



HOKKAIDO UNIVERSITY

Title	Studies of the neural substrates of temporal-difference error in domestic chicks
Author(s)	温, 琛涛
Degree Grantor	北海道大学
Degree Name	博士(生命科学)
Dissertation Number	甲第12729号
Issue Date	2017-03-23
DOI	https://doi.org/10.14943/doctoral.k12729
Doc URL	https://hdl.handle.net/2115/67152
Type	doctoral thesis
File Information	Chentao_Wen.pdf



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)

A DISSERTATION

submitted to the Graduate School of Life Science,

Hokkaido University

in partial fulfillment of the requirements for the degree of

DOCTOR OF LIFE SCIENCE

Chentao Wen

2016

CONTENTS

1	Introduction	1
2	Material and Methods.....	4
2.1	Abbreviations of neural structures	4
2.2	Subjects and ethical note.....	5
2.3	Apparatus	6
2.4	Behavioral tasks	6
2.5	Recording of single unit activity.....	8
2.6	Pre-processing of spike data	9
2.7	Categorizing neurons based on discrete-time state-space models	10
2.8	Graphical illustration of neuronal activity	13
2.9	Tract tracing by BDA and DiI.....	14
3	Results	17
3.1	Simulation of neuronal representations of temporal-difference learning	17
3.2	MSt neuron activities	18
3.3	Tegmentum neuron activities.....	23
3.4	Linear summation model of the TD error signal	27
3.5	Reciprocal connections between MSt and tegmental DA-ergic neurons.....	28
4	Discussion	29
4.1	Striatal representations of the target signal and the prediction signal	29
4.2	Striatal and tegmental representations of TD error	31
4.3	Tegmental neurons may also contribute to TD-error computation	32
4.4	Neuroanatomical bases of TD error computation	33
4.5	TD learning for updating state value and behavioral execution	34
5	Figures	36
6	Table	63
	Acknowledgements	64
	References.....	65
	List of other publications	73

1 Introduction

To cope with the ever-changing environment, adaptive agents generate an internal representation of the value associated with their present state. Appropriate updating of state value is achieved through trial-and-error and model-free interactions with the environment. Based on the psychology of animal learning, a variety of reinforcement learning methods have been developed in the ongoing search for efficient updating processes. One such method is temporal-difference (TD) learning.

Historically, TD learning has stemmed from the finding of second-order conditioning (also called secondary conditioning) (Pavlov, 1927), which the Rescorla-Wagner rule could hardly explain (Niv and Montague, 2009). Initially, before being paired, the secondary reinforcer has no value. Once paired with a primary reinforcer, such as a food reward, the secondary reinforcer strengthens behavior in a manner that is similar to that of the primary reinforcer. TD learning makes higher-order learning tractable, so that something that signals a predictor will also act as a predictor (Niv and Montague, 2009). According to the canonical formulation (Sutton and Barto, 1998), the value of the preceding state is updated by the TD error that is computed in the current state. The update process is parameterized by a learning rate and a discounting rate (See formulas (3-2) and (3-3) in Results). Here, time represents real time steps rather than a discrete number of trials, as would be assumed in most psychological tasks. In TD learning, agents do not wait until the end of a trial at which point a payoff is finally received. Instead, agents wait until the next time step to update the state value. This feature is particularly useful when agents are required to improve their strategy in a multi-step task, such as foraging in an uncertain environment (animals) or playing a board game (humans). Variant algorithms have been developed to play games such as checkers (Samuel, 1959), backgammon (Tesauro, 1995), and Go (Silver et al., 2016). Some of these have reached or even exceeded the level of expert human players.

In the mid-1990s, new developments in artificial intelligence were accompanied by breakthroughs regarding the neuronal mechanisms of reinforcement learning via a series of studies with rhesus monkeys. In a multiple trial schedule, progress towards a forthcoming reward was represented in the striatum (Shidara et al., 1998) and the

anterior cingulate cortex (Shidara and Richmond, 2002). Action-specific value representations have also been documented in the striatum (Samejima et al., 2005). Schultz *et al.* discovered that the dopaminergic neurons (DA neurons) in the ventral tegmental area (VTA) and the substantia nigra (SN) show a characteristic firing pattern that resembles TD-error signals (Montague et al., 1996; Schultz et al., 1997). In a later study, researchers found that the signs of TD error signals were inverted with respect to those of DA responses in the lateral habenula in monkeys (Matsumoto and Hikosaka, 2007). Furthermore, the TD-error signals were also found in other brain regions; see (Schultz, 2015) for a comprehensive review. These results support the idea that TD learning occurs in living organisms, although how TD-error signals are computed is unclear. More specifically, questions remain regarding: (1) how the primary reinforcement and the current state value, i.e. the target signal, are represented; (2) how the preceding state value, i.e. the prediction signal, is represented; and (3) how these signals are merged to compute TD-error.

I sought to address these questions using domestic chicks as subjects. Within days of hatching, chicks can learn to peck conspicuous visual cues (i.e., a colored bead) to gain an associated reward (food and water) (Matsushima et al., 2003). In neuronal recordings from chicks performing the operant pecking task, reinforced by delayed rewards, both the medial striatum (MSt) (Yanagihara et al., 2001) and arcopallium (Arco, an associative isocortex) (Aoki et al., 2003) have been found to code rewards and prediction of rewards. Particularly, some MSt neurons show a sequence of characteristic burst activities during the cue period, the post-operant delay period, and/or the reward period of the task (Amita and Matsushima, 2014; Izawa et al., 2005). Localized MSt lesions cause impulsive choices (Izawa et al., 2003) while Arco lesions cause cost-averse choices (Aoki et al., 2006), suggesting the involvement of these areas in foraging decision making. However, the functional roles of these multi-phase MSt activities remain unclear.

Several aspects of chick behavior are further puzzling. Generally, once acquired, pecks at rewarding cues are barely extinguished, although MSt lesions partially impair the process of extinction (Ichikawa et al., 2004). Even after the associated reward is

omitted, chicks fail to stop pecking for several hours, as if the state value was not updated. Conversely, in another task in which chicks actively forage between two feeders placed at opposite ends of an I-shaped maze ([Ogura et al., 2015](#); [Ogura and Matsushima, 2011](#)), chicks quickly changed their stay time when the profitability of the feeders changed ([Xin & Matsushima, unpublished](#)). In this case, chicks behave as if the state value was flexibly updated.

In this study, I specifically focused on the neuronal activities that occurred during the reward period when a predicted food was omitted. Specifically, I sought to distinguish the representations of the reward from those of the predicted reward. Preliminary recordings suggested that some MSt neurons could code the prediction even during the reward period, similar to GABAergic neurons in the VTA ([Cohen et al., 2012](#)). As the first step of my analysis, I constructed mathematical simulations of the critical signals of TD learning in the extinction task. As the second step, I classified recorded MSt neurons into three types according to changes in reward-period activity in the omission block. I then compared the simulated TD learning signals and the classified neurons. I then analyzed neurons in the tegmentum around the substantia nigra (SN) and the formatio reticularis medialis mesencephali (FRM) in the same manner. As a third step, I examined the assumed connections from the MSt descending to the tegmentum via tract-tracing combined with immunostaining for thyroxine hydroxylase (TH, a marker of DA-ergic neurons). Based on these results, I propose a novel hypothetical process in which TD learning for foraging behavior is accomplished via interactions between the MSt and the midbrain DA system.

2 Material and Methods

2.1 Abbreviations of neural structures

Ac	nucleus accumbens
APrH	prehippocampal area
BCS	brachium colliculi superioris
BSTl	lateral part of the bed nucleus of the stria terminalis
EW	nucleus of Edinger-Westphal; nucleus nervi oculomotorii, pars accessoria (ICAAN)
FRM	formatio reticularis medialis mesencephalic
GP	globus pallidus
HA	apical part of the hyperpallium
Imc	nucleus isthmi, pars magnocellularis
LSt	lateral striatum
M	mesopallium
MSt	medial striatum
NI	nidopallium intermediate
NIII	nervus oculomotorius
NCL	caudolateral nidopallium
OMd	nucleus nervi oculomotorii, pars dorsalis
Omv	nucleus nervi oculomotorii, pars ventralis
OV	nucleus ovoidalis
PTM	nucleus pretectalis medialis
Ru	nucleus ruber
SN	substantia nigra

SP	nucleus subpretectalis
SpL	nucleus spiriformis lateralis
VP	ventral pallidum
VTa	ventral tegmental area

2.2 Subjects and ethical note

Experiments were conducted according to the guidelines and approval of the Committee on Animal Experiments of Hokkaido University. The guidelines are based on national regulations for animal welfare in Japan (Law for Humane Treatment and Management of Animals; after a partial amendment No.68, 2005). A total of 62 unsexed domestic chicks (*Gallus domesticus*, White Leghorn strain) were used in this experiment; 45 chicks for neuronal recording in a freely behaving condition and 17 for neuroanatomical tract-tracing experiments. Fertilized eggs were purchased from a local supplier (Iwamura Poultry Ltd., Yubari, Japan) and incubated in the laboratory. I also used newly hatched male chicks from the same poultry supplier.

Training started on the day the chicks hatched (post-hatching day 0, or PH0). Prior to undergoing a surgical operation for chronic electrode implantation (PH7–8), pairs of chicks were housed in transparent plastic cages (30 × 17 × 13 cm) in a chamber lit by LED lamps on a 12-h light/12-h dark cycle, with the light phase starting at 09:00 am. After electrode implantation, the chicks were individually housed in the same chamber in transparent plastic cages (29 × 18 × 18 cm), so that chicks were mutually visible. Water was freely available from a drinking bottle, while food was strictly controlled. The restricted diet served to ensure that: (i) chicks actively fed during the behavioral tasks, and (ii) they increased in body weight (BW) gradually such that they reached 45 g or higher on PH7–8.

For the surgical operations, chicks were anesthetized via an intra-muscular injection of ketamine/xylazine cocktail (a 1:1 mixture of 10 mg/ml ketamine [Daiichi Sankyo Co., Ltd., Tokyo, Japan] and 2 mg/ml xylazine [Sigma-Aldrich Co., St. Louis, USA]) at a dose of approximately 0.1 ml per 10 g BW. Supplementary injections (0.1

ml) were given if necessary. When the brain was not sampled, chicks were euthanized via exposure to carbon dioxide.

2.3 Apparatus

I used an operant box ($30 \times 28 \times 38$ cm, illuminated by LEDs and maintained at approximately 25–30 °C) to train chicks and record single neuron activity in the freely behaving condition. The box was made of metal and electrically shielded to reduce noise. The subject chicks were monitored via a CCD camera on the ceiling of the enclosure, which enabled us to observe behaviors without being seen by the chicks. The ceiling was also equipped with a rotary slip ring, which enabled us to connect the implanted electrodes (tetrodes) to differential amplifiers located outside of the box.

The front panel of the box was equipped with a pair of multi-color LEDs placed side by side (3.2 cm apart) for cue color presentation, a pair of holes through which the response bar protruded below the LEDs, and a feeder (a food-dispensing tube and dish) at the center of the panel; see Fig. 2B. The LEDs, response bars, and feeder were driven by a micro-robot (RCX 1.0, Lego Co., Billund, Denmark) controlled via LabView (National Instruments Co., Austin, Texas, USA).

2.4 Behavioral tasks

2.4.1 Habituation and pre-training

On PH0 and 1, pairs of chicks received one habituation session per day in the operant training apparatus. During each 20-min session, a multi-color LED (emitting red, green, or blue light) was circularly lit continuously, changing color in a fixed sequence every 40 s. Grains of millet seed were intermittently delivered in an unpredictable way that was not associated with the color of the lit LED.

On PH2, chicks were pre-trained to associate one of the LED colors (cue1) with a food reward (four grains of millet). In one trial, the LED was lit and the response bar was simultaneously protruded. The lit LED and the response bar were maintained until the chick pecked the bar. When the chick pecked the bar, the LED was immediately turned off and the bar was retracted, and the food was delivered without a delay. I

conducted two pre-training sessions (20 min each) on PH2. In the first session, chicks were trained in pairs. In the second session, chicks were individually trained. After these two sessions, individual chicks were tested for their responses to the simultaneously presented LED and bar in 20 consecutive trials. Those chicks that pecked the bar in 15 or more trials were subsequently trained in the color-food association task.

2.4.2 Color-food association training

On PH3–5, chicks were trained to associate the LED colors with the following food rewards: four grains (cue1), one grain (cue2), and no food (cue3). See [Fig. 1A and B](#) for the schedule of trials and the color-reward associations. The assignment of the LED colors (red, green, or blue) to cue1–3 was randomized among individuals. In each trial, one of the cue LEDs was lit, and its onset was defined as $t = 0$ s. After 0.5 s ($t = 0.5$ s), a response bar protruded for the chick to peck. The chick had 1 s to peck the bar. Irrespective of whether the chick pecked, the LED turned off and the bar was retracted at $t = 1.5$ s. After a delay period of 1.5–2.0 s ($t = 3.0$ – 3.5 s), chicks received the associated food. If chicks did not peck the bar, food was not delivered. If chicks responded incorrectly (no peck at cue1 and cue2, or peck at cue3), correction trials were repeated, with up to five additional trials.

Individual chicks received two training sessions per day. One session comprised 60 trials, excluding correction trials: 20 trials for each cue1, cue2, and cue3, with a pseudo-random order of presentation. Inter-trial intervals ranged from 12 to 20 s. On PH6 or afterward, i.e., after 3 days of training, a final test was conducted. The test procedure was identical to the training sessions on PH3–5, except that no correction trials were given. If chicks pecked the bar in $\geq 17/20$ trials for both cue1 and cue2, and $\leq 10/20$ trials for cue3, they proceeded to the electrophysiological experiment. If chicks failed to meet these criteria on PH6, they were repeatedly trained and tested up to PH8. Those chicks that met the criteria on PH8 also proceeded to the electrophysiological experiment.

2.4.3 Behavioral task during electrophysiological recording

To investigate how the neuronal correlates of food reward are updated, I recorded extracellular single unit activities from freely behaving chicks performing a food omission task ([Fig. 1B](#)). The recording session comprised an initial control block followed by an omission block. Cue1 was associated with four grains of food in the control block. In the omission block, food was omitted (in 42 chicks) or delivered after a longer delay period of 3.5 s (in three chicks). After electrophysiological recording in the omission block, chicks were re-trained in the same condition as the control block (termed the reacquisition block) before the next recording session. If not stated otherwise, the association for cue2 and cue3 did not change.

2.5 Recording of single unit activity

2.5.1 Chronic implantation of tetrode

I recorded neuronal activity using tetrodes, which were hand-made by twisting 4 formvar-insulated nichrome wires (bare diameter: 18 μm ; coated diameter: 25 μm ; A-M System Co., Sequim, Washington, USA). The tip of each tetrode was gold-plated and its resistance was reduced to 100-300 k Ω when measured at 1 kHz in a saline solution. I used a metal electrode impedance tester (Model IMP-2, Bak Electronics, Inc., Umatilla, Florida, USA) for the impedance measurements. The plated tetrodes were inserted in thin stainless steel tubes, implanted into the brain tissue, and connected to a micro-driver.

On PH7 or 8, chicks were anesthetized as described above. The anesthetized chicks were fixed on a rat stereotaxic apparatus (type SR-5N, Narishige, Co. Tokyo, Japan) modified such that it was possible to secure the beak of a chick. Using micromanipulators (type SM-15M), a tetrode was inserted into either the medial striatum or midbrain tegmentum. The coordinates of the tetrode tips are shown in [Table S1](#). After the tetrode reached the coordinates, the micro-driver was chronically fixed to the skull surface with dental cement, allowing us to gradually insert the tetrode.

2.5.2 Amplifiers for extracellular recording

Recording started on the day after tetrode implantation. Neuronal signals were buffered by a head-amplifier (FET input operational amplifier, TA75074F, Toshiba, Tokyo, Japan) and then amplified by an AC-coupled differential amplifier. The cut-off frequency was set at 0.3 kHz, amplification $\times 2,000$, and the band-pass filter was set at 0.5–1.5 kHz (18 dB per octave). Signals were A/D-converted at a sampling rate of 16.6–25.0 kHz (Micro1401, CED Co., Cambridge, UK) and stored in a PC.

2.5.3 Histological examination of recording sites

After the recording experiment, chicks were given an overdose of ketamine/xylazine cocktail (0.6–0.7 ml of a 1:1 mixture) and transcardially perfused with a fixative (4% paraformaldehyde in 0.1 M PB; PFA). The entire brain was dissected out and post-fixed for up to 1 week in the same fixative at 4 °C. The brain tissue was then trimmed, embedded in egg yolk, and fixed for an additional 3 days. The tissue was subsequently cut into a complete series of 50- μ m-thick frontal sections using a vibrating microtome (DTK-3000, Dosaka Co., Kyoto, Japan). Sections were mounted on glass slides coated with APS (amino silane), stained with cresyl violet, cover-slipped, and examined using a microscope and a drawing tube. The recording sites were estimated based on the complete reconstruction of tetrode tracks and record of tetrode advancement. The coordinates conformed to the chick brain atlas ([Kuenzel and Masson, 1988](#)), and neural nuclei terminology conformed to the nomenclature reform ([Reiner et al., 2004](#)).

2.6 Pre-processing of spike data

I performed initial off-line analysis using Spike-2 software (CED Co., Cambridge, UK). To extract single unit activity, a threshold was arbitrarily determined so that the amplitude of the peak (or trough) of spikes exceeded the background noise by 2-fold or more. Spikes were further sorted using a template-matching method and stored as smr files, which were exported to Matlab (version 2009b, The MathWorks Inc., USA) via a library SON provided by CED company.

Spikes were timed in reference to the cue onset as $t = 0$ s in each trial. Spikes were thus counted in each of 100 ms-long bins between $t = -5 - 15$ s, hence there were 200 bins in total for each trial. Trials were sequentially numbered as T , with the first trial of the omission block denoted as $T = 1$. I measured firing rate (spikes per s) during the reward period (for 2 s or 20 bins during $t = 3-5$ s, if not stated otherwise), denoted as FR_{R_T} for each T . The FR_{R_T} was further normalized as a ratio to the corresponding baseline firing rate (FR_{Base} during the period preceding the cue presentation, $t = -5 - 0$ s) as:

$$FR_{R_Nor_T} = \frac{FR_{R_T}}{lowess(FR_{Base})_T} \quad (2-1)$$

Here, to smooth the fluctuating baseline activity along the procession of trials, I used a built-in R function (`lowess`) to adopt a robust locally-weighted regression method. In the following, I denote $FR_{R_Nor_T}$ as Y_T for simplicity. $Y_T = 1$ means that the reward period was indistinguishable from the baseline inter-trial period in terms of neuronal firing. This normalized reward-period activity Y_T is shown in e.g. [Fig. 4A\(c\)](#) and [4C](#).

2.7 Categorizing neurons based on discrete-time state-space models

I categorized neurons based on their activity during the omission block. To this end, I used the normalized reward-period activities (Y_T) in the cue1 trials. I constructed three types of state-space models; the actual reward model (AR), reward prediction model (RP), and prediction error model (PE). Following the framework employed by [Petris et al. \(2007\)](#), I assumed that each model had one hidden state variable (θ_T) along the discrete trial number, T . The firing rate Y_T thus stems from the hidden θ_T in each trial (see each model below). I estimated the parameters of the models using the maximum likelihood method, and chose the best one with the smallest AICs (Akaike Information Criteria) ([Akaike, 1974](#)). For model fitting and statistical analysis, I used R (version 3.1.3, R, The R Foundation for Statistical Computing, Vienna, Austria, <http://www.r-project.org>) in addition to MatLab.

2.7.1 Actual Reward (AR) model

In this model, I assumed the neurons to be a sensory representation of the actual reward (AR) in each trial, regardless of whether the food was perceived via visual, acoustical, or gustatory clues (i.e. r_4). The neuronal activity returns to the baseline level as soon as the food is omitted (therefore $\theta_T = 1$ for $T \geq 1$), and Y_T is given as:

$$Y_T = 1 + \epsilon \quad (T \geq 1) \quad (2-2)$$

This model therefore does not have a free parameter. Here and the following models, ϵ is zero-mean Gaussian-distributed noise. The AR model corresponds to the target signal (type-1), specifically in the reward period (Fig. 2C).

2.7.2 Reward Prediction (RP) model

In this model, I assumed the neurons to represent reward prediction (RP) during the reward period of the trials (i.e. $\hat{V}(S_3)$). Therefore, the hidden variable is $\theta_T = p_T$ (for $T \geq 1$). Following the updating rule (3-2') in the results of the main text, the p_T asymptotically approaches 1 (meaning no predicted reward) as the omission block proceeds. The observed Y_T is thus given as:

$$Y_T = p_T + \epsilon \quad (T \geq 1) \quad (2-3)$$

p_T is given as follows. The initial value of p_T (p_0) is given by the observed Y_T in the last trial of the control rewarding block, after smoothing via the lowess function.

$$p_T = p_0 \quad (T = 1) \quad (2-4-1)$$

p_T is subsequently updated at every trial by the product of a learning rate $\alpha \in [0,1]$ and the term representing the prediction error, $(1 - p_{T-1})$.

$$p_T = p_{T-1} + \alpha \cdot (1 - p_{T-1}) \quad (T \geq 2) \quad (2-4-2)$$

I may reasonably assume that each neuron has a distinct α value as a free parameter. If α is high and close to 1, quick changes will emerge. Alternatively, if α is 0, p_T will remain unchanged throughout the omission block. The RP model corresponds to the prediction signal (type-3), specifically in the reward period (Fig. 2C).

2.7.3 Prediction Error (PE) model

In this model, I assumed the neurons to represent the prediction error (PE) in the reward period (i.e. δ_4). Similar to the RP model, I assumed the prediction error pe_T to be the hidden variable θ_T , and the observed firing rate Y_T is given by pe_T as follows:

$$Y_T = 1 + pe_T + \epsilon \quad (T \geq 1) \quad (2-5)$$

In the first trial of the omission block ($T = 1$), pe_T drops considerably (ΔR , reward difference), because the animal does not gain the food that the animal learned to expect during the control block.

$$pe_T = \Delta R + (1 - \alpha) \cdot pe_0 \quad (T = 1) \quad (2-6-1)$$

In subsequent trials ($T \geq 2$), pe_T will be updated as follows:

$$pe_T = (1 - \alpha) \cdot pe_{T-1} \quad (T \geq 2) \quad (2-6-2)$$

(2-6-1) and (2-6-2) were derived from the updating rule (2-4-2) in the RP model. See [Box 1](#) below for details. The PE model has two free parameters, ΔR and α . The PE model corresponds to the TD-error signal (type-2), specifically in the reward period ([Fig. 2C](#)).

2.7.4 Fitting neuronal data to the models and selecting the best model

I tested the fit of these three models for the recorded data from each neuron. The free parameters were estimated using the maximum-likelihood estimation method. In the RP model, α was estimated using the built-in function (optimize) in the R programming environment. In the PE model, the two free parameters (α and ΔR) were estimated using the optimx function in the optimx package ([Nash, 2014; Nash and Varadhan, 2011](#)). The function (optimx) supplied us with a set of different estimates (14 sets) after applying 14 different optimization algorithms. I chose the one with the highest likelihood among these estimates. To avoid inappropriate estimates, such as local maxima, I graphically checked the fitting curves.

Box 1. Derivation of PE model

By definition:

$$pe_T = R_T - p_T \quad (b1)$$

This is identical to:

$$p_T = R_T - pe_T \quad (b2)$$

According to RP model (2-4-2):

$$p_T = p_{T-1} + \alpha \cdot pe_{T-1} \quad (b3)$$

By substituting p_T and p_{T-1} in (b3) with (b2), I get:

$$R_T - pe_T = R_{T-1} - pe_{T-1} + \alpha \cdot pe_{T-1} \quad (b4)$$

To solve for pe_T , I get:

$$pe_T = R_T - R_{T-1} + (1 - \alpha) \cdot pe_{T-1} = \Delta R + (1 - \alpha) \cdot pe_{T-1} \quad (b5)$$

After fitting all three models, I sought to select the best one. I calculated Akaike's information criterion (AIC) for the three models for each neuronal activity data set. The model that yielded the smallest AIC value was thus chosen for the categorization of the neuron under study. To be strict, I added to following criteria post-hoc. Among those neurons classified as AR and RP neurons, the Y_T data sets (normalized firing rate during the reward period) were examined for significant activity in the control block. Neurons were ruled out when one-sample t -test failed to detect a statistical significance at $p > 0.05$, and the neuron was classified as being an 'other' type. The same test allowed us to further classify the neurons (irrespective of AR, RP, and PE type) into two groups: excitation type (positive t -value) and inhibition type (negative t -value).

2.8 Graphical illustration of neuronal activity

To graphically illustrate the averaged neuron activities, I calculated the z-scores of the spike-counts plotted against the 100-ms bins, averaged across trials. The z-score in the i -th bin is given as:

$$z_i = \frac{\text{mean}(C_{i,T}) - \text{mean}(C_Base)}{SD(C_Base)/\sqrt{n}} \quad (2-7)$$

Here, the $C_{i,T}$ is the spike count in trial T in the i -th bin. The C_Base is the set of spike-counts during the period before cue onset (50 bins, $t = -5$ to 0 s). Data obtained for each trial type (e.g. cue1 trials in the control block) were included. I calculated the $\text{mean}(C_Base)$ and the $SD(C_Base)$ across all bins in the corresponding trial type. SD represents standard deviation, and n denotes the number of included trials. The number n was fixed constant ($n = 20$) to reliably compare among different trial types; I excluded neurons if I had recording data for fewer than 20 trials. With this adjusted z-score, positive values indicated the excitatory response and negative values the inhibitory response, in reference to the baseline activity. z-scores above 1.96 or below -1.96 implied a significant excitatory or inhibitory response ($p < 0.05$; no adjustment for multiple comparisons) against the baseline firing rate.

2.9 Tract tracing by BDA and DiI

To reveal the efferent terminals from the MSt, I used biotinylated dextran amine (BDA, 0.1 μ l per injection, 10% in distilled water, 10 kDa; D22910, Molecular Probes®, Thermo Fisher Scientific Inc., USA) as an anterograde tracer. To reveal the MSt neurons projecting to the SN, I used 1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate (DiI, 30 nl per injection, 7% solution in N,N-Dimethylformamide) as a retrograde tracer. I used a micro injection instrument (Nanoject II, Drummond Scientific Co., Broomall, Pennsylvania, USA) to inject the tracer into chicks aged approximately PH9. The injection was performed under ketamine/xylazine anesthesia, as described above. Either 7 days (BDA) or 11 days (DiI) after the operation, chicks were transcardially perfused with 4% PFA. Brains were dissected out and post-fixed in the same fixative at 4 °C overnight (BDA) or for ≥ 3 days (DiI).

2.9.1 Histochemistry for visualizing anterograde labeling with BDA

After 1 day of cryo-protection in PBS with 20% sucrose, the brains were frozen and stored at -30°C until sectioning. I used a sliding microtome with a freezing stage (TU-213, Yamato Kohki Industrial Co. Ltd., Saitama, Japan) to cut the brains into sagittal sections for single or double histochemical labeling.

For single labeling to visualize BDA, 60- μm -thick sections were cut and incubated in avidin-biotinylated horseradish peroxidase complex reagent (PK-6100, Vectastain® Elite ABC Kit, Vector Laboratories Co., USA) and DAB (SK-4100, DAB Peroxidase Substrate Kit, Vector Laboratories) as a chromogen. Sections were mounted on APS coated glass slides (S8441, Matsunami Glass Ind. LTD., Osaka, Japan), cover-slipped in Permount™ mounting medium (SP15-500, Thermo Fisher Scientific Inc., USA), and stored at room temperature.

For double labeling, 24- μm -thick sections were initially soaked in Alexa Fluor®488 streptavidin conjugate at room temperature for one hour to visualize BDA (S32354, Molecular Probes®; dilution by 1:400). The sections were then processed with a primary antibody; rabbit anti-TH (1:1000, 4°C , overnight; AB152, Chemicon®, EMD Millipore Co., USA) or rabbit anti-GAD65 (1:1000, 4°C , 3 days; bs-0400R, Bioss Inc., USA). As the secondary antibody, I used goat anti-rabbit IgG - Alexa Fluor® 568 conjugate (1:400, A11011, Molecular Probes®) at room temperature for 1 hour. Sections were then mounted on APS coated glass slides and cover-slipped in Prolong® Diamond antifade mountant with DAPI (P36962, Thermo Fisher Scientific Inc.) and stored at room temperature.

2.9.2 Procedure for retrograde labeling with DiI

The fixated brains were embedded into yolk, post-fixated in 4% PFA for an additional ≥ 3 days, and cut into 50- μm -thick sections using a vibrating microtome (DTK-1000). Sections were collected in PBS, mounted onto APS coated glass slides, and cover-slipped in PBS. The cover glass was sealed with a transparent nail polish to prevent drying.

2.9.3 Microscopic observations

At low magnification, stained sections were photographed using a bright-field light microscope (Olympus BH-2) and fluorescence microscopes (Leica MZ16F, filter set DsRED, for DiI; EVOS® FL Imaging System, light cubes GFP/RFP, for BDA/anti-TH). I used a confocal microscope (Zeiss LSM 510) to examine the connectivity between BDA-positive terminals and tegmentum neurons at a high magnification. The combination of beam splitters and filters is as follow: (1) Main dichroic beam splitter: HFT 405/488/561; (2) Secondary dichroic beam splitters: NFT 565 and NFT 490; (3) Filters: 572-690 for anti-TH/GAD65, BP 505-550 for BDA, BP 420-480 for DAPI. Scanned images were examined using Zeiss LSM 5 Image. Images of interest were edited using a free graphics editor GIMP2.8 (GNU Image Manipulation Program; URL: <https://www.gimp.org/>).

3 Results

3.1 Simulation of neuronal representations of temporal-difference learning

To simulate critical signals involved in TD learning, I assumed a discrete step-time procedure that mimicked the behavioral task (Fig. 2A). I adopted an algorithm that followed the standard formulation of the one-step TD method (TD(0) method) (Sutton and Barto, 1998). In this simulation, a trial is a finite sequence composed of states S_0, S_1, S_2, S_3, S_4 and $S_{terminal}$, corresponding to a pre-trial period ($t = 0$), cue period (1), peck-operant period (2), delay period (3), and reward period (4), respectively, followed by the terminal. At the transition to each state S_t , reward is received ($r_t = 1$) or not ($r_t = 0$). In the control block, as the reward was delivered in the reward period, I set $r_4 = 1$ and $r_t = 0$ when $t \neq 4$, with an arbitrary unit (Fig. 2B).

Generally, the state value $V(S_t)$ is given by the expected sum of the discounted future rewards after S_t , such as:

$$V(S_t) = E[r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots] \quad (3-1)$$

For simplicity, I hypothesize no temporal discounting, so $\gamma = 1$. $V(S_t)$ is hidden, and subjects must learn to estimate it through experience. In Fig. 2B, $\hat{V}(S_t)$ denotes the subjective estimate of $V(S_t)$. I assumed that subjects had been fully trained, so $\hat{V}(S_t) = V(S_t)$ in the control block. I therefore assume $V(S_t)$ (and thus $\hat{V}(S_t)$) = 1 for $t = 1$ to 3 (Fig. 2B). In the first trial of the omission block, $\hat{V}(S_t)$ is equal to that in the control block (Fig. 2C), even though r_4 turns = 0 (Fig. 2C). In subsequent trials, $\hat{V}(S_t)$ is gradually updated according to the TD(0) method, so that:

$$\hat{V}(S_{t-1}) \leftarrow \hat{V}(S_{t-1}) + \alpha \delta_t \quad (3-2)$$

where $\alpha \in [0,1]$ is the learning rate. In this scheme, α is set as 0.03. The TD error δ_t is given by:

$$\delta_t = r_t + \gamma \hat{V}(S_t) - \hat{V}(S_{t-1}) \quad (3-3)$$

I assume that neurons in the medial striatum and tegmentum represent the critical signals in the formula (3-3) (Fig. 2B, C). Thus, in addition to the target of TD learning $r(t) + \gamma \hat{V}(S_t)$, predicted rewards are also represented in terms of delayed and inhibitory activity in the form of $-\hat{V}(S_{t-1})$. In other words, reward prediction signal can also appear in the reward period (i.e. $-\hat{V}(S_3)$), and is represented as inhibition, or suppressed neuronal activity. As formula (3-3) indicated, the simple summation of these two signals will yield the TD error signal δ_t . In the following, these signals are referred to as type-1, -2, and -3. I compared these signals: $r_t + \gamma \hat{V}(S_t)$, $-\hat{V}(S_{t-1})$, and δ_t , with activities recorded from neurons in the medial striatum (Figs. 4, 5) and tegmentum (Figs. 7, 8). I paid particular attention to the characteristic temporal patterns of neuronal activities in each trial, and their changes in the omission block.

In the control block (Fig. 2B), the type-1 signal is turned on at the cue period, and shows sustained activity through the peck/delay/reward periods. The activity of the type-2 signal is similar, but the delayed inhibition is due to the type-3 signal which cancels out the peck/delay/reward period activities, while the initial transient activity still remains as the TD error δ_t . In the omission block (Fig. 2C), as the reward signal r_4 turns = 0, both type-1 and -2 signals will show a rapid drop in activity during the reward period. Type-1 activity will drop to the level of baseline, whereas type-2 activity will drop below baseline. Conversely, activity in type-3 signals will remain unchanged in the first trial of the omission block. In subsequent trials, the activities of all three types of signals will be gradually updated. Specifically, the state value in the reward period will be updated as follows:

$$\hat{V}(S_3) \leftarrow \hat{V}(S_3) + \alpha \delta_4 \quad (\alpha = 0.03) \quad (3-2')$$

Based on the updating rule (3-2'), I simulated 3 dynamic values: r_4 , δ_4 , and $\hat{V}(S_3)$ in the omission block. I constructed 3 corresponding statistical models for classifying real neurons: the Actual Reward (AR) model for type-1, Prediction Error (PE) model for type-2, and Reward Prediction (RP) model for type-3. See the section 2.7 for details regarding the statistical models.

3.2 MSt neuron activities

3.2.1 Habitual pecking responses in the omission block

Throughout the initial part of the omission block, for at least 20 trials, the chicks pecked the response bar for both cue1 and cue2. In Fig. 1C, the percentage of sessions in which the subject chick pecked (y-axis) was plotted against the trial number (x-axis; -19 to 0 in the control block, 1 to 20 in the subsequent omission block). The figure shows data obtained from 12 chicks in 23 recording sessions in which neuronal activities were successfully recorded. In the omission block, for both cue1 (red) and cue2 (green) trials, chicks pecked the response bar in $\approx 90\%$, even though food was omitted for cue1 but not for cue2. In contrast, the pecking response in the cue3 trials (non-rewarding trials) monotonically decreased through both blocks (blue). I therefore assumed that the pecking response was habituated at least for the initial 20 trials in which neuronal activities were recorded. In the following, if not stated otherwise, I analyzed activity during the initial 20 trials of the omission block.

3.2.2 General properties of MSt neurons

I recorded 88 neurons in 68 recording sessions from 30 chicks. Histological examination revealed that these neurons were located in the MSt, NI, and M (Fig. 3A, B). In the present study, I focused on MSt neurons, and pallium neurons were disregarded. In 34 out of the 49 neurons in the MSt, I recorded activity for a sufficiently long duration, enabling us to classify the neurons by type. Of 29 neurons, I found 13 to be type-1, 6 to be type-2, and 10 to be type-3; 5 neurons failed to match any model, and were assigned to an ‘other’ category. A one-way unbalanced ANOVA revealed a difference in the laterality of the recording sites among the three types ($F = 4.34$, $df = 2$, $p = 0.0236$). Type-1 neurons were located more laterally than type-3 neurons according to a *post hoc* Tukey test ($p = 0.0228$; type-1: 1.92 ± 0.58 mm; type-3: 1.28 ± 0.55 mm, mean $\pm$ SD). In the anterior-posterior level, I found no significant difference among the three types (ANOVA: $F = 1.08$, $p = 0.355$). In terms of baseline firing rate and spike width, an ANOVA revealed no significant difference among the three types (Fig. 3C, firing rate: $F = 0.28$, $p = 0.755$), (Fig. 3D, spike width: $F = 1.02$, $p = 0.375$). In the following sections (Figs. 4 and 5), I show the neuronal activities in terms of (1) z scores

of the averaged firing rate in each block, and (2) temporal changes in the normalized firing rate in the reward period, plotted across trials.

3.2.3 Type-1 neurons

Figure 4A shows a representative example. In the control block, after a brief period of transient activity after cue1 onset, tonic responses appeared in the delay period and the reward period. In the omission block, the reward-period response disappeared immediately from the first trial. The cue1 activity in the reward period was identical to that of cue3, in which the food reward was also absent. I thus assumed that this neuron coded the actual reward in the reward period. The actual reward model gave rise to the smallest AIC (inlet table), thus confirming my assumption as type-1. In contrast, the transient activity after cue onset and the tonic activity in the delay period remained in the omission block, although I observed a slight decrease in amplitude. This feature supports the idea that the neuron also coded reward prediction prior to food delivery. However, this neuron also fired, although weakly, in the delay period of the cue3 trials in which no reward followed. Despite this, the simulated target signal $r_t + \gamma \hat{V}(S_t)$ (Fig. 2) was a good fit for this neuron.

I categorized a total of 13 MSt neurons as type-1 neurons based on their reward-period activities; 11 neurons showed excitation and two showed inhibition during the reward period. Of these 11 excitation subtype neurons, I averaged the activities of eight neurons for their firing rate (z scores, Fig. 4B). The other three neurons were not included as the recording time was less than 20 trials in the omission block. The rapid drop in reward-period activity was also reproduced (Fig. 4C). The actual-reward activity was preceded by a reward-prediction signal during the delay period. Note that the delay-period activity declined in the omission block, as indicated by the upward arrow. In contrast, no decline was observed in the cue/peck periods. In an example neuron, shown in Fig. S1, a normalized firing rate in the delay period (1.5–3.0 s) gradually declined during the omission block, suggesting a gradual change in the reward prediction, in concert with the simulated $r_t + \gamma \hat{V}(S_t)$ signal (Fig. 2C). Two neurons showed inhibition during the reward period. These neurons showed complex firing patterns that were markedly different from those of the other neurons (see Fig. S2).

3.2.4 Type-2 neurons

Figure 4D shows a representative example. In the control block, similar to the type-1 example, transient activity upon cue1 onset was followed by tonic responses in the delay and reward periods. However, in the omission block, the reward-period activity changed its sign immediately from excitatory to inhibitory. Cue3 trials were not conducted in this recording. I thus assumed that this neuron coded the negative prediction error in the reward period. The prediction error model gave rise to the smallest AIC (inlet table), thus confirming the status of type-2 neuron. The sign of the tonic activity in the delay period was also inverted in the omission block. However, the transient activity after the cue onset remained. This feature indicates that the neuron also coded reward prediction in the cue period. The prediction signal in the cue period and the negative prediction error signal in the reward period fit the simulated TD-error signal δ_t . However, the excitatory response to the predicted-reward conflicted with the simulated δ_t (Fig. 2B).

I categorized a total of six MSt neurons as type-2 based on their reward-period activities. All six neurons showed excitation during the reward period. Of these six neurons, I averaged the firing rate of four neurons (z scores, Fig. 4E); the other two neurons were not included due to insufficient recording time. The rapid drop and inverted sign of the reward-period activity were reproduced (Fig. 4F). Similar to the type-1 neurons, I observed the reward-prediction signal during the delay period, and found that this signal declined in the omission block (upward arrow). In contrast, cue/peck period activity did not decline. The rapid drop, inverted-sign, and declined delay-period activity fit the simulated δ_t (Fig. 2C). However, both (1) the excitatory reward-period activity in the cue1 trials in the control block and (2) the inhibitory reward-period activity in the cue3 trials in the control block conflicted with the simulated δ_t (Fig. 2B), which showed no response in the reward period for these two cases.

3.2.5 Type-3 neurons

A total of 10 MSt neurons were categorized as type-3 neurons based on their reward-period activities: six neurons showed excitation and the other four neurons showed inhibition during the reward period.

3.2.5.1 Excitation subtype.

Figure 5A shows a representative example. In the control block, tonic responses appeared in the peck, delay, and reward periods. In the omission block, the reward-period response gradually disappeared. The cue1 activity was still higher than the cue3 activity in the delay and reward periods, but it was lower than the cue1 activity in the control block. This neuron was thus assumed to code for reward prediction both prior to and after food delivery. The reward prediction model gave rise to the smallest AIC (inlet table), thus confirming the status of type-3 neuron. In cue3 trials with no reward, this neuron also fired in the delay period, but much more weakly. The simulated prediction signal $\hat{V}(S_{t-1})$ (Fig. 2) fit this neuron well.

Of the six excitation subtype neurons that I found, I averaged the firing rate of five neurons (z scores, Fig. 5B); the other one neuron was discarded due to insufficient recording time. The gradual decrease in reward-period activity was reproduced (Fig. 5E, green). The reward-prediction activity was preceded by the reward-prediction signal during the cue/peck periods, which remained in the omission block. However, the delay-period response was absent. This conflicts with the simulated $\hat{V}(S_{t-1})$ signal (Fig. 2).

3.2.5.2 Inhibition subtype

Figure 5C shows a representative example. In the control block, a brief inhibitory transient activity in the cue period was followed by inhibitory tonic responses in the delay and reward periods. In the omission block, the reward-period response remained and gradually disappeared. In contrast, the cue3 activity in the reward period was near baseline levels. I thus expected this neuron to code for reward prediction in the reward

period. The reward prediction model gave rise to the smallest AIC (inlet table), thus confirming the status of type-3 neuron. In contrast, the transient activity after cue onset and tonic activity in the delay period remained in the omission block, with nearly the same amplitude. This finding supports the idea that the neuron also coded the reward prediction prior to the food delivery. This neuron showed nearly no response in cue3 trials in which no reward followed the cue. The simulated prediction signal $-\hat{V}(S_{t-1})$ (Fig. 2) fit this neuron well.

I tested the response of this neuron when the number of grains of food increased from 1 to 4 in cue2 trials. The amplitude of the inhibitory response in the reward period gradually increased, although the speed of this increase was quicker than the decrease in the omission block (Fig. S3). This result supports my expectation that reward-period activity in this neuron codes reward prediction.

I then averaged the firing rate of the four inhibition subtype neurons (z scores, Fig. 5D). I found that the gradual change in reward-period activity in the omission block was reproduced (Fig. 5E, black). The reward-prediction activity was preceded by a reward-prediction signal during the delay period, which remained in the omission block. The weak cue/peck period responses also remained in the omission block. All of these features are consistent with the simulated $-\hat{V}(S_{t-1})$ signal (Fig. 2).

3.3 Tegmentum neuron activities

3.3.1 General properties of tegmentum neurons

I recorded 39 neurons in 36 recording sessions from 15 chicks. Histological examination revealed that these neurons were located in the SN, FRM, and other regions rich in DA neurons (Fig. 6A). In 25 out of the 39 neurons, activity was recorded for a sufficiently long period, enabling us to classify the neuronal type. Of the 25, 14 were type-1, 4 were type-2, and 3 were type-3. Four neurons failed to match any model, and were assigned to an ‘other’ category. The type-1 neurons were widely distributed in all five anterior-posterior levels. The type-2 neurons were found in the [A4.8] and [A4.4] levels. The type3 neurons were sparsely distributed in the [A4.8] and [A3.2] levels. In terms of baseline firing rate and spike width, an ANOVA revealed no significant

differences among the three types (Fig. 6B, firing rate: $F = 0.31$, $p = 0.735$), (Fig. 6C, spike width: $F = 0.63$, $p = 0.543$). The tegmentum neurons were recorded in the omission condition or in the delay condition (i.e., the delay period increased by 2 s). In the following sections (Figs. 7 and 8), I describe the neuronal activities in these 2 conditions separately.

3.3.2 Type-1 neurons

3.3.2.1 Omission condition

Figure 7A shows a representative example. In the control block, tonic responses appeared in the cue/peck period and the reward period. In the omission block, the reward-period activity disappeared immediately from the first trial onwards. In the following reacquisition block, the reward-period activities reappeared within the first trial. I found no cue3 activity in the reward period. I thus expected this neuron to code for the actual reward in the reward period. The actual reward model gave rise to the smallest AIC (inlet table), thus confirming the status of type-1 neuron. Conversely, the tonic cue/peck period activity remained in the omission block. The cue3 activity in the cue/peck period was weaker than that for cue1. These features support the idea that the neuron also coded the reward prediction prior to the food delivery. However, this neuron lacked a response in the delay period, which conflicts with the simulated target signal $r_t + \gamma \hat{V}(S_t)$ (Fig. 2).

3.3.2.2 Delay condition

Figure 7B shows a representative example. In the control block, two tonic responses appeared in the cue/peck period and the reward period. In the omission (delay) block, the reward-period response disappeared immediately from the first trial onwards. Simultaneously, a novel reward-period response appeared during the new food-delivering phase (5.0–7.0s). I found no cue3 activity in all periods. This neuron was thus expected to code for the actual reward in the reward period. The actual reward model gave rise to the smallest AIC (inlet table), thus confirming the status of type-1 neuron. In contrast, the tonic cue/peck period activity remained in the omission (delay) block, thus supporting the idea that the neuron also coded the reward prediction prior to

the food delivery. However, the lack of a delay-period response conflicts with the simulated target signal $r_t + \gamma \hat{V}(S_t)$ (Fig. 2).

Fig. S4A shows the averaged firing rates of 10 excitation subtype type-1 neurons. The data for neurons recorded in the omission and delay conditions were grouped together for the period before $t=5$ s (left figure) and grouped separately afterwards (right figure).. The rapid drop in reward-period activity was reproduced (Fig. 7C). I thus expected these neurons to code the actual reward in the reward period, and to code the reward prediction in the cue/peck period. However, the lack of response in the delay period conflicts with the simulated $r_t + \gamma \hat{V}(S_t)$ signal (Fig. 2).

3.3.3 Type-2 neurons

3.3.3.1 Omission condition

Figure 7D shows a representative example. In the control block, two tonic responses appeared in the cue/peck period and the reward period. In the omission block, the sign of the reward-period activity changed immediately from excitatory to inhibitory. Unlike my observations regarding the MSt type-2 neurons (Fig. 4E), I observed little or no cue3 activity in the reward period. Thus, this neuron was expected to code the negative prediction error in the reward period. The prediction error model gave rise to the smallest AIC (inlet table), thus confirming the status of type-2 neuron. In contrast, in the omission block, the tonic cue/peck period activity remained while the delay period activity was inhibited. I did not observe any cue3 activity prior to reward delivery. With the exception of the excitatory reward-period response for cue1 in the control block, all features fit the simulated TD-error signal δ_t well (Fig. 2).

3.3.3.2 Delay condition

Figure 7E shows a representative example. In the control block, a strong tonic response appeared in the cue/peck period, while a weak tonic response appeared in the reward period. In the omission (delay) block, the sign of the reward-period activity changed immediately from excitatory to inhibitory. Simultaneously, a novel but weak reward-period response appeared during the new food-delivering phase (5.0–7.0 s). I observed a weak transient cue response to cue3 but no response in the other periods.

This neuron was thus expected to code the negative prediction error in the reward period. The prediction error model gave rise to the smallest AIC ([inlet table](#)), thus confirming the status of type-2 neuron. Conversely, the tonic cue/peck period activity remained in the omission (delay) block, although the amplitude declined slightly. These features fit the simulated TD-error signal δ_t well ([Fig. 2](#)).

[Fig. S4B](#) shows the averaged firing rate of 4 type-2 neurons. The rapid drop and inverted sign were reproduced ([Fig. 7F](#)). The response patterns of these neurons are in concert with the simulated δ_t signal ([Fig. 2](#)).

3.3.4 Type-3 neurons

3.3.4.1 Omission condition

[Figure 8A](#) shows the only neuron observed in this category. In the control block, an inhibitory tonic response appeared from cue1 onset and continued until the end of the reward period. In the omission block, the reward-period response remained and gradually disappeared. In the reacquisition block, the attenuated response quickly recovered when food delivery was reinstated. I did not observe any cue3 activity in the reward period. This neuron was thus expected to code the reward prediction in the reward period. The reward prediction model gave rise to the smallest AIC ([inlet table](#)), thus confirming the status of type-3 neuron. Conversely, the cue/peck/delay period activity remained in the omission block, with nearly the same amplitude. I also found a response to cue3 in the cue/peck/delay periods, but the amplitude was weaker. These features support the idea that the neuron coded the reward prediction prior to the food delivery. With the exception of the clear cue-period response, this neuron fits the simulated prediction signal $-\hat{V}(S_{t-1})$ well ([Fig. 2](#)).

3.3.4.2 Delay condition

[Fig. S5](#) shows a representative example and [Fig. S6](#) shows the averaged firing rate from two neurons. In the control block, these neurons showed responses in the cue period and reward period. In the omission (delay) block, the reward-period response remained and gradually disappeared. These responses were rather noisy, so I have chosen not to discuss the details of the firing patterns. Given the lack of a delay period

response, I argue that the activity of these neurons conflicts with that of the simulated $\hat{V}(S_{t-1})$ signal (Fig. 2).

3.4 Linear summation model of the TD error signal

As detailed above, type-1 neurons in the MSt (excitation subtype; Fig. 4A, B) appear to code the target signal $r_t + \gamma\hat{V}(S_t)$. In contrast, type-3 neurons in the MSt (inhibition subtype; Fig. 5C, D) may code the prediction signal $-\hat{V}(S_{t-1})$. I examined whether these two populations of neurons could sufficiently account for the TD-error signal δ_t of type-2 neurons in the tegmentum (Fig. 7D, E; Fig. S4B). To this end, I constructed a simple model of linear summation. The averaged z-score of excitatory type-1 neurons in the MSt was assigned as AR_{str} (actual reward in the striatum, target signal). Similarly, the averaged z-score of inhibitory type-3 neurons in the MSt was termed RP_{str} (reward prediction in the striatum, prediction signal). I expected the weighed sum of these two values to yield the z-score of type-2 neurons in the tegmentum (PE_{teg} , TD-error) as expressed by:

$$PE_{teg} = \beta_1 \cdot AR_{str} + \beta_2 \cdot RP_{str} \quad (4)$$

Coefficients ($\beta_1 = 0.7006$, $\beta_2 = 0.6623$) were estimated using the least squares method without assuming a constant term. The linear sum (orange) was superimposed on the PE_{teg} (black) in the bottom traces of Fig. 9A. Although the sum slightly underestimated the cue/peck period activities, it fit fairly well with the recorded PE_{teg} signal. Here, the trial period (from 0 s to 5 s) was composed of 50 bins, as the bin width was 100 ms. For each of the trial types, the PE_{teg} value of these 50 bins (y-axis) were plotted against the corresponding sum (x-axis) in Fig. 9B, with a considerable degree of correlation $R^2 = 0.5216$. The paired PE_{teg} value was also color-plotted on a AR_{str} vs RP_{str} plane in Fig. 9C; contour plot in C(a), and linear plot in C(b). Discrepancies between these two plots indicate that the linear model was limited to a first-order approximation.

Alternatively, the PE_{teg} signal may be an appropriate fit for the sum of tegmental neurons, namely excitatory type-1 neurons (as AR_{teg}) and the inhibitory type-3 neuron (as RP_{teg}), as expressed by:

$$PE_{teg} = \beta'_1 \cdot AR_{teg} + \beta'_2 \cdot RP_{teg} \quad (5)$$

Here, the coefficients were estimated as: $\beta'_1 = 0.5158$ and $\beta'_2 = 0.1152$. This model (5) fit similarly to (4) (Fig. S7), although the correlation for this model ($R^2 = 0.6067$) was slightly higher. Thus, both striatal and tegmental representations of the reward and its prediction could be involved in the computation of TD error.

3.5 Reciprocal connections between MSt and tegmental DA-ergic neurons

After micro-infusion of BDA to the MSt, I found dense anterogradely labeled fibers in the FRM and SN of the ipsilateral tegmentum (Fig. 10A), and less dense fibers in the VTA (not shown). Branching fibers and varicosities (A(b), (c)) indicate the presence of MSt neuron synaptic terminals in the FRM and SN. BDA and anti-TH double labeling indicated a high degree of overlap between MSt terminals and DA-ergic neurons in the tegmentum (Fig. 10B). High magnification observation using confocal microscopy revealed close apposition (arrowheads and arrows) between the MSt terminals (green) and the TH-positive neurons and proximal dendrites (red) in the FRM, SN, and VTA (Fig. 10C). Some cases of varicosity in BDA positive terminal boutons were co-localized with instances of anti-GAD65 labeling, indicating that some MSt terminals are GABA-ergic (Fig. 10D).

After micro-infusion of DiI to the MSt, I found retrogradely labeled cell bodies in the FRM, SN, and VTA (Fig. 11A). The projection neurons were dense in the SN and VTA, while I only found a few neurons in the FRM. When DiI was injected into the SN, I found retrogradely labeled neurons in several areas in the ipsilateral striatum, such as the MSt, Ac, BSTl, and VP (Fig. 11B). The medial part of the MSt contained more labeled neurons than the lateral MSt, suggesting a functional separation between these two sub-regions.

4 Discussion

4.1 Striatal representations of the target signal and the prediction signal

The neuronal mechanisms involved in the computation of TD error have been intensively studied. The mechanisms for one TD method, termed the actor-critic method, have been localized in the basal ganglia (Barto, 1995; Houk et al., 1995). Specifically, DA-ergic neurons, together with striatal neurons, have been assumed to play a critical role as the ‘critic’ in the computation of TD error (Doya, 2007; Houk et al., 1995; Joel et al., 2002). Several lines of supporting evidence have been developed in mammals and birds. First, the striatum provides one of the major projections descending to DA-ergic neurons in the tegmentum (Anderson et al., 1991; Mezey and Csillag, 2002; Watabe-Uchida et al., 2012). Second, localized lesion and pharmacological manipulation studies have reported critical involvement of the striatum in reinforcement learning (Annett et al., 1989; Castañé et al., 2010; Clarke et al., 2008; Ichikawa et al., 2004; Izawa et al., 2001; Ogura et al., 2015; Rueda-Orozco et al., 2008). Third, during reinforcement tasks, striatal neurons show reward-related activities both before and after mammals (Apicella et al., 2009; Janak et al., 2004; Kim et al., 2009; Tremblay et al., 1998) and birds (Amita and Matsushima, 2014; Izawa et al., 2005; Yanagihara et al., 2001) receive a reward.

Despite the above-mentioned efforts, the detailed mechanisms of TD error computation have not been fully elucidated at the neuronal level. Previous studies in mammals have suggested the possibility that the striatum send the state value signal to midbrain DA regions via direct or indirect pathway (Houk et al., 1995; Doya, 2000; Kawato and Samejima, 2007). However, the present results suggest for the first time that both critical signals of TD learning are represented by striatum neurons. Based on my present results, Fig. 12 illustrates my proposed neuronal network underlying TD error computation. The sign-inverted signal of the predicted value of the state S_{t-1} represents the prediction $-\hat{V}(S_{t-1})$. In other words, the striatum retains the reward prediction signal even after the food is delivered. The signal $r_t + \gamma\hat{V}(S_t)$ represents the target of $\hat{V}(S_{t-1})$. Through the course of learning, the prediction signal $\hat{V}(S_{t-1})$ approaches the target signal $r_t + \gamma\hat{V}(S_t)$ according to the difference between these two

signals, i.e. the TD-error signal. This update of the prediction may be achieved through the modulation of the cortex/pallium-striatum synaptic transmission by DA inputs from tegmentum (Schultz 2015). The main difference between my model and previous models is the target signal assumed to be represented in the striatum. In my model, only striatum supplies the signals for computing TD-error. However, it is also possible that other regions together with striatum supply the signals, e.g. pedunclopontine tegmental nucleus may also be involved (Okada et al., 2007).

The prediction signal that I observed in chick striatal neurons (Fig. 5D) is similar to those found in the GABA-ergic neurons of the mouse VTA (Cohen et al., 2012). Mouse GABA-ergic neurons were found to code prediction in the reward period. The firing gradually increased after the onset of a reward-predictive cue, and sustained even after the reward was received. The activity in the reward period remained unaltered even in omission trials. Similar neuronal signals have been reported in the striatum in mammals (Kim et al., 2009; Oyama et al., 2015; Tremblay et al., 1998). Note, however, that in the study by Kim et al. (2009), the researchers expected the neurons to code the action value rather than the state value. See below for discussions on the distinction between these two forms of value representation.

The target signal $r_t + \gamma \hat{V}(S_t)$ found in this study has two components, i.e., (1) the actual reward r_t , and (2) the expected value of the current state $\gamma \hat{V}(S_t)$. Similar *prediction followed by reward* activity has been reported in monkeys (Tremblay et al., 1998), although this finding has not been associated with TD learning theory. Instead, I suggest that the same neuron may represent these two components as a critical signal in TD learning.

In this respect, it is worth noting that type-1 and type-3 neurons differed in terms of recording site (Fig. 3A). The type-3 neurons (the prediction signal) were found in the medial part of the MSt and the Ac, whereas the type-1 neurons (the target signal) were located in the lateral part of the MSt. In a neuroanatomical study in pigeons, the medial part of the MSt was found to receive afferents from several pallial regions (Veenman et al., 1995). Of these, two regions are important in reinforcement learning, i.e., the central arcopallium and the prehippocampal area (APrH). Some neurons in the central

arcopallium showed sustained responses during reward omission (Aoki et al., 2003), similar to the prediction signal found in the MSt. The APrH is thought to be analogous to the mammalian cingulate cortex (Veenman et al., 1995), which also codes actual reward, prediction, and prediction error in monkeys (Seo and Lee, 2007). Thus, these two regions may supply the prediction signal to the medial MSt.

In contrast, the lateral part of the MSt receives inputs from other pallial regions (Kröner and Güntürkün, 1999; Veenman et al., 1995). Of these, two regions may be critical. The first is the apical part of the hyperpallium (HA), which is part of the Wulst, one of the major visual centers (Ocklenburg and Güntürkün, 2012). The HA may supply the lateral MSt with the necessary information regarding color cues and food. The second critical region is the rostral part of the caudolateral nidopallium (rostral NCL). Neurons in the NCL show sustained responses to reward-predictive cues and responses to actual rewards (Diekamp et al., 2002), similar to the target signal found in this study in the MSt. These two regions may thus converge onto lateral MSt neurons, giving rise to the target signal.

4.2 Striatal and tegmental representations of TD error

In the present study, I found that type-2 neurons in the MSt (Fig. 4D–F) and tegmentum (Fig. 7D–F) fit the model of prediction error better than the alternatives. In the omission block, both of these neurons showed (1) excitation in the cue/peck period, and (2) inhibition (lower firing below the baseline) in the following reward period. Both of these features match those assumed to be involved in the TD error signal. However, in the control block, these neurons showed excitation in the reward period. If chicks had been over-trained such that they accurately predicted the reward amount, such excitatory activity should not occur. Because of this conflict, I argue that the inhibition in the reward period found in the type-2 neurons may not represent the prediction error.

One possible explanation for the above finding is that the chicks were not fully trained to discriminate color cues. As shown in the behavioral data (Fig. 1C), even after intensive training for 3 days or longer, the chicks still pecked cue3 (non-rewarding color cue) in 25–50% of the trials. Similarly, in the cue1 and cue2 trials, the chicks might not

have predicted the food with 100% certainty. This may explain why the TD signal was positive in the reward period, in which food was only partially predicted. Indeed, DA-ergic neurons in monkeys showed a similar pattern of excitation in response to a predicted reward (Fiorillo et al., 2003; Morris et al., 2004).

An alternative explanation is that the type-2 neurons code the target signal $r_t + \gamma \hat{V}(S_t)$ rather than the TD error δ_t , similar to the type-1 neurons. The inhibition observed during the reward period could be due to food omission, rather than the prediction error signal. This is particularly plausible in the striatal type-2 neurons (Fig. 4E), in which similar inhibition occurred in the cases in which the omission of food was predicted (cue3: dashed blue line) and unpredicted (cue1: dashed red line).

However, it is not possible to explain the pattern of tegmental type-2 neuron activity in this manner, because distinct activities occurred in cue1 (omission/delay) and cue3 (control) trials (Fig. 7D–E). Thus, this neuron type might fit the explanation that these code the TD-error. However, I had no evidence as to whether the tegmental type-2 neurons are DA-ergic neurons. In an electrophysiological study of zebra finches, DA-ergic neurons in the VTA and SNc (substantia nigra, pars compacta) exhibited wider spikes and a lower firing rate compared with non-DA-ergic neurons in the same regions (Gale and Perkel, 2006). In my present study, on the other hand, I found no significant differences in spike width and firing rate among the three neuron types (Fig. 6B). Importantly, the prediction error signal in previous studies has also been found in non-DA-ergic neurons (Schultz, 2015), including those in the striatum in rats (Kim et al., 2009; Oyama et al., 2010) and monkeys (Apicella et al., 2009). The different firing patterns observed in the type-2 neurons in the MSt and tegmentum may imply that these regions have different functionality.

4.3 Tegmental neurons may also contribute to TD-error computation

Although most type-1 and type-3 neurons in the tegmentum did not fit the TD learning signals well, one tegmental neuron warrants attention (Fig. 8). The reward-period inhibition gradually decayed in the omission block, and resumed in the subsequent reacquisition (rewarding) block. This activity pattern resembles the

simulated prediction signal $-\hat{V}(S_{t-1})$ (Fig. 2). Simple summation of the type-1 (Fig. 7A–C) and type-3 neuron can yield the TD error signal. This may be locally computed within the tegmentum, similar to the local GABAergic neurons in the mouse VTA (Cohen et al., 2012; Eshel et al., 2015); see the discussion below and Fig. 12.

4.4 Neuroanatomical bases of TD error computation

4.4.1 Direct inhibitory pathway

My tract-tracing experiments were consistent with previous reports regarding the connectivity between the MSt and the DA rich tegmentum nuclei in the avian brain. As previously reported in chicks (Székely et al., 1994), I confirmed that descending MSt neurons have direct synaptic contacts onto DA-ergic neurons in the FRM, SN, and VTA (Fig. 10C). My GABA immunostaining data (Fig. 10D) also supported the previous finding that striatal projection neurons in pigeons are GABAergic (Reiner and Anderson, 1990). It is therefore reasonable to suggest that MSt projection neurons have an inhibitory effect on DA neurons. However, my hypothetical algorithm (Fig. 2) and the linear summation model (Fig. 9) assumes that excitatory type-1 and inhibitory type-3 MSt neurons have an excitatory effect on type-2 neurons in the tegmentum. Thus, how the descending inhibitory pathway mediates the summation of the two striatal signals in the tegmentum requires further explanation.

4.4.2 Indirect pathway for disinhibition

In addition to the direct inhibitory pathway, striatal neurons may indirectly affect DA-ergic neurons through local interneurons within the tegmental nuclei. A immunohistochemical study in pigeons showed that DA-ergic neurons in the SN receive inputs from both SP-positive striatal neurons and SP-negative neurons, which may come from other regions (Anderson et al., 1991). The authors also reported that SP-positive striatal terminals contacted both DA-ergic and non DA-ergic neurons in the SN. A recent study in mice proposed the functional involvement of the indirect pathway, as nucleus accumbens neurons in the ventral striatum dis-inhibit DA-ergic neurons in the VTA by inhibiting GABA-ergic local inter-neurons (Bocklisch et al., 2013). Additionally, DA-

ergic neuron activity in the VTA is suppressed by local GABA-ergic inter-neurons in mice (Eshel et al., 2015).

Similar disinhibitory action may occur in chicks. My present tracing experiment is consistent with a previous study in chicks (Bálint et al., 2011), which reported that the above-mentioned DA-ergic tegmentum nuclei receive efferents from the MSt and Ac. However, it is important to know how and where the descending GABA-ergic inhibition is converted. As reported in mammals, candidates include the local GABA-ergic inter-neurons in the VTA, FRM and SN pars reticulata in the avian brain (Veenman and Reiner, 1994) (Fig. 12). In future research, it will be critically important to determine whether the descending GABA-ergic MSt efferents have synaptic contacts with the presumed GABA-ergic local interneurons in the tegmental nuclei. Also, this disinhibition effect should be examined using electrophysiology.

4.5 TD learning for updating state value and behavioral execution

4.5.1 Two types of TD errors for state value and action value

Generally, two different types of TD error signals have been studied using theoretical approaches (Sutton and Barto, 1998). The first type focuses on the TD of the state value. The classical actor-critic method adopts this type, which was assumed in the early studies of DA-ergic neurons (Montague et al., 1996; Schultz et al., 1997). Actually, neuronal activities in the ventral striatum and anterior cingulate of monkeys coded the progress of a task comprising a series of trials prior to a reward (Shidara et al., 1998; Shidara and Richmond, 2002). Thus, aspects of state may be coded in these regions. On the other hand, the second type focuses on the TD of the action value. Methods such as Q-learning and SARSA adopt this type of TD error. In recent studies, the second type also proved to be plausible, as DA-ergic neuron activity in a decision making task was accounted for by TD error via the SARSA (Morris et al., 2006) and Q-learning methods (Roesch et al., 2007). In the present study, I assumed the first type of TD error signals, and found neuronal activities that matched the simulated signals. Whether the second type is also implemented by striatal/tegmental neurons in decision making tasks is still unknown.

4.5.2 Representation of the action value in the striatum

In addition to reinforcement learning, the striatum is involved in the modulation of locomotor movements ([Grillner et al., 2005](#)). It is thus important to determine whether other striatal neurons code the action value, or the quality of several different actions ([Sutton and Barto, 1998](#)). As mentioned above, striatal networks may be critical for computing the second type of TD errors, and thus may code action value. Furthermore, the action value can guide the action selection in the actor-critic method ([Barto, 1995](#)), in which the action with a larger action value tends to be chosen more frequently. Action value signals have been found in the striatum in monkeys ([Kawagoe et al., 1998](#); [Samejima et al., 2005](#)) and rats ([Kim et al., 2009](#)), and these may be modified by TD error signals issued by DA-ergic neurons ([Doya, 2007](#)). In the present study, chicks did not choose from multiple options, so I did not focus on the action value. As a future project, it will be important to determine whether striatal/tegmental neurons also code the action values for tasks in which subjects must choose from multiple targets or actions.

5 **Figures**

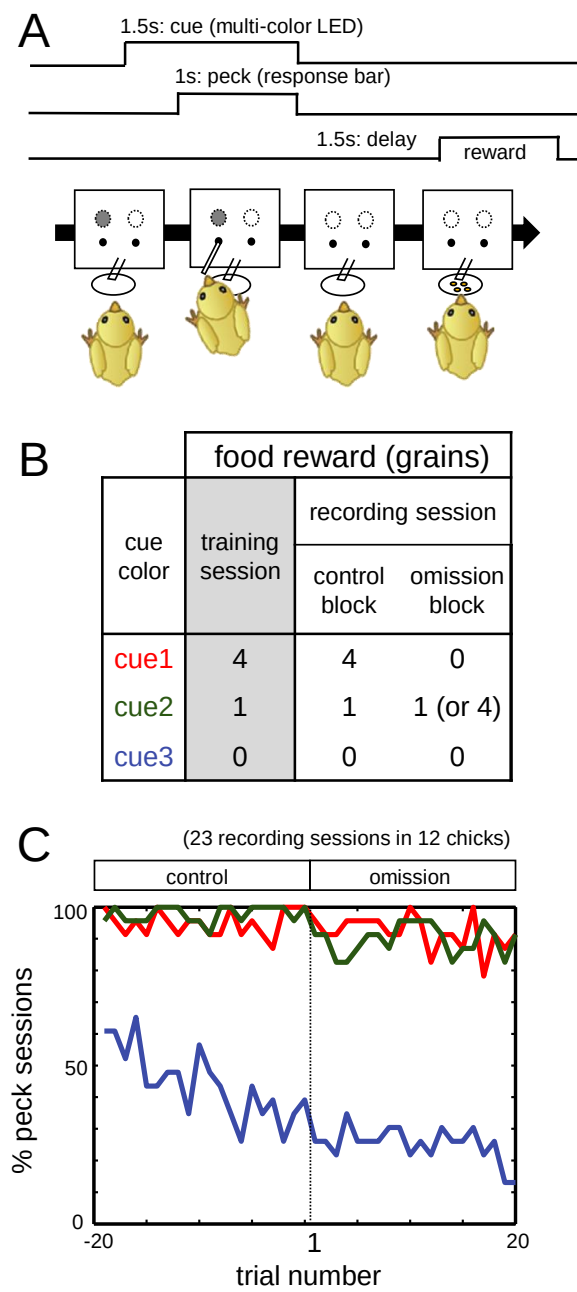
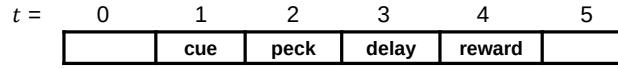


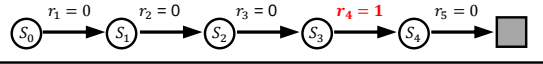
Figure 1. (A) Procedure of operant task reinforced by a delayed reward. (B) Color-food associations in training and recording sessions. (C) Percentage of the sessions in which chicks pecked. Red, green, and blue lines denote responses in cue1, cue2, and cue3 trials, respectively.

A



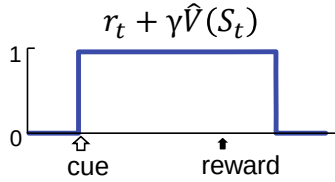
B

control block

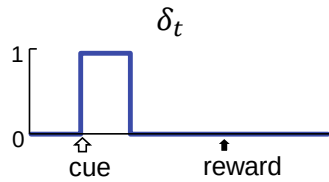


r_t	0	0	0	0	1	0
$\hat{V}(S_t)$	0	1	1	1	0	
$\hat{V}(S_{t-1})$		0	1	1	1	0
$\delta_t = r_t + \gamma \hat{V}(S_t) - \hat{V}(S_{t-1})$		1	0	0	0	

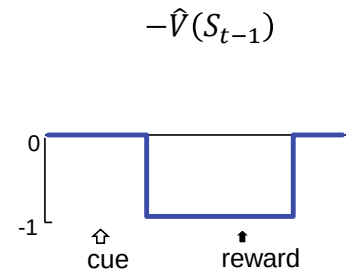
type-1 (target)



type-2 (TD-error)



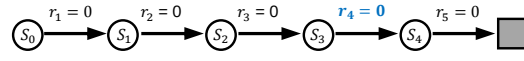
type-3 (prediction)



assumption
 $\hat{V}(S_t) \approx V(S_t)$
 $\gamma=1$

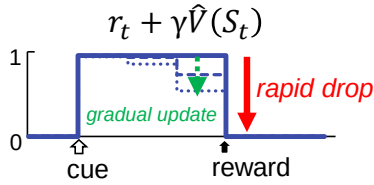
C

omission block

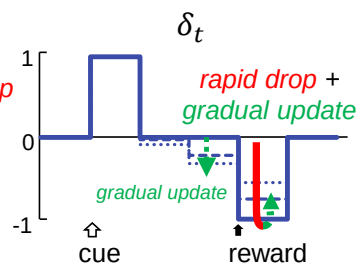


r_t	0	0	0	0	0	0
$\hat{V}(S_t)$	0	1	1	1	0	
$\hat{V}(S_{t-1})$		0	1	1	1	0
$\delta_t = r_t + \gamma \hat{V}(S_t) - \hat{V}(S_{t-1})$		1	0	0	-1	

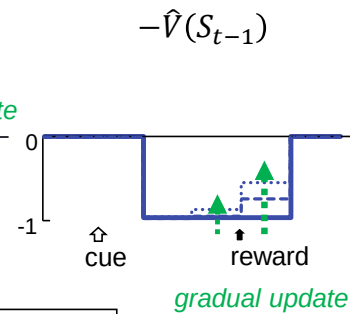
type-1 (target)



type-2 (TD-error)



type-3 (prediction)



— 1st trial
 --- 10th trial
 20th trial
 assumption
 $\alpha=0.03$

Figure 2. Variables of TD learning were simulated according to the task. **(A)** Trials were mimicked as finite sequences composed of 6 discrete states. **(B)** Control block. Signals are shown in the table and also schematically illustrated below. S_t and r_t denote state and reward, and $\hat{V}(S_t)$ represents the estimated state value at t . δ_t represents the temporal difference error (TD error). Temporal discounting of reward was not assumed, thus $\gamma = 1$. **(C)** Omission block. Signals in the first trial of the omission block are shown in the table, and their updating processes are illustrated below. Signals at the 1st, 10th, and 20th trial are shown. I adopted TD(0) method, and assumed the learning rate to be $\alpha = 0.03$.

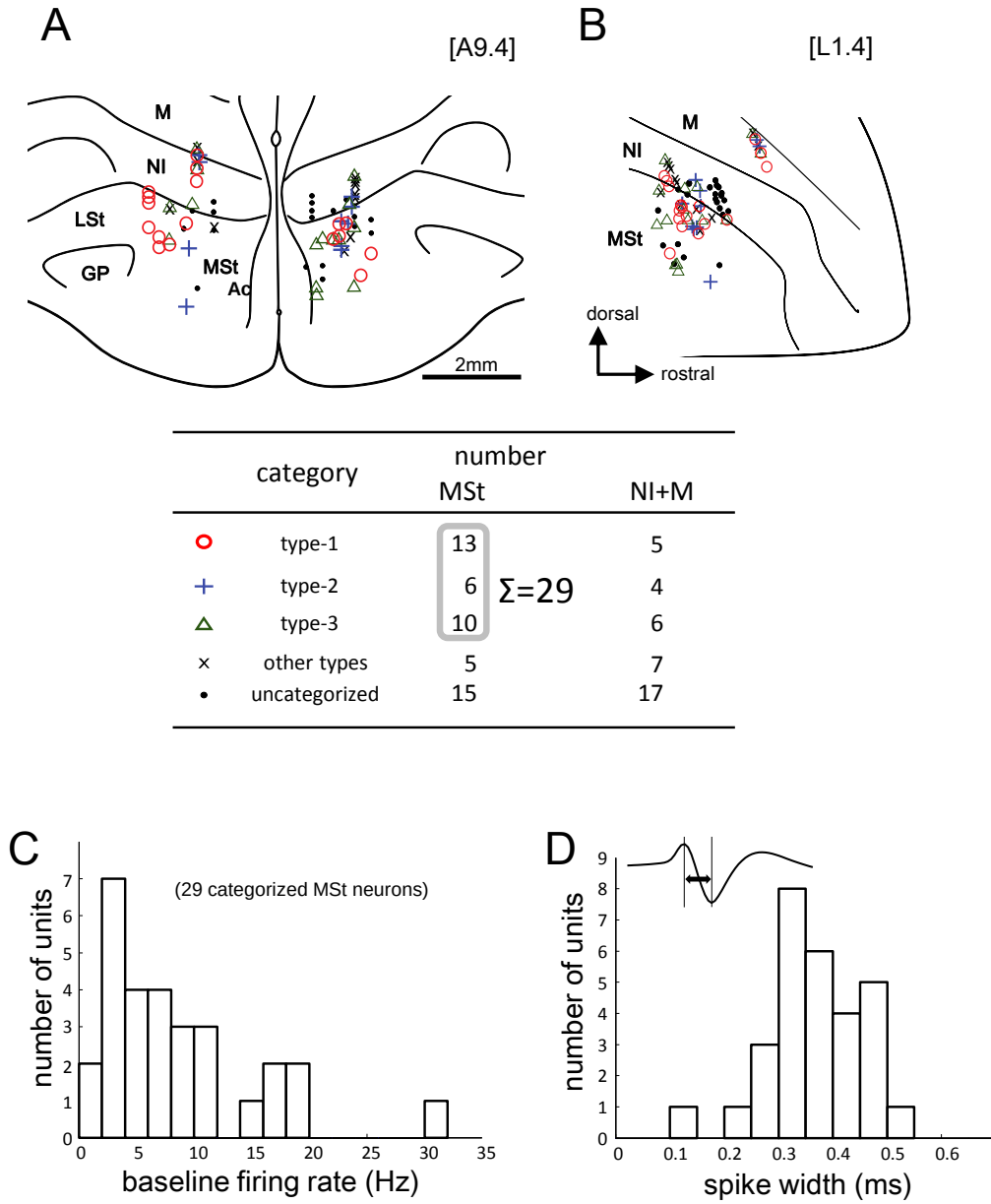


Figure 3. Task-related neurons in the medial striatum (MSt) and surrounding regions. **(A, B)** Histological reconstruction of recording sites on frontal **(A)** and sagittal **(B)** planes. Anterior-posterior level [A9.4] and laterality level [L1.4] follow the atlas by Kuenzel and Masson (Kuenzel and Masson, 1988); see section 2.1 for abbreviations. Symbols denote different neuron types (inlet table). Neurons were categorized as type-1, -2, -3, or other, according to the reward period activities of the cue1 trials. **(C)** Baseline firing rates (pre-trial period) of 29 MSt categorized neurons in a histogram. **(D)** Spike width, as measured by the peak-to-peak duration (inlet figure), in a histogram.

type-1 & -2 neurons in MSt
[prediction in cue/peck/delay period + actual reward in reward period]

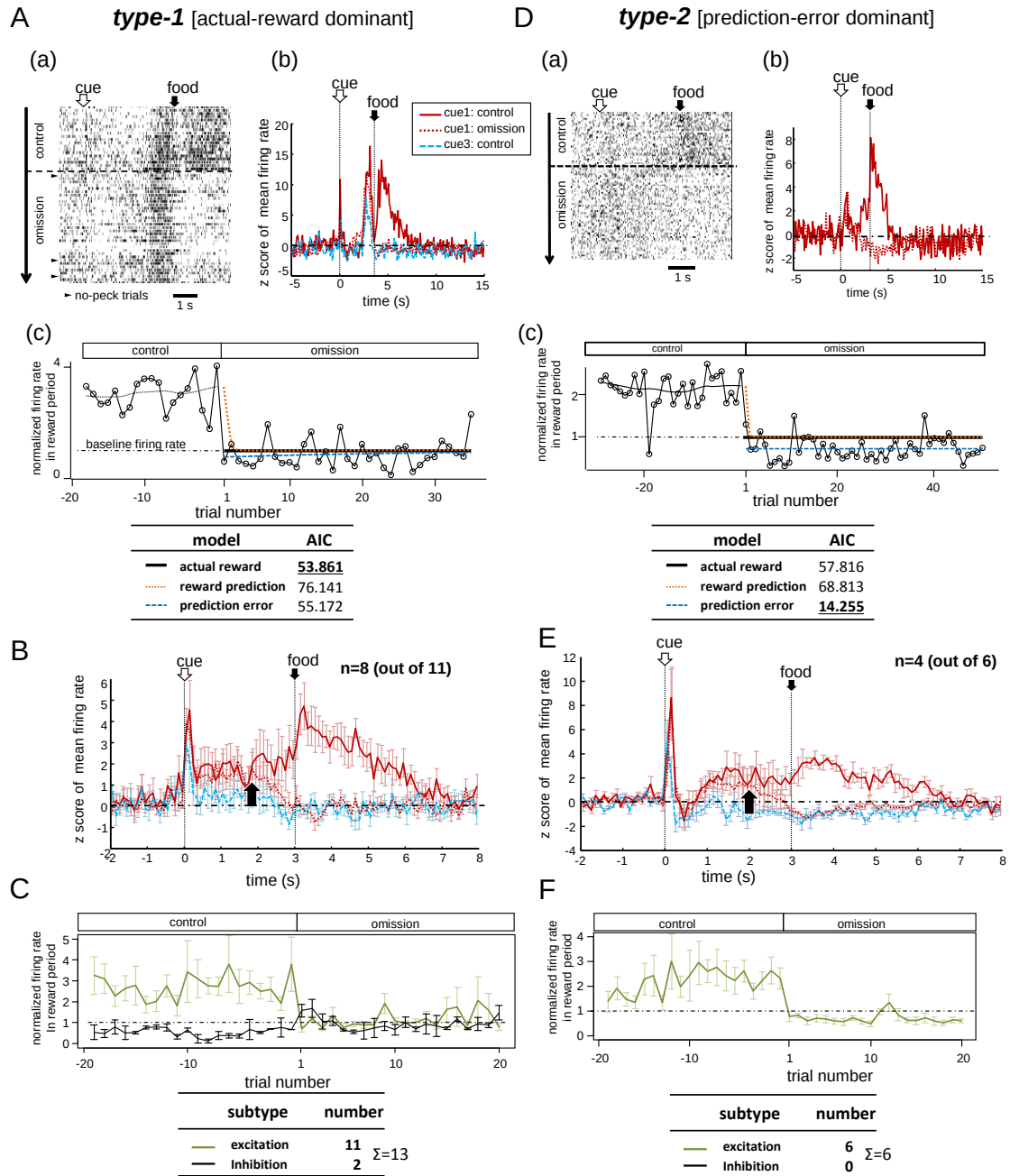


Figure 4. MSt neurons representing reward prediction in the cue/peck/delay periods, and actual reward in the reward period. These neurons were categorized into two types; type-1 (actual-reward dominant; **A-C**) and type-2 (prediction-error dominant; **D-F**). **A** shows a representative example. In **A(a)**, activities in cue1 (rewarding) trials are shown as a rastergram. Arrowheads indicate the few no-peck trials in which chicks did not peck the response bar. In **A(b)**, the averaged firing rate (z-score) in the cue1 trials is compared between the control block (4 grains: red line) and the omission block (0 grain: dashed red line). Activity in the non-rewarding cue3 trials in the control block is superimposed (dashed blue line). In **A(c)**, the normalized firing rate in the reward period (open circles connected with lines) is plotted against the trial number, with the number = 1 denoting the first trial of the omission block. The dashed line superimposed in the control block represents the smoothed activity. I tested the fit of the firing rate in the omission block to three models: actual reward (thick dark line), reward prediction (orange dotted line), and prediction error (blue dashed line). The table below shows the AIC value of each fit curve; see the text for details regarding the models. **B** shows population data; mean firing rate (z score) of type-1 neurons (n = 8, excitation type; mean and s.e.m.) in cue1 (control; red line), cue1 (omission; dashed red line), and cue3 (control; dashed blue line) trials. The upward arrow indicates the divergence point between the control and omission block for cue1. **(C)** Normalized firing rates of type-1 neurons in the reward period. Excitation (green line) and inhibition type (black line) neurons are shown separately. A representative example (**D**) and population data (**E-F**) for type-2 neurons, which had significant inhibition in the reward period, thus fitting the prediction error model well. **D-F** follow the same conventions as **A-C**.

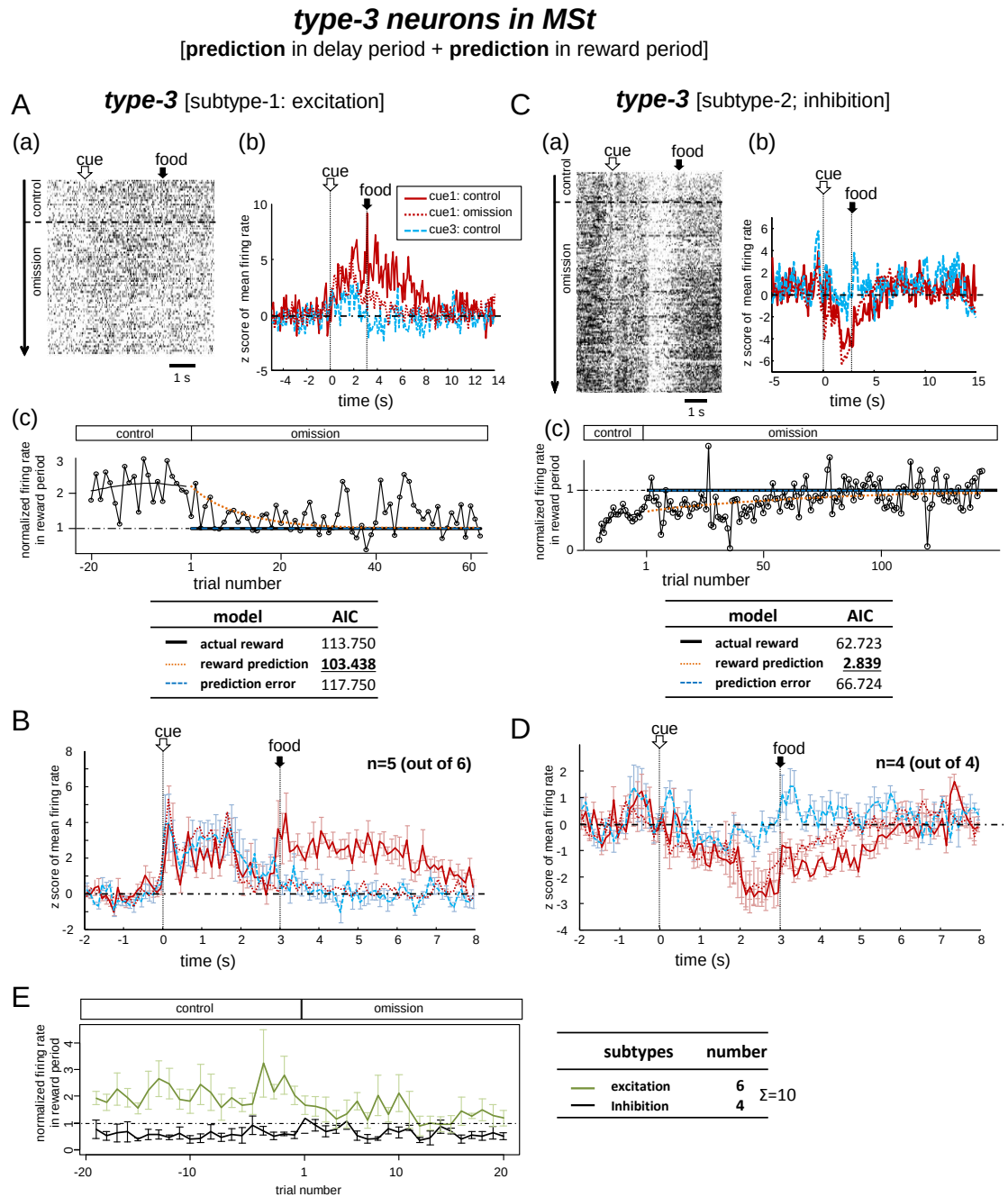


Figure 5. MSt neurons with representations of reward prediction in the delay period and the reward period (type-3, reward-prediction dominant). Activity in the cue1 trials of the omission block fit best to the reward prediction model. (**A**, **B**) Excitation subtype. (**C**, **D**) Inhibition subtype. (**E**) Changes across the trial number. In contrast to the type-1/2 neurons in **Figure 4**, the delay-period activity did not diverge between the control and omission blocks.

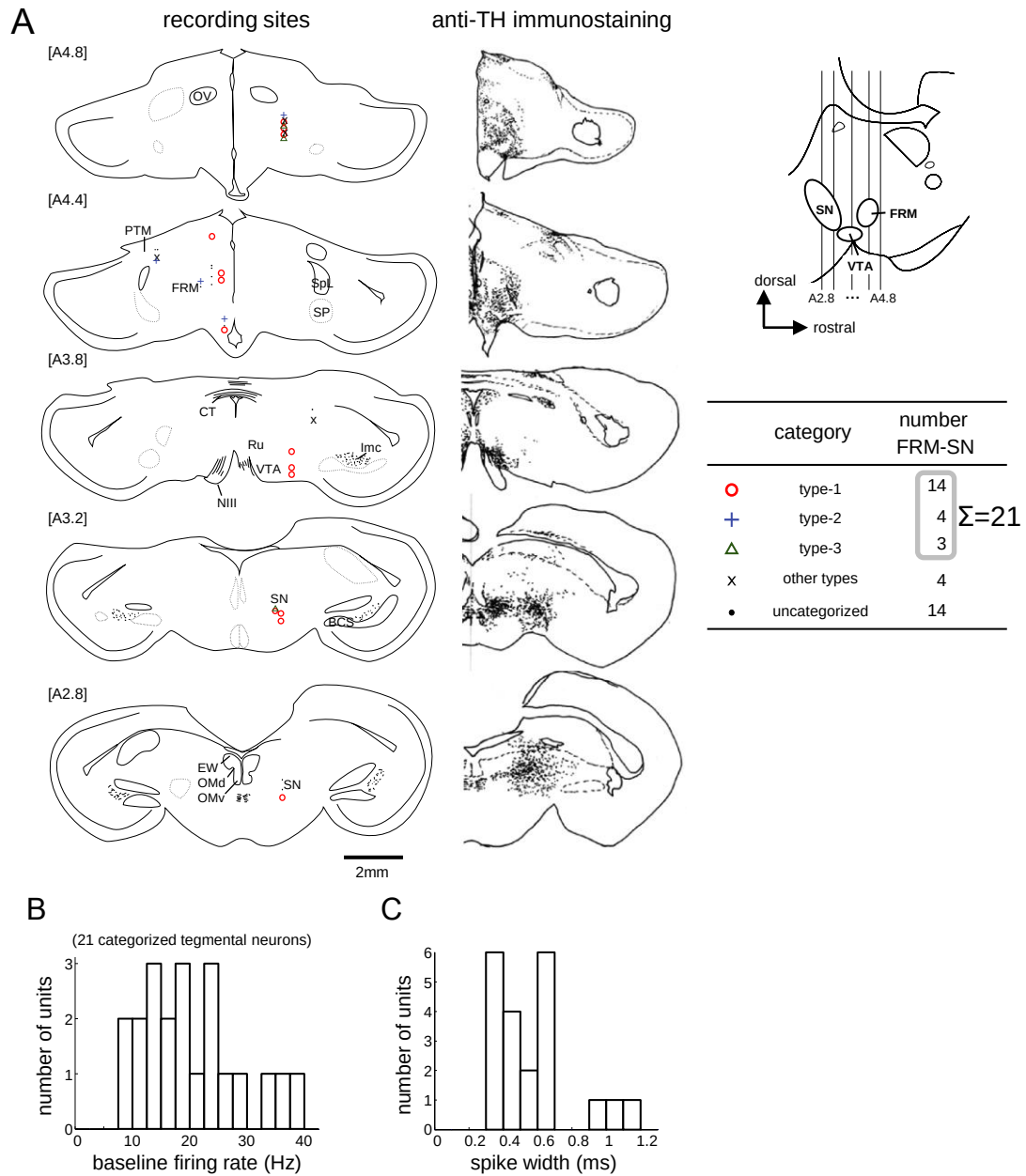


Figure 6. Task-related neurons in the midbrain tegmentum. **(A)** Histological reconstruction of recording sites on frontal planes, corresponding to the anti-TH immunostaining on the right; see section 2.1 for abbreviations. The anterior-posterior level of sections (A2.8 to A4.8) is shown in the inlet. **(B, C)** Baseline firing rate and spike width histograms of the 21 categorized tegmentum neurons.

type-1 & -2 neurons in tegmentum

[prediction in cue/peck period + actual reward in reward period]

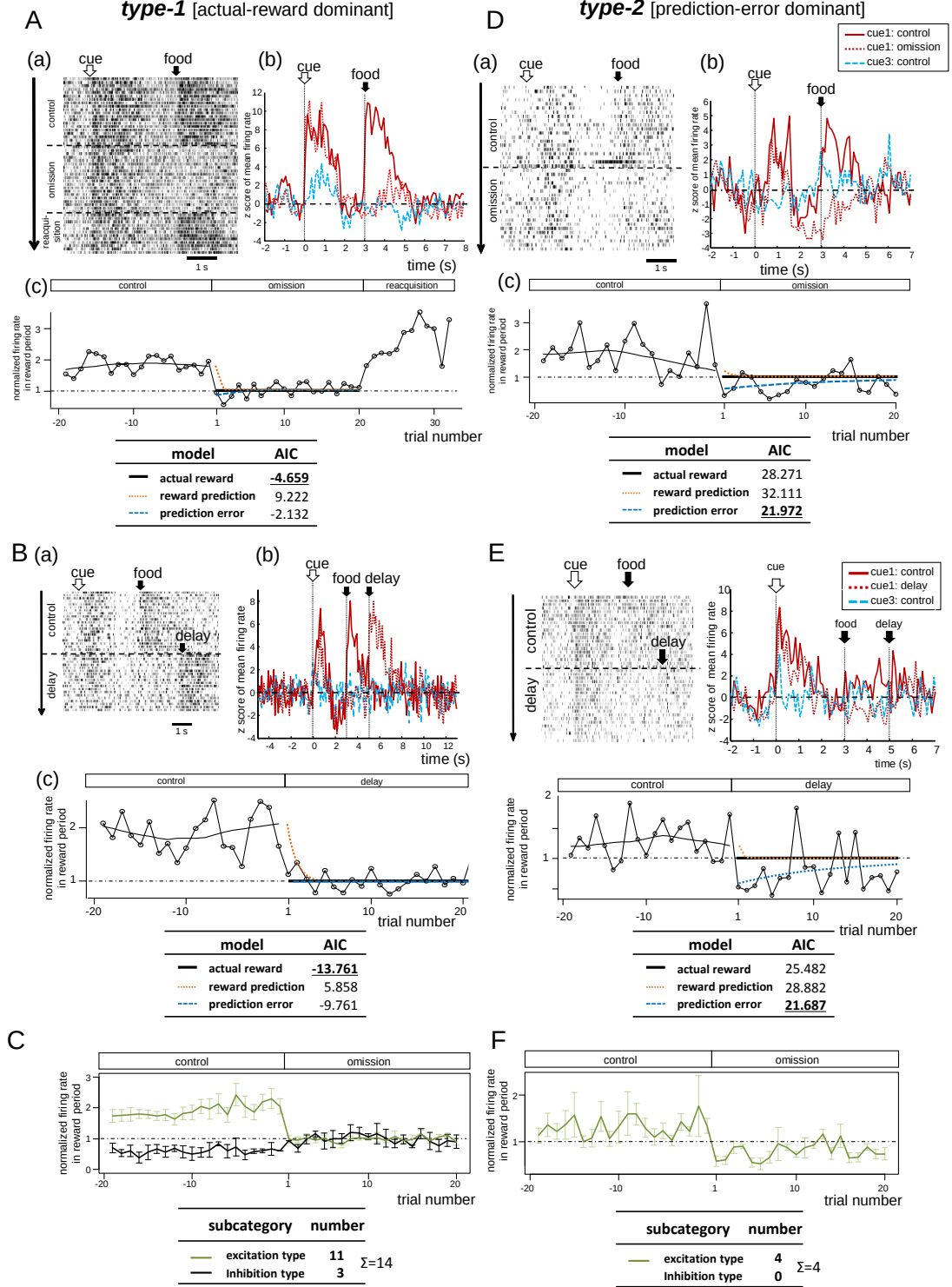


Figure 7. Tegmental neurons representing reward prediction in the cue/peck period, and actual reward in the reward period. These neurons were categorized into two types; type-1 (actual-reward dominant) and type-2 (prediction-error dominant). Two example neurons are shown for each type. **A** and **B** show two type-1 neurons tested in the omission (**A**) or in the condition with delayed delivery of reward (**B**). Similarly, **D** and **E** show two type-2 neurons tested in the omission (**D**) or delayed reward condition (**E**).

type-3 neurons
[prediction in cue/peck/delay period + prediction in reward period]

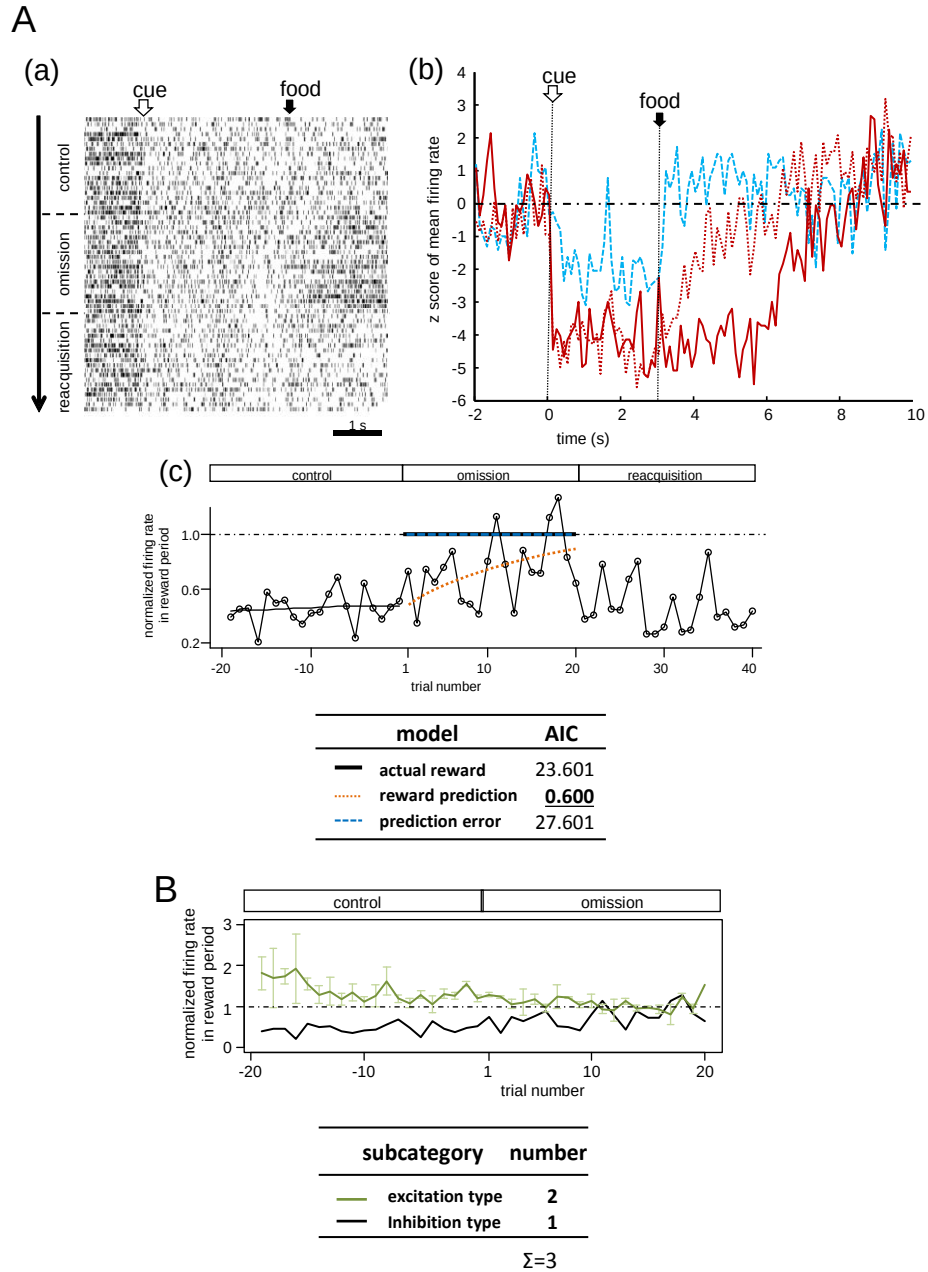


Figure 8. Tegmental neurons representing reward prediction in the cue/peck/delay period and the reward period (type-3, reward-prediction dominant). The omission block was followed by another rewarding block (reacquisition) in the example shown in **A**. Note the gradual decline in reward-period inhibition in the omission block in **A(b–c)**.

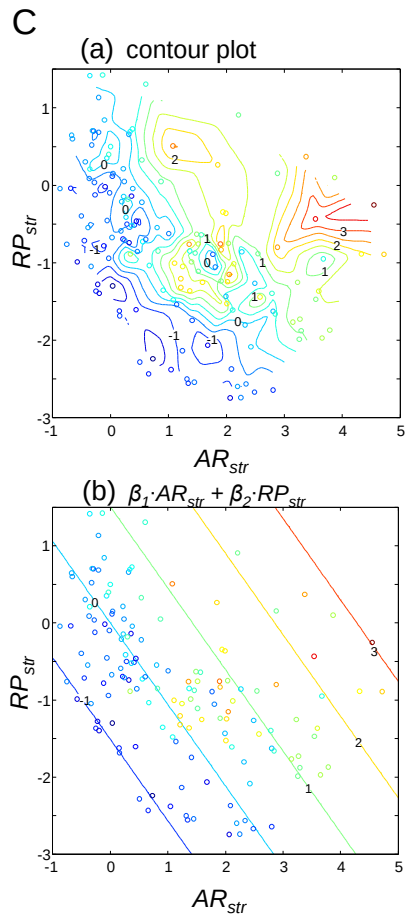
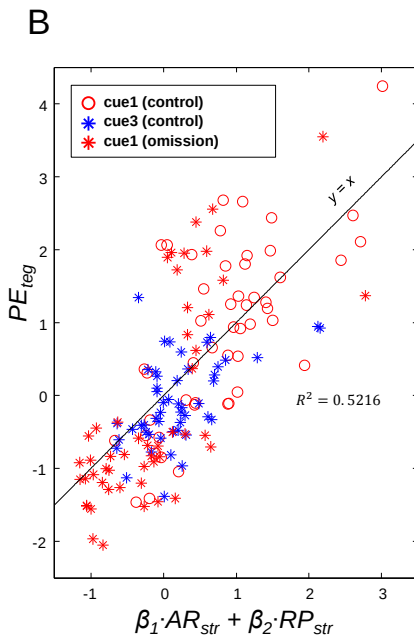
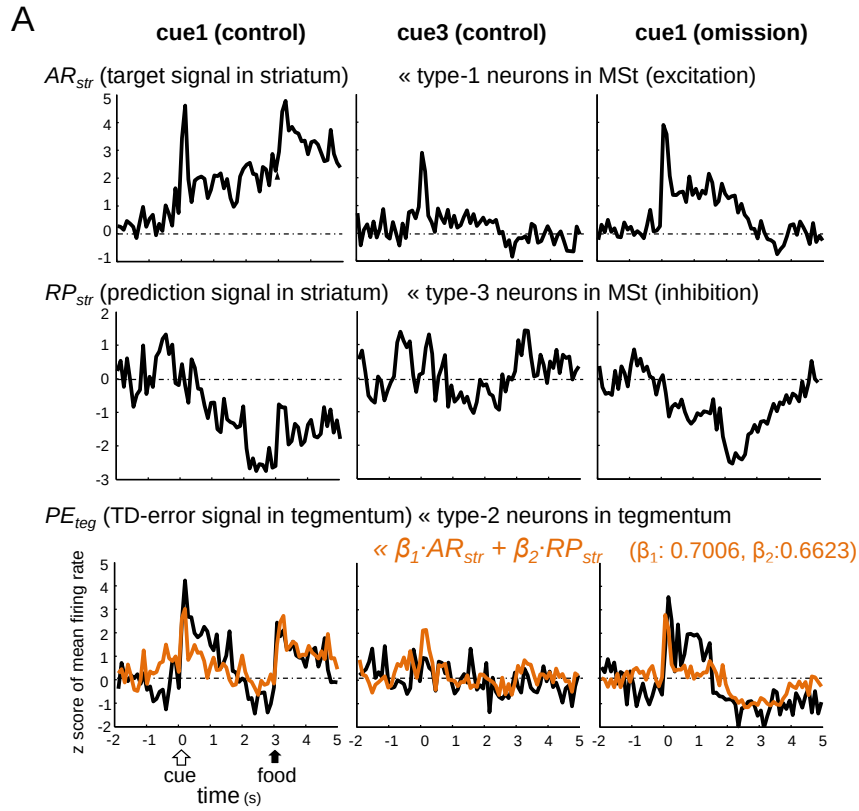
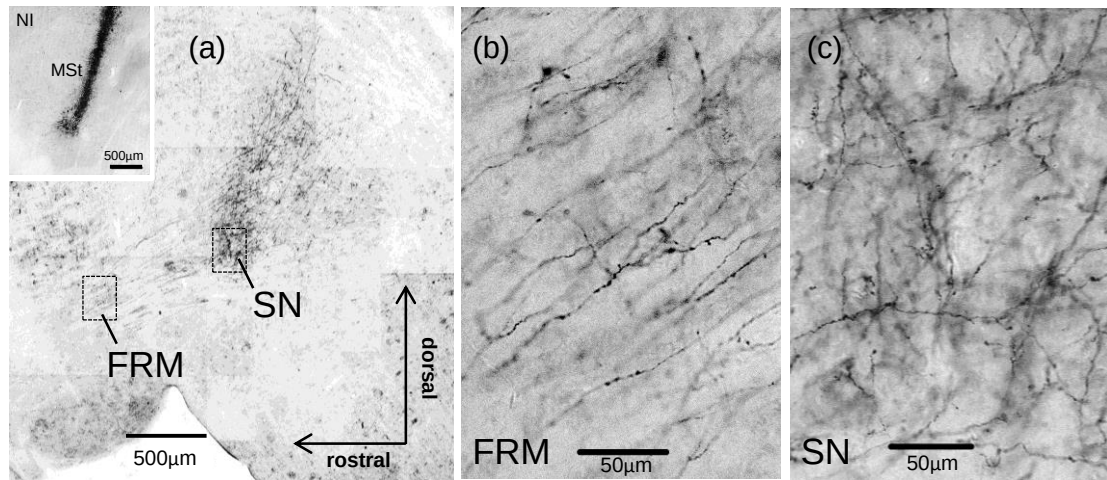
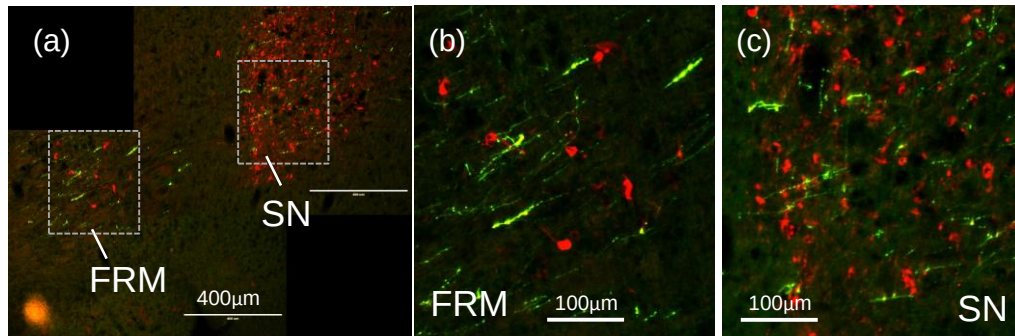


Figure 9. Type-2 neuronal activity in the tegmentum (presumptive TD-error signal, PE_{teg}) fitted as a linear sum of the activities observed in a subset of MSt neurons. **(A)** Averaged activity of type-1 (AR_{str}) and type-3 (RP_{str}) MSt neurons, shown together with that of type-2 tegmentum neurons (PE_{teg}) for 3 blocks of trials, cue1 (control), cue3 (control), and cue1 (omission). Superimposed orange lines on the PE_{teg} signal denote the linear sum of AR_{str} and RP_{str} . **(B)** Scatter plot of PE_{teg} versus the linear sum. **(C).** Pseudo 3-D plots of AR_{str} (x-axis), RP_{str} (y-axis), and PE_{teg} (color code) with the interpolated contour plot **(a)** and linear summation **(b)**.

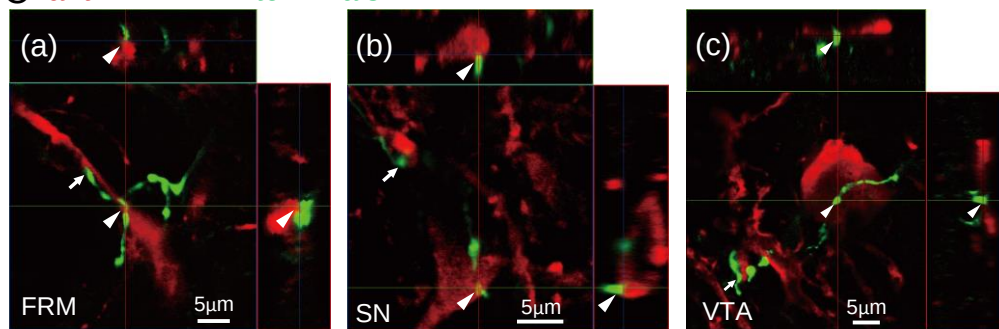
A Fibers and terminals labelled after BDA microinfusion to MSt



B anti-TH + BDA terminals



C anti-TH + BDA terminals



D anti-GAD65 + BDA terminals

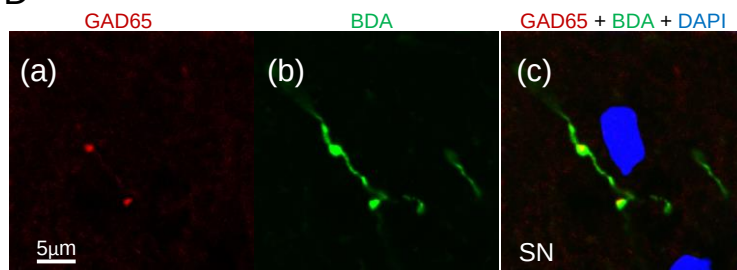
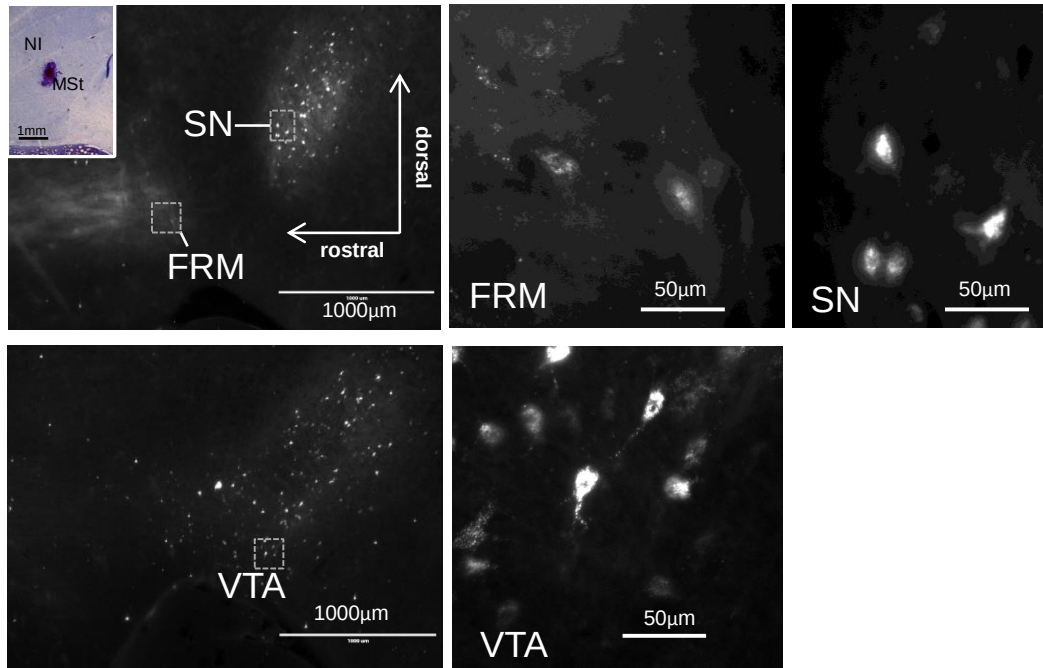


Figure 10. Direct contacts of MSt terminals on DA-ergic neurons in the tegmentum. **(A)** Dense arborizations of MSt efferent fibers were found in the FRM and SN (sagittal plane); low magnification **(a)**, high magnification in the FRM **(b)**, and the SN **(c)**. The inlet figure shows the injection site in the MSt. **(B)** BDA/TH double labeling in the tegmentum; BDA, green; TH, red. **(C)** Confocal images of direct contacts between BDA-positive terminal boutons and TH-positive dendrites and soma. Reconstructed on 3 orthographic planes. Arrowheads and arrows indicate the close appositions. **(D)** BDA/GAD65 double labeling in the SN, indicating co-localization on the terminal boutons. Sagittal sections with laterality: L1.4 in **Aa-c**; L0.9 in the **A** inlet; L1.4 in **B(a-c)**; L1.5, 1.1, and 1.0 in **C(a-c)**; L1.3 in **D(a-c)**.

A Retrogradely labeled tegmental neurons after DiI microinfusion to MSt



B Retrogradely labeled striatal neurons after DiI microinfusion to SN

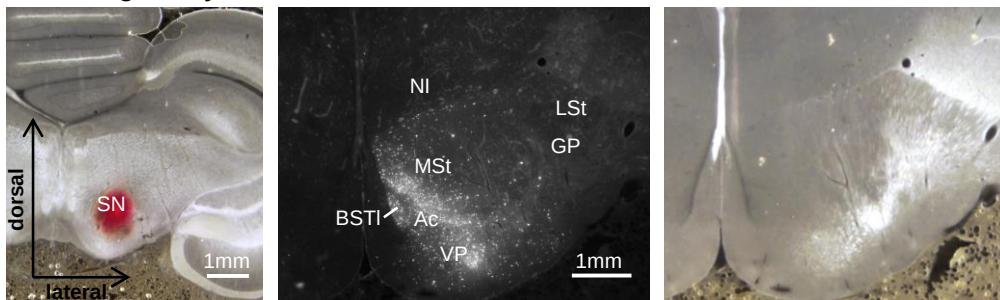


Figure 11. Reciprocal projections between the MSt and tegmentum. **(A)** Retrogradely labeled neurons in the FRM, SN, and VTA after micro-infusion of DiI into the MSt. The inlet indicates the injection site. **(B)** Retrogradely labeled neurons in the striatum after DiI infusion into the SN (left). DiI image (middle) and corresponding bright-field photo (right). See section 2.1 for abbreviations. Sagittal planes with laterality: L1.2 in the injection site, L1.4 in the FRM and SN, and L0.8 in the VTA. Frontal planes with A-P level: A3.4 in the injection site, A9.4 in the striatum.

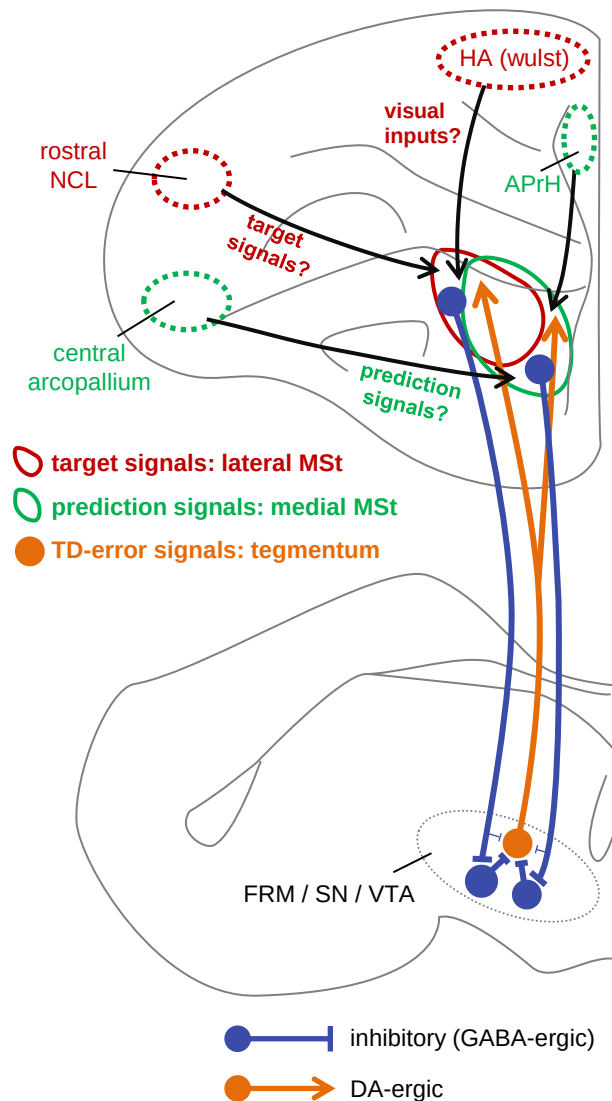


Figure 12. Hypothetical neuronal mechanism for TD error computation. The target signal and the prediction signal are coded by different but anatomically overlapping populations of neurons in the lateral MSt (solid red circle) and medial MSt (solid green circle), respectively. These two signals are sent to the tegmentum (FRM/SN/VTA) by GABA-ergic projection neurons (blue). In the tegmentum, these two signals converge via inhibitory local GABA-ergic inter-neurons. Through this convergent disinhibition, the TD error signal appears in DA-ergic neurons (orange filled circle). The DA-ergic neurons project back to a wide range of striatal areas, including the MSt (orange line arrow). In contrast, the lateral and medial MSt receive inputs from different pallial regions, such as the central arcopallium, rostral NCL, HA (Wulst), and APrH in the

hippocampal complex. Red circles (HA and rostral NCL) indicate regions that project mainly to the lateral MSt, while green circles (central arcopallium and APrH) indicate regions that project mainly to the medial MSt.

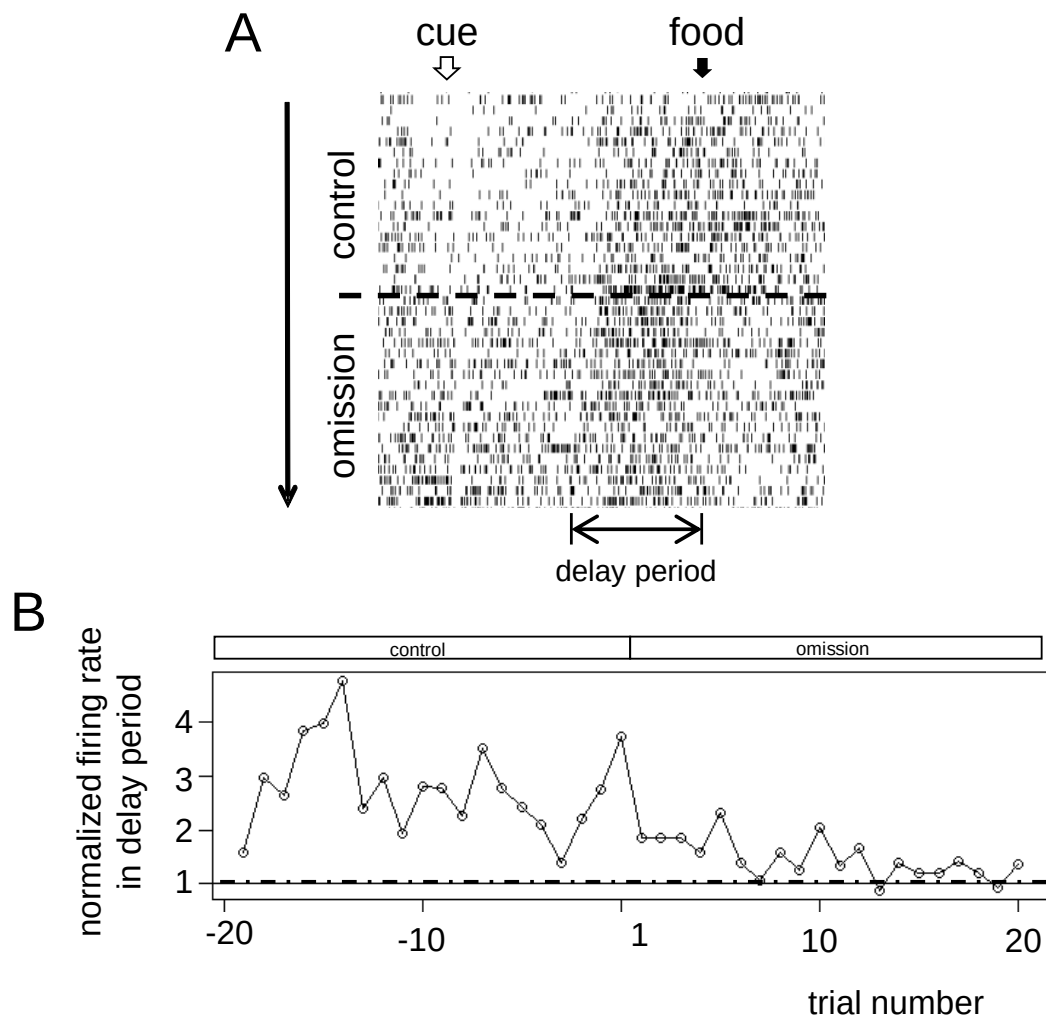


Figure S1. Delay-period activity of a type-1 MSt neuron. (A) Rastergram.(B) Normalized firing rate of the delay-period activities (1.5s~3.0s) is plotted against the trial number.

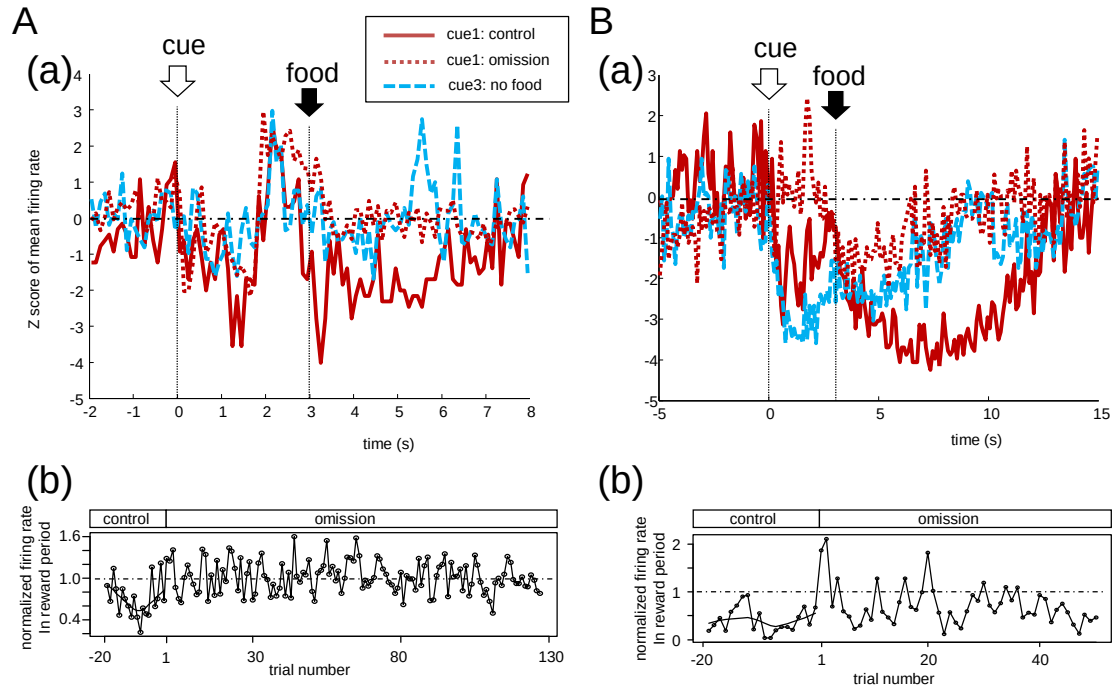


Figure S2. Two representative examples of the inhibitory type-1 neurons in MSt. (A) Neuron#78. (a) Averaged firing rate (z-score) in the cue1 trials is compared between the control block (red line) and the omission block (red dashed line). Data obtained in cue3 trials (blue dashed line) is superimposed. (b) Normalized firing rate in the reward period is plotted against the trial number. (B) Neuron #5. Conventions are as in (A). (a) Cue-/peck-/delay-period inhibition was followed by a further imbibition in the reward-period.

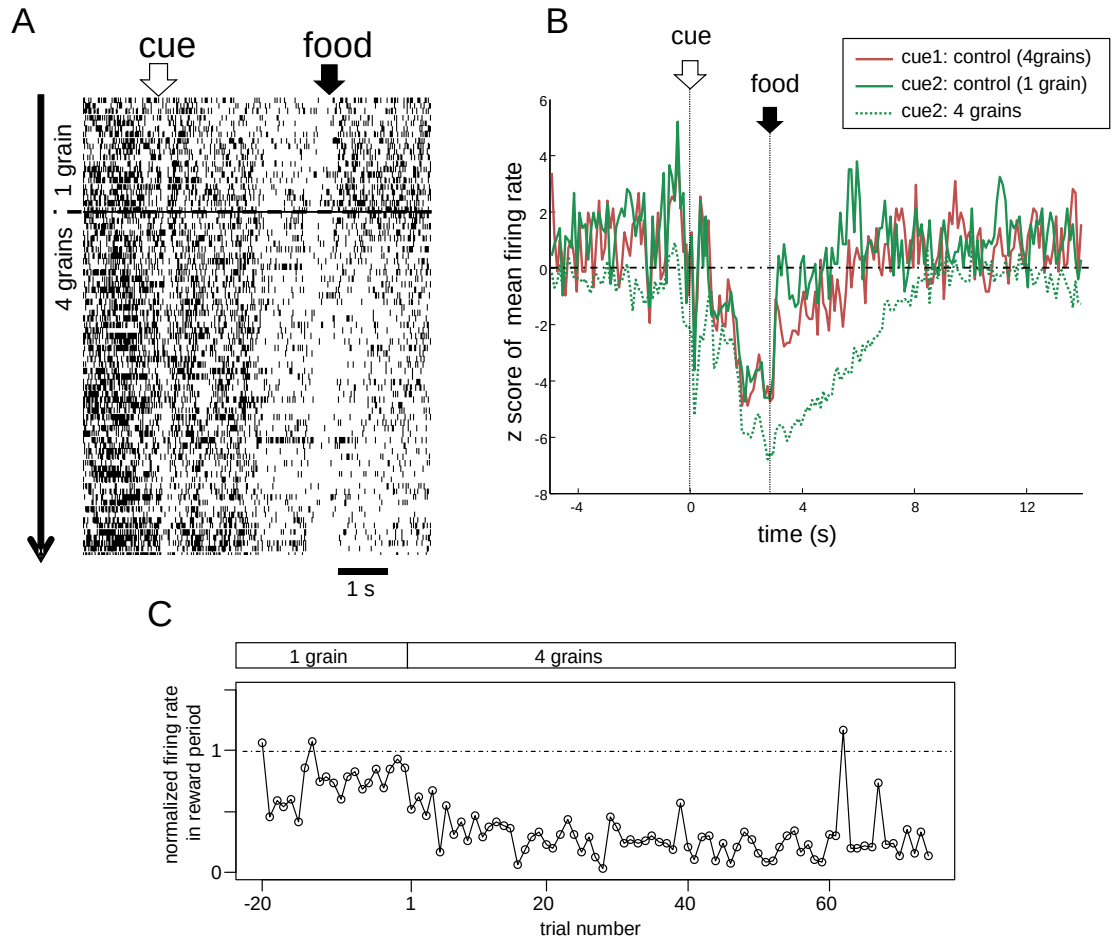


Figure S3. An example of inhibitory subtype of the type3 neuron in MSt. In cue2 trials, reward increased from 1 grain to 4 grains. The same neuron as shown in Fig. 5C. After increase in the food reward, the amplitude of the reward-period activity gradually increased. (A) Rastergram of the cue2 trials. (B) Averaged firing rate (z-score). (C) Normalized firing rate in reward period of cue2 trials is plotted against the trial number.

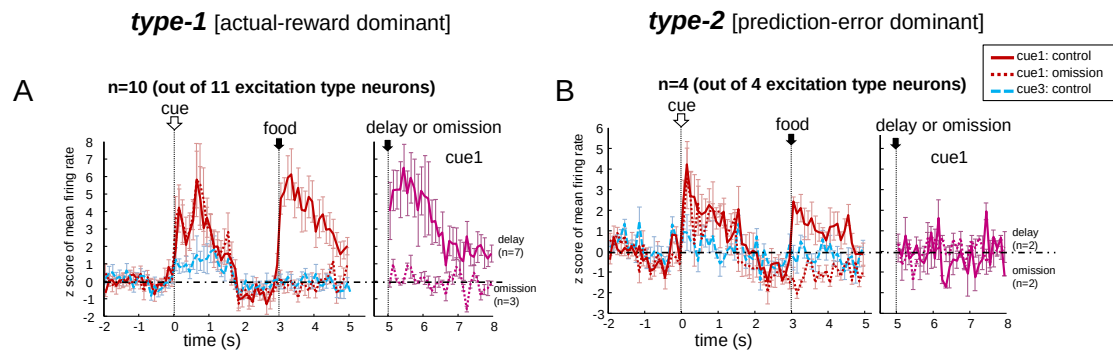


Figure S4. Population average of the type-1 (A) and type-2 (B) neuron activities in tegmentum.

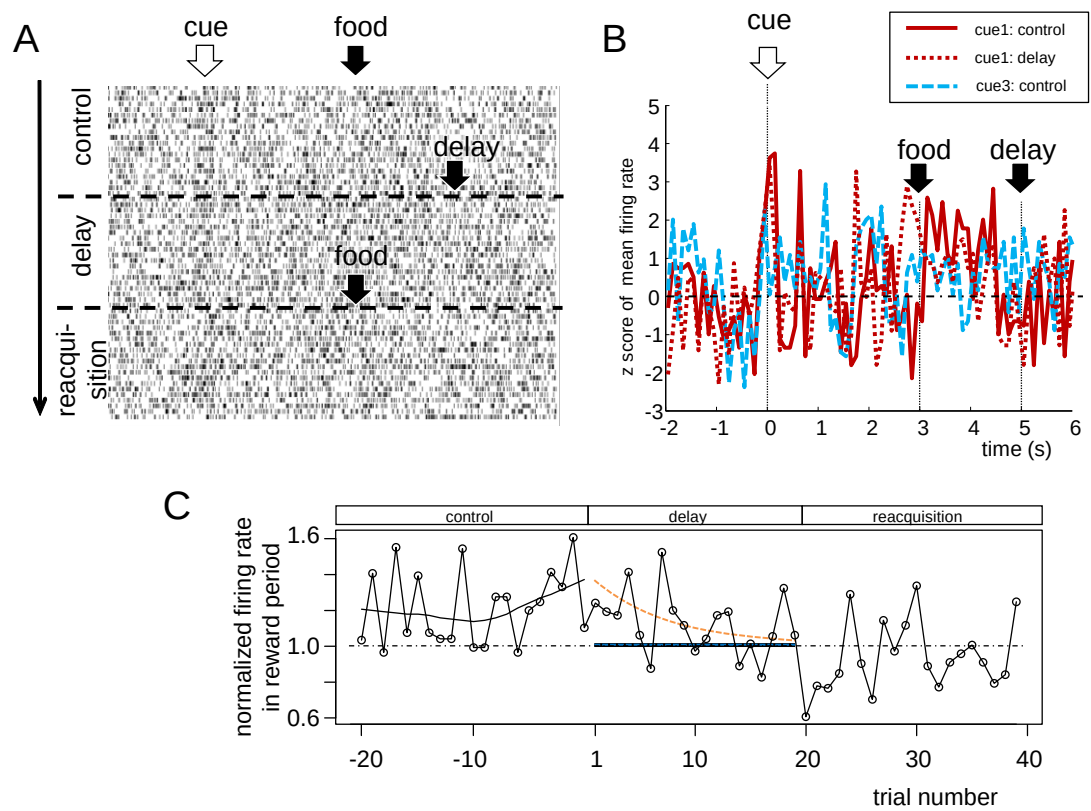


Figure S5. A representative example of excitatory subtype of the type-3 neuron. (A) Rastergram. (B) Averaged firing rate. (C) The reward-period excitation plotted against the trial number.

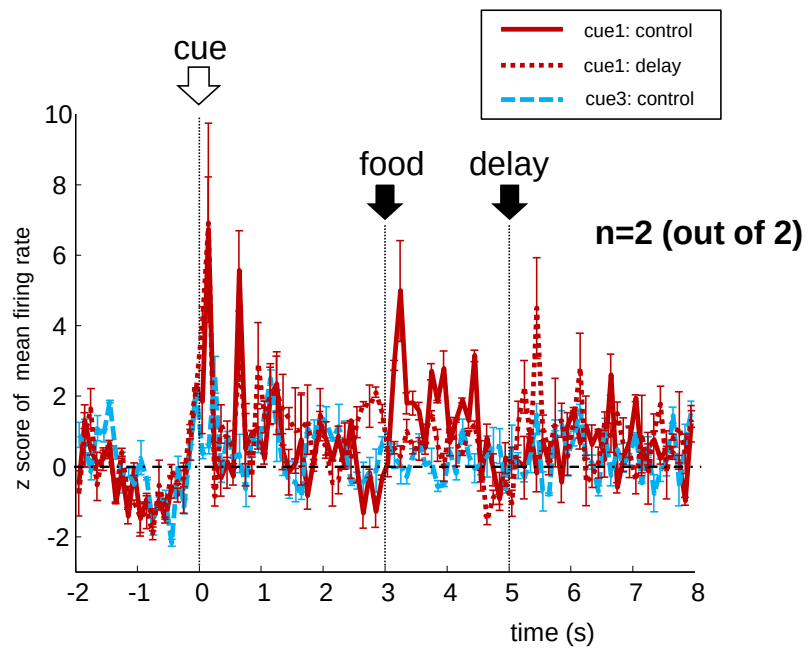


Figure S6. Averaged firing rate of two excitation subtype of the type-3 neurons.

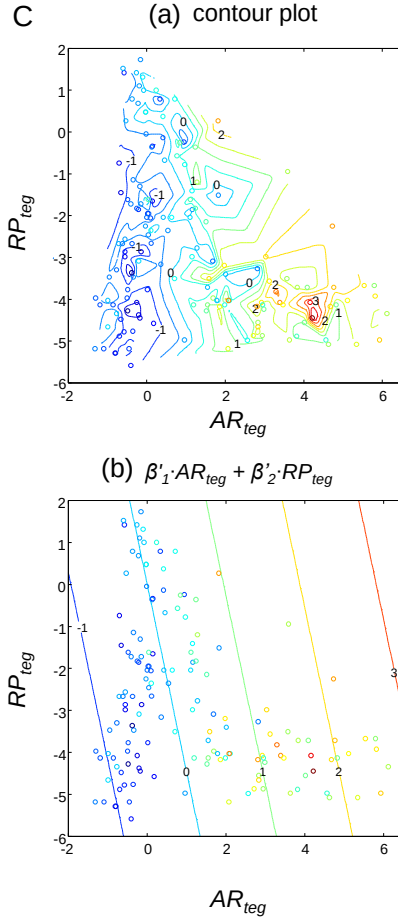
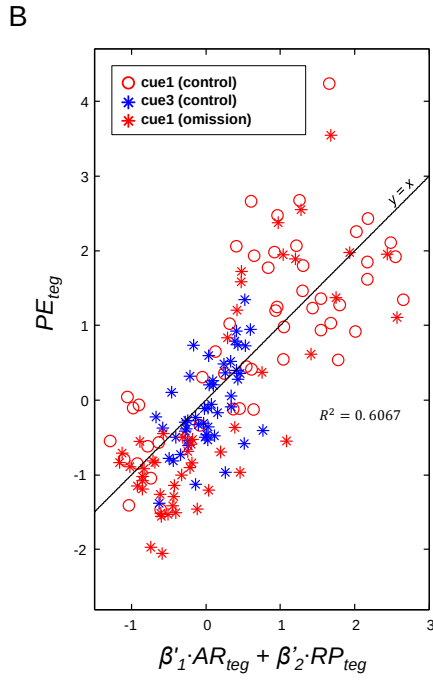
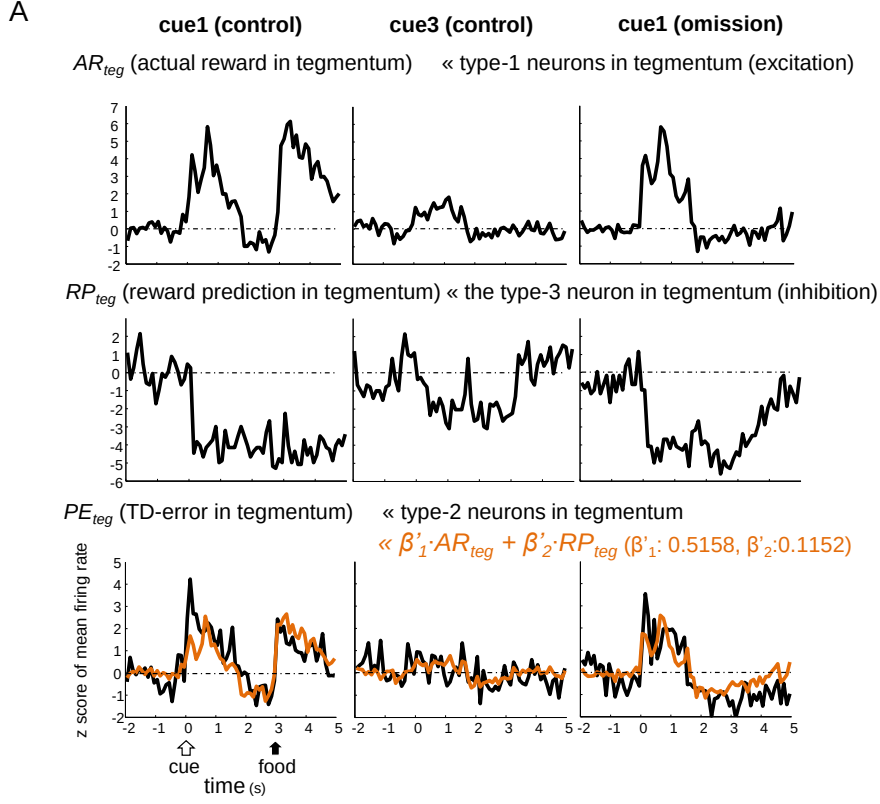


Figure S7. Type-2 neuronal activity in tegmentum (presumptive TD-error signal, PE_{teg}) fitted as a linear sum of the activities observed in a subset of the tegmentum neurons. **(A)** Averaged activities of the type-1 (AR_{teg}) and type-3 (RP_{teg}) tegmentum neurons are shown together with those of the type-2 tegmentum neurons (PE_{teg}) for 3 blocks of trials, cue1 (control), cue3 (control) and cue1 (omission). Superimposed orange lines on the PE_{teg} signal indicate the linear sum of AR_{teg} and RP_{teg} . **(B)** Scatter plot of PE_{teg} versus the linear sum. **(C)**. Pseudo 3-D plots of AR_{teg} (x-axis), RP_{teg} (y-axis) and PE_{teg} (color code) with the interpolated contour plot **(a)** and linear summation **(b)**.

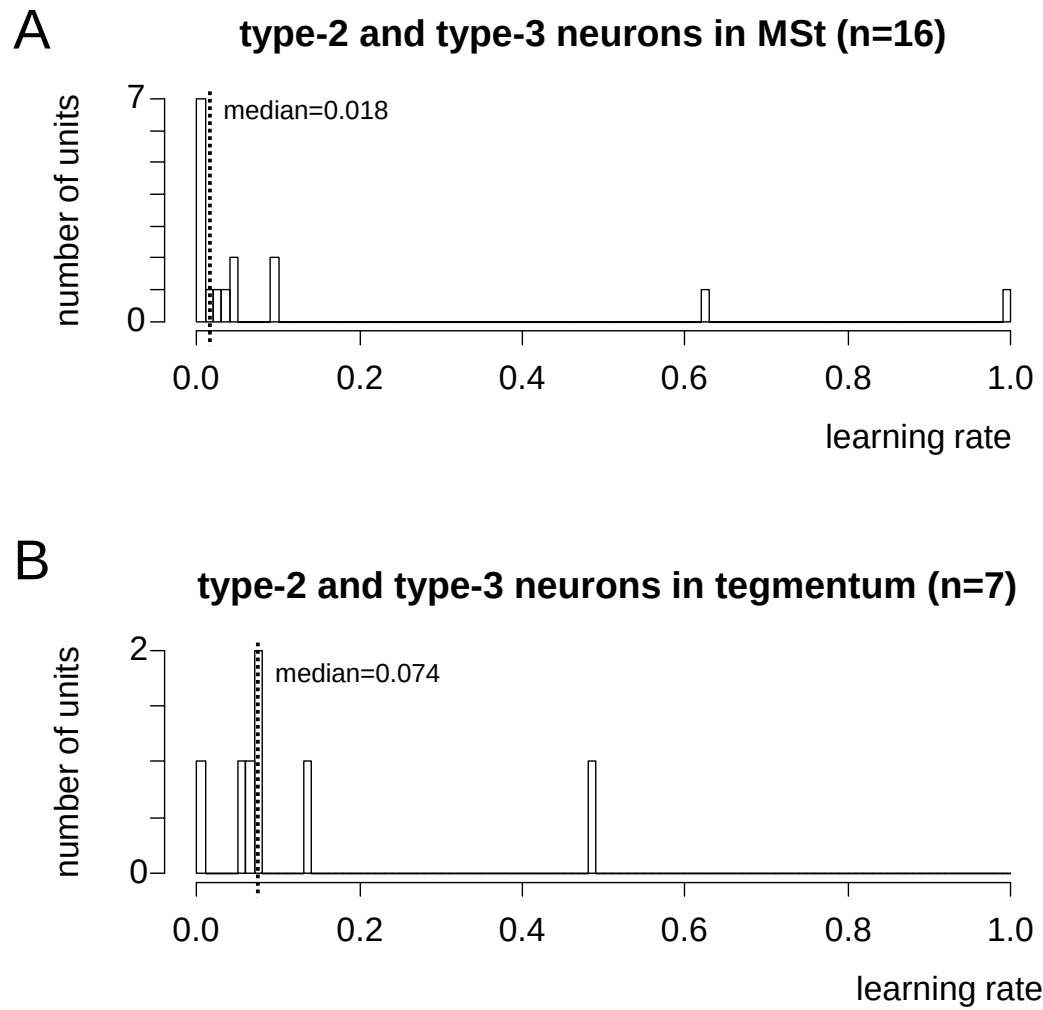


Figure S8. Estimated learning rate (α) of individual neurons. Only the reward-period activities in the type-2 and type-3 neurons are included. (A) MSt. (B) tegmentum.

6 Table

Target region	Distance (mm)			Angle between insertion and vertical line
	Anterior from bregma	Lateral from midline	Depth from brain surface	
MSt	1.5~2.0	1.25~1.5	5.0~5.5	least 25° caudally
Anterior tegmentum	-0.5~1.0	0.9~2.75	6.8~9.0	14° rostrally
Posterior tegmentum	-3.3~-2.5	1.5	8.5~9.0	vertically

Table S1. The coordinates of the target regions for tetrodes implantations and tracer injection.

Acknowledgements

First and foremost I am grateful to my supervisor Prof. Toshiya Matsushima. He gave me sufficient support to complete this PhD study. He also taught me lots including knowledges and experimental skills of neuroscience. I also thank for his patient instructions and assistances in writing this thesis.

I would also like to thank the members of the dissertation committee: Prof. Makoto Mizunami, Hiroto Ogawa, and Kazuyuki Samejima, for the critical discussions and important suggestions. I thank Dr. Kazuhiro Wada for his assistances with my experiment and critical comments. Thank Dr. Masayo Soma for her critical comments on my study.

I am grateful to those previous and current members in Matsushima's lab, and other members in the sub-department of Behavioural Neurobiology. Thank Dr. Hidetoshi Amita for his instruction and assistances with my electrophysiological experiment. Thank Dr. Yukiko Ogura for her assistances with my immunohistochemical experiment. Thank Qihong Xin for her assistances with my tract-tracing experiment. Thank other members for their help and advices.

Finally, I must thank my family for their encouragement and patience. Without their support, I may not complete such an exciting study in a foreign country.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723.
- Amita, H., and Matsushima, T. (2014). Competitor suppresses neuronal representation of food reward in the nucleus accumbens/medial striatum of domestic chicks. *Behav. Brain Res.* 268, 139–149.
- Anderson, K. D., Karle, E. J., and Reiner, A. (1991). Ultrastructural single-and double-label immunohistochemical studies of substance P-containing terminals and dopaminergic neurons in the substantia nigra in pigeons. *J. Comp. Neurol.* 309, 341–362.
- Annett, L., McGregor, A., and Robbins, T. (1989). The effects of ibotenic acid lesions of the nucleus accumbens on spatial learning and extinction in the rat. *Behav. Brain Res.* 31, 231–242.
- Aoki, N., Csillag, A., and Matsushima, T. (2006). Localized lesions of arcopallium intermedium of the lateral forebrain caused a handling-cost aversion in the domestic chick performing a binary choice task. *Eur. J. Neurosci.* 24, 2314–2326.
- Aoki, N., Izawa, E. I., Yanagihara, S., and Matsushima, T. (2003). Neural correlates of memorized associations and cued movements in archistriatum of the domestic chick. *Eur. J. Neurosci.* 17, 1935–1946.
- Apicella, P., Deffains, M., Ravel, S., and Legallet, E. (2009). Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. *Eur. J. Neurosci.* 30, 515–526.
- Bálint, E., Mezey, S., and Csillag, A. (2011). Efferent connections of nucleus accumbens subdivisions of the domestic chicken (*Gallus domesticus*): an anterograde pathway tracing study. *J. Comp. Neurol.* 519, 2922–2953.

- Barto, A. G. (1995). “Adaptive critics and the basal ganglia,” in *Models of Information Processing in the Basal Ganglia*, eds. J. C. Houk, J. Davis, and D. Beiser (MIT Press), 215–232.
- Bocklisch, C., Pascoli, V., Wong, J. C., House, D. R., Yvon, C., De Roo, M., et al. (2013). Cocaine disinhibits dopamine neurons by potentiation of GABA transmission in the ventral tegmental area. *Science* 341, 1521–1525.
- Castañé, A., Theobald, D. E., and Robbins, T. W. (2010). Selective lesions of the dorsomedial striatum impair serial spatial reversal learning in rats. *Behav. Brain Res.* 210, 74–83.
- Clarke, H. F., Robbins, T. W., and Roberts, A. C. (2008). Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *J. Neurosci.* 28, 10972–10982.
- Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482, 85–88.
- Diekamp, B., Kalt, T., and Güntürkün, O. (2002). Working memory neurons in pigeons. *J. Neurosci.* 22, RC210 (1–5).
- Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Curr. Opin. Neurobiol.* 10, 732–739.
- Doya, K. (2007). Reinforcement learning: Computational theory and biological mechanisms. *HFSP J.* 1, 30–40. doi:10.2976/1.2732246.
- Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., and Uchida, N. (2015). Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* 525, 243–246.

- Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299, 1898–1902.
- Gale, S. D., and Perkel, D. J. (2006). Physiological properties of zebra finch ventral tegmental area and substantia nigra pars compacta neurons. *J. Neurophysiol.* 96, 2295–2306.
- Grillner, S., Hellgren, J., Ménard, A., Saitoh, K., and Wikström, M. A. (2005). Mechanisms for selection of basic motor programs—roles for the striatum and pallidum. *Trends Neurosci.* 28, 364–370.
- Houk, J. C., Adams, J. L., and Barto, A. G. (1995). “A model of how the basal ganglia might generate and use neural signals that predict reinforcement,” in *Models of Information Processing in the Basal Ganglia*, eds. J. C. Houk, J. Davis, and D. Beiser (MIT Press), 249–270.
- Ichikawa, Y., Izawa, E.-I., and Matsushima, T. (2004). Excitotoxic lesions of the medial striatum delay extinction of a reinforcement color discrimination operant task in domestic chicks; a functional role of reward anticipation. *Cogn. Brain Res.* 22, 76–83.
- Izawa, E.-I., Aoki, N., and Matsushima, T. (2005). Neural correlates of the proximity and quantity of anticipated food rewards in the ventral striatum of domestic chicks. *Eur. J. Neurosci.* 22, 1502–1512.
- Izawa, E.-I., Yanagihara, S., Atsumi, T., and Matsushima, T. (2001). The role of basal ganglia in reinforcement learning and imprinting in domestic chicks. *Neuroreport* 12, 1743–1747.
- Izawa, E.-I., Zachar, G., Yanagihara, S., and Matsushima, T. (2003). Localized lesion of caudal part of lobus parolfactorius caused impulsive choice in the domestic chick: evolutionarily conserved function of ventral striatum. *J. Neurosci.* 23, 1894–1902.

- Janak, P. H., Chen, M.-T., and Caulder, T. (2004). Dynamics of neural coding in the accumbens during extinction and reinstatement of rewarded behavior. *Behav. Brain Res.* 154, 125–135.
- Joel, D., Niv, Y., and Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Netw.* 15, 535–547.
- Kawagoe, R., Takikawa, Y., and Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* 1, 411–416.
- Kawato, M., and Samejima, K. (2007). Efficient reinforcement learning: computational theories, neuroscience and robotics. *Curr. Opin. Neurobiol.* 17, 205–212.
- Kim, H., Sul, J. H., Huh, N., Lee, D., and Jung, M. W. (2009). Role of striatum in updating values of chosen actions. *J. Neurosci.* 29, 14701–14712.
- Kröner, S., and Güntürkün, O. (1999). Afferent and efferent connections of the caudolateral neostriatum in the pigeon (*Columba livia*): a retro-and anterograde pathway tracing study. *J. Comp. Neurol.* 407, 228–260.
- Kuenzel, W. J., and Masson, M. (1988). *A Stereotaxic Atlas of the Brain of the Chick (Gallus Domesticus)*. Johns Hopkins University Press.
- Matsumoto, M., and Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* 447, 1111–1115.
- Matsushima, T., Izawa, E.-I., Aoki, N., and Yanagihara, S. (2003). The mind through chick eyes: memory, cognition and anticipation. *Zoolog. Sci.* 20, 395–408.
- Mezey, S., and Csillag, A. (2002). Selective striatal connections of midbrain dopaminergic nuclei in the chick (*Gallus domesticus*). *Cell Tissue Res.* 308, 35–46.
- Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947.

- Morris, G., Arkadir, D., Nevet, A., Vaadia, E., and Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 43, 133–143.
- Morris, G., Nevet, A., Arkadir, D., Vaadia, E., and Bergman, H. (2006). Midbrain dopamine neurons encode decisions for future action. *Nat. Neurosci.* 9, 1057–1063.
- Nash, J. C. (2014). On best practice optimization methods in R. *J. Stat. Softw.* 60, 1–14. Available at: <http://www.jstatsoft.org/v60/i02/>.
- Nash, J. C., and Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for R. *J. Stat. Softw.* 43, 1–14. Available at: <http://www.jstatsoft.org/v43/i09/>.
- Niv, Y., and Montague, P. R. (2009). “Theoretical and empirical studies of learning,” in *Neuroeconomics: Decision Making and the Brain*, eds. P. W. Glimcher, C. F. Camerer, E. Fehr, and R. A. Poldrack (Academic Press), 331–351.
- Ocklenburg, S., and Güntürkün, O. (2012). Hemispheric asymmetries: The comparative view. *Front. Psychol.* 3. doi:10.3389/fpsyg.2012.00005.
- Ogura, Y., Izumi, T., Yoshioka, M., and Matsushima, T. (2015). Dissociation of the neural substrates of foraging effort and its social facilitation in the domestic chick. *Behav. Brain Res.* 294, 162–176.
- Ogura, Y., and Matsushima, T. (2011). Social facilitation revisited: increase in foraging efforts and synchronization of running in domestic chicks. *Front. Neurosci.* 5. doi:10.3389/fnins.2011.00091.
- Okada, K., Toyama, K., Inoue, Y., Isa, T., and Kobayashi, Y. (2009). Different pedunculo pontine tegmental neurons signal predicted and actual task rewards. *J. Neurosci.* 29, 4858–4870.
- Oyama, K., Hernádi, I., Iijima, T., and Tsutsui, K.-I. (2010). Reward prediction error coding in dorsal striatal neurons. *J. Neurosci.* 30, 11447–11457.

- Oyama, K., Tateyama, Y., Hernádi, I., Tobler, P. N., Iijima, T., and Tsutsui, K.-I. (2015). Discrete coding of stimulus value, reward expectation, and reward prediction error in the dorsal striatum. *J. Neurophysiol.* 114, 2600–2615.
- Pavlov, I. P. (1927). *Conditioned Reflexes: an Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford university press.
- Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic Linear Models with R*. Springer.
- Reiner, A., and Anderson, K. D. (1990). The patterns of neurotransmitter and neuropeptide co-occurrence among striatal projection neurons: conclusions based on recent findings. *Brain Res. Rev.* 15, 251–265.
- Reiner, A., Perkel, D. J., Bruce, L. L., Butler, A. B., Csillag, A., Kuenzel, W., et al. (2004). Revised nomenclature for avian telencephalon and some related brainstem nuclei. *J. Comp. Neurol.* 473, 377–414.
- Roesch, M. R., Calu, D. J., and Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.* 10, 1615–1624.
- Rueda-Orozco, P. E., Montes-Rodriguez, C. J., Soria-Gomez, E., Méndez-Díaz, M., and Prospéro-García, O. (2008). Impairment of endocannabinoids activity in the dorsolateral striatum delays extinction of behavior in a procedural memory task in rats. *Neuropharmacology* 55, 55–62.
- Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science* 310, 1337–1340.
- Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* 3, 210–229.
- Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. *Physiol. Rev.* 95, 853–951.

- Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.
- Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* 27, 8366–8377.
- Shidara, M., Aigner, T. G., and Richmond, B. J. (1998). Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J. Neurosci.* 18, 2613–2625.
- Shidara, M., and Richmond, B. J. (2002). Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* 296, 1709–1711.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489.
- Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT press.
- Székely, A. D., Boxer, M. I., Stewart, M. G., and Csillag, A. (1994). Connectivity of the lobus parolfactorius of the domestic chicken (*Gallus domesticus*): an anterograde and retrograde pathway tracing study. *J. Comp. Neurol.* 348, 374–393.
- Tesauro, G. (1995). Temporal difference learning and TD-Gammon. *Commun. ACM* 38, 58–68.
- Tremblay, L., Hollerman, J. R., and Schultz, W. (1998). Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J. Neurophysiol.* 80, 964–977.
- Veenman, C. L., and Reiner, A. (1994). The distribution of GABA-containing perikarya, fibers, and terminals in the forebrain and midbrain of pigeons, with particular

reference to the basal ganglia and its projection targets. *J. Comp. Neurol.* 339, 209–250.

Veenman, C. L., Wild, J. M., and Reiner, A. (1995). Organization of the avian “corticostriatal” projection system: a retrograde and anterograde pathway tracing study in pigeons. *J. Comp. Neurol.* 354, 87–126.

Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A., and Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74, 858–873.

Yanagihara, S., Izawa, E.-I., Koga, K., and Matsushima, T. (2001). Reward-related neuronal activities in basal ganglia of domestic chicks. *Neuroreport* 12, 1431–1435.

List of other publications

Article

1. Matsushima, T., **Wen, C.**, Ogura, Y., and Amita, H.

「競争・同調・社会的促進」計測自動制御学会誌（計測と制御），52 巻，3 号，pp. 189–194（2013）

Conference presentations (oral)

1. **Wen, C.**, and Matsushima, T. Neurons of nucleus accumbens code the value of gained reward in the domestic chick, 日本動物学会北海道支部第 57 回大会、札幌、2012 年 8 月
2. **Wen, C.**, and Matsushima, T. Medial striatum simultaneously codes actual reward and reward prediction in domestic chicks, 日本動物学会第 84 回大会、岡山、2013 年 9 月

Conference presentations (poster)

1. **Wen, C.**, and Matsushima, T. Neuronal activities in the medial striatum of domestic chicks in the reward period may code the expectation of reward to calculate the prediction error, Society for Neuroscience 43rd annual meeting, San Diego, California, USA, November, 2013
2. Amita, H., **Wen, C.**, and Matsushima, T. Pseudo-competition suppresses subjective values of food in terms of predicted and gained food rewards represented by the nucleus accumbens/medial striatum, Society for Neuroscience 43rd annual meeting, San Diego, California, USA, November, 2013
3. **Wen, C.**, and Matsushima, T. Neuronal activities in the medial striatum of domestic chicks in the reward period may code the expectation of reward to calculate the prediction error, 「脳と心のメカニズム」第 14 回 冬のワークショップ, 留寿都, 北海道, 2014 年 1 月
4. **Wen, C.**, and Matsushima, T. Neurons in medial striatum of domestic chicks may code actual reward and reward prediction to compute prediction error signal, 11th International Congress of Neuroethology, Sapporo, Japan, July, 2014
5. **Wen, C.**, and Matsushima, T. Computation of reward prediction error by projections from medial striatum to midbrain dopaminergic neurons in domestic chicks, Society for Neuroscience 45rd annual meeting, Chicago, Illinois, USA, October, 2015

Appendix in response to the referees' comments

On the preliminary examination on my initial version of the thesis (16 June, 2016), Prof Makoto Mizunami (Hokkaido University), Prof Hiroto Ogawa (Hokkaido University) and Prof Kazuyuki Samejima (Tamagawa University) gave me a series of instructive comments and suggestions. In response to these comments, I added an appendix that includes further analyses on the obtained data.

In Figures A1 and A2, as Prof. Samejima and Prof. Mizunami required, the neuronal activities in cue2 trials were compared between control block and omission block. As figures show, those activities did not change when the associated amount of the cue2 reward was unchanged. On the other hand, when the cue2 amount increased, the type-3 neurons showed a larger inhibitory responses in the reward period, while type-1,-2 tegmental neurons did not show change.

In Figure A3, as Prof. Samejima required, I examined whether the pure reward signal r_t exist. As figures show, the type-1, -2 neurons in MSt and the type-1 neurons in tegmentum did not code r_t , because of their responses in pre-reward periods.

In Figure A4, as Prof. Samejima required, I examined the cross-correlations of 3 pairs of categorized neurons. The results imply possible local-interactions between those neurons. It is of particularly interesting, though in a small number of successful examination, the type-1 neuron inhibited the concomitantly recorded type-2 and -3 neurons. This fact is not contradictory with my idea that the type-1 neurons are inhibitory in nature.

In Figure A5, as Prof. Samejima required, I counted the putative contacts between boutons of MSt terminal and the TH-ir cells in tegmentum. Those boutons with direct contacts with the proximal dendrites and cell bodies constituted ca. 10% of the total boutons counted. As the TH staining failed to visualize the fine dendritic trees, this figure is supposed to underestimate the proportion of direct contacts. On the other hand, I have no direct evidence for the contacts on non-TH positive (non-dopaminergic) neurons in the tegmentum.

MSt

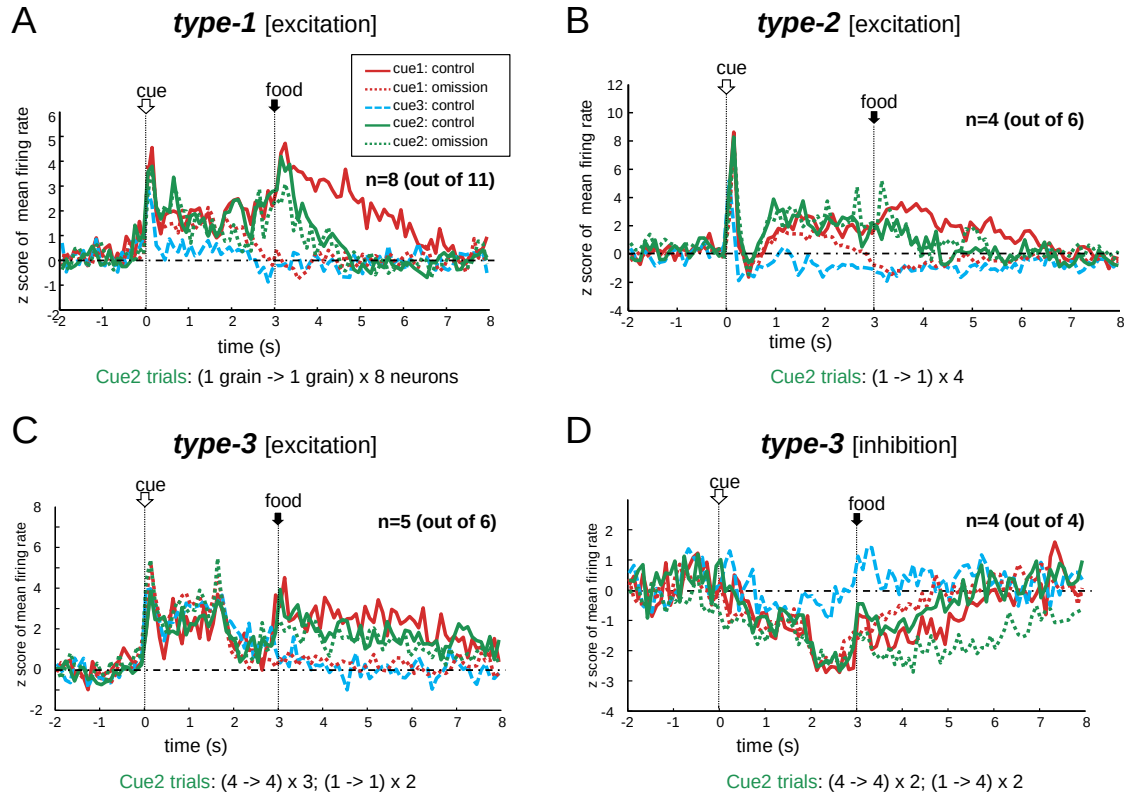


Figure A1. Activities of MSt neurons including cue2 trials.

The averaged z scores of mean firing rate are plotted as curves. MSt neurons were recorded from #1 to #88. The associations in cue2 trials of these neurons are as follows: (1) in neurons #1-8, 4 grains to 4 grains; (2) in neurons #25 and #61-88, 1 grains to 4 grains; (3) in other neurons, 1 grain to 1 grain. In (A), (B), and (C), the amount of food in cue2 trials did not change. The corresponding neuronal activities also kept unchanged in all of the periods. In (D), the amount of food increased in two out of four neurons in cue2 trials. Consequently, the amplitude of the inhibitory response in reward period was strengthened.

Tegmentum

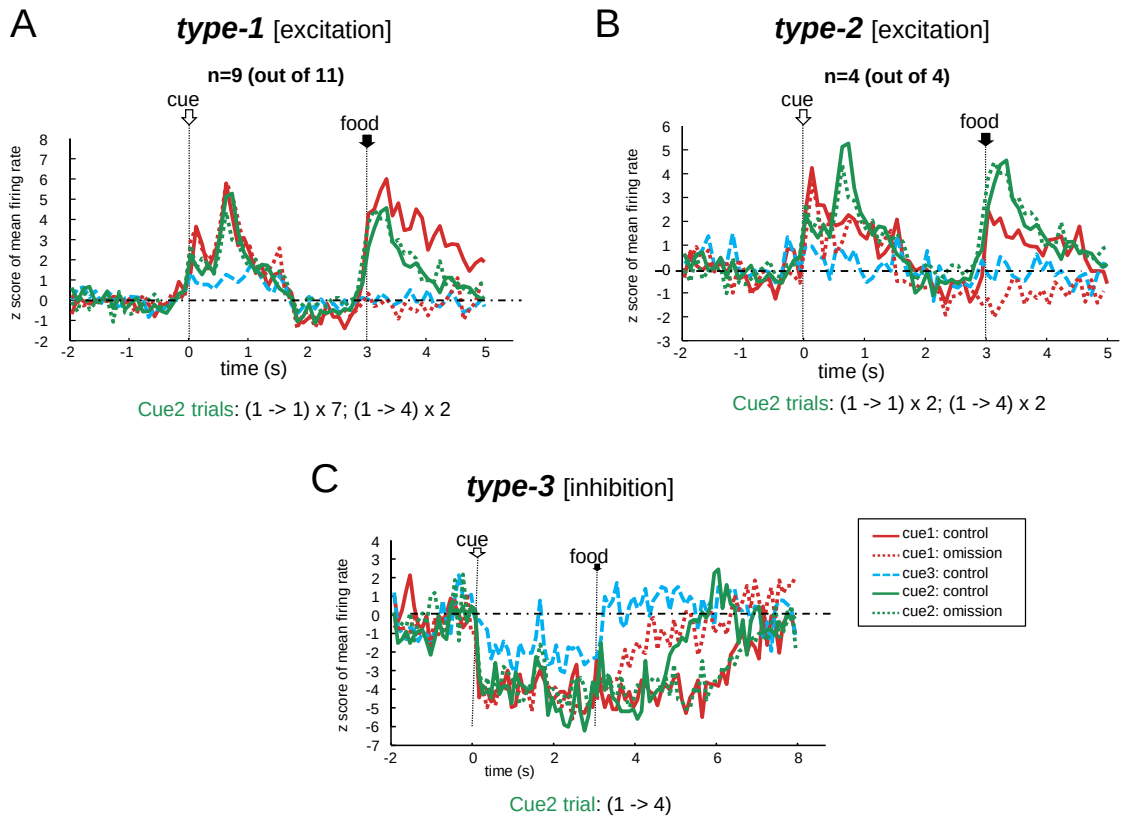
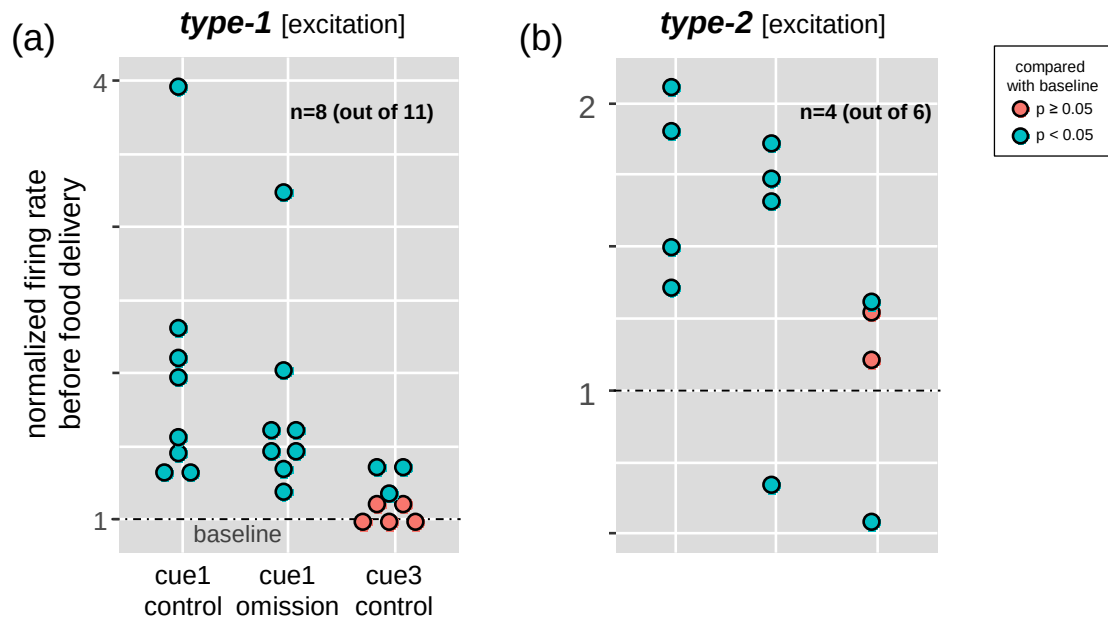


Figure A2. Activities of tegmental neurons including cue2 trials.

The averaged z scores of mean firing rate are plotted as curves. Tegmental neurons were recorded from #1 to #39. The associations in cue2 trials of these neurons are as follows: (1) in neurons #11-29, 1 grain to 1 grain. (2) in other neurons, 1 grain to 4 grains. In (A) and (B), the responses in cue2 trials kept unchanged, although the food increased for some neurons. This indicates these neurons may not be sensitive to the amount, but only to the existence of the food. In (C), the inhibitory response in reward period was strengthened after the food increased.

A MSt



B Tegmentum

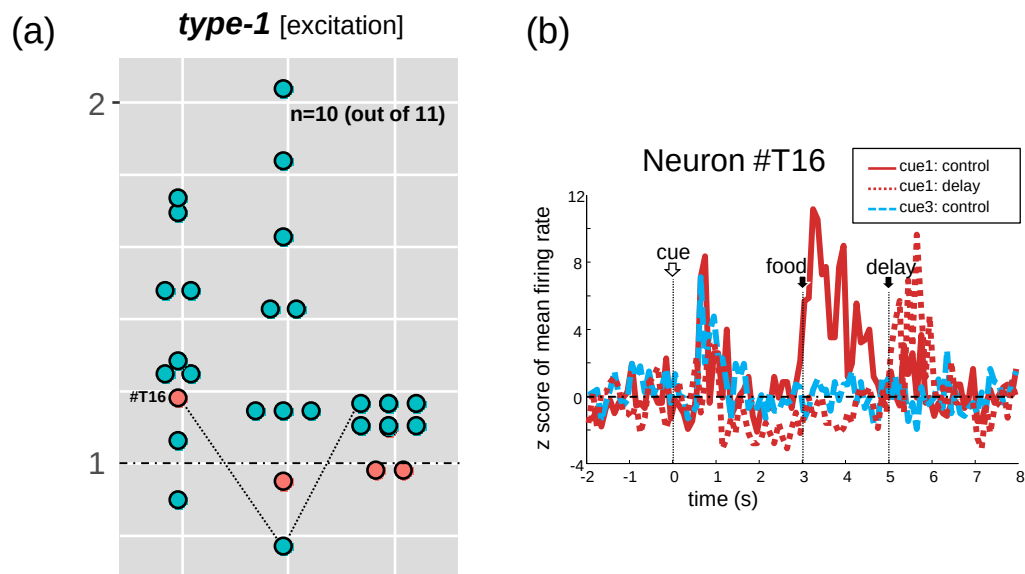


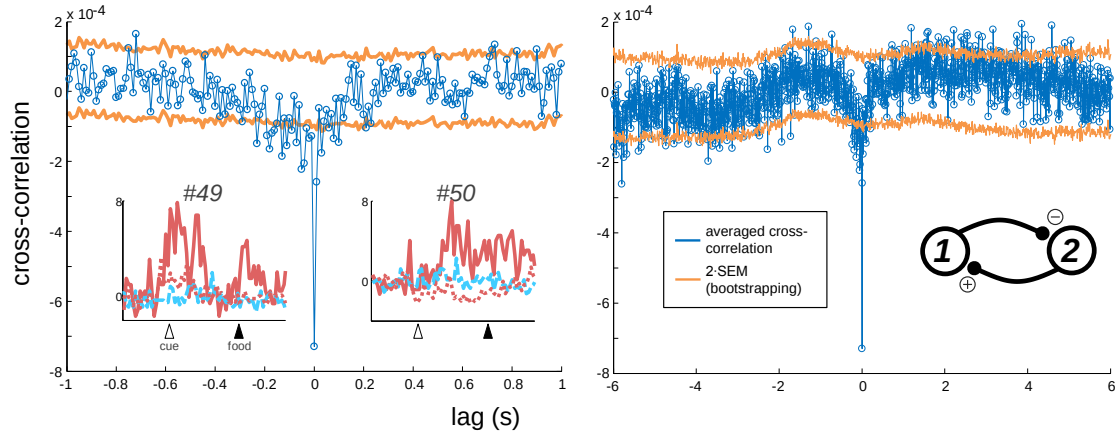
Figure A3. Pre-reward activities of those reward-coding neurons.

According to the results and discussions in the main text, type-1, -2 neurons in MSt and type-1 neurons in tegmentum code actual reward in reward period. Their pre-reward activities (from $t=0$ s to $t=3$ s) were calculated and were compared with the baseline firing rate (from $t=-5$ s to $t=0$ s). Blue circles indicate those neurons with significant response before food delivery (one sample t test compared with the baseline, 1). Red circles indicate the neurons with non-significant activities. In (A), all type-1 and type-2 MSt neurons (excitation subtype) show significant responses in cue1 control block, thus these neurons do not code pure reward signal. The inhibition subtype of type-1 MSt neurons also do not code the pure reward signal, see Fig. S2. In (Ba), only one excitation subtype of type-1 tegmental neuron shows non-significant response in cue1 control block. However, the firing pattern of this neuron (Bb) indicates a short delayed response to the cues, thus this neuron does not code the pure reward signal. The inhibition subtype of type-1 tegmental neurons also does not code the pure reward signal, according to their firing patterns (not shown here).

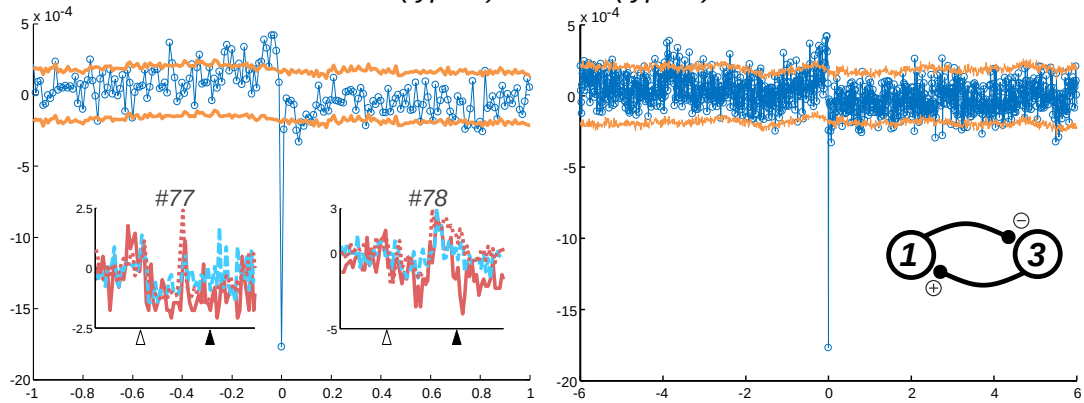
Cross-correlation between simultaneously recorded neurons

MSt

#49 (type-1) and #50 (type-2)



#77 (type-3) and #78 (type-1)



Tegmentum

#22 (type-1) and #23 (type-3)

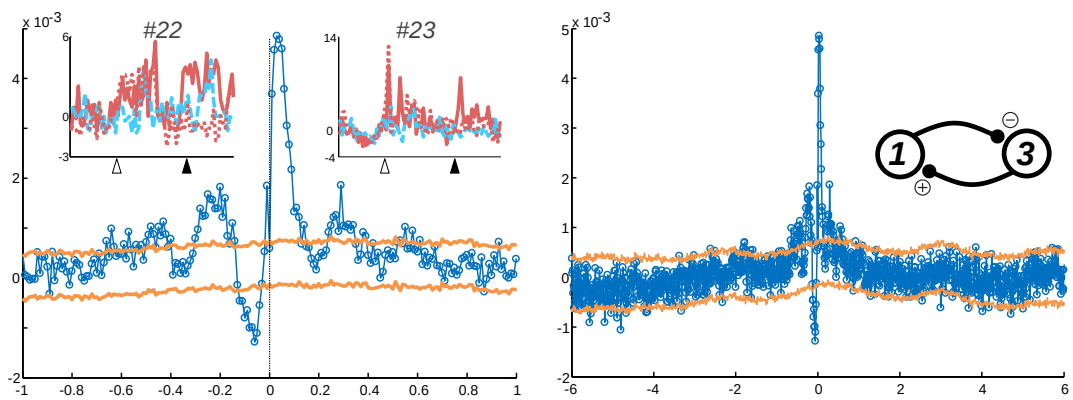
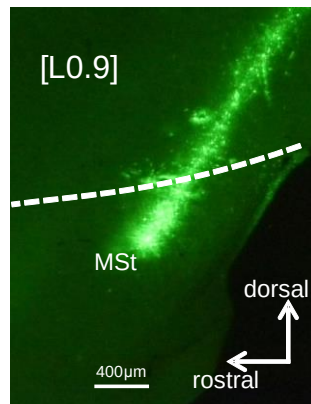


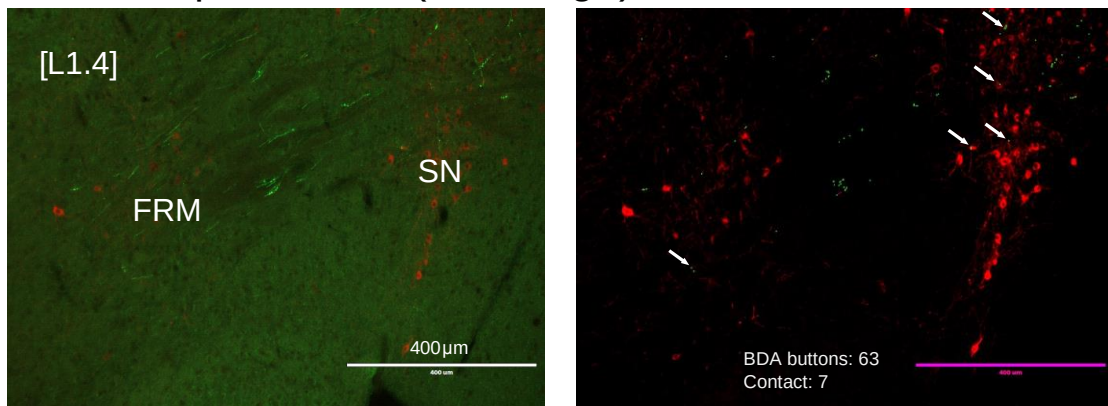
Figure A4. Cross-correlation between 3 pairs of simultaneously recorded neurons.

Only 3 pairs of simultaneously recorded neurons are of the categorized neurons. They are shown here. The cross-correlation was calculated between the binned spike trains of neuron_A($t+\text{lag}$) and neuron_B(t). The bin width is 10ms. Left side and right side show the same results in different temporal scales. Blue circles together with the blue lines indicate the averaged cross-correlation across trials. Orange curves indicate the $\text{mean} \pm 2 \cdot \text{SEM}$ (standard error of the mean) of the cross-correlation, which is an approximate estimate of the 95% confidence interval. The mean and the SEM are estimated using the resampling method, bootstrapping, based on spike trains of neuron_A and neuron_B from different trials. The inlet figures of left side are the firing patterns from the corresponding neurons. The inlet figures of right side are the possible connectivity of these 3 pairs of neurons. The numbers in circles indicate the neuron types. Plus and minus symbols indicate the excitatory and inhibitory effects, respectively. Although in these figures neurons were directly connected with each other, the real connections are not necessarily monosynaptic. It is also possible that those neurons receive common inputs, which made the effects. These preliminary analyses imply that the local computation of TD error may occur in MSt or in tegmentum.

A injection site (150309g3)



B an example section (150309g3)



C statistics of the contacts

brain	BDA buttons (varicosities)	contacts between buttons and TH cells			
		SN	VTA	FRM	total
(a)150309g3	355	19	3	13	35
(b)150115g1	1056	88	3	6	97

Figure A5. Statistics of the contacts between BDA buttons and the TH-cells.

(A) The injection site of BDA (green) in brain 150309g3. (B) an example section of double staining of BDA (green) terminals and TH-ir (red) cells in the tegmentum. Left: original photo. Right: the same photo processed for analyzing contacts. Buttons were selected by a threshold and marked by green ellipses. TH-ir staining was enhanced. Arrows indicate the sites of contacts. (C) Statistics of the contacts in tegmentum of two brains. Laterality of injection sites are: (a) L0.9 and (b) L1.2. Range of brain sections for analysis: (a) L0.3-L1.4, 14 photos from 12 sections; (b) L1.0-L1.7, 16 photos from 8 sections. Around 10% of buttons of MSt terminals appears to contact with the TH-ir cells or dendrites in tegmentum. However, the analyses may overestimate the number of contacts with TH-ir soma, because they only confirmed the contacts on the x-y plane but not on the z axis. On the other hand, the number of contacts with TH-ir dendrites may be underestimated, because of the limited spatial resolution of the non-confocal microscopy.