NAIST-IS-DD1461011

Doctoral Dissertation

A Physiological Study of the Striatum Based on Reinforcement Learning Models of the Basal Ganglia

Tomohiko Yoshizawa

March 15, 2018

Graduate School of Information Science
Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of SCIENCE

Tomohiko Yoshizawa

Thesis Committee:

Professor Kazushi Ikeda	(Supervisor)
Professor Shigehiko Kanaya	(Co-supervisor)
Associate Professor Junichiro Yoshimoto	(Co-supervisor)
Professor Kenji Doya	(OIST)

"Scientia et potentia humana in idem coincidunt, quia ignoratio causae destituit effectum."

-Francis Bacon

A Physiological Study of the Striatum Based on Reinforcement Learning Models of the Basal Ganglia*

Tomohiko Yoshizawa

Abstract

The reinforcement learning models of the basal ganglia assume that the striatum plays roles of value functions for reward prediction. In this thesis, I designed rodent's behavioral tasks based on the reinforcement learning theory and physiologically examined the roles of the striatum for decision making or learning by recording their neural activities during the task.

The striatum consists of striosome and matrix compartments. It is hypothesized that the striosomes and matrix perform the roles of reward prediction (critic) and action selection (actor), respectively. Using a selective *in vivo* calcium imaging method of striosomal neurons for mice, I recorded striosomal neural activities from the dorsomedial striatum during an odor-classical conditioning task and found that some striosomal neurons responded to odors associated with rewards. The amplitude was proportional to expected reward size. These activities encode values of odors, therefore suggest that the striosomes perform the role of critic.

The striatum is not only involved in decision making but also motor control. However, previous studies focused on them independently. To investigate parallel neural representations of information related to decision making and motor control at the single neuron level, I recorded electrophysiologically rat's neural activities during a free-choice task and at the same time captured rat's motions using a motion tracking system, then analyzed neural representations of task-, space-, and motor-related variables in the dorsomedial and dorsolateral striatum.

*Doctoral Dissertation, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1461011, March 15, 2018.

The results of regression analysis of neural activities indicated that majorities of striatal neurons represented these variables in parallel. This suggests that the striatum is also involved in motor control during decision making.

Keywords:

Reinforcement learning, Basal ganglia, Striatum, Striosome, Reward prediction

大脳基底核の強化学習モデルに基づく線条体の生理学的研究*

吉澤 知彦

内容梗概

大脳基底核の強化学習モデルは、価値関数に基づく報酬予測の機能を線条体に仮定してきた。本研究では、学習や意思決定に果たす線条体の生理学的な役割を強化学習との対比で理解することを目的に、行動課題中のげっ歯類線条体ニューロンの活動計測実験を行い、ニューロンの情報表現を検証した。

線条体は全体の約 15%を占める striosome と約 85%を占める matrix と呼ばれる 2 種類のコンパートメントから構成されており、それぞれが報酬予測と行動選択に関与すると想定されている。マウス striosome ニューロンの神経活動を選択的に記録できる脳深部 *in vivo* カルシウムイメージング手法を用いて、匂いと報酬による古典的条件付け時の背内側線条体の striosome ニューロンの活動を観察したところ、一部の striosome ニューロンは報酬に先行する匂い刺激に対して、期待される報酬量に比例した応答をしていた。さらに、matrix ニューロンの近似として線条体ニューロンをコンパートメント非選択的に記録した場合には、striosome ニューロン選択的に記録した場合と比較して、報酬予測的な活動を観察する頻度が低くなった。これは striosome ニューロンが匂いの価値をコードすることを示すとともに、striosome が matrix よりも強く報酬予測に関与することを示唆する。

また、線条体は意思決定のみならず、運動制御にも関与している。意思決定、運動制御に関わる情報の単一ニューロンにおける表現を明らかにするため、自由選択課題中のラットの線条体神経活動と身体の動きを電気生理学的手法とモーショントラッキング手法により同時記録し、課題に関する情報、空間情報、運動情報の背内側・背外側線条体における神経表現を検証した。その結果、大部分の線条体ニューロンは課題・空間・運動の情報を同時に表現していた。一方で、背内側・

*奈良先端科学技術大学院大学 情報科学研究科 博士論文, NAIST-IS-DD1461011, 2018 年 3 月 15 日.

背外側線条体で表現に差は認められなかった。この結果は、線条体ニューロンは意思決定に関わる処理と運動制御に関わる処理を並列的に行うことを示唆する。

キーワード

強化学習, 大脳基底核, 線条体, ストリオソーム, 報酬予測

Contents

List of Figures

1. General Introduction

The striatum is a major input site of the basal ganglia, which has been hypothesized as a brain region for prediction of future rewards as state or action values of reinforcement learning [1, 2]. This thesis aims to clarify physiological functions of the striatum for learning and decision making from the sight of reinforcement learning. In this chapter, I firstly introduce animal behavioral learning from psychological view. After that, I explain the reinforcement learning theory, and the anatomy and the physiology of neural circuits of the basal ganglia. Combined these knowledge, I describe reinforcement learning models of the basal ganglia.

1.1 Reward-based behavioral learning of animals

In the psychology field, learning is broadly classified into two groups, which are the non-associative learning and the associative learning. The non-associative learning including the habituation and the sensitization is the learning of a stimulus itself when subjects receive it once or repeatedly, whereas the associative learning, which is classified classical conditioning and operant (instrumental) conditioning, is the learning of relations between stimuli.

The classical conditioning is also called the Pavlovian conditioning because its property was found by Pavlov using dogs (Figure 1.1). When you give a hungry dog a feed, the dog will flow saliva. This phenomenon is an unconditioned response (UR) called the salivary reflex. The feed is an unconditioned stimulus (US). Next, you make the dog hear sound of bell before giving a feed. After repeating this procedure, the dog will become to flow saliva only hearing the sound of bell. Here, the sound of bell and the salivation are called a conditioned stimulus (CS) and a conditioned response (CR), respectively. In the classical conditioning, animals associate the conditioned stimulus with the unconditioned stimulus.

However, behaviors of humans or other animals are not only acquired by learning associations between a stimulus and another stimulus. They learn behaviors adaptively by observing transition of their environment as the result of their previous behaviors. This learning process is called an operant (instrumental) conditioning. For example, when rats are putted on an experimental box in

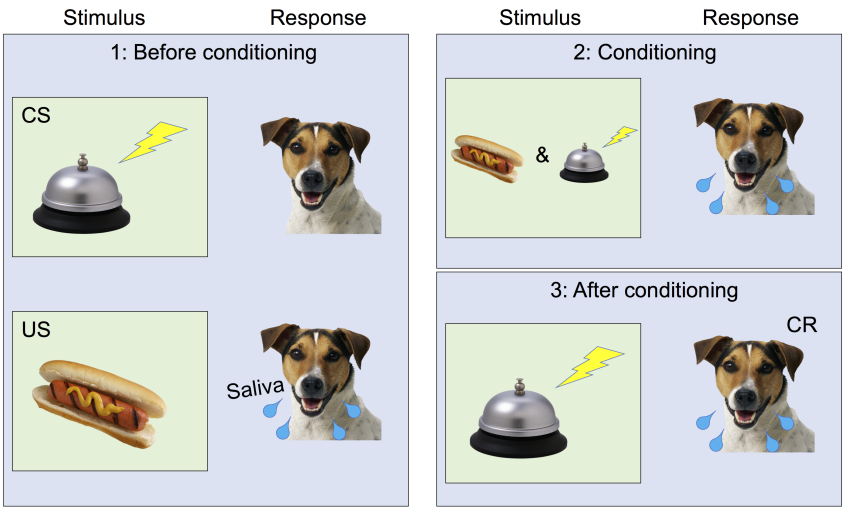

Figure 1.1: Classical conditioning. The dog does not respond to a conditioned stimulus (CS), which is the sound of bell in this case, before conditioning, whereas he shows saliva reflex to an unconditioned stimulus (US), which is visual stimulus of a food. To condition him, we need to present the sound of bell and the food at the same time. After conditioning, he becomes to show saliva reflex only when he hears the sound of bell by learning the association between the sound and food.

which they will get a feed by pushing a lever in the box, they firstly explore in the box, then incidentally push the lever and get a feed (reinforcer). As the result, the frequency of lever pushing after CS gradually increases (reinforce). When a certain action triggers a good result, such as reward, animals become to select the same action more frequently under the same situation. Thorndike, a psychologist in 19-20th century, named this phenomenon "law of effect". The "law of effect" expresses the basic mechanism of the reinforcement learning theory using natural language.

1.2 Reinforcement learning

Our learning processes are conducted through interactions with environments, as the operant conditioning. The reinforcement learning (RL) is an approach

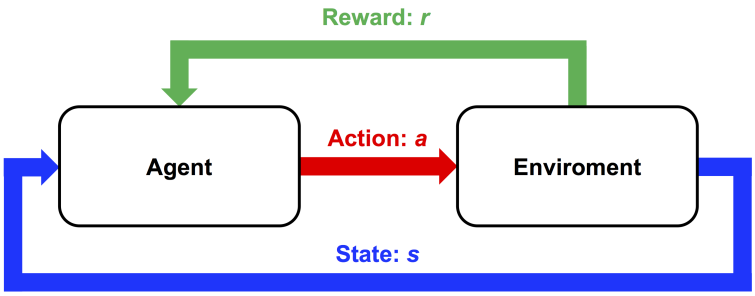

Figure 1.2: Architecture of the reinforcement learning. An agent observes an environment, then takes one of possible actions. Depending on agent's action, it can get a reward.

focusing on goal-directed learning based on such interactions. An agent in the RL learns what should do to maximize digitized reward signal. To achieve this goal, an agent need to find an action in which it can earn more rewards through try and error because it does not be taught the best action like the supervised learning. As illustrated in Figure 1.2, the RL assumes a situation that an agent (animal, human, robot, program etc.) monitors a state of environment (s), choices an action (a) and then receives a reward (r).

1.2.1 Elements of the RL

The RL consists of three major elements, those are a policy, reward function and value function. The policy is a mapping from a state, which an agent perceives in an environment, to an action that it should select. It corresponds to stimulus-response rules or associations in psychology. The reward function maps the state on a single number, a reward. The reward thus defines what are the good and bad events for the agent. A goal of the agent is to maximize the total reward it receives over the long run. The value function is the total amount of reward an agent can expect to accumulate over the future, starting from a state. Whereas rewards determine the immediate, intrinsic desirability of environmental states, values indicate the long-term desirability of states after taking into account the

states that are likely to follow, and the rewards available in those states. To select an action leading to the maximum total rewards, an agent tries to find the action resulting to a state with the highest value rather than immediate reward. In fact, the most important component of almost all reinforcement learning algorithms is a method for efficiently estimating values.

1.2.2 State and Action values

The state and action values are defined as functions of a state and of a state-action pair, respectively. The state value evaluates goodness of a state, in which an agent is. The action value evaluates goodness of an action that an agent takes in a state. The goodness is assessed based on expected future rewards. The value functions are defined in terms of a policy π because the future rewards depend on actions selected by an agent. The policy π is a probability $p(a|s)$ on which an agent takes an action a under a state s. The value of a state s under a policy π, denoted $V^\pi(s)$, is the expected return when starting in s and following πthereafter. $V^\pi(s)$ is defined formally as

$$V^\pi(s) = E_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s \right\} \tag{1}$$

where $E_\pi \{\}$ denotes the expected value of a random variable given that the agent follows policy π, and t is any time step. The function $V\pi(s)$ is called the state-value function for policy π. Similarly, the value of taking an action a in a state s under a policy π, denoted $Q^\pi(s, a)$, as the expected return starting from s, taking the action a, and thereafter following policy π. $Q^\pi(s, a)$ is defined formally as

$$Q^\pi(s, a) = E_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a \right\} \tag{2}$$

1.2.3 Actor-Critic method and TD error

The Actor-Critic method is one of the reinforcement learning algorithms. The Actor plays a role of action selection based on current estimated value functions. The Critic evaluates goodness of a transition state after taking an action and updates value functions. The result of evaluation takes a form of TD error as

following;

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \tag{3}$$

Using this TD error as a learning signal, a state value can be updated with satisfying the formula (1) as following;

$$V(s_t) \leftarrow V(s_t) + \alpha\delta_t \tag{4}$$

The TD error is also used for evaluation of a selected action. If the TD error takes a positive value, the action becomes to be taken more frequently, since the action increases the value of selected action. On the other hand, if the TD error takes a negative value, the action becomes to be taken less frequently. The action selection is expressed as following;

$$\pi_t(s, a) = p(a_t = a|s_t = s) = \frac{e^{p(s,a)}}{\sum_b e^{p(s,a)}} \tag{5}$$

Then, an update of the actor is as following;

$$p(s_t, a_t) \leftarrow p(s_t, a_t) + \beta\delta_t \tag{6}$$

Here, β is a positive step-size parameter.

1.3 RL in animals

1.3.1 Anatomical structures and circuits of the basal ganglia

A brain is anatomically divided into the brainstem, cerebellum, diencephalon and cerebrum. The cerebrum is divided into the cortex and basal ganglia consisting of the striatum, globus pallidus, subthalamic nucleus and substantia nigra (Figure 1.3). The striatum receives glutamatergic and dopaminergic inputs from various cortex areas and the substantia nigra pars compacta (SNc) respectively, and sends inhibitory projections to the internal globus pallidus (GPi) and substantia nigra pars reticulate (SNr), which are the major output site of the basal ganglia, resulting in a net disinhibition or excitation of the thalamus. This pathway is called the direct pathway. On the other hand, the indirect pathway originates in the striatum and inhibits the external globus pallidus (GPe), resulting in disinhibition of the GPi which is then free to inhibit the thalamus. Therefore, loop circuits called the cortico-basal ganglia loop are formed between the cortex and basal ganglia.

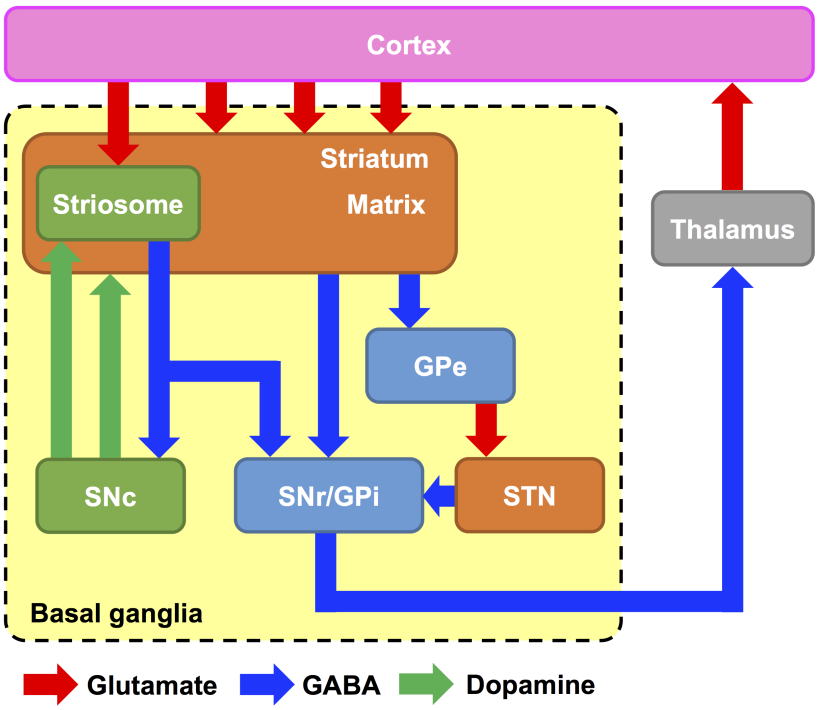

Figure 1.3: Neural circuit of the basal ganglia. The basal ganglia consists of the striatum, globus pallidus (GP), subthalamic nucleus (STN) and substantia nigra (SN). The striatum is an input site, whereas the SNr and GPi are output site of the basal ganglia.

1.3.2 Striatum

Types of cells in the striatum include medium spiny neurons, cholinergic interneurons and GABAergic interneurons. The medium spiny neurons (MSN) are middle sized GABAergic projection neurons having many spines on their dendrite and occupy about 80-95% of neurons in the striatum. While activities of the MSNs are normally silent, they fire only when there are inputs. The MSNs are classified according to their projection sites and expressing types of dopamine receptors as follows; 1) the direct pathway neurons expressing dopamine D1 receptors project to the SNr/GPi and contain the substance P and dynorphin, 2) the indirect pathway neurons expressing dopamine D2 receptors project to the GPe and contain the enkephalin. The interneurons receive inputs from the cortex, thalamus or

SNc and make synapses to projection neurons. Therefore, the interneurons are thought to modify activities of projection neurons.

These neurons do not form layer structures like the cortex and cerebellum. Although they exist randomly in the striatum at a glance, they are scattered across embryologically different two compartments called the striosomes and matrix. The striosomes are embryologically older than the matrix and appears with receiving dopaminergic inputs. On the other hand, the matrix appears after forming the striosomes, however, it finally becomes to occupy about 85% of the entire striatum. Both of them contain direct and indirect pathway MSNs and other interneurons. Inputs to the striosomes derive from the limbic cortex such as the orbitofrontal cortex or insular cortex and those to the matrix derive from broad cortical areas including motor, somatosensory and parietal cortices. Further, unlike the matrix, direct pathway neurons in the striosomes project not only to GPi/SNr but also directly to SNc, where dopaminergic neurons are present [3-6].

1.3.3 Substantia nigra pars compacta and dopamine

Major neurons in the SNc are dopaminergic and especially send projections to the striatum. Dopamine receptors are one of the G-protein-coupled receptors and have D1-D5 subtypes. The adenylate cyclase is activated and inhibited by binding of dopamine to the D1 (also D5) and D2 (also D3 and D4) receptors respectively. The direct and indirect pathway MSNs in the striatum express the D1 and D2 receptors, respectively. Therefore, actions of dopamine are excitatory to the direct pathway, while inhibitory to the indirect pathway.

Schultz and his colleagues found that dopaminergic neurons in the SNc of monkeys showed activities like the TD signals used as learning signals in the RL [7]. They recorded activities of dopaminergic neurons during an operant task in which monkeys could get juice reward by pushing a button at lighting of a lamp. Dopaminergic neurons responded to juice itself in the beginning of learning, then were activated by lighting of a lamp that indicate pushing a button after sufficient learning. Moreover, when monkeys could not get juice reward even if they pushed a button at lighting of a lamp, dopaminergic neurons were inhibited. This finding indicates that dopaminergic neurons do not respond to reward itself but discrepancy between expected and actual reward. The activities

of dopaminergic neurons are much similar to the TD signals in the RL.

1.3.4 RL models of the basal ganglia

The striatum, in which reward-predictive neural activities have been recorded, is a brain site that received dopaminergic inputs most strongly. Samejima and his colleagues recorded neural activities of monkey's dorsal striatum during a free-choice task [8]. In the task, monkeys turned a handle to the left or right, then were able to get probabilistically large or small water rewards depending on their actions. They found many neurons whose activities were different between reward probability 90% and 50% to the left action when monkeys turned a handle to the left. Moreover, they estimated action value functions using a reinforcement learning model and calculated correlations between estimated action values and firing of striatal neurons. The results indicated that activities of many striatal neurons before turning a handle highly correlated with action values of the left or right action.

Based on the above findings, the actor-critic model was proposed as a reinforcement learning model of basal ganglia (Figure 1.4), which assumes that the striosomes and matrix perform the roles of reward prediction (critic) and action selection (actor), respectively [1]. According to this model, the striosomes represents the state values, while the reward prediction error is calculated in the SNc dopaminergic neurons using output from striosomal neurons. The matrix contributes to action selection through acquisition of the rule (policy) to choose the best action, depending on environmental states. However, lateral inhibitory connections between striatal neurons have been shown not to be very strong, in order to effect winner-take-all type action selection [9-11]. Another hypothesis is that the striosome represents the state value whereas the matrix represents the action value, with action selection realized downstream of matrix projections [2, 12].

The hierarchical reinforcement learning model is another model of the striatum [13, 14]. In the striatum, the more dorsolateral parts receive sensorimotor-related information, whereas the more ventromedial areas receive associative and motivational information. Reflecting the anatomical connections, it is known that subdivided striatal areas have different roles for learning, such as the DMS is as-

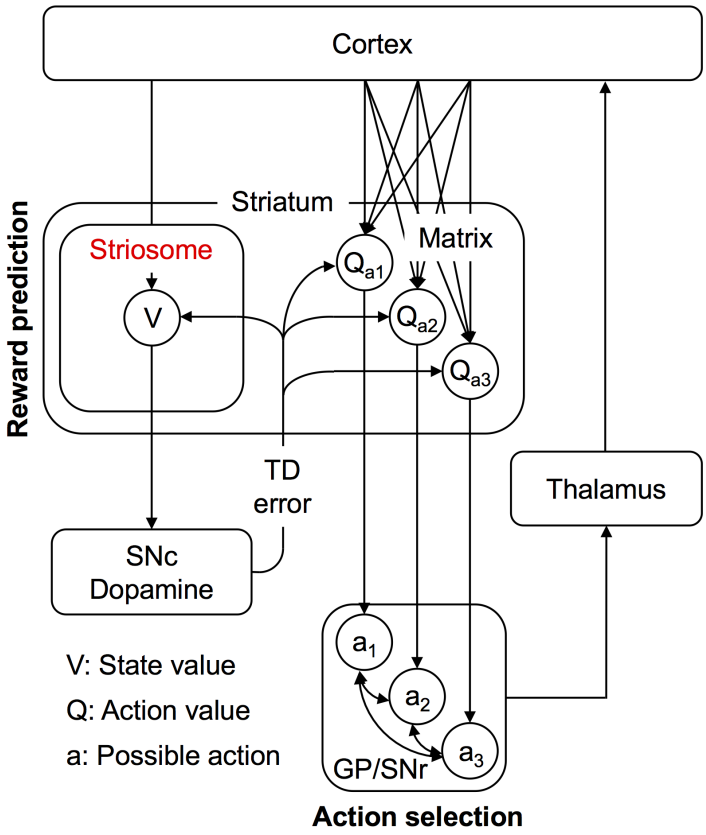

Figure 1.4: Actor-critic model of the basal ganglia. This model hypothesizes that neurons in striosomes and matrix represent state and action values, respectively.

sociated with goal-directed behaviors, whereas the DLS contributes to habitual behaviors [15]. The model hypothesizes that VS, DMS and DLS are involved in actions at different physical and temporal scales. That is, VS is the coarsest module in charge of the action of the whole animal, such as aiming for a goal, avoiding a danger, or just take a rest. DMS is the middle module in charge of abstract actions, such as turn left or go straight. DLS is the finest module in charge of physical actions, such as the control of each limb.

1.4 Aims and composition of this thesis

Many lines of research, including functional brain imaging [16, 17] and neural recording [8, 18-20] have demonstrate that the striatum plays a critical role in reward-based learning and decision making. However, they could not show specific roles of striatal compartments, because the discrimination of striosomes and matrix is difficult due to their mosaic-like structure in the striatum. In addition, I can list another problem that they only focused on neural representations of task-related variables, such as states, actions or rewards, therefore, ignored the striatal functions for motor control.

Aims of this thesis are 1) to examine specific roles of stratal compartments for reward-based learning and 2) to demonstrate parallel processing of task- and motor-related information in the striatum during decision making. In order to address these research questions, I conducted behavioral and neural recording experiments using rodents. For the striatal compartments, I conducted Ca^{2+} imaging experiments during an odor-classical conditioning using striosome-selective cre-expression transgenic mice and found that reward prediction was conducted in striosomal neurons. This is described in Chapter 2. For the striatal parallel information processing, I conducted electrophysiological recoding experiments during a free-choice decision-making task using rats, and got the data supporting parallel information processing in the striatum. This is described in Chapter 3. As a conclusion, I conduct general discussions and show future research directions in Chapter 4.

2. Calcium imaging experiment of striatal neurons in striosomes

2.1 Introduction

The striatum consists of two neurochemically and anatomically distinct compartments: the striosomes (also known as patches), which are rich in mu opioid receptors (MOR), receive inputs from the limbic cortex, and project monosynaptically to midbrain dopaminergic neurons, and the matrix, which receives inputs from the sensorimotor and associative cortices [21-24]. Many lines of research, including functional brain imaging [16, 17] and neural recording [8, 18-20] have demonstrate that the striatum plays a critical role in decision making and reinforcement learning. In the process of reinforcement learning, prediction of forthcoming rewards from the present sensory state and possible actions such as 'state value' and 'action value', respectively, comprise the basis for learning and action selection [25]. These values are updated by a reward-prediction error, defined as the discrepancy between the predicted and actual rewards. The striatum is a major cortical-input site of the basal ganglia and also receives inputs from midbrain dopaminergic neurons encoding the reward-prediction error [7]. Cortico-striatal synapses show dopamine-dependent plasticity that is suitable for reinforcement learning [26]. From these observations, the striatum has been hypothesized as the brain region that predicts future rewards as state or action values [16, 27-29]. In fact, electrophysiological studies have shown that striatal neurons encode state or action values [8, 18-20, 30, 31], but they could not identify whether recorded neurons belonged to striosomes or matrix, because these compartments form a mosaic-like structure [32-34]. Because striosomal neurons comprise only about 15% of striatal neurons, it is particularly unclear whether striosomal neurons are engaged in reward prediction. It is important to characterize their activities during reward-based learning because almost all striatal neurons directly projecting to midbrain dopaminergic neurons belong to striosome compartments [5, 6, 35, 36].

Recently, a transgenic mouse line became available that selectively expresses Cre protein, which is a site-specific DNA recombinase, in striosomal neurons [37,

38]. In combination with optical neural imaging, it is possible to image deep brain structures using endoscopic microscopes [39-41]. In this study, to test whether striosomal neurons show reward-predictive activities, we recorded activities of neurons in striosomes during classical conditioning using endoscopic *in vivo* calcium imaging of transgenic mice with selective calcium indicator expression in their striosomal neurons.

2.2 Materials and Methods

2.2.1 Subjects

Male Sepw1-NP67 [37] mice (n = 8; 25-35 g body weight; 8-12 weeks old) were housed individually under a 12-hr light/dark cycle (lights on at 07:00; off at 19:00). Experiments were performed during the light phase. Water was restricted to 1-2 mL per day for two weeks before experimental initiation and during the experimental period. Food was provided *ad libitum* for the entire period. The Okinawa Institute of Science and Technology Graduate University Animal Research Committee approved the study.

2.2.2 Surgery

Mice were anesthetized with isoflurane (1.0-3.0%) and placed in a stereotaxic frame. The skull was exposed, a hole (diameter: 1.0 mm) was drilled in the skull, and the dura was removed over the imaging site. For calcium imaging, 0.4-0.6 μ L of AAV2/9.Syn.Flex.GCaMP6s (n = 5 mice) or AAV2/9.Syn.GCaMP6s (n = 3, Penn Vector Core) were injected into the striatum (AP: +0.50 mm; ML: $\pm$ 1.75 mm; DV: 2.85 mm from brain surface). Three weeks after virus injection, an endoscope (GRIN lens; PartID: 130-000151; diameter: 0.5mm; length: 6.1 mm; Inscopix) with a custom endoscope holder was slowly implanted at the following coordinates: AP: +0.50 mm; ML: $\pm$ 1.75 mm; DV: 2.60 mm. The endoscope was fixed with UV adhesive (LOCTITE 4305, Henkel) and clear dental cement (Super bond, Sun Medical) and protected by a PCR tube. A head plate (CF-10, NARISHIGE) was fixed with pink dental cement. Two to four weeks after endoscope implantation, awake mice were head-fixed with a head plate holder. A baseplate (Part ID: 100-000279; Inscopix) attached to the miniature micro-

scope was positioned above the endoscope. The focal plane (100-300 μm working distance) was adjusted until neuronal structures and GCaMP6s responses were clearly observed. After mice were anesthetized with isoflurane, the baseplate was fixed with black-painted dental cement (CLEARFIL MAJESTY ES Flow; Kuraray Noritake Dental) and a baseplate cover (Part ID: 100-000241; Inscopix) was secured to the baseplate with a set screw to protect the lens until imaging.

2.2.3 Behavioral task

Mice were head-fixed using the head plate and habituated for 3-5 days before task training. A custom-built olfactometer (O'Hara) delivered a 1:9 mixture of air saturated with one of four odors (isoamyl acetate, citral, eugenol, or (-)carvone) and clean air. The olfactometer constantly delivered clean air during inter-trial intervals (ITIs). ITIs were randomly selected from 10 to 20 sec. In each trial, we delivered one of four odors, selected pseudorandomly, for 2 sec., followed by a delay of 0.5 sec. and an outcome. Each odor was associated with a different outcome: a big drop of water (4μL), a small drop of water (2 μL), no outcome, or an air puff delivered to the animal's face. These outcomes were randomly omitted with a 20% probability. The combination of odor and outcome differed for different mice. A daily session consisted of 100 trials. Licks were detected by interruptions of an infrared beam placed in front of the water tube. 1 g of water gel (HydroGel; ClearH2O) was provided after daily sessions.

2.2.4 Calcium imaging

In each daily session, we first head-fixed mice using the head plate and holder. Then we connected the microscope to the magnetic baseplate, and fixed it in place with the baseplate set screw. Fluorescence images were acquired at 20 fps with LED power at 20% of 1.2 mW/mm^2 maximum and the image sensor gain at 1.0-4.0 before A/D conversion. To compare calcium activity in different sessions, image acquisition parameters were held constant for each mouse across days. An external signal (5V TTL) from the control device triggered the start or end of recording. Neural activities in each trial were recorded from 2.5 sec. before odor onset to 5 sec. after US onset (total: 10 sec./trial) in order to minimize photo toxicity.

2.2.5 Image processing

All image processing was performed in Mosaic (version 1.1.3; Inscopix) and Matlab (version 2016b; Mathworks). First, the raw image of each frame was translated into a 16-bit tiff image. In order to reduce data size and processing time, spatial down-sampling (spatial binning factor: 4) was applied to each tiff image. After image sequences of all trials for each session were concatenated, a motion correction process was applied to remove movement artefacts and to compensate for shifts in microscope positioning. After removing the post-registration black borders, average fluorescence F was calculated over the whole motion-corrected image sequence and percentage-change-over-baseline ($\Delta F/F = (F_n - F)/F$) images were generated for each frame. Here, F_n was the motion-corrected image at n-th frame. Finally, $\Delta F/F$ image sequences of all sessions for each animal were concatenated, and temporally down-sampled (temporal binning factor: 4), then spatial filters to extract activities of single neurons were calculated with a cell-sorting algorithm using independent and principal component analyses [42].

2.2.6 Extraction of calcium signals and event detection

To extract calcium signals of each neuron at 20 Hz, spatial filters were applied to the original $\Delta F/F$ image sequence of each session. The extracted calcium signal of each neuron was normalized to: mean=0, variance=1 (normalized $\Delta F/F$) for each session because the expression levels of GCaMP6s could have differed between neurons and sessions. Then, " Ca^{2+} events " [43, 44] were detected by applying the following procedure. For the normalized $\Delta F/F$ trace in each trial i, all local maxima were detected and for j-th local maximum (M_{ij}), the preceding local minimum (m_{ij}) was registered. When the difference ($\Delta m_{ij} = M_{ij} - m_{ij}$) between the local maximum and the preceding minimum exceeded a threshold (4x the median absolute deviation, 4 MAD), Δm_{ij} was registered as a Ca^{2+} event of amplitude (y_{ik}) at the midpoint time (t_{ik}) between the time of M_{ij} and m_{ij}, where k is the index of the event in a trial.

2.2.7 Experimental design and statistical analysis

To show that a neuron encodes outcomes expected from odor stimuli rather than odor natures, changing of CS-US combinations between mice is effective. Therefore, we needed at least 2 mice each from the striosome and control groups. We actually used 5 and 3 mice from the striosome and control groups, respectively, to collect enough samples to analyze their properties.

Two-sample t-tests were employed for statistical tests for frequencies of licking or Ca^{2+} events between task conditions. In order to evaluate neural representations of behavioral variables, we carried out regression analyses of Ca^{2+} events during the CS-delay period (2.5 sec. between CS onset and US onset) and the US period (2.5 sec. following US onset). Regression analysis employed the variables licking frequency ($Lick$), prediction of reward (Vr), air puff (Va), delivery of reward (Rwd), and air puff (Air). The variables Vr and Rwd took one of three levels: 0 (0 μL), 0.5 (2 μL) and 1 (4 μL) while Va and Air took 0 or 1. Note that Rwd and Air took 0 in omission trials, so that they were different from Vr and Va. The sum of the amplitudes of all Ca^{2+} events during the CS-delay or US period of i-th trial was registered as, $y(i,\mathrm{CS})$ and $y(i,\mathrm{US})$. First, to remove the effects of licking on neural activities, we performed the following regression analysis and obtained the residual activities z:

$$y(i,s) = \beta_0 + \beta_{Lick} Lick(i,s) + z(i,s) \tag{7}$$

where $s = \mathrm{CS}$ or US denotes the time period. We then analyzed residual activities in the CS and US periods using the following regression models.
For big, small, and no reward conditions:

$$z(i,\mathrm{CS}) = \beta_1 + \beta_{Vr}\, Vr(i) \tag{8}$$

$$z(i,\mathrm{US}) = \beta_2 + \beta_{Rwd}\, Rwd(i) \tag{9}$$

For air-puff and no reward conditions:

$$z(i,\mathrm{CS}) = \beta_3 + \beta_{Va}\, Va(i) \tag{10}$$

$$z(i,\mathrm{US}) = \beta_4 + \beta_{Air}\, Air(i) \tag{11}$$

When the p-value of the regression coefficient was <0.05, we concluded that neural activity and the explanatory variable were significantly correlated. Chi-squared tests were used for comparison of proportions of predictive/responsive neurons between groups or stages.

For the decoding analysis, we used $n = 1$ to 10 simultaneously recorded neurons. Since the number of simultaneously recorded neurons differed between mice, we randomly selected $n$neurons from simultaneously recorded populations and regressed Vr or Va, and Rwd or Air with the sum of amplitudes of Ca^{2+} events of them during the CS-delay and US period.
For big, small, and no reward conditions:

$$Vr(i) = w_{Vr,0} + \sum_{j=1}^{n} w_{Vr,j} x_j(i, \mathrm{CS}) \tag{12}$$

$$Rwd(i) = w_{Rwd,0} + \sum_{j=1}^{n} w_{Rwd,j} x_j(i, \mathrm{US}) \tag{13}$$

For air-puff and no reward conditons:

$$Va(i) = w_{Va,0} + \sum_{j=1}^{n} w_{Va,j} x_j(i, \mathrm{CS}) \tag{14}$$

$$Air(i) = w_{Air,0} + \sum_{j=1}^{n} w_{Air,j} x_j(i, \mathrm{US}) \tag{15}$$

where $x_j(i, \mathrm{CS})$ and $x_j(i, \mathrm{US})$ are the sum of amplitudes Ca^{2+} events during the CS-delay and US period, and $w_{Vr,j}$, $w_{Rwd,j}$, $w_{Va,j}$, $w_{Air,j}$ are weights for j-th neuron out of n neurons. After 100 iterations of these procedures for each population size n, we averaged MSEs of each group's mouse in order to compare the population coding of expected and actual US between two groups, and tested them by paired t-test.

2.2.8 Immunohistochemistry

We adapted an immunohistochemical protocol for identifying striosomes in rats [45] for use with mice. After all experiments were completed, mice were deeply anesthetized with pentobarbital sodium and then perfused with 4% paraformaldehyde (PFA). Brains were carefully removed so that endoscopes would not cause

tissue damage, post-fixed in 4% PFA at 4°C overnight, and then transferred to a 30% sucrose/PBS solution at 4°C until brains sank to the bottom. Coronal or horizontal sections were cut at 30 μm on an electrofreeze microtome (REM-710; Yamato) and stored in wells containing PBS at 4°C. Free-floating sections were washed in PBS for 5 min and placed in blocking buffer (5% normal donkey serum and 0.4% Triton X-100 in PBS) for 2 h at room temperature (RT). Sections were simultaneously incubated in primary antibody-rabbit anti-MOR (ab10275; Abcam) diluted 1:500 in blocking buffer, for 48 h at 4°C. Two days later, sections were washed 6x for 10 min in PBS and placed in blocking buffer for 1 h at RT. Sections were simultaneously incubated in secondary antibody donkey anti-rabbit (Alexa Fluor 594; Invitrogen) diluted 1:250 in blocking buffer for 2 h at RT. Sections were washed 6x for 10 min in PBS, mounted on glass slides and coverslipped with VECTASHIELD Mounting Medium with DAPI (Vector Laboratories). To inspect stained tissue, a confocal microscope (LSM780; Carl Zeiss) was used and pictures were taken using ZEN software.

2.3 Results

2.3.1 Spout-licking behavior during odor conditioning

We employed classical odor conditioning, a standard reward-based learning task for rodents [46, 47]. Water-deprived mice were classically conditioned with different odor cues predicting water (reward) or air puffs (aversive stimuli) under head-restrained conditions (Figure 2.1A). Daily training sessions were composed of 100 trials. Each trial began with a conditioned stimulus (CS; odor, 2 sec.), followed by a delay period (0.5 sec.) and an unconditioned stimulus (US; water 4 μL/water 2 μL/air puff/nothing, Figure 2.1B). For each mouse, the CS was randomly selected from four odor cues that the mouse had to associate with different US, and the CS was fixed for all days. The combination of CS-US was varied among mice. In order to evaluate reward-prediction performances of the mice, we counted the number of licks toward the water-delivery spout.

In early training, mice licked the spout immediately after reward onset in some trials. After days of conditioning, they began licking during the CS-delay period before rewards arrived (Figure 2.1C). In order to detect stages of learning,

we quantified each mouse's mean daily licking frequency during the CS-delay period. Licking frequency showed no significant differences between the four odor conditions until day 5. Then commencing at day 6, it became significantly higher in the big-reward condition than in other conditions (Figure 2.1D). By day 11, licking frequencies in big-reward, small-reward, and no-reward conditions differed significantly. Although the numbers of days for CS-US learning differed depending on the mouse, all 8 mice displayed similar behavior. Therefore, we defined two learning stages: 'Early stage', comprising the first three days that licking frequency in the CS-delay period became significantly faster in the big-reward condition than in the no-reward condition ($p < 0.05$, two-sample t-test), and 'Late stage', comprising the first three days that licking frequencies during the CS-delay period in big-reward, small-reward, and no-reward conditions all differed significantly ($p < 0.05$). The number of days from training initiation to the Early stage was 4.6 ± 0.71 (average $\pm$ standard error) and to the Late stage was 12 ± 1.1. Licking frequency during the CS-delay period increased monotonically with reward size in both stages (Figure 2.1E). This result indicates that mice predicted forthcoming rewards from odor stimuli by learning CS-US associations.

2.3.2 Selective in vivo calcium imaging of neurons in striosomes

We used transgenic mice (Sepw1-NP67) expressing Cre selectively in their striosomal neurons [37, 38, 48]. In order to express GCaMP6s selectively in striosomal neurons using the Cre-loxP system, AAV2/9.Syn.Flex.GCaMP6s was injected unilaterally (left hemisphere: 2 mice, right hemisphere: 3 mice) into the dorsomedial striatum (DMS) of transgenic mice (striosome group, Figure 2.2A). MOR immunohistochemistry of virus-injected brain slices confirmed that GCaMP6s was selectively expressed in striosomes (Figure 2.2B). We also prepared mice expressing GCaMP6s in both striosomes and matrix as the control group by injecting AAV2/9.Syn.GCaMP6s (not contain the loxP sequences, left hemisphere: 2 mice, right hemisphere: 1 mouse) to the DMS (Figure 2.2CD).

An endoscope (GRIN lens, diameter: 0.5 mm) was implanted into the DMS, and neural activities were recorded through the endoscope using a miniature microscope integrating an LED light source and an image sensor [39] (Figure 2.2E). 122 neurons were recorded from 5 mice in the striosome group and 83 neurons

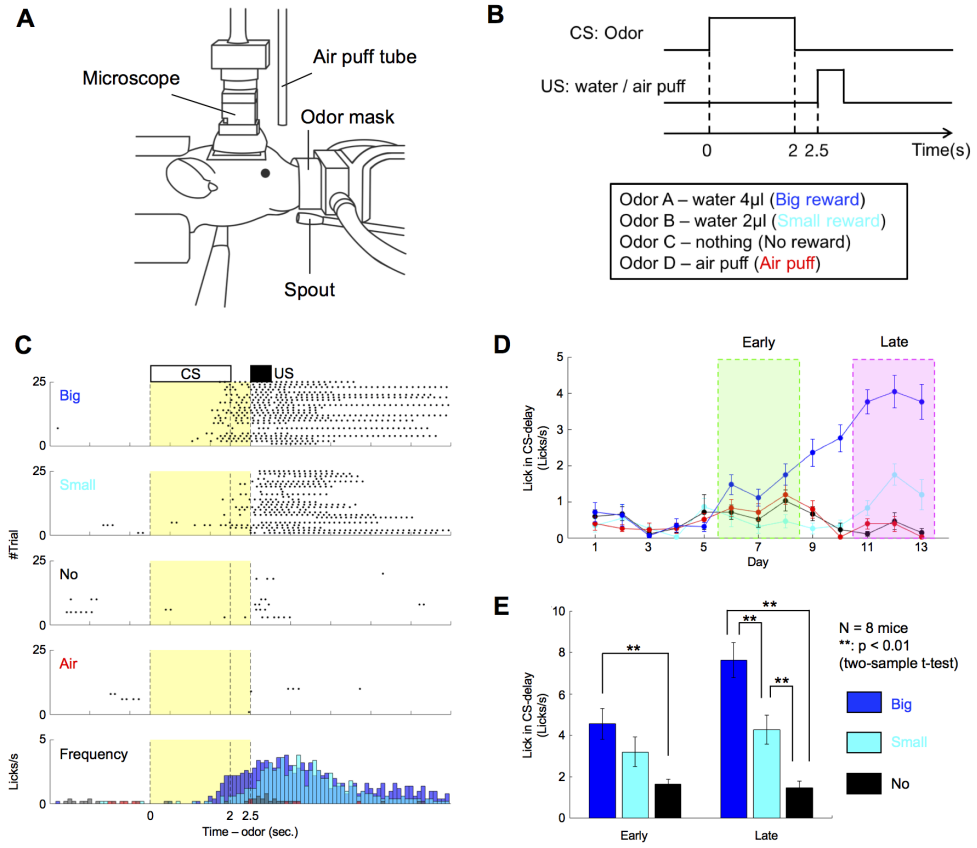

Figure 2.1: Mice showed odor-induced reward-predictive licking behavior proportional to expected reward size. (A) Schematic illustration of the behavioral apparatus. Mice were restricted, head and body, by the metal frame and tube. The odor mask, water spout and air-puff tube were set in front of their noses, mouths, and eyes. Spout-licking behaviors were monitored using an infrared sensor. The miniature microscope was mounted on their heads. (B) Time sequence of a classical conditioning task. (C) An example of reward-predictive spout-licking behaviors after sufficient learning. In trials of reward conditions, spout-licking behaviors started during odor presentation periods. Black dots indicate spout-licking behaviors. Yellow areas show CS-delay periods. (D) Daily changes of spout-licking frequency during CS-delay periods of the mouse illustrated in C. Early and late stages were defined based on the appearance of reward-predictive licking. Error bars indicate standard errors. (E) Average spout-licking frequencies during CS-delay periods of all 8 mice. Error bars indicate standard errors.

from 3 mice in the control group. On average, we were able to simultaneously record 24 neurons (maximum 45) from one mouse in the striosome group and 28 neurons on average (maximum 36) in the control group. Because the advantage of this imaging method is that we can continuously observe the same neurons for several weeks [40, 41], calcium imaging was performed in all mice every day from the first to the final day of behavioral experiments (Figure 2.2F). We measured fluorescence intensity of each neuron during a resting state (for 2.5 sec. before odor onset in each trial) to check changes GCaMP6s expression level. Although 7% and 8% maximum increases in the median rate of change of fluorescence intensity were observed in the striosome and control groups, respectively, differences between sessions had no significant effect upon the rate of change in either group (striosome: $p = 0.69$, control: $p = 0.64$, Kruskal Wallis test). This indicates that neural activities were stably recorded throughout early and late stages.

After the imaging experiment, we made coronal brain slices including the trace of the endoscope and checked GCaMP6s expression and MOR immunohistochemistry. In all 5 mice of the striosome group, we confirmed that the GCaMP6s-expressing neurons were located within the working distance of the endoscope (250-300 μm) and that they were included in the MOR-positive striosome compartments (Figure 2.2G).

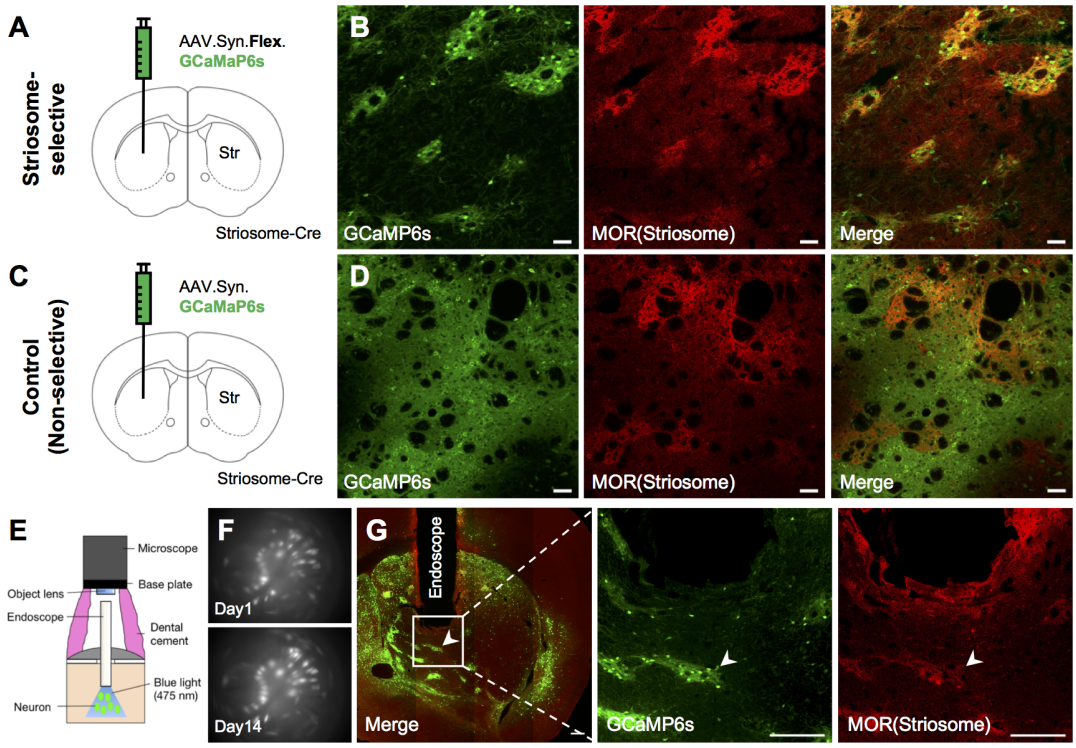

Figure 2.2: An endoscopic microscope was used for selective *in vivo* calcium imaging of striosomal neurons in the striata of Sepw1-NP67 mice expressing Cre-dependent GCaMP6s. (A) Striosome group. In order to express GCaMP6s selectively in striosomal neurons, AAV.Syn.Flex.GCaMP6s was injected into the DMS. (B) GCaMP6s (green) was selectively expressed in striosomes (red) 3 weeks after virus injection. Scale bar: 50μm. (C) Control group. To express GCaMP6s in both striosomes and matrix, AAV.Syn.GCaMP6s was injected to DMS. (D) GCaMP6s expressed in both striosomes and matrix 3 weeks after virus injection. (E) Schematic illustration of endoscopic *in vivo* calcium imaging. (F) Averaged fluorescence images recorded by miniature microscope. White dots indicate neurons. The same neurons in striosomes were stably observed over 2 weeks. (G) Images showing endoscope placement and cre-dependent GCaMP6s-expressing neurons within the striatum. The focal plane in tissue is 250-300 μm from the bottom of the endoscope, as indicated by the white arrow heads. Scale bar: 200 μm.

2.3.3 Reward-predictive neural activities

We first examined responses of striosomal neurons to odor stimuli. After normalizing the $\Delta F/F$ trace of recoded neurons (normalized $\Delta F/F$), we detected "Ca^{2+} events" [43, 44], which estimate the strength of neural activity while taking into account the slow decay time of GCaMP6s [49] (See Methods). In the early stage, the normalized $\Delta F/F$ of a representative striosomal neuron (Figure 2.3AB) rose with the presentation of odor stimuli associated with the big reward, whereas no rise was observed in the no-reward condition. The sum of amplitudes of Ca^{2+} events during the CS-delay period in the early stage was significantly larger in the big-reward condition than in the no-reward condition ($p = 1.2e-04$, two-sample t-test, Figure 2.3C), while the amplitude in the late stage displayed no significant difference between the big-reward condition and the no-reward condition ($p = 0.61$), as the response to the odor stimulus associated with the big reward became weak. The amplitude correlated positively with forthcoming reward size in the early stage ($r = 0.25, p = 1.6e-04$, Figure 3D), but not in the late stage ($r = -0.038, p = 0.58$).

In contrast, the sum of amplitudes of Ca^{2+} events in another striosomal neuron during the CS-delay period in the early stage showed no significant difference between the big-reward condition and the no-reward condition ($p = 0.62$, Figure 2.3E-G), while the response in the late stage was significantly larger in the big-reward condition ($p = 8.2e-06$). The amplitude did not significantly correlate with forthcoming reward size in the early stage ($r = -0.035, p = 0.60$, Figure 2.2H), but positively in the late stage ($r = 0.31, p = 1.7e-06$). Neurons in which the sum of amplitudes of Ca^{2+} events during the CS-delay period correlated with forthcoming reward size in one of the learning stages were found in the control group as well.

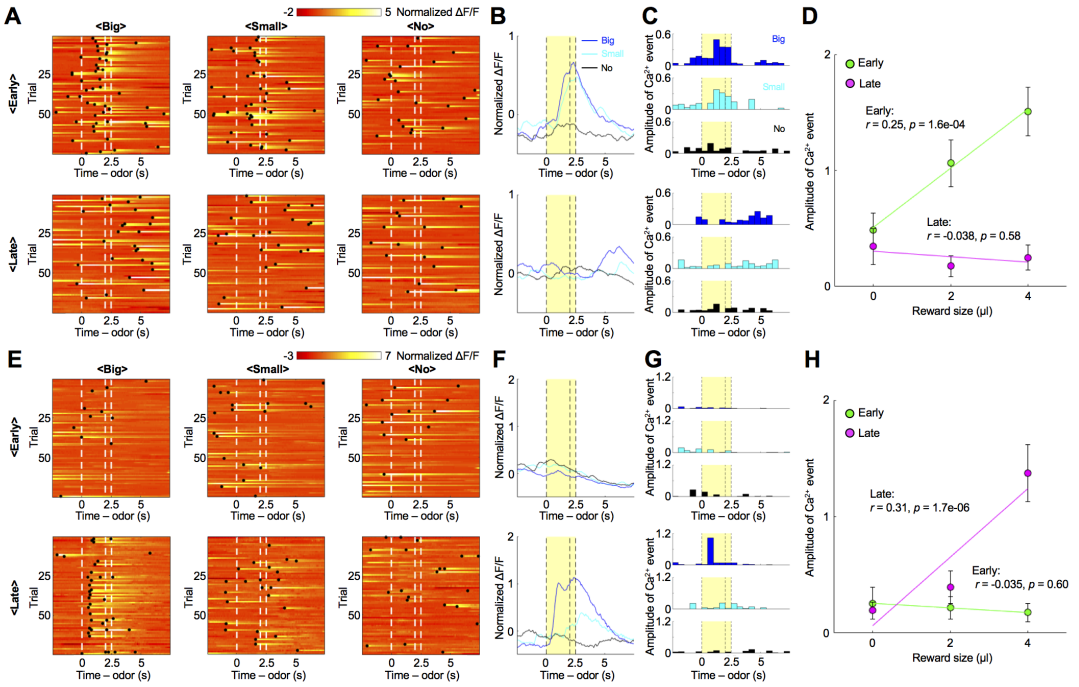

Figure 2.3: Reward-associated odors activated striosomal neurons in a specific learning stage. (A) Normalized $\Delta F/F$ of a striosomal neuron showing reward-predictive activity specifically in the early stage. Black dots indicate detected Ca^{2+} events. (B, C) Averaged $\Delta F/F$ and Ca^{2+} events of the striosomal neuron illustrated in A. Yellow areas show the CS-delay period. (D) Amplitudes of CS-delay period Ca^{2+} events of the striosomal neuron illustrated in A were averaged over trials and plotted against reward size. In the early stage, Ca^{2+} events show a positive correlation with reward size ($r = 0.25, p = 1.6e-04$). On the other hand, this correlation disappeared in the late stage (r=-0.038, p=0.58). Error bars and lines indicate standard errors and regression lines. (E) Normalized $\Delta F/F$ of another striosomal neuron showing reward-predictive activity specifically in the late stage. (F, G) Averaged $\Delta F/F$ and Ca^{2+} events of the striosomal neuron illustrated in E. (H) Amplitudes of CS-delay period Ca^{2+} events of the striosomal neuron illustrated in E were averaged over trials and plotted against the reward size. In the early stage, Ca^{2+} events show no significant correlation with reward size ($r = -0.035, p = 0.60$). However, a positive correlation was observed in the late stage ($r = 0.31, p = 1.7e-06$).

To quantify proportions of reward-predictive neurons in the striosome, we performed a regression analysis of the sum of amplitudes of Ca^{2+} events during the CS-delay period. In order to eliminate neural activities directly related to licking movements, we first conducted a regression analysis with licking frequencies and then analyzed residual components with the reward (Vr) predicted from the odor cues (See Methods). In most neurons of both striosome and control groups, reward-predictive activities that had significant regression coefficients to Vr were observed specifically in the early or the late stage (Figure 2.4A). 8% of striosomal neurons (10 of 122) and 13% of control neurons (11 of 83) were reward-predictive in the early stage, but not in the late stage. On the other hand, 10% of striosomal neurons (12 of 122) and 1% of control neurons (1 of 83) were reward-predictive in the late stage, but not in the early stage. In the striosome group, only 2% (2 of 122) of the neurons were reward-predictive in both learning stages. Therefore, total proportion of the striosome group was not significantly different from that of the control group in the early stage, while it was larger in the late stage (Early: 10%, striosome, and 13%, control, $p = 0.45$, Late: 11%, striosome, and 1%, control, $p = 0.0056$, Chi-squared test, Figure 2.4B). Compared with the early stage, reward-predictive neurons in the control group decreased in the late stage ($p = 0.0027$). Moreover, the majority of reward-predictive neurons had positive regression coefficients to Vr (Early: 50%, striosome, and 82%, control, Late: 93%, striosome, and 100%, control).

Furthermore, to study neural representation of expected reward at the population level, we performed a decoding analysis of forthcoming reward size from simultaneously recorded neuronal activities. Since the numbers of simultaneously recorded neurons were different in each mouse, we randomly selected n neurons from each simultaneously recorded population and used their activities during the CS-delay period for linear regression of forthcoming reward size (Figure 2.4C, See, Methods). We varied the sub-population size n from 1 to 10 and for each n, we took 100 random combinations of neurons and compared the mean squared errors (MSE) for striosome and control groups in early and late stages (Figure 2.4D). The results indicated that MSEs of the striosome group were significantly larger in the early stage and smaller in the late stage than those of the control group (Early: $p = 1.1e-04$, Late: $p = 0.0020$, paired t-test).

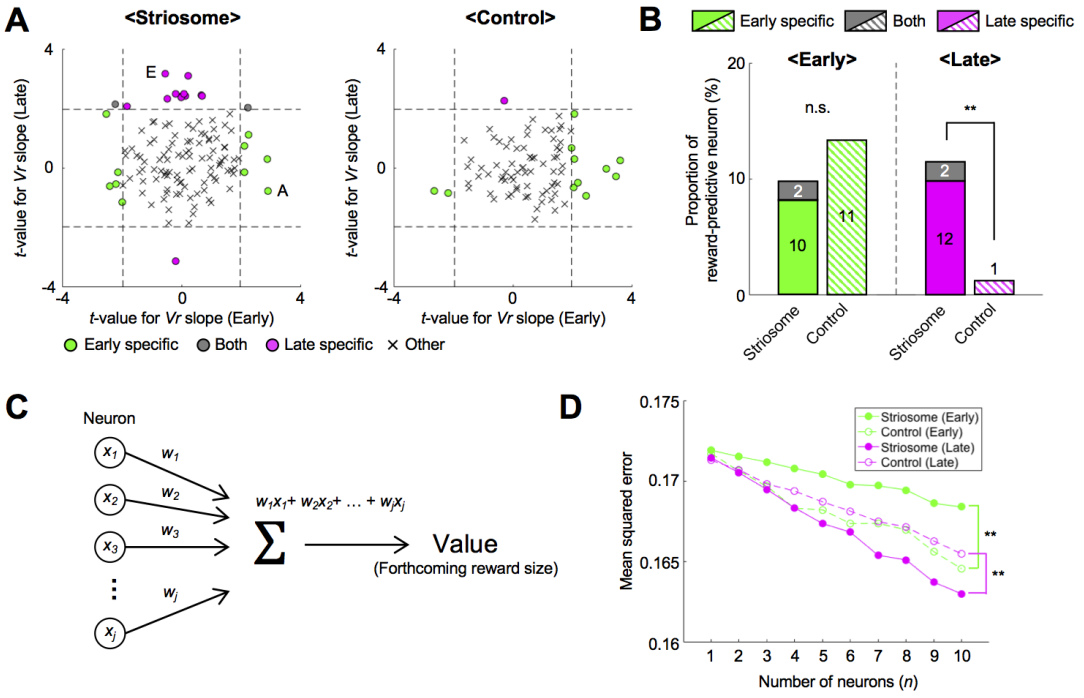

Figure 2.4: During each learning stage, different neural ensembles participated in reward prediction and population coding of expected reward differed between two groups. (A) In order to remove effects of motor behavior on neural activities, we first performed a regression analysis of the sum of amplitudes of Ca^{2+} events during the CS-delay period with frequencies of licking. Then we analyzed the residual component using prediction of reward (Vr). Scatter plots of t-values for regression coefficients of Vr in each learning stage. Dashed lines indicate levels of significant Vr slope at $p = 0.05$. Letters A and E indicate the example neurons in Figures 2.3A and E. (B) Proportions of reward-predictive neurons in each learning stage. Numbers in bars indicate actual counts of reward-predictive neurons. **: $p < 0.01$, n.s.: $p \geq 0.05$, Chi-squared test. (C) Schematic illustration of neural decoding analysis. Forthcoming reward size was estimated from the sum of weighted neuronal activities. x_j: sum of amplitudes of Ca^{2+} events during the CS-delay period. w_j: weight for j-th neuron out of n neurons. (D) Mean squared errors between actual and decoded reward sizes at each number of neurons used for analyses. **: $p < 0.01$, paired t-test.

These analyses of reward-predictive neural activities revealed that neurons in striosomes represent reward values of odor stimuli in specific learning stages, and that reward-predictive striosomal neurons are more dominant in the late learning stage.

2.3.4 Air-puff-predictive neural activities

We next examined whether recorded neurons responded to air-puff-predictive odor stimuli. In the early stage, the normalized $\Delta F/F$ of a representative striosomal neuron (Figure 2.5AB) rose with the presentation of odor stimuli associated with an air puff, whereas this rise was not observed in the no-reward condition. The sum of amplitudes of Ca^{2+} events during the CS-delay period in the early stage was significantly larger in the air-puff condition than in the no-reward condition ($p = 0.036$, two-sample t-test Figure 2.5C). On the other hand, CS-delay period activity in the late stage showed no significant difference between the air-puff condition and the no-reward condition ($p = 0.98$) as the $\Delta F/F$ response to odor stimuli associated with the air puff became weak.

Contrastingly, the sum of amplitudes of Ca^{2+} events in another striosomal neuron (Figure 2.5DE) during the CS-delay period in the early stage showed no significant difference between air-puff and no-reward conditions ($p = 0.35$), while amplitudes in the late stage were significantly larger in the air-puff condition than in the no-reward condition ($p = 1.0e-04$, Figure 2.5F). Neurons in which the sum of amplitudes of Ca^{2+} events during the CS-delay period differed significantly between the air-puff and no-reward conditions in one of the learning stages were also found in the control group.

Next, we analyzed air-puff-predictive activity using the predicted delivery of an air puff (Va) as the regressor. As in the case of reward-predictive activities, air-puff-predictive activities were observed specifically in one learning stage or the other (Figure 2.5G). 11% of striosomal neurons (13 of 122) and 1% of control neurons (1 of 83) were air-puff-predictive in the early stage, but not in the late stage. On the other hand, 10% of striosomal neurons (12 of 122) and 2% of control neurons (2 of 83) were air-puff-predictive in the late stage, but not in the early stage. 3% of striosomal neurons (4 of 122) and 1% of control neurons (1 of 83) were air-puff-predictive in both learning stages. This means that total

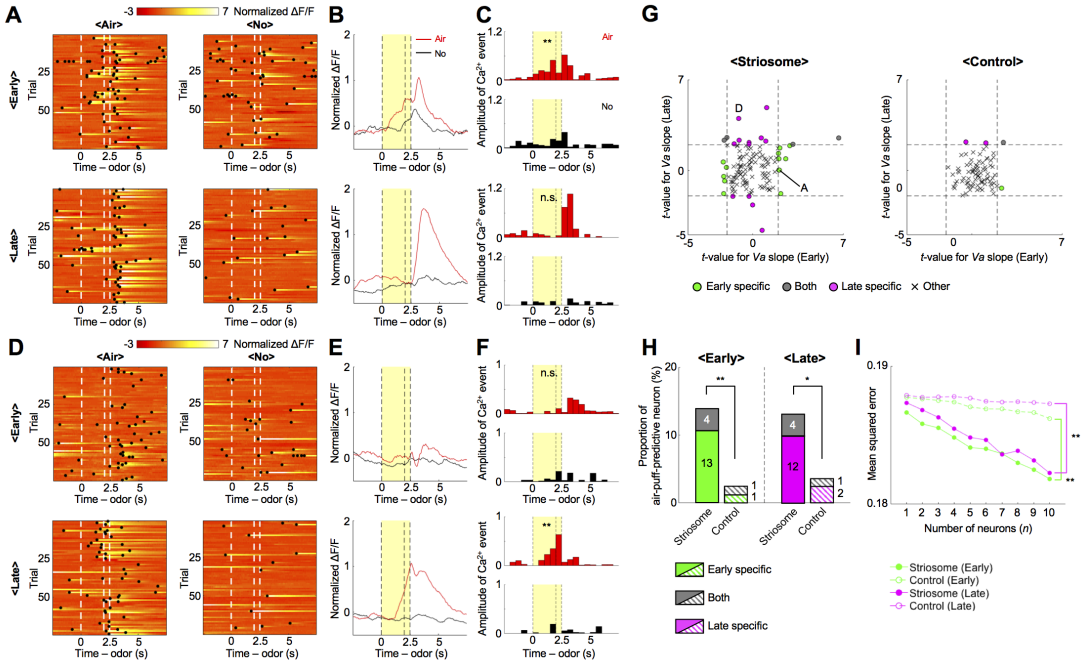

Figure 2.5: During each learning stage, different neural ensembles in the striosome predicted air-puff stimuli. (A) Normalized $\Delta F/F$ of a striosomal neuron showing air-puff-predictive activities specifically in the early stage. Black dots indicate detected Ca^{2+} events. (B, C) Averaged $\Delta F/F$ and Ca^{2+} events of the striosomal neuron illustrated in A. Yellow areas show the CS-delay period. **: $p < 0.01$, n.s.: $p \geq 0.05$, two-sample t-test. (D) Normalized $\Delta F/F$ of another striosomal neuron showing air-puff-predictive activities specifically in the late stage. (E, F) Averaged $\Delta F/F$ and Ca^{2+} events of the striosomal neuron illustrated in D. **: $p < 0.01$, n.s.: $p \geq 0.05$. (G) Scatter plots of t-values for regression coefficients of prediction of air puff (Va) in each learning stage. Dashed lines indicate levels of significant Va slope at $p = 0.05$. Letters A and D indicate the example neurons in Figures 2.5A and D. (H) Proportions of air-puff-predictive neurons in each learning stage. Numbers in bars indicate actual counts of air-puff-predictive neurons. n.s.: $p \geq 0.05$, Chi-squared test. (I) Mean squared errors between actual and decoded air-puff values at each number of neurons used for analyses. **: $p < 0.01$, paired t-test.

proportions of the striosome group were significantly larger than those of the control group in both learning stages (Early: 14%, striosome, and 2%, control, p=0.0052, Late: 13%, striosome, and 4%, control, p=0.021, Chi-squared test, Figure 2.5H). Moreover, the majority of air-puff-predictive striosomal neurons had positive regression coefficients to Air (Early: 58%, striosome, and 100%, control, Late: 75%, striosome, and 100%, control).

Furthermore, to compare the population neural coding of expected aversive stimulus between two groups, we decoded forthcoming air-puff stimuli from the activities of various sizes of sub-populations of simultaneously recorded neurons (Figure 2.5I). In both learning stages, MSEs of the striosome group were significantly smaller than those of the control group (Early: $p = 1.1e-05$, Late: $p = 2.1e-04$, paired t-test). These analyses of air-puff-predictive neural activities showed that neurons in striosomes also represent aversive values of odor stimuli in learning-stage specific ways, as is the case with reward values, and suggest that aversive values are more strongly encoded in the striosomes than in the matrix.

2.3.5 Reward- and air-puff-responsive neural activities

The normalized $\Delta F/F$ of a representative striosomal neuron (Figure 2.6AB) rose with reward presentation, whereas that rise was not observed in the absence of a reward. The sum of amplitudes of Ca^{2+} events during the US period in rewarded trials was significantly larger in the big-reward condition than with no-reward ($p = 1.35e-10$, two-sample t-test, Figure 2.6C). On the other hand, amplitudes in reward-omitted trials did not differ significantly between big-reward and no-reward conditions ($p = 0.25$). Amplitude positively correlated with reward size in rewarded trials ($r = 0.42, p = 9.4e-10$, Figure 2.6D), but not in reward-omitted trials ($r = 0.16, p = 0.096$). This indicates that striosomal neurons responded to the rewards themselves. Reward-responsive activities were also observed in neurons of the control group.

After subtracting the licking component (see Methods), regression analyses of the sum of amplitudes of Ca^{2+} events during the US period revealed that most reward-responsive neurons, which had significant regression coefficients to the acquired reward size *Rwd*, had learning-stage-specific properties, similar to those of reward-predictive neurons (Figure 2.6E). 13% of striosomal neurons (16

of 122) and 13% of control neurons (11 of 83) were reward-responsive in the early stage, but not in the late stage. On the other hand, 11% of striosomal neurons (13 of 122) and 13% of control neurons (11 of 83) were reward-responsive in the late stage, but not in the early stage. 7% of all neurons showed reward-responsive activities in both learning stages in the striosome group (9 of 122), but only 1% in the control group (1 of 83). Therefore, total proportions of the striosome group were not significantly different from those of the control group in either learning stage (Early: 20%, striosome, and 14%, control, $p = 0.27$, Late: 18%, striosome, and 14%, control, $p = 0.50$, Chi-squared test, Figure 2.6F).

In addition, we decoded acquired reward size from various numbers of simultaneously recorded neuronal activities during the US period (Figure 2.6G). In both learning stages, MSEs of the striosome group were significantly smaller than those of the control group (Early: $p = 0.0034$, Late: $p = 5.9e-04$, paired t-test). This decoding result also shows that the reward acquisition is more robustly presented by the striosome neurons.

The normalized $\Delta F/F$ of another striosomal neuron (Figure 2.6HI) rose with the presentation of an air-puff stimulus, whereas this rise was not observed without the air puff. The sum of amplitudes of Ca^{2+} events during the US period was significantly larger in the air-puff condition than in the no-reward condition ($p = 1.49e-08$, Figure 2.6J), whereas the response in the air-puff-omitted trials was not significantly different from that in the no-reward condition ($p = 0.28$). This indicated that the striosomal neuron respond to the air-puff stimulus itself. The air-puff-responsive activities were observed in neurons of the control group as well.

We analyzed air-puff-responsive activity using received air puff (Air) as a regressor in much the same way as with reward-responsive activities (Figure 2.6K). 25% (31 of 122) of striosomal neurons and 19% of control neurons (16 of 83) were air-puff-responsive in the early stage, but not in the late stage. On the other hand, 11% of striosomal neurons (14 of 122) and 14% of control neurons (12 of 83) were air-puff-responsive in the late stage, but not in the early stage. 16% of striosomal neurons (20 of 122) and 17% of control neurons (14 of 83) were air-puff-responsive in both learning stages. This means that the two groups did not differ significantly in total proportions of air-puff-responsive neurons in either learning

stage (Early: 42%, striosome, and 36%, control, $p = 0.42$, Late: 28%, striosome, and 31%, control, $p = 0.59$, Chi-squared test, Figure 6L). Finally, we decoded received air-puff stimuli from various numbers of simultaneously recorded neuronal activities during the US period (Figure 6M). MSEs of the striosome group were significantly larger in the early stage and smaller in the late stage than those of the control group (Early: $p = 6.2e-04$, Late: $3.2e-06$, paired t-test).

These results indicate that some striosomal neurons respond directly to reward or air-puff stimuli.

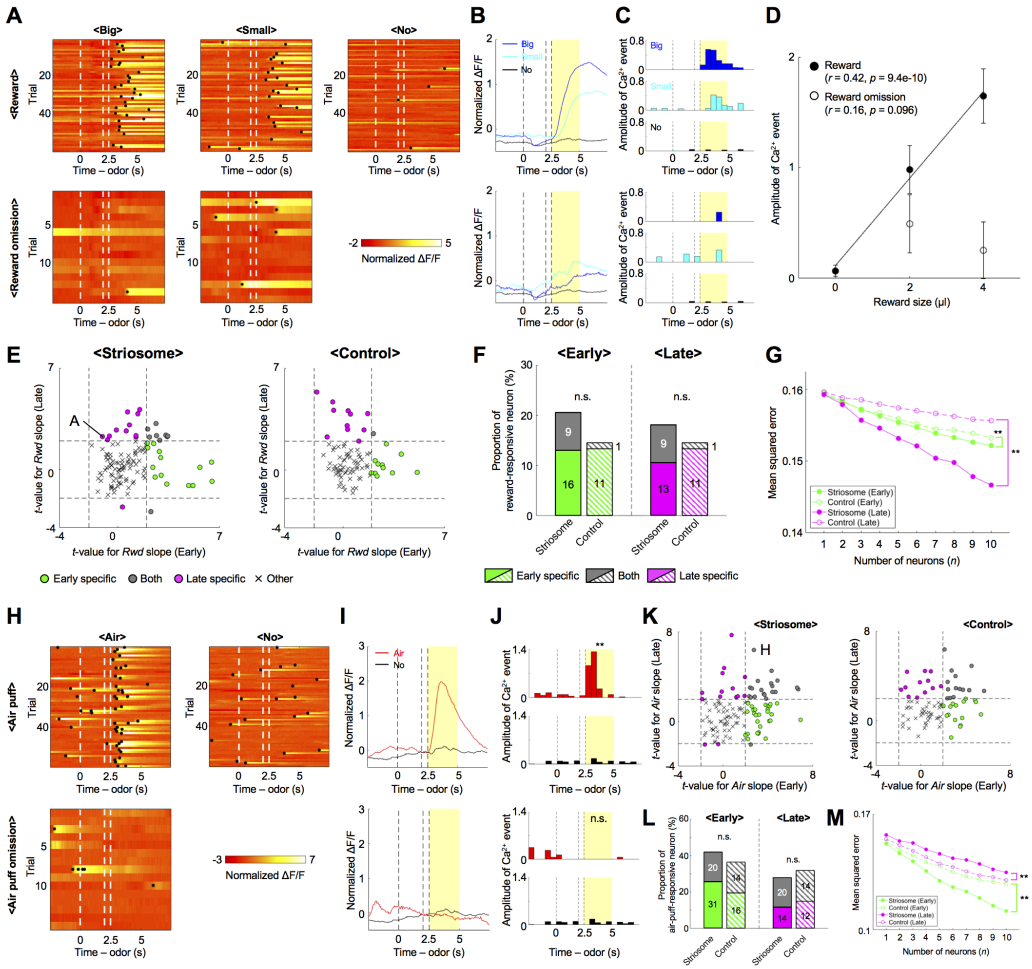

Figure 2.6: Both rewards and air puffs activated striosomal neurons. (A) Normalized $\Delta F/F$ of a striosomal neuron showing reward-responsive activities. This is $\Delta F/F$ in the late stage. Black dots indicate detected Ca^{2+} events.

Figure 2.6: (B, C) Averaged $\Delta F/F$ and Ca^{2+} events of the striosomal neuron illustrated in A. Yellow areas show the US period. (D) Amplitudes of US period Ca^{2+} events of the striosomal neuron illustrated in A were averaged over trials and plotted against reward size. In rewarded trials, Ca^{2+} events show a positive correlation with reward size ($r = 0.42$, $p = 9.4e - 10$). On the other hand, there was no significant correlation in reward-omitted trials ($r = 0.16$, $p = 0.096$). Error bars and lines indicate standard errors and regression lines. (E) Scatter plots of t-values for regression coefficients of delivery of reward (Rwd) in each learning stage. Dashed lines indicate levels of significant Rwd slope at $p = 0.05$. Letter A indicates the example neuron in Figure 2.6A. (F) Proportions of reward-responsive neurons in each learning stage. Numbers in bars indicate actual counts of reward-responsive neurons. n.s.: $p \geq 0.05$, Chi-squared test. (G) Mean squared errors between actually received and decoded reward size at each number of neurons used for analyses. **: $p < 0.01$, paired t-test. (H) Normalized $\Delta F/F$ of a striosomal neuron showing air-puff-responsive activities. This is also $\Delta F/F$ in the late stage. Black dots indicate detected Ca^{2+} events. (I, J) Averaged $\Delta F/F$ and Ca^{2+} events of the striosomal neuron illustrated in G. **: p¡0.01, n.s.: $p \geq 0.05$, two-sample t-test. (K) Scatter plots of t-values for regression coefficients of delivery of air puff (Air) in each learning stage. Dashed lines indicate levels of significant Air slope at $p = 0.05$. Letter H indicate example neurons in Figure 2.6H. (L) Proportions of air-puff-responsive neurons in each learning stage. Numbers in bars indicate actual counts of air-puff-responsive neurons. n.s.: $p \geq 0.05$. (M) Mean squared errors between actually received and decoded air-puff stimuli at each number of neurons used for analyses. **: $p < 0.01$, paired t-test.

2.4 Discussion

We performed selective *in vivo* calcium imaging of neurons in striosomes and monitored neural activities of mice performing a classical odor-conditioning task. To the best of our knowledge, this is the first report to characterize striosomal neuronal activities of living animals. The major findings were as follows:

1. Striosomal neurons showed reward- or air-puff-predictive activities; therefore, they encoded the values of odor stimuli.

2. Most reward or air-puff-predictive activities were specific to early or late learning stages.

3. Some striosomal neurons responded to presentation of a reward or an air puff.

4. Striosomal neurons have more significant roles in reward and air-puff prediction than randomly recorded striatal neurons.

2.4.1 Predictive neural activities in striosomes

Although previous electrophysiological studies reported that striatal neurons represent value information [8, 18, 20], they did not distinguish between striosomal and matrix neurons. In this study, we found that neurons in striosomes show reward- or air-puff-predictive activities that matched the definition of value, both by regression of single neuron activities and by decoding from population activities. We also found about 10% of non-selectively recorded neurons in the DMS showed reward-predictive activities in the early stage. This proportion is consistent with a recent electrophysiological study [20]. Since the licking frequency in cue period correlated with forthcoming reward size, it was possible that reward-predictive striosomal activity might represent motor behavior instead of reward size expected from odor stimuli. However, those activities represented the reward size even after removing the effects of licking. Thus, the striosome encodes values of odor stimuli.

This result that striosomal neurons encode values of present sensory states, supports reinforcement learning models that postulate that striosomal neurons

learn state values [1, 2, 12]. These models postulated that matrix neurons are involved in either action selection (actor) or action value learning. An alternative view, based on human brain imaging or lesion experiments, is that the dorsal and ventral striatum assume the roles of actor and critic, respectively [16]. However, the striosomes comprise a larger portion of the ventral striatum than of the dorsal striatum; whereas the matrix constitutes a smaller portion of the ventral striatum and a larger portion of the dorsal striatum [50]. Therefore, the striosome-matrix difference may contribute to ventral-dorsal functional differences. A recent rabies tracing study indicated that both striosomal and matrix neurons project to dopaminergic neurons, with a higher density of SNc projecting neurons in the striosome, but a larger number in the matrix, given its larger volume [38]. This new finding raises the possibility that matrix neurons are also directly involved in computation of reward prediction error signals. In order to test those hypotheses, we will need to record and analyze the activities of striosomal and matrix neurons during an operant conditioning task that involves choices between multiple actions. It would also be desirable to record selectively not only striosomal neurons, but also matrix neurons from the ventral, dorsomedial and dorsolateral striatum.

In both learning stages, the proportion of air-puff-predictive neurons was larger in the striosomes than in the control. Air-puff stimuli are widely used as aversive stimuli in rodents and known to cause avoidance behaviors such as predictive eye blinks [46, 51-53]. A recent study revealed anatomical connections to striosomes from the bed nucleus of the stria terminalis [38], which is known to be involved in fear or anxiety [54, 55]. Furthermore, optogenetic inhibition of axon terminals of prefrontal neurons projecting to the striosomes reduced sensitivity to aversive light exposure in a cost-benefit conflict situation [56]. Air-puff-predictive neurons in striosomes might link aversive signals to avoidance behaviors through their projection to the SNr and the internal globus pallidus and fear or anxiety through their projection to the stria terminalis.

In the Sepw1-Cre mouse line used in this study, 83.2% of Cre-expressing neurons were D1 medium spiny neurons (MSNs), projecting monosynaptically to dopaminergic neurons in the SNc, while matrix neurons that do not express Cre had no such projections [38]. It was shown in the same Sepw1-Cre line that striatonigral fibers originating from the striosome form bouquet-like arborizations

innervating clusters of dopamine-containing neurons with tightly bundled dendrites [48]. Therefore, it is expected that the majority of striosomal neurons that showed reward- and air-puff-predictive activities in this study have monosynaptic projections to dopaminergic neurons in the SNc, which encode reward-prediction errors [7]. Our present discovery that the majority of reward-predictive striosomal neurons showed activities positively correlated with reward values suggests that they contribute to subtraction of predicted reward in computing reward prediction errors. On the other hand, striosomal neuronal activities that were correlated negatively with reward or positively with air puffs might contribute to computation of saliency, including both reward and aversive information, which is represented by a subset of dopaminergic neurons [57].

2.4.2 Learning-stage-specific neural ensembles coding value information

Since the endoscopic *in vivo* calcium imaging method made it possible to observe activities of the same neurons over long periods, we were able to compare value representations of each striatal neuron across learning stages. It was an unexpected finding that reward- or air-puff-predictive activities observed in the early stage disappeared in the late stage. It was also surprising that there were few neurons that showed reward- or air-puff-predictive activities in both early and late learning stages. This result indicates that value-coding neurons form unique ensembles depending on the learning stage. Combined with the finding of Thorn et al. [15] that population activities of the striatum change with learning, the ensemble representation of value information in the early stage might contribute to goal-directed behavior, while that in the late stage might support habitual behavior.

2.4.3 Differences in reward-related neural coding in striosomes and matrix

Different parts of the striatum, especially near its ventromedial to dorsolateral axis, have different roles in goal-directed and habitual behaviors [58]. It was reported that population activities of DMS neurons become weaker after acquisition of habitual behavior [15]. In this study, we implanted endoscopes in the DMS

and monitored their neural activities during reward-based learning. Our regression analyses show that the number of reward-predictive neurons in the control group in the late stage decreased from that in the early stage. This is consistent with the result of non-selective recording of DMS neurons. In the late stage, the proportion of reward-predictive neurons was larger in the striosome group than in the control group. Our decoding analyses also showed that population neural activities of striosomes represented expected rewards more strongly than the control in the late stage. It is expected that recorded neural activities from the control group are mostly derived from the matrix, since roughly 85% of striatum neurons belong to the matrix. This suggests a possibility that striosomal neurons assume a more dominant role in reward prediction after habituation than do matrix neurons. On the other hand, roughly 80% of neurons in striosomes are D1-MSNs and another 20% are D2-MSNs, whereas the proportion is around 50%-50% in matrix [5]. Therefore, the differences between the striosome and control groups may reflect the difference in D1/D2 percentages.

Our finding of reward- and air-puff-predictive activities of neurons in striosomes contributes to understanding of mechanisms of reinforcement learning in the brain. The next important issues to clarify are whether striosomal neurons encode the state value rather than the action value in a choice task, and to test whether and how striosomal neurons contribute to computation of reward-prediction error.

3. Electrophysiological experiment of the striatum and the cortex

3.1 Introduction

It is well known that the cortico-basal ganglia circuits are involved in motor control, whereas recent studies have demonstrated that they are also engaged in reward prediction and decision making [7, 8]. Especially, the striatum, which is a major cortical input site of the basal ganglia, have important roles for both motor control and reward prediction. Anatomically, it is subdivided into the dorsolateral striatum (DLS), the dorsomedial striatum (DMS), and the ventral striatum (VS). Because the DLS and the DMS are received inputs from sensorimotor and frontal cortices, respectively [59], it is supposed that they have different roles for motor control or decision making. In fact, recent lesion and recording studies of DLS and DMS suggest that the DLS is necessary for habitual behaviors, whereas that the DMS is important for goal-directed behaviors with reward prediction [60]. For example, when rats learned a reward-based choice task with two options, such as left and right, neurons in the DMS were activated depending on expected reward to each option, whereas those in the DLS showed short-term activations in various timing of trials [20]. These DLS neural activities are thought to be involved in motor behaviors and contribute to habituation of motor patterns [60-62]. These studies only focused on neural responses to one modality, such as reward, animal's position or locomotion speed, however, it is possible that neurons in the basal ganglia including the striatum do not purely represent one modality but also parallely encode multiple modalities.

In this study, we recorded rat's neural activities from DLS, DMS, M1 and PL during a reward-based free-choice task, and captured rat's motions in parallel. Then, we tested whether the recorded neural activities purely encoded one modality or parallely encoded multiple modalities.

3.2 Materials and Methods

3.2.1 Subjects

Male Long-Evans rats (n = 6; 260-310 g body weight; 16-37 weeks old at the first recording session) were housed individually under a light/dark cycle (lights on at 7:00, off at 19:00). Experiments were performed during the light phase. Food was provided after training and recording sessions so that body weights dipped no lower than 90% of initial levels. Water was supplied ad libitum. The Okinawa Institute of Science and Technology Graduate University Animal Research Committee approved the study.

3.2.2 Apparatus

All training and recording procedures were conducted in a $40 \times 40 \times 50$ cm experimental chamber placed in a sound-attenuating box (O'Hara & Co.). The chamber was equipped with three nose-poke holes in one wall and a pellet dish on the opposite wall (Figure 3.1A). Each nose-poke hole was equipped with an infrared sensor to detect head entry, and the pellet dish was equipped with an infrared sensor to detect the presence of a sucrose pellet (25 mg) delivered by a pellet dispenser. The chamber top was open to allow connections between electrodes mounted on the rat ' s head and an amplifier. House lights, two video cameras, two infrared (IR) LED lights and a speaker were placed above the chamber. A computer program written with LabVIEW (National Instrument) was used to control the speaker and the dispenser, and to monitor states of the infrared sensors.

3.2.3 Behavioral task

Animals were trained to perform a choice trial and a no-choice trial using nose-poke responses. In either task, each trial began with a tone presentation (start tone: 3000 Hz, 1000 ms). When the rat performed a nose-poke in the center hole for 500-1000 ms, one of two cue tones (choice tone: white noise, 1000-1500 ms; no-choice tone: 900 Hz; 1000-1500 ms) was presented (Figure 3.1B).

After onset of the tone A (choice trials), the rat was required to perform a nose-poke in either the left or right hole within 2 s after the exit from the center hole. If the rat exited the center hole before the offset of the choice tone, the choice

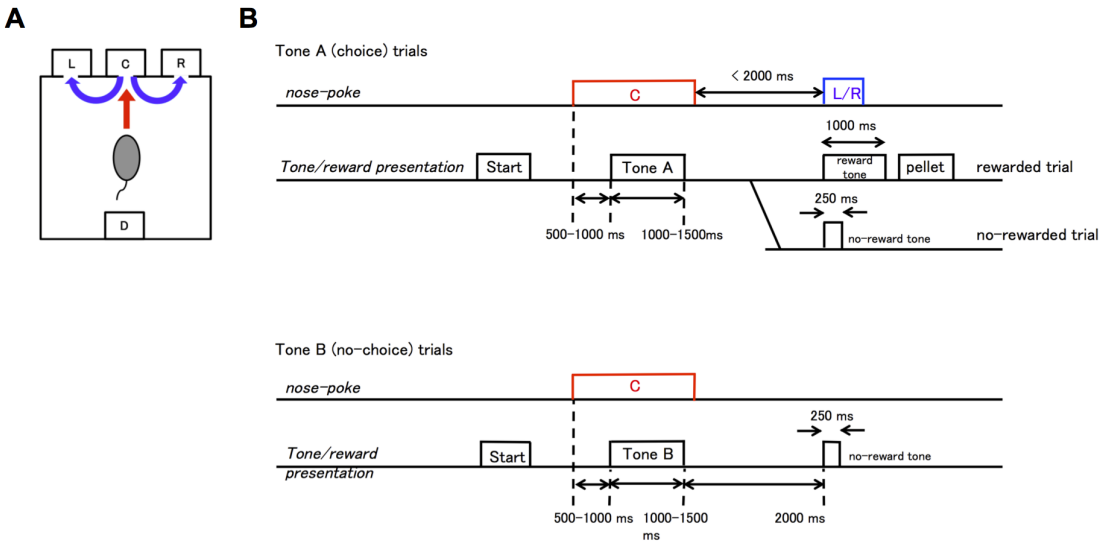

Figure 3.1: Apparatus and behavioral task. (A) Experimental chamber was equipped with three holes for nose poking and a pellet dish. (B) Time sequences of choice task. The behavioral task consisted of choice and no-choice trials.

tone was stopped. When the rat performed left or right poking, ether a reward tone (500 Hz, 1000 ms) or a no-reward tone (500 Hz, 250 ms) was probabilistically determined depending on the selected action, and presented. The probabilities were ether (Left, Right) = (75%, 25%) or (25%, 75%) and switched for every block. The reward tone was followed by delivery of a sucrose pellet (25 mg) in the food dish. If the rat did not perform neither left nor right nose-poke within 2 s, the trial was ended as an error trial after the error tone (9500 Hz, 1000 ms).

For the tone B (no-choice trials), the rat was required not to perform left nor right nose pokes during 2 s after the exit from the center hole. Then, the trial was correctly finished by the presentation of the no-reward tone. In this no-choice trial, the rat could not obtain any pellets, but if the rat could not perform this trial correctly, that is, if the rat incorrectly performed left or right nose-poke for the no-choice tone, the trial was ended as an error trials after the error tone presentation, and the no-choice trial was repeated again in the next trial.

We designed the continuous condition (CC) consisting of only choice trials, and the intermittent condition (IC) where no-choice trial was inserted after every

choice trial. Because it was hard for the rats to keep performing IC, we had to limit the number of choice trials to 20 trials in one sequential of IC condition. First three blocks were CC and the subsequent two blocks were IC. A block is defined as a sequence of the trials under the same reward probabilities. The probabilities of the first block were randomly selected from (Left, Right) = (75%, 25%) or (25%, 75%) for every recording session.

To adjust the block-change conditions in CC and IC, the first CC and the third IC blocks were ended when the choice frequency of the rats in the last 10 choice trials reached 80% optimal. The second IC, and, fourth and fifth CC blocks were ended when 10 choice trials had been conducted. By this setting, the first 20 choice trials in the second and the third CC blocks, and in the fourth and fifth IC blocks could be comparable; starting from 80% biased choice and switching reward probabilities after 10 choice trials. These set of five blocks were repeated basically six times in one day recording session.

3.2.4 Surgery

After rats mastered the choice tasks, they were anesthetized with pentobarbital sodium (50 mg/kg, i.p.) and placed in a stereotaxic frame. The skull was exposed and holes were drilled in the skull over the recording site. Three drivable electrode bundles were implanted into DLS in the right hemisphere (1.0 mm anterior, 3.5 mm lateral from bregma, 3.3 mm ventral from the brain surface), M1 in the right hemisphere (1.0 mm anterior, 2.6 mm lateral from bregma, 0.4 mm ventral from the brain surface), DMS in the left hemisphere (1.0 mm posterior, 1.6 mm lateral from bregma, 3.7 mm ventral from the brain surface), and PL in the left hemisphere (3.2 mm anterior, 0.7 mm lateral from bregma, 2.0 mm ventral from the brain surface).

An electrode bundle was composed of eight Formvar-insulated, 25 μm bare diameter nichrome wires (A-M Systems) and was inserted into a stainless-steel guide cannula (0.3 mm outer diameter; Unique Medical). Tips of the microwires were cut with sharp surgical scissors so that $\sim$ 1.5 mm of each tip protruded from the cannula. Each tip was electroplated with gold to obtain an impedance of 100-200 kΩ at 1 kHz. Electrode bundles were advanced by 125 μm per recording day to acquire activity from new neurons.

3.2.5 Electrophysiological recoding

Recordings were made while rats performed the choice tasks. Neuronal signals were passed through a head amplifier at the head stage and then fed into the main amplifier through a shielded cable. Signals passed through a band pass filter (50∼3000 Hz) to a data acquisition system (Power1401; CED), by which all waveforms that exceeded an amplitude threshold were time-stamped and saved at a sampling rate of 20 kHz. The threshold amplitude for each channel was adjusted so that action potential-like waveforms were not missed while minimizing noise. After a recording session, the following off-line spike sorting was performed using a template-matching algorithm and principal component analysis by Spike2 (Spike2; CED): recorded waveforms were classified into several groups based on their shapes, and a template waveform for each group was computed by averaging. Groups of waveforms that generated templates that appeared to be action potentials were accepted, and others were discarded. Then, to test whether accepted waveforms were recorded from multiple neurons, principal component analysis was applied to the waveforms. Clusters in principal component space were detected by fitting a mixture Gaussian model, and each cluster was identified as signals from a single neuron. This procedure was applied to each 50 min data segment; and if stable results were not obtained, the data were discarded. Then, gathered spike data were refined by omitting data from neurons that satisfied at least one of the five following conditions:

(1) The amplitude of waveforms was $< 7\times$ the SD of background noise.

(2) The firing rate calculated by perievent time histograms (PETHs) (from -4.0 s to 4.0 s with 100 ms time bin based on the onset of cue tone, the exit of the center hole, or the entrance of the left or right hole) was $<$ 1.0 Hz for all time bins of all PSTHs.

(3) The estimated recording site was considered to be outside the target.

Furthermore, considering the possibility that the same neuron was recorded from different electrodes in the same bundle, we calculated cross-correlation histograms with 1 ms time bins for all pairs of neurons that were recorded from different electrodes in the same bundle. If the frequency at 0 ms was 10 $\times$ larger than the mean frequency (from -200 ms to 200 ms, except the time bin at 0 ms) and their PETHs had similar shapes, either one of the pair was removed from

the database.

3.2.6 Motion tracking

Three IR-reflection markers for motion tracking were attached on rat's head, back and tail. We used sphere markers (diameter; 15 mm) for head and back and fixed by adhesive (Aron Alpha, Konishi) and a screw implanted in their back, respectively. IR-reflection seal (width; 8mm) was used as a tail marker and fixed on base of tail. Rat's motion during the task was recorded at 30 fps using two monochrome cameras set on the top of chamber. After experiments, marker's positions were calculated from two movies in three dimensions using Move-tr/3d (Library). In the procedures, we firstly calibrate coordinates using six reference points in the chamber. The calibration was conducted in every session since the chamber might be moved by cleaning etc.

3.2.7 Histology

After all experiments were completed, rats were anesthetized as described in the surgery section, and a 10 μA positive current was passed for 30 s through one or two recording electrodes of each bundle to mark the final recording positions. Rats were perfused with 10% formalin containing 3% potassium hexacyanoferrate (II), and brains were carefully removed so that the microwires would not cause tissue damage. Sections were cut at 60 μm on an electrofreeze microtome and stained with cresyl violet. Final positions of electrode bundles were confirmed using dots of Prussian blue. The position of each recorded neuron was estimated from the final position and the moved distance of the bundle of electrodes. If the position was outside DLS, M1, DMS, or PL, recorded data were discarded.

3.3 Results

3.3.1 Behavioral performance

We trained the free-choice task to 6 rats and conducted 28 sessions totally. In the continuous condition (CC), rats became to choose an optimal side with large reward probability (75%) with their reward experiences (Figure 3.2A). When the side switched to the opposite after some trials, their choice tendencies also changed. In the CC, the probability of choosing an optimal side in choice trials was 0.65 on average and was significantly larger than the chance level ($p = 1.8e-04$, two-sample t-test, Figure 3.2B). However, in the intermittent condition (IC), the probability was 0.53 on average and was not significantly larger than the chance level ($p = 0.40$). In addition, the probability was significantly larger in the CC than in the IC ($p = 1.4e-13$, paired t-test). These results indicate that rats recognized a side with large reward probability and reflected it to their action selections in the CC, whereas that no-choice trials in the IC inhibited their optimal action selections.

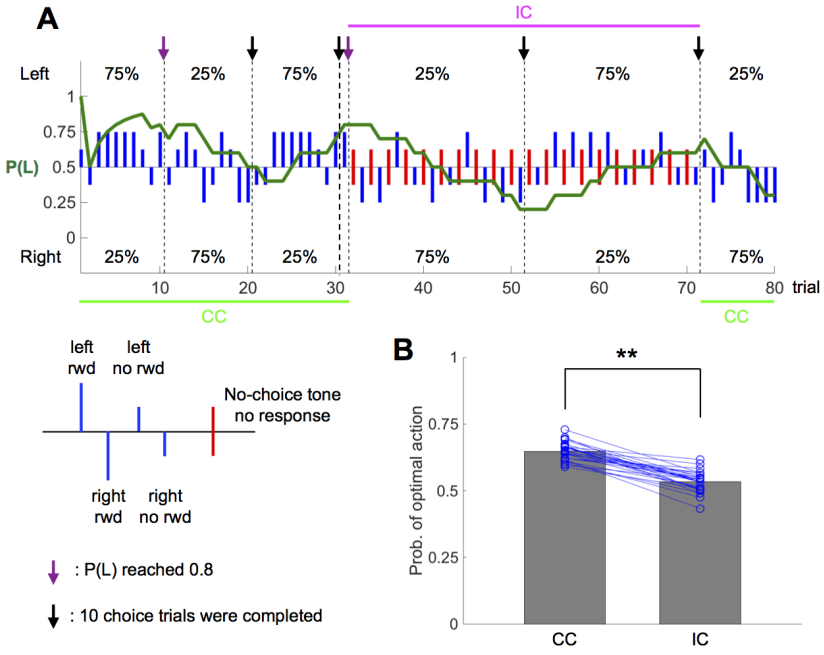

Figure 3.2: Behavioral results.

Figure 3.2: (A) A representative example of a rat's performance. The blue vertical lines indicate individual choices in choice trials. The red vertical lines indicate no-choice trials. The long lines and short lines represent rewarded and no-reward trials, respectively. The green trace in the middle indicates the probability of a left choice in choice trials (average of the last 10 choice trials). (B) The gray bars indicate average probabilities of optimal action selection. The blue circles indicate the probabilities in each session.

3.3.2 Neural responses to task-, space- and motor-related variables

We succeeded in recording of 185, 71, 81 and 171 neurons from DMS, DLS, PL and M1, respectively. To investigate parallel neural representation during the task, we distinguished behavioral variables into following three classes. Task-related variables were presented cues (tone A/B), rat ' s choices (left/right), and delivery of reward. Space-related variables were rat ' s head positions and head directions in the chamber. Motor-related variables were rat ' s head velocities (lateral/anterior), accelerations (lateral/anterior) and rotation. Note that there is a possibility that the task-, space- and motor-related variables are correlated with each other. As an example, when a cue tone was presented, rats' motion stopped, since they entered their noses into the center hole.

We firstly checked neural responses to the task-related variables. The averaged activities of an example of DMS neuron rose with the presentation of tone A, whereas no rise was observed to tone B (Figure 3.2A). To statistically test these responses, we compared firing rates during 0.5 sec. following cue onset of tone A or B. The result showed significance difference ($p < 0.05$, Mann-Whitney U test) at $p = 1.0e - 13$, therefore, we classified this neuron as a tone-correlated neuron. Another example DMS neuron was more strongly activated when a rat chose the right poke hole than when he chose the left one (Figure 3.2B). The firing rate during 0.5 sec. following poke C offset was significantly different ($p < 0.05$) between in action L and R ($p = 4.3e - 29$). This type of neuron was named an action-correlated neuron. In addition, another example of neural activities recorded from DMS was activated by deliveries of reward, whereas not activated in no-rewarded trials (Figure 3.2C). We compared firing rates during 0.5 sec.

following reward or no-reward tone onset between rewarded and no-rewarded trials. As the result, these activities were significantly different ($p < 0.05$) at $p = 0.0073$, therefore this neuron was a reward-correlated neuron.

In summary of neural representations of task-related variables, the proportions of tone-correlated neurons (Figure 3.2D) were 39% (73 of 185 neurons; DMS), 49% (35 of 71; DLS), 49% (40 of 81; PL) and 29% (50 of 171; M1). The proportion of the PL was significantly larger than that of the M1 ($p = 0.0016$, Chi-squared test). The proportions of action-correlated neurons (Figure 3.2E) were 45% (84 of 185; DMS), 55% (39 of 71; DLS), 26% (21 of 81; PL) and 59% (101 of 171; M1). The proportion of the M1 was significantly larger than that of the PL ($p = 1.1e - 06$). The proportions of reward-correlated neurons (Figure 3.2F) were 14% (26 of 185; DMS), 8% (6 of 71; DLS), 16% (13 of 81; PL) and 22% (37 of 171; M1).

We next checked neural responses to the space-related variables. To investigate neural activities to rat's head position, we calculated average firing rates in 160 (40×40) subdivided places, therefore, the area of each subdivide place was 1 cm^2. The result of a DMS neuron was presented as a heat map (Figure 3.3A). It seemed that this neuron had place preference around right-poking hole. In order to quantify its place preference, we changed the numbers of subdivide areas from 160 to 12 (3×3 in chamber + 3 poking holes), then calculated averaged firing rates in each area. The averaged firing rates in the place #9 and #R were over the entirely averaged firing rate plus 0.5 standard deviation (Figure 3.3B). We defined the space-preference neuron as neuron whose averaged firing rates in each area was over the entirely averaged firing rate plus 0.5 standard deviation (s.d.) at least one area.

We also examined whether recorded neurons showed activities depending on rat's head directions in the chamber. Head directions were defined as the facing side of line drawn from back to head markers. We showed an example of activities of a DMS neuron (Figure 3.3D). Here, we discretized head directions to eight directions in order to simplify analyses, then calculated average firing rates at each direction. The averaged firing rate at 90 ° was over the entirely averaged firing rate plus 0.5 s.d. (Figure 3.3E). We defined the head-direction neuron as

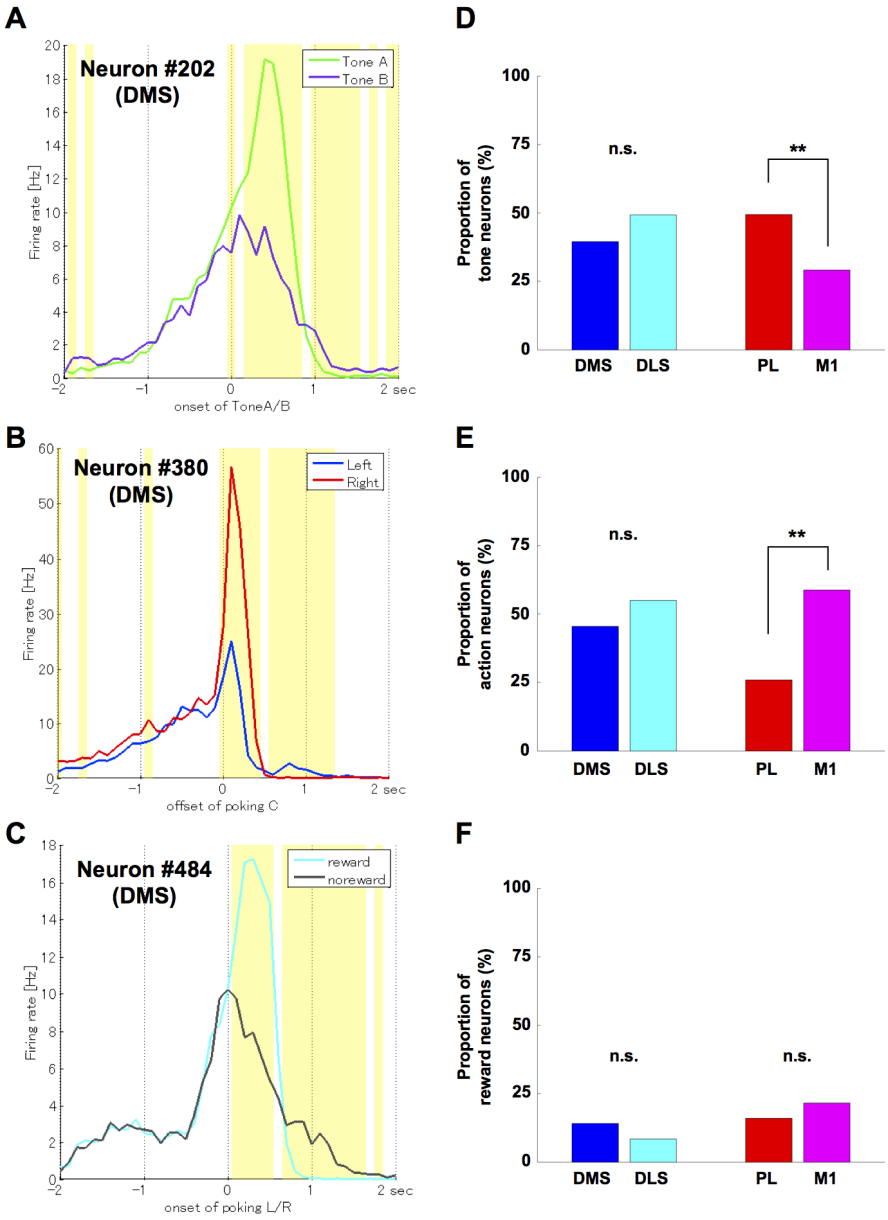

Figure 3.3: Examples of tone-, action- and reward-correlated neurons. (A) Averaged firing rate of a tone-correlated DMS neuron. Yellow areas indicate significant difference between tone A and B. (B) Averaged firing rate of an action-correlated DMS neuron. (C) Averaged firing rate of a reward-correlated DMS neuron. (D-F) The proportions of tone-correlated neurons (D), action-correlated neurons (E) and reward-correlated neurons (F).

neuron whose averaged firing rates at each direction was over the entirely averaged firing rate plus 0.5 s.d. at least one direction.

In summary of neural representations of space-related variables, the proportions of place-preference neurons (Figure 3.3C) were 32% (59 of 185; DMS), 8% (6 of 71; DLS), 33% (27 of 81; PL) and 26% (45 of 171; M1). The proportions of head-direction neurons (Figure 3.3F) were 17% (31 of 185; DMS), 1% (1 of 71; DLS), 10% (8 of 81; PL) and 9% (16 of 171; M1). In both cases of space and head direction, the proportions of the DMS were significantly larger than those of the DLS (space; $p = 1.1e-04$, head direction; $p = 8.9e-04$).

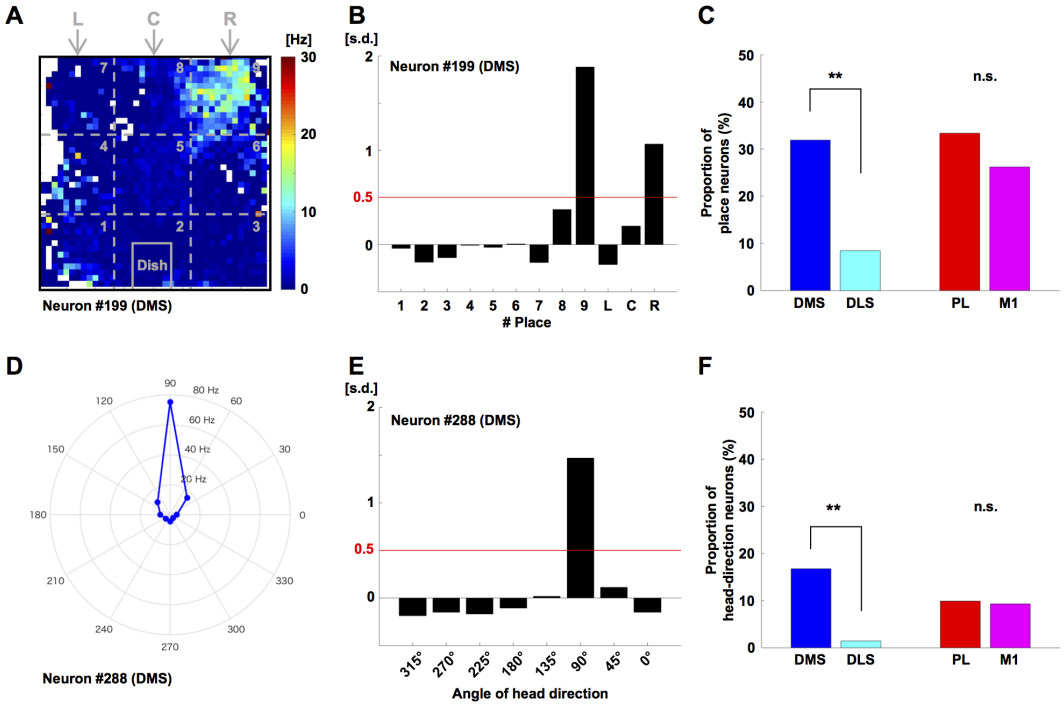

Figure 3.4: Examples of neural activities with preferences for place or head direction. (A) Heat map of averaged firing rate of a DMS neuron at each place. The gray numbers indicate # of place. (B) This neuron was strongly activated in the place #9 and #R. Each bar indicates differences from the averaged firing rate of entire place. (C) The proportions of neurons with place preference. **; $p < 0.01$, n.s.; $p \geq 0.05$, Chi-squared test. (D) Averaged firing rates at each head direction.

Figure 3.4: (E) This neuron was strongly activated at 90°. Each bar indicates differences from the averaged firing rate of all directions. (F) The proportions of neurons with head-direction preference.

Thirdly, to check neural responses to the motor-correlated variables, we calculated head velocities in each 100 ms time bin on rat's egocentric axes and analyzed correlations between the head velocities and firing rates. These velocities took positive values when head moved to right or anterior side, whereas negative values when it moved to left or posterior side. An example of neural activities of the DLS was shown in Figure 3.4A and C. The firing rate of this neuron became higher when the head moved to left or posterior side. The correlation coefficient between lateral velocities and firing rate (rlateral) was significantly negative ($r_{lateral} = -0.25$, $p < 1.0e-256$). And the correlation coefficient between anterior velocities and firing rate (ranterior) was also significantly negative ($r_{anterior} = -0.24$, $p < 1.0e-256$).

Moreover, rat's rotations were detected using head and back markers. For the detection, we drew lines connecting head and back markers at each time bin, then measured angles between lines before and after rotation. Here, positive angle means right rotation, whereas negative angle means left rotation. An example of neural activities of the DMS was presented in Figure 3.4E. This neuron was more strongly activated when the rat rotated left side. The correlation coefficient between rotation angle and firing rate ($r_{rotation}$) was significantly negative ($r_{rotation}$=-0.34, $p < 1.0e-256$).

In summary, the medians of $r_{lateral}$ were DMS; 0.024, DLS; 0.016, PL; 0.017 and M1; 0.032 (Figure 3.4B). The medians of $r_{anterior}$ were DMS; 0.023, DLS; 0.022, PL; 0.021 and M1; 0.041 (Figure 3.4D). The medians of $r_{rotation}$ were DMS; 0.024, DLS; 0.018, PL; 0.014 and M1; 0.033 (Figure 3.4F). In the cortex, the M1 had significant larger $r_{lateral}$ ($p = 4.2e-04$, Mann-Whitney U test), $r_{anterior}$ ($p = 1.4e-05$) and $r_{rotation}$ ($p = 1.2e-04$) than the PL, whereas in the striatum, there were no significant differences between the DMS and DLS.

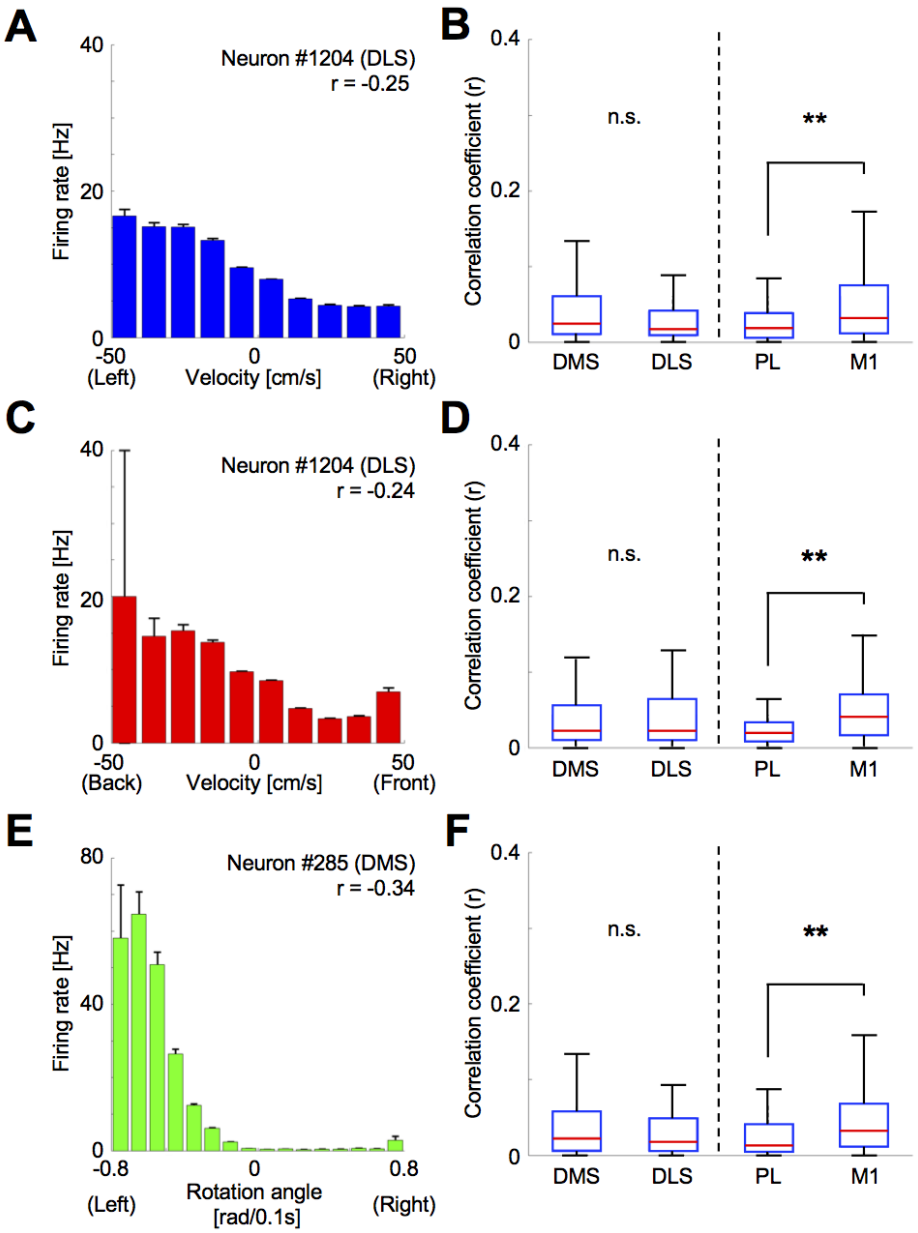

Figure 3.5: Examples of neural activities correlated with head velocities or rotation. (A) Averaged firing rate of a DLS neuron at each velocity range (lateral motion). Right motion has positive velocity. Error bars indicate standard errors. (B) Medians of correlation coefficients between lateral velocity and firing rate. **; $p < 0.01$, n.s.; $p >= 0.05$, Mann-Whitney U test.

Figure 3.5: (C) Averaged firing rate of same DLS neuron at each velocity range (anterior motion). Anterior motion has positive velocity. Error bars indicate standard errors. (D) Medians of correlation coefficients between anterior velocity and firing rate. (E) Averaged firing rate of a DMS neuron at each rotation angle. Right rotation has positive. Error bars indicate standard errors. (E) Medians of correlation coefficients between rotation angle and firing rate.

3.3.3 Parallel neural representations of task-, space- and motor-related information in the striatum and cortex

As I showed in the preceding sections, we confirmed neurons in four recorded brain regions responded to task-, space-, or motor-related variables. In this section, to examine parallel neural representations of these variables, we designed multiple regression models, and tested which model did most explain each neural activity. The regression models were 1) task model including information of conditions (CC or IC), type and presentation of cues, rat's choice, and delivery of reward, 2) space model including information of rat's place and head directions, 3) motor model including rat's head velocities (lateral/anterior), acceleration (lateral/anterior) and angle of rotation, 4) task-space model combining task model and space model, 5) task-motor model combining task model and motor model, 6) space-motor model combining space model and motor model, 7) task-space-motor model combining task model, space model and motor model, 8) null model.

After fitting firing rates of each neuron to these models, we got BICs as model criteria, then selected a model with the minimum BIC. The proportions of each model were presented in Figure 3.7A. In all recorded areas, the task-space-motor model was most frequently selected (DMS; 64%, 119 of 185, DLS; 56%, 40 of 71, PL; 72%, 58 of 81, M1; 74%, 128 of 172). In the DMS, DLS and M1, the space-motor model was second-frequently selected (DMS; 21%, 39 of 185, DLS; 24%, 17 of 71, PL; 5%, 4 of 81, M1; 16%, 28 of 172). In the PL, the task-motor model was second-frequently selected (DMS; 4%, 7 of 185, DLS; 6%, 4 of 71, PL; 12%, 10 of 81, M1; 5%, 9 of 172). In total, 68% of DMS neurons (126 of 185), 61% of DLS neurons (44 of 71), 84% of PL neurons (68 of 81) and 80% of M1 neurons (137 of 172) were represented task-related variables (Figure 3.7B).

Focusing on space-related variables, 85% of DMS neurons (158 of 185), 80% of DLS neurons (57 of 71), 77% of PL neurons (62 of 81) and 91% of M1 neurons (156 of 172) represented them. In the case of motor-related variables, 90% of DMS neurons (166 of 185), 87% of DLS neurons (62 of 71), 90% of PL neurons (73 of 81) and 97% of M1 neurons (167 of 172) represented motor-related variables. In the cortex, the total proportions of neurons coding space- or motor-related variables were significantly larger in the M1 than in the PL (task; $p = 0.42$, space; $p = 0.0023$, motor; $p = 0.019$, Chi-squared test), whereas there were no significant differences in the striatum (task; $p = 0.35$, space; $p = 0.32$, motor; $p = 0.58$). These results indicate that majorities of striatal and cortical neurons encoded multiple modalities in parallel and that the M1 encoded space and motor information more strongly than the PL.

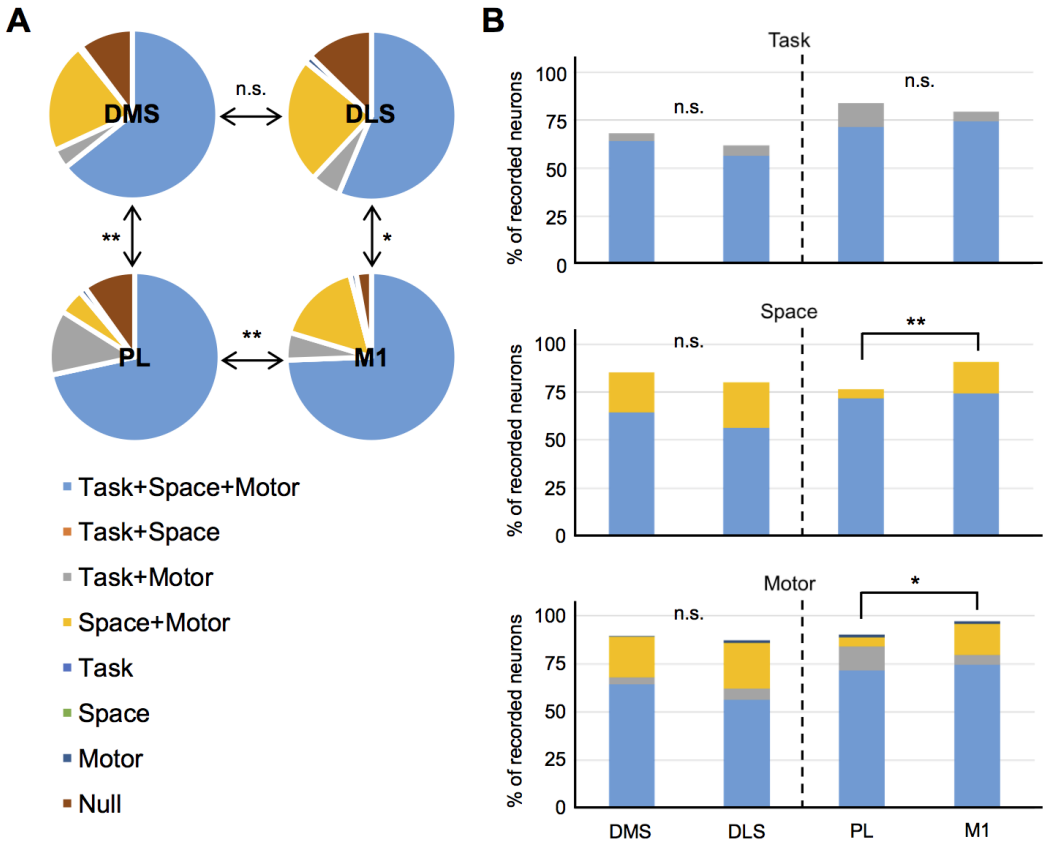

Figure 3.6: Results of model selection. (A) The proportions of selected models. **: $p < 0.01$, *: $p < 0.05$, n.s.: $p \geq 0.05$, Chi-squared test. (B) The total proportions of neurons coding task-, space- or motor-related variables. **: $p < 0.01$, *: $p < 0.05$, n.s.: $p \geq 0.05$, Chi-squared test.

3.4 Discussion

We captured motions of rats performing a choice task and performed electrophysiological recording of neurons in DMS, DLS, PL and M1 in parallel. The major findings were as follows;

1. Neurons in DMS, DLS, PL and M1 responded to task-, space-, or motor-related variables.

2. Majorities of striatal and cortical neurons encoded multiple modalities in parallel.

3. Compared with neurons in the PL, those in the M1 encoded space or motor information more strongly.

In analyses of neural activities based on the task-related variables, the activities of neurons modulated by types of cue tones (tone-correlated neurons) may code either the difference in expected reward (i.e., no pellet for the no-choice tone, and less than one pellet for the choice tone) or the difference in behaviors after the tone. In another case, the activities of neurons modulated by different actions during the execution of the poking (action-correlated neurons) might code either differences in the physical movements or differences in the spatial position of rats. In the current study, we recorded rat's physical motions and neural activities in parallel, therefore, could analyze the activities on space- and motor-related variables as well. The results demonstrated that many neural activities would not only be explained by task-related variables, but also by other variables, such as place, head velocity or rotation and so on. This is because variables in each modality were related to each other. For example, when rats were poking to a hole or eating a pellet, their head velocities became nearly zero. Therefore, we tried to determine which combination of task-, space- and motor-related variables did most explain neural activities. For this purpose, we compared BICs of eight linear regression models and selected the model with the minimum BIC. In all recorded brain regions, task-space-motor coding neurons were majorities. Moreover, in the DMS, DLS and M1, space-motor coding neurons were second-frequently observed. On the other hand, neurons coding only one modality of

variables were rare. These results suggest the parallel information processing in the cortico-basal ganglia loops.

In the striatum, there were no significant differences on the distribution of selected model. This is an unexpected result, since the DMS and the DLS are received anatomical inputs from frontal and sensorimotor cortices, respectively [59]. We just focused on the proportions of selected models in this analysis, therefore, the amount of each information might be different between the DMS and the DLS. On the other hand, the total proportions of neurons coding space- or motor-related variables were larger in the M1 than in the PL. These results are consistent with well-known roles of cortex; that is the M1 is involved in motor function.

In conclusion, neurons in the cortico-basal ganglia loops including striatum, prefrontal cortex and motor cortex processes multi-modal information in parallel. This observation encourages to record and analyze neural activities with the animal ' s motion.

4. Conclusion and future directions

4.1 Conclusion

Based on the reinforcement theory, I examined physiological functions of the striatum for reward-based learning or decision making. Current major research directions of the striatum are 1) striosome/matrix, 2) subregions and 3) direct/indirect pathway. This thesis covered two topics as the following. Chapter 2 covered the topic of striosome/matrix. Using a cell-type specific calcium imaging method for striosomal neurons, I could record activities of neurons in striosomes during a classical conditioning task and found reward-predictive activities. According to the reinforcement learning theory, these activities encode values of sensory states. Chapter 3 covered the topic of striatal subregions. By conducting electrophysiological experiments of behaving rats and linear regression analysis of recorded neural activities, I found that majorities of striatal neurons parallery encoded multiple modalities. However, I could not find any differences of information coding between the DMS and the DLS unfortunately.

This thesis would give knowledge about neural representations of the striatum during reward-based learning or decision making and a suggestion that the reinforcement learning theory is an useful framework to understand physiological functions of the cortico-basal ganglia circuits.

4.2 Future directions

In the Ca^{2+} imaging experiment of striosomal neurons, there was no concept of action selection since its behavioral task was the classical conditioning. For this reason, I could not approach a question whether reward prediction and action selection are separated between striosomes and matrix. In addition, I could not record matrix neural activities selectively due to technical difficulties at that time. However, a matrix-cre line (Plxnd1-OG1) has been developed recently [37, 38] and become available from mutant mouse research resource center (MMRRC). Using same Ca^{2+} imaging method applied to striosomal neurons, we are able to record matrix neurons selectively. By recording and manipulating striosome/matrix neurons during an operant conditioning task that involves choices between multiple

actions (Figure 4.1), we might be able to answer the question.

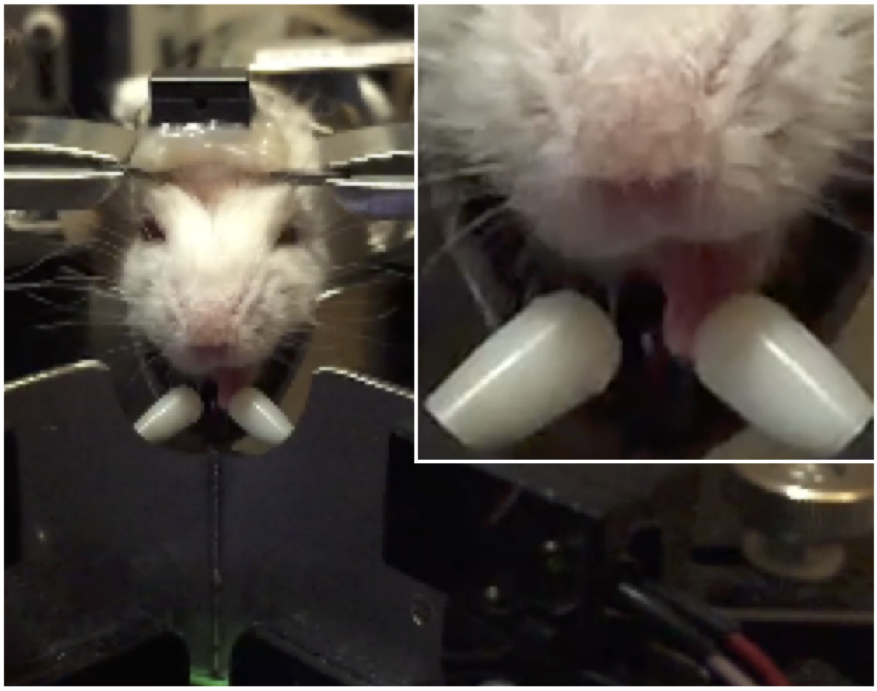

Figure 4.1: Head-fixed dual-licking choice task. Mice choose the left or right by licking spouts and acquire water rewards.

As I mentioned in the previous chapter, it is known that the ratios of D1-MSNs and D2-MSNs were different between striosomes and matrix [5]. I think that selective recording of striosomal and matrix neurons is not enough to future study. I am predicting that one of the next research trends in the field of the striatum is to record or manipulate neurons with discrimination of striosome-D1, striosome-D2, matrix-D1 and matrix-D2. A strategy is to combine transgenic animals and retrograde viruses. These viruses retrogradely infects neurons via the axon terminals. AAV2-retro is one of the retrograde virus and its properties are non-pathogenic and high-efficient infection [63]. It is theoretically possible that we conduct selective recording and manipulation of striosome-D1 and matrix-D1 MSNs by infecting AAV2-retro from SN to striatum (Figure 4.2).

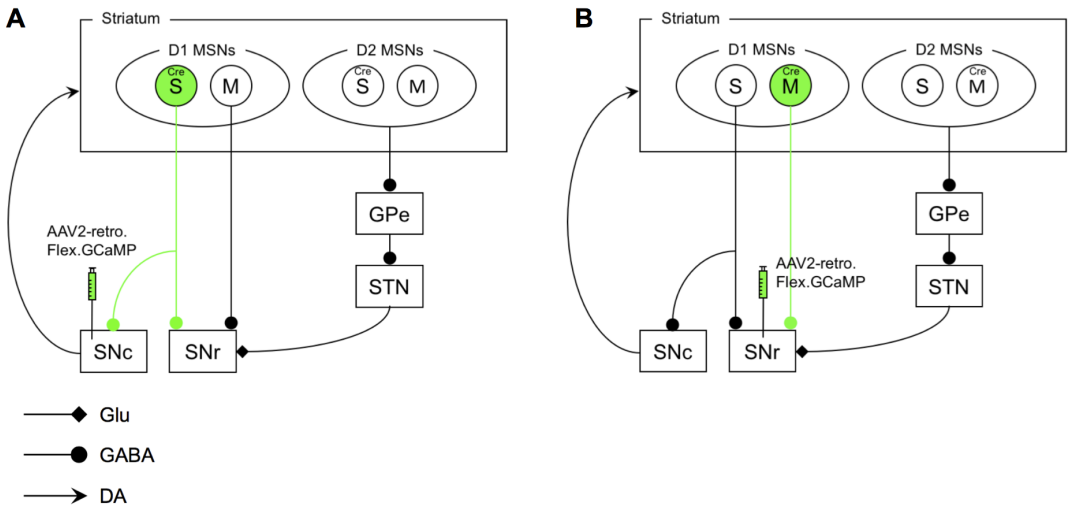

Figure 4.2: Concept of striatal compartment-pathway selective virus infection. (A) Striosomal D1-MSNs selective GCaMP expression. (B) Matrix D1-MSNs selective GCaMP expression. S: striosome, M: matrix, SNc: substantia nigra pars compacta, SNc: substantia nigra pars reticulate, GPe: external globus pallidus, STN: subthalamic nucleus.

With respect to parallel information coding in the striatum, we should next record neurons of the VS. The VS receives inputs from prefrontal cortex [59]. Thus, it is expected that its neurons represent more strongly task-related variables, such as context [13], than the others. On the other hand, neurons in each subarea are not specialized in information processing at a specific modality, and it does not know whether it really affects decision making. Although it is known that lesions of DMS and DLS trigger inhibition of acquisition of goal-directed and habitual behaviors, respectively [15], it is necessary to conduct lesion or optogenetic experiment from the angle of parallel information processing. It is expected that activation or inhibition of neurons in VS, DMS and DLS effect to task initiation, abstract action selection and physical movement, because the hierarchical reinforcement learning model hypothesizes that they represent states or actions at different physical and temporal scales.

Acknowledgements

はじめに、博士論文の審査をご担当いただきました、奈良先端科学技術大学院大学（NAIST）池田 和司教授、金谷 重彦教授、吉本 潤一郎准教授および、沖縄科学技術大学院大学（OIST）銅谷 賢治教授に御礼申し上げます。

本研究は OIST 神経計算ユニットにて実施いたしました。同ユニットを主宰する銅谷 賢治教授に深く感謝をいたします。銅谷先生には、実験計画の立案から実施、実験データの解析、論文執筆に至るすべての段階で数々のご指導を賜りましたこと厚く御礼申し上げます。さらに研究を実施するための場所や予算を充分に与えて下さったことに関しても心より感謝いたします。銅谷先生のもとで博士の学位を取得することは、今後の研究者人生における最大の財産となります。本当にありがとうございました。伊藤 真博士（神経計算ユニット システム生物学グループ 元グループリーダー）には、げっ歯類の行動実験、電気生理実験といった動物実験の手技や、ニューロンデータの解析手法、プログラミング、論文執筆といった研究の基盤となる技術を懇切丁寧にご指導いただきました。また、伊藤博士と実験の合間にした世間話は大変面白く、忙しない日々の中に癒しを与えるオアシスのような存在でした。吉本 潤一郎准教授には、OIST 在職中のみならず NAIST へご栄転されてからも、データの解析やその解釈について統計学的な観点から助言をいただきました。さらに中間発表や学位審査等に関わる事務的な処理も助けていただきました。宮崎 勝彦博士をはじめとするシステム生物学グループの皆様には、経験に頼る部分も多い動物実験を行うに当たり、数々の助言を頂きプロジェクトを前進させることができました。動的システムグループ、適応システムグループの皆様にはラボミーティング等の場面で、機械学習やロボット工学の観点からプロジェクトに対してコメントを頂きました。リサーチアドミニストレーターの安里 恵美子氏、松尾 起久子氏には、研究資材の管理や購入、出張などに関わる事務手続きで多大なご助力を頂きました。

カルシウムイメージング実験で使用した遺伝子組み換えマウス (Sepw1-NP67) は、National Institute of Mental Health, Charles R. Gerfen 博士に、ウイルスベクター (AAV.2/9.Syn.Flex.GCaMP6s, AAV.2/6.Syn.GCaMP6s) は、University of Pennsylvania, Penn Vector Core にご提供いただきました。御礼申し上げます。

最後に、学部から博士課程まで 11 年間にもおよぶ長い時間を学生として過ごしてきた私を、離れた地より応援してくれた両親に感謝いたします。

References

1. Barto, A., Adaptive critics and the basal ganglia. Models of Information Processing in the Basal Ganglia, 1995: p. 215-232.

2. Doya, K., Complementary roles of basal ganglia and cerebellum in learning and motor control. Curr Opin Neurobiol, 2000. 10(6): p. 732-9.

3. Gerfen, C.R., The neostriatal mosaic: compartmentalization of corticostriatal input and striatonigral output systems. Nature, 1984. 311(5985): p. 461-4.

4. Levesque, J.C. and A. Parent, GABAergic interneurons in human subthalamic nucleus. Mov Disord, 2005. 20(5): p. 574-84.

5. Fujiyama, F., et al., Exclusive and common targets of neostriatofugal projections of rat striosome neurons: a single neuron-tracing study using a viral vector. Eur J Neurosci, 2011. 33(4): p. 668-77.

6. Watabe-Uchida, M., et al., Whole-brain mapping of direct inputs to midbrain dopamine neurons. Neuron, 2012. 74(5): p. 858-73.

7. Schultz, W., P. Dayan, and P.R. Montague, A neural substrate of prediction and reward. Science, 1997. 275(5306): p. 1593-9.

8. Samejima, K., et al., Representation of action-specific reward values in the striatum. Science, 2005. 310(5752): p. 1337-40.

9. Czubayko, U. and D. Plenz, Fast synaptic transmission between striatal spiny projection neurons. Proc Natl Acad Sci U S A, 2002. 99(24): p. 15764-9.

10. Jaeger, D., H. Kita, and C.J. Wilson, Surround inhibition among projection neurons is weak or nonexistent in the rat neostriatum. J Neurophysiol, 1994. 72(5): p. 2555-8.

11. Tunstall, M.J., et al., Inhibitory interactions between spiny projection neurons in the rat striatum. J Neurophysiol, 2002. 88(3): p. 1263-9.

12. Doya, K., Metalearning and neuromodulation. Neural Netw, 2002. 15(4-6): p. 495-506.

13. Ito, M. and K. Doya, Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. Curr Opin Neurobiol, 2011. 21(3): p. 368-73.

14. Balleine, B.W., et al., Hierarchical control of goal-directed action in the cortical-basal ganglia network. Current Opinion in Behavioral Sciences, 2015. 5(Supplement C): p. 1-7.

15. Thorn, C.A., et al., Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. Neuron, 2010. 66(5): p. 781-95.

16. O'Doherty, J., et al., Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science, 2004. 304(5669): p. 452-4.

17. Tanaka, S.C., et al., Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat Neurosci, 2004. 7(8): p. 887-93.

18. Ito, M. and K. Doya, Validation of decision-making models and analysis of decision variables in the rat basal ganglia. J Neurosci, 2009. 29(31): p. 9861-74.

19. Kim, H., et al., Role of striatum in updating values of chosen actions. J Neurosci, 2009. 29(47): p. 14701-12.

20. Ito, M. and K. Doya, Parallel Representation of Value-Based and Finite State-Based Strategies in the Ventral and Dorsal Striatum. PLoS Comput Biol, 2015. 11(11): p. e1004540.

21. Jimenez-Castellanos, J. and A.M. Graybiel, Subdivisions of the dopamine-containing A8-A9-A10 complex identified by their differential mesostriatal innervation of striosomes and extrastriosomal matrix. Neuroscience, 1987. 23(1): p. 223-42.

22. Gerfen, C.R., The neostriatal mosaic: striatal patch-matrix organization is related to cortical lamination. Science, 1989. 246(4928): p. 385-8.

23. Eblen, F. and A.M. Graybiel, Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. J Neurosci, 1995. 15(9): p. 5999-6013.

24. Kincaid, A.E. and C.J. Wilson, Corticostriatal innervation of the patch and matrix in the rat neostriatum. J Comp Neurol, 1996. 374(4): p. 578-92.

25. Sutton, R.S. and A.G. Barto, Reinforcement learning. 1998, Cambridge, MA: MIT Press.

26. Reynolds, J.N., B.I. Hyland, and J.R. Wickens, A cellular mechanism of reward-related learning. Nature, 2001. 413(6851): p. 67-70.

27. Shidara, M., T.G. Aigner, and B.J. Richmond, Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. J Neurosci, 1998. 18(7): p. 2613-25.

28. Kawagoe, R., Y. Takikawa, and O. Hikosaka, Expectation of reward modulates cognitive signals in the basal ganglia. Nat Neurosci, 1998. 1(5): p. 411-6.

29. Pagnoni, G., et al., Activity in human ventral striatum locked to errors of reward prediction. Nat Neurosci, 2002. 5(2): p. 97-8.

30. Lau, B. and P.W. Glimcher, Value representations in the primate striatum during matching behavior. Neuron, 2008. 58(3): p. 451-63.

31. Pasquereau, B., et al., Shaping of motor responses by incentive values through the basal ganglia. J Neurosci, 2007. 27(5): p. 1176-83.

32. Pert, C.B., M.J. Kuhar, and S.H. Snyder, Opiate receptor: autoradiographic localization in rat brain. Proc Natl Acad Sci U S A, 1976. 73(10): p. 3729-33.

33. Graybiel, A.M. and C.W. Ragsdale, Jr., Histochemically distinct compartments in the striatum of human, monkeys, and cat demonstrated by acetylthiocholinesterase staining. Proc Natl Acad Sci U S A, 1978. 75(11): p. 5723-6.

34. Herkenham, M. and C.B. Pert, Mosaic distribution of opiate receptors, parafascicular projections and acetylcholinesterase in rat striatum. Nature, 1981. 291(5814): p. 415-8.

35. Jimenez-Castellanos, J. and A.M. Graybiel, Compartmental origins of striatal efferent projections in the cat. Neuroscience, 1989. 32(2): p. 297-321.

36. Tokuno, H., et al., Efferent projections from the striatal patch compartment: anterograde degeneration after selective ablation of neurons expressing mu-opioid receptor in rats. Neurosci Lett, 2002. 332(1): p. 5-8.

37. Gerfen, C.R., R. Paletzki, and N. Heintz, GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. Neuron, 2013. 80(6): p. 1368-83.

38. Smith, J.B., et al., Genetic-Based Dissection Unveils the Inputs and Outputs of Striatal Patch and Matrix Compartments. Neuron, 2016. 91(5): p. 1069-84.

39. Ghosh, K.K., et al., Miniaturized integration of a fluorescence microscope. Nat Methods, 2011. 8(10): p. 871-8.

40. Resendez, S.L., et al., Visualization of cortical, subcortical and deep brain neural circuit dynamics during naturalistic mammalian behavior with head-mounted microscopes and chronically implanted lenses. Nat Protoc, 2016. 11(3): p. 566-97.

41. Ziv, Y., et al., Long-term dynamics of CA1 hippocampal place codes. Nat Neurosci, 2013. 16(3): p. 264-6.

42. Mukamel, E.A., A. Nimmerjahn, and M.J. Schnitzer, Automated analysis of cellular signals from large-scale calcium imaging data. Neuron, 2009. 63(6): p. 747-60.

43. Kirschen, G.W., et al., Active Dentate Granule Cells Encode Experience to Promote the Addition of Adult-Born Hippocampal Neurons. J Neurosci, 2017. 37(18): p. 4661-4678.

44. Okuyama, T., et al., Ventral CA1 neurons store social memory. Science, 2016. 353(6307): p. 1536-1541.

45. Jedynak, J.P., C.M. Cameron, and T.E. Robinson, Repeated methamphetamine administration differentially alters fos expression in caudate-putamen patch and matrix compartments and nucleus accumbens. PLoS One, 2012. 7(4): p. e34227.

46. Cohen, J.Y., et al., Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature, 2012. 482(7383): p. 85-8.

47. Oyama, K., et al., Reward prediction error coding in dorsal striatal neurons. J Neurosci, 2010. 30(34): p. 11447-57.

48. Crittenden, J.R., et al., Striosome-dendron bouquets highlight a unique striatonigral circuit targeting dopamine-containing neurons. Proc Natl Acad Sci U S A, 2016. 113(40): p. 11318-11323.

49. Chen, T.W., et al., Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature, 2013. 499(7458): p. 295-300.

50. Gerfen, C.R., The neostriatal mosaic: multiple levels of compartmental organization in the basal ganglia. Annu Rev Neurosci, 1992. 15: p. 285-320.

51. Piochon, C., et al., Cerebellar plasticity and motor learning deficits in a copy-number variation mouse model of autism. Nat Commun, 2014. 5: p. 5586.

52. Heiney, S.A., et al., Cerebellar-dependent expression of motor learning during eyeblink conditioning in head-fixed mice. J Neurosci, 2014. 34(45): p. 14845-53.

53. Kloth, A.D., et al., Cerebellar associative sensory learning defects in five mouse autism models. Elife, 2015. 4: p. e06085.

54. Jennings, J.H., et al., Distinct extended amygdala circuits for divergent motivational states. Nature, 2013. 496(7444): p. 224-228.

55. Kim, S.-Y., et al., Diverging neural pathways assemble a behavioural state from separable features in anxiety. Nature, 2013. 496(7444): p. 219-223.

56. Friedman, A., et al., A Corticostriatal Path Targeting Striosomes Controls Decision-Making under Conflict. Cell, 2015. 161(6): p. 1320-33.

57. Matsumoto, M. and O. Hikosaka, Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature, 2009. 459(7248): p. 837-41.

58. Pennartz, C.M., et al., Corticostriatal Interactions during Learning, Memory Processing, and Decision Making. J Neurosci, 2009. 29(41): p. 12831-8.

59. Voorn, P., et al., Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci, 2004. 27(8): p. 468-74.

60. Barnes, T.D., et al., Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature, 2005. 437(7062): p. 1158-61.

61. Kimchi, E.Y. and M. Laubach, The dorsomedial striatum reflects response bias during learning. J Neurosci, 2009. 29(47): p. 14891-902.

62. Schmitzer-Torbert, N. and A.D. Redish, Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. J Neurophysiol, 2004. 91(5): p. 2259-72.

63. Tervo, D.G., et al., A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. Neuron, 2016. 92(2): p. 372-382.