REVIEW

DECISION MAKING

nature neuroscience

Modulators of decision making

Kenji Doya[1,2]

Human and animal decisions are modulated by a variety of environmental and intrinsic contexts. Here I consider computational factors that can affect decision making and review anatomical structures and neurochemical systems that are related to contextual modulation of decision making. Expectation of a high reward can motivate a subject to go for an action despite a large cost, a decision that is influenced by dopamine in the anterior cingulate cortex. Uncertainty of action outcomes can promote risk taking and exploratory choices, in which norepinephrine and the orbitofrontal cortex appear to be involved. Predictable environments should facilitate consideration of longer-delayed rewards, which depends on serotonin in the dorsal striatum and dorsal prefrontal cortex. This article aims to sort out factors that affect the process of decision making from the viewpoint of reinforcement learning theory and to bridge between such computational needs and their neurophysiological substrates.

Our daily life is a chain of decisions. For example, when you go out to lunch, you choose which restaurant to go to, what dish to eat, whom to go with and even whether or not to take a lunch at all. Such decisions vary day to day depending on external and internal factors, such as where you ate yesterday, how hungry you are and whom you hope to see.

The process of decision making can be decomposed into four steps. First, one recognizes the present situation (or state). Second, one evaluates action candidates (or options) in terms of how much reward or punishment each potential choice would bring. Third, one selects an action in reference to one's needs. Fourth, one may reevaluate the action based on the outcome. Although these steps may not always be followed explicitly, the normative theoretical models of how these steps should be carried out are useful in understanding of how these steps are realized in the brain.

In situations such as detecting the motion direction of a noisy visual stimulus, the first step would be the main determinant of the choice. Such a process of perceptual decision making has been studied within the theoretical framework of bayesian inference[1]. In the following sections, I first review theoretical models of the other three steps—evaluation, action selection and learning—based on the framework of reinforcement learning[2,3]. I next discuss how the parameters for those steps should be modulated by environmental factors and the decision maker's needs and experiences. Finally, I review recent literature related to decision making and try to link factors affecting decision making with their neurobiological substrates.

Computational model of decision making

Evaluation of action candidates. What makes everyday decisions so difficult is that decisions can result in rewards or punishments of different amounts at different timings with different probabilities. In a general form, the value of a reward given by an action at a state is a function of reward amount, delay and probability. Although there is no general guarantee that the three factors are independent, it is often assumed they work multiplicatively[4]:

$$V = f(\text{amount}) \times g(\text{delay}) \times h(\text{probability})$$

Figure 1 illustrates different models of the effects of these components. The function f is called the 'utility function'. For example, the value of a perishable food should saturate depending on how much the animal can eat. Thus the function is often regarded as a saturating nonlinear function.

The second function, g, determines 'temporal discounting' of delayed rewards, which is a decreasing function of the delay. In many animal and human choice experiments, subjects tend to be more sensitive to differences in shorter delays than longer delays, which is well modeled by hyperbolic functions[5,6].

The last function, h, represents the possible over- or undervaluation of stochastic outcomes. When there are multiple possible outcomes, the value of the option is their sum,

$$V = \sum_i f(\text{amount}_i) \times g(\text{delay}_i) \times h(\text{probability}_i)$$

In standard 'expected utility' theory[7], h is assumed to be identity, resulting in a simpler form:

$$V = E[f(\text{amount}) \times g(\text{delay})]$$

where E denotes expectation. However, subjects often undervalue probabilistic outcomes, which is better modeled by a function h that is smaller than unity for probability < 1 (except at very small probabilities, which can be overvalued). Such deviations from expected utility theory are summarized in 'prospect theory'[8].

Action selection. After evaluating the value of each action candidate, the next issue is how to select an appropriate one. Given the values for action candidates $V(a_1),\ldots, V(a_n)$, the most straightforward way is to choose the one with the highest value. This is called greedy action selection.

[1]Neural Computation Unit, Okinawa Institute of Science and Technology, 12-22 Suzaki, Uruma, Okinawa, 904-2234, Japan. [2]ATR (Advanced Telecommunications Research) Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika, Soraku, Kyoto 619-0288, Japan. Correspondence should be addressed to K.D. (doya@oist.jp).

Published online 26 March 2008; doi:10.1038/nn2077

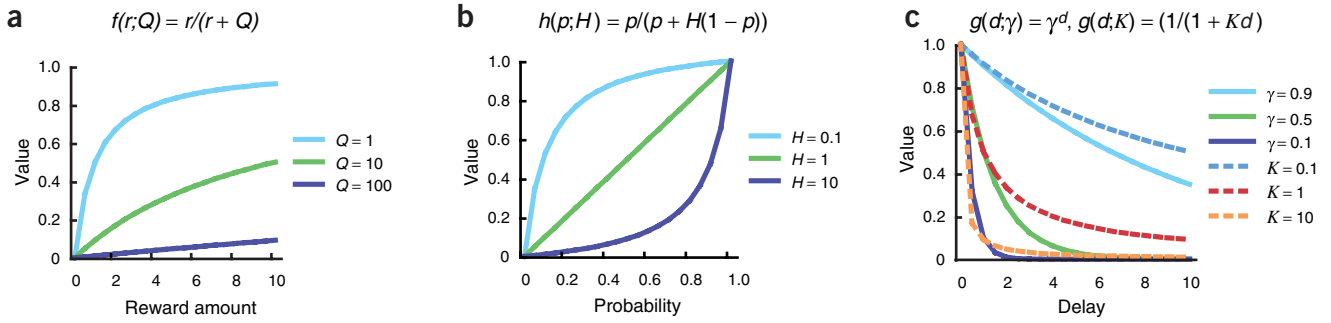

Figure 1 Standard models of evaluation of amount, delay and probability of reward[4]. (**a**) The utility function for the amount of reward is modeled by a saturating function: $f(r) = r/(r + Q)$, where r is the size of the reward and Q determines the amount with which the utility curve saturates. (**b**) A common way to express nonlinear evaluation of probability is take a hyperbolic function $f(\theta) = 1/(1 + H\theta)$, where θ is called the odds against: $\theta = (1 - p)/p$, where p is the probability[4,88]. When expressed as a function of the probability, the function is $h(p) = p/[p + H(1 - p)]$. The parameter $H = 1$ means linear evaluation and $H > 1$ means under-evaluation of stochastic rewards. (**c**) In the standard theory of economics and reinforcement learning[3], the temporal discounting function for delay d is supposed to be exponential: $g(d) = g\gamma^d$, where the parameter γ is called the discount factor; the larger the γ, the longer delayed rewards are taken into account. An alternative model supported by psychology experiments is a hyperbolic function[6]: $g(d) = 1/(1 + K \times d)$, where the parameter K determines the steepness of discounting, large K meaning rapid discounting. Hyperbolic functions decay rapidly near zero and slowly as the delay increases.

In real life, the true value of each action candidate is rarely known exactly. When the values of actions are being learned by experience, exploration of actions is necessary. A simple way is to take a random action at probability ε and take the greedy action with respect to the current estimates of the value function the rest of the time. This is called ε-greedy action selection. Another common solution is 'Boltzmann selection' in which selection probabilities p are proportional to the exponentials of the estimated values:

$$p(\text{action} = a_1) \propto e^{\beta V(a_i)}$$

By an analogy with thermodynamics[9], the scaling parameter β is called the 'inverse temperature'; $\beta = 0$ means all actions are taken with an equal probability of $1/n$, and the larger the β, the greedier the selection.

In animal behavior studies, a well known principle is the matching law[10,11], in which an action is selected in proportion to its value:

$$p(\text{action} = a_1) \propto V(a_i)$$

This is a nearly optimal strategy in 'baited' tasks, in which a reward becomes available at a given location with a certain probability and will stay there until it is taken[12]. In such an environment, a less rewarded action becomes more profitable after a long interval.

Learning. In learning the values of actions in dynamic environments, a critical issue is to identify which action in time caused a given outcome. For example, if you feel sick after lunch, you wonder what the cause was: something you ate at the meal? the cold wind on the way there? or words you heard from your dining companion? This is the problem of 'temporal credit assignment'.

There are three basic ways for learning values in dynamic environments[3]. First, keep in memory which action was taken at which state in the form of 'eligibility traces', and when a reward is given, reinforce the state-action associations in proportion to the eligibility traces. Second, use so-called temporal difference learning. In the case of exponential temporal discounting, this involves following a model using a recursive relationship of the values of subsequent states and actions

$$V(\text{state, action}) = E[\text{reward} + \gamma V(\text{new state, new action})]$$

to update the previous state-action pair. Third, learn a model of action-dependent state-transition probability and, given the present state, predict the future rewards for hypothetical actions in order to select the best evaluated[13].

These three methods have pros and cons. Learning by eligibility trace is simple and robust but not very efficient for delayed rewards. Temporal difference learning is more efficient but depends on the appropriate choice of the state variable. Model-based planning requires more contrived operations but can provide flexible adaptation to changes in behavioral goals.

Factors that affect decisions and learning

Let us now consider how valuation, action selection and learning should be tuned depending on the environment and the needs of the decision maker.

Needs and desires. The utility curve f should reflect the decision maker's physiological or economic needs. Suppose you found that you had lost your wallet: picking up a penny on the road would not help you with bus fare back home. The utility of any amount exceeding the maximal consumption should also saturate. Thus utility functions often have sigmoid shape with threshold and saturation. In people, different desires leads to different thresholds of nonlinear valuation. To enter a good school, attaining a certain score in the exam is a must. Different life goals, such as buying a home or starting a company, put different thresholds and saturation into utility curves.

Flattening of the utility curve, for instance by satiety, is called devaluation and is a useful tool for assessing the mechanism of valuation[14,15].

Risk and uncertainty. Buying insurance is supposed to be a rational behavior, even though it leads to a loss on average. The main reason for buying insurance is to improve the value of the worst-case outcome. A conservative choice for the worst-case scenario can result from min-max evaluation: to minimize the maximal punishment or to maximize the minimal possible reward. Another example of deviation from average evaluation is buying a lottery ticket. If the utility function has a high threshold—for example, enough for a person of humble income to own a house—playing the lottery may be the only choice for going above the threshold at a nonzero probability. Such deviations from simple linear evaluation can be regarded as 'risk-averse' or 'risk-seeking' decisions and be modeled by nonlinearity in either the utility function f or the probability evaluation function h.

Knowledge and uncertainty about the environment are also important in decision making. There are different kinds of

uncertainties: from the stochasticity inherent in the environmental dynamics, from unexpected variation of the environment and from the limited knowledge possessed by the decision maker.

Stochastic environmental dynamics limit the predictability of the future state of the environment, which should affect the optimal setting of temporal discounting. Looking too far ahead can make prediction more difficult, leading to slower learning. In reinforcement learning, the temporal horizon needs to be set long enough, but not too long. In addition, predictability depends not only on the features of the environment but also on the knowledge and actions of the subject. Even a regular environment can appear random if someone fails to capture an essential sensory cue; conversely, experienced individuals may become skilled at reading very subtle cues. For example, a good surfer may be able to plan maneuvers ahead, while a novice on the same wave will end up being tumbled into the water.

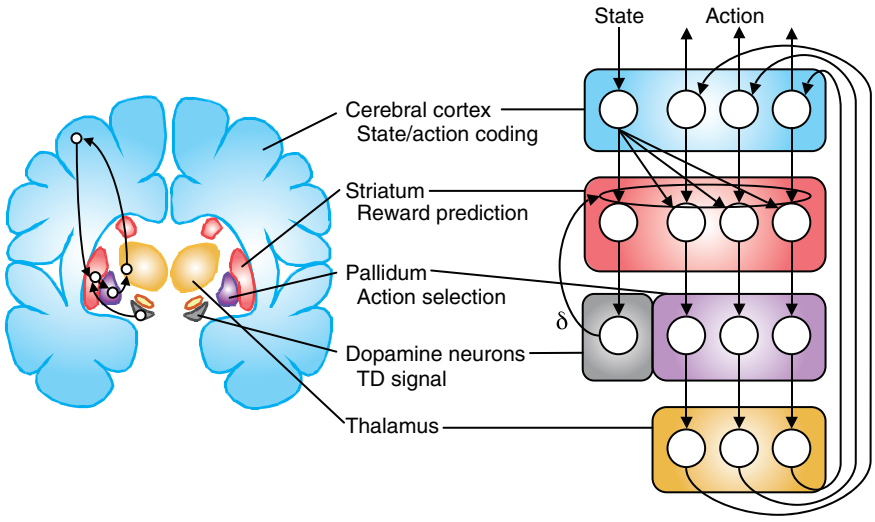

Figure 2 A hypothetical model of realization of reinforcement learning in the cortex–basal ganglia network[2]. Left, coronal section of the brain. Right, functional model, where δ denotes the reward prediction error carried by the midbrain dopamine neurons.

Unexpected changes in the environment and the degree of the subject's knowledge also affect the optimal degree of exploration and memory updating[16]. In a familiar, reliable environment, there is no need for exploration and learning. In contrast, if the subject is aware that his knowledge of the environment could be improved, or that the environment could have changed, he needs to take exploratory actions and increase the learning rate. Uncertainty in prediction can also favor one learning framework over another—for example, model-free temporal difference learning and model-based planning[17].

Time spent and time remaining. A general issue in learning is how fast one should learn from new experiences and how stably old knowledge should be retained. The appropriate choice of the learning rate depends on both the character of the environment and the experience of the subject. In a constant environment, the theoretically optimal way to learn is to start with rapid memory updating and then to decay the learning rate as an inverse of the number of experiences. When the dynamics of the environment change over time, the setting of the learning rate should depend on the estimate of the time for which the past experiences remain valid.

The time left for learning and foraging should also affect temporal discounting and exploration. Most animals have to find food before sunset (or dawn in the case of nocturnal animals) and before they starve to death. Such deadlines for reward acquisition naturally set the upper limit on appropriate temporal discounting.

Another important factor is the exclusiveness of commitment. Your decision of whether to wait for a table at a popular restaurant may depend on whether you have to keep standing in line or you can just sign up and spend the time shopping or drinking. In deciding between actions with less-than-average rewards, an action with smaller reward in shorter time can be more appropriate because it allows moving on to the next action earlier, resulting in an apparently impulsive choice[18,19]. Such 'opportunity cost' can manifest as a linear temporal discounting of reward. Whereas most choice experiments in animals involve real, exclusive waiting, most choice experiments in humans using imaginary questionnaires (such as "Which do you prefer, $10 now or $11 tomorrow?") implicitly assume that the subjects can do whatever they want during the waiting period.

Neural substrates modulating decision making

Now let us consider how the environmental factors and the decision maker's needs and experiences can modulate the neurobiological process of decision making. To begin, it is worthwhile to examine the key players in the neurobiology of decision making. **Figure 2** depicts our current view on how decisions are made in the circuit linking the cerebral cortex and the basal ganglia[2,20]. Reward-predictive neural activities are found in a variety of areas in the cortex[21–24], the striatum[25,26], the globus pallidus[27] and the thalamus[28,29]. Neural recording experiments in animals reveal that midbrain dopamine neurons encode reward prediction errors[30,31], whereas functional brain imaging in humans show activity related to reward prediction error in the striatum[32–35], which receives strong dopaminergic projections. Dopamine-dependent plasticity in the striatum seems to be important in learning of reward-predictive neural activities[36,37]. The dynamic interaction of these areas composing the cortex–basal ganglia loops, as well as other subcortical structures, especially the amygdala, is believed to result in reward-dependent selection of particular actions[20,27]. The network is affected by the sensory and contextual information represented in the cortex, as well as in diffuse neurochemical systems, such as serotonin, norepinephrine and acetylcholine[38]. **Table 1** lists recent studies on the roles of these anatomical structures and neuromodulatory systems in decision making.

Gains and losses. The amygdala is involved in processing of aversive stimuli and avoidance learning. Human brain imaging shows response of the amygdala to expectation of losses as opposed to gains[39]. However, a neural recording study in nonhuman primates showed that neurons in the amygdala respond to reward expectation as well[40]. Human functional brain imaging reveals that different parts in the striatum respond to gains and losses[41].

Midbrain dopamine neurons respond to reward predictive cues or unexpected delivery of rewards. However, because of their low spontaneous firing rate, their inhibitory response to prediction of no reward or omission of reward is weak[31,42]. A hypothetical medium of such aversive prediction is serotonergic neurons[43]. However, little evidence directly supports loss- or aversion-specific response of serotonergic neurons. On the other hand, a recent recording study has clearly demonstrated that neurons in the lateral habenula

Table 1 Roles of neurochemical and anatomical systems in valuation and decision

System	Function	Task	Method	Ref.
	Gain and loss			
Amygdala	Response to expected loss	Gambling	Human, fMRI	39
Amygdala	Expected reward and punishment	Pavlovian	Monkey, recording	40
Lateral habenula	Prediction of no reward omission	Saccade	Monkey, recording	44
	Cost and effort			
Dopamine	Effort for reward	Lever press	Rat, antagonist	62
Dopamine (D_2 receptor)	Effort and delayed rewards	T-maze	Rat, antagonist	45
Dopamine, tonic	Opportunity cost	Lever press	Model	18
ACC	Effort for reward	T-maze	Rat, lesion	54
ACC	Effort for reward	T-maze	Rat, lesion	63
	Risk or variance			
Lateral OFC	High variance	Gambling	Human, fMRI	47
Norepinephrine (β receptor)	Variance of losses	Gambling	Human, antagonist	64
Nucleus accumbens, anterior insula	Risk seeking	Investment	Human, fMRI	46
Nucleus accumbens core	Large, uncertain reward	Lever choice	Rat, lesion	65
	Delay discounting			
Glutamate (NMDA receptor)	Delayed reward	Lever press	Rat, NMDA antagonist	62
Serotonin	Delayed reward	T-maze	Rat, serotonin synthesis blocker	45
Serotonin	Withholding pavlovian response	Approach behavior	Rat, neurotoxin	66
Serotonin in medial PFC	Delay discounting task	Lever choice	Rat, microdialysis	55
Serotonin in dorsal striatum	Activity for long-term value	Liquid reward	Human, ATDL, fMRI	56
Serotonin in ventral striatum	Less activity for short-term value	Liquid reward	Human, ATDL, fMRI	56
Dopamine (D_1 receptor)	Large delayed reward	Nose poke	Rat, antagonist	67
Dopamine in OFC	Large delayed reward	Lever choice	Rat, neurotoxin	68
Dopamine in OFC	Delay discounting task	Lever choice	Rat, microdialysis	55
Norepinephrine	Large delayed reward	Lever choice	Rat, reuptake inhibitor	69
OFC	Delayed reward	T-maze	Rat, lesion	70
OFC	Delayed or uncertain reward	Lever press	Rat, lesion	71
OFC	Large delayed reward	Lever choice	Rat, lesion	72,73
OFC	Small immediate reward	Lever choice	Rat, lesion	74
Nucleus accumbens core	Large delayed reward	Lever choice	Rat, lesion	75
Nucleus accumbens core	Large delayed reward	Lever choice	Rat, lesion	76
Basolateral amygdala	Large delayed reward	Lever choice	Rat, lesion	74
Nucleus accumbens, medial OFC, ACC, PCC	Immediate reward	Liquid reward	Human, fMRI	52
Dorsolateral PFC, PPC, anterior insula	Delayed and immediate reward	Liquid reward	Human, fMRI	52
Ventral striatum	Small immediate reward	Questionnaire	Human, fMRI	77
Ventral striatum, medial PFC, PCC	Subjective value	Questionnaire	Human, fMRI	53
	Learning rate			
ACC	Volatility and learning rate	Gambling	Human, fMRI	57
ACC	Sustaining rewarded actions	Joystick, reversal	Monkey, lesion	78
	Switching and exploration			
Norepinephrine	Reset of network dynamics	Review	Rat, monkey	58
Norepinephrine, tonic	Unexpected uncertainty	Model	Rat, monkey	79
Norepinephrine, phasic	Interrupt for unexpected events	Model	Rat, monkey	59
Norepinephrine, tonic	Exploration-exploitation	Review	Monkey	60
Norepinephrine, phasic	Temporal filtering	Review	Monkey	60
Serotonin	Extradimensional shift	Visual discrimination	Human, ATD	80
Serotonin	Greedy choice	Gambling	Human, ATD	81
Serotonin	Probabilistic choice reversal	Probabilistic learning	Human, SSRI	82
Serotonin in PFC	Choice reversal	Visual discrimination	Marmoset, neurotoxin	61
Serotonin in OFC	Inhibiting perseverative choice	Visual discrimination	Marmoset, neurotoxin	83
Serotonin in medial PFC	Response to reward change	Spatial and odor discrimination	Rat, neurotoxin	84
Frontal polar cortex, IPC	Exploratory decisions	Gambling	Human, fMRI	85
OFC	Greedy choice	Gambling	Human, focal damage	81
Medial PFC	Goal-directed learning	Devaluation	Rat, lesion	86
Dorsomedial striatum	Goal-directed learning	Devaluation	Rat, lesion	87

ACC, anterior cingulate cortex; ATD, acute tryptophan depletion; ATDL, acute tryptophan depletion and loading; IPC, intraparietal cortex; OFC, orbitofrontal cortex; PCC, posterior cingulate cortex; PFC, prefrontal cortex; PPC, posterior parietal cortex.

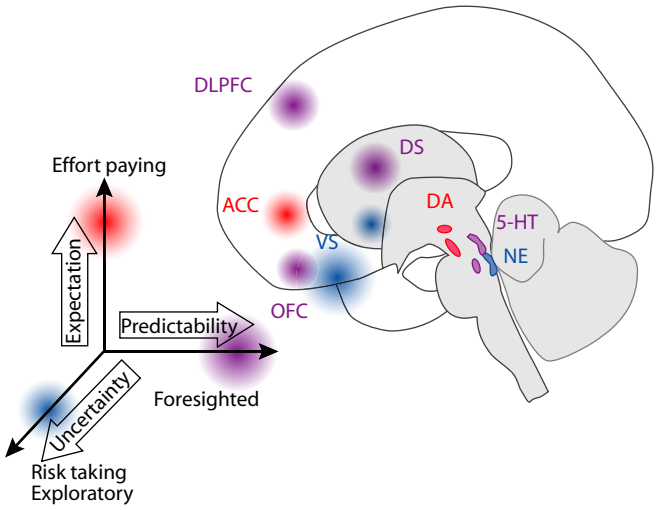

Figure 3 Possible links between computational factors and parameters of decision making and learning, and their neurobiological substrates. 5-HT, serotonin; ACC, anterior cingulate cortex; DA, dopamine; DLPFC, dorsolateral prefrontal cortex; DS, dorsal striatum; NE, norepinephrine; OFC, orbitofrontal cortex; VS, ventral striatum.

respond to no-reward predictive cues as well as reward omission, in exactly the opposite way as dopamine neurons, and that stimulation of the lateral habenula causes inhibition of dopamine neurons[44]. The result highlights the lateral habenula as a possible center of aversive learning, and its role in guiding decision making would be a very interesting question.

Cost and effort. In a T-maze with a small reward behind a low wall on one side and a large reward behind a high wall on the other, lesions of the anterior cingulate cortex (ACC) cause choices of small rewards obtained by smaller effort. Choice of a larger reward with a larger effort is impaired by a dopamine D2 receptor antagonist[45].

It has been proposed that the tonic level of dopamine represents the 'opportunity cost', predicting that animals will work more vigorously when the average expected reward is high, which is signaled by the tonic firing of dopamine neurons[18].

Risk and variance. Brain imaging shows that the striatum, especially the ventral striatum, is involved in expectation of rewards. How is variance or uncertainty of reward represented? Imaging studies show activity in the anterior insula[46] and the lateral orbitofrontal cortex (OFC)[47] in response to variance in the predicted reward. Risk-seeking choice also activates the ventral striatum[46].

Delay discounting. Deficits in the serotonergic system are implicated in impulsivity, both in suppression of maladaptive motor behaviors and in choices of larger but delayed rewards. However, the results of lesion and pharmacology studies are by no means simple, possibly because of autoregulatory feedback mechanisms in the serotonergic system (see ref. 48 for in-depth review).

The rat T-maze experiment mentioned above showed dissociation of the dopaminergic and serotonergic systems in choices of larger reward after more effort and longer delay, respectively[45].

In functional brain imaging studies using a game in a dynamic environment, the dorsolateral prefrontal cortex, dorsal premotor cortex, parietal cortex and insula are more activated in conditions requiring long-term prediction of rewards rather than in conditions requiring short-term predictions[49,50]. Also, functional magnetic resonance imaging (fMRI) experiments using a monetary questionnaire[51] or liquid rewards[52] found activation of the ventral striatum, medial OFC, ACC and posterior cingulate cortex for expectation of immediate rewards. However, another study found these areas to be activated by the subjective value after discounting, irrespective of the delay of the reward[53]. The rat T-maze experiments also found regional dissociation of ACC and OFC in effort and delay discounting[54].

How are these regional specializations in evaluation of immediate and delayed rewards related to the serotonergic system? Serotonin efflux increases in medial prefrontal cortex while rats perform a delay discounting task[55]. Acute tryptophan depletion and loading has shown that activation of the ventral striatum for immediate reward prediction is enhanced during a low-serotonin condition, whereas activation of the dorsal striatum for delayed reward prediction is enhanced during a high-serotonin condition[56].

Learning and exploration. The optimal setting of the learning rate depends on how quickly the world is changing. Subjects' learning rates vary depending on the volatility of the task environment, which is also correlated with the activity of ACC[57].

After an abrupt change of the environment, it is more appropriate to totally reset what has been learned (or switch to another learning module) and start over. Norepinephrine is implicated in such 'resets' of ongoing activities[58,59]. Norepinephrine is also suggested to be important in regulating the decision to explore alternatives versus exploiting a known resource[60]. Deficits in serotonin, especially in the medial prefrontal cortex, disturb adaptation to changes in the required action for a given cue (reversal learning) by making the subjects more likely to stick to prelearned behaviors[61].

Conclusion

I have reviewed the computational mechanisms that should affect decision making and the neurobiological substrates that are related to regulation of decision making. Reports for a variety of systems in varieties of tasks and species are rather equivocal. Functional correspondence of frontal cortical areas between rodents, monkeys and humans poses a further challenge for any unified view. Nevertheless, there are some common observations (**Fig. 3**).

Expectation of a high reward motivates subjects to choose an action despite a large cost, for which dopamine in the anterior cingulate cortex is important. Uncertainty of action outcomes can promote risk taking and exploratory choices, in which norepinephrine and the orbitofrontal cortex seem to be involved. Predictable environments promote consideration of longer-delayed rewards, for which serotonin and the dorsal part of the striatum as well as the dorsal prefrontal cortex are key. Much work will be required to build quantitative models of how decision parameters should be regulated depending on the environment and experience, and then to elucidate how they could be realized by network, cellular and neurochemical mechanisms.

ACKNOWLEDGMENTS

The author thanks P. Dayan, N. Daw and N. Schweighofer for discussions on temporal discounting. This research was supported by a Grant-in-Aid for Scientific Research on Priority Areas, Ministry of Education, Culture, Sports, Science and Technology of Japan.

Published online at http://www.nature.com/natureneuroscience
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Doya, K., Ishii, S., Pouget, A. & Rao, R. *Bayesian Brain: Probabilistic Approach to Neural Coding and Learning* (MIT Press, Cambridge, Massachusetts, USA, 2007).
2. Doya, K. Reinforcement learning: computational theory and biological mechanisms. *HFSP J.* **1**, 30–40 (2007).
3. Sutton, R.S. & Barto, A.G. *Reinforcement Learning* (MIT Press, Cambridge, Massachusetts, USA, 1998).
4. Ho, M.Y., Mobini, S., Chiang, T.J., Bradshaw, C.M. & Szabadi, E. Theory and method in

the quantitative analysis of "impulsive choice" behaviour: implications for psychopharmacology. *Psychopharmacology (Berl.)* **146**, 362–372 (1999).

5. Berns, G.S., Laibson, D. & Loewenstein, G. Intertemporal choice–toward an integrative framework. *Trends Cogn. Sci.* **11**, 482–488 (2007).
6. Laibson, D.I. Golden eggs and hyperbolic discounting. Q. *J. Econ.* **62**, 443–477 (1997).
7. von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, New Jersey, USA, 1944).
8. Kahneman, D. & Tversky, A. Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–291 (1979).
9. Ishii, S., Yoshida, W. & Yoshimoto, J. Control of exploitation-exploration meta-parameter in reinforcement learning. *Neural Netw.* **15**, 665–687 (2002).
10. Sugrue, L.P., Corrado, G.S. & Newsome, W.T. Matching behavior and the representation of value in the parietal cortex. *Science* **304**, 1782–1787 (2004).
11. Herrnstein, R.J. Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* **4**, 267–272 (1961).
12. Baum, W.M. Optimization and the matching law as accounts of instrumental behavior. *J. Exp. Anal. Behav.* **36**, 387–403 (1981).
13. Puterman, M.L. *Markov Decision Processes: Discrete Dynamic Stochastic Programming* (Wiley, New York, 1994).
14. Balleine, B.W., Delgado, M.R. & Hikosaka, O. The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* **27**, 8161–8165 (2007).
15. Balleine, B.W. & Killcross, S. Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci.* **29**, 272–279 (2006).
16. Kakade, S. & Dayan, P. Acquisition and extinction in autoshaping. *Psychol. Rev.* **109**, 533–544 (2002).
17. Daw, N.D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
18. Niv, Y., Daw, N.D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191**, 507–520 (2007).
19. Schweighofer, N. *et al.* Humans can adopt optimal discounting strategy under real-time constraints. *PLOS Comput. Biol.* **2**, e152 (2006).
20. Doya, K. Complementary roles of basal ganglia and cerebellum in learning and motor control. *Curr. Opin. Neurobiol.* **10**, 732–739 (2000).
21. Matsumoto, K., Suzuki, W. & Tanaka, K. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* **301**, 229–232 (2003).
22. Platt, M.L. & Glimcher, P.W. Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238 (1999).
23. Schultz, W., Tremblay, L. & Hollerman, J.R. Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereb. Cortex* **10**, 272–284 (2000).
24. Watanabe, M. Reward expectancy in primate prefrontal neurons. *Nature* **382**, 629–632 (1996).
25. Kawagoe, R., Takikawa, Y. & Hikosaka, O. Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* **1**, 411–416 (1998).
26. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of action-specific reward values in the striatum. *Science* **310**, 1337–1340 (2005).
27. Pasquereau, B. *et al.* Shaping of motor responses by incentive values through the basal ganglia. *J. Neurosci.* **27**, 1176–1183 (2007).
28. Komura, Y., Tamura, R., Uwano, T., Nishijo, H. & Ono, T. Auditory thalamus integrates visual inputs into behavioral gains. *Nat. Neurosci.* **8**, 1203–1209 (2005).
29. Minamimoto, T., Hori, Y. & Kimura, M. Complementary process to response bias in the centromedian nucleus of the thalamus. *Science* **308**, 1798–1801 (2005).
30. Montague, P.R., Dayan, P. & Sejnowski, T.J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
31. Schultz, W., Dayan, P. & Montague, P.R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
32. McClure, S.M., Berns, G.S. & Montague, P.R. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* **38**, 339–346 (2003).
33. O'Doherty, J. *et al.* Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
34. O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H. & Dolan, R.J. Temporal difference models and reward-related learning in the human brain. *Neuron* **38**, 329–337 (2003).
35. Seymour, B. *et al.* Temporal difference models describe higher-order learning in humans. *Nature* **429**, 664–667 (2004).
36. Reynolds, J.N. & Wickens, J.R. Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw.* **15**, 507–521 (2002).
37. Wickens, J.R., Begg, A.J. & Arbuthnott, G.W. Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience* **70**, 1–5 (1996).
38. Doya, K. Metalearning and neuromodulation. *Neural Netw.* **15**, 495–506 (2002).
39. Yacubian, J. *et al.* Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J. Neurosci.* **26**, 9530–9537 (2006).
40. Belova, M.A., Paton, J.J., Morrison, S.E. & Salzman, C.D. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* **55**, 970–984 (2007).
41. Seymour, B., Daw, N., Dayan, P., Singer, T. & Dolan, R. Differential encoding of losses and gains in the human striatum. *J. Neurosci.* **27**, 4826–4831 (2007).
42. Satoh, T., Nakai, S., Sato, T. & Kimura, M. Correlated coding of motivation and outcome of decision by dopamine neurons. *J. Neurosci.* **23**, 9913–9923 (2003).
43. Daw, N.D., Kakade, S. & Dayan, P. Opponent interactions between serotonin and dopamine. *Neural Netw.* **15**, 603–616 (2002).
44. Matsumoto, M. & Hikosaka, O. Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* **447**, 1111–1115 (2007).
45. Denk, F. *et al.* Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology (Berl.)* **179**, 587–596 (2005).
46. Kuhnen, C.M. & Knutson, B. The neural basis of financial risk taking. *Neuron* **47**, 763–770 (2005).
47. Tobler, P.N., O'Doherty, J.P., Dolan, R.J. & Schultz, W. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.* **97**, 1621–1632 (2007).
48. Cardinal, R.N. Neural systems implicated in delayed and probabilistic reinforcement. *Neural Netw.* **19**, 1277–1301 (2006).
49. Tanaka, S.C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.* **7**, 887–893 (2004).
50. Tanaka, S.C. *et al.* Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. *Neural Netw.* **19**, 1233–1241 (2006).
51. McClure, S.M., Laibson, D.I., Loewenstein, G. & Cohen, J.D. Separate neural systems value immediate and delayed monetary rewards. *Science* **306**, 503–507 (2004).
52. McClure, S.M., Ericson, K.M., Laibson, D.I., Loewenstein, G. & Cohen, J.D. Time discounting for primary rewards. *J. Neurosci.* **27**, 5796–5804 (2007).
53. Kable, J.W. & Glimcher, P.W. The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* **10**, 1625–1633 (2007).
54. Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M. & Rushworth, M.F. Separate neural pathways process different decision costs. *Nat. Neurosci.* **9**, 1161–1168 (2006).
55. Winstanley, C.A., Theobald, D.E., Dalley, J.W., Cardinal, R.N. & Robbins, T.W. Double dissociation between serotonergic and dopaminergic modulation of medial prefrontal and orbitofrontal cortex during a test of impulsive choice. *Cereb. Cortex* **16**, 106–114 (2006).
56. Tanaka, S.C. *et al.* Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum. *PLoS ONE* **2**, e1333 (2007).
57. Behrens, T.E., Woolrich, M.W., Walton, M.E. & Rushworth, M.F. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
58. Bouret, S. & Sara, S.J. Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends Neurosci.* **28**, 574–582 (2005).
59. Dayan, P. & Yu, A.J. Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network* **17**, 335–350 (2006).
60. Aston-Jones, G. & Cohen, J.D. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28**, 403–450 (2005).
61. Clarke, H.F., Dalley, J.W., Crofts, H.S., Robbins, T.W. & Roberts, A.C. Cognitive inflexibility after prefrontal serotonin depletion. *Science* **304**, 878–880 (2004).
62. Floresco, S.B., Tse, M.T. & Ghods-Sharifi, S. Dopaminergic and glutamatergic regulation of effort- and delay-based decision making. *Neuropsychopharmacology*, published online 5 September 2007 (doi:10.1038/sj.npp.1301565).
63. Walton, M.E., Bannerman, D.M. & Rushworth, M.F. The role of rat medial frontal cortex in effort-based decision making. *J. Neurosci.* **22**, 10996–11003 (2002).
64. Rogers, R.D., Lancaster, M., Wakeley, J. & Bhagwagar, Z. Effects of beta-adrenoceptor blockade on components of human decision-making. *Psychopharmacology (Berl.)* **172**, 157–164 (2004).
65. Cardinal, R.N. & Howes, N.J. Effects of lesions of the nucleus accumbens core on choice between small certain rewards and large uncertain rewards in rats. *BMC Neurosci.* **6**, 37 (2005).
66. Winstanley, C.A., Dalley, J.W., Theobald, D.E. & Robbins, T.W. Fractionating impulsivity: contrasting effects of central 5-HT depletion on different measures of impulsive behavior. *Neuropsychopharmacology* **29**, 1331–1343 (2004).
67. van Gaalen, M.M., van Koten, R., Schoffelmeer, A.N. & Vanderschuren, L.J. Critical involvement of dopaminergic neurotransmission in impulsive decision making. *Biol. Psychiatry* **60**, 66–73 (2006).
68. Kheramin, S. *et al.* Effects of orbital prefrontal cortex dopamine depletion on inter-temporal choice: a quantitative analysis. *Psychopharmacology (Berl.)* **175**, 206–214 (2004).
69. Robinson, E.S. *et al.* Similar effects of the selective noradrenaline reuptake inhibitor atomoxetine on three distinct forms of impulsivity in the rat. *Neuropsychopharmacology*, published online 18 July 2007 (doi:10.1038/sj.npp.1301487).
70. Rudebeck, P.H., Buckley, M.J., Walton, M.E. & Rushworth, M.F. A role for the macaque anterior cingulate gyrus in social valuation. *Science* **313**, 1310–1312 (2006).
71. Mobini, S. *et al.* Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology (Berl.)* **160**, 290–298 (2002).
72. Kheramin, S. *et al.* Effects of quinolinic acid-induced lesions of the orbital prefrontal cortex on inter-temporal choice: a quantitative analysis. *Psychopharmacology (Berl.)* **165**, 9–17 (2002).
73. Kheramin, S. *et al.* The effect of orbital prefrontal cortex lesions on performance on a progressive ratio schedule: implications for models of inter-temporal choice. *Behav. Brain Res.* **156**, 145–152 (2005).
74. Winstanley, C.A., Theobald, D.E., Cardinal, R.N. & Robbins, T.W. Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. *J. Neurosci.* **24**, 4718–4722 (2004).
75. Cardinal, R.N., Pennicott, D.R., Sugathapala, C.L., Robbins, T.W. & Everitt, B.J. Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science* **292**, 2499–2501 (2001).
76. Pothuizen, H.H., Jongen-Relo, A.L., Feldon, J. & Yee, B.K. Double dissociation of the effects of selective nucleus accumbens core and shell lesions on impulsive-choice behaviour and salience learning in rats. *Eur. J. Neurosci.* **22**, 2605–2616 (2005).
77. Hariri, A.R. *et al.* Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *J. Neurosci.* **26**, 13213–13217 (2006).
78. Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J. & Rushworth, M.F.

Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* **9**, 940–947 (2006).
79. Yu, A.J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–692 (2005).
80. Rogers, R.D. *et al.* Tryptophan depletion impairs stimulus-reward learning while methylphenidate disrupts attentional control in healthy young adults: implications for the monoaminergic basis of impulsive behaviour. *Psychopharmacology (Berl.)* **146**, 482–491 (1999).
81. Rogers, R.D. *et al.* Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: evidence for monoaminergic mechanisms. *Neuropsychopharmacology* **20**, 322–339 (1999).
82. Chamberlain, S.R. *et al.* Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science* **311**, 861–863 (2006).
83. Clarke, H.F., Walker, S.C., Dalley, J.W., Robbins, T.W. & Roberts, A.C. Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. *Cereb. Cortex* **17**, 18–27 (2007).
84. van der Plasse, G. *et al.* Medial prefrontal serotonin in the rat is involved in goal-directed behaviour when affect guides decision making. *Psychopharmacology (Berl.)* **195**, 435–449 (2007).
85. Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B. & Dolan, R.J. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
86. Corbit, L.H. & Balleine, B.W. The role of prelimbic cortex in instrumental conditioning. *Behav. Brain Res.* **146**, 145–157 (2003).
87. Balleine, B.W. Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. *Physiol. Behav.* **86**, 717–730 (2005).
88. Rachlin, H., Raineri, A. & Cross, D. Subjective probability and delay. *J. Exp. Anal. Behav.* **55**, 233–244 (1991).