enables learning by dopamine-dependent synaptic plasticity in the striatum [7]. Recent studies, however, have shown that the reality may be more complex. Discriminating the environmental state behind noisy observation is in itself a hard problem, known as perceptual decision making [8,9]. Activities related to action values are found not only in the striatum, but also in the pallidum [10,11•] and the cortex [12••]. Different parts of the striatum, especially in its ventromedial to dorsolateral axis, have different roles in goal-directed and habitual behaviors [13]. Action selection may be performed not just in one locus in the brain but by competition and agreement among distributed decision networks [14]. Finally, a subset of midbrain dopamine neurons located in the dorsolateral part signal not only rewarding but also aversive signals [15••].

Based primarily on primate studies, Samejima and Doya [16] proposed that different cortico-basal ganglia subloops realize decisions in motivational, context-based, spatial, and motor domains. In this article, we consider how different algorithms of decision making, such as model-based and hierarchical reinforcement learning algorithms, can be implemented in the cortico-basal ganglia circuit with a focus on the ventromedial to dorsolateral axis in the rodent striatum.

Computational axes in action learning

In looking into the computational mechanisms of decision making and reinforcement learning, there are several axes that are useful for sorting out the process.

From state recognition through valuation to action selection

Where in the brain is the locus of decision making? Exactly how an action is selected may depend on the modality and clarity of sensory evidence, available methods of value estimation, and the preparedness in implementation of the action.

From flexible learning to efficient execution

Early in learning, actions have to be exploratory and require a high cognitive load. As behavior becomes well learned, actions can be smooth, stereotyped, and needing less cognitive demands. Such a transition can be due to changes in the parameters like the 'temperature' for action selection [3], but also the shift in the recruitment of the brain loci implementing different algorithms, most notably model-based, predictive strategies and model-free, retrospective strategies [17].

From whole life to movement details

Animal behaviors have hierarchical structure in time and space. To satisfy the fundamental needs of life like eating, drinking and mating, an animal has to organize a temporal sequence of exploration, approaching, and manipulation involving movements of the whole body to body parts and individual muscles. Dealing with such temporal and physical hierarchy requires proper mechanisms for coordination and credit assignment. Appropriate weighting is needed for the reward attained at the goal and cost or danger incurred along the way, and also for immediate and long-term outcomes.

In the following sections, we will first look into the computational frameworks to deal with such complexities. We will then review experimental works that provide clues as to how they could be implemented in the cortico-basal ganglia network. We will finally propose a working hypothesis on the implementation of hierarchical reinforcement learning in the ventral-dorsal axis of the basal ganglia.

Algorithms for reinforcement learning

Model-free reinforcement learning algorithms

In the basic theory of reinforcement learning, the learning agent does not initially know how its actions affect the environmental state or how much rewards are given in what state. The action value-based algorithms, including Q-learning and SARSA, use actual experience of state, action, and reward to estimate the action value function Q(state,action), which evaluates how much future reward is expected by taking a particular action at a given state. An action can be selected greedily or stochastically by comparing the action values of the candidate actions.

Another popular algorithm for model-free reinforcement learning is the actor-critic, in which the critic learns to predict the future rewards in the form of the state value function V(state) and the actor improves action policy P(action|state) using the reward prediction error as the reinforcement signal. A good feature of actor-critic method is that after sufficient learning, only the actor part is needed for real-time control with less computational demand.

Model-based algorithms

In the classical theory of optimal control and decision making, the knowledge of the environment, i.e. the state transition probability P(new state|state,action) is supposed to be available. Such a model of the environmental dynamics can facilitate decisions and learning in multiple ways.

The typical way is the search for a good action or a sequence of actions that gives the largest rewards. This enables flexible adaptation to a new reward setting. However, as the numbers of available actions, possible outcomes, and steps to reach the goal increase, running a full tree search requires a lot of time and working memory load. The evaluation of intermediate states in the form of state value function can help truncating a deep search.

In addition to such an on-line use for action selection, dynamic models can be used for learning value functions and/or action policies by off-line through simulated experience either forward or backward in time. Another important use of environmental dynamic models is for estimating the state of the environment given past actions and sensory observations [9].

Hierarchical architectures

Hierarchical reinforcement learning algorithms (e.g. [18,19]) have been proposed for dividing a large complex problem into small simpler problems and reusing the solutions for sub tasks to better cope with new situations. A typical way of dividing a task is for an 'action' in a higher level to serve as the context or activation signal for the lower level [20,21]. A proper reward signal should be given upon completion of a specified subtask, even if the entire behavior may not be rewarded. In some cases, the state value for a subtask can be seen as the action value for choosing the subtask by the upper level, which can blur the distinction between the two.

Model-based analysis of learner's variables

In order to describe how an animal's choices change dynamically depending on the reward experience, a straight forward way is to take a Markov model in which the conditional probability of action choice given previous state, action, and reward is computed. Such non-parametric, hypothesis-neutral description is helpful in measuring goodness of more elaborate model-based explanation [11•]. Recent use of normative models, especially those by reinforcement learning algorithms, has turned out to be powerful tools for characterizing subjects' decision strategies and parameters, and searching for their neural correlates [22–25].

One problem in normative model-based behavior analysis is the choice of the learning algorithm and the estimation algorithm. In order to describe the subject's learning process, most studies assume just one or a few learning algorithms, such as Q-learning [5] and its variants [11•,26], or semi-parametric models [27,28]. Different studies use different algorithms for estimation of the model parameters, such as maximal likelihood estimate, Kalman filter [29], and particle filter [5,11•]. Establishment of standard methods and tools is strongly in demand.

Another issue is validation of the models. Common criteria for evaluating model performance are Akaike's Information Criterion (AIC; [28,30,31•,32••]) and Bayesian information criterion (BIC; [29,31•,33–37]). However, a care must be taken as the validity of these measures is under some assumptions, such as incremental complexity. A more robust comparison can be made by cross validation [11•,31•,38]. Furthermore, the best fit models to a subject's sample behavioral sequence may not reproduce the subject's behavior when it is run on its own in the same task. It is important to check if the model reproduces some statistics of the subject's behavior, such as the total reward and learning speed.

Possible implementation in the cortico-basal ganglia network

Based on the computational requirements and possible reinforcement learning algorithms, we now review neural recording and brain imaging results, many of which through the model-based analysis described above, that shed light on how they could be implemented in the cortico-basal ganglia network.

From state recognition through valuation to action selection

Information about states, actions, and rewards has been found in the striatum [11•,31•,39–43]. Furthermore, information of past actions and rewards is also represented in the striatum [11•,31•,42,44,45].

State value

Monkey studies using a free choice task reported that state-value coding neurons were few in the dorsal striatum (DS) [5,46]. In rat studies, the state value coding neurons was found in both DS and the vental striatum (VS), although the proportion was not large [11•,31•]. Human imaging studies suggest that VS play a role of critic in the actor-critic algorithm [47].

Action value

In primates, the action value coding has been reported in DS [5,10,12••,41,46], the internal pallidum [10], and the supplemental motor area [12••]. In rodents, however, the population of action-value coding neurons was reported to be significant but small in both DS and VS [11•,31•,48].

Action command

While the representation of upcoming action has been reported in DS [5,10,31•,49], its representation in the ventral striatum is not consistent, with both positive [48] and negative reports [11•,31•,44], Action command is also represented in the cortical areas in the loop circuit with the striatum, such as the supplementary and presupplementary motor area [50], the prefrontal cortex [49] and the parietal cortex [51].

Chosen value

The action value for selected action Q(selected action), named chosen value, is necessary for comparison with the actually delivered reward for learning. Chosen-value coding was also found in DS in monkeys [46] and in both DS and VS in rats [31•].

Reward prediction error (RPE)

Functional MRI studies have reported RPE in VS [29,32••,34,47,52] and DS [29,47,52]. This is consistent with the RPE coding of the dopamine neurons projecting to the striatum because the fMRI signal is known to respond strongly to the presynaptic inputs [53]. It is reported in rats that a fraction of the striatal neurons coded RPE [54]. Recently Dickerson and Delgado [55] reported in a probabilistic learning task with rats that RPE signal was observed in not only DS but also hippocampus, suggesting the involvement of episodic memory system in decision making. Recently Nomoto *et al.* [56•] showed in monkeys using a random dot motion discrimination task with different rewards for different directions that the midbrain dopamine neurons show two-phase cue responses, the early one in proportion to the average reward and the later one reflecting the deviation from the average.

While dopamine neurons code RPE, serotonin neurons have been hypothesized to code aversive prediction error [57]. However, Miyazaki *et al.* [58] reported increased serotonin activity during waiting for delayed rewards, but not for unexpected reward omission. Aversive prediction error signal was found instead in the lateral habenula neurons [59] and a subset of midbrain dopamine neurons [15••].

From flexible learning to efficient execution

Studies using the devaluation paradigm in rodents showed the distinct roles of dorsomedial and dorsolateral striatum in goal-directed and habitual behaviors, respectively [60–64].

Brain imaging studies investigated possible realization of model-based action and learning strategies, such as a Kalman filter [29] or a hidden Markov model [34], in the cortico-basal ganglia system [17]. Recently, a hybrid model of the model-based and the model-free strategies was reported to show the higher prediction accuracy than that of the single model-base or model-free strategy [32••].

From whole life to movement details

A possible implementation of hierarchical reinforcement learning in the cortico-basal ganglia loop is along the ventro-dorsal axis within the striatum. The ventral striatum is connected with the limbic system, which represent primary reward information and regulates the affects and motivation of the whole animal. The dorsal striatum, on the contrary, is connected with the sensory-motor cortices that control detailed body movements required for acquisition of reward and avoidance of punishment.

A possible reason why action command and action value signals have not been found in VS (but [48]) is that the actions treated in the VS was not like a left and right, but 'do the task' or 'do not'. The nucleus accumbens core in VS has been thought to be an important site for the motivation [65,66] (for review, see [67]).

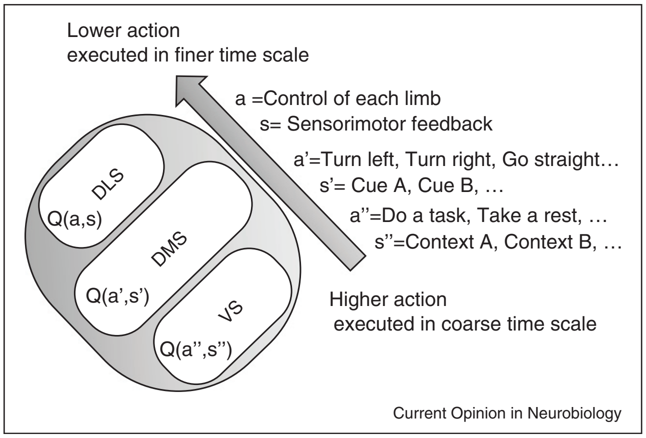
The spiral organization of the connections between the striatum and the dopamine neurons might be used for passing reward signal from the higher level learner to the lower level learner [68]. Recently it was found that some dopamine neurons located dorsally in the substantia nigra and VTA respond to both reward and aversive stimuli [15••]. Although such bidirectional response is generally regarded as encoding salience, another possibility is that they represent reinforcing signal for avoidance behaviors.

Hierarchical reinforcement learning in the cortico-basal ganglia loops

Anatomically and neurophysiologically, DS and VS have the same basic structure and there is no clear boundary [69], suggesting a possibility that DS and VS work with the same mechanism. On the contrary, input from the cortex has a dorsolateral–ventromedial gradient in the modality: the more dorsolateral striatum receives sensorimotor-related information and the more ventromedial part receives associative and motivational information [69]. These striatal subdivisions send their output through the pallidum, the substantia nigra, and the thalamus to the cortical areas to form parallel but partly overlapped loops. A possible reason for such gradient in the input and the output is for implementation of hierarchical reinforcement learning in the striatum [16,68].

Here, we propose a working hypothesis that the ventral striatum (VS), the dorsomedial striatum (DMS), and the dorsolateral striatum (DLS) are parallel and hierarchical learning modules that are in charge of actions at different physical and temporal scales (Figure 1). VS is the coarsest module in charge of the action of the whole animal, such as aiming for a goal, avoiding a danger, or just take a rest. The decision is related to the overall goodness of the choice at a coarse time scale. DMS is the middle module in charge of abstract actions, such as turn left or go straight. DLS is the finest module in charge of physical actions, such as the control of each limb. The finer-level behaviors are more distally and conditionally linked with primary rewards because their sequentially and parallelly coordinated executions are necessary for achieving a goal or avoiding a danger. They also have to be memorized longer term for possible reuse in different contexts. These modules work in parallel [70•], which is why bilateral lesions of one among DLS, DMS and VS, do not impair a performance in a simple instrumental learning [60,71].

Figure 1


Lower action
executed in finer time scale
a =Control of each limb
s= Sensorimotor feedback
a'=Turn left, Turn right, Go straight…
s'= Cue A, Cue B, …
a''=Do a task, Take a rest, …
s''=Context A, Context B, …
Higher action
executed in coarse time scale
DLS
Q(a,s)
DMS
Q(a',s')
VS
Q(a'',s'')
Current Opinion in Neurobiology

A working hypothesis that the dorsolateral (DLS), the dorsomedial (DMS), and the ventral striatum are parallel and hierarchical Q-learning modules that are in charge of actions at different physical and temporal scales.

Such differences in the physical and temporal scales may lead to the differential use of model-free and model-based algorithms in the ventromedial to dorsolateral axis. Model-based search algorithms are useful with coarsely discretized action and state representations in the VS and DMS connected with the prefrontal cortex. Model-free algorithms are more appropriate for DLS that has to deal with finer motor actions in shorter latency.

Conclusion

We reviewed computational issues and possible algorithms for decision making and reinforcement learning and recent findings on the neural correlates of the variables in those algorithms. Then we proposed a working hypothesis: the dorsolateral, the dorsomedial, and the ventral striatum comprise a parallel and hierarchical reinforcement learning modules that are in charge of actions at different physical and temporal scales. The parallelism of the decision modules has been suggested also in the prefrontal cortex [17,32••] and the hippocampal system [55,72,73•]. Manipulation of the different subparts of the parallel networks combined with careful task design and computation model would be necessary to further clarify their specialization and coordination.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Doya K: **Reinforcement learning: computational theory and biological mechanisms**. *HFSP J* 2007, **1**:30-40.

2. Doya K: **Modulators of decision making**. *Nat Neurosci* 2008, **11**:410-416.

3. Sutton RS, Barto AG: *Reinforcement Learning*. Cambridge, MA: MIT Press; 1998.

4. Doya K: **Complementary roles of basal ganglia and cerebellum in learning and motor control**. *Curr Opin Neurobiol* 2000, **10**:732-739.

5. Samejima K, Ueda Y, Doya K, Kimura M: **Representation of action-specific reward values in the striatum**. *Science* 2005, **310**:1337-1340.

6. Schultz W, Dayan P, Montague PR: **A neural substrate of prediction and reward**. *Science* 1997, **275**:1593-1599.

7. Reynolds JN, Hyland BI, Wickens JR: **A cellular mechanism of reward-related learning**. *Nature* 2001, **413**:67-70.

8. Kiani R, Shadlen MN: **Representation of confidence associated with a decision by neurons in the parietal cortex**. *Science* 2009, **324**:759-764.

9. Rao RP: **Decision making under uncertainty: a neural model based on partially observable markov decision processes**. *Front Comput Neurosci* 2010, **4**:146.

10. Pasquereau B, Nadjar A, Arkadir D, Bezard E, Goillandeau M, Bioulac B, Gross CE, Boraud T: **Shaping of motor responses by incentive values through the basal ganglia**. *J Neurosci* 2007, **27**:1176-1183.

11. • Ito M, Doya K: **Validation of decision-making models and analysis of decision variables in the rat basal ganglia**. *J Neurosci* 2009, **29**:9861-9874.

The authors showed that a modified version of Q-learning with a forgetting term was able to predict rats' choice behavior with the highest accuracy among 11 algorithms. They also found that the information about action value was coded in the ventral striatum and the ventral pallidum but was less dominant than information about state, action, and reward.

12. •• Wunderlich K, Rangel A, O'Doherty JP: **Neural computations underlying action-based decision making in the human brain**. *Proc Natl Acad Sci USA* 2009, **106**:17199-17204.

Using a special choice task where subjects were required to select an option by different movements (hand or eye), neuronal activities correlating with the action value were detected for the first time in imaging studies, in the supplementary motor area, the lateral parietal cortex, the anterior cingulate cortex, and the dorsal putamen.

13. Pennartz CM, Berke JD, Graybiel AM, Ito R, Lansink CS, van der Meer M, Redish AD, Smith KS, Voorn P: **Corticostriatal interactions during learning, memory processing, and decision making**. *J Neurosci* 2009, **29**:12831-12838.

14. Cisek P: **Cortical mechanisms of action selection: the affordance competition hypothesis**. *Philos Trans R Soc Lond B Biol Sci* 2007, **362**:1585-1599.

15. •• Matsumoto M, Hikosaka O: **Two types of dopamine neuron distinctly convey positive and negative motivational signals**. *Nature* 2009, **459**:837-841.

The authors found that a large proportion of dopamine neurons located in the dorsolateral parts of the substantia nigra and the ventral tegmental area were excited by not only reward-predicting stimuli but also punishment-predicting stimuli. The signal coded by these neurons might used for signaling saliency, or avoidance learning rather than value learning in the striatum.

16. 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.

17. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control**. *Nat Neurosci* 2005, **8**:1704-1711.

18. 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.

19. Dietterich TG: **Hierarchical reinforcement learning with the MAXQ value function decomposition**. *J Artif Intell Res* 1999, **13**:227-303.

20. Dayan P, Hinton GE: **Feudal reinforcement learning**. In *Advances in Neural Information Processing Systems.* Edited by Cowan JD, Tesauro G, Alspector J. Morgan Kaufmann; 1993.

21. Morimoto J, Doya K: **Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning**. *Rob Auton Syst* 2001, **36**:37-51.

22. Daw ND, Doya K: **The computational neurobiology of learning and reward**. *Curr Opin Neurobiol* 2006, **16**:199-204.

23. Corrado G, Doya K: **Understanding neural coding through the model-based analysis of decision making**. *J Neurosci* 2007, **27**:8178-8180.

24. O'Doherty JP, Hampton A, Kim H: **Model-based fMRI and its application to reward learning and decision making**. *Ann N Y Acad Sci* 2007, **1104**:35-53.

25. Doya K, Ito M, Samejima K: **Model-based analysis of decision variables**. In *Decision Making, Affect, and Learning: Attention and Performance XXIII.* Edited by Delgado MR, Phelps EA, Robbins TW. Oxford University Press; 2011.

26. Barraclough DJ, Conroy ML, Lee D: **Prefrontal cortex and decision making in a mixed-strategy game**. *Nat Neurosci* 2004, **7**:404-410.

27. Sugrue LP, Corrado GS, Newsome WT: **Matching behavior and the representation of value in the parietal cortex**. *Science* 2004, **304**:1782-1787.

28. Lau B, Glimcher PW: **Dynamic response-by-response models of matching behavior in rhesus monkeys**. *J Exp Anal Behav* 2005, **84**:555-579.

29. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ: **Cortical substrates for exploratory decisions in humans**. *Nature* 2006, **441**:876-879.

30. Akaike H: **A new look at the statistical model identification**. *IEEE Trans Autom Control* 1974, **19**:716-723.

31. • Kim H, Sul JH, Huh N, Lee D, Jung MW: **Role of striatum in updating values of chosen actions**. *J Neurosci* 2009, **29**:14701-14712.

Authors recorded neuronal activity from the ventral and dorsal striatum of rats during a choice task. They demonstrated that in both areas the neuronal signal of chosen action value was increased and persisted after animal's choice. The signals of reward prediction error and updated action value were represented after the outcome was revealed.

32. •• Iascher J, Daw N, Dayan P, O'Doherty JP: **States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning**. *Neuron* 2010, **66**:585-595.

A hybrid model combining model-based and model-free learners was used for the first time to explain the choice behavior in a decision task. Authors demonstrated that state prediction error, which is used in the model-based learner, is encoded by fMRI signals in the interparietal sulcus and lateral prefrontal cortex, while reward prediction error, which is used for model-free strategy, is encoded in the ventral striatum.

33. Schwarz G: **Estimating the dimension of a model**. *Ann Stat* 1978, **6**:461-464.

34. Hampton AN, Bossaerts P, O'Doherty JP: **The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans**. *J Neurosci* 2006, **26**:8360-8367.

35. Seo H, Lee D: **Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game**. *J Neurosci* 2007, **27**:8366-8377.

36. Seo H, Barraclough DJ, Lee D: **Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game**. *J Neurosci* 2009, **29**:7278-7289.

37. Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW: **Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task**. *J Neurosci* 2009, **29**:15104-15114.

38. Bishop CM: In *Pattern recognition and machine learning.* Edited by Jordan M, Kleinberg J, Scholkopf B. New York: Springer; 2006.

39. Hollerman JR, Tremblay L, Schultz W: **Influence of reward expectation on behavior-related neuronal activity in primate striatum**. *J Neurophysiol* 1998, **80**:947-963.

40. Lau B, Glimcher PW: **Action and outcome encoding in the primate caudate nucleus**. *J Neurosci* 2007, **27**:14502-14514.

41. Hori Y, Minamimoto T, Kimura M: **Neuronal encoding of reward value and direction of actions in the primate putamen**. *J Neurophysiol* 2009, **102**:3530-3543.

42. Kimchi EY, Laubach M: **The dorsomedial striatum reflects response bias during learning**. *J Neurosci* 2009, **29**:14891-14902.

43. Cohen MX, Axmacher N, Lenartz D, Elger CE, Sturm V, Schlaepfer TE: **Neuroelectric signatures of reward learning and decision-making in the human nucleus accumbens**. *Neuropsychopharmacology* 2009, **34**:1649-1658.

44. Kim YB, Huh N, Lee H, Baeg EH, Lee D, Jung MW: **Encoding of action history in the rat ventral striatum**. *J Neurophysiol* 2007, **98**:3548-3556.

45. Yamada H, Matsumoto N, Kimura M: **History- and current instruction-based coding of forthcoming behavioral outcomes in the striatum**. *J Neurophysiol* 2007, **98**:3557-3567.

46. Lau B, Glimcher PW: **Value representations in the primate striatum during matching behavior**. *Neuron* 2008, **58**:451-463.

47. O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ: **Dissociable roles of ventral and dorsal striatum in instrumental conditioning**. *Science* 2004, **304**:452-454.

48. Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G: **Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards**. *J Neurosci* 2009, **29**:13365-13376.

49. Pasupathy A, Miller EK: **Different time courses of learning-related activity in the prefrontal cortex and striatum**. *Nature* 2005, **433**:873-876.

50. Hoshi E, Tanji J: **Differential roles of neuronal activity in the supplementary and presupplementary motor areas: from information retrieval to motor planning and execution**. *J Neurophysiol* 2004, **92**:3482-3499.

51. Roitman JD, Shadlen MN: **Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task**. *J Neurosci* 2002, **22**:9475-9489.

52. Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S: **Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops**. *Nat Neurosci* 2004, **7**:887-893.

53. Schonberg T, O'Doherty JP, Joel D, Inzelberg R, Segev Y, Daw ND: **Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study**. *Neuroimage* 2010, **49**:772-781.

54. Oyama K, Hernadi I, Iijima T, Tsutsui K: **Reward prediction error coding in dorsal striatal neurons**. *J Neurosci* 2010, **30**:11447-11457.

55. Dickerson KC, Li J, Delgado MR: **Parallel contributions of distinct human memory systems during probabilistic learning**. *Neuroimage* 2010.

56. • Nomoto K, Schultz W, Watanabe T, Sakagami M: **Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli**. *J Neurosci* 2010, **30**:10692-10702.

In this study, midbrain dopamine neurons were recorded during a random-dot motion detection task. For weakly coherent motions, dopamine neurons showed two-phase responses to the stimuli, the first for any stimulus and the second for a rewarding stimulus, which were consistent with the time course required for estimation of expected reward value that parallels the motion discrimination processing.

57. Daw ND, Kakade S, Dayan P: **Opponent interactions between serotonin and dopamine**. *Neural Netw* 2002, **15**:603-616.

58. Miyazaki K, Miyazaki KW, Doya K: **Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards**. *J Neurosci* 2011, **31**:469-479.

59. Matsumoto M, Hikosaka O: **Lateral habenula as a source of negative reward signals in dopamine neurons**. *Nature* 2007, **447**:1111-1115.

60. Yin HH: **The sensorimotor striatum is necessary for serial order learning**. *J Neurosci* 2010, **30**:14719-14723.

61. 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.

62. Yin HH, Knowlton BJ, Balleine BW: **Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning**. *Eur J Neurosci* 2005, **22**:505-512.

63. Yin HH, Knowlton BJ, Balleine BW: **Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning**. *Behav Brain Res* 2006, **166**:189-196.

64. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW: **The role of the dorsomedial striatum in instrumental conditioning**. *Eur J Neurosci* 2005, **22**:513-523.

65. Shiflett MW, Martini RP, Mauna JC, Foster RL, Peet E, Thiels E: **Cue-elicited reward-seeking requires extracellular signal-regulated kinase activation in the nucleus accumbens**. *J Neurosci* 2008, **28**:1434-1443.

66. Talmi D, Seymour B, Dayan P, Dolan RJ: **Human pavlovian-instrumental transfer**. *J Neurosci* 2008, **28**:360-368.

67. Cardinal RN, Parkinson JA, Hall J, Everitt BJ: **Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex**. *Neurosci Biobehav Rev* 2002, **26**:321-352.

68. 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.

69. 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.

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

The functional dissociation between the dorsomedial and the dorsolateral striatum during learning is an important issue. Authors demonstrate that the dorsomedial striatum developed ensemble spike activity mainly during the action phase during learning of a T-maze task, while the dorsolateral striatum develops the activity that is heightened at action boundaries of the task.

71. Cardinal RN, Cheung TH: **Nucleus accumbens core lesions retard instrumental learning and performance with delayed reinforcement in the rat**. *BMC Neurosci* 2005, **6**:9.

72. Packard MG, Knowlton BJ: **Learning and memory functions of the Basal Ganglia**. *Annu Rev Neurosci* 2002, **25**:563-593.

73. • van der Meer MA, Johnson A, Schmitzer-Torbert NC, Redish AD: **Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task**. *Neuron* 2010, **67**:25-32.

Identification of multiple systems in the decision making is a crucial issue. Authors demonstrated that future paths of rats from the decision point were represented in the ventral striatum and the hippocampus but not in the dorsal striatum. On the contrary, the future reward from the decision point was represented only in the dorsal striatum.