Self-Organization in the Basal Ganglia with Modulation of Reinforcement Signals

Hiroyuki Nakahara
hiro@brain.riken.go.jp
Shun-ichi Amari
amari@brain.riken.go.jp
Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute
2-1 Hirosawa, Wako, Saitama, 351-0198, Japan

Okihide Hikosaka
hikosaka@med.juntendo.ac.jp
Department of Physiology, School of Medicine, Juntendo University, 2-1-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

Self-organization is one of fundamental brain computations for forming efficient representations of information. Experimental support for this idea has been largely limited to the developmental and reorganizational formation of neural circuits in the sensory cortices. We now propose that self-organization may also play an important role in short-term synaptic changes in reward-driven voluntary behaviors. It has recently been shown that many neurons in the basal ganglia change their sensory responses flexibly in relation to rewards. Our computational model proposes that the rapid changes in striatal projection neurons depend on the subtle balance between the Hebb-type mechanisms of excitation and inhibition, which are modulated by reinforcement signals. Simulations based on the model are shown to produce various types of neural activity similar to those found in experiments.

1 Introduction

The basal ganglia (BG) are well known to contribute to sequential motor and cognitive behaviors (Knopman & Nissen, 1991; Graybiel, 1995). Almost the entire cerebral cortex projects to the BG, and the BG project mainly back to the frontal cortex through the thalamus and to the superior colliculus. A striking fact about the BG is a vast convergent projection from the cerebral cortex to the striatum, a major input zone of the BG (Oorschot, 1996) and another convergent projection from the striatum to the output nuclei of the BG, that is, the global pallidus internal segments (GPi), and the substantia nigra pars reticulata (SNr). Given this fact, it is expected that the BG

have efficient representations of cortical inputs to interact effectively with the cerebral cortex (Graybiel, Aosaki, Flaherty, & Kimura, 1994). We should note that the majority of the neurons in the striatum, including the projection neurons and some types of interneurons, are GABAergic, working most likely as inhibitory within the striatum as well as to target neurons in projection areas (Kita, 1993; Kawaguchi, Wilson, Augood, & Emson, 1995; Wilson, 1998).

The BG, the striatum in particular, receive rich reinforcement signals from dopaminergic (DA) neurons, which originate in the substantia nigra pars compacta (SNc). It has been observed that the neural responses in the striatum are strongly modulated by DA neurons or reward conditions (Aosaki et al., 1994; Schultz, Apicella, Romo, & Scarnati, 1995; Kawagoe, Takikawa, & Hikosaka, 1998). DA neurons exhibit a phasic activity when an unexpected reward occurs or when a conditioning stimuli appears that allows the subject to anticipate a coming reward (Schultz, 1998). Driven by these reinforcement signals carried by DA neurons, the striatum has been

---

Figure 1: *Facing page*. (A) Example of the memory-guided saccade task in a one-direction-rewarded condition (1DR). In experiments, there were two reward conditions: 1DR and ADR (all-direction-rewarded condition) conditions. The task procedure in each trial is the same as a memory-guided saccade task between both conditions except reward conditions: a task trial started with the onset of a central fixation point, which the monkeys had to fixate. A cue stimulus (spot of light) then came at one of the four directions. After the fixation point turned off, the monkeys had to make a saccade to the cued location. In ADR, all directions are rewarded after a saccade to the cued location in each trial. In 1DR, throughout a block of the experiment (60 trials), only one direction was rewarded among four directions. In the example shown here, the right direction is rewarded. Even for nonrewarded directions in 1DR, the monkeys had to make correct saccades; otherwise, the same trial was repeated. 1DR was performed in four blocks, in each of which a different direction was rewarded. Other than the actual reward, no indication was given to the monkeys as to which direction was to be rewarded. (B) Examples of the three types of neural responses observed in the experiment are shown over four 1DR blocks (taken from Kawagoe et al., 1998). Data obtained in each block of 1DR (left) are shown as a polar diagram indicating the magnitudes of the responses for four cue directions. Rewarded direction is indicated by a bull's-eye mark. (Top) Flexible type (the most frequently observed type) changes its preferred direction quickly to the rewarded direction in each 1DR block. (Middle) Conservative type maintains its preferred direction (rightward for this neuron) across 1DR blocks (at least in two of four 1DR blocks), while the response toward each direction is enhanced when it is rewarded in each 1DR block. (Bottom) Reverse type (less frequently observed than the other two types) shows the smallest response to the rewarded direction in each 1DR block.

considered to undergo heterosynaptic plasticity (Calabresi, Maj, Pisani, Mercuri, & Bernardi, 1992; Wickens & Kötter, 1995) and to contribute to skill memory formation through the cortico-basal ganglia loops (Marsden, 1980; Alexander, Crutcher, & DeLong, 1990; Knowlton, Mangels, & Squire, 1996; Hikosaka et al., 1999; Nakahara, Doya, & Hikosaka, 2001).

In a recent experiment (Kawagoe et al., 1998), reward-modulated changes of neural responses in the caudate, a part of the striatum, were investigated in a systematic manner, using asymmetrically rewarded memory-guided saccade tasks, in which one of the four directions was randomly chosen as the saccade target in a trial and the monkeys had to make a memory-guided saccade in the trial (see Figure 1A). In this task, there were two reward
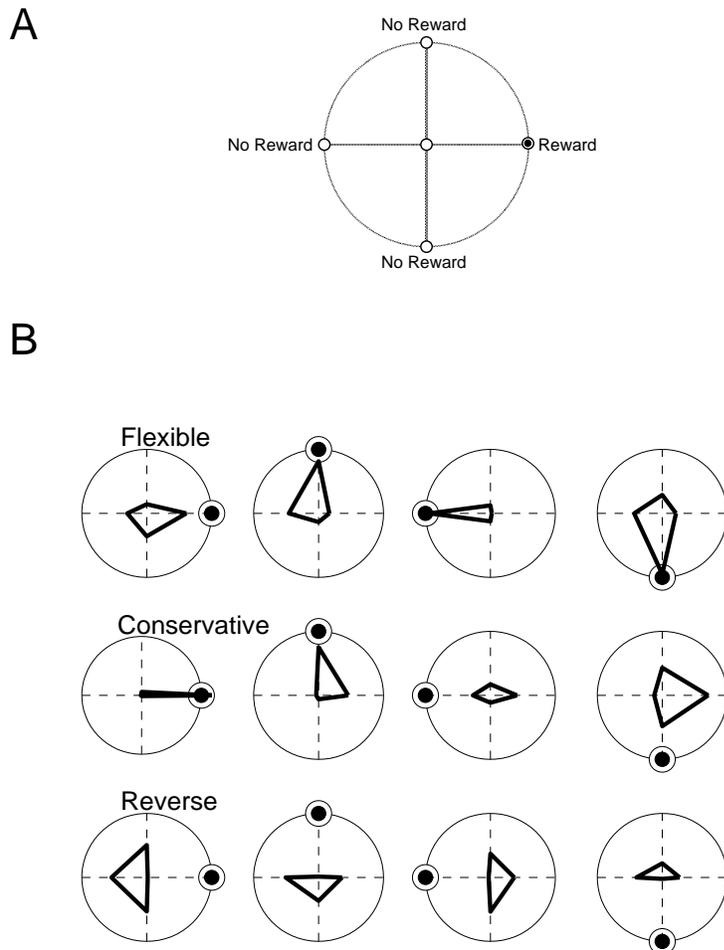
conditions: all-directions-rewarded condition (ADR), where all four directions were rewarded in a block of experiments after a correct saccade in a trial, and one-direction-rewarded condition (1DR), where only one fixed direction was rewarded after a correct saccade in a trial (the other three directions were not rewarded throughout a block even when a correct saccade was made). In visual and memory-related periods, the preferred directions of the caudate neural responses, observed in ADR (Kawagoe et al., 1998), were typically found to be contralateral, as previously reported (Hikosaka, Sakamoto, & Usui, 1989), so that the caudate neurons exhibited spatial-directional selective responses. The caudate neural responses, however, were strongly modulated by the rewarded directions in the 1DR blocks (see Figure 1B). Three typical patterns were found in the response changes: flexible, conservative, and reverse type patterns (see Figure 1B; their definitions are provided in the legend). When a 1DR block is altered, the caudate responses change within ten to a few tens of trials and develop much slower than the DA activities (Kawagoe et al., 1998; Kawagoe, Takikawa, & Hikosaka, 1999) (see section 2.1). Such changes are supposed to be caused by the synaptic plasticity under the influence of the DA activity (Calabresi et al., 1992; Wickens & Kötter, 1995; Reynolds & Wickens, 2000), while the DA-induced changes in the internal states of the striatal neurons may play a role as well (Surmeier, Song, & Yan, 1996). These results suggest that

---

Figure 2: *Facing page*. (A) Schematic diagram of the relationship of the cerebral cortex and the caudate. Spatial information is conveyed from the cerebral cortex to the caudate, while reinforcement signals are provided via dopaminergic projections (DA). Black and white arrowheads indicate excitatory and inhibitory connections, respectively. (B) Scheme of self-organization in the caudate. Directional information is topographically represented in the cerebral cortex with two components: one reflecting each directional selectively (N_i) and the other reflecting some overlap between different directions (M). Common inputs (M) are shared by cortical representations for different directions (x_i), and in the figure, cortical representations for two directions are shown. Reinforcement signals (α) carried by dopamine (DA) neurons influence cortical inputs. An integrated inhibitory input to caudate neurons is denoted by x_0. (C) Schematic example of a conservative-type neuron at equilibrium, responding to its preferred (but non-rewarded) direction (x_1) and a rewarded direction (x_2). This neuron does not fire to x_3 (x_4 is dropped here for simplicity). After the learning converges, the synaptic weight converges to $\bar{w} = \frac{1}{2}(x_1 + x_2)$. By assuming $\| w \| = 1$, each of the inner products between the input (x_i) and the weight ($\bar{w}$) with and without reinforcement signal modulation is indicated on the right. Inhibitory effect is summarized by λ, which is also indicated on the right. Since $\bar{w} \cdot x_3 < \lambda < \bar{w} \cdot x_1$, $\bar{w} \cdot x_2(1 + \alpha_2)$, the neuron responds to x_1, and x_2 with reinforcement signal, but not to x_3. If α_2 is not provided, the neuron does not respond to x_2. The dashed cone region indicates the receptive region of this neuron. The neuron responds to all of the inputs, which can be modulated by reinforcement signals, in this region.

the caudate neurons rapidly alter their response properties guided by the DA activity when the reward condition changes (see the next section and section 4). Importantly, the pattern of changes in a response can vary in each neuron and is classified roughly into one of three categories.

We propose that the caudate neurons self-organize their responses under the control of both the intrinsic cortical inputs representing each cue direction and the DA reinforcement signals reflecting the reward conditions (Schultz, 1998; see Figure 2A). We therefore study a simplified self-organization model with
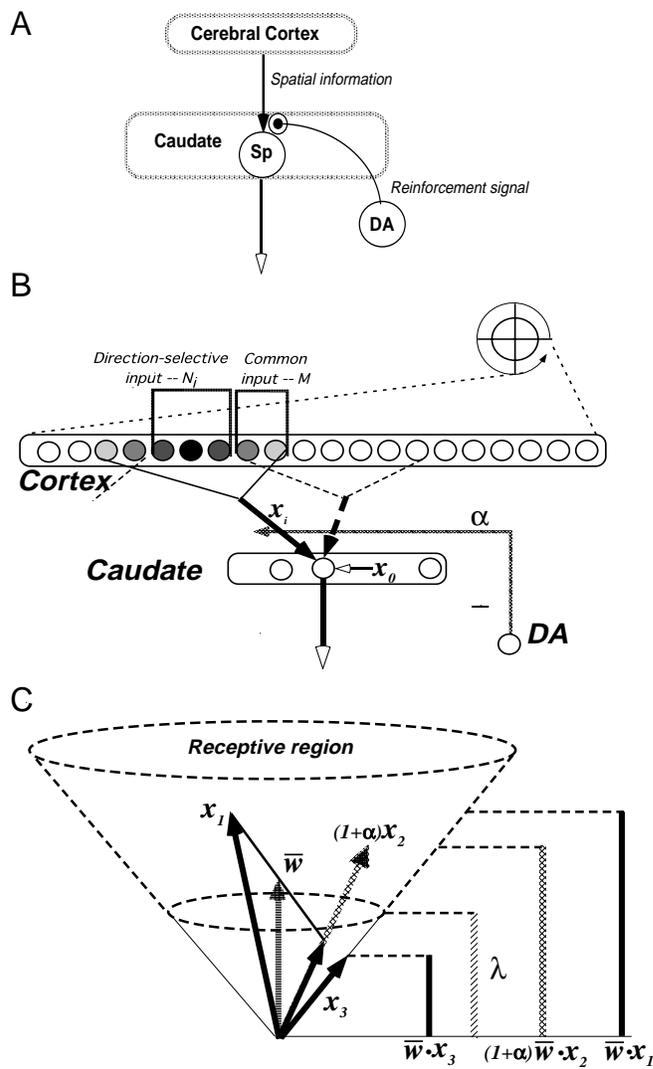
reinforcement signals (see Figure 2B).It is possible to analyze its behaviors rigorously by giving conditions that guarantee the requested behaviors (Kawagoe et al., 1998). Our model includes plastic changes in both the excitatory and inhibitory synapses, and their subtle balance generates a variety of phenomena, as seen in experiments. Under each condition, we can determine which pattern of neural responses appears, using the parameters of both the reinforcement signal and the inhibitory effect with respect to the cortical representations. Our model can explain these typical patterns of modulation by the same simple mechanism in a unified way, which would help us predict the relationship between the neural response modulations and the cortico-striatal and nigra-striatal connections of the neurons.

2 A Theoretical Model

2.1 A Self-Organization Neuron Model. Emergence of various types of self-organization has been studied previously in a unified manner (Amari, 1977, 1983; Amari & Takeuchi, 1978). Our model of the striatum neurons is a new version in that the internal state of a neuron $u(t)$ at time t is enhanced by the reinforcement signal $\alpha(t)$,

$$u(t) = \boldsymbol{w}(t) \cdot \boldsymbol{x}(t)\{1 + \alpha(t)\} - w_0(t)x_0(t), \tag{2.1}$$

where the vector $\boldsymbol{x}$ stands for excitatory cortical inputs to a neuron in the striatum, corresponding to the cue signal, which is one of the four directions, while x_0 summarizes a population of the inhibitory inputs as a single variable in this model and is assumed to be a constant for simplicity in the later analysis. Here, $\boldsymbol{w}$ denotes the weights of the cortico-striatal connections of this neuron for $\boldsymbol{x}$, and w_0 denotes the weight for the inhibitory input x_0.

The term $\alpha(t)$ denotes the effect of the reinforcement signal, carried by dopaminergic (DA) neurons, on the striatal projection neuron firing rates. Experimentally, the DA effects on striatal firing rates have been found to be facilitatory (i.e., $\alpha > 0$ in our model) or inhibitory ($\alpha < 0$). Facilitatory and inhibitory effects may be mediated by D1 and D2 receptors, respectively (Gerfen, 1992; Cepeda, Buchwald, & Levine, 1993). Alternatively, both effects can be mediated by the bistable nature of D1 receptors (Nicola, Surmeier, & Malenka, 2000; Gruber, 2000).

The phasic DA activities occur in general to an unexpected reward or a conditioning stimulus preceding a rewards r, once the conditional relationship is well established (Schultz, Apicella, & Ljungberg, 1993). While this phasic DA activity is hypothesized to provide a reward prediction error, the striatal neurons are considered to change their neural activities, using this reinforcement signal by the DA neurons (Schultz, 1998). Correspondingly, experiments in 1DR and ADR (Kawagoe et al., 1999) have shown that there is the phasic DA activity locked with a rewarded direction cue in each 1DR (ADR) block. When a new 1DR block is started, the phasic DA activity quickly shifts to the rewarded cue within a few trials and stays unchanged throughout the block, possibly having a very small, but negligible, decay toward the end of the block (Kawagoe et al., 1999). Accordingly, we denote the phasic DA activity by α and assume

that α changes without delay between experimental blocks and stays the same throughout a block.

The caudate neurons also change their activities when a 1DR block is altered. Interestingly, the changes in the caudate neural activities are much slower, taking a few tens of trials, than those in the DA activity. This suggests that the plastic caudate activities may be induced by synaptic changes in the cortico-caudal projection, under the DA modulation, rather than by the direct DA-induced changes in the caudate response properties, although the latter effect may also play a role, particularly in initiating such plastic caudate activities (see sections 3.1 and 4).

Let us now look into the time course of the DA activity in a single trial. The phasic DA activity (α) starts to rise roughly around 100 milliseconds after the cue presentation, when it is rewarded. Next, it starts to drop, first rapidly, until around 400 milliseconds and then gradually to the resting level around 700 milliseconds, during which a very weak, but still larger than at the resting level, DA activity exists (Kawagoe et al., 1999; also see Schultz et al., 1993; Schultz, Romo, Ljungberg, Minenowicz, Hollerman, & Dickenson, 1995). To distinguish from the phasic peak activity α, we denote this weak DA activity by α', which can be regarded as nearly, but not exactly, zero (i.e., at the resting level) (therefore, $\alpha' \sim 0$). In contrast, the post-cue caudate neural activities, depending on each neuron, are confined not only during the time of the phasic DA period but also during the time of the weak DA activity (and even later) (Kawagoe et al., 1998, 1999). In other words, the plastic changes in the caudate neural activities possibly carry over not only in the time of α but also in the time of α' at least, suggesting that the DA-modulated synaptic changes induce the plastic changes in the caudate neural activities.

Note that when $\alpha(t) = 0$, the neuron model in Eq. (2.1) is equivalent to the primitive self-organization model (Amari & Takeuchi, 1978). When the reinforcement signal exists (i.e., $|\alpha| > 0$), the internal state of the neuron $u(t)$ increases or decreases, and controls the self-organizing process, so that the reinforcement signal works in a modulatory manner.

The output $y(t)$ of the neuron is given by the transfer function

$$y(t) = f\{u(t)\}. \tag{2.2}$$

To obtain an analytical solution, we first choose $f(\cdot)$ as the step function, given as $\mathbf{1}(u) = 1$ (if $u > 0$), and 0 otherwise. This is the simplest choice, allowing us to derive an explicit analytical solution for the dynamical behavior of learning in the model. This simplification implies that the state of the neuron is binary, that is, the neuron fires ($y = 1$) or does not ($y = 0$). Later, we let $f(\cdot)$ be the sigmoid function $f(a) = 1/(1 + e^{-\gamma a})$ where a is a real value and γ is a scaling parameter.

The dynamics of a Hebbian-type learning rule is chosen to show changes in synaptic efficacies,

$$\begin{cases} \tau \dot{\boldsymbol{w}}(t) = -\boldsymbol{w}(t) + cy(t)\boldsymbol{x}(t) \\ \tau \dot{w}_0(t) = -w_0(t) + c'y(t)x_0, \end{cases} \tag{2.3}$$

where $\dot{\boldsymbol{w}}$ and $\dot{w}_0$ represent the time derivatives $d\boldsymbol{w}/dt$ and dw_0/dt, and c and c' are learning rates. In equation 2.3, $y(t)$ depends on $\alpha(t)$, because $y(t) = f\{u(t)\}$

and $u(t)$ is a function of $x(t)$ and $\alpha(t)$. In other words, DA signals, particularly the phasic ones, initiate and guide the learning process, where the phasic activity (α) quickly changes to the weaker one ($\alpha' \sim 0$) in the time course of a trial. In the first equation, the second term yx is Hebbian, indicating that the weight increases in proportion to the input x when the output y of the neuron is positive. The above model is different from the ordinary Hebb-type neuron in that the inhibitory weight w_0 is also modifiable, as shown in the second equation (Amari & Takeuchi, 1978; Amari, 1983).

The intrinsic mechanism of the model is based on the balance between the excitatory and inhibitory effects, mediated by modifiable weights under the initial modulation of the reinforcement signal. To understand the behavior of models of this kind, we need to examine their dynamical behaviors of learning, typically the equilibrium states and their stability under a stationary environment from which input signals $x(t)$ are supplied. Many stable equilibrium states exist in general, and this multistability is important to allow an ensemble of neurons, exposed to the same environment, to differentiate with different neural responses and capture various features of the environment.

2.2 Analysis of Learning Dynamics. When inputs x_i are presented with probabilities $p_i \equiv p(x_i)$ ($i = 1, 2, 3, 4$) and reinforcement signals $\alpha(x_i)$, which quickly change into $\alpha'(x_i)$, the averaged learning equation is given by

$$\begin{cases} \tau \dot{w} = -w(t) + c \sum_i p_i y_i x_i \\ \tau \dot{w}_0 = -w_0(t) + c' \sum_i p_i y_i x_0. \end{cases} \tag{2.4}$$

The synaptic weights converge to the equilibrium state ($\bar{w}, \bar{w}_0$), satisfying $\dot{w} = \dot{w}_0 = 0$:

$$\begin{cases} \bar{w} = c \sum_i p_i y_i x_i \\ \bar{w}_0 = c' \sum_i p_i y_i x_0. \end{cases} \tag{2.5}$$

Equation 2.5 is not the explicit solution for $\bar{w}$ and $\bar{w}_0$, since y_i on the right-hand side depends on $\bar{w}$ and $\bar{w}_0$, so that it is the equation to be solved for them. After the weights have converged by learning, the internal state of the neuron for input x with accompanying reinforcement signal α (or α') is

$$\bar{u} = \bar{w} \cdot x(1 + \alpha) - \bar{w}_0 x_0. \tag{2.6}$$

If $\bar{u} > 0$ for input x, this neuron fires. In order to analyze the characteristic of this neuron, we define the receptive field of this neuron R as

$$R = \{x \mid \bar{u}(x) > 0\}, \tag{2.7}$$

and using R,

$$p_R = \sum_{x_i \in R} p_i \tag{2.8}$$

$$w_R = \left(\sum_{x_i \in R} p_i x_i \right) / p_R. \tag{2.9}$$

Intuitively speaking, p_R stands for the probability mass of the signals in region R, and $\boldsymbol{w}_R$ stands for the center of gravity in region R. Recall that $y_i = \boldsymbol{1}(u(\boldsymbol{x}_i))$ so that $\sum y_i \boldsymbol{x}_i$ reduces to the sum over all $\boldsymbol{x}$ in the receptive field R. Hence, equation 2.6 can be rewritten as

$$\bar{u}_j = \bar{u}(\boldsymbol{x}_j) = c p_R (1 + \alpha_j) \left(\boldsymbol{w}_R \cdot \boldsymbol{x}_j - \frac{c'}{c(1 + \alpha_j)} x_0^2 \right), \tag{2.10}$$

where $\alpha_j = \alpha(\boldsymbol{x}_j)$. When $\boldsymbol{x}_j$ is rewarded, $\alpha_j = \alpha$ in the beginning and becomes α' later. Therefore, by defining

$$\lambda \equiv \frac{c'}{c} x_0^2, \tag{2.11}$$

the condition for a neuron to fire in response to $\boldsymbol{x}_j$ is given by $K(\boldsymbol{x}_j) > 0$, where

$$K(\boldsymbol{x}_j) \equiv \boldsymbol{w}_R \cdot \boldsymbol{x}_j - \frac{\lambda}{1 + \alpha_j}, \tag{2.12}$$

and the condition to be silent in response to $\boldsymbol{x}_j$ is given by $K(\boldsymbol{x}_j) \leq 0$. The two conditions are the necessary condition but are not sufficient. Note that in the later period of a trial, i.e., when $\alpha_j = \alpha'$, the neuron still fires and $K(\boldsymbol{x}_j) > 0$ for $\alpha_j = \alpha' \approx 0$. Thus, equation 2.12 provides a mathematical criterion for studying the receptive region generated by learning. We can see that λ is the important parameter that controls the size of R. As λ increases, the receptive field becomes smaller. To ensure sufficiency, we need to check

$$\text{for } \forall \boldsymbol{x} \in R, \qquad K(\boldsymbol{x}) > 0, \quad \text{and} \quad \text{for } \forall \boldsymbol{x} \notin R, \qquad K(\boldsymbol{x}) \leq 0.$$

This is because $\boldsymbol{w}_R$ depends on R, and R is defined as the set of inputs $\boldsymbol{x}$ by which the neuron should fire (Amari & Takeuchi, 1978; Amari, 1983).

2.3 Analysis of Caudate Neurons. The emergence of the three response types of neurons is explained here. There are two types of 1DR blocks in the actual experiment: an exclusive 1DR and a relative 1DR (Kawagoe et al., 1998). We discuss only the former here, because our analysis can be easily applied to the latter with slight modifications.

The cortical inputs $\{\boldsymbol{x}_i\}_{i=1}^4$, corresponding to the four cue directions, are given with probability $p(\boldsymbol{x}_i) = \frac{1}{4}$ in the experiment (Kawagoe et al., 1998). In the exclusive 1DR, only one of the four directions is a rewarded direction (say, $\boldsymbol{x}_i$ with the reward $r_i = r$, $r > 0$); the other three directions are not rewarded ($r_j = 0$, $j \neq i$). When one block of experiments starts, we have $\alpha_i = \alpha$ ($|\alpha| > 0$; α becomes α' shortly in each trial) and $\alpha_j = 0$ ($j \neq i$), respectively, as the effect of the phasic reinforcement signal by DA neurons on the striatal projection neuron (see section 2.1).

We now describe how the four cue directions are represented in the cortical signal $\boldsymbol{x}$ to a caudate neuron in our model. The signal consists of a bundle of inputs in which some parts are common to all of the cue directions, just representing the appearance of a cue, and the other parts include specific exclusive

information for each direction. Let M be the number of common inputs, and let N_i be the number of inputs specific to x_i. We rearrange the components such that the common M inputs appear first. We then have the following representation,

$$
\begin{aligned}
x_1 &= (\overbrace{1, \ldots, 1,}^{M} \quad \overbrace{1, \ldots, 1,}^{N_1} \quad \overbrace{0, \ldots, 0,}^{N_2} \quad \overbrace{0, \ldots, 0,}^{N_3} \quad \overbrace{0, \ldots, 0)}^{N_4} \\
x_2 &= (1, \ldots, 1, \quad 0, \ldots, 0, \quad 1, \ldots, 1, \quad 0, \ldots, 0, \quad 0, \ldots, 0) \\
x_3 &= (1, \ldots, 1, \quad 0, \ldots, 0, \quad 0, \ldots, 0, \quad 1, \ldots, 1, \quad 0, \ldots, 0) \\
x_4 &= (1, \ldots, 1, \quad 0, \ldots, 0, \quad 0, \ldots, 0, \quad 0, \ldots, 0, \quad 1, \ldots, 1).
\end{aligned} \tag{2.13}
$$

The inner products among $\{x_i\}_{i=1}^4$ are given by

$$
x_i \cdot x_j = \begin{cases} M + N_i & (i = j) \\ M & (i \neq j). \end{cases} \tag{2.14}
$$

This definition of the cortical representations for each direction is only for presentational simplicity. The following theorems can be proved, as is evident in their proofs, with a general definition of inner products,

$$
x_i \cdot x_j = g_{ij},
$$

if we wish. Through this general form of the inner product, the magnitude of the cortical representation for each direction is determined, and furthermore, the proximity between the cortical representations for different directions is determined. These properties are essential to determine each response type with the other two parameters, λ and α, as shown in the three theorems below, where equation 2.14 is used.

Theorem 1. *A neuron behaves as if the flexible type in all four 1DR blocks if its parameters satisfy*

$$
M + \frac{N_{\max}}{2} \leq \lambda < \min\{M(1 + \alpha), (M + N_{\min})(1 + \alpha')\}, \tag{2.15}
$$

where $N_{\max} = \max_{i \in I} N_i$, $N_{\min} = \min_{i \in I} N_i$ and $I = \{1, 2, 3, 4\}$.

Proof. Without loss of generality, we assume that the rewarded direction is x_1 and, hence, $\alpha_1 = \alpha$ ($\alpha > 0$) in the beginning and $\alpha_j = 0$ ($j \neq 1$). We assume that the initial weight w is large enough such that the neuron is excited by x_1. We search for the condition that $R = \{x_1\}$ is an equilibrium state of the equation, implying that the neuron is excited only by x_1. If this is the case, $p_R = \frac{1}{4}$ and $w_R^1 = x_1$. In equation 2.12, we further require $K(x_1) > 0$ and $K(x_i) \leq 0$ ($i \neq 1$) even when α is reduced to α' in the process. They are equivalently rewritten as

$$
M \leq \lambda < (M + N_1)(1 + \alpha').
$$

Now suppose that the reward direction is changed to x_2 in the next block. We need to obtain the condition to ensure that the neuron changes its behavior

to respond only to x_2. By noting that the initial state of the weight vector of the neuron in the new block is the same as the final equilibrium state of the weights in the previous block, we require $w_R^1 \cdot x_2 - \frac{\lambda}{1+\alpha} > 0$ for the neuron to respond to x_2 in the beginning of the new block, where the reward direction changed and $\alpha_2 = \alpha$. This gives

$$\lambda < M(1 + \alpha).$$

Even when this condition is satisfied, it may happen that this neuron responds to both x_1 and x_2 in the beginning of the second block, though x_1 is no more rewarded, that is, $\alpha_1 = 0$. However, this changes as learning takes place in the new situation. We then need a condition such that while the neuron may initially respond to both x_1 and x_2, the neuron stops responding to x_1 before reaching the equilibrium state of $R = \{x_2\}$. This condition is given by

$$M + \frac{N_2}{2} \leq \lambda.$$

By summarizing the above conditions and taking all 1DR blocks into account, the theorem is proved.

The theorems for the other two types are given as follows.

Theorem 2.   *A neuron behaves as if the conservative type in all four 1DR blocks if its parameters satisfy*

$$M + \frac{N'_{\max}}{2}$$

$$\leq \lambda < \min \left\{ M(1 + \alpha), \left(M + \frac{N'_{\min}}{2} \right)(1 + \alpha'), M + \frac{N_1}{2} \right\}, \qquad (2.16)$$

where we assume that x_1 is the preferred direction (i.e., $N_1 = N_{\max}$) and we have $N'_{\max} = \max_{i \in I'} N_i$, $N'_{\min} = \min_{i \in I'} N_i$ and $I' = \{2, 3, 4\}$.

Theorem 3.   *A neuron behaves as if the reverse type in all four 1DR blocks if its parameters satisfy*

$$\left(M + \frac{N_{\max}}{4} \right)(1 + \alpha') \leq \lambda < M, \qquad (2.17)$$

where $N_{\max} = \max_{i \in I} N_i$ and $\alpha < \alpha' < 0$.

See the appendix for the proofs of theorems 2 and 3.

Figure 3A demonstrates each response type proved in the three theorems above. Note that a binary output neuron is used in the theorems, since the step function, as the transfer function of neurons, allows neurons only to fire or to be silent. Hence, each response type is defined accordingly: Flexible-type neurons respond only for a rewarded direction in each 1DR block; conservative-type
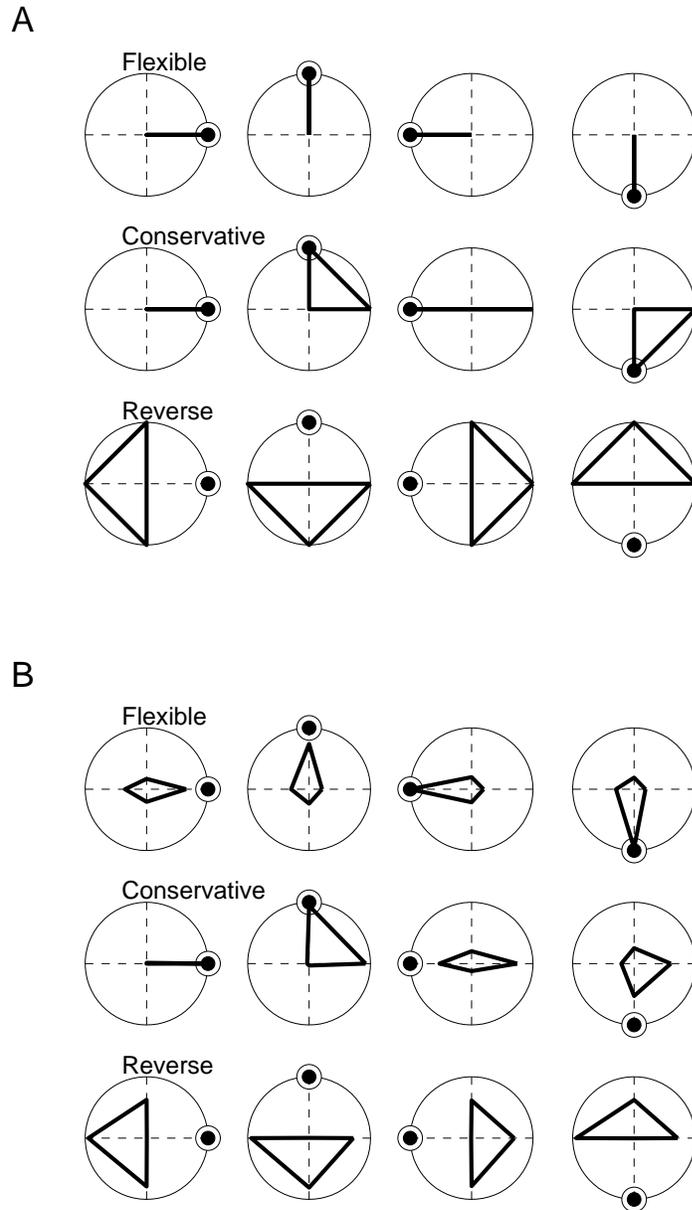
Figure 3: The three types of simulated neural responses shown over four 1DR blocks. The same parameter setting is used in *A* and *B* except transfer functions. (A) Step function is used as the transfer function to generate a neural output. (B) Sigmoid function as the transfer function.

neurons respond for both rewarded and their intrinsic preferred directions in each 1DR block; reverse-type neurons respond for nonrewarded directions in each 1DR block.

The three types proved in the theorems are chosen to stand for the most typical types of neural response changes across 1DR blocks observed in experiments. At the same time, in the rich variety of experimental neural responses, there are some types that are not included in the three types and other types that can be considered as subtypes in one of the three types. For example, a very few neurons behave as if they were reverse-conservative type, which has a larger response to each direction when it is not rewarded, while its intrinsic preferred direction is somewhat maintained in each 1DR block. Other neurons behave as if they were the super-conservative type, which has a response almost only for its intrinsic preferred direction in any 1DR block. It is possible to prove the conditions for these types, but we focus on the typical three types for clarity in this study.

2.4 Fine Characteristics of Neural Responses. The step function allows us to obtain exact analytical solutions; however, simulation results based on the step function differ from experimental ones in that neurons can only be binary mode: to fire or to be silent (Figure 3A). To simulate experimental results further, an analog sigmoid function can be employed, by which a normalized mean firing rate can be represented (Figure 3B). In this alteration of the transfer function, qualitative behaviors are expected to be similar because the sigmoid function becomes similar to the step function as the steepness increases. Once the sigmoid function is used, the difference in the magnitudes of the responses ($y = f(u)$) can be reflected in the difference in the internal states (i.e., u). Hence, by adjusting the proximity of the cortical inputs (i.e., the inner product in equation 2.14 or g_{ij} in general), the fine characteristics in 1DR responses can be represented. In other words, the directional selectivity in the caudate neural responses can be reflected (see Figure 3B; for example, the flexible type is set as having the leftward preferred direction).

For flexible-type neurons in experiments, the response to each direction in ADR tends to be smaller than the corresponding rewarded response in 1DR. This tendency is observed in the model. We first note that the total amount of rewards in one block was the same in ADR and 1DR in experiments (Kawagoe et al., 1998). In other words, the amount of rewards in each reward direction is four times larger in any 1DR block than in ADR. Under this condition, our model shows the observed tendency that neural responses become less distinctive in ADR than in 1DR. Given the step function as the transfer function, the condition for neurons of the flexible type to have responses in ADR as well as responses in 1DR is summarized by

$$M + \frac{N}{2} < \lambda < \min \left\{ M(1 + \alpha), \left(M + \frac{N}{4} \right) \left(1 + \frac{\alpha'}{4} \right) \right\},$$

where we set $N = N_i$ for simplicity. The difference of the internal states in 1DR and ADR ($\bar{u}^{1DR}$ and $\bar{u}^{ADR}$) is given as

$$\bar{u}^{1DR} - \bar{u}^{ADR} = \frac{3c}{16}(N\alpha + 4\lambda - 4M).$$

From the above two equations, we can conclude

$$\bar{u}^{1DR} - \bar{u}^{ADR} > 0,$$

that is, we find the above-mentioned tendency in our model. However, the observed tendency might simply be due to the difference in the amount of rewards per trial. But preliminary experimental results suggest that when a two-directional version of 1DR task is employed, the response amplitude for one direction is influenced by the other chosen direction, indicating an interactive effect between the different directional inputs on the response amplitudes (Takikawa, Kawagoe, & Hikosaka, 2001). Hence, the tendency may not be a simple result of the difference in the amount of reward per a trial.

3 Remarks

3.1 Neuron Model and Learning Rule. The neuron model in this study is given by equations 2.1 and 2.2, while the learning rule is given by equation 2.3. In this formulation, the reinforcement signal is not directly evident in the learning rule but has an indirect effect on the learning through

$$y(t) = f(u) = f[\boldsymbol{w}(t) \cdot \boldsymbol{x}(t)\{1 + \alpha(t)\} - w_0(t)x_0(t)].$$

Accordingly, the DA-induced modulation α influences the neural state u at first. It then works as a modulator, indirectly affecting the learning process of the synaptic weights $\boldsymbol{w}$ and w_0.

DA-induced changes in a neural state (see equation 2.1) eventually lead to selective changes of the synaptic efficacy through the learning process in the model (see equation 2.3). This may illustrate the issue on whether short-latency DA responses are for reinforcement learning or attentional switching (Redgrave, Prescott, & Gurney, 1999). In our model, when a new phasic DA response occurs to an unexpected reward, the phasic DA activity directly affects the striatal neural response properties (attentional switching) and consequently initiates a new learning process ("reinforcement learning").

3.2 Emergence of Three Neural Response Types. The results in the previous section demonstrate that all three types of neuronal behaviors emerge from the same mechanism, depending on the values of the underlying parameters (see Figure 3A). Table 1 summarizes the conditions for each type of neuron. All of the conditions are expressed by the two factors, the term λ and the reinforcement signal α, with respect to the cortical representations (i.e., M, N_i).

The term $\lambda = (c'/c)x_0^2$ (see equation 2.11) is composed of the learning rates for the excitatory and inhibitory weights (i.e., c and c') and the magnitude of inhibitory input x_0, and, roughly speaking, it summarizes the efficacy of learning in the inhibitory effect relative to that in the excitatory effect (see Figures 2B and 2C) (see the next section).

The term α indicates the effect of the reinforcement signal, carried by DA neurons, on the striatal firing rate. As shown in Table 1, the analysis indicates

Table 1: Conditions for Different Types of the Caudate Neurons.

Neuron Type	Condition	
Flexible	$M + \frac{N_{\max}}{2} \leq \lambda < \min\{M(1+\alpha),$	$\alpha \geq \alpha' > 0$
	$(M + N_{\min})(1 + \alpha')\}$	
Conservative	$M + \frac{N'_{\max}}{2} \leq \lambda < \min\{M(1+\alpha),$	$\alpha \geq \alpha' > 0$
	$(M + \frac{N'_{\min}}{2})(1+\alpha'), M + \frac{N_1}{2}\}$	
	$(N_1 > N_i, i \in I')$	
Reverse	$(M + \frac{N_{\max}}{4})(1 + \alpha') \leq \lambda < M$	$\alpha \leq \alpha' < 0$
	$N_{\max} \equiv \max_{i \in I} N_i, \quad N_{\min} \equiv \min_{i \in I} N_i, I \equiv \{1, 2, 3, 4\}$	
	$N'_{\max} \equiv \max_{i \in I'} N_i, \quad N'_{\min} \equiv \min_{i \in I'} N_i, I' \equiv \{2, 3, 4\}$	
	$\lambda \equiv \frac{c'}{c} x_0^2, \quad x_i \cdot x_j \equiv \begin{cases} M + N_i & (i = j) \\ M & (i \neq j) \end{cases}$	

that the reinforcement signal α should work as facilitatory ($\alpha > 0$) for the flexible and conservative types and as inhibitory ($\alpha < 0$) for the reverse type.

In our simplified definition of the inner product of the cortical representations (see equations 2.14 and 2.13), M represents the overlap between the cortical inputs, while N_i represents a part specific to each directional input. More generally, $M + N_i$ corresponds to the square of the magnitude of each direction (i.e., $|x_i|^2$) and M corresponds to $x_i \cdot x_j = |x_i||x_j| \cos \theta_{x_i, x_j}$ ($i \neq j$), where θ_{x_i, x_j} is the angle between x_i and x_j. In Table 1, therefore, the terms such as $M + N_i$, $M + \frac{N_i}{2}$, and so on, express how the proximity in the cortical representations, or their overlap, influences the emergence of each type. For example, a critical difference in the flexible and conservative types is that a preferred cue direction input (x_1 for N_1 in Table 1) satisfies $\lambda < M + \frac{N_1}{2}$ for the conservative type but $M + \frac{N_1}{2} < \lambda$ for the flexible type, while there are some other different conditions between the two types. Now, we can write $M + \frac{N_1}{2} = \frac{1}{2}(\| x_1 \|^2 + x_1 \cdot x_i)$, where $i \in \{2, 3, 4\}$. Hence, the term $M + \frac{N_1}{2}$ is related to both the magnitude of the preferred cue direction input x_1 and the angle between x_1 and other cue direction inputs. When the cortical representations $\{x_i\}$ are chosen as in equation 2.13, the conditions in Table 1 explicitly relate each response type to the magnitudes of the cortical inputs and their overlap, which is essentially the number of cortico-striatal connections reflecting each directional input (see equation 2.13). Since the cortical representations in equation 2.13 and their inner products in equation 2.14 used in Table 1 are only for presentation purposes, in a more general term, the conditions in Table 1 suggest that each response type is determined by the magnitudes and the proximity of the directional inputs conveyed through the cortico-striatal projections (i.e., via the inner product g_{ij}) in relation to λ and α. Our model may be generic enough to capture microscopic properties such as small cluster properties (Hikosaka et al., 1989; Flaherty &

Graybiel, 1994; Jaeger, Kita, & Wilson, 1994; Kincaid, Zheng, & Wilson, 1998) with a more detailed analysis on the cortical input proximities, along with 1DR experiments.

3.3 Effect of Inhibitory Weight Modifiability. The term λ, roughly speaking, summarizes the efficacy of inhibitory learning dynamics. If λ is too small, all neurons start to fire for any directional input at equilibrium, whereas all neurons become silent for any directional input at equilibrium if λ is too large. For an ensemble of neurons to acquire discriminative power by self-organization, λ should be set relevantly so that the neurons can start to respond to different stimuli with differential response properties, governed by their cortical input proximities and reinforcement signals. Thus, while $\lambda = \frac{c'}{c}x_0^2$ is kept constant for each neuron in our simple model, neurons with different values of λ lead to different response types.

As an example, let us consider how a neuron of the flexible type can maintain its response type in transition between 1DR blocks. In the transition, a neuron of the flexible type should start to respond to a new rewarded direction in the beginning of the second block (see the proof of theorem 1). In this situation, the inhibitory effect $\bar{w}_0 x_0$ in a neural state u is given as $\bar{w}_0 x_0 = c' p_R x_0^2 = q p_R \lambda$. Thus, if λ is so large that it makes $u = \bar{w} \cdot x(1 + \alpha) - \bar{w}_0 x_0 < 0$, then the neuron fails to respond to the rewarded direction in the transition. In this sense, λ should be set relatively small (Jaeger et al., 1994). On the other hand, if λ is so small that it allows a neuron to respond to both the previous reward direction and the current reward direction even at its equilibrium, then the neuron fails to maintain the response property of the flexible type after the transition.

Figure 4 shows an example of the correspondence of the three neural response types with different parameter values. In each graph, the regions for the flexible, conservative, and reverse types are drawn with light gray, gray, and dark gray, respectively, and are determined by λ and α for fixed $N_{\max}$ and N (or $N/N_{\max}$) (in other words, we set $N_1 = N_{\max}$, say, and then $N_j = N$ ($j \neq 1$) for simplicity; see the figure legend). In this figure, when $N_{\max}$ is large, there is a small common input M and, when $N/N_{\max}$ becomes smaller, the preferred direction input gets larger relative to the other direction inputs. There is a nonlinear effect to determine the region of each response type. Depending on $N_{\max}$ and N, different values of α and λ provide each response type. We note that the example shown in Figure 4 is just one example, chosen to indicate the coexistence of the three types under some parameter regime. For example, there can be a parameter regime where only the flexible and inhibitory types, but not the conservative type, exist.

The analysis in this study can predict possible changes in the self-organization of neural responses when inhibitory effects are altered. For example, decreasing λ leads neurons to lose their directional-selective responses of the flexible and conservative types (as exemplified in Figure 4). Decreasing λ can be achieved in several ways, for example, by decreasing the inhibitory input (x_0). We wait for experimental examinations on this issue.
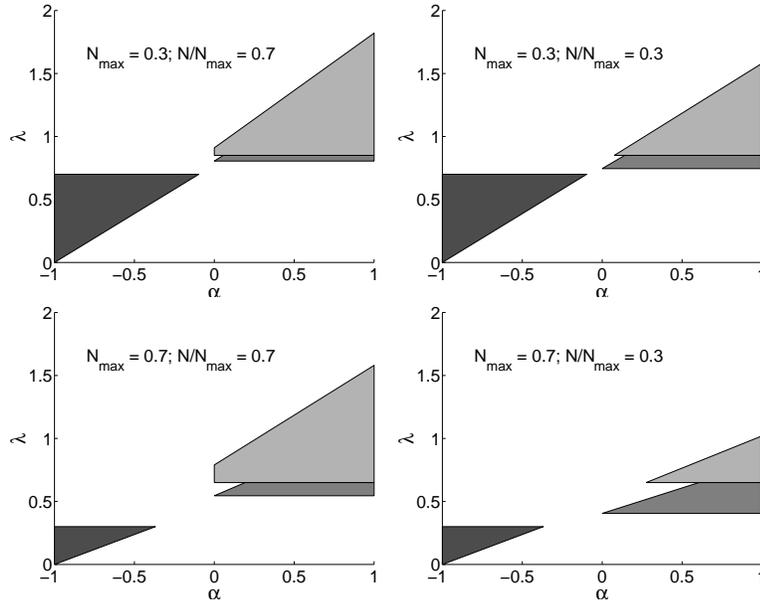
Figure 4: Examples of correspondences of the three neural response types with different parameter regions. Flexible, conservative, and reverse type regions are drawn in light gray, gray, and dark gray, respectively. Graphs are generated by assuming $N_{\max} > N'_{\max} = N'_{\min} = N_{\min} = N$ (in other words, given $N_i = N_{\max}$, $N_j = N$ ($j \neq i$)) and $\alpha = \alpha'$, and normalizing the conditions by $M + N_{\max} = 1$. Hence, we reduced the three conditions in Table 1 to those of the following parameters: $\lambda, \alpha, N_{\max}, N$. In each graph, the regions of each type are drawn with respect to λ and α for fixed $N_{\max}$ and N (or $N/N_{\max}$).

4 Discussion

This study theoretically investigated, as a model of striatum neurons in the basal ganglia, a simple self-organization neuron model under the modulation of reinforcement signals. This model is motivated by experimental observations: directional-selective responses of the striatal neurons, abundant reinforcement signals carried by dopaminergic neurons, a rich source of inhibitory neurons in the striatum, and a vast convergence of the cortico-basal ganglia projection. By a theoretical analysis, the model is shown to explain, in a unified manner, that seemingly different neural response patterns, observed in experiments (Kawagoe et al., 1998), can emerge from the same model with different parameter values. Due to the choice of a simple model, our analysis can explicitly relate each response type with two factors, the reinforcement signal (α) and the inhibitory effect (λ), in conjunction with the magnitudes and the proximity of the cortical input representations (M, N_i).

Various types of self-organization rules have been proposed to investigate mainly the cerebral cortical formation (von der Malsburg, 1973; Amari & Takeuchi, 1978; Willshaw & von der Malsburg, 1979; Bienenstock, Cooper, & Munro, 1982; Obermayer, Ritter, & Schulten, 1990; Földiák, 1991). Experimental support has been largely obtained from the developmental formation of neural circuits in the sensory cortices (Wiesel & Hubel, 1963; LeVay, Stryker, & Shatz, 1978) and from the reorganization of the sensory cortical maps (Buonomano & Merzenich, 1988). The BCM theory in particular has been a very successful model in the developmental formation (Bienenstock et al., 1982; Kirkwood, Rioult, & Bear, 1996). All of these rules use the Hebbian synapse as a mathematical core, although controversy remains in the details. Their differences lie in the way each stabilizes learning such as competition among synapses (von der Malsburg, 1973), floating threshold (Bienenstock et al., 1982), or others.

Different from others, the modifiability of inhibitory synapses (Amari & Takeuchi, 1978) plays a fundamental role in regulating the self-organization in our model (see Figure 2C). Without the reinforcement signal α in equation 2.1, previous studies (Takeuchi & Amari, 1979; Amari, 1980, 1983) have indicated that this inhibitory synapse modifiability, along with the excitatory synapse modifiability, allows efficient input representations, such as a receptive field self-organized with respect to various features of environment inputs, a formation of topographic maps, patch structures, and so on (Takeuchi & Amari, 1979; Amari, 1980, 1983). Hence, the current model with the reinforcement signal α can be expected to construct a cortical input representation effectively modulated by reinforcement signals, although a rigorous study remains to be investigated. This property is useful in the face of the strong anatomical convergence in the cortico-basal projections (Oorschot, 1996). Any divergent manner of the cortico-striatal projections within this convergence (Parthasarathy, Schall, & Graybiel, 1992; Graybiel et al., 1994) may provide further combinatorial benefits for the efficiency in the self-organization. Generally, it is important to construct efficient representations of the state information (inputs) under the modulation of reinforcement signals (Dayan, 1991). A notable experiment has indicated that such a remapping under the influence of a diffused neuromodulator occurs in the sensory cortex (Bakin & Weinberger, 1996; Kilgard & Merzenich, 1998; Sachdev, Lu, Wiley, & Ebner, 1998). Our model can be a primitive model for this issue.

Abundant inhibitory sources exist in the striatum (Kawaguchi et al., 1995). The plasticity of inhibitory synapses is found in the hippocampus (Nusser, Hajos, Somogyi, & Mody, 1998), in the cerebellum (Kano, Rexhausen, Dreessen, & Konnerth, 1992), the cerebral cortex (Komatsu & Iwakiri, 1993; Komatsu, 1996) and other areas (Kano, 1995); recent findings have suggested an important role of inhibitory connections in self-organization even in the cortex, for example, the early visual cortex (Hensch et al., 1998; Fagiolini & Hensch, 2000). To our knowledge, there is no direct evidence of modifiable inhibitory synapses in the striatum, which is an important future study.

We have not specified a neural origin of inhibitory effects. According to experimental findings, a neural origin of inhibitory effects could be either inhibitory interneurons (Jaeger et al., 1994; Bennett & Bolam, 1994; Koós & Tepper, 1999) or the collaterals of projection neurons (Groves, 1983; Wickens, 1993). If

the former is the case, an inhibitory effect may work possibly in a feedforward manner under the influence of cortical projections. If the latter is the case, an inhibitory effect may work as feedback and mutual inhibition and possibly in a manner similar to a winner-take-all mechanism (Groves, 1983; Wickens, 1993). It is possible to extend the current analysis with a recurrent connection. In this perspective, this study can be regarded as an extension of the winner-take-all mechanism. The current analysis of our model, however, treats the effect of modifiable excitatory and inhibitory weights in a feedforward manner, because we do not commit ourselves to specify the inhibitory effect as the collaterals. In this sense, our model shares a feature of reward-modulated feedforward network with previous work by Barto and his colleagues (Barto, Sutton, & Brouwer, 1981; Barto, 1985). A third possibility is that inhibitory interneurons and the collaterals of projection neurons (Groves, 1983; Wickens, 1993) jointly work as an inhibitory source (Wickens, 1997). A recent demonstration of a weak collateral interaction between projection neurons also suggests this possibility (Tunstall, Kean, Wickens, & Oorschot, 2001). In the future, it is important to incorporate our model parameters with the experimentally estimated quantitative nature of these inhibitory resources (Jaeger et al., 1994; Wickens, 1997; Tunstall et al., 2001) as well as of the cortical input magnitudes and proximity (Oorschot, 1996; Kincaid et al., 1998).

Generally, the DA modulation on the caudate neurons can be considered in two aspects: the synaptic efficacy (Calabresi et al., 1992; Wickens & Kötter, 1995) and the response property (Nicola et al., 2000). Caudate response changes are much slower than DA response changes over trials in one 1DR block. In addition, caudate response changes possibly occur beyond the DA phasic response in the time course of a single trial. Hence, we considered that the DA-modulated synaptic efficacy is involved in the emergence of the three response types and investigated how the DA-modulated synaptic plasticity can lead to these response types, while we considered the DA effect on the caudate firing rates to be complementary (see section 3.1). Yet this DA effect on the firing rates may play a larger role in accounting for the three response types (Gruber, 2000). Enhanced caudate activities long after the DA phasic response can be, partially at least, due to a prolonged DA effect that possibly sustains longer than the period of the phasic DA response (Gonon, 1997; Durstewitz, Seamans, & Sejnowski, 2000); The combined effect of D1 and D2 receptors on the caudate firing rates may further help shape the nature of the three response types (Gerfen, 1992; Cepeda et al., 1993; Nicola et al., 2000). How the two DA modulations, on the synaptic efficacy and on the response property, are integrated remains to be investigated. The dynamical aspect of the interaction between these modulations, possibly with the regulation of different DA receptors, is of particular interest because their timescales presumably are different.

Finally, in our model, we treated only the current reinforcement signal, not delayed ones. To enjoy the full power of reinforcement learning, it is important to extend our model to include delayed-reinforcement signals (Barto, 1995; Houk, Adams, & Barto, 1995; Montague, Dayan, & Sejnowski, 1996; Berns & Sejnowski, 1998; Schultz, Dayan, & Montague, 1997; Trappenberg, Nakahara, & Hikosaka, 1998; Nakahara, Trappenberg, Hikosaka, Kawagoe, & Takikawa, 1998; Monchi

& Taylor, 1999; Hikosaka et al., 1999). For example, it is possible to maintain the same response property in our model even when the magnitude of α is reduced if the receptive region $\mathcal{R}$ stays the same with $K(x) > 0$ for $x \in \mathcal{R}$ and with $K(x) < 0$ for $x \notin \mathcal{R}$ (see section 2.2). This kind of property is useful in transferring an effect of a delayed-reinforcement signal to a preceding stimulus. We are currently investigating this issue.

Appendix

A.1 Proof of Theorem 2. Recall that the order of the 1DR blocks is randomized in the experiment. Provided that the direction of x_1 is the most preferred direction of the neuron, we need to consider how the behaviors of a neuron change in the three cases of transition of reward directions over 1DR blocks: (1) from the 1DR block where the reward direction is x_1 to another, say x_2, (2) from the 1DR block where the reward direction is x_j ($j = 2, 3, 4$), say x_2, to another 1DR block where the reward direction is not x_1, say, x_3, and (3) from the 1DR block where the reward direction is not x_1, say, x_3, to the 1DR block where the reward direction is x_1. Note also that similar to the proof of theorem 1, we assume that the neuron is excited by x_1 initially.

We begin with the condition guaranteeing that the neuron is responsive to only x_1 at the equilibrium. In this case, the receptive region is $R_1 = \{x_1\}$ and hence $p_{R_1} = \frac{1}{4}$ and $w_R^1 = x_1$. In equation 2.12, we require $K(x_1) > 0$ and $K(x_i) \leq 0$ ($i \neq 1$), which yield

$$M \leq \lambda < (M + N_1)(1 + \alpha').$$

In the equilibrium states of the other three 1DR blocks (here, we treat the case where x_2 is the reward direction), the receptive region should be $R_2 = \{x_1, x_2\}$. In this case, $p_{R_2} = \frac{1}{2}$ and $w_R^2 = \frac{1}{2}(x_1 + x_1)$. We require $K(x_i) > 0$ ($i = 1, 2$) and $K(x_i) \leq 0$ ($i = 3, 4$), which leads to the following condition:

$$M \leq \lambda < \min \left\{ M + \frac{1}{2}N_1, \left(M + \frac{1}{2}N_2 \right)(1 + \alpha') \right\}.$$

As for the condition guaranteeing the transition from $R = \{x_1\}$ to $R = \{x_1, x_2\}$ as the reward direction changes, the neuron should respond to x_2 in addition to x_1 as the reinforcement signal is accompanied by x_2. Hence, we need to have $w_1^R \cdot x_2 - \frac{\lambda}{1+\alpha} > 0$, which is equivalent to

$$\lambda < M(1 + \alpha).$$

For case 2, the neuron has to respond to x_3 in the beginning of the next block, that is, $w_2^R \cdot x_3 - \frac{\lambda}{1+\alpha} > 0$, which is equivalent to $\lambda < M(1 + \alpha)$. When this condition is satisfied, the neuron responds to all of x_1, x_2, x_3 when a new 1DR block starts. However, $R_3 = \{x_1, x_2, x_3\}$ is not stable under a certain condition, as stated in the following. In this case, competition should occur among these three inputs to let the neural response to the nonrewarded input x_2 be silent. For $R_3 = \{x_1, x_2, x_3\}$, $p_{R_3} = \frac{3}{4}$ and $w_R^3 = \frac{1}{3}(x_1 + x_2 + x_3)$. Therefore, the response

to x_2 becomes silent when $w_R^3 \cdot x_2 - \lambda \leq 0$, or equivalently,

$$M + \frac{1}{3}N_2 \leq \lambda.$$

In case 3, the neuron responds to both x_1 and x_3 in the first block, and in the next block, the neuron changes to respond only to x_1. To ensure this requirement, we impose the condition under which the equilibria of $R_4 = \{x_1, x_3\}$ in the next block are not stable, which is written as

$$M + \frac{1}{2}N_3 \leq \lambda.$$

By summing up all these conditions and taking all combinations of 1DR blocks into account, we get leads to

$$M + \frac{N'_{max}}{2} \leq \lambda < \min \left\{ M(1 + \alpha), \left(M + \frac{N'_{min}}{2} \right)(1 + \alpha'), M + \frac{N_1}{2} \right\},$$

where $N'_{max} = \max_{i \in I'} N_i$, $N'_{min} = \min_{i \in I'} N_i$, and $I' = \{2, 3, 4\}$.

A.2 Proof of Theorem 3. It is sufficient to consider the condition for equilibrium states in one 1DR block and for transition from one block to another block. Suppose that the reward direction is x_1. A reverse-type neuron should have $R = \{x_2, x_3, x_4\}$ as the equilibrium. This implies $p_R = \frac{3}{4}$ and $w_R^1 = \frac{1}{3}(x_2 + x_3 + x_4)$. In equation 2.12, we further require $K(x_1) \leq 0$ and $K(x_i) > 0$ $(i \neq 1)$, which are equivalently rewritten together as

$$M(1 + \alpha') \leq \lambda < M + \frac{1}{3}N_i, \text{ where } i = 2, 3, 4.$$

At the start of the next block (where the reward direction is, say, x_2), we first require that the neuron starts to respond to x_1 to which the reinforcement signal α is no longer given. This requires $w_R^1 \cdot x_1 - \lambda > 0$, or equivalently,

$$\lambda < M.$$

Provided that this condition is satisfied, the neuron responds to all of the directions in this block. Hence, we impose another condition under which $R = \{x_1, x_2, x_3, x_4\}$ is not a stable equilibrium so that the neuron stops firing to the input x_2. This condition can be rewritten, given $p_R = 1$ and $w_R^1 = \frac{1}{4}(x_1 + x_2 + x_3 + x_4)$, as

$$w_R \cdot x_2 - \frac{1}{1 + \alpha'}\lambda \leq 0,$$

or equivalently,

$$\left(M + \frac{1}{4}N_2 \right)(1 + \alpha') \leq \lambda.$$

Since the equilibrium condition of this block is the same as that of the first block, summing up the above conditions leads to the theorem.

Acknowledgments

References

Alexander, G. E., Crutcher, M. D., & DeLong, M. R. (1990). Basal ganglia-thalamocortical circuits: Parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. In H. B. M. Uylings, C. G. Van Eden, J. P. C. De Bruin, M. A. Corner, & M. G. P. Feenstra (Eds.), *Progress in brain research* (Vol. 85, pp. 119–146). Amsterdam: Elsevier.

Amari, S. (1977). Neural theory of association and concept-formation. *Biological Cybernetics, 26*, 175–185.

Amari, S. (1980). Topographic organization of nerve fields. *Bulletin of Mathematical Biology, 42*, 339–364.

Amari, S. (1983). Field theory of self-organizing neural nets. *IEEE Transaction on Systems, Man and Cybernetics, SMC-13*(9 & 10), 741–748.

Amari, S., & Takeuchi, A. (1978). Mathematical theory on formation of category detecting nerve cells. *Biological Cybernetics, 29*, 127–136.

Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M., & Kimura, M. (1994). Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *J. Neurosci., 6*, 3969–3984.

Bakin, J. S., & Weinberger, N. M. (1996). Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proceedings of the National Academy of Sciences, 93*, 11219–11224.

Barto, A. G. (1985). Learning by statistical cooperation of self-interested neuron-like computing elements. *Human Neurobiology, 4*, 229–256.

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 215–232). Cambridge, MA: MIT Press.

Barto, A. G., Sutton, R. S., & Brouwer, P. S. (1981). Associative search network: A reinforcement learning associative memory. *Biological Cybernetics, 40*, 201–211.

Bennett, B. D., & Bolam, J. P. (1994). Synaptic input and output of parbalbumin-immunoreactive neurons in the neostriatum of the rat. *Neuroscience, 62*, 707–719.

Berns, G. S., & Sejnowski, T. J. (1998). A computational model of how the basal ganglia produce sequences. *Journal of Cognitive Neuroscience, 10*(1), 108–121.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience, 2*, 32–48.

Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience, 21*, 149–186.

Calabresi, P., Maj, R., Pisani, A., Mercuri, N. B., & Bernardi, G. (1992). Long-term synaptic depression in the striatum: Physiological and pharmacological characterization. *Journal of Neuroscience, 12*, 4224–4233.

Cepeda, C., Buchwald, N. A., & Levine, M. S. (1993). Neuromodulatory actions of dopamine in the neostriatum are dependent upon the excitatory amino acid receptor subtypes activated. *Proceedings of the National Academy of Sciences of the United States of America, 90*, 9576–9580.

Dayan, P. (1991). Navigating through temporal difference. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in neural information processing systems, 3* (pp. 464–470). San Mateo, CA: Morgan Kaufmann.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *Journal of Neurophysiology, 83*, 1733–1750.

Fagiolini, M., & Hensch, T. K. (2000). Inhibitory threshold for critical-period activation in primary visual cortex. *Nature, 404*(6774), 183–186.

Flaherty, A. W., & Graybiel, A. M. (1994). Input-output organization of the sensorimotor striatum in the squirrel monkey. *J. Neurosci., 2*, 599–610.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation, 3*, 194–200.

Gerfen, C. R. (1992). The neostriatal mosaic: Multiple levels of compartmental organization. *Trends in Neurosciences, 15*(4), 133–139.

Gonon, F. (1997). Prolonged and extrasynaptic excitatory action of dopamine mediated by D1 receptors in the rat striatum in vivo. *Journal of Neuroscience, 17*, 5972–5978.

Graybiel, A. M. (1995). Building action repertoires: Memory and learning functions of the basal ganglia. *Current Opinion in Neurobiology, 5*, 733–741.

Graybiel, A. M., Aosaki, T., Flaherty, A., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science, 265*, 1826–1831.

Groves, P. M. (1983). A theory of the functional organization of the neostriatum and the neostriatal control of voluntary movement. *Brain Research, 286*(2), 109–132.

Gruber, A. (2000). *A computational study of D1 induced modulation of medium spiny neuron response properties*. Unpublished master's thesis, Northwestern University.

Hensch, T. K., Fagiolini, M., Mataga, N., Stryker, M. P., Baekkeskov, S., & Kash, S. F. (1998). Local GABA circuit control of experience-dependent plasticity in developing visual cortex. *Science, 282*(5393), 1504–1508.

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., Miyachi, S., & Doya, K. (1999). Parallel neural networks for learning sequential procedures. *Trends in Neuroscience, 22*(10), 464–471.

Hikosaka, O., Sakamoto, M., & Usui, S. (1989). Functional properties of monkey caudate neurons. II. visual and auditory responses. *Journal of Neurophysiology, 61*, 799–813.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge, MA: MIT Press.

Jaeger, D., Kita, H., & Wilson, C. J. (1994). Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. *Journal of Neurophysiology, 72*(5), 2555–2558.

Kano, M. (1995). Plasticity of inhibitory synapses in the brain: A possible memory mechanism that has been overlooked. *Neuroscience Research, 21*, 177–182.

Kano, M., Rexhausen, U., Dreessen, J., & Konnerth, A. (1992). Synaptic excitation produces a long-lasting rebound potentiation of inhibitory synaptic signals in cerebellar Purkinje cells. *Nature, 356*, 601–604.

Kawagoe, R., Takikawa, Y., & Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nature Neuroscience, 1*(5), 411–416.

Kawagoe, R., Takikawa, Y., & Hikosaka, O. (1999). Change in reward-predicting activity of monkey dopamine neurons: Short-term plasticity. *Society for Neuroscience Abstracts, 25*, 1162.

Kawaguchi, Y., Wilson, C. J., Augood, S. J., & Emson, P. C. (1995). Striatal interneurones: Chemical, physiological and morphological characterization. *Trends in Neurosciences, 18*(12), 527–535.

Kilgard, M. P., & Merzenich, M. M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science, 279*, 1714–1718.

Kincaid, A. E., Zheng, T., & Wilson, C. J. (1998). Connectivity and convergence of single corticostriatal axons. *Journal of Neuroscience, 18*(12), 4722–4731.

Kirkwood, A., Rioult, M. C., & Bear, M. F. (1996). Experience-dependent modification of synaptic plasticity in visual cortex. *Nature, 381*(6582), 526–528.

Kita, K. (1993). GABAergic circuits of the striatum. In G. W. Arbuthnott & P. C. Emson (Eds.), *Chemical signalling in the basal ganglia* (pp. 51–72). Amsterdam: Elsevier.

Knopman, D., & Nissen, M. J. (1991). Procedural learning is impaired in Huntington's disease: Evidence from the serial reaction time task. *Neuropsychologia, 29*(3), 245–254.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*, 1399–1402.

Komatsu, Y. (1996). $GABA_b$ receptors, monoamine receptors, and postsynaptic inositol trisphosphate-induced CA2+ release are involved in the induction of long-term potentiation at visual cortical inhibitory synapses. *Journal of Neuroscience, 16*(20), 6342–6352.

Komatsu, Y., & Iwakiri, M. (1993). Long-term modification of inhibitory synaptic transmission in developing visual cortex. *NeuroReport, 4*(7), 907–910.

Koós, T., & Tepper, J. M. (1999). Inhibitory control of neostriatal projection neurons by GABAergic interneurons. *Nature Neuroscience, 2*(5), 467–472.

LeVay, S., Stryker, M. P., & Shatz, C. J. (1978). Ocular dominance columns and their development in layer IV of the cat's visual cortex: A quantitative study. *Journal of Comparative Neurology, 179*(1), 223–244.

Marsden, C. D. (1980). The enigma of the basal ganglia and movement. *Trends in Neuroscience*, pp. 284–287.

Monchi, O., & Taylor, J. G. (1999). A hard wired model of coupled frontal working memories for various tasks. *Information Sciences, 113*(3), 221–243.

Montague, R., Dayan, P., & Sejnowski, T. J. (1996). Framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16*, 1936–1947.

Nakahara, H., Doya, K., & Hikosaka, O. (2001). Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuo-motor sequences—a computational approach. *Journal of Cognitive Neuroscience*, *13: 5*, 626–647.

Nakahara, H., Trappenberg, T., Hikosaka, O., Kawagoe, R., & Takikawa, Y. (1998). Computational analysis on reward-modulated activities of caudate neurons. *Society for Neuroscience Abstracts, 28*, 1651.

Nicola, S. M., Surmeier, J., & Malenka, R. C. (2000). Dopaminergic modulation of neuronal excitability in the striatum and nucleus accumbens. *Annual Review of Neuroscience, 23*, 185–215.

Nusser, Z., Hajos, N., Somogyi, P., & Mody, I. (1998). Increased number of synaptic GABA(A) receptors underlies potentiation at hippocampal inhibitory synapses. *Nature, 395*(6698), 172–177.

Obermayer, K., Ritter, H., & Schulten, K. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proceedings of the National Academy of Sciences, 87*(21), 8345–8349.

Oorschot, D. E. (1996). Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: A stereological study using the cavalieri and optical disector methods. *Journal of Comparative Neurology, 366*, 580–599.

Parthasarathy, H. B., Schall, J. D., & Graybiel, A. M. (1992). Distributed but convergent ordering of corticostriatal projections: Analysis of the frontal eye field and the supplementary eye field in the macaque monkey. *Journal of Neuroscience, 12*, 4468–4488.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error? *Trends in Neuroscience, 22*(4), 146–151.

Reynolds, J. N. J., & Wickens, J. R. (2000). Substantia nigra dopamine regulates synaptic plasticity and membrane potential fluctuations in the rat neostriatum, in vivo. *Neuroscience, 99*, 199–203.

Sachdev, R. N. S., Lu, S.-M., Wiley, R. G., & Ebner, F. F. (1998). Role of the basal forebrain cholinergic projection in somatosensory cortical plasticity. *Journal of Neurophysiology, 79*, 3216–3228.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1–27.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience, 13*(3), 900–913.

Schultz, W., Apicella, P., Romo, R., & Scarnati, E. (1995). Context-dependent activity in primate striatum reflecting past and future behavioral events. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 11–27). Cambridge, MA: MIT Press.

Schultz, W., Dayan, P., & Montague, R. (1997). A neural substrate of prediction and reward. *Science, 275*, 1593–1599.

Schultz, W., Romo, R., Ljungberg, T., Mirenowicz, J., Hollerman, J. R., & Dickinson, A. (1995). Reward-related signals carried by dopamine neurons. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 231–248). Cambridge, MA: MIT Press.

Surmeier, D. J., Song, W.-J., & Yan, Z. (1996). Coordinated expression of dopamine receptors in neostriatal medium spiny neurons. *Journal of Neuroscience, 16*, 6579–6591.

Takeuchi, A., & Amari, S. (1979). Formation of topographic maps and columnar microstructures. *Biological Cybernetics, 35*, 63–72.

Takikawa, Y., Kawagoe, R., & Hikosaka (2001). *Reward-dependent spatial selection of anticipatory activity in monkey caudate neurons*. Manuscript submitted for publication.

Trappenberg, T., Nakahara, H., & Hikosaka, H. (1998). Modeling reward dependent activity pattern of caudate neurons. In *International Conference on Artificial Neural Network (ICANN98)* (pp. 973–978). Skövde, Sweden.

Tunstall, M. J., Kean, A., Wickens, J. R., & Oorschot, D. (2001). Inhibitory interaction between spiny projection neurons of the striatum: a physiological and anatomical study. In *Abstracts for International Basal Ganglia Society VIIth International Triennial Meeting* (p. 38). Waitangi, New Zealand.

von der Malsburg, C. (1973). Self-organization of orientation selective cells in the striate cortex. *Kybernetik, 14*, 85–100.

Wickens, J. (1993). *A theory of the striatum*. Oxford: Pergamon Press.

Wickens, J. (1997). Basal ganglia: Structure and computations. *Network: Computation in Neural Systems, 8*, R77–R109.

Wickens, J., & Kötter, R. (1995). Cellular models of reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 187–214). Cambridge, MA: MIT Press.

Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology, 26*, 1003–1017.

Willshaw, D. J., & von der Malsburg, C. (1979). A marker induction mechanism for the establishment of ordered neural mappings: Its application to the retinotectal problem. *Philos. Trans. R. Soc. Lond. B. Biol. Sci., 287*(1021), 203–243.

Wilson, C. J. (1998). Basal ganglia. In G. M. Shepherd (Ed.), *The synaptic organization of the brain* (4th ed., pp. 279–316). Oxford: Oxford University Press.