

# Serotonin and the Evaluation of Future Rewards

## Theory, Experiments, and Possible Neural Mechanisms

NICOLAS SCHWEIGHOFER,<sup>a</sup> SAORI C. TANAKA,<sup>b</sup> AND KENJI DOYA<sup>b,c</sup>

<sup>a</sup>*Department of Biokinesiology and Physical Therapy, University of Southern California, Los Angeles, California, USA*

<sup>b</sup>*Computational Neuroscience Laboratories, Advanced Telecommunications Research Institute, Kyoto, Japan*

<sup>c</sup>*Initial Research Project, Okinawa Institute of Science and Technology, Okinawa, Japan*

**ABSTRACT:** The ability to select an action by considering both delays and amount of reward outcome is critical for survival and well-being of animals and humans. Previous animal experiments suggest a role of serotonin in action choice by modulating the evaluation of delayed rewards. It remains unclear, however, through which neural circuits, and through what receptors and intracellular mechanisms, serotonin affects the evaluation of delayed rewards. Here, we review experimental studies and computational theory of decisions under delayed rewards, and propose that serotonin controls the timescale of reward prediction by regulating neural activity in the basal ganglia.

**KEYWORDS:** discounting; impulsivity; reinforcement learning; discount rate; basal ganglia

### INTRODUCTION

A neuromodulator, such as serotonin, is a neurotransmitter that has spatially distributed and temporally extended effects on the recipient neurons and circuits.<sup>1-3</sup> Neuromodulators have traditionally been assumed to be involved in the control of general arousal.<sup>2,4</sup> Recent studies in molecular biology and neuroscience, however, have provided a more complex picture, with sometimes hard-to-reconcile data on the spatial localization and physiological effects of

Address for correspondence : Nicolas Schweighofer, Department of Biokinesiology and Physical Therapy, University of Southern California, 1450 E. Alcazar Street, Los Angeles, CA, 90089. Voice: 323-442-1838; fax: 323-442-1515.  
schweigh@usc.edu

Ann. N.Y. Acad. Sci. 1104: 289–300 (2007). © 2007 New York Academy of Sciences.  
doi: 10.1196/annals.1390.011

different neuromodulators and their receptors. In particular, despite numerous physiological and pharmacological studies, the role of serotonin is unclear. Such an understanding of the role of serotonin is all the more needed that serotonin dysfunction is thought to be linked to a variety of common mood and behavior disorders, such as depression<sup>5</sup> and impulsivity.<sup>6</sup>

Here, we focus on one important (though definitely *not* exclusive) aspect of serotonin, suggested by lesion and pharmacological data, in choice behaviors with delayed rewards. Specifically, we propose a theory on the role of serotonin from the viewpoint that it is a medium for signaling specific global parameters, which controls the timescale of the evaluation of delayed rewards. The article is organized as follows. We begin by discussing the concepts of reward values and reward discounting, first in animals and humans, and then in light of the framework of reinforcement learning theory developed in artificial intelligence. We then review the role of the serotonergic system in impulsive behavior, and show that existing data point to a role of serotonin in regulating the rate of discounting of delayed rewards. Finally, we suggest functional–anatomical models of the role of serotonin in the evaluation of future rewards.

## REWARDS DISCOUNTING IN ANIMALS AND HUMANS

When choosing between a larger but delayed reward, and a smaller but more immediate reward, we compare the “values” associated with each reward, and often choose the reward associated with the larger value.<sup>7</sup> Critical to these choices are the *shape* and the *steepness* of the reward values, which monotonically decrease as a function of the delay: the rewards are said to be discounted as a function of the delays (FIG. 1).

Two models that characterize the *shape* of reward discounting have been proposed: exponential<sup>8–10</sup> and hyperbolic.<sup>11–19</sup> The exponential discounting model leads to maximal gain under the assumption of a constant probability of reward loss per unit time and exact estimate of the time of the future reward delivery. The reward value  $V$  is then an exponential function of the delay:

$$V = R \exp(-k_e D), \quad (1)$$

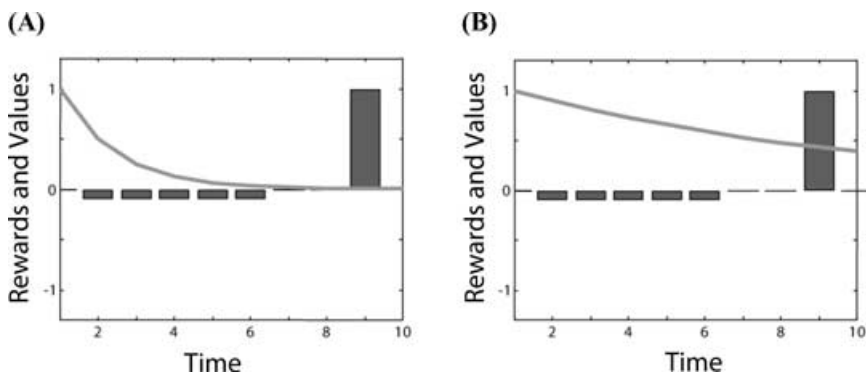
where  $k_e \geq 0$  is the decay rate, or equivalently

$$V = R \gamma^D, \quad (2)$$

where  $\gamma$  the discount factor ( $0 \leq \gamma < 1$ ). A small discount factor,  $\gamma$ , which is equivalent to a large decay rate  $k_e$  as  $\gamma = \exp(-k_e)$ , results in steeper discounting.

In repeated reward choice animal experiments, assuming a constant intertrial interval, if the animal consistently makes a choice that gives the same reward  $R$  after the same delay  $D$ , the average reward rate is the hyperbolic function of the delay<sup>20</sup>:

$$V = R/(T + D), \quad \text{with } T > 0, \quad (3)$$



**FIGURE 1.** The effect of the discount factor  $\gamma$  in decision making. The lines show the discounted value and the bars the potential rewards. In a scenario where repeated small negative rewards (costs) are expected before receiving a large positive reward, the cumulative future reward  $V$  becomes slightly negative if the discount factor is small (**A**, at Time 10) and positive if the discount factor is large enough (**B**, at Time 10). In the situation in **A**, if there is baseline behavior in which the expected rewards and costs are zero, the agent will take no action.

where  $T$  is the sum of all times except the delay in each trial (which is often equal to the intertrial interval). When  $T$  is large, discounting is steep. In human experiments, where subjects make a number of one off choices, the hyperbolic function is often given by

$$V = R/(1 + k_y D), \text{ with } k_y > 0, \tag{4}$$

where  $k_y \geq 0$  is the discounting parameter. When  $k_y$  is large, discounting is steep. A remarkable feature of hyperbolic discounting is that the rate of growth of the value, that is,  $V(D)/V(D + 1) = (1 + k_y (D + 1))/(1 + k_y D) = 1 + 1/(D + 1/k_y)$ , increase as the delay  $D$  becomes close to zero, while the growth rate is constant in exponential model ( $V(D)/V(D + 1) = 1/\gamma$ ).<sup>21</sup> As a consequence, hyperbolic discounting (but not exponential discounting) can result in an “irrational” preference reversal. For instance, a person may prefer one apple today to two apples tomorrow, but at the same time prefer two apples in 51 days to one apple in 50 days.<sup>22</sup> Thus, hyperbolic discounting is often presented as a struggle between oneself and one’s alter ego in the future, or similarly, between a myopic doer and a far-sighted planner.<sup>23,24</sup>

Most behavioral studies that have directly compared the two types of discounting in animals or humans have concluded that hyperbolic discounting better fits delayed reward choice data than does exponential discounting, for example.<sup>12–14,25–27</sup> We have, however, recently questioned the notion of hyperbolic reward discounting as a universal principle in humans.<sup>28</sup> In a reward decision task with temporal constraints in which each choice affects the time remaining for later trials, and in which the delays vary at each trial, we

demonstrated that most subjects adopted exponential discounting, and by doing so maximized their total gain.

The *steepness* of discounting specifies how far in the future delayed rewards should be considered. A large discount rate (the  $k_y$  parameter in equations 4, or similarly the  $k_e$  parameter in equation 1, which corresponds to small discount factor  $\gamma$  in equation 2), which results in steep discounting, biases individuals to acquire small and more immediate rewards, as delayed rewards have very small values. Individuals with impulse-control disorders, as well as heroin-, alcohol-, cigarette-, and cocaine- addicted individuals, have steeper discounting functions than controls.<sup>16,29-32</sup> To maximize future gains, the discount rate should be carefully adjusted, as we have shown in,<sup>28</sup> and as we will further discuss below.

### REWARDS DISCOUNTING IN ARTIFICIAL AGENTS

The theory of reinforcement learning, developed in artificial intelligence but initially loosely based on psychology,<sup>10</sup> is a particularly attractive model of animal learning, as it provides a normative model of how an adaptive agent should update its evaluation of values and behavioral policy. Furthermore, dopamine neurons have the formal characteristics of the teaching signal known as the temporal difference (TD) error.<sup>33</sup> Thus, reinforcement learning has both desirable theoretical properties and offers a good model of the basal ganglia and dopaminergic system.<sup>34-36</sup> The main issue in the theory of reinforcement learning is to maximize the long-term cumulative reward. Thus, central to reinforcement learning is the estimation of the value function

$$V(s(t)) = E \left[ \sum_{k=0}^{\infty} \gamma^k r(t+k) \right], \quad (5)$$

where  $r(t)$ ,  $r(t+1)$ ,  $r(t+2)$ ... are the rewards acquired by following a certain action policy  $P(a|s)$  starting from the state  $s(t)$ , and  $\gamma$  is a discount factor such that  $0 \leq \gamma < 1$ . We note here that this formulation of the value function reduces to that of the exponential value of a single reward in animal and human experiments (see above). The value function for the states before and after the transition should satisfy the consistency equation

$$V(s(t-1)) = E [r(t) + \gamma V(s(t))]. \quad (6)$$

Therefore, any deviation from the consistency equation, expressed as

$$\delta(t) = r(t) + \gamma V(s(t)) - V(s(t-1)), \quad (7)$$

should be zero on average. This signal is the TD error and is used as the teaching signal to learn the value function

$$\Delta V(s(t-1)) = \alpha \delta(t), \quad (8)$$

where  $\alpha$  is a learning rate.

The policy is usually defined via the action value function  $Q(s(t), a)$ , which represents how much future rewards the agent would get by taking the action  $a$  at state  $s(t)$  and following the current policy in subsequent steps. One common way for stochastic action selection that encourages exploitation is to compute the probability to take an action by the soft-max function

$$P(a_i | s(t)) = \frac{e^{\beta Q(s(t), a_i)}}{\sum_{j=1}^M e^{\beta Q(s(t), a_j)}}, \quad (9)$$

where the meta-parameter  $\beta$  is called the inverse temperature.

Crucial to successful reinforcement learning is the careful setting of the three meta-parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ .

- The learning rate  $\alpha$  controls the speed of learning, as small learning rates induce slow learning, and large learning rates induce instability of memory update.
- The inverse temperature  $\beta$  controls the exploitation–exploration trade-off. Ideally,  $\beta$  should initially be low to allow large exploration, when the agent does not have a good mapping of which actions will be rewarding, and gradually increase as the agent reaps higher and higher rewards.
- The discount factor  $\gamma$  determines how far into the future the agent should consider reward prediction and action selection. The setting of the discount factor is particularly important when there is a conflict between the immediate and long-term outcomes (FIG. 1). In real life, it is often the case that one would have to pay some immediate cost (negative reward) to achieve a larger future reward, for example, long travels in foraging, or daily cultivation for harvest. It is also the case that one should avoid positive immediate reward if it is associated with a large negative reward in the future. If  $\gamma$  is small, the agent learns to behave only for short-term rewards. Although a large  $\gamma$  (close to 1) promotes the agent to learn to act for long-term rewards, there are at least three reasons why  $\gamma$  should not be too large. First, any real learning agent, either artificial or biological, has a limited lifetime. Thus, a discounted value function is equivalent to a nondiscounted value function for an agent with a constant death rate of  $1 - \gamma$ . Second, an agent has to acquire some rewards in time; for instance, an animal must find food before it starves; a robot must recharge its battery before it is exhausted. Third, if the environmental dynamics is highly stochastic or the dynamics is nonstationary, long-term prediction is unreliable. The complexity of learning a value function has been shown to increase with the increase of  $1/(1 - \gamma)$ .<sup>37</sup>

## SEROTONIN AND THE TIMESCALE OF REWARD PREDICTION

In a seminal paper, Soubrie<sup>6</sup> suggested that serotonergic neurons are brought into play whenever behavioral inhibition is required. He pointed out that reduced serotonin was linked to impulsive behavior: the animal is less able to “wait.” Low levels of serotonin are often associated with behaviors regarded as impulsive, such as aggression,<sup>38,39</sup> or failures to not respond in response to a stimulus in a no-go trial.<sup>40</sup> “Impulsivity” is a multidimensional phenomenon, however; Evenden described three varieties of impulsivity<sup>41</sup> (1) unreliable sensory discrimination (attention/preparation), (2) making premature responses in situations that require the postponement of actions (execution), and (3) choosing a smaller immediate reinforcer rather than a larger delayed reinforcer (outcome).

Serotonin dysfunction seems to specifically lead to the third type of impulsivity, as shown by delayed reward choice experiments. Using Mazur’s<sup>42</sup> adjusting delay paradigm, Wogar *et al.*<sup>43</sup> showed that serotonin is involved in maintaining effectiveness of delayed reinforcers: rats with lesioned ascending serotonergic system did not wait for the large reinforcer when offered a choice between immediate small and large delayed reinforcers. Further, it has been shown that serotonin depletion results in failure of delayed rewards to motivate behavior.<sup>44</sup> On the other hand, increased serotonin levels decrease impulsive choice.<sup>45,46</sup> Two variables can possibly lead to this effect of serotonin on delayed reward choices: the reinforcer magnitude or the delay to the reward. Mobini and coworkers showed that serotonin depletion in the forebrain steepens hyperbolic discounting, that is, lower serotonin resulted in higher value of the parameter  $k_y$  in equation 4.<sup>47</sup> These authors found no modulation of the magnitude of the reward, however. Although serotonergic fibers arise from both dorsal and median raphe nuclei, the dorsal raphe nucleus seems to be the source of serotonergic neurons involved in impulsivity.<sup>48</sup>

These experimental data are consistent with the hypothesis that serotonin neurons in the dorsal raphe nucleus control the timescale of reward evaluation. In this hypothesis, serotonin controls the discount factor (larger  $\gamma$  in the exponential model of equation 2 and smaller  $k_y$  in the hyperbolic model of equation 4 with the higher level of serotonin), which controls the evaluation of future reward. With low serotonin levels, because delayed rewards have a low value, agents choose the small immediate reward over the large delayed reward, characteristics of impulsivity.

### MODULATION OF THE DISCOUNT RATE BY SEROTONIN

#### *Serotonin Regulation of the Discount Rate*

A body of experimental evidence strongly suggests that learning of the value functions occurs in the central nervous system, presumably in the basal ganglia.

In particular, it has been found that dopamine neuron activity resembles closely the temporal difference (TD) error (e.g., Refs. 33,35). Further, recent experimental studies suggest that the striatum computes value functions,<sup>49–53</sup> and it has been suggested that action selection occurs in the globus pallidus.<sup>54,55</sup> Dopamine induces long-lasting plasticity in corticostriatal synapses,<sup>56,57</sup> and thus could allow learning of the value functions. Serotonergic neurons project to the basal ganglia,<sup>58–60</sup> and control dopamine release in the striatum,<sup>61,62</sup> and thus could modulate the computation of value function (see role of the discount factor  $\gamma$  in equation 5), and the dopaminergic activity (see role of the discount factor  $\gamma$  in equation 7).

The control of the timescale of reward prediction could be achieved by activating or deactivating multiple reward prediction pathways in the basal ganglia. A parallel learning mechanism in the corticobasal ganglia loops used for reward prediction at a variety of timescales would have the merit of enabling flexible selection of a relevant timescale appropriate for the task and the environment at the time of decision making.

This view is supported by our previous brain imaging study,<sup>50</sup> in which we developed a “Markov decision task” to probe decision making in a dynamic context, with small losses followed by a large positive reward (as in FIG. 1). By analyzing subjects’ performance data using a reward value model with different discount factors (as in equation 2), we found a gradient of activation within the striatum for prediction error of rewards at different timescales. The graded maps are consistent with the topographic corticostriatal organization,<sup>63</sup> and suggest that areas that project to the more dorsoposterior part of the striatum are involved in reward prediction at a longer timescale. These results are also consistent with the observations that localized damages within the limbic and cognitive corticobasal ganglia loops manifest as deficits in evaluation of future rewards<sup>64–68</sup> and learning of multistep behaviors.<sup>69</sup>

A possible mechanism underlying these observations is that these different corticobasal ganglia subloops are differentially activated by the ascending serotonergic system from the dorsal raphe nucleus. Although serotonergic projections are relatively diffuse and global, differential expression of serotonergic receptors in the cortical areas and in the ventral and dorsal striatum<sup>58,60</sup> could result in differential modulation. The distribution of serotonin receptor subtypes is not uniform within the striatum, as various subtypes receptors subtypes, with different affinities and intracellular effects, are differentially distributed in the ventral and dorsal parts of the striatum.<sup>58,59</sup> Such differential distributions could allow differential striatal modulation of activities under different serotonin levels. A positron emission tomography (PET) experiment using particular receptor radioligands may shed light on these mechanisms.<sup>70,71</sup>

### ***Regulation of Serotonin Levels***

How could serotonergic neurons themselves be regulated? In a previous computational study, we proposed a simple, yet robust and biologically plausible

algorithm that regulates the reinforcement learning meta-parameters, which include the discount factor.<sup>72</sup> The algorithm is based on Stochastic Real Value Units (SRV) algorithm.<sup>73</sup> An SRV unit output is produced by adding to the weighted sum of its input pattern a small random perturbation that provides the unit with the variability necessary to explore its activity space. When a perturbation results in increased probability of receiving extra rewards, the unit's input synaptic efficacies are adjusted such that the output moves in the direction in which it was perturbed. We expanded the idea of SRV units and take it on to a level higher—that is, we proposed that neuromodulator neurons are themselves SRV-like units. According to this hypothesis, serotonergic, or “gamma” neuron, would be governed by both a slowly varying mean activity term and a noise term. The noise term corresponds to a spontaneous change in the tonic firing of the gamma neuron. We proposed that mean activity of the serotonergic neuron is updated by Hebbian-like learning rule that correlates with the random perturbation and the difference between a short-term and a long-term running reward average. If a positive perturbation in the neurons' firing rate yields a state of affair slightly superior to that the animal expects, then the discount rate  $\gamma$  is increased, and vice versa.

Although untested, this simple algorithm is biologically plausible. Spontaneous fluctuations of the tonic firing of the neuromodulator neuron may arise naturally with the wake–sleep cycle and/or the level of activity of the animal.<sup>74</sup> The difference between a short-term and a long-term running average of the reward could be carried by dopaminergic neuron activity.<sup>75</sup> As dopaminergic neurons send projections to serotonergic neurons,<sup>76</sup> we predicted that dopamine-dependent plasticity is present in these neurons.

## CONCLUDING REMARKS

Serotonin seems to play a major role in depression, as selective serotonin reuptake inhibitors (SSRI) and other serotonin-enhancing drugs are known to be effective for unipolar depression and bipolar disorders. The therapeutic mechanisms of these drugs are still not well understood, however.<sup>5</sup> Our theory of the role of serotonin, although primarily aimed at explaining impulsive behavior, may also perhaps explain certain aspects of depressive behavior: low serotonin levels could lead to the situation shown in the left of FIGURE 1, in which the optimal policy is not to act. Future experiments using delayed reward paradigms could be designed to study impulsivity in depressed patients.

## ACKNOWLEDGMENTS

This work was supported in part by CREST, and by grants NIH P20 RR020700–02 and NSF IIS 0535282 to NS.

## REFERENCES

1. KATZ, P.S. 1999. *Beyond Neurotransmission: Neuromodulation and Its Importance for Information Processing*. Oxford University Press. Oxford, UK.
2. SAPER, C.B. 2000. Brain stem modulation of sensation, movement and consciousness. *In Principles of Neural Science* E.R. Kandel, J.H. Schwartz & T.M. Jessel, Eds.: 889–909. McGraw-Hill. New York.
3. MARDER, E. & V. THIRUMALAI. 2002. Cellular, synaptic and network effects of neuromodulation. *Neural Netw.* **15**: 479–493.
4. ROBBINS, T.W. 1997. Arousal systems and attentional processes. *Biol. Psychol.* **45**: 57–71.
5. WONG, M.L. & J. LICINIO. 2001. Research and treatment approaches to depression. *Nat. Rev. Neurosci.* **2**: 343–351.
6. SOUBRIE, P. 1986. Serotonergic neurons and behavior. *J. Pharmacol.* **17**: 107–112.
7. PLATT, M.L. 2002. Neural correlates of decisions. *Curr. Opin. Neurobiol.* **12**: 141–148.
8. SAMUELSON, P.A. 1937. A note on measurement of utility. *Rev. Econ. Stud.* **4**: 155–161.
9. KAGEL, J.H., L. GREEN & T. CARACO. 1986. When foragers discount the future: constraints or adaptation? *Anim. Behav.* **34**: 271–283.
10. SUTTON, R.S. & A.G. BARTO. 1998. *Reinforcement Learning*. The MIT Press. Cambridge, MA.
11. AINSLIE, G. 1975. Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychol. Bull.* **82**: 463–496.
12. MAZUR, J.E. 1987. An adjusting procedure for studying delayed reinforcement. *In Quantitative Analysis of Behavior*. Vol. V: The Effect of Delay and Intervening Events. M.L. Commons, *et al.*, Eds. Erlbaum. London.
13. RODRIGUEZ, M.L. & A.W. LOGUE. 1988. Adjusting delay to reinforcement: comparing choice in pigeons and humans. *J. Exp. Psychol. Anim. Behav. Process* **14**: 105–117.
14. RACHLIN, H., A. RAINERI & D. CROSS. 1991. Subjective probability and delay. *J. Exp. Anal. Behav.* **55**: 233–244.
15. BATESON, M. & A. KACELNIK. 1996. Rate currencies and the foraging starling: the fallacy of the averages revisited. *Behav. Ecol.* **7**: 341–352.
16. BICKEL, W.K., A.L. ODUM & G. J. MADDEN. 1999. Impulsivity and cigarette smoking: delay discounting in current, never, and ex-smokers. *Psychopharmacology (Berl.)* **146**: 447–454.
17. HO, M.Y. *et al.* 1999. Theory and method in the quantitative analysis of “impulsive choice” behaviour: implications for psychopharmacology. *Psychopharmacology (Berl.)* **146**: 362–372.
18. KIRBY, K.N., N.M. PETRY & W. K. BICKEL. 1999. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *J. Exp. Psychol. Gen.* **128**: 78–87.
19. PETRY, N.M. 2001. Pathological gamblers, with and without substance use disorders, discount delayed rewards at high rates. *J. Abnorm. Psychol.* **110**: 482–487.
20. KACELNIK, A. 1997. Normative and descriptive models of decision making: time discounting and risk sensitivity. *In Characterizing Human Psychological Adaptations*: 51–70. Wiley. Chichester.

21. LAIBSON, D.I. 2003. Intertemporal decision making. *In* Encyclopedia of Cognitive Science. Nature Publishing Group. London.
22. THALER, R.H. & H.M. SHEFRIN. 1981. An economic theory of self-control. *J. Pol. Economy* **89**: 392–410.
23. AINSLIE, G. 2005. Precipitous breakdown of will. *Behav. Brain Sci.* **28**: 635–650.
24. THALER, R.H. 1981. Some empirical evidence on dynamic inconsistency. *economic letters*. **8**: 201–207.
25. KIRBY, K.N. & N.N. MARAKOVIC. 1995. Modeling myopic decisions: evidence for hyperbolic delay-discounting within subjects and amounts. *Org. Behav. Human Decision Proc.* **64**: 22–30.
26. MYERSON, J. & L. GREEN. 1995. Discounting of delayed rewards: models of individual choice. *J. Exp. Anal. Behav.* **64**: 263–276.
27. ANGELETOS, G.M. *et al.* 2001. The hyperbolic consumption model: calibration, simulation, and empirical evaluation. *J. Eco. Prospect.* **15**: 47–68.
28. SCHWEIGHOFER, N. *et al.* 2006. Humans can adopt optimal discounting strategy under real-time constraints. *Plos Comp. Biol.* **11**: 1349–1356.
29. CREAN, J.P., H. DE WIT & J. B. RICHARDS. 2000. Reward discounting as a measure of impulsive behavior in a psychiatric outpatient population. *Exp. Clin. Psychopharmacol.* **8**: 155–162.
30. MADDEN, G.J. *et al.* 1997. Impulsive and self-control choices in opioid-dependent patients and non-drug-using control participants: drug and monetary rewards. *Exp. Clin. Psychopharmacol.* **5**: 256–262.
31. VUCHINICH, R.E. & C.A. SIMPSON. 1998. Hyperbolic temporal discounting in social drinkers and problem drinkers. *Exp. Clin. Psychopharmacol.* **6**: 292–305.
32. COFFEY, S.F. *et al.* 2003. Impulsivity and rapid discounting of delayed hypothetical rewards in cocaine-dependent individuals. *Exp. Clin. Psychopharmacol.* **11**: 18–25.
33. SCHULTZ, W. 1998. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**: 1–27.
34. HOUK, J.C., J. DAVIS & D. BEISER, Eds. 1995. Models of Information Processing in the Basal Ganglia. 249–270. The MIT Press. Cambridge, MA.
35. MONTAGUE, P.R., P. DAYAN & T. J. SEJNOWSKI. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**: 1936–1947.
36. DOYA, K. 2000. Complementary roles of basal ganglia and cerebellum in learning and motor control. *Curr. Opin. Neurobiol.* **10**: 732–739.
37. LITTMAN, M.L., T.L. DEAN & L. P. KAEHLBLING. 1995. On the complexity of solving Markov decision problems. Eleventh International Conference on Uncertainty in Artificial Intelligence.
38. BUHOT, M.C. 1997. Serotonin receptors in cognitive behaviors. *Curr. Opin. Neurobiol.* **7**: 243–254.
39. ROBBINS, T.W. 2000. From arousal to cognition: the integrative position of the prefrontal cortex. *Prog. Brain Res.* **126**: 469–483.
40. HARRISON, A.A., B.J. EVERITT & T. W. ROBBINS. 1999. Central serotonin depletion impairs both the acquisition and performance of a symmetrically reinforced go/no-go conditional visual discrimination. *Behav. Brain Res.* **100**: 99–112.
41. EVENDEN, J.L. 1999. Varieties of impulsivity. *Psychopharmacology (Berl.)* **146**: 348–361.

42. MAZUR, J.E., M. SNYDERMAN & D. COE. 1985. Influences of delay and rate of reinforcement on discrete-trial choice. *J. Exp. Psychol. Anim. Behav. Process* **11**: 565–575.
43. WOGAR, M.A., C.M. BRADSHAW & E. SZABADI. 1993. Effect of lesions of the ascending 5-hydroxytryptaminergic pathways on choice between delayed reinforcers. *Psychopharmacology (Berl.)* **111**: 239–243.
44. RAHMAN, S. *et al.* 2001. Decision making and neuropsychiatry. *Trends Cogn. Sci.* **5**: 271–277.
45. BIZOT, J. *et al.* 1999. Serotonin and tolerance to delay of reward in rats. *Psychopharmacology (Berl.)* **146**: 400–412.
46. POULOS, C.X., J.L. PARKER & A.D. LE. 1996. Dexfenfluramine and 8-OH-DPAT modulate impulsivity in a delay-of-reward paradigm: implications for a correspondence with alcohol consumption. *Behav. Pharmacol.* **7**: 395–399.
47. MOBINI, S. *et al.* 2000. Effect of central 5-hydroxytryptamine depletion on intertemporal choice: a quantitative analysis. *Psychopharmacology (Berl.)* **149**: 313–318.
48. CARLI, M. & R. SAMANIN. 2000. The 5-HT(1A) receptor agonist 8-OH-DPAT reduces rats' accuracy of attentional performance and enhances impulsive responding in a five-choice serial reaction time task: role of presynaptic 5-HT(1A) receptors. *Psychopharmacology (Berl.)* **149**: 259–268.
49. O'DOHERTY, J. *et al.* 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**: 452–454.
50. TANAKA, S.C. *et al.* 2004. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.* **7**: 887–893.
51. SHIDARA, M., T.G. AIGNER & B.J. RICHMOND. 1998. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J. Neurosci.* **18**: 2613–2625.
52. TREMBLAY, L. & W. SCHULTZ. 2000. Reward-related neuronal activity during gogo task performance in primate orbitofrontal cortex. *J. Neurophysiol.* **83**: 1864–1876.
53. SAMEJIMA, K. *et al.* 2005. Representation of action-specific reward values in the striatum. *Science* **310**: 1337–1340.
54. BERNS, G.S. & T.J. SEJNOWSKI. 1998. A computational model of how the basal ganglia produce sequences. *J. Cogn. Neurosci.* **10**: 108–121.
55. DOYA, K. 2000. Metalearning and neuromodulation. *Math. Sci.* **38**: 19–24.
56. REYNOLDS, J.N., B.I. HYLAND & J.R. WICKENS. 2001. A cellular mechanism of reward-related learning. *Nature* **413**: 67–70.
57. REYNOLDS, J.N. & J.R. WICKENS. 2002. Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw.* **15**: 507–521.
58. COMPAN, V. *et al.* 1998. Selective increases in serotonin 5-HT1B/1D and 5-HT2A/2C binding sites in adult rat basal ganglia following lesions of serotonergic neurons. *Brain Res.* **793**: 103–111.
59. VARNAS, K., C. HALLDIN & H. HALL. 2004. Autoradiographic distribution of serotonin transporters and receptor subtypes in human brain. *Hum. Brain Mapp.* **22**: 246–260.
60. MIJNSTER, M.J. 1997. Regional and cellular distribution of serotonin 5-hydroxytryptamine<sub>2a</sub> receptor mRNA in the nucleus accumbens, olfactory tubercle, and caudate putamen of the rat. *J. Comp. Neurol.* **389**: 1–11.

61. SERSHEN, H., A. HASHIM & A. LAJTHA. 2000. Serotonin-mediated striatal dopamine release involves the dopamine uptake site and the serotonin receptor. *Brain Res. Bull.* **53**: 353–367.
62. DE DEURWAERDERE, P & U. SPAMPINATO. 1999. Role of serotonin(2A) and serotonin(2B/2C) receptor subtypes in the control of accumbal and striatal dopamine release elicited *in vivo* by dorsal raphe nucleus electrical stimulation. *J. Neurochem.* **73**: 1033–1042.
63. MIDDLETON, F.A. & P. L. STRICK. 2000. Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res. Brain Res. Rev.* **31**: 236–250.
64. BECHARA, A., H. DAMASIO & A.R. DAMASIO. 2000. Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex* **10**: 295–307.
65. CARDINAL, R.N. *et al.* 2001. Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science* **292**: 2499–2501.
66. ROLLS, E.T. 2000. The orbitofrontal cortex and reward. *Cereb. Cortex* **10**: 284–294.
67. EAGLE, D.M. *et al.* 1999. Effects of regional striatal lesions on motor, motivational, and executive aspects of progressive-ratio performance in rats. *Behav. Neurosci.* **113**: 718–731.
68. PEARS, A. *et al.* 2003. Lesions of the orbitofrontal but not medial prefrontal cortex disrupt conditioned reinforcement in primates. *J. Neurosci.* **23**: 11189–11201.
69. HIKOSAKA, O. 1999. Parallel neural networks for learning sequential procedures. *Trends Neurosci.* **22**: 464–471.
70. HUANG, Y. *et al.* 2005. Synthesis of potent and selective serotonin 5-HT<sub>1B</sub> receptor ligands. *Bioorg. Med. Chem. Lett.* **15**: 4786–4789.
71. LARISCH, R. *et al.* 2003. Influence of synaptic serotonin level on [<sup>18</sup>F]altanserin binding to 5HT<sub>2</sub> receptors in man. *Behav. Brain Res.* **139**: 21–29.
72. SCHWEIGHOFER, N. & K. DOYA. 2003. Meta-learning in reinforcement learning. *Neural Netw.* **16**: 5–9.
73. GULLAPALLI, V. 1990. A stochastic reinforcement learning algorithm for learning real-valued functions. *Neural Netw.* **3**: 671–692.
74. JACOBS, B.L. & C.A. FORNAL. 1993. 5-HT and motor control: a hypothesis. *Trends Neurosci.* **16**: 346–352.
75. DAW, N.D., S. KAKADE & P. DAYAN. 2002. Opponent interactions between serotonin and dopamine. *Neural Netw.* **15**: 603–616.
76. HAJ-DAHMANE, S. 2001. D<sub>2</sub>-like dopamine receptor activation excites rat dorsal raphe 5-HT neurons *in vitro*. *Eur. J. Neurosci.* **14**: 125–134.