Studies of the neural substrates of temporal-difference error in domestic chicks
(ニワトリ雛におけるTD誤差の神経基盤に関する研究)

【序論】Introduction
To ensure survival, animals must update the internal representations of their environment in a trial-and-error fashion. Based on the psychology of animal learning, a variety of reinforcement learning methods have been developed. One such method is temporal-difference (TD) learning (Sutton & Barto, 1998). The TD learning allows agents to update their prediction before they actually receive a reward at its final step. It is therefore useful in a multi-step task, such as foraging in an uncertain environment (animals) or playing a board game (humans).

Electrophysiological studies in mammals have suggested that the dopaminergic (DA) neurons code TD-error, which serve a teaching signal in the TD learning (Schultz et al., 1997). This finding supports the idea that TD learning is actually adopted in living organisms. However, the mechanisms responsible for the computation of the TD-error signal have not been fully understood at the neuronal level. Previous lesion and neuronal recording experiments in chicks have indicated the medial striatum (MSt) is involved in reinforcement learning, thus this is one of the candidates for computing TD error (Izawa et al., 2003, 2005; Ichikawa et al., 2004). To investigate the neural substrates of TD error, I recorded activities in MSt and tegmentum neurons in chicks, and compared these activities with the signals that are theoretically simulated in the TD learning.

【方法】Methods
I trained domestic chicks to associate color cues with delayed food as the reward. Single unit activities were recorded from freely behaving chicks via chronically implanted tetrode. During the recording, the reward was temporally omitted in one of the rewarding trial types, and I examined how the reward omission altered the neuronal activities. Based on the observed alterations, neurons in MSt and tegmentum were classified and compared with the signals simulated based on the TD learning.

I also examined the neuronal connections between striatum and tegmentum, by using anterograde tracer (biotinylated dextran amine, BDA) and retrograde tracer (DiI). The BDA tracing was also combined with immunostaining of DA neurons (using anti-tyrosine hydroxylase, TH) and GABAergic terminals (anti-glutamic acid decarboxylase 65,
GAD65). I paid particular attention to whether the MSt terminals had direct contacts on the DA neurons in tegmentum, and whether the MSt terminals contained GABA as neurotransmitter.

【結果】Results

As the first step, TD learning signals were examined in a mathematical simulation, which mimicked the actual behavioral task of the chick. The simulation consisted of repeated trials, and each trial was a finite sequence composed of five discrete steps of time and a terminal step. Three signals were simulated, namely, (1) target signal for updating, (2) TD-error signal, and (3) reward prediction signal. Here, the TD-error signal (2) is a difference between the target signal (1) and the prediction signal (3). When the reward is omitted, we assumed that these signals will be subsequently updated according to TD(0) method, a simple form of the TD method.

As the second step, neurophysiological recordings were performed. In both MSt and tegmentum, majority of the recorded neurons were classified into one of the three types, based on their fitting to one of the corresponding three statistical models. Specifically, two types of the striatal neurons successfully mimicked, i.e., (1) the target signal and (3) the prediction signal. A linear summation of activities in these two types of striatal neurons was a good fit for the activity of one type of tegmental neurons, which mimicked (2) the TD-error signal. However, the tegmental TD-error signal was also well explained as a summation of type (1) and (3) neurons in tegmentum.

As the third step, I examined the neuronal connections. The BDA-positive terminals descending from MSt formed dense arborizations onto the TH-positive neurons in the medial reticular formation (FRM), substantia nigra (SN), and ventral tegmental nucleus (VTA) in tegmentum. Some of the MSt terminals were also GAD65-positive. Retrograde tracing using DiI indicated the MSt receives ascending afferents from the FRM, SN, and VTA. Reciprocal connections between the MSt and tegmental nuclei were confirmed.

【考察】Discussion

The present study demonstrates a possibility that medial striatum and tegmentum convey the signals that are critical for the TD learning. I assume that a convergent summation of the target signal and the prediction signals in MSt account for the TD-error signal in the tegmentum. However, the direct contact of the GABA-ergic MSt terminals would give rise to an opposite pattern of activities in the tegmental DA neurons. However, anatomical study in pigeons (Anderson et al., 1991) and electrophysiological study in mice (Bocklisch et al., 2013) suggest that MSt neurons may disinhibit tegmental DA neurons via GABA-ergic inhibitory local inter-neurons. The local inhibitory interneurons may finally make the TD error signal in the tegmental DA neurons.