



ELSEVIER

# The computational neurobiology of learning and reward

Nathaniel D Daw<sup>1</sup> and Kenji Doya<sup>2,3</sup>

Following the suggestion that midbrain dopaminergic neurons encode a signal, known as a 'reward prediction error', used by artificial intelligence algorithms for learning to choose advantageous actions, the study of the neural substrates for reward-based learning has been strongly influenced by computational theories. In recent work, such theories have been increasingly integrated into experimental design and analysis. Such hybrid approaches have offered detailed new insights into the function of a number of brain areas, especially the cortex and basal ganglia. In part this is because these approaches enable the study of neural correlates of subjective factors (such as a participant's beliefs about the reward to be received for performing some action) that the computational theories purport to quantify.

## Addresses

<sup>1</sup> Gatsby Computational Neuroscience Unit, UCL, Alexandra House, 17 Queen Square, London, WC1N 3AR, UK

<sup>2</sup> Initial Research Project, Okinawa Institute of Science and Technology, 12-22 Suzaki, Uruma, Okinawa 904-2234, Japan

<sup>3</sup> ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika, Soraku, Kyoto 619-0288, Japan

Corresponding author: Daw, Nathaniel D ([daw@gatsby.ucl.ac.uk](mailto:daw@gatsby.ucl.ac.uk))

**Current Opinion in Neurobiology** 2006, **16**:199–204

This review comes from a themed issue on  
Cognitive neuroscience  
Edited by Paul W Glimcher and Nancy Kanwisher

Available online 24th March 2006

0959-4388/\$ – see front matter  
© 2006 Elsevier Ltd. All rights reserved.

DOI [10.1016/j.conb.2006.03.006](https://doi.org/10.1016/j.conb.2006.03.006)

## Introduction

Reinforcement learning (RL) [1] is the branch of artificial intelligence that concerns how an agent, such as a robot, can learn by trial and error to make decisions in order better to obtain rewards and avoid punishments. Such computational theories are also increasingly becoming central to thinking about the neural substrates for similar learning and decision functions in humans and animals [2–4]. This viewpoint particularly builds on the observation [2] that the phasic responses of midbrain dopamine neurons recorded from primates behaving for rewards resemble a major learning signal used in RL, called a 'temporal-difference error signal'. Recently, neuroscientists have begun to integrate such models directly into the design and analysis of experiments, quantitatively studying the models' fit to behavioral responses and neural

signals from individual subjects and trials. This approach enables the study of the neural substrates of inherently subjective quantities that the models purport to quantify, such as the 'value' or 'utility' of an action, meaning the degree of reward that a subject expects to receive for executing that action. Furthermore, because the models are grounded in algorithms describing how ideal subjects should optimally behave, they help to explain not just the mechanisms underlying observed data, but also why they are the way they are.

Many RL accounts of learned decision-making center around the repeated application of variations of the following three steps (Figure 1): first, predict the reward values of action candidates; second, select the action that maximizes the predicted value; third, learn from experience to update the predictions. Here, we review recent studies of the neural substrates for these functions, focusing particularly on work combining theory and experiment in a field in which (happily, in our view) these approaches are fast becoming inseparable.

## Model-based analysis

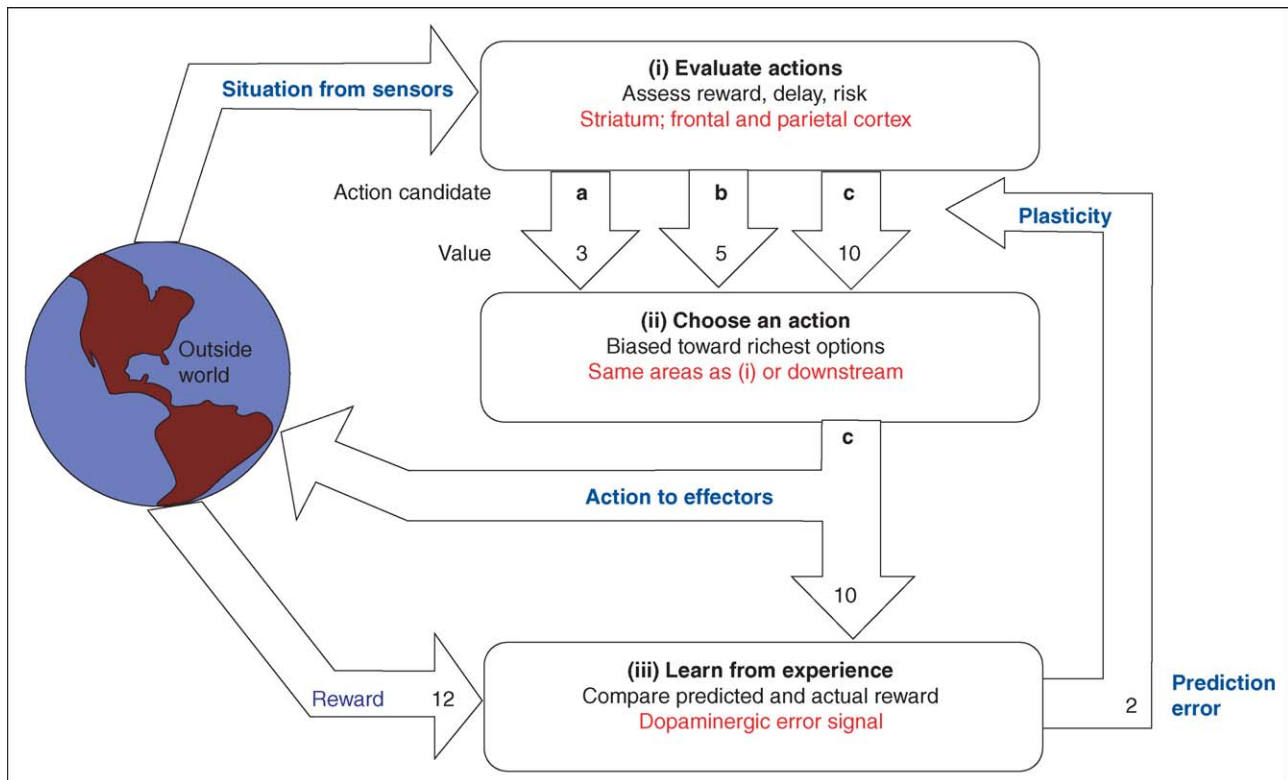
A recent trend, exemplified by many of the studies we review below, is the use of computational models to estimate unobservable time-varying variables, such as how much reward a subject expects on each trial. Behavioral and neural correlates of such variables can then be sought, using quantitative trial-by-trial comparisons rather than the more qualitative analogies common in earlier computational modeling. In such analyses, two crucial issues are how to compare different possible models and how to set various free parameters of the models, such as those controlling how quickly a model learns from experience. Both of these problems can be addressed using standard Bayesian statistical methods to assess which candidate models and parameters best predict the data that are actually observed.

Such model-based analytical methods, when based on RL models, can probe the neural substrates for learned decision-making. We now turn to the results of such studies.

## How to evaluate actions

The first step in our simple decision-making strategy is evaluation. Candidate actions can be evaluated by predicting the long-term future utility, or 'value', expected after taking a particular action in a particular situation. There has been much interest in determining where in the brain such values are represented and how they control behavior. A possible mechanism for this is an array or 'map' of neurons (or groups of neurons) that each

Figure 1



The three basic stages of many reinforcement learning accounts of learned decision-making. **(i)** Predict the rewards expected for candidate actions (here a, b, c) in the current situation. **(ii)** Choose and execute one by comparing the predicted rewards. **(iii)** Finally, learn from the reward prediction error to improve future decisions. Numbers indicate the predicted action values, the obtained reward, and the resulting prediction error.

represent the value of a particular action candidate. Such neurons could compete, on the basis of their value predictions, for their preferred action to be selected. The striatum and various cortical areas have been suggested as possible substrates for such a map.

### Striatum

Given the hypothesized role of dopamine as a signal controlling reward learning [2], its most prominent target, the striatum, is an obvious candidate site for that learning. Supporting this identification, the striatum is associated with motor pathologies, with well-learned, so-called 'habitual' actions [5], and with dopamine-dependent synaptic plasticity [6]. Also, neuronal responses in striatum are modulated by both actions and their anticipated outcomes [7,8]. In a recent study, striatal neurons were recorded while monkeys chose whether to turn a handle leftward or rightward to receive (usually) different probabilities of water reward. (This is called a 'free choice' task, to distinguish it from ones in which animals are instructed which action to take.) The recordings were studied quantitatively to test whether responses encode action values prior to a choice being entered [9••]. During block-by-block changes in the probability that the turns would be rewarded, responses in the majority of striatal

neurons with reward-related movement activities correlated with the block-wise value of either one of the two options. Many fewer neurons were modulated by the relative value of one action over another (which, in the RL model used for analysis, is more directly linked to the probability that the action will be chosen).

Also, in some human functional imaging experiments, the blood-oxygenation level dependent (BOLD) signal in the striatum correlates with predicted reward [10,11]; in other studies, however, it instead correlates with prediction errors for reward [12–14] (and punishment [15]). This difference might be explained if value correlations reflect cortical input or intrinsic activity, whereas the error signal reflects dopaminergic input.

### Cortex

Reward-predictive neural responses have also been observed in a variety of cortical areas, including prefrontal cortex [16–19] and its orbital division [20]. One theoretical proposal [21] (see also [3]) to explain this proliferation of value information is that prefrontal and striatal systems subserve distinct RL methods for action evaluation. In particular, prefrontal cortex might be distinguished by the use of more cognitive methods to plan

actions by sequentially contemplating their consequences [22], whereas dorsolateral striatum is associated with more stimulus-triggered responses [5]. A Bayesian model based on this division of labor accounts for a number of behavioral and lesion results concerning the differential recruitment of these systems [21].

For decisions about where to saccade, another area that has received extensive attention as a potential action value map is the lateral intraparietal area (LIP), where microstimulation evokes saccades and neurons have visuo-spatial receptive fields related to saccade targets. In primates, LIP firing is modulated by the magnitude and probability of the reward expected for an instructed saccade into a neuron's receptive field, and (in an early example of theory-driven data analysis) by a model-generated estimate of subjective target value when monkeys could freely choose, for reward, which of two targets to fixate [23]. Recent work has developed this approach using more elaborate theoretical models. In another free-choice saccade study [24<sup>••</sup>], monkeys' behavioral decisions could satisfactorily be explained by choice according to a relative value measure (called 'local fractional income'); furthermore, LIP neuron responses were modulated by this index for the target in their receptive fields. Another saccade choice task [25] was used to study whether LIP firing is more closely related to the value expected for a saccade or its propensity to be chosen; these were distinguished using trial blocks in which value and choice probability were dissociated using principles from game theory. Block-wise, LIP activity followed value expectancy.

### Variability and delay

In general, candidate actions might differ according to the expected magnitude, probability or delay to rewarding outcomes, and all of these factors might influence the valuation of actions. There is some neural evidence that these factors are accounted for by dissociable neural systems. For instance, variability or 'risk' in an outcome (\$20 with 50 percent probability) can make it either more or less subjectively desirable compared with a different outcome with the same average value (\$10 with certainty). In a functional magnetic resonance imaging (fMRI) study of human financial decision-making, risk-seeking or risk-averse choices were preceded by activation of ventral striatum or anterior insular cortex, respectively [26]. Also, in a primate free-choice study, both behavioral choices and neurons in posterior cingulate were modulated by the variability of the outcome [27].

Similarly, the value of a reward might be modulated by its delay — money received immediately might be more desirable than money promised in a year, a phenomenon known as 'temporal discounting'. The steepness of such discounting — how quickly anticipated reward loses value with delay — is a preference that, in computational theories,

must be tuned to respect particular circumstances (interest rates, hunger) and to improve the efficiency of learning [28]. A possible neural substrate for this function was suggested by an fMRI study of a decision task with delayed rewards [10]. Reward valuations from RL models with different time discounting preferences correlated with BOLD signals arranged in an orderly map along insular cortex and striatum, with ventroanterior value signals discounting delay more steeply compared with those of dorsoposterior signals. In another fMRI study of discounting [11], a dissociation was found between areas (such as ventral striatum and medial prefrontal cortex) differentially activated during choices involving immediate reward and others (such as lateral prefrontal areas) activated during choices regardless of delay. This result was interpreted in terms of choice models from economics that assume a person's overall valuation of a delayed reward comprises contributions from distinct delay-sensitive and -insensitive components.

### How to select actions

The purpose of evaluating the long-term value of action candidates is to simplify the next step in decision-making: choosing between them. The neural substrates for action choice might overlap with those for evaluation — for instance, if neurons representing action values compete for selection through mechanisms such as mutual inhibition — or choice could occur downstream from the value centers.

### Choice rules

Given value estimates for all action candidates, choice might be as simple as 'greedily' selecting the action expected to deliver the greatest future value. However, this strategy might miss out on more valuable actions that have not yet been explored or have previously been unlucky. For this reason, RL models generally assume that there should be some degree of randomness in the choices. In RL, this is often accomplished by a 'softmax' decision rule that chooses randomly but is biased toward the seemingly richest options; in behavioral psychology, the venerable 'matching law' [29] achieves a similar effect using a somewhat different mathematical form.

Many data sets discussed in this review are fit well by the softmax rule (e.g. [9<sup>••</sup>,17,27]). Although the matching rule has also been used to model behavioral and neural data [24<sup>••</sup>], reanalysis [30] (see also [31<sup>••</sup>]) indicated that the softmax rule provided a better fit to behavior. This is partly because (in the case of two actions) it correctly predicts that choice probability depends on the difference, rather than the ratio, of action values.

As already mentioned, an important computational problem facing a decision-maker is tuning the parameters of his or her choice algorithm [28]. In a recent study dramatically demonstrating such 'metalearning', monkeys

playing a 'matching pennies' choice game adjusted the degree to which their choices were random according to the payoff algorithm used by the computer opponent [17]. As the payoff rules were changed to punish predictable responses, monkeys' choices changed from relatively deterministic to more random.

### Actor and critic

Where in the brain are actions actually selected? In recent simultaneous recordings in primate prefrontal cortex and dorsal striatum during a learning task in which the learned associations between stimuli and actions were periodically reversed [32], striatal and prefrontal neurons were both closely tied to animals' behavioral responses. However, the time course of change in behavioral responses over trials following a reversal was more closely related to the change in prefrontal responses than to that of striatal responses. The authors interpreted this result as suggesting that the prefrontal region was more likely to be controlling behavior. Alternatively, given the finding of action value coding in the striatum [9<sup>••</sup>], it is possible that action selection might occur yet elsewhere in the cortico-basal ganglionic loop, downstream from striatum and controlled by its value output [4].

Another possible organization is suggested by RL methods that subdivide choice into parallel prediction and decision subtasks (rather than the serial approach discussed so far; these are called 'actor-critic' methods). Recent fMRI results provide some support for the long-standing suggestion [33] that choice and prediction might be localized in adjacent dorsal and ventral striatum, respectively. Specifically, experiments showed that whereas ventral striatum is implicated for prediction learning regardless of whether rewards are action-contingent, dorsal striatum is recruited only when actions are chosen to obtain reward [14,34].

### How to learn from experience

Finally, experience with the consequences of a chosen action can be used for learning: to update beliefs about the value of the action and thereby improve future decisions. The temporal-difference (TD) learning algorithm updates action values to more accurate ones according to the 'prediction error' or mismatch between received and predicted reward. The observation [2] that the phasic firing of dopamine neurons qualitatively resembles such a prediction error signal is now classic; dynamic behavioral and neural responses during learning are also well fit by this update rule in many of the studies we have reviewed (e.g. [9<sup>••</sup>,10,13–15,17]).

### More news on dopamine

The prediction error hypothesis of dopamine has been tested more quantitatively in several recent experiments. In tasks in which a cue is reinforced probabilistically, excitatory dopamine responses to cues and rewards vary

linearly with the reward probability [35–37], as predicted by the model. The dopamine response to a cued reward also correlates with the number of preceding non-rewarded presentations [38], as would be expected with learning. Perhaps most impressively, a trial-by-trial regression analysis of dopamine responses in a task with varying reward magnitudes showed that the response dependence on the magnitude history has the same form as that expected from TD learning [39].

Another line of evidence is emerging from fast-scan cyclic voltammetry in rodents. This method can detect transient changes in dopamine concentration in the striatum with subsecond temporal resolution. Such recordings exhibit transient dopamine surges in circumstances similar to those that phasically excite dopamine neurons; particularly noteworthy are several reports of timelocking between dopamine surges and cue-evoked or self-initiated behavioral responses [40,41] (see also [36]). Because it measures dopamine release rather than spiking, this technique might also prove particularly useful in studying self-administration of cocaine and brain stimulation reward [41,42<sup>••</sup>], which are both pathological choice behaviors hypothesized to follow from direct interference with the dopaminergic prediction error signal [43].

There is an alternative school of thought that dopamine activity instead flags attentionally relevant or salient events, regardless of their value [44]. In support of the attentional hypothesis, dopamine neurons exhibit some anomalous excitatory responses such as those to novel, neutral stimuli; microdialysis also provides evidence for enhanced dopamine activity on a slow timescale in salient aversive situations, such as footshock. However, an attentional account offers no clear explanation as to why dopamine is so strongly implicated in reinforcement, nor why dopamine neurons are inhibited rather than excited by some salient non-rewarding events, such as the omission of reward [2]. Also problematic from an attentional view, a recent experiment in anesthetized rats verified that dopaminergic neurons are also inhibited rather than excited by salient aversive stimuli [45<sup>••</sup>]. There is also a theoretical account of why dopamine neurons should respond to novel, neutral stimuli (if they encode a reward prediction error) [46].

Another intriguing finding is that dopamine neurons show slowly building excitation when a reward might or might not arrive stochastically [35]. (In this case, spiking ramps up prior to the time reward is expected, earlier to and more slowly than the burst responses we have discussed so far.) This activity was originally interpreted as a distinct signal of uncertainty (i.e., variability) co-located with the prediction error signal, but an alternative explanation is that the ramp is a side effect of the standard TD error signal, related to the anticipatory nature of the learned predictions [47]. The original experimenters disagree

with this account [48], but resolving this question will require further experiments or a more conclusive reanalysis of the existing data.

### What about serotonin?

Unlike the situation with positive prediction error resulting from rewards or reward prediction, neuronal recordings [35–39] have revealed less evidence for a quantitative relationship between the degree of dopaminergic inhibition and the level of negative error from punishment or omitted reward. This might be due to the low background firing rates of the neurons. A natural question is how aversive signals are carried. A suitable candidate is the dorsal raphe serotonin system, many of the functions of which seem to be in opposition to those of the dopamine system [49]. Because this system projects to the striatum, it could be the source of BOLD responses correlated with aversive prediction error in the ventral striatum [15].

However, this proposal clearly does not account for all of the many functions of serotonin. Serotonergic antagonism is associated with diverse deficits including, importantly, impulsive choices, which led to the proposal that the serotonin system regulates the steepness of temporal discounting [28]. For example, in a rat experiment comparing the effects of dopamine and serotonin antagonists on decisions about cost and delay, serotonergic antagonism encouraged impulsive choice of small immediate rewards [50]. Differential activation of the dorsal raphe and the dorsal striatum in fMRI during task conditions that require shallow temporal discounting [10] is also consistent with the hypothesis. However, elucidation of the role of the serotonergic system most urgently requires more evidence from recordings of serotonergic neurons, which have so far been scarce because of technical difficulties.

### Conclusions

The study of reward-guided decision-making is an exemplary area for the integration of theory and experiment. The recent development of this field is characterized by the use of models that are noteworthy for three reasons. First, they are normative — that is, more than phenomenological simulations, they are grounded in and shed light on sound computational principles. Second, they enable the study of the objective correlates of subjective phenomena such as value. Finally, they are closely integrated into the experimental and analytical process, often detailed enough to fit trial by trial to raw experimental data.

An important future direction in the study of learning and decision is how to deal with choice under various sorts of uncertainty. Normative, statistical models of this issue have been proposed to explain such diverse phenomena as the differential roles of acetylcholine and norepinephrine [51] and of prefrontal and striatal systems [21], and

behavioral and neural data from discriminations between noisy sensory stimuli [52]. However, as yet, there has been relatively little experimental work aimed at testing these theories.

### Acknowledgements

N Daw was supported during preparation of this manuscript by the Royal Society, the Gatsby Foundation, and the EU Bayesian Inspired Brain and Artefacts (BIBA) project. K Doya was supported by the National Institute of Information and Communication Technology of Japan, and Grant-in-Aid for Scientific Research on Priority Areas, Ministry of Education, Culture, Sports, Science, and Technology of Japan.

### 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. Sutton RS, Barto AG: *Reinforcement Learning: an Introduction*. MIT Press; 1998.
2. Schultz W, Dayan P, Montague PR: **A neural substrate of prediction and reward**. *Science* 1997, **275**:1593-1599.
3. Doya K: **What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex**. *Neural Netw* 1999, **12**:961-974.
4. Doya K: **Complementary roles of basal ganglia and cerebellum in learning and motor control**. *Curr Opin Neurobiol* 2000, **10**:732-739.
5. 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.
6. Reynolds JNJ, Hyland BI, Wickens JR: **A cellular mechanism of reward-related learning**. *Nature* 2001, **413**:67-70.
7. Shidara M, Aigner TG, Richmond BJ: **Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials**. *J Neurosci* 1998, **18**:2613-2625.
8. Watanabe K, Lauwereyns J, Hikosaka O: **Neural correlates of rewarded and unrewarded eye movements in the primate caudate nucleus**. *J Neurosci* 2003, **23**:10052-10057.
9. Samejima K, Ueda K, Doya K, Kimura M: **Representation of action-specific reward values in the striatum**. *Science* 2005, **310**:1337-1340.
- This study is the first controlled experimental examination of the theoretically motivated suggestion that striatal neurons encode action values. The experimenters used a choice task designed to tease apart a number of different potential RL signals. The methodology is also noteworthy for introducing sophisticated model fit procedures to the study of the dynamics of learning parameters.
10. 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.
11. McClure SM, Laibson DI, Loewenstein G, Cohen JD: **Separate neural systems value immediate and delayed monetary rewards**. *Science* 2004, **306**:503-507.
12. McClure SM, Berns GS, Montague PR: **Temporal prediction errors in a passive learning task activate human striatum**. *Neuron* 2003, **38**:339-346.
13. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ: **Temporal difference models and reward-related learning in the human brain**. *Neuron* 2003, **38**:329-337.
14. 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.

15. Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS: **Temporal difference models describe higher-order learning in humans.** *Nature* 2004, **429**:664-667.
16. Matsumoto K, Suzuki W, Tanaka K: **Neuronal correlates of goal-based motor selection in the prefrontal cortex.** *Science* 2003, **301**:229-232.
17. Barraclough DJ, Conroy ML, Lee D: **Prefrontal cortex and decision making in a mixed-strategy game.** *Nat Neurosci* 2004, **7**:404-410.
18. Roesch MR, Olson CR: **Neuronal activity related to reward value and motivation in primate frontal cortex.** *Science* 2004, **304**:307-310.
19. Amemori KI, Sawaguchi T: **Contrasting effects of reward expectation on sensory and motor memories in primate prefrontal neurons.** *Cereb Cortex* 2005, Oct 12 [Epub ahead of print].
20. Tremblay L, Schultz W: **Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex.** *J Neurophysiol* 2000, **83**:1864-1876.
21. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control.** *Nat Neurosci* 2005, **8**:1704-1711.
22. Owen AM: **Cognitive planning in humans: neuropsychological, neuroanatomical, and neuropharmacological perspectives.** *Prog Neurobiol* 1997, **53**:431-450.
23. Platt ML, Glimcher PW: **Neural correlates of decision variables in parietal cortex.** *Nature* 1999, **400**:233-238.
24. Sugrue LP, Corrado GS, Newsome WT: **Matching behavior and the representation of value in the parietal cortex.** *Science* 2004, **304**:1782-1787.
- The authors build on classic 'matching' models from behavioral psychology in a careful analysis of monkey choices and associated LIP activity. Intriguingly, for the restricted class of models considered, monkeys exhibit learning parameters near optimal for harvesting rewards. This result invites future study of the dynamic tuning of such parameters.
25. Dorris MC, Glimcher PW: **Activity in posterior parietal cortex is correlated with the subjective desirability of an action.** *Neuron* 2004, **44**:365-378.
26. Kuhnen CM, Knutson B: **The neural basis of financial risk taking.** *Neuron* 2005, **47**:763-770.
27. McCoy AN, Platt ML: **Risk-sensitive neurons in macaque posterior cingulate cortex.** *Nat Neurosci* 2005, **8**:1220-1227.
28. Doya K: **Metalearning and neuromodulation.** *Neural Netw* 2002, **15**:495-506.
29. Herrnstein RJ: **Relative and absolute strength of response as a function of frequency of reinforcement.** *J Exp Anal Behav* 1961, **4**:267-272.
30. Corrado GS, Sugrue LP, Seung SH, Newsome WT: **Linear-nonlinear-Poisson models of primate choice dynamics.** *J Exp Anal Behav* 2005, **84**:581-617.
31. Lau B, Glimcher PW: **Dynamic response-by-response models of matching behavior in rhesus monkeys.** *J Exp Anal Behav* 2005, **84**:555-579.
- To date, this is one of the most quantitatively sophisticated analyses of monkey choice behavior. The authors argue that existing models, including most of those reviewed here, fail to capture autocorrelation structure in choices (such as tendencies to alternate or persevere, regardless of rewards), and experimental analyses might therefore be biased.
32. Pasupathy A, Miller EK: **Different time courses of learning-related activity in the prefrontal cortex and striatum.** *Nature* 2005, **433**:873-876.
33. Dayan P, Balleine BW: **Reward, motivation, and reinforcement learning.** *Neuron* 2002, **36**:285-298.
34. Tricomi EM, Delgado MR, Fiez JA: **Modulation of caudate activity by action contingency.** *Neuron* 2004, **41**:281-292.
35. Fiorillo CD, Tobler PN, Schultz W: **Discrete coding of reward probability and uncertainty by dopamine neurons.** *Science* 2003, **299**:1898-1902.
36. Satoh T, Nakai S, Sato T, Kimura M: **Correlated coding of motivation and outcome of decision by dopamine neurons.** *J Neurosci* 2003, **23**:9913-9923.
37. Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H: **Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons.** *Neuron* 2004, **43**:133-143.
38. Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O: **Dopamine neurons can represent context-dependent prediction error.** *Neuron* 2004, **41**:269-280.
39. Bayer HM, Glimcher PW: **Midbrain dopamine neurons encode a quantitative reward prediction error signal.** *Neuron* 2005, **47**:129-141.
40. Roitman MF, Stuber GD, Phillips PE, Wightman RM, Carelli RM: **Dopamine operates as a subsecond modulator of food seeking.** *J Neurosci* 2004, **24**:1265-1271.
41. Phillips PE, Stuber GD, Heien ML, Wightman RM, Carelli RM: **Subsecond dopamine release promotes cocaine seeking.** *Nature* 2003, **422**:614-618.
42. Montague PR, McClure SM, Baldwin PR, Phillips PE, Budygin EA, Stuber GD, Kilpatrick MR, Wightman RM: **Dynamic gain control of dopamine delivery in freely moving animals.** *J Neurosci* 2004, **24**:1754-1759.
- The authors combine dopaminergic neuron stimulation and fast-scan cyclic voltammetry to study transmitter release at striatal targets. The experiments reveal complex depression and facilitation dynamics of a sort not anticipated in TD models of dopamine function, which are based mainly on recordings of dopaminergic spiking activity.
43. Redish AD: **Addiction as a computational process gone awry.** *Science* 2004, **306**:1944-1947.
44. Horvitz JC: **Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum.** *Behav Brain Res* 2002, **137**:65-74.
45. Ungless MA, Magill PJ, Bolam JP: **Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli.** *Science* 2004, **303**:2040-2042.
- By combining juxtacellular labeling and unit recordings, the authors help to settle a longstanding and theoretically important debate about whether dopamine neurons are excited by aversive stimulation. The results suggest that neurons previously reported to be excited by noxious stimuli could have been misidentified as dopaminergic, and the authors suggest caution and more stringent neuronal identification criteria for future studies of dopaminergic units.
46. Kakade S, Dayan P: **Dopamine: generalization and bonuses.** *Neural Netw* 2002, **15**:549-559.
47. Niv Y, Duff MO, Dayan P: **Dopamine, uncertainty and TD learning.** *Behav Brain Funct* 2005, **1**:6.
48. Fiorillo CD, Tobler PN, Schultz W: **Evidence that the delay-period activity of dopamine neurons corresponds to reward uncertainty rather than backpropagating TD errors.** *Behav Brain Funct* 2005, **1**:7.
49. Daw ND, Kakade S, Dayan P: **Opponent interactions between serotonin and dopamine.** *Neural Netw* 2002, **15**:603-616.
50. Denk F, Walton ME, Jennings KA, Sharp T, Rushworth MF, Bannerman DM: **Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort.** *Psychopharmacology (Berl)* 2005, **179**:587-596.
51. Yu AJ, Dayan P: **Uncertainty, neuromodulation, and attention.** *Neuron* 2005, **46**:681-692.
52. Gold JI, Shadlen MN: **Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward.** *Neuron* 2002, **36**:299-308.