Prefrontal Modulation of Midbrain Dopamine Systems during Navigation-based Decision Tasks

Yong Sang Jo

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Sheri J.Y. Mizumori, Chair

Jeansok J. Kim

Paul E.M. Phillips

Program Authorized to Offer Degree:

Department of Psychology

University of Washington

Abstract

Prefrontal modulation of midbrain dopamine systems during navigation-based decision tasks

Yong Sang Jo

Chair of the supervisory committee:
Professor Sheri J. Y. Mizumori
Department of Psychology

Midbrain dopamine (DA) systems are central for reinforcement learning. DA cells encode discrepancies between expected and received rewards in a phasic fashion. These errors in reward prediction may be used as a teaching signal by other brain regions for the learning of reward-directed behavior. Computation of prediction errors requires the value of future rewards estimated in a given situation or state. The literature suggests that the prefrontal cortex (PFC) represents reward expectancy in associative learning, so the PFC may be one of the critical structures that convey expected reward values to DA cells for signaling prediction errors. In order to address this possibility, the current dissertation investigated whether temporary inactivation of each of two major PFC, the orbitofrontal cortex (OFC) and the medial PFC (mPFC) disrupted DAergic prediction errors in the ventral tegmental area (VTA) while rats performed a delay-based decision making task on a T-maze. Significant alternations in firing of DA and non-DA cells commonly indicate that the OFC provides the VTA with value signals in the task. On the other hand, mPFC inactivation also induced significant changes in DA activity, but non-DA cells remained unaltered. These results suggest that the mPFC modulates DAergic prediction errors by conveying temporal information about time delays, rather than value signals.

Thus, the OFC, but not the mPFC, is the major prefrontal source of expected reward values to the VTA in the task.

Recently, it has become controversial as to whether the rodent OFC encodes value signals or state information about the current choice and its consequence. To examine the nature of neural representations in the OFC, single unit activity was recorded directly from the OFC of rats performing the same delay discounting task as in the prior experiments. Different groups of OFC cells showed excited responses to a series of task-relevant events and periods. In each state, individual neurons signaled a preferred reward condition by firing stronger than in the other reward conditions. In addition, their population activity reflected outcome values evaluated in each task state. These results provide compelling evidence that the OFC represents both specific outcome information and their relative values at the individual and population levels, respectively.

It has long been suggested that the midbrain reticular formation (MRF) encodes motivational value in expectation of future rewards. To determine its functional role in delay-based decision making, single unit activity was monitored from the MRF using the same delay discounting task. Consistent with the previous reports, a large number of MRF cells signaled information about expected rewards during waiting periods by continuously ramping their firing to different levels depending on the size of upcoming rewards. Thus, considering the direct projections from the MRF to the VTA, it is likely that the MRF is another source of value signals to midbrain DA systems.

Table of contents

Table of figures

Table of tables

Acknowledgement

I would like to thank my advisor Sheri Mizumori for continual support and guidance throughout my doctoral training at University of Washington. I am also thankful to my supervisory committee members, Drs. Jaime Diaz, Jeansok Kim, Paul Phillips, and John Neumaier for their critical insights and thoughtful comments and insights on my research. In addition, I wish to express my sincere gratitude to all the past and current lab members.

I would like to dedicate my Ph.D. dissertation to my wife Minyoung Jo and my son Hadden Jo who have supported and encouraged me.

Chapter 1: Introduction

Coding of prediction errors by midbrain dopamine cells

Midbrain dopamine (DA) cells play a central role in reinforcement learning (Schultz, 2002; Wise, 2004; Fields et al., 2007). A large body of experimental and computational studies indicates that DA phasic responses in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) mimic reward prediction errors of the temporal difference (TD) algorithm for machine learning (Sutton and Barto, 1990; Montague et al., 1996; Schultz et al., 1997; Pan et al., 2005; Day et al., 2007; Clark et al., 2010; Flagel et al., 2011; Cohen et al., 2012). Specifically, DA cells initially exhibit brief bursts of activity in response to primary rewards, but not predictive cues, early in Pavlovian conditioning. Such phasic responses are thought to act as a teaching signal to learn the predictive or incentive value of the cues. Subsequent associative learning produce a response shift from the reward delivery to the presentation of the predictive cues, and DA cells are no longer responsive to fully expected rewards. The magnitude of phasic responses to the reward-predicting cues has been shown to reflect the relative value of expected outcomes (Fiorillo et al., 2003; Tobler et al., 2005). Furthermore, DA cells also encode negative prediction errors by suppressing firing when expected rewards are omitted (Schultz et al., 1997; Roesch et al., 2007). However, it remains to be determined how DA cells compute prediction errors.

Remarkable evidence for the use of DA phasic activity as prediction errors to drive reinforcement learning came from a recent optogenetic study using a blocking paradigm (Steinberg et al., 2013). Blocking is a well-known learning phenomenon where the association between a conditioning cue and an outcome is prevented or blocked if the outcome is already

predicted by another cue (Kamin, 1969; Rescorla and Wagner, 1972). No learning of the blocked cue is thought to occur due to the absence of prediction errors after obtaining the fully expected outcome (Waelti et al., 2001). To determine whether artificial prediction errors were able to increase the associative strength of the block cue, Steinberg et al (2013) optogenetically stimulated DA cells briefly at the time of rewards after the reliably predictive cue and the blocked cue were simultaneously presented. In a subsequent probe test, the blocked cue successfully elicited conditioned responses. This unblocking result clearly indicates that the associative learning circuits in the brain use the phasic activation of DA cells as prediction errors.

In the TD model, a prediction error δ at the current time t is calculated as $\delta_t = r_t + V(S_t) - V(S_{t-1})$, where r_t is the available, if any, reward, $V(S_t)$ is the value of total future rewards expected in the current situation or state S_t of the task (current state value), and $V(S_{t-1})$ is the value of all future rewards estimated in the previous task state (previous state value). After the TD framework and existing evidence were combined, a number of theoretical models have postulated that the pedunculopontine tegmental nucleus (PPTg), the striatum, and the prefrontal cortex (PFC) are essential for conveying the three ingredients to compute prediction errors by DA cells (Doya, 2002; Joel et al., 2002; Worgotter and Porr, 2005; Kawato and Samejima, 2007; Morita et al., 2012). The PPTg is suggested to send the current reward information, based on the electrophysiological findings that PPTg neurons signal the value of obtained primary rewards in rats and primates (Okada et al., 2009; Norton et al., 2011). Given cholinergic and glutamatergic projections to the VTA and the SNc (Woolf, 1991; Futami et al., 1995; Sesack et al., 2003), the reward signals in the PPTg can induce phasic reward activity of DA cells (Floresco et al., 2003; Pan and Hyland, 2005; Zweifel et al., 2009). As one of the major target structures of DA cells,

the striatum, especially the ventral striatum (VS), is known to store the value of expected outcomes by accumulating DAergic prediction errors and in turn to provide the value signals to DA cells for subsequent computations of prediction errors (Apicella et al., 1992; Suri and Schultz, 2001; O'Doherty et al., 2004; Roesch et al., 2009; Cai et al., 2011; Day et al., 2011). It has been posited that the current state value is sent to midbrain DA systems through striatonigral direct pathway neurons that eventually excite DA cells, whereas the previous state value is fed via striatopallidal indirect pathway neurons that inhibit DA cells (Doya, 2002; Morita et al., 2012). However, it is unclear how the PFC contributes to DAergic prediction errors. The theoretical models have commonly hypothesized that the PFC plays a critical role in conveying information about the current and the previous state to the striatum, so that expected values in the corresponding states can be generated in the striatum (Doya, 2002; Morita et al., 2012). Although two recent studies claimed state representations in the PFC, especially the orbitofrontal cortex (OFC) (Takahashi et al., 2011; Wilson et al., 2014), the majority of previous studies suggested that the OFC encodes the subjective value of expected outcomes (Tremblay and Schultz, 1999; O'Doherty et al., 2001; Gottfried et al., 2003; Izquierdo et al., 2004; Padoa-Schioppa and Assad, 2006). The following section primarily considers what information the PFC conveys to DA cells.

Value signals in the PFC

Reinforcement learning models have considered two different strategies to update value signals: model-free and model-based strategies (Daw et al., 2005; Niv et al., 2006). The former learns reward values according to prediction errors generated after each action, so that value signals can be updated by trial and error without building an explicit knowledge or model of the current behavioral task. Reward prediction errors play a primary role in changing the value signals in the

model-free strategy. Empirical evidence has suggested that the striatum is central for encoding model-free value attached to or cached in reward-predicting cues or task states (Glascher et al., 2010). By contrast, the model-based strategy uses prior knowledge of the task to construct a cognitive model that describes transitions between task states. The model can be accessed to flexibly adjust value signals whenever internal and external factors (e.g, motivation, outcome, task rule, and environment) change even without experiencing actual outcomes. The literature indicates a prefrontal involvement in model-based reinforcement learning (Glascher et al., 2010; McDannald et al., 2011; Smittenaar et al., 2013). Neural activity related to model-based value signals has been found in the OFC (Takahashi et al., 2013; Stalnaker et al., 2014). Specifically, when two sensory cues that were initially associated with the same amount of rewards in separate trials were later presented together as a compound cue in an overexpectation task (Takahashi et al., 2013), individual neurons in the OFC signaled a new, higher outcome value inferred from the compound cue even before obtaining the actual reward. Similar findings were also observed in a reinforcer devaluation task in which a particular type of reward was devalued by overfeeding or pairing with illness. Normal rats and monkeys reduced responding to the cue predictive of the devalued reward, whereas animals with damage to the OFC failed to exhibit such a devaluation effect due to the inability to immediately update associative information between the cue and the devalued outcome (Gallagher et al., 1999; Izquierdo et al., 2004).

The OFC has been shown to carry not only value signals, but also non-value-related information. For example, OFC neurons represent complex variables associated with outcomes such as specific sensory features of the outcomes and their predictive cues (Rolls and Baylis, 1994; Padoa-Schioppa and Assad, 2006), temporal delay to outcome delivery (Roesch et al., 2006), and probability of obtaining the outcome (Kennerley and Wallis, 2009). In addition,

rodent OFC neurons encode spatial location as well as response direction leading to outcomes (Feierstein et al., 2006; Roesch et al., 2006; Furuyashiki et al., 2008). Based on these multiple neuronal representations, Schoenbaum and his colleagues hypothesize that the OFC plays an essential role in determining the current state from other similar task states (Schoenbaum et al., 2011; Takahashi et al., 2011; Wilson et al., 2014). Taken together, a unified view of the two OFC functions raises the possibility that the OFC may signal model-based value signals that are estimated in different task states.

Prefrontal modulation of DA activity

PFC influences on burst firing of DA cells in the VTA and the SNc have been extensively studied, more because the PFC has been implicated in the pathophysiology of DA-related disorders rather than in coding of value and state information. In particular, a hypoactive PFC function is associated with elevated striatal DA transmission in schizophrenia (Weinberger, 1987; Davis et al., 1991; Grace, 2000; Meyer-Lindenberg et al., 2002). To understand the casual relationship between the two abnormalities in PFC and DA systems, previous studies electrically or chemically stimulated the rodent medial PFC (mPFC) while monitoring changes in DA activity (Gariano and Groves, 1988; Overton et al., 1996; Tong et al., 1996; Lodge, 2011). It was commonly reported that mPFC stimulation resulted in both brief excitatory and longer-lasting inhibitory effects on different DA cells and the inhibited DA cells were proportionally greater than the phasically excited ones. These findings were supported by the anatomical literature showing that the mPFC sends excitatory glutamatergic projections directly the VTA and indirectly via GABAergic interneurons in the VTA or other structures such as the ventral striatum (Carr and Sesack, 2000; French and Totterdell, 2002; Geisler and Wise, 2008;

Omelchenko and Sesack, 2009; Watabe-Uchida et al., 2012). In line with this view, mPFC stimulation at physiologically relevant frequencies (around 10 Hz) decreases extracellular DA transmission measured in the ventral striatum (Jackson et al., 2001). These results indicate that the mPFC exerts much stronger inhibitory control on DA cells than excitatory control.

As another prefrontal source of glutamatergic projections to the VTA (Geisler and Wise, 2008; Vazquez-Borsetti et al., 2009; Watabe-Uchida et al., 2012), the OFC has been shown to bidirectionally modulate DA activity similar to the mPFC (Lodge, 2011; Takahashi et al., 2011). OFC stimulation led to either phasic elevation or prolonged suppression in firing of dopamine cells, and the activated DA cells outnumbered the inhibited ones. These electrophysiological findings contributed to understanding basic functional connections from the PFC to midbrain DA systems.

However, it remains elusive what information the PFC provides to DA cells for signaling prediction errors. In an attempt to address this issue, the present dissertation examined whether temporary inactivation of either OFC or mPFC induced significant alternations in firing of DA and non-DA cells in the VTA as rats were required to choose between two differently delayed rewards on an elevated T-maze. Based on the findings indicating that the OFC is the major prefrontal source of expected reward values to the VTA, OFC neurons were directly recorded in the delay discounting task in order to understand the exact nature of their representations. Finally, as a likely candidate for another source of value signals to the VTA, neuronal activity in the midbrain reticular formation (MRF) was also monitored in the same behavioral task.

Chapter 2: Prefrontal regulation of neuronal activity in the ventral tegmental area

Introduction

Midbrain dopamine (DA) cells encode differences between expected and received rewards and between previous and current reward expectations (Montague et al., 1996; Schultz et al., 1997; Bayer and Glimcher, 2005; Pan et al., 2005). These DAergic prediction error signals play an essential role in assigning values to reward-predicting stimuli in reinforcement learning (Fiorillo et al., 2003; Tobler et al., 2005; Flagel et al., 2011). Such acquired or cached values are thought to be stored in a common neural currency in the striatum which, in turn, provides the value information to dopamine neurons in preparation for subsequent computation of prediction errors (Apicella et al., 1992; Suri and Schultz, 2001; O'Doherty et al., 2004; Daw et al., 2005).

As a different valuation system, the prefrontal cortex, particularly the orbitofrontal cortex (OFC), is suggested to encode value signals that are derived or inferred from specific features of the outcome predicted in a given situation or state of the current learning task (Tremblay and Schultz, 1999; Roesch et al., 2006; Schoenbaum et al., 2011). The literature indicates that the OFC is able to flexibly estimate the value of a novel outcome even before experiencing the outcome (Izquierdo et al., 2004; Jones et al., 2012; Takahashi et al., 2013). Due to this feature (updating values independent of DAergic prediction errors), little attention has been paid to the possibility whether or not the OFC may directly contribute to DAergic error signals. Recently, Schoenbaum and his colleagues reported that unilateral damage to the OFC disrupted a sudden change in prediction error by ipsilateral DA cells when rewards were unexpectedly increased or decreased in value (Takahashi et al., 2011). Based on the results from computational modeling, they concluded that the OFC provides DA cells with state information, rather than expected

values per se. However, the partial damage left open the possibility that the intact contralateral OFC indirectly conveyed value signals to the VTA. In the current study, therefore, bilateral OFC function was temporarily inactivated by microinjection of muscimol (MUS, a GABA receptor agonist) while single unit activity was recorded from the VTA of rats performed a delay discounting task. In addition, the medial prefrontal cortex (mPFC), another prefrontal subregion that is well known to regulate DA activity (Gariano and Groves, 1988; Tong et al., 1996; Jo et al., 2013) and represent value signals, although weaker than the OFC (Sul et al., 2010), was also manipulated in the same animals. DAergic prediction errors as well as non-DA responses in the VTA were first determined in the delay-based decision making task, and then significant changes in their activity were compared before and after OFC or mPFC dysfunction.

Materials and methods

Subjects

Thirteen male Long-Evans rats (320-400 g; Simonson Labs, Gilroy, CA) were individually housed and initially allowed free access to food and water. Then food was restricted to maintain their body weights at 85% of free-feeding weights. All experiments were conducted during the light phase of a 12 hour light/dark cycle (lights on at 7:00 am), in accordance with the University of Washington's Institutional Animal Care and Use Committee guidelines.

Behavioral apparatus

An elevated T-maze, consisting of one start arm (the middle stem) and two goal arms (58 × 5.5 cm each), was located at the center of a circular curtained area. Each goal arm contained a metal food cup (0.7 cm in diameter × 0.6 cm deep) at the end, in front of which was a wooden barrier

(10 × 4 × 15 cm) to control access to reward. The arm was hinged such that its proximal end closest to the maze center could be lowered by remote control if needed.

Presurgical training

Each rat was placed on the maze and allowed to forage for chocolate milk drops scattered throughout the maze. Afterward, the rat was shaped to retrieve a reward delayed by 3 s (0.15 ml) only from two goal arms. Specifically, the animal was trained to run down the start arm and freely choose either goal arm. Upon arrival in front of the barrier, the rat had to wait for 3 s. As the barrier was removed by an experimenter who measured the elapsed time using a digital stopwatch, the rat was allowed to approach and consume the reward. After re-baiting the food cup and putting the barrier back in place, the experimenter gently guided the rat to the start arm for the next trial. Once the rat could finish 16 trials within 20 min, it underwent stereotaxic surgery.

Surgery

Tetrodes were made from 20μm lacquer-coated tungsten wires (California Fine Wire, Grover Beach, CA) and final impedance of each wire was adjusted to 0.2-0.4 MΩ (tested at 1 kHz). Six individually drivable tetrodes were chronically implanted in the right hemisphere dorsal to VTA (5.3 mm posterior to bregma, 0.8 mm lateral to midline, and 7.0 mm ventral to the brain surface). Six rats also received bilateral implantation of guide cannulae (25 gauge) aimed at the OFC (3.2 mm anterior, 3.1 mm lateral, and 4.6 mm ventral to bregma) and the mPFC (3.2 mm anterior, 0.7 mm lateral, and 3.3 mm ventral to bregma). A 33 gauge dummy cannula was inserted into each guide cannula to prevent blockade.

Delay discounting task

After a week of recovery, two groups of rats performed slightly different decision making tasks on the maze. The seven rats implanted with recording electrodes alone performed a delay-based decision making task in which they were required to choose between a sooner small (SS) reward and a later large (LL) reward. The delay to SS reward (0.05 ml) was held constant at 3 s throughout the experiments, whereas three different delays (10, 20, and 40 s) before LL reward (0.3 ml) were used to test possible changes in choice preference as a function of delay to LL reward. Thus, a daily testing session consisted of three blocks of trials to which the three delays were randomly assigned and only one delay was presented in a given block. To inform rats as to which delay was imposed before LL reward, each block began with 10 forced-choice trials followed by 6 or 8 free-choice trials. During the forced-choice trials, 5 SS and 5 LL reward trials were pseudorandomly ordered and only one goal arm was made available in each trial by lowering the other goal arm. Both goal arms were presented during the free-choice trials in which animals' choice performance was measured. As the rats selected and entered the goal arm associated with LL reward, an additional barrier was located at its entrance to confine them in the chosen arm during the longer delays. The three blocks were separated by an interblock interval of 5-10 min during which the animals were placed on a holding area adjacent to the maze. The location of SS and LL rewards in the goal arms remained constant within each rat but was counterbalanced across rats.

The second group of six rats implanted with bilateral cannulae as well as recording electrodes was trained in a modified version of the delay discounting task in which the delay to LL reward remained unvaried at 10 s across blocks. In a recording session, the location of LL and SS rewards was randomly selected and the rats were required to choose between a 3 s-delayed SS reward and a 10 s-delayed LL reward in the first block of trials. After either SAL or

MUS were injected into one of the PFC subregions, the second block of trials was tested. A total of 8 drug testing sessions were given per subregion of each rat.

Single-unit recording

Neural activity was monitored daily using a Cheetah data acquisition system (Neuralynx, Bozeman, MT). Unit signals were amplified, filtered at 600-6000 Hz, and digitized at 16 kHz. The animal's head position was also recorded at 30 Hz by tracking two light-emitting diodes mounted on the headstage. To increase the chance of recording from DA cells, single-units displaying low discharge frequencies and wide spike waveforms were prescreened and recorded subsequently. If such units were not detected, tetrodes were advanced in 40 μm increments, up to 160 μm per day. High-firing units were recorded only if a DA-like unit was present on the same tetrode. All tetrodes were advanced at the end of each recording session to find new cells. For recording sessions involving drug infusion, however, tetrodes were lowered in an attempt to hold the same units for two consecutive days. In this way, the same units could be tested with either SAL or MUS injections into one PFC subregion. During the recording session, three salient events such as delay onset (DO), delay termination (DT), and reward were inserted into the data stream online. Specifically, timestamps for DO and DT were marked when an experimenter pressed the stopwatch buttons. The time of reward encounters was automatically detected by 'lick-detectors' (custom designed by Neuralynx) when the animals first licked the chocolate milk in the food cups.

Intracranial microinjection

A 33 gauge injection cannula extending 1 mm below the tip of the guide cannula was connected to a 10 μl syringe (Hamilton, Reno, NV) via polyethylene tubing (PE 20). Either 0.3 μl of MUS (1 μg/μl) dissolved in saline (SAL) or its vehicle was bilaterally injected at 10 μl/h rate using a

microinfusion pump (KD Scientific, Holliston, MA) while rats were under light gas anesthesia by isoflurane. The injection cannula was left in place for an additional 1 min to ensure proper diffusion from its tip. Then they were returned to their home cage and the behavioral recording resumed in 30 min.

Histology

After completion of all experiments, small marker lesions were made by passing a 10 μA current for 10 s through two wires of each tetrode. All rats were transcardially perfused and their brains were stored in a 10% formalin/30% sucrose solution and cut in coronal sections (40 μm) on a freezing microtome. The sections were stained with cresyl violet and examined under light microscopy to reconstruct tetrode tracks through the VTA and cannula tip locations in the PFC. Data recorded only from the VTA were analyzed.

Unit classification

Single-units (> 2:1 signal-to-noise ratio) were isolated by clustering various spike waveform parameters using Offline sorter (Plexon, Dallas, TX). For some units that were recorded more than one day, the session in which the units were most clearly isolated from other units and background noise was used for analysis. DA cells in the VTA were identified by a cluster analysis that was performed on two distinct waveform features of DA cells, spike duration and amplitude ratio, as previously described (Roesch et al., 2007; Takahashi et al., 2009; Takahashi et al., 2011). To verify whether the classified DA cells were sensitive to D2 agonists such as quinpirole, a subset of these cells were additionally recorded and their spontaneous activity was compared before and after quinpirole injection (0.4 mg/kg, s.c.).

Data analysis

Spontaneous firing properties of DA cells were calculated from data collected while rats were placed in a holding area prior to the first block of trials and between blocks; these included mean firing rate and the percentage of spikes that occurred in bursts. A burst was defined as successive spiking with an interspike interval of <80 ms followed by an interspike interval of >160 ms (Grace and Bunney, 1984). To examine DAergic phasic firing, peri-event time histograms (PETHs; bin width, 50 ms) were constructed separately for the 5 s period around DT and reward events in each block of all trials. A DA cell was considered to show phasic responses to one of the two events or both if it passed the following two criteria: 1) its peak firing was observed within the 200 ms epoch around DT or within the 100 ms epoch after reward and 2) its average activity within the corresponding epochs was ≥ 200% of its mean firing rate over each block. To determine a linear relationship between DA activity and delay to LL reward, Spearman's rank correlation coefficient was calculated in each time bin of the PETHs. The significance of correlation was estimated using a permutation test in which firing rates of each bin were randomly shuffled across blocks for 1000 times. A 99% confidence interval was calculated from correlation coefficients of the shuffled data.

Using PETHs, a non-DA cell was also categorized as reward-related if it met the following three criteria: 1) its highest firing occurred in a 400 ms window after reward (-50 to 350 ms), 2) the average activity during the window was ≥ 200% of its mean firing for the block of all trials, and 3) its late delay activity (-1.5 s from DT) was significant higher than the early delay activity (1.5 s from DO) during at least one of four delays as tested by a Wilcoxon Signed Rank test. For non-DA cells displaying delay-dependent changes in firing, discharge rates during the entire delay of each trial were converted to z-scores relative to the mean firing of all trials

within the block. A non-DA cell was defined as delay-excited or delay-inhibited if the average z-score during at least one of the delays was greater or less than 2, respectively.

Statistical analysis

ANOVAs with repeated measures were mainly used to test statistical significance of neural activity. Significant differences in firing between two reward conditions in a given block were analyzed with *t*-tests. Pearson's correlation tests were performed to establish a relationship between two variables. Two-tailed p values < 0.05 were considered statistically significant. Data are expressed as mean ± SEM.

Results

Choice behavior

Seven rats were trained to choose between SS and LL rewards in a delay discounting task on an elevated T-maze (Fig. 2.1A). To investigate the animals' choice performance as a function of delay to LL reward, three different lengths of delay were randomly imposed prior to LL delay in separate blocks of trials (Fig. 2.1B). However, the delay to SS reward was kept constant. Each block began with 10 forced-choice trials and choice preference for LL reward was measured in the following 6 or 8 free-choice trials. A wooden barrier located in front of the food cup of each goal arm was used to prevent access to reward during the delays. Since an experimenter manually removed the barrier at the time of delay termination (DT) while monitoring the elapsed time using a digital stopwatch, there were slight variations in delay length. In a total of 42 behavioral recording sessions (3-12 sessions per rat), SS reward was delayed by 3.27 ± 0.26 s (mean ± SD) and LL reward was delayed by 10.19 ± 0.36, 20.25 ± 0.47, or 40.34 ± 0.58 s.

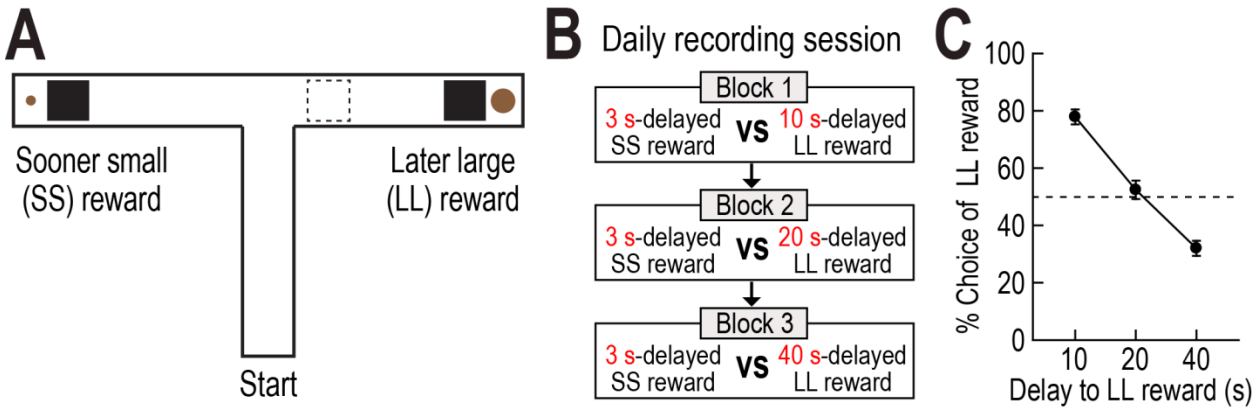

Figure 2.1. Choice performance on a delay-based decision making task. (**A**) Illustration of the T-maze. Rectangular wooden barriers (black squares) were placed before food cups on the goal arms to control animals' access to SS and LL rewards. When rats chose a goal arm associated with LL reward, an additional barrier (dashed rectangle) was placed at its entrance to prevent the animals from exiting the goal arm during the waiting period. (**B**) Daily behavioral recording procedures. Three different lengths of delay to LL reward were randomly ordered and tested in separate blocks of trials. The delay to SS reward remained unchanged. Each block consisted of forced-choice trials, followed by free-choice trials. (**C**) Choice preference for LL reward as a function of delay to LL reward. Error bars indicate SEM.

Rats' choice performance indicated the value of LL reward was discounted by the delay preceding it (Fig. 2.1C). Specifically, they exhibited a strong preference to a 10 s-delayed LL reward. However, the animals were indifferent between a SS reward and a 20 s-delayed LL reward. Further extension of the delay to LL reward to 40 s reversed choice preference such that they chose SS reward more frequently. A Pearson's correlation test demonstrated a significant negative relationship between choice performance and delay length prior to LL reward (r = -0.94, $p < 0.001$).

Responses of DA cells during the delay discounting task

Single-unit activity was recorded from 364 neurons in the VTA of rats performing the task (Fig. 2.2A, B). Of these neurons, 116 cells were classified as DAergic by clustering a group of cells showing a wide spike duration and low amplitude ratio (Fig. 2.2C). Previous studies reported that DA cells identified by the cluster analysis encode reward prediction errors (Roesch et al., 2007; Takahashi et al., 2009; Takahashi et al., 2011). After daily recording sessions, a subset of 67 neurons was further tested to verify whether DA cells were inhibited by the D2 receptor agonist quinpirole (Fig. 2.2D). All of the cells identified as DAergic (n = 18) were suppressed (ranged from -97 to -15%) after injecting quinpirole, whereas non-DA cells showed heterogeneous responses to quinpirole. On average, t-tests revealed that quinpirole injections significantly inhibited spontaneous activity of DA cells ($t_{(17)} = 10.43$, $p < 0.001$), but significantly enhanced firing rates of non-DA cells ($t_{(48)} = 3.61$, $p = 0.001$).

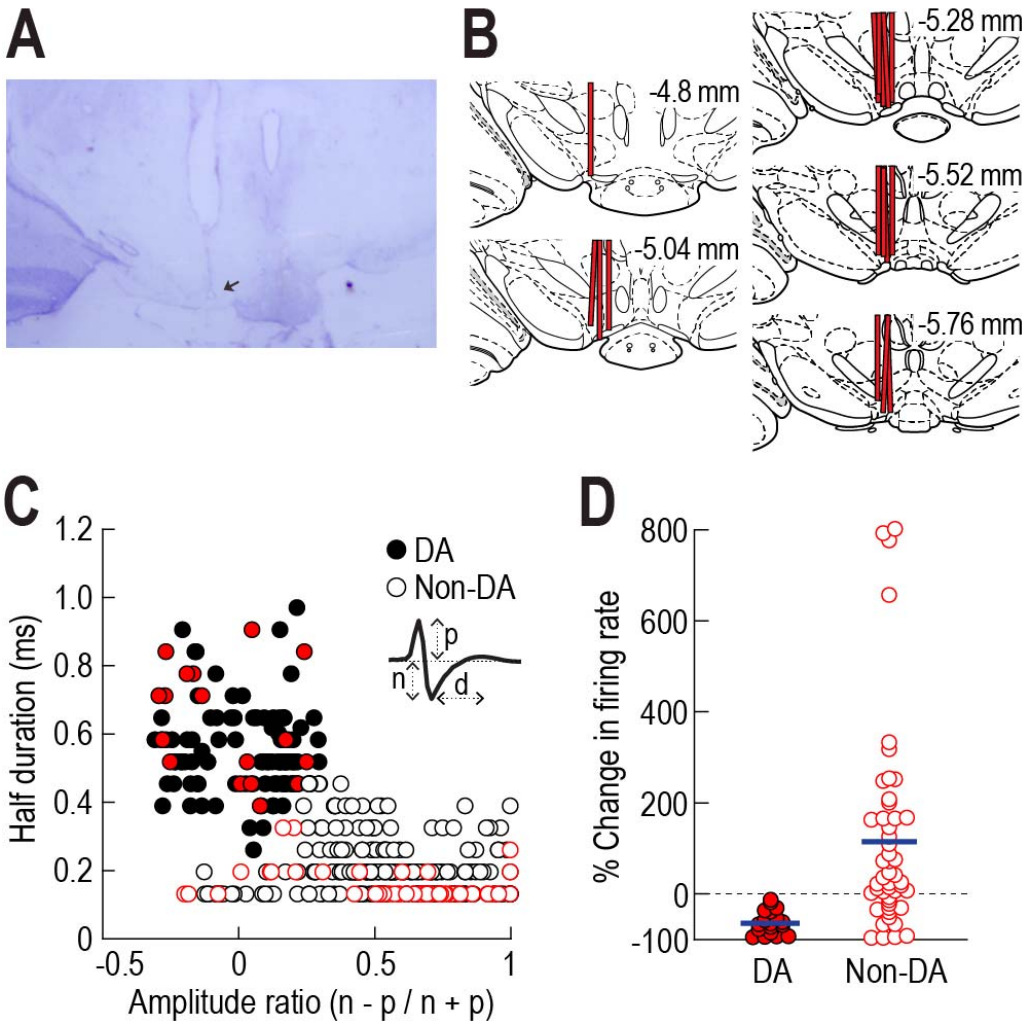

Figure 2.2. Histological verification of tetrode locations and classification of DA cells in the VTA. (**A**) Nissl-stained section showing the final location of a tetrode tip in the VTA. (**B**) Reconstruction of all tetrode tracks. (**C**) Cluster analysis for all VTA neurons. DA cells in black were identified using two waveform features: half spike duration (d) and the amplitude ratio of the first positive peak (p) and negative valley (n). (**D**) Changes in spontaneous firing after quinpirole injections. VTA neurons in red (C) were tested with the D2 receptor agonist after daily behavioral recording. Blue lines indicate the average changes for DA and non-DA cells.

DA cells that were initially activated by the receipt of primary reward have been shown to respond to both a reward-predicting cue and actual reward delivery in the middle of Pavlovian conditioning, and exhibit phasic responses only to the predictive cue after learning (Pan et al., 2005; Clark et al., 2010; Cohen et al., 2012). Consistent with the findings, DA cells in the current task showed a response shift in time from reward encounters to DT when the barriers on the goal arms were removed. For instance, a representative DA cell in the left column of Figure 2.3A responded to both DT and reward. In the following day, the same cells showed phasic responses only to DT but no longer to reward (Fig. 2.3A, right column), which indicated that the removal of barriers at the end of delays fully predicted the upcoming rewards. Overall, 25 DA cells (21.6%) responded to both reward and DT (Fig. 2.3B) and the fraction of these cells decreased as recording sessions progressed ($r = -0.79$, $p = 0.001$; Fig. 2.3C). A larger number of DA cells (81/116, 69.8%) were phasically excited only at the time of DT and their fraction increased across recording sessions ($r = 0.58$, $p < 0.05$). The remaining DA cells (10/116, 8.6%) did not display task-related activity. A DA cell responding only to reward was not found in the study, presumably due to the extensive pre-training before the recording sessions.

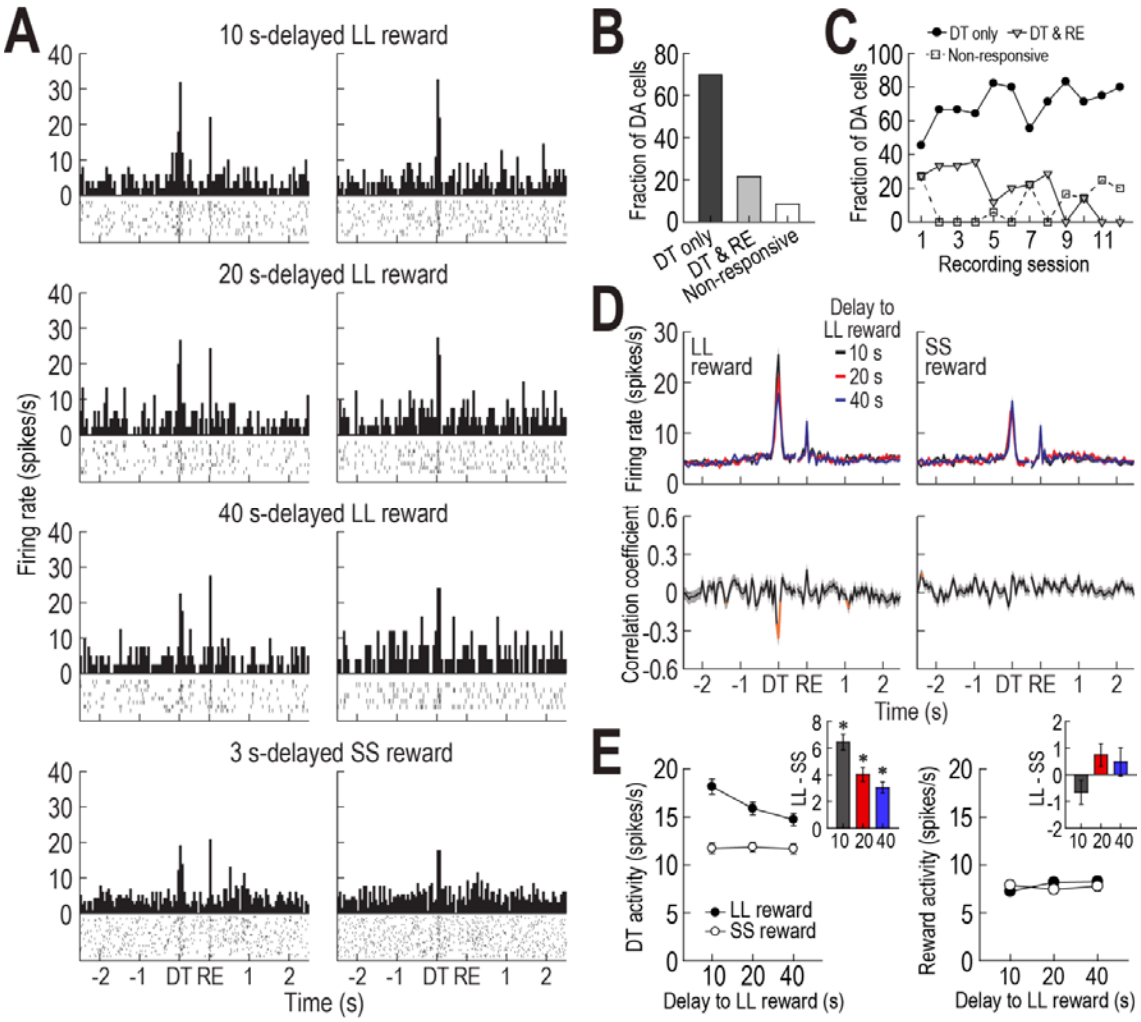

Figure 2.3. Phasic DA activity in the delay discounting task. (**A**) A representative DA cell recorded for two consecutive days. The DA cell initially responded to both delay termination (DT) and reward (RE) in the left column (bin width, 50 ms), but the same cell no longer exhibited phasic responses to reward on the next day (right column). Histograms are aligned on DT and RE. (**B, C**) Fractions of DA neurons showing phasic activity at different times. Most DA cells were excited only at the time of DT. Non-responsive DA cells were not included in the subsequent analyses. (**D**) Population activity of all task-related DA cells. Correlation coefficients for individual neuronal responses across three blocks of trials were calculated per each time bin. Orange data points that fell outside the 99% confidence interval obtained from a permutation test for at least two consecutive bins were considered significantly correlated. (**E**) Average DT and reward responses. Inset bar graphs show differential firing between LL and SS reward conditions within blocks of trials (*$p < 0.001$; *t*-test). Shaded areas and error bars indicate SEM.

The population histograms of all task-relevant DA cells indicated that DT, but not reward, responses were influenced by delay length (Fig. 2.3D). Specifically, when Spearman's rank correlation coefficients for individual neuronal responses across three blocks of trials were calculated per time bin of 50 ms, significant negative correlations were found around the time of DT (-50 to 100 ms). No significant relationships were found around the time of DT before SS reward or around the time of reward. These results were also supported by the magnitude of population responses (Fig. 2.3E). A repeated measures ANOVA performed on average reward activity found no significant effects of block ($F_{(2,210)} = 1.88$, $p = 0.16$) or reward size ($F_{(1,105)} = 0.29$, $p = 0.59$), but an interaction between the variables was significant ($F_{(2,210)} = 3.84$, $p = 0.02$). Another ANOVA comparing average DT activity demonstrated significant effects of block ($F_{(2,210)} = 15.52$, $p < 0.001$; Fig. 2.3E) and reward size ($F_{(1,105)} = 113.67$, $p < 0.001$) and a significant interaction ($F_{(2,210)} = 24.98$, $p < 0.001$). Interestingly, *post hoc* comparisons showed that DT responses prior to LL reward were consistently greater than those before SS reward across blocks (p values < 0.001), even during the block where rats preferred a 3 s-delayed SS reward over a 40 s-delayed LL reward. These results indicated that although DT activity was discounted by imposed delays, this activity did not reflect animals' decision value in the task.

Responses of non-DA cells during the delay discounting task

Recently GABAergic neurons in the VTA have been suggested to encode the value of expected reward (Cohen et al., 2012). These neurons ramped up their activity when a future reward was predicted, and then reached the peak firing rates at the time of reward delivery. In the current study, 15 reward-related non-DA cells showed similar firing patterns (Fig. 2.4A) even though they were briefly inhibited at the time of DT when DA cells were phasically excited. In an early

phase of four different delays (the 1.5 s epoch after DO), reward-related non-DA cells displayed no differential responses across different delays or between SS and LL rewards (Fig. 2.4B) as indicated by no significant effects of block (ANOVA with repeated measures, $F_{(2,28)} = 0.02$, $p = 0.98$) or reward size ($F_{(1,14)} = 0.67$, $p = 0.43$) and no significant interaction between the variables ($F_{(2,28)} = 2.85$, $p = 0.07$). In a late phase of the delays (the 1.5 s epoch before DT), however, their ramping activity prior to LL reward became higher than that before SS reward in the late phase of the delays (Fig. 2.4C). The final responses were also more highly elevated with shorter delays to LL reward as significant negative correlations (-800 to -100 ms before DT) were observed in the individual neurons (Fig. 2.4A). A repeated measures ANOVA confirmed a significant effect of reward size ($F_{(1,14)} = 38.8$, $p < 0.001$). Although no effect of block ($F_{(2,28)} = 1.27$, $p = 0.29$) was found, there was a significant interaction between the factors ($F_{(2,28)} = 5.06$, $p = 0.01$). *Post hoc* comparisons demonstrated a significant difference between 10 s- and 40 s-delayed LL rewards ($p < 0.05$), but no differences in the other pairwise comparisons (p values > 0.1). When the rats subsequently obtained reward, the non-DA cells showed robust activation that lasted longer for LL reward than for SS reward. A repeated measures ANOVA revealed a significant effect of reward size ($F_{(1,14)} = 9.09$, $p = 0.009$; Fig. 2.4D), whereas neither a significant effect of block ($F_{(2,28)} = 0.59$, $p = 0.56$) nor an interaction between the variables ($F_{(2,28)} = 0.87$, $p = 0.43$) was found. These firing patterns suggest that reward-related non-DA cells represent expected reward values in the task.

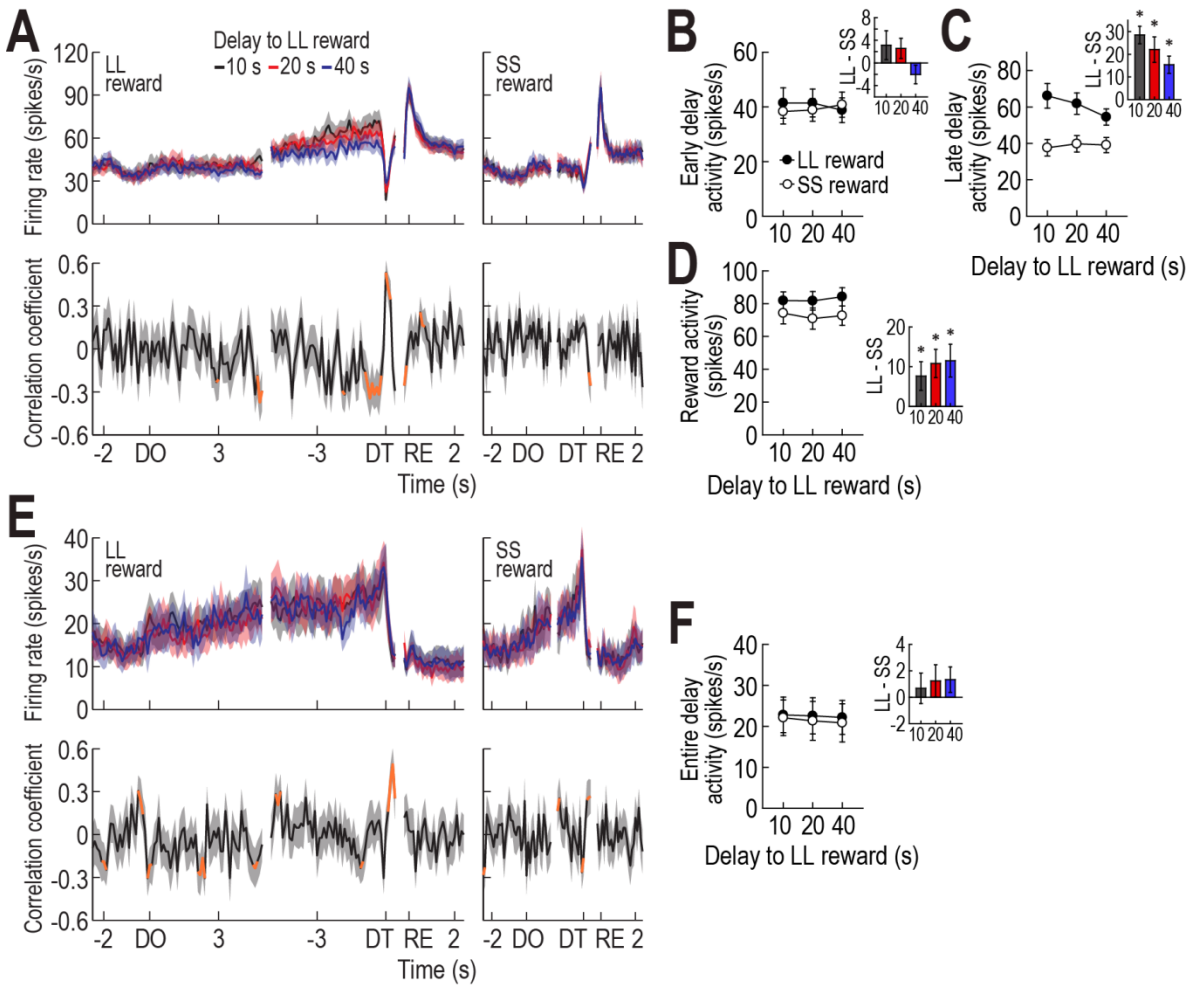

Figure 2.4. Activity of non-DA cells in the delay discounting task. (**A**) Population histograms (bin width, 100 ms) of reward-responsive non-DA cells and correlation coefficients over the course of delays. Data are aligned on delay onset (DO) and termination (DT). Significant correlation coefficients for more than two consecutive bins were depicted in orange. (**B, C**) Average responses during an early phase (1.5 s after DO, B) and a late phase (1.5 s before DT, C) of delays. Inset bar graphs indicate differential firing between LL and SS reward conditions within blocks of trials (*$p < 0.01$; *t*-test). (**D**) Average reward activity. (**E**) Population responses of delay-excited non-DA cells and correlation coefficients around delays. (**E**) Average activity during the entire delays. All graphs show mean ± SEM.

In addition, 17 delay-excited non-DA cells were observed in the VTA. They developed ramping activity throughout delays, but this excitation rapidly declined around the time of DT (Fig. 2.4E). Delay-excited non-DA cells showed no differential firing across delays or between two reward sizes (Fig. 2.4F), as a repeated measures ANOVA found no significant effects of block ($F_{(1,16)} = 0.29$, $p = 0.59$) or reward size ($F_{(2,32)} = 0.82$, $p = 0.45$) with no significant interaction between the factors ($F_{(2,32)} = 0.02$, $p = 0.98$). Thus, it appears that the delay-excited activity reflects a gradual increase in general motivation in expectation of upcoming outcomes.

PFC contributions to choice performance

To test whether the PFC modulated choice behavior and VTA neural activity in the task, six rats were implanted with bilateral cannulae targeted at two PFC subregions (i.e., the OFC and the mPFC) as well as recording electrodes in the VTA (Fig. 2.6A-D). The animals were trained in a modified version of the delay discounting task in which they had to choose between a 3 s-delayed SS reward and a 10 s-delayed LL reward (Fig. 2.5A). In each testing session, baseline behavioral performance and neural activity were measured in the first blocks of trials. After either SAL or MUS was injected into one subregion, the second block was given to compare changes from the baseline. The location of two rewards was pseudorandomly switched between sessions.

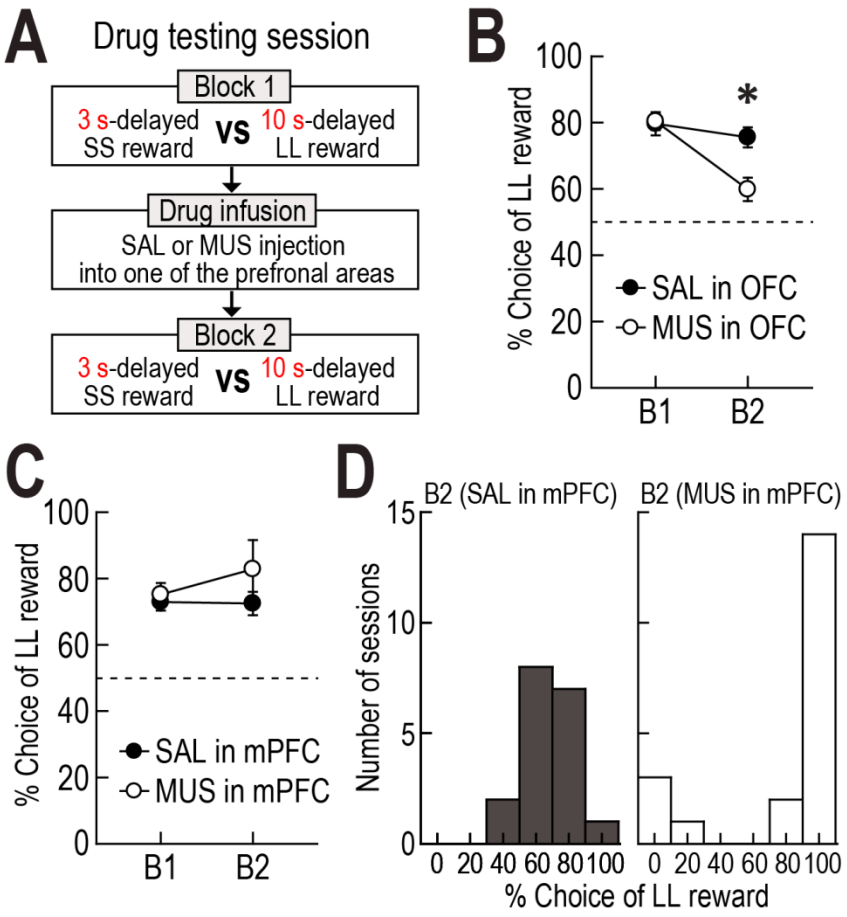

Figure 2.5. Effects of OFC and mPFC inactivation on choice performance. (**A**) Drug testing procedures. The goal arms associated 3 s-delayed SS and 10 s-delayed LL rewards were randomly selected on each testing day. Either SAL or MUS was infused into one of the prefrontal subregions between two blocks of trials. (**B, C**) Changes in choice preference for LL reward after OFC (B) and mPFC (C) inactivation (*$p < 0.01$). (**D**) Distributions of behavioral performance tested after injecting drugs into the mPFC. Error bars represent SEM.

These rats normally showed a strong preference for LL reward over SS reward (Fig. 2.5B). When drugs were injected into the OFC during 42 behavioral recording sessions (20 SAL and 22 MUS sessions; 2-5 SAL and 2-5 MUS sessions per rat), a repeated measures ANOVA found significant effects of drug ($F_{(1,5)} = 10.9$, $p = 0.02$) and block ($F_{(1,5)} = 12.29$, $p = 0.01$) and a significant interaction between drug and block ($F_{(1,5)} = 51.86$, $p < 0.001$). *Post hoc* comparisons demonstrated that MUS injections significantly decreased the preference for LL reward relative both to both baseline trials within the first block and to the performance after SAL injections (p values < 0.01), whereas behavioral performance was comparable between blocks in SAL sessions ($p = 0.79$). This result was consistent with previous evidence indicating OFC dysfunction induces impulsive choice on a T-maze (Rudebeck et al., 2006).

By contrast, drug injections into the mPFC during 38 sessions (18 SAL and 20 MUS sessions; 1-4 SAL and 2-4 MUS sessions per rat) did not cause significant changes in choice performance (Fig. 2.5C). This observation was confirmed by no effects of drug ($F_{(1,5)} = 2.47$, $p = 0.18$) or block ($F_{(1,5)} = 0.44$, $p = 0.54$) and no significant interaction between the variables ($F_{(1,5)} = 0.75$, $p = 0.43$). The relative larger error bar for the MUS-injected block prompted a closer investigation which found a response bias to one goal arm (Fig. 2.5D). In 4 testing sessions, for example, mPFC-inactivated rats persistently chose the location associated with SS reward even though they sampled LL reward in the preceding forced-choice trials. The abnormal choice behavior may result from a deficit in responding according to the previously acquired knowledge or rules of the task after mPFC inactivation (Stefani and Moghaddam, 2005; Jung et al., 2008).

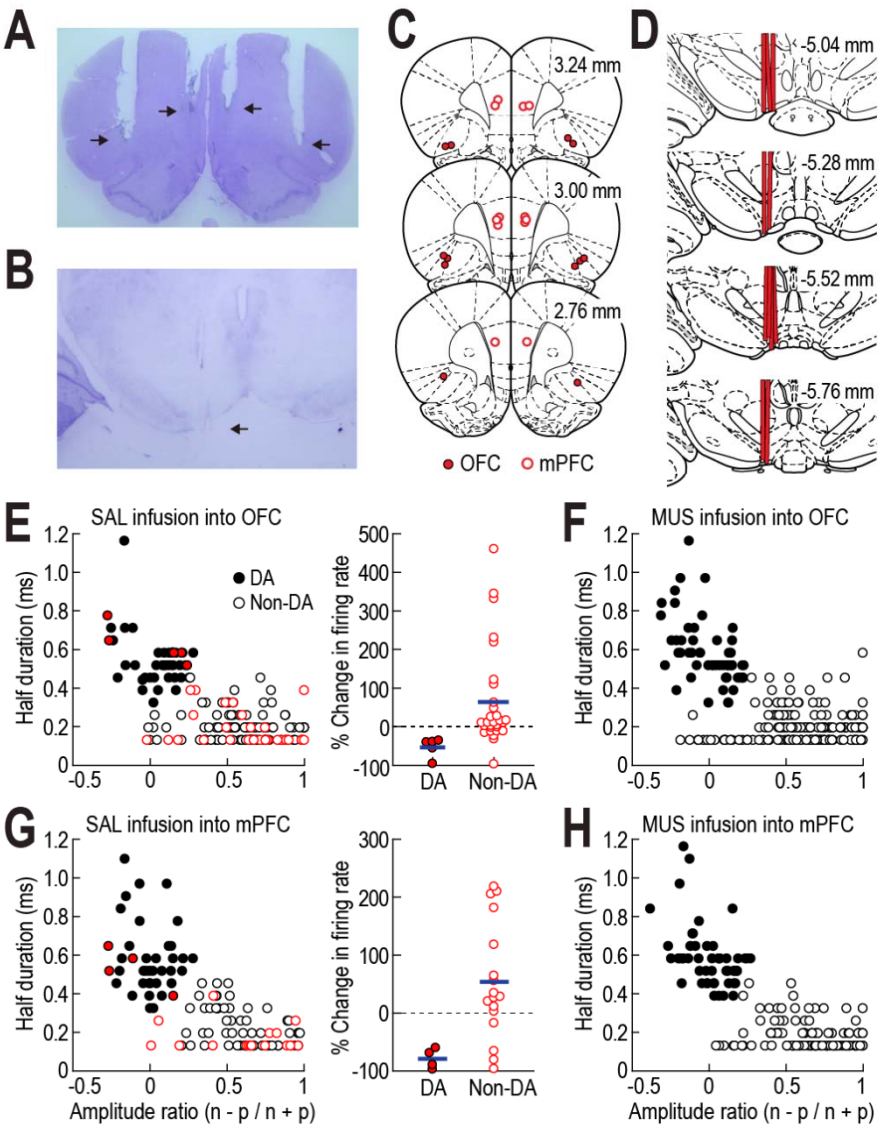

Figure 2.6. Injection sites, tetrode locations, and identification of DA cells. (**A, B**) Nissl-stained photomicrographs for cannula placement in the OFC and the mPFC (A) and for a tetrode tip (B). (**C**) Illustration of all microinjection sites. (**D**) Reconstruction of all tetrode tracks. (**E-H**) Cluster analyses for VTA neurons recorded with prefrontal manipulations. Some neurons in red (E, G) were further tested with quinpirole injections and changes in spontaneous firing after quinpirole are depicted on the right. Blue lines indicate the average changes for two groups of cells in the VTA.

Prefrontal regulation of DA activity

A total of 176, 221, 123, and 155 cells were recorded with SAL and MUS injections into the OFC and the mPFC, respectively. Of these VTA cells, 45, 54, 39, and 52neurons were identified as DAergic using the cluster analysis (Fig. 2.6E-H) in SAL/OFC, MUS/OFC, SAL/mPFC, and MUS/mPFC sessions, respectively. After SAL sessions, a subgroup of neurons was additionally recorded to compare their discharge rates before and after quinpirole infusion. *T*-tests showed that quinpirole significantly inhibited DA cells recorded in both SAL/OFC ($t_{(4)} = 4.79$, $p = 0.009$) and SAL/mPFC sessions ($t_{(3)} = 9.15$, $p = 0.003$), whereas the D2 receptor agonist significantly excited non-DA cells recorded in the two SAL-injected conditions (OFC, $t_{(28)} = 2.65$, $p = 0.01$; mPFC, $t_{(16)} = 2.2$, $p = 0.04$).

To first examine the effects of PFC inactivation on spontaneous activity of DA cells, mean firing rates and percentages of spikes in bursts were measured while rats were not engaged in the task (Table 2.1). Paired *t*-test found that the two measurements were significantly reduced after MUS injections into the OFC ($t_{(53)}$ values > 6.84, p values < 0.001) and the mPFC ($t_{(51)}$ values > 3.89, p values < 0.001), but no alternations in spontaneous activity were found in SAL sessions irrespective of the PFC subregions (t values < 1.66, p values > 0.1). It is noteworthy that compared to mPFC inactivation, OFC inactivation more severely decreased the percentage of spikes in burst ($t_{(104)} = 2.15$, $p = 0.04$), which is in line with anatomical evidence showing that the OFC sends stronger projections to VTA DA cells than the mPFC (Watabe-Uchida et al., 2012).

Table 2.1. Effects of prefrontal inactivation on spontaneous activity of DA cells

Subregion	Drug injected before B2	Mean firing rate		% Spikes in bursts	
		B1	B2	B1	B2
OFC	SAL	4.4 ± 0.3	4.6 ± 0.4	40.6 ± 3.6	40.0 ± 3.7
	MUS	4.4 ± 0.2	3.5 ± 0.2**	42.7 ±2.9	18.8 ± 2.5**
mPFC	SAL	3.9 ± 0.3	4.1 ± 0.3	40.1 ± 3.9	40.3 ± 4.2
	MUS	4.4 ± 0.3	3.9 ± 0.3*	40.1 ± 3.2	27.8 ± 3.5**

Asterisks indicate a significant difference between before (B1) and after drug infusion (B2) on a paired *t*-test (*$p < 0.01$, ** $p < 0.001$; mean ± SEM).

Next, 38 and 51 DA cells that exhibited task-related responses to DT or reward during SAL/OFC and MUS/OFC sessions, respectively, were analyzed for the effects of OFC manipulations on their positive prediction errors in the task. As seen from representative examples and population responses (Fig. 2.7A-C), MUS, but not SAL, injections into the OFC caused a marked reduction in DAergic phasic activity at the time of DT but, conversely, an elevation in response to reward as if the removal of barriers at the time of DT failed to properly predict an upcoming reward and therefore the reward was unexpectedly encountered. Repeated measures ANOVAs separately performed on DT and reward activity during MUS/OFC sessions demonstrated significant effects of block ($F_{(1,50)}$ values > 44.77, p values < 0.001; Fig. 2.7E) and reward size ($F_{(1,50)}$ values > 12.72, p values < 0.001). A significant interaction between the factors was found for reward activity ($F_{(1,50)} = 24.76$, $p < 0.001$), but not for DT activity ($F_{(1,50)} = 0.7$, $p = 0.41$). During SAL/OFC sessions, both DT and reward responses significantly differed between two reward sizes ($F_{(1,37)}$ values > 8.06, p values < 0.01; Fig. 2.7D), but no effect of block ($F_{(1,37)}$ values < 0.78, p values > 0.38) and no interactions between the variables ($F_{(1,37)}$ values < 0.78, p values > 0.38) were found. To further determine a correlation between changes in DT and reward responses in MUS/OFC sessions, differential firing of the two responses across blocks was compared within individual neurons (Fig. 2.7F). Significant inverse correlations for both LL and SS reward conditions (Pearson's correlation, r values < -0.36, p values < 0.01) indicated that the less DA cells responded at the time of DT, the stronger they were excited by reward. These alternations after OFC dysfunction suggest that the OFC may convey the information about expected reward to DA cells. However, it seems unlikely that the OFC is the only source for outcome expectancies, because although reduced after OFC inactivation, the DT activity was still stronger in anticipation of LL over SS rewards (Fig. 2.7E).

Other value signals may be fed to DA cells by different brain structures such as the ventral striatum (Roesch et al., 2009; Day et al., 2011; Clark et al., 2012).

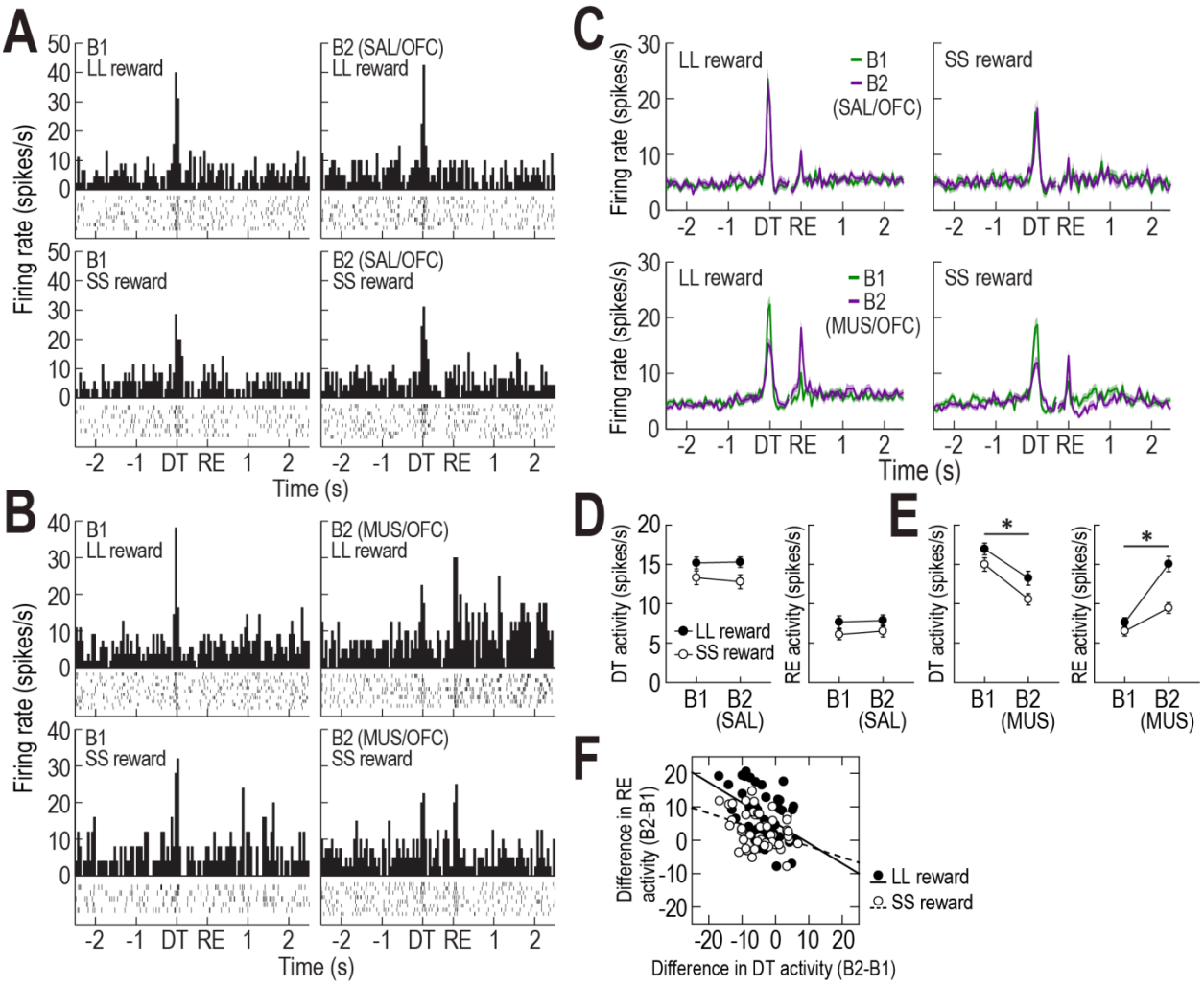

Figure 2.7. Effects of OFC inactivation on DA cells. (**A, B**) Examples of representative DA cells recorded with SAL (A) and MUS (B) injections. Histograms (bin width, 50 ms) are aligned on delay termination (DT) and reward (RE). (**C**) Population responses of all task-related DA cells. (**D, E**) Average DT and reward responses in SAL/OFC (D) and MUS/OFC (E) sessions (*$p <$ 0.001). (**F**) Correlations between altered DT and reward responses across blocks in MUS/OFC sessions. Shaded areas and error bars indicate SEM.

When mPFC function was manipulated with SAL and MUS, 32 and 46 DA cells, respectively, showed task-related responses. MUS infusions into the mPFC increased DAergic phasic responses to both DT and reward, whereas SAL infusions did not alter the two responses (Fig. 2.8A-C). Repeated measures ANOVAs revealed that both DT and RE responses in MUS/mPFC sessions significantly differed before and after mPFC inactivation ($F_{(1,44)}$ values > 11.77, p values < 0.01; Fig. 2.8E) and between SS and LL rewards ($F_{(1,44)}$ values > 11.4, p values < 0.01), but interactions between the variables were not significant ($F_{(1,44)}$ values < 0.58, p values > 0.45). During SAL/mPFC sessions, the two responses differed between two reward sizes ($F_{(1,31)}$ values > 7.2, p values < 0.05; Fig. 2.8D) with neither effects of block ($F_{(1,31)}$ values < 2.67, p values < 0.11) nor interactions ($F_{(1,31)}$ values < 0.3, p values < 0.58). As a representative DA cell in the Figure 2.8B displayed elevated activity at the time of both DT and reward after mPFC inactivation, such overall increases in the two responses was evident in the population of DA neurons in LL reward trials (Pearson's correlation, $r = 0.34$, $p = 0.02$, Fig. 2.8F), but not in SS reward trials ($r = 0.11$, $p = 0.48$). Nevertheless, these results suggest that mPFC inactivation disinhibits DAergic prediction error signaling.

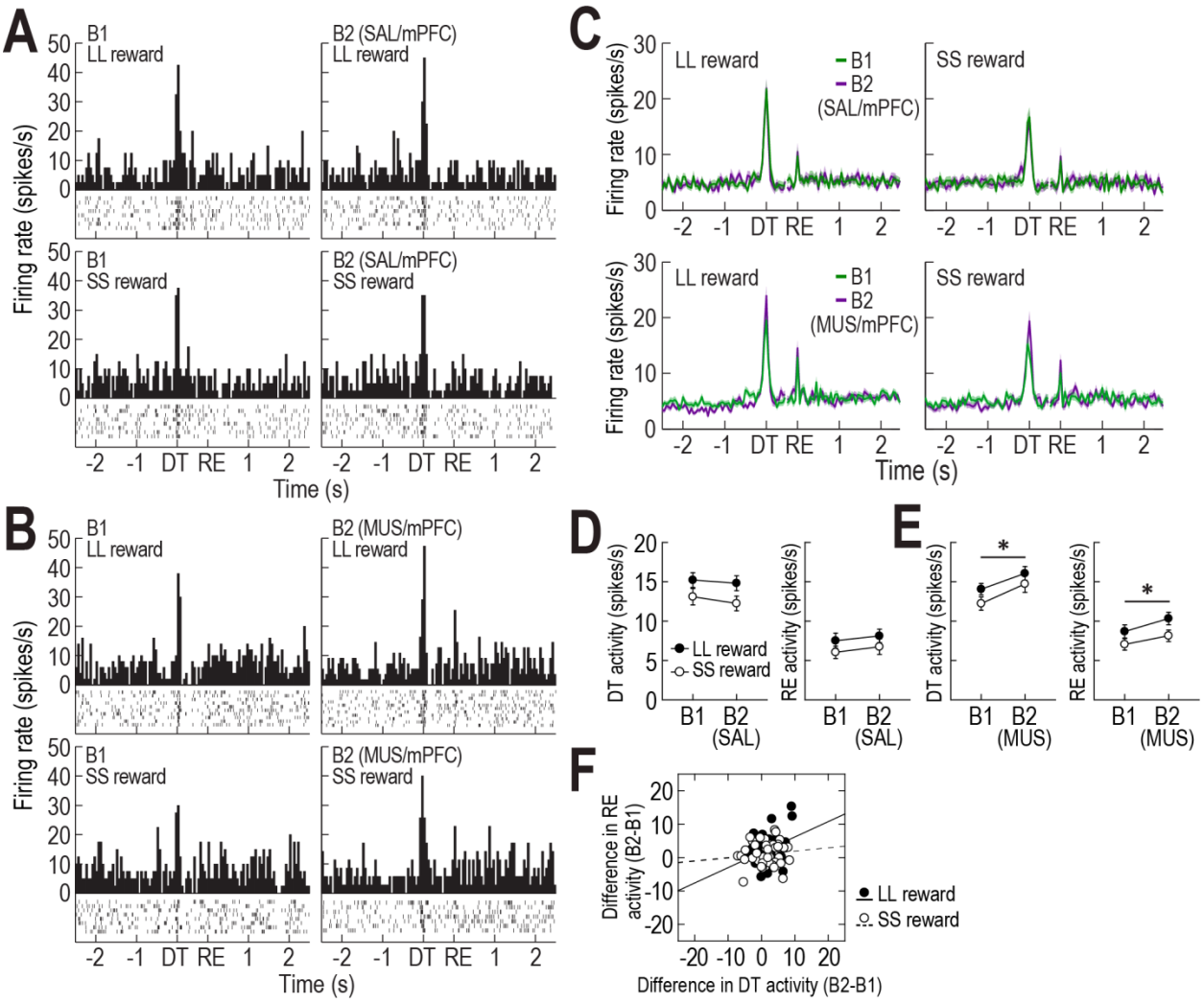

Figure 2.8. Effects of mPFC inactivation on DA cells. (**A, B**) Examples of representative DA cells recorded with SAL (A) and MUS (B) injections. Histograms (bin width, 50 ms) are aligned on delay termination (DT) and reward (RE). (**C**) Population responses of all task-related DA cells. (**D, E**) Average DT and reward responses in SAL/mPFC (D) and MUS/mPFC (E) sessions (*$p < 0.01$). (**F**) Correlations between altered DT and reward responses across blocks in MUS/mPFC sessions. Shaded areas and error bars show SEM.

Prefrontal regulation of non-DA activity

Alternations in non-DA activity after PFC manipulations were also investigated. Similar to DA cells, non-DA cells showed significant decreases in spontaneous activity when MUS was infused into both PFC subregions (Table 2.2). Paired *t*-tests demonstrated significant differences in firing before and after MUS injections into the OFC and the mPFC (t values > 3.21, p values < 0.01), whereas no significant changes in spontaneous activity were found in SAL sessions regardless of the subregions (t values < 0.33, p values > 0.74).

Table 2.2. Effects of prefrontal inactivation on spontaneous activity of non-DA cells

Subregion	Drug injected before B2	Mean firing rate	
		B1	B2
OFC	SAL	10.6 ± 1.3	10.5 ± 1.3
	MUS	9.7 ± 1.1	7.7 ± 0.9**
mPFC	SAL	12.1 ± 1.8	12.1 ± 1.8
	MUS	10.4 ± 1.6	8.6 ± 1.3*

Asterisks indicate a significant difference between before (B1) and after drug infusion (B2) on a paired *t*-test (*$p < 0.01$, ** $p < 0.001$; mean ± SEM).

As previously defined, 10, 11, 9, and 13 reward-related non-DA cells that signaled expected reward values were found during SAL/OFC, MUS/OFC, SAL/mPFC, and MUS/mPFC sessions, respectively. MUS injections into the OFC strikingly disrupted the development of ramping activity during both delays prior to LL and SS rewards (Fig. 2.9A). This observation was supported by significant effects of block (repeated measures ANOVA, $F_{(1,10)} = 10.35$, $p < 0.01$; Fig. 2.9B) and size ($F_{(1,10)} = 12.02$, $p < 0.01$), but no significant interaction between the factors ($F_{(1,10)} = 0.25$, $p = 0.63$). Moreover, a planned comparison revealed that the decreased delay responses after OFC inactivation was not different between SS and LL reward conditions ($t_{(10)} = 2.04$, $p = 0.07$). The reward activity was also slightly decreased, but a repeated measures ANOVA failed to find an effect of block ($F_{(1,10)} = 1.97$, $p = 0.19$). Instead, an effect of reward size was only significant ($F_{(1,10)} = 50.72$, $p < 0.001$) without an interaction between the variables ($F_{(1,10)} = 0.01$, $p = 0.91$). After MUS injections into the mPFC (Fig. 2.9C) or SAL injections into either subregion, any distinct alternations were not observed relative to the baseline firing. Repeated measures ANOVAs found that delay and reward activity in these sessions significantly differed between two reward sizes (F values > 7.29, p values < 0.05; Fig. 2.9D), but there were no effects of block (F values < 4.31, p values > 0.06) and no interactions between the variables (F values < 1.73, p values > 0.21). These results provide compelling evidence that expected reward values signaled by non-DA cells require the OFC, but not mPFC.

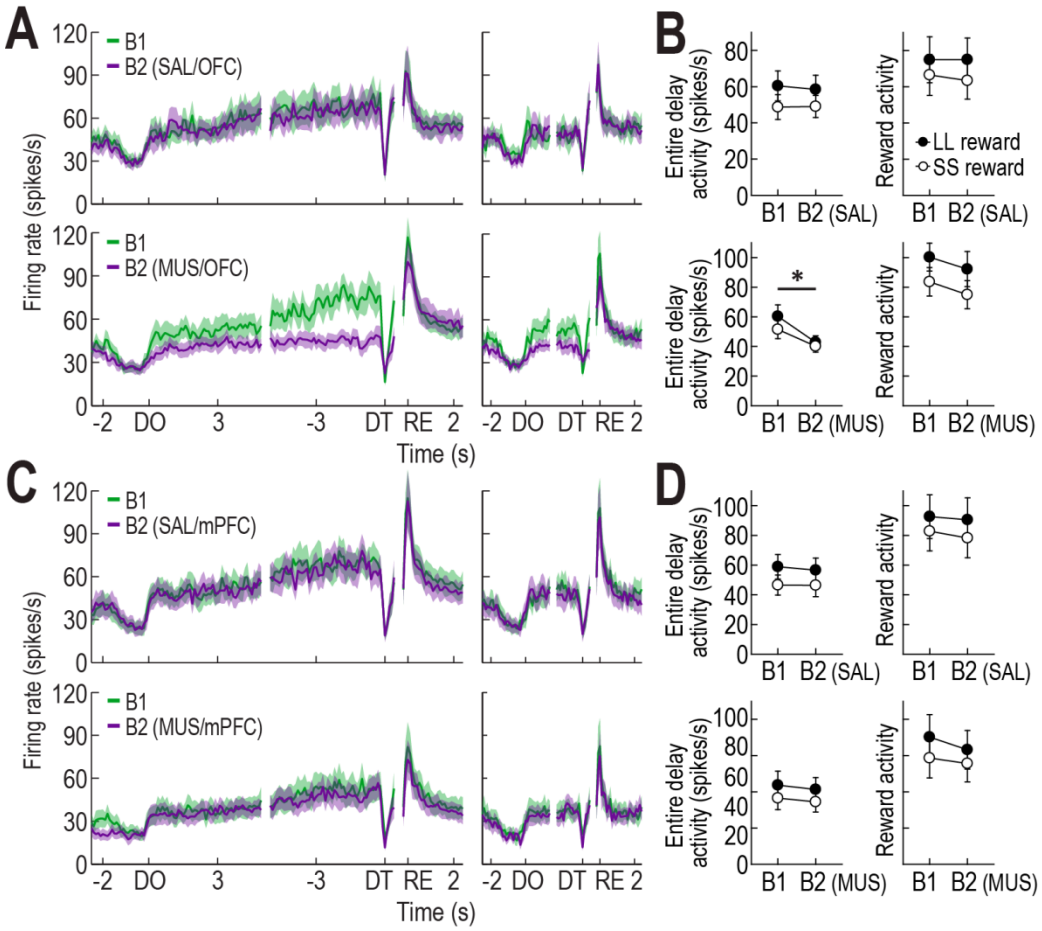

Figure 2.9. Effects of OFC and mPFC inactivation on reward-responsive non-DA cells. (**A**) Population histograms (bin width, 100 ms) of all reward-responsive non-DA cells recorded with OFC manipulations. Data are aligned on delay onset (DO) and termination (DT). (**B**) Average activity during the entire delays and reward activity with SAL or MUS injections into the OFC (*$p < 0.01$). (**C**) Population activity of all reward-responsive non-DA cells recorded with mPFC manipulations. (**D**) Average activity during the entire delays and reward activity with SAL or MUS injections into the mPFC. All graphs represent mean ± SEM.

In SAL/OFC, MUS/OFC, SAL/mPFC, and MUS/mPFC sessions, 14, 16, 13, and 13 non-DA cells, respectively, were identified as delay-excited. Similar to reward-related non-DA cells, delay-excited non-DA cells exhibited weaker ramping activity after OFC inactivation relative to the baseline (Fig. 2.10A). A repeated measures ANOVA revealed a significant effect of block ($F_{(1,15)} = 10.57$, $p < 0.01$; Fig. 2.10B), but no effect of reward size ($F_{(1,15)} = 1.74$, $p = 0.21$) and no interaction between the factors ($F_{(1,15)} = 2.16$, $p = 0.16$). Any noticeable alternations were not found when MUS was infused into the mPFC or SAL was injected into both PFC subregions (Fig. 2.10C), as indicated by no significant effects of block (repeated measures ANOVA, F values < 2.72, p values> 0.12; Fig. 2.10D) and no interactions between block and reward size (F values < 0.36, p values > 0.56). Thus, OFC inactivation disturbed an anticipatory increase in general motivation encoded by delay-excited non-DA cells.

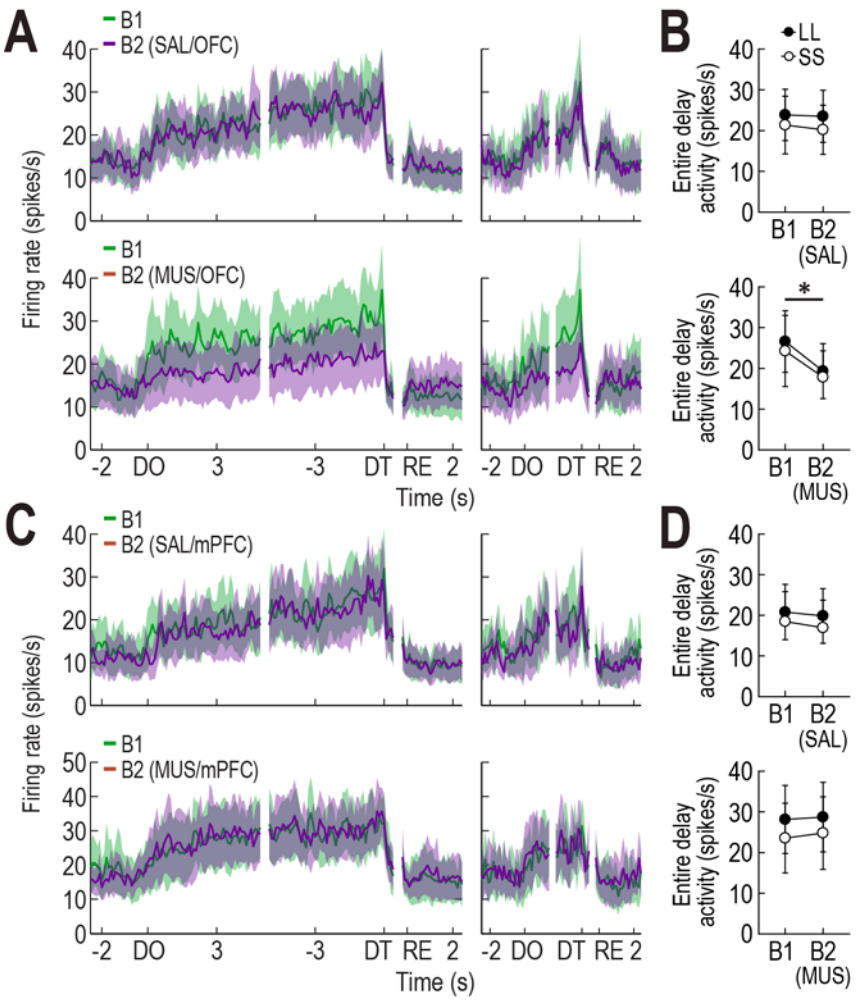

Figure 2.10. Effects of OFC and mPFC inactivation on delay-excited non-DA cells. (**A**) Population histograms (bin width, 100 ms) of all delay-excited non-DA cells recorded with OFC manipulations. Data are aligned on delay onset (DO) and termination (DT). (**B**) Average activity during the entire delays with SAL or MUS injections into the OFC (*$p < 0.01$). (**C**) Population activity of all delay-excited non-DA cells recorded with mPFC manipulations. (**D**) Average responses during the entire delays with SAL or MUS injections into the mPFC. All graphs show mean ± SEM.

Discussion

The current study characterized how DA and non-DA cells in the VTA responded in a delay discounting task. Consistent with prediction error signals observed in Pavlovian conditioning paradigms (Schultz et al., 1997; Pan et al., 2005; Clark et al., 2010), DA cells initially showed phasic responses to reward but such responses were no longer present as recording sessions progressed. Instead, these neurons continuously responded at the time of DT, which indicated that the removal of the wooden barriers served as the reward-predicting cue in the task. The barriers themselves were not reward-predicting stimuli because DA cells displayed no phasic responses when rats first arrived in front of them at the time of DO. Among non-DA cells, reward-related neurons signaled expected reward values by gradually ramping up their firing over the course of waiting periods and peaking at the time of obtaining delayed rewards. Recent evidence indicates that the information about reward expectancy is represented by GABAergic neurons in the VTA (Cohen et al., 2012). Another group of non-DA cells also exhibited ramping activity only during delays, but not during reward consumption. These firing patterns may reflect elevated motivation in anticipation of delayed rewards.

Previous primate studies suggest that DA cells signal the subject value of delayed rewards when cues predicting the future outcomes were presented prior to waiting periods (Fiorillo et al., 2008; Kobayashi and Schultz, 2008).In support of the findings, DA activity at the time of DT decreased as the length of delay before LL reward increased. Given that expected reward value in the current state exerts a positive influence on DAergic prediction errors, the graded phasic activity may result from differential reward expectancies signaled by reward-related non-DA cells during the late phase of delays (Fig. 2.4A, C). Thus, the present study suggests that DA phasic responses triggered by predictive cues encode the discounted value of

future rewards regardless of whether the predictive cues are presented before or after waiting periods.

OFC contribution to generating expected reward value

The OFC has long been known to encode the value of expected reward (Tremblay and Schultz, 1999; O'Doherty et al., 2001; Gottfried et al., 2003; Izquierdo et al., 2004; Padoa-Schioppa and Assad, 2006; Roesch et al., 2006). However, it was recently proposed that the OFC is critical for representing various states of the behavioral task from which estimates of expected future outcomes are derived, rather than representing values *per se* (Schoenbaum et al., 2011; Wilson et al., 2014). Although the nature of its role remains elusive, the present study found that OFC inactivation disrupted expected reward values signaled by non-DA cells during delays. Consequently, DAergic prediction errors were severely altered as if delayed rewards were less expected than before OFC inactivation. In the absence of OFC inputs, specifically, DA cells were less excited at the time of removing the barriers, but they exhibited greater phasic responses to actual rewards. These results provide compelling evidence that the OFC is necessary for proper prediction error signaling by contributing to accurately computing the value of expected rewards in the current state.

mPFC regulation of DA activity

Previous work indicates that the mPFC conveys value signals (Sul et al., 2010) and can alter expected reward signals by non-DA cells in a mPFC-dependent task (Jo et al., 2013). However, mPFC inactivation in the delay discounting task resulted in no changes in activity of non-DA cells during waiting periods for future rewards. Hence, it appears that the mPFC is not the major

prefrontal subregion that provides reward expectancies to the VTA in the task. In support of the idea, the finding that mPFC-inactivated rats showed perseverative choice biases to one spatial location (Fig. 2.5D) implies that the mPFC contributes to responding based on previous knowledge of task rules such as place-outcome associations, rather than value-based responding (Stefani and Moghaddam, 2005; Jung et al., 2008).

Despite no effects on non-DA cells, mPFC dysfunction increased DA phasic responses to both reward-predicting cues at the time of DT and actual rewards. The elevated prediction errors were not due to the general disinhibition after reduced inputs from the mPFC, because mPFC inactivation lowered spontaneous burst firing of DA cells. Instead, among three major components (i.e, currently available reward and two future reward values estimated in the current and the previous state) that are required for DA cells to compute prediction errors (Montague et al., 1996; Schultz et al., 1997; Pan et al., 2005), it is possible that the mPFC may contribute to maintaining the previous state value in its working memory function. Since this information negatively modulates DA activity, mPFC dysfunction can lead to increases in prediction error signaling. In line with this hypothesis, it is suggested that mPFC neurons temporarily store representations about past choice and its consequence (Sul et al., 2010). Alternatively, it is possible that the mPFC may convey temporal information to DA cells. It is well known that DAergic prediction errors are influenced by the precise temporal relationships between reward-predicting cues and reward delivery (Montague et al., 1996; Fiorillo et al., 2008; Kobayashi and Schultz, 2008). For example, DA cells exhibit phasic responses to a well-expected reward if it is delivered at a different time than scheduled. Therefore, if mPFC-inactivated rats perceived the lengths of delay to be shorter than the actual delay due to the inability to keep track of the elapse of time during waiting periods after mPFC dysfunction, DAergic prediction errors at the times of

DT and reward would increase. Indeed, a recent study found that the mPFC is crucial for processing time intervals (Kim et al., 2013).

Chapter 3: Prospective, concurrent, and retrospective evaluation of outcomes in the orbitofrontal cortex

Introduction

A large body of literature indicates that the orbitofrontal cortex (OFC) in primates and humans plays an essential role in value-based decision making by signaling the subjective value of expected outcomes (Tremblay and Schultz, 1999; O'Doherty et al., 2001; Gottfried et al., 2003; Izquierdo et al., 2004; Padoa-Schioppa and Assad, 2006; Hare et al., 2008; Kennerley et al., 2011). In contrast, it has been suggested that the rodent OFC does not encode general affective values *per se*, but rather specific information about the expected outcomes (e.g., odor, magnitude, spatial location, and temporal delay to delivery) (Feierstein et al., 2006; Roesch et al., 2006; van Duuren et al., 2007). Schoenbaum and his colleagues further claim that after combining the outcome information and other multisensory inputs, the OFC is critical for determining the current situation or state of an ongoing behavioral task (Schoenbaum et al., 2011; Takahashi et al., 2011; Wilson et al., 2014). These discrepancies are partially due to the functional heterogeneity of the OFC across species and different testing procedures (Padoa-Schioppa and Cai, 2011; Wallis, 2012), but further investigation is needed to elucidate the exact role of the OFC.

It was previously observed that OFC inactivation severely altered neural activity in the ventral tegmental area (VTA) due to a loss of expected reward values while rats performed a delay discounting task on a T-maze (chapter 1). To examine whether the OFC conveyed relative value signals or different state information, the major ingredient to compute value, to other brain regions such as the VTA, single units in the OFC were recorded in the same delay discounting

task as well as in a simple reward discrimination task. Neuronal responses to various salient states that animals went through during performance were identified and compared between the two behavioral tasks.

Materials and methods

Subjects

Four Long-Evans rats (320-380 g; Simonson Labs, Gilroy, CA) were individually housed in cages with free access to food and water. Before behavioral training, daily food access was restricted to maintain 80% of their free-feeding weight. All experiments were conducted during the light phase of a 12 hour light/dark cycle (lights on at 7:00 am), in compliance with the University of Washington's Institutional Animal Care and Use Committee guidelines.

Behavioral apparatus

An elevated T-maze that consisted of one start arm (the middle stem) and two goal arms (58 × 5.5 cm each) was used in the study. A metal food cup (0.7 cm in diameter × 0.6 cm deep) was located at the end of each goal arm. A wooden barrier (10 × 4 × 15 cm) was located in front of the food cup to prevent immediate access to reward. The proximal end of the goal arms was hinged to be lowered by remote control if one goal arm had to be presented alone.

Presurgical training

Each rat was allowed to freely explore and consume chocolate milk drops scattered throughout the maze for familiarization. Afterward, the animal was trained to obtain a 3 s-delayed reward (0.15 ml) only from the goal arms. In each trial, specifically, the rat was required to choose one of the goal arms after placed in the start arm. As it arrived in front of the barrier, the elapsed time was measured by an experimenter using a digital stopwatch. After the barrier was removed by

the experimenter at the end of the delay, the animal could approach and consume the reward. The experimenter gently guided the rat to the start arm for next trial after re-baiting the food cup and replacing the barrier back in place. Once the rat finished 16 trials within 20 min, it received surgical implantation for recording electrodes.

Surgery

Tetrodes were constructed from 20μm lacquer-coated tungsten wires (California Fine Wire, Grover Beach, CA) and tetrode tips were plated with gold to a final impedance of 0.2-0.4 MΩ (tested at 1 kHz). A drivable bundle of six tetrodes was chronically implanted under stereotaxic guidance in left hemisphere dorsal to OFC (3.0 mm anterior to bregma, 3.1 mm lateral to midline, and 4.0 mm ventral to the brain surface).

Delay discounting and reward discrimination tasks

After at least 10 days of recovery, all rats performed two different behavioral tasks on the T-maze per day. The first task was a delay discounting task that required each rat to choose between a sooner small (SS) reward and a later large (LL) reward. The delay to SS reward (0.05 ml) was kept constant at 3 s throughout the task, whereas three different delays (10, 20, and 40 s) before LL reward (0.3 ml) were used to examine choice behavior as a function of delay to LL reward. The three delays to LL reward were randomly assigned to one of three blocks of trials and only one selected delay was used in a given block. To inform which delay was associated with LL reward, each block began with 10 forced-choice trials followed by 8 free-choice trials. During the forced-choice trials, 5 SS and 5 LL reward trials were pseudorandomly given and only one goal arm was made available in each trial by lowering the other goal arm. Both goal arms were present during the free-choice trials in which animals' choice preference for LL reward was analyzed. After the rats chose and entered the goal arm associated with LL reward,

an additional barrier was placed at its entrance to prevent them from exiting the chosen arm while waiting for LL reward. The three testing blocks were separated by an interblock interval of 5-10 min during which the animals were placed on a holding area adjacent to the maze. The location of SS and LL rewards in the goal arms remained constant within each rat but was counterbalanced across rats.

The other daily task was a reward discrimination task in which rats had to discriminate two goal arms associated with a small (0.05 ml) and a large (0.3 ml) reward. The testing procedures were similar to the delay discounting task except for two modifications. First, no delay was imposed before acquiring reward. As soon as rats arrived in front of the barrier in a goal arm, it was immediately removed by an experimenter. Second, the locations associated with two different sizes of reward were randomly chosen in the first blocks of trials and were reversed in the next block. The testing order of the two tasks was randomized in each day.

Single-unit recording

Prior to the daily behavioral tasks, neural activity in the OFC was monitored using a Cheetah data acquisition system (Neuralynx, Bozeman, MT). Unit signals were amplified, filtered at 600-6000 Hz, and digitized at 32 kHz. The rat's head position was also recorded at 30 Hz by tracking two LEDs mounted on the tetrode assembly. If no stable and isolated units (> 2:1 signal-to-noise ratio) were detected, tetrodes were advanced in 40 μm increments, up to 160 μm per day. Once such units were found, behavioral recording was conducted, and all tetrodes were advanced to find different units the end of the recording. While the recording was progressing, three salient events such as delay onset (DO), delay termination (DT), and reward (RE) were inserted into the data stream online. Specifically, timestamps for DO and DT were flagged when an experimenter pressed the stopwatch buttons. The time of obtaining reward was automatically detected by 'lick-

detectors' (custom designed by Neuralynx) when the animal first lick chocolate milk in the food cups.

Histology

After completion of all experiments, small electrolytic lesions were made by passing a 10 μA current for 10 s through two wires of each tetrode. Then, all rats were transcardially perfused and their brains were stored in a 10% formalin/30% sucrose solution and cut in coronal sections (40 μm) on a freezing microtome. The sections were stained with cresyl violet and examined under light microscopy to reconstruct tetrode tracks. Recording data histologically confirmed to be recorded only from the VTA were further analyzed.

Data analysis

Spikes from single units were isolated by cluster cutting of multiple waveform parameters using Offline sorter (Plexon, Dallas, TX). Subsequent analysis of the sorted units was performed with Matlab software (Mathworks, Natick, MA). To classify OFC cells responding to various salient events in the current tasks, peri-event time histograms (PETHs; bin width, 50 ms) were constructed for the 5 s epoch around DO and DT events and for a 10 s window around RE encounters (-2.5 to 7.5 s) in each block of all trials. An OFC cell was considered as phasically excited in response to each event if it passed the following two criteria: 1) its peak firing was observed within the 300 ms window around the event 2) its average activity within the corresponding window was ≥ 200% of its mean firing rate over each block of trials. In the same way, a OFC cell showing long-lasting consummatory reward activity or post-reward activity were also classified if its excited responses were observed in the 1.5 s epoch after RE or a 1 s epoch after reward consumption (0.5 to 1.5 s after small reward and 3.2 to 4.2 s after large reward), respectively. Additionally an OFC cell displaying elevated responses during the four

different lengths of delay was identified if its average z-score during at least one of the delays, but not during the 2.5 s period before DO, was greater than 2 after discharge rats during the entire delay of each trial were converted to z-scores relative to the mean firing of all trials within the block. To examine a linear relationship between neural activity and delay to LL reward, Spearman's rank correlation coefficient was calculated in each time bin of the PETHs. The significance of correlation was estimated using a permutation test in which discharge rates of each bin were randomly shuffled across the three blocks for 1000 times. A 99% confidence interval was obtained based on the distribution of the shuffled correlation coefficients.

The animals' movement during recording sessions was analyzed by measuring the instantaneous velocity, the distance between two consecutive head positions sampled at 30 Hz.

Statistical analysis

Repeated-measures ANOVAs and *t*-tests were used to evaluate statistical significance of neural activity across different reward conditions. Pearson's correlation tests were performed to establish a relationship between two variables. Two-tailed p values < 0.05 were considered statistically significant. Data are expressed as mean ± SEM.

Results

Choice behavior

Four rats were tested daily in both a delay discounting task and a reward discrimination task on an elevated T-maze (Fig. 3.1A). The former required the animals to choose between SS and LL rewards. To evaluate behavioral preference for LL reward as a function of the delay preceding it, three different delays prior to LL reward were randomly ordered and tested in separate blocks of trials, whereas the delay to SS reward remained constant (Fig. 3.1B). Each block started with 10

forced-choice trials, followed by 8 free-choice trials in which choice performance was measured. A wooden barrier was placed in front of the food cup of each goal arm to prevent rats from having access to reward during delays. While monitoring the elapsed time using a digital stopwatch, an experimenter removed the barrier at the time of delay termination (DT). The manual measurement of delays caused slight variations in delay length. In a total of 75 behavioral recording sessions (11-29 sessions per rat), SS reward was delayed by 3.2 ± 0.14 s (mean ± SD) and LL reward was delayed by 10.17 ± 0.24, 20.26 ± 0.44, or 40.27 ± 0.36 s. The location of SS and LL rewards in the goal arms remained stable within rats but was counterbalanced between rats. For the second task, animals had to discriminate between small and large rewards. The barriers were immediately removed upon rats' arrival in front of them. Although no delay was imposed before obtaining reward, the barriers were continuously used to ensure that the animals' approaching behavior to the end of two goal arms was similar as in the delay discounting task (Fig. 3.1C, D). The goal arms associated with SS and LL rewards were randomly selected in the first block of trials and were switched in the second block.

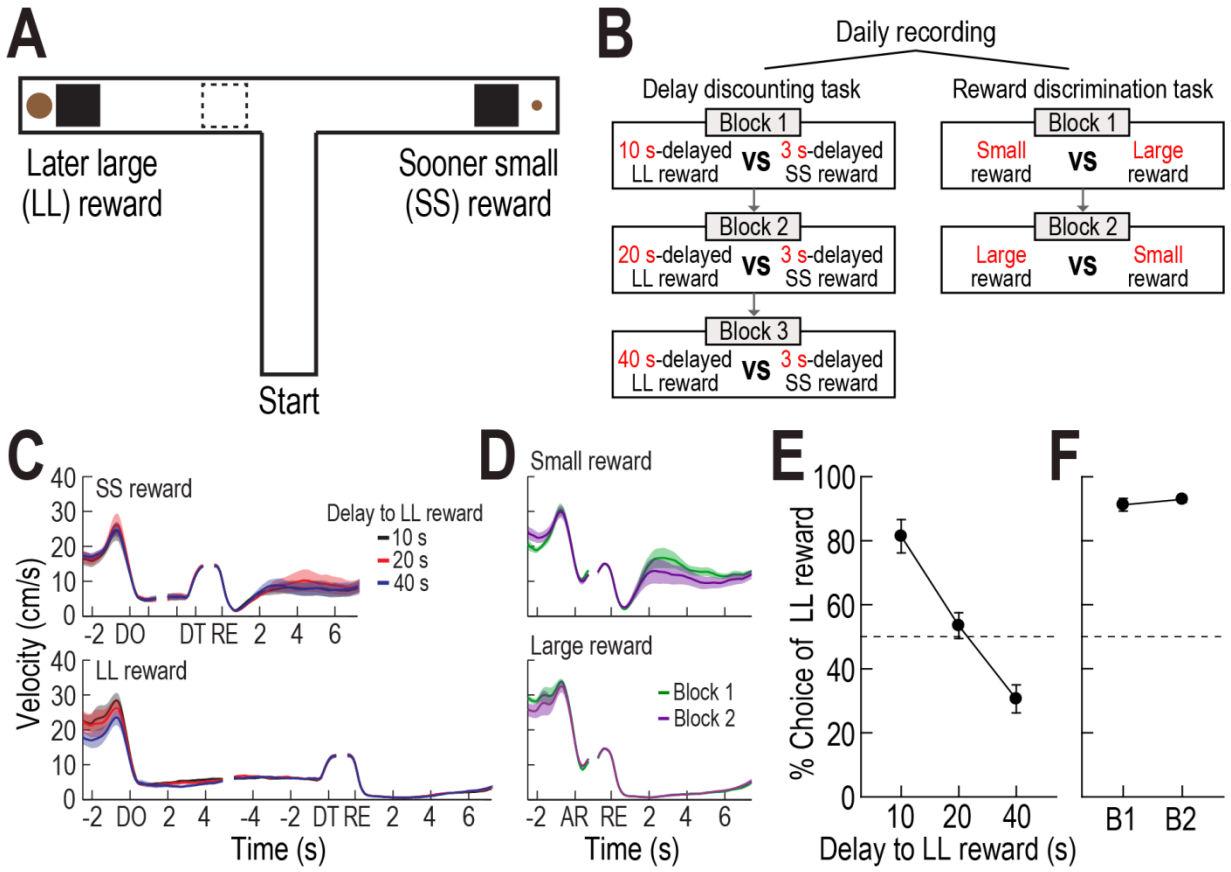

Figure 3.1. Behavioral performance on choice tasks. (**A**) Illustration of the T-maze. Rectangular wooden barriers (black squares) were placed before food cups on the goal arms to control animals' access to SS and LL rewards. When rats chose a goal arm associated with LL reward, an additional barrier (dashed rectangle) was placed at its entrance to prevent the animals from exiting the chosen arm during the waiting period. (**B**) Daily behavioral recording procedures. A daily recording sessions consisted of delay discounting and reward discrimination tasks. In the former task, three different lengths of delay to LL reward were randomly ordered and tested in separate blocks of trials. The delay to SS reward remained unchanged. In the latter task, small and large rewards were immediately accessible without delays, but the goal arms associated with the rewards were switched between blocks of trials. (**C, D**) Changes in instantaneous velocity. Data in the delay discounting task (C) are aligned on delay onset (DO), delay termination (DT), and reward (RE). Velocities in the reward discrimination task are aligned on arrival in front of the barriers (AR) and RE. (**E, F**) Choice preference for LL reward in the delay discounting (E) and reward discrimination task (F). Shaded areas and error bars indicate SEM.

During the delay discounting task, rats strongly preferred a 10 s-delayed LL reward over a 3 s-delayed SS reward (Fig. 3.1E). However, the preference disappeared as the delay to LL reward increased to 20 s. Further extension of the delay to 40 s induced the rats to prefer the SS reward. A significant inverse correlation between behavioral preference and delay to LL reward (Pearson's correlation, $r = -0.94$, $p < 0.001$) indicated temporal discounting of LL reward. When tested in the reward discrimination task, the same animals also showed value-guided choice behavior (Fig. 3.1F). Despite the reversal of reward location in the second block, they exhibited comparable levels of behavioral preference for large reward (paired *t*-test, $t_{(3)} = 1.56$, $p = 0.22$).

Neural correlates in the OFC during the delay discounting task

While performing the task, rats went through a series of salient events and periods such as delay onset (DO), delay period, DT, reward, and post-reward period after reward consumption. Out of a total of 884 neurons recorded in the OFC (Fig. 3.2), 513 (58%) neurons responded to these various task states. Most of the task-related cells exhibited specific responses to one event or period, but some neurons had multiple correlates (Fig. 3.3). Thus, subsequent analysis for each event or period included any task-related cells that were responsive to it.

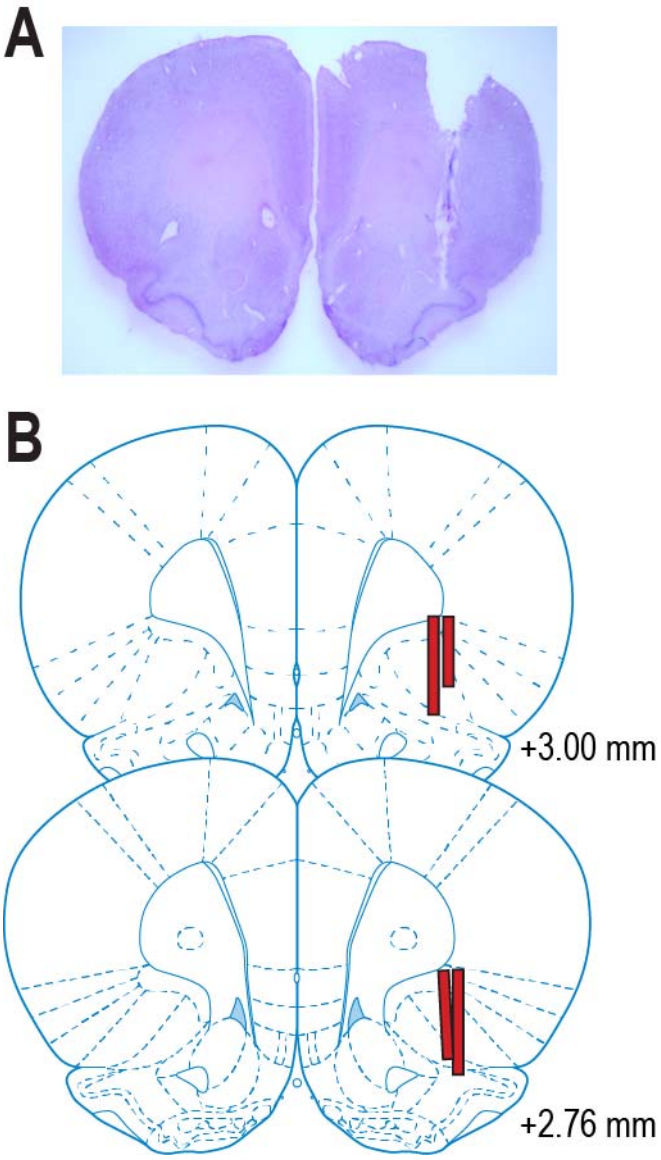

Figure 3.2. Histological verification of recording sites. (**A**) Nissl-stained section showing the final location of a tetrode tip in the OFC. (**B**) Reconstruction of all tetrode tracks.

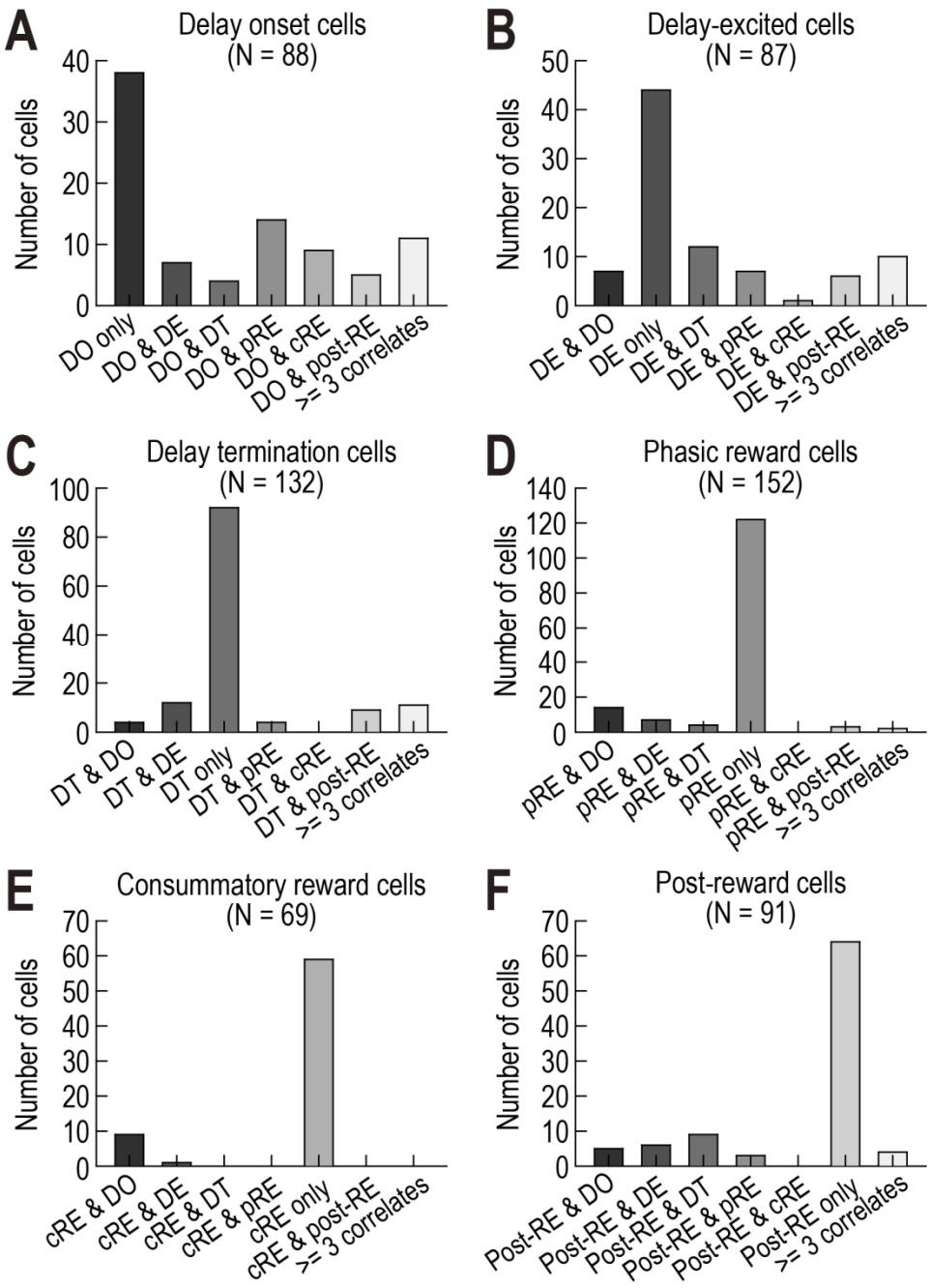

Figure 3.3. Summary of neural correlates in the OFC during the delay discounting task. (**A-F**) Number of OFC cells responding to multiple events and periods. Some OFC cells were phasically activated at the time of delay onset (DO, A), delay termination (DT, C), and reward (pRE, D), while other OFC cells were long-lastingly excited during delays (DE, B) and reward consumption (cRE, E) and shortly after reward consumption (post-RE, F).

At the time of DO when rats arrived in front of the barriers on the goal arms, 88 OFC cells showed phasic activity. Individual DO cells responded stronger in one reward condition than in the other (Fig. 3.4A). For instance, some DO cells preferentially responded when a SS reward (12/88, 13.6%), a 20 s-delayed LL reward (13/88, 14.8%), or a 40 s-delayed LL reward (18/88, 20.5%) was expected at the time of DO. A large number of DO cells (45/88, 51.1%) displayed stronger responses in choice trials for a 10 s-delayed LL reward. Therefore, each DO cell appeared to encode a specific reward condition. However, the population activity of all DO cells represented the subjective value of SS and LL rewards (Fig. 3.4B). DO activity in LL reward conditions decreased as LL reward was discounted by longer delays. Despite the same delay to SS reward, DO activity in SS reward conditions also changed across three blocks of trials; it was especially increased as the value of the alternative LL reward decreased. When Spearman's rank correlation coefficients for individual neuronal responses across blocks were calculated per time bin of 50 ms, significant negative and positive correlations were found around the time of DO in LL (-100 to 600 ms) SS reward conditions (-100 to 150 ms), respectively. Direct comparisons in average population activity between two reward options within blocks also supported the idea that DO cells reflected the relative desirability of a chosen option at the time of DO (Fig. 3.4C). A repeated measures ANOVA found a significant interaction between block and reward size ($F_{(2,174)} = 22.08$, $p < 0.001$) without main effects of the two factors (F values < 3.51, p values > 0.064). Post hoc tests revealed that when rats biased their choices toward a 10 s-delayed LL reward, DO activity was significantly greater in the LL reward condition than in the alternative SS reward condition ($p < 0.001$). As rats became indifferent between SS and a 20 s-delayed LL rewards, DO activity was comparable between the reward conditions ($p = 0.54$). Finally when the animals chose a SS reward more often than a 40

s-delayed LL reward, DO activity was significantly lower in the LL reward condition (p = 0.012).

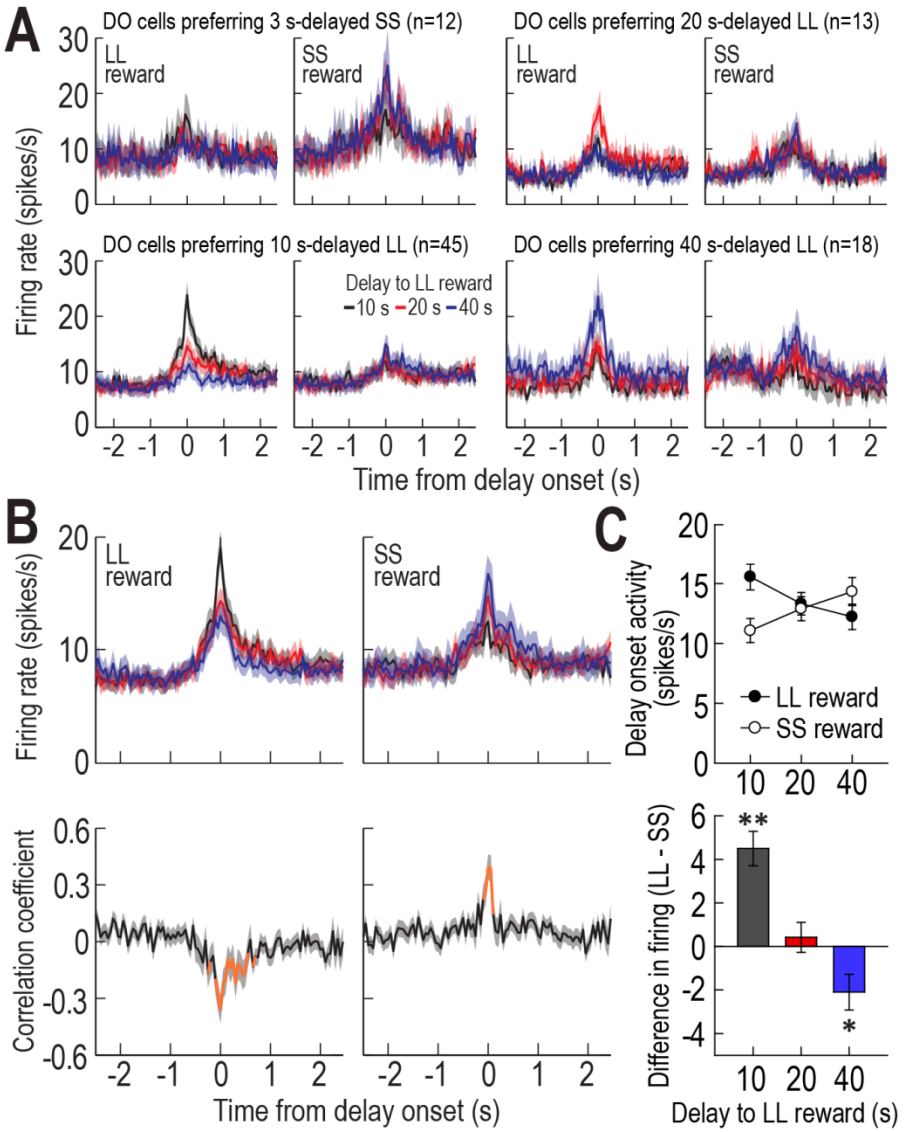

Figure 3.4. Delay onset activity in the delay discounting task. (**A**) Population histograms (bin width, 50 ms) of four groups of DO cells preferring different reward conditions. The reward condition in which a DO cell showed the strongest response was designated as preferred. (**B**) Population responses of all DO cells. Correlation coefficients for individual neuronal responses across three blocks of trials were calculated per each time bin. Orange data points that fell outside the 99% confidence interval obtained from a permutation test for at least two consecutive bins were considered significantly correlated. (**C**) Average DO activity and differential firing between LL and SS reward conditions within blocks of trials (*$p < 0.05$, **$p < 0.001$; *t*-test). Shaded areas and error bars indicate SEM.

During the waiting periods for delayed rewards, 87 OFC cells ramped up their firing. Individual delay-excited (DE) cells showed stronger ramping activity in a preferred delay (Fig. 3.5A). While 33.3% DE cells preferentially fired during the delay prior to SS reward, the other DE cells preferred one of the three delays to LL reward, especially the 10-s delay. These DE cells signaled relative value information at the population level, instead of the information about specific delays (Fig. 3.5B). DE responses before LL reward were less excited with longer delays as indicated by inverse correlations between DE activity and the three delays. DE activity prior to SS reward tended to increase as the alternative LL reward became less valuable, but significant positive correlations were observed briefly in the middle of the delay (0.8-1.2 s after DO). When differential firing between LL and SS reward conditions within blocks were analyzed during the early phase of delays (Fig. 3.5C), an ANOVA with repeated measures detected a significant interaction between block and reward size ($F_{(2,172)} = 20.01$, $p < 0.001$), but no main effects (F values < 3.23, p values > 0.076). *Post hoc* comparisons demonstrated a significant increase, no difference, and a significant decrease in the blocks involving LL rewards delayed by 10 s ($p = 0.005$), 20 s ($p = 0.11$), and 40 s ($p < 0.001$), respectively. These decision values got weaker during the later phase of delays (Fig. 3.5D). Specifically, *post hoc* tests following a significant interaction between block and reward size ($F_{(2,172)} = 17.49$, $p < 0.001$) consistently found a significant increase and no difference in firing within the blocks containing 10 s-delayed ($p < 0.001$) and 20 s-delayed LL rewards ($p = 0.44$), respectively. However, no significant decrease in DE activity was found in the block containing a 40 s-delayed LL reward ($p = 0.19$).

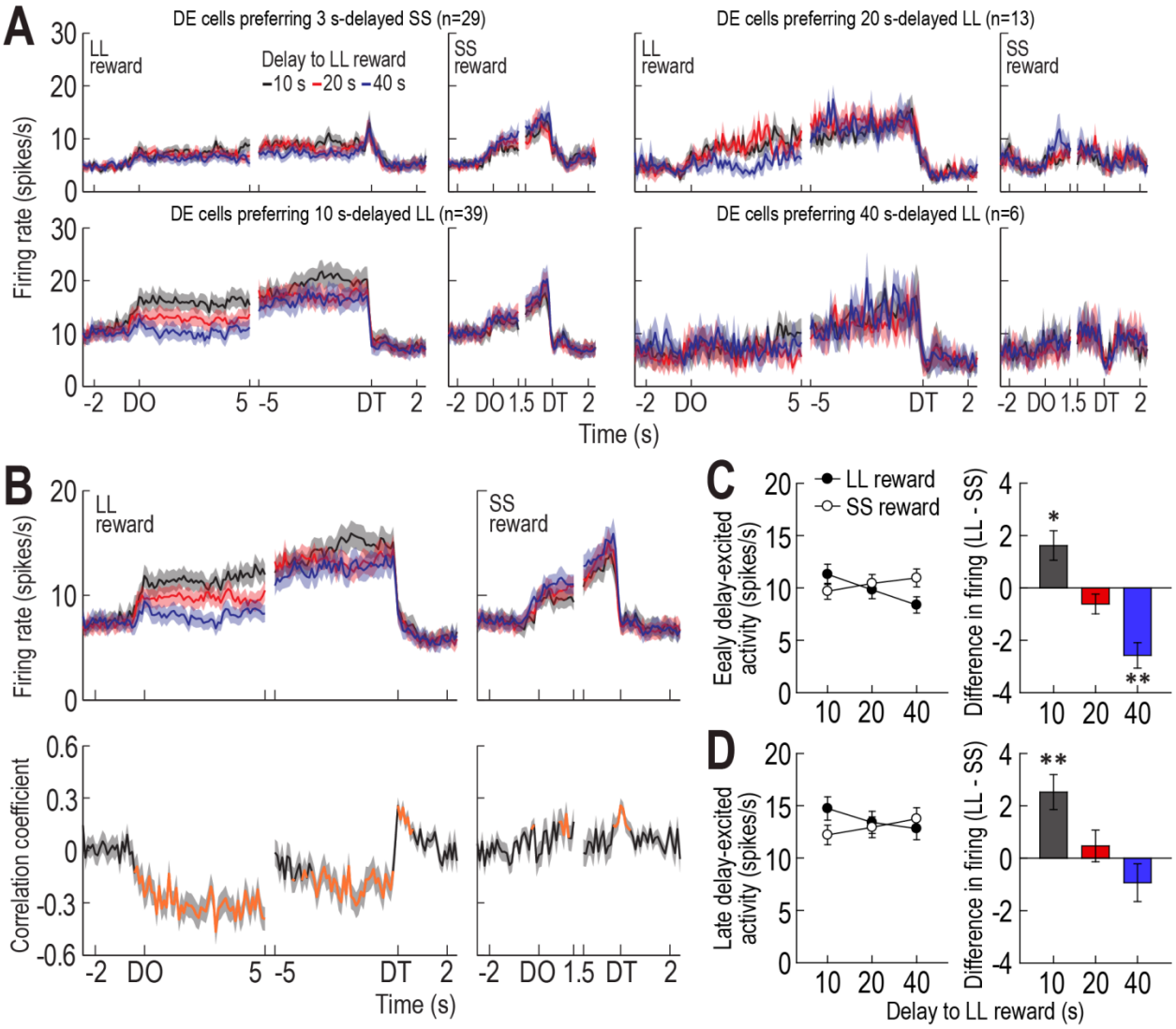

Figure 3.5. Delay-excited activity in the delay discounting task. (**A**) Population histograms (bin width, 100 ms) of four groups of DE cells preferring different reward conditions. Data are aligned on DO and DT. The reward condition in which a DE cell showed the strongest response during the entire delay was designated as preferred. (**B**) Population activity of all DE cells and correlation coefficients over the course of delays. Significant correlation coefficients for more than two consecutive bins were depicted in orange. (**C, D**) Average DE responses during an early phase (1.5 s after DO, C) and a late phase (1.5 s before DT, D). Differential firing rates between LL and SS reward conditions within blocks of trials were also measured during the two phases (*$p < 0.01$, **$p < 0.001$; *t*-test). Shaded areas and error bars represent SEM.

At the time of delay termination (DT) when the barriers were removed to allow rats to obtain reward, 132 OFC cells exhibited phasic responses. Individual DT neurons had a preferred reward condition by showing stronger responses in one reward condition over the other (Fig. 3.6A). Their population activity was modulated by the size of expected rewards regardless of delay lengths preceding the rewards (Fig. 3.6B). DT responses were not different across blocks within either reward condition, but were more elevated in anticipation of LL reward. A repeated measures ANOVA demonstrated a significant effect of reward size ($F_{(1,131)} = 19.51$, $p < 0.001$; Fig. 3.6C), whereas no effect of block ($F_{(2,262)} = 0.03$, $p = 0.97$) and no interaction between the two factors ($F_{(2,262)} = 0.25$, $p = 0.78$) were found.

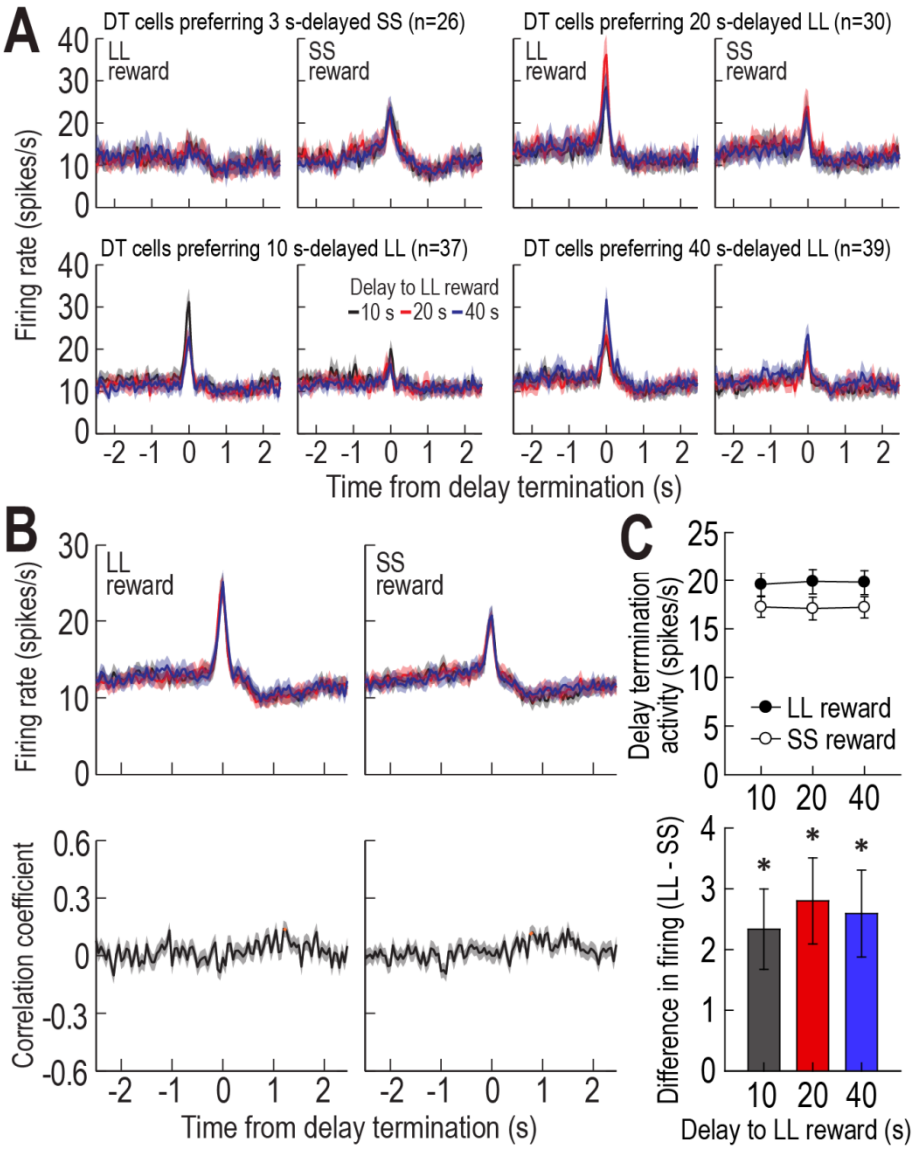

Figure 3.6. Delay termination activity in the delay discounting task. (**A**) Population histograms (bin width, 50 ms) of four groups of DT cells preferring different reward conditions. (**B**) Population responses of all DT cells and correlation coefficients around the time of DT. (**C**) Average DT responses and differential firing between LL and SS reward conditions within blocks of trials (*$p < 0.001$; *t*-test). Shaded areas and error bars show SEM.

When rats encountered reward, 152 OFC cells were phasically activated. Similar to DT cells, individual phasic reward (pRE) cells encoded a preferred reward as shown in Figure 3.7A, but population activity signaled the value of encountered rewards (Fig. 3.7B). A repeated measures ANOVA performed on population pRE activity showed a significant effect of reward size ($F_{(1,151)} = 26.95$, $p < 0.001$; Fig. 3.7C), but neither an effect of block ($F_{(2,302)} = 0.98$, $p = 0.38$) nor an interaction between the variables ($F_{(2,302)} = 0.03$, $p = 0.97$) was significant. Another group of 69 OFC cells displayed long-lasting excitation during reward consumption. Considering the animals' movement in the task (Fig. 3.1C), they finished reward consumption and started moving in about 1 s after obtaining SS reward and in about 3 s after encountering LL reward. The durations of elevated firing in response to SS and LL rewards closely matched the times taken to consume the two rewards. Some individual consummatory reward (cRE) cells (16/69, 23.2%) showed higher responses during the consumption of SS reward, whereas the other cells were more excited while consuming the LL reward. Two different amounts of reward were well represented at the population level. A repeated measures ANOVA revealed a significant effect of reward size ($F_{(1,68)} = 24.42$, $p < 0.001$; Fig. 3.8C), but no effect of block ($F_{(2,136)} = 2.08$, $p = 0.13$) and no interaction between the factors ($F_{(2,136)} = 0.14$, $p = 0.87$).

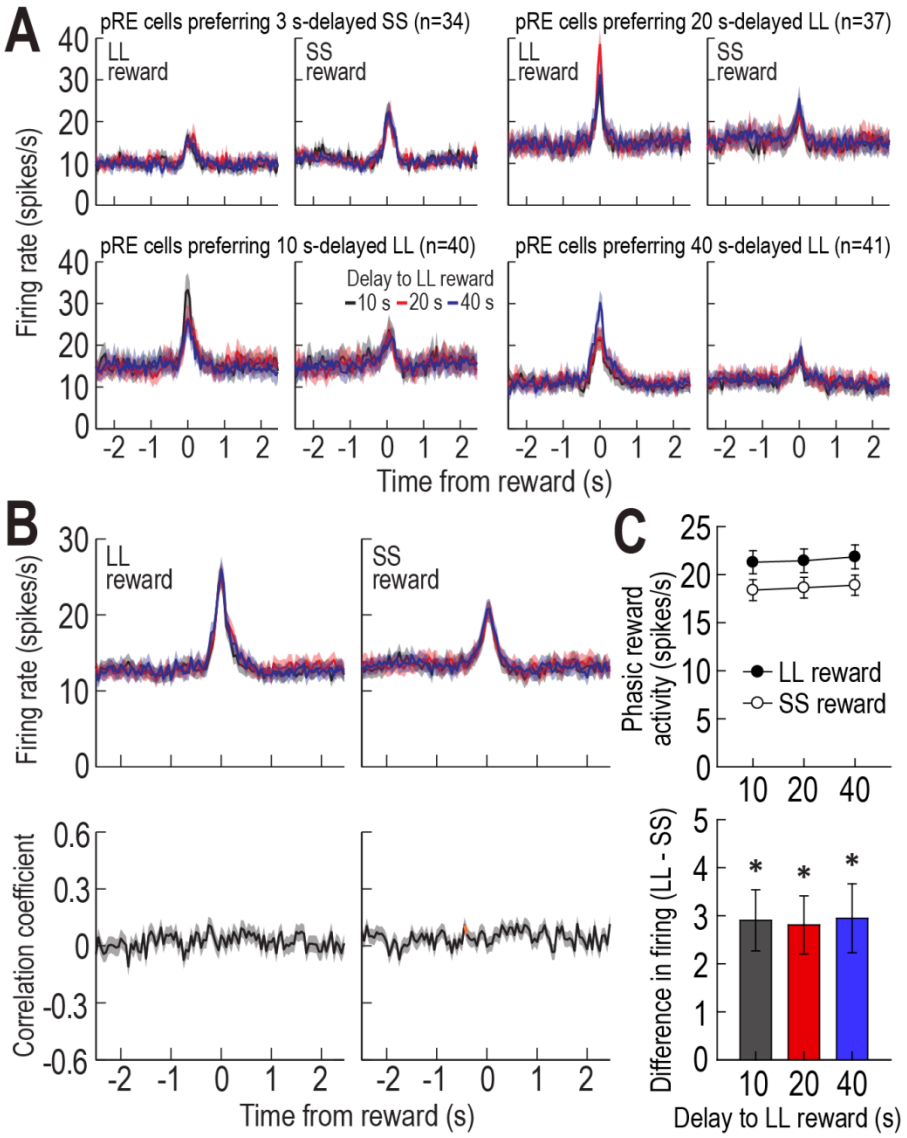

Figure 3.7. Phasic reward activity in the delay discounting task. (**A**) Population histograms (bin width, 50 ms) of four groups of pRE cells preferring differently delayed rewards. (**B**) Population responses of all pRE cells and correlation coefficients around the time of obtaining reward. (**C**) Average pRE responses and differential firing between LL and SS rewards within blocks of trials (*$p < 0.001$; *t*-test). Shaded areas and error bars indicate SEM.

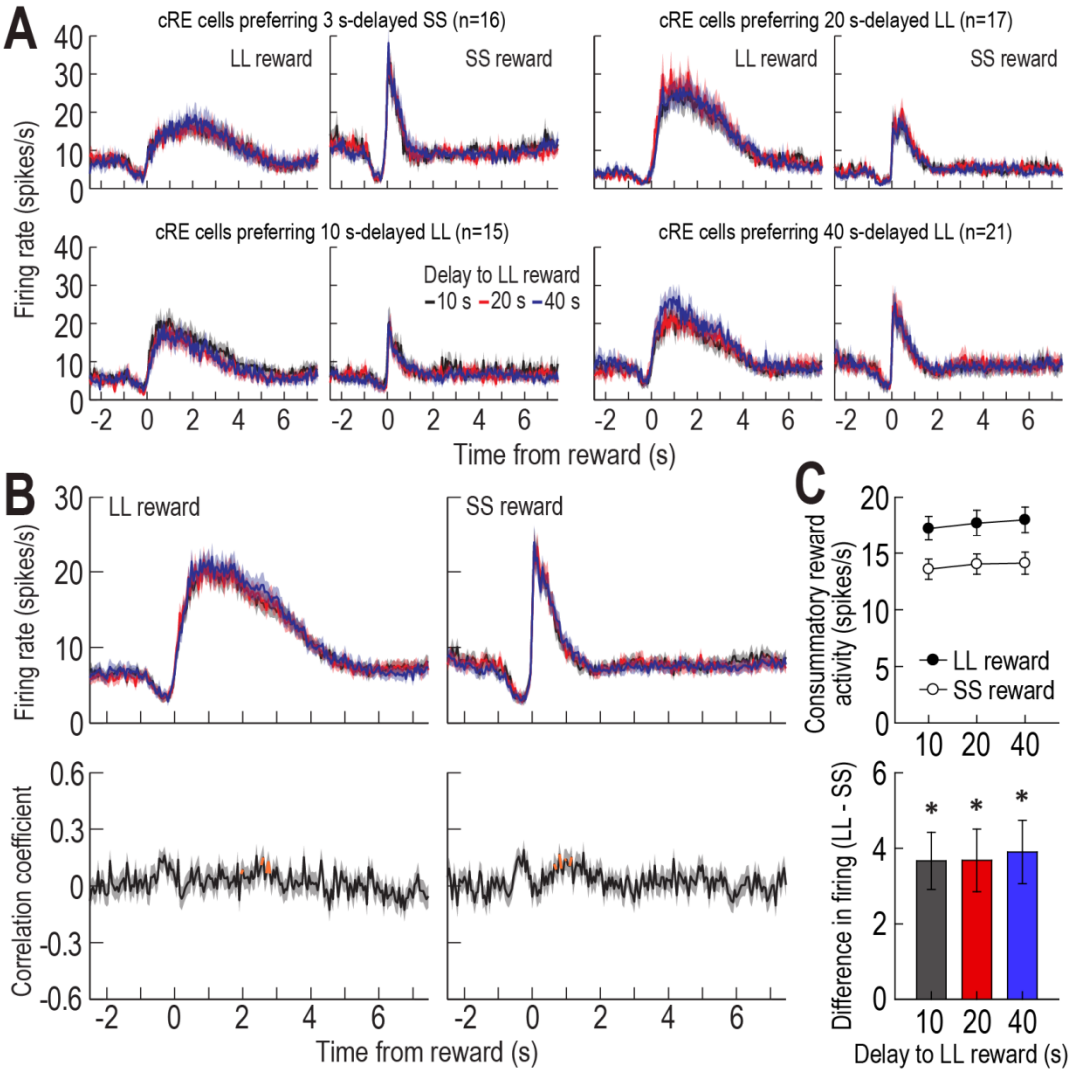

Figure 3.8. Consummatory reward activity in the delay discounting task. (**A**) Population histograms (bin width, 50 ms) of four groups of cRE cells preferring differently delayed rewards. (**B**) Population responses of all cRE cells and correlation coefficients while consuming reward. (**C**) Average cRE responses and differential firing between LL and SS rewards within blocks of trials (*$p < 0.001$; *t*-test). Shaded areas and error bars represent SEM.

Finally, 91 OFC cells showed post-reward (post-RE) activity which gradually developed during reward consumption and reached their peak after finishing consumption. Consistent with the pRE and cRE cells, individual post-RE cells were preferentially excited after the consumption of a particular reward (Fig. 3.9A). Their population activity encoded the value of consumed rewards (Fig. 3.9B). A repeated measure ANOVA found that post-RE responses were significantly greater after LL reward consumption than after SS reward consumption ($F_{(1,90)}$ = 12.38, p = 0.001; Fig. 3.9C), whereas post-RE firing rates were not different across three blocks ($F_{(2,180)}$ = 0.16, p = 0.85) and no interaction between reward size and block was found ($F_{(2,180)}$ = 2.98, p = 0.053). The post-reward activity was not attributed to rats' locomotor movement since it peak occurred while rats remained at the end of the goal arms, especially after consuming LL reward, typically engaging in exploratory behaviors such as licking and sniffing around the food cups. Rather, post-RE cells appeared to retrospectively evaluate the value of rewards that were just consumed.

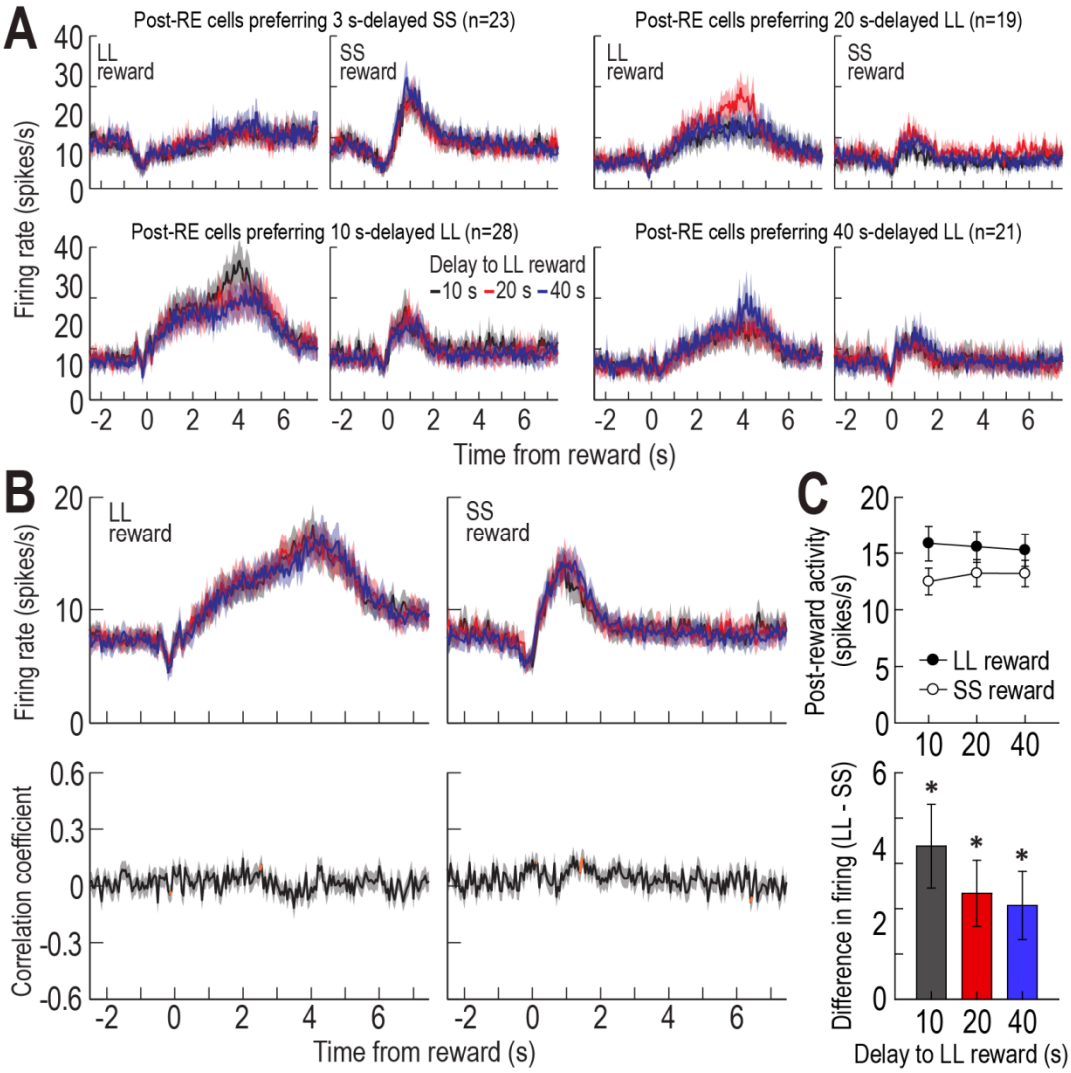

Figure 3.9. Post-reward activity in the delay discounting task. (**A**) Population histograms (bin width, 50 ms) of four groups of post-RE cells preferring different reward conditions. (**B**) Population responses of all post-RE cells and correlation coefficients after reward consumption. (**C**) Average post-RE responses and differential firing between LL and SS reward conditions within blocks of trials (*$p < 0.01$; *t*-test). Shaded areas and error bars show SEM.

Responses of OFC cells in the reward discrimination task

The same OFC cells were recorded during the reward discrimination task to determine task-dependent changes in neural activity. It was hypothesized that DO, DE, and DT cells would exhibit significant alternations in their firing, since the lack of delays in the task should cause the delay-related events and period less salient, compared to the delay discounting task. Indeed, 40.9% of individual DO cells became non-responsive when rats arrived in front of the barriers (Fig. 3.10A). The other DO cells phasically responded in only one goal arm, especially in the goal arm associated with LL reward during the delay discounting task. The spatial selectivity was task-dependent because these DO cells did not showed spatially biased firing in the delay discounting task. The population activity of DO cells no longer represented the value of expected rewards, as indicated by no effects of reward size ($F_{(1,87)} = 0.62$, $p = 0.43$) and block ($F_{(1,87)} = 0.28$, $p = 0.59$), and no significant interaction between the factors ($F_{(1,87)} = 0.82$, $p = 0.37$). In addition, similar altered firing patterns were observed from DE cells. Half of individual DE neurons selectively fired after arrival at the barrier in one goal location, the other half showed no responses (Fig. 3.10C). Their population responses failed to signal value information (Fig. 3.10D). A repeated measures ANOVA found no effects of reward size ($F_{(1,86)} = 3.59$, $p = 0.06$) and block ($F_{(1,86)} = 0.57$, $p = 0.45$), and no significant interaction between the variables ($F_{(1,86)} = 0.16$, $p = 0.69$). In the case of DT cells, no significant excitation was observed before obtaining reward (Fig. 3.10E). Overall, DO, DE, and DT responses did not signal which goal locations were associated with small and large rewards in a given block of trials.

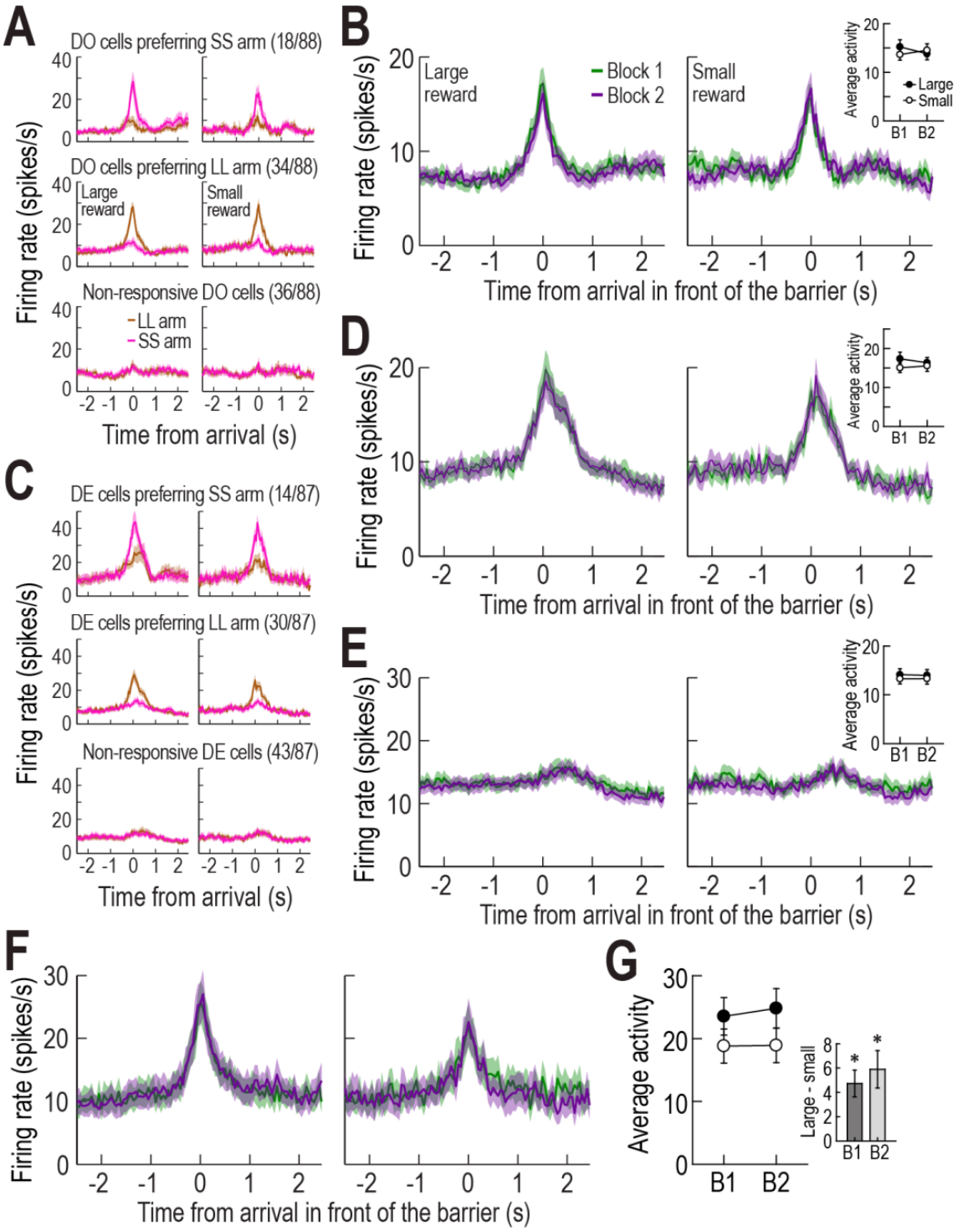

Figure 3.10. Activity of DO, DE, and DT cells in the reward discrimination task. (**A, C**) Population histograms (bin width, 50 ms) of DO (A) and DE cells (C) showing three different responses. Data are aligned on arrival in front of the barriers (AR). Some of the two groups of OFC neurons became non-responsive and the others display strong response biases to one goal

arm associated with either SS or LL reward in the delay discounting task. (**B, D, E**) Population activity of all DO (B), DE (D), DT cells (E). Inset graphs indicate average responses at the time of AR. (**F**) Population responses of OFC neurons that encode expected reward values at the time of AR in the reward discrimination task. (**G**) Average activity of the cells in (F). Differential responses between large and small rewards within two blocks were plotted in the inset (*$p <$ 0.001; *t*-test). Shaded areas and error bars indicate SEM.

Since the location-outcome associations were switched between blocks in the reward discrimination task, the ability to reverse responding in accordance of outcome value was necessary. The OFC has been suggested to be critical to such reversal learning (Schoenbaum et al., 2003; Hornak et al., 2004; Izquierdo et al., 2004). In particular, previous studies found that OFC cells changed their selective responses to new cues that predicted better outcomes during reversal learning (Rolls et al., 1996; Schoenbaum et al., 1999). Thus, it was possible that different OFC cells other than DO, DE, and DT cells might encode expected reward values irrespective of spatial location at the time of arrival in front of the barriers. Among the OFC cells showing no task-related activity in the delay discounting task, 36 neurons were activated differently depending on the upcoming reward sizes in the reward discrimination task (Fig. 3.10F). A repeated measures revealed that these cells exhibited greater activity in anticipation of large rewards ($F_{(1,35)} = 27.39$, $p < 0.001$; Fig. 3.10G), but there were no effect of block ($F_{(1,35)} = 0.23$, $p = 0.64$) and no interaction between reward size and block ($F_{(1,35)} = 0.45$, $p = 0.51$). These results suggest that different groups of OFC cells are involving in signaling outcome expectancy in different behavioral tasks.

Task-dependent changes were not observed for reward-related OFC cells. As previously described in the delay discounting task, individual pRE, cRE, and post-RE cells consistently represented a specific reward condition (Fig. 3.11A-C). The cells which preferred SS reward showed greater responses to small rewards, whereas the cells preferring one of three LL reward conditions responded more to large rewards. The population responses of the three groups also signaled the magnitude of small and large rewards (Fig. 3.11D-F). ANOVAs with repeated measures demonstrated that the three groups of OFC cells exhibited significantly higher responses to LL reward than to SS reward (F values > 20.57, $p < 0.001$). However, there were no

differences between blocks (F values < 1.78, $p > 0.18$) and no interactions between reward size and block (F values < 0.73, $p > 0.39$).

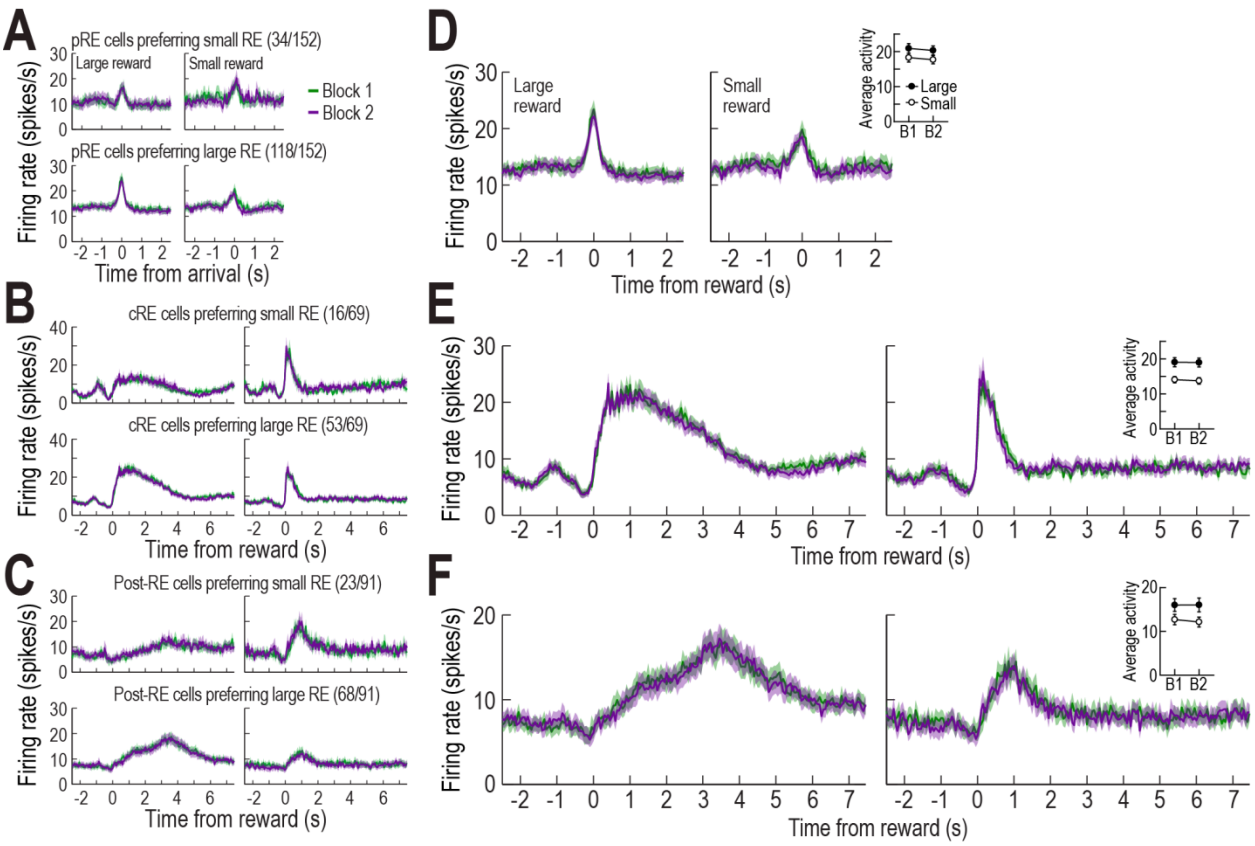

Figure 3.11. Activity of pRE, cRE, and post-RE cells in the reward discrimination task. (**A-C**) Population histograms (bin width, 50 ms) of pRE (A), cRE (B), and post-RE cells (C) that preferred either small or large reward condition. (**D-F**) Population activity of all pRE (D), cRE (E), and post-RE cells (F). Inset graphs indicate average responses. Shaded areas and error bars show SEM.

Discussion

It has been debated whether the OFC signals the relative value of outcomes or specific information about them in various task states (Padoa-Schioppa and Cai, 2011; Schoenbaum et al., 2011; Wallis, 2012). To address this issue, single-unit activity was investigated in the OFC of rats required to make a choice between differently delayed rewards on a T-maze. Different groups of OFC cells showed excited responses to a series of task-relevant events and periods, such as DO, waiting period, DT, reward, or post-reward period after consumption. Individual neurons responding in each state signaled a preferred reward condition by firing stronger than in the other conditions. However, the population activity of these cells reflected outcome values evaluated in each state of the task. Specifically, when future rewards were estimated at the time of DO and during delays, population responses represented the relative value of the differently delayed rewards. After the waiting periods, when the rewards were evaluated shortly before (i.e., at the time of DT), during, and shortly after reward consumption, population activity of other OFC cells encoded the magnitude of the delayed rewards, irrespective of different delays preceding them. These results provide compelling evidence that the OFC represents both specific outcomes and their relative values at the individual and population levels, respectively. The value signals within each state seem to be modulated by changing the proportion of individual cells preferring a more valuable outcome relative to other cells preferring less valuable options.

The two signals in the OFC were not always computed in every state. As DO, delay, and DT became irrelevant because of immediate access to rewards in the reward discrimination task, OFC cells responding in the three states of the delay discounting task no longer encoded the information about specific outcomes and expected reward values. Instead, these neurons were non-responses in the task states or robustly modulated by the spatial location of chosen goal

arms. Such spatial selectivity was also reported in previous studies (Feierstein et al., 2006; Roesch et al., 2006). Given that individual OFC cells encoded specific information about expected outcomes, it appears likely that the location-selective cells may represent the specific features of the goals arms associated with SS and LL rewards in the other task, rather than pure spatial information *per se* (Padoa-Schioppa and Cai, 2011). Consequently, population activity of DO, DE, and DT cells did not represent value information in the reward discrimination task. This finding suggests that the OFC computes subjective values only in the task states relevant to current decision making processes.

Chapter 4: A role for the medial reticular formation in delay-based decision making

Introduction

The reticular formation (RF) consists of a variety of functionally different, but net-like interconnected nuclei that extend rostrally from the medulla, through the pons, to the midbrain. The RF has been implicated in coordinating motor activity (Siegel and McGinty, 1977; Fabre et al., 1983; Peterson, 1984), modulating transitions between sleep and wakefulness (Moruzzi and Magoun, 1949; Steriade and McCarley, 1990), and regulating arousal, vigilance, and attention states (Pragay et al., 1978; Kinomura et al., 1996). In particular, the midbrain portion of the RF (MRF) has been suggested to signal elevated motivation in anticipation of positively reinforcing events in rodents (Olds et al., 1969; Phillips and Olds, 1969) and primates (Pragay et al., 1978; Ray et al., 1982). For example, Phillips and Olds (1969) reported that individual neurons in the MRF continuously increased their firing while rats deprived of food, but not water, were waiting for impending food rewards. The amplitude of such anticipatory responses was low when less desirable water was expected. Conversely, after the same rats were later water-deprived with free access to food, the elevated neuronal responses during waiting periods were evident in expectation of water, but not food.

The presence of MRF neurons that signaled reward expectancies during waiting periods raises an important question of whether the MRF contributes to delay-based decision making. To answer this idea, single units were recorded from the MRF of rats performing a delay discounting task in which they were required to choose between sooner small (SS) and later large (LL) rewards. Consistent with previous findings (Olds et al., 1969; Phillips and Olds, 1969), a substantial number of MRF neurons increased their firing during delays in expectation of

upcoming rewards. It was investigated how their delay-excited responses were influenced by delay length and reward size. Additionally, muscimol, a GABA receptor agonist, was infused into the MRF to determine whether MRF inactivation altered behavioral performance in the delay discounting task.

Materials and methods

Subjects

Fifteen male Long-Evans rats (340-400 g; Simonson Labs, Gilroy, CA) were housed individually in Plexiglas cages and were kept on a 12 hour light/dark cycle (lights on at 7:00 am). Each rat was maintained on a restricted diet at 85 % of its free-feeding weight with water freely available. All animal care and experiments were conducted during the light phase, in accordance with the University of Washington's Institutional Animal Care and Use Committee guidelines.

Behavioral apparatus

An elevated T-maze (79 cm from the floor) was used throughout the experiments. The maze was made of black Plexiglas and was composed of one start (the middle stem) and two goal arms (58 × 5.5 cm each). There was a metal food cup at the end of each goal arm. A black wooden barrier was placed before the food cup to control animals' access to reward during various lengths of waiting periods. Each maze arm was hinged such that its proximal end closest to the maze center could be raised and lowered by remote control. The maze was encircled by black curtains that were decorated with spatial cues.

Presurgical training

All rats were acclimated to the T-maze for 3-5 days. During the habituation phase, they were allowed to freely forage for chocolate milk drops that were randomly scattered on three maze

arms. Then, they were shaped to collect a shortly delayed reward (0.15 ml) only from the goal arms. Specifically, each rat was placed on the start arm in a given trial and was encouraged to choose one of the goal arms. Upon arrival at the barrier, the animals had to wait for 3 s before acquiring reward. The elapsed time was measured by an experimenter using a digital stopwatch. As the barrier was removed by the experimenter at the termination of the delay, the rat could approach and consume the reward. After replacing the barrier and then re-baiting the food cup, the experimenter gently guided the animal to the start arm for the next trial. Once the rat was able to finish 16 trials within 20 min, it underwent the surgical implantation of recording electrodes or guide cannulae.

Surgery

Under anesthesia with isoflurane (4% mix with oxygen at a flow rate of 1L/min), rats were mounted on a stereotaxic instrument (David Kopf Instruments, Tujunga, CA). Subsequently the isoflurane concentration was reduced to 1-3%. The skull was exposed and adjusted to place bregma and lambda on the same horizontal plane. Five rats had six individually drivable tetrodes chronically implanted in the right hemisphere dorsal to the MRF (5.2 mm posterior to bregma, 1.3 mm lateral to midline, and 5.4 mm ventral to the brain surface). Another two rats were implanted with a single drivable bundle loaded with six tetrodes in the same area. Each tetrode was made by twisting four 20μm lacquer-coated tungsten wires (California Fine Wire, Grover Beach, CA) and its tips were plated with gold to a final impedance of 0.2-0.4 MΩ (tested at 1 kHz). The remaining eight rats received bilateral implantation of guide cannulae (25 gauge) aimed at the MRF (5.2 mm anterior, 1.2 mm lateral, and 5.8 mm ventral to bregma). A 33 gauge dummy cannula was inserted into each guide cannula to prevent clogging.

Delay discounting and reward discrimination tasks

After a week of recovery, all rats were put back on a food-restricted diet. Three separate groups of animals performed different decision making tasks on the maze. The first group of 5 rats implanted with recording tetrodes were trained in a delay discounting task in which they should choose between a sooner small (SS) reward and a later large (LL) reward. To assess choice performance as a function of delay to LL reward, three different lengths of delay (10, 20, and 40 s) prior to LL reward (0.3 ml) were tested in separate blocks of trials. However, the delay to SS reward (0.05 ml) remained constant at 3 s throughout the experiments. In a daily testing session, the three delays before LL reward were randomly assigned to different blocks of trials and only one delay was used in a given block. Since the rats did not initially know how long they needed to wait for LL reward, each block began with 10 forced-choice trials followed by 6 or 8 free-choice trials. During the forced-choice trials, five SS and five LL reward trials were pseudorandomly ordered and only one goal arm was presented in a given trial after lowering the other goal arm. During the free-choice trials, both goal arms were made available and their choice preference for LL reward was measured. A choice was considered to be made, when the entire body (excluding the tail) entered a goal arm. To prevent the choice reversal during longer delays, an additional barrier was located at the entrance of the LL reward arm after the rats entered the goal. The three blocks of trials were separated by an interblock interval of 5-10 min, during which they were placed on a holding area adjacent to the T-maze. The spatial location of SS and LL rewards in the maze was held constant within each rat, but was counterbalanced across rats.

To further examine whether neural activity in the MRF was influenced by the magnitude of expected rewards when delays to the rewards were held constant, the second group of two rats implanted with a tetrode bundle was trained in a reward discrimination task in which the animals

were required to discriminate two goal arms baited with a small (0.05 ml) and a large (0.3 ml) reward. The testing procedures were identical to the delay discounting task except that both small and large rewards were equally delayed by 5 or 10 s. One of the two delays was randomly used in the first block and the other delay was tested in the second block. The location of small and large rewards was randomly selected in each day. Each block consisted of 10 forced-choice (5 small and 5 large rewards) and 10 free-choice trials.

The last group of eight rats implanted with bilateral cannulae into the MRF was initially trained in the delay discounting task. Once the rats showed similar delay-discounting performance as the first group, the MRF was manipulated with SAL or MUS injections on four different days. In each drug testing session, four different delays (3, 10, 20, and 40 s) were imposed prior to LL reward in separate blocks of trials. The 3-s delay was included to examine whether MRF inactivation altered the ability to discriminate large and small rewards that were equally delayed. In order to limit the number of drug injections and minimize possible effects of the presentation order of the four delays, each drug was tested on two consecutive days with either ascending (3 to 40 s) or descending sequences (40 to 3 s). These behavioral data were combined into single means. The order of drug injections and two delay sequences were counterbalanced across rats.

Intracranial microinjection

MUS (1 μg/μl dissolved in saline) was used to temporarily inactivate the MRF (Jo and Lee, 2010). A 33gauge injection cannula extending 1 mm below the tip of the guide cannula was connected to a 10 μl syringe (Hamilton, Reno, NV) via polyethylene tubing (PE 20). Either 0.3 μl of MUS or SAL was bilaterally infused at 10 μl/h rate using a microinfusion pump (KD Scientific, Holliston, MA). The injection cannula was left in place for an additional 1 min to

allow diffusion of the drugs from its tip. After the drug injections, rats were carefully observed for any behavioral abnormalities in their home cages for 20 min before they were placed on the maze.

Single-unit recording

Neural activity was monitored prior to each recording session while rats were located in the holding area (Norton et al., 2011; Jo et al., 2013). Recording tetrodes were connected to a preamplifier and neural data were transferred to a Cheetah data acquisition system (Neuralynx). Unit signals were digitalized at 16 kHz, amplified 500-6000 times, and filtered between 0.6 and 6 kHz. Neuronal spikes were acquired for a 2 ms sampling period whenever a voltage deflection from any tetrode channel exceeded a user-defined threshold. Two LEDs were attached to the preamplifier. The LED signals were captured by a camera mounted on the ceiling at a sampling rate of 30 Hz and subsequent position data were fed to the acquisition system. Once clearly isolated and stable units were found, a daily recording session started. Acceptable signals were at least twice the amplitude as the background activity. While the rats were performing the task, three salient events including delay onset, delay termination, and reward were recorded in parallel with neural activity. Specifically, timestamps for delay onset and termination were fed into the data stream when an experimenter operated the stopwatch to measure elapsed time during waiting periods. Reward events were timestamped by 'lick-detectors' (custom designed by Neuralynx) when the animals first licked chocolate milk in the food cups. At the end of each session, tetrodes were advanced in 40 μm increments, up to 160 μm per day to find new cells. These recording procedures continued until, according to the distance traveled, tetrodes passed through the MRF.

Histology

After completion of all experiments, the final position of each tetrode was marked by electrolytic lesions (15 μA current for 12 s) while all rats were anesthetized under isoflurane. Afterward, the animals were perfused transcardially with physiological saline followed by 10% formalin. Their brains were extracted and stored in a 10% formalin-30% sucrose solution at 4°C for 72 h. The brains were cut in coronal sections (40 μm) on a freezing microtome. The serial sections were stained with cresyl violet. Tetrode tracks and marker lesions were identified using photomicrographs taken under digital microscopy. Only cells verified to be recorded in the MRF were included in the data analysis. Cannula placements were also verified in the same way except for the absence of electrolytic lesions.

Data analysis

Unit isolation was made using an Offline Sorter (Plexon, Dallas, TX). Various waveform features such as peak, valley, width, principle component, and energy were compared across simultaneously recorded units from four tetrode wires. Only units showing good recording stability across blocks were included. Further analysis of sorted units and statistics were performed with Matlab software (Mathworks, Natick, MA).

To examine delay-dependent changes in firing, spike rates during the entire delay of each trial were converted to z-scores relative to the mean firing within each block of all trials. A MRF cell was classified as delay-excited or delay-inhibited if it passed the following two criteria: 1) its average activity during at least one of the delays was greater or less than a z-score of 2, respectively, and 2) no such increased or decreased activity was not observed during the 2.5 s window prior to the delay onset. In addition, reward responses were investigated using peri-event time histograms (PETHs; bin width, 50 ms) that were constructed for the 5 s period around reward encounters. A MRF cell was categorized as reward-responsive if it had a peak firing rate

during a 300 ms epoch after obtaining reward (-50 to 250 ms) and the average activity was ≥ 200% of its mean firing for the block of all trials. For the three groups of MRF cells, spearman's rack correlation coefficient was calculated in each time bin of PETHs in order to test whether their activity changed in a linear fashion as the delay to LL reward increased. The significance of correlation was estimated using a permutation test in which firing rates of each bin were randomly shuffled across blocks for 1000 times. A confidence interval of $p < 0.99$ was calculated from correlation coefficients of the shuffled data.

The rats' movement during performance of the tasks was analyzed by calculating the distance between two consecutive head positions sampled at 30 Hz (instantaneous velocity). The spatial firing rate maps of individual cells were also depicted by dividing the number of spikes with the total time spent in each pixel (2.5 cm × 2.5 cm) of position data.

Statistical analysis

Differential firing of individual neurons across different reward conditions was analyzed with *t*-tests and repeated-measures ANOVAs followed by Bonferroni's *post hoc* pairwise comparisons. Pearson's correlation tests were performed to establish a relationship between two variables. Two-tailed p values < 0.05 were considered statistically significant. Data are expressed as mean ± SEM.

Results

Choice behavior in the delay discounting task

Five rats were trained to choose between SS and LL rewards on an elevated T-maze (Fig. 4.1A). Three different delays to LL reward (10, 20, and 40 s) were randomly ordered and tested in separate blocks of trials to examine the animals' decision making performance as a function of

delay to LL reward (Fig. 4.1B). The delay to SS reward (3 s) was constant throughout the task. In each testing block, 10 forced-choice trials (5 SS and 5 LL rewards) were given to inform rats as to which delay was associated with LL reward, and then 6 or 8 free-choice trials were tested to assess their choice behavior. Each delay started as the rat arrived in front of the wooden barrier that prevented immediate access to reward. An experimenter manually removed the barrier at the end of the delay while monitoring the elapsed time using a digital stopwatch, which caused slight variations in delay length. During all 41 behavioral recording sessions, SS reward was delayed by 3.19 ± 0.39 s (mean ± SD) and LL reward was delayed by 10.43 ± 0.49 s, 20.44 ± 0.79 s, or 40.35 ± 0.95 s. During these delays, the animals mostly engaged the barrier with minimal movement by sniffing and biting it as well as rearing against it (Fig. 4.1D). Overall, average velocity during the entire delays before LL reward was lower than that during the entire delays prior to SS reward (repeated measures ANOVA, $F_{(1,4)} = 9.51$, $p = 0.037$), whereas there were no differences in velocity across three delays within the same reward conditions ($F_{(2,8)} = 1.24$, $p = 0.34$) and no interaction between reward size and block ($F_{(2,8)} = 0.14$, $p = 0.87$).

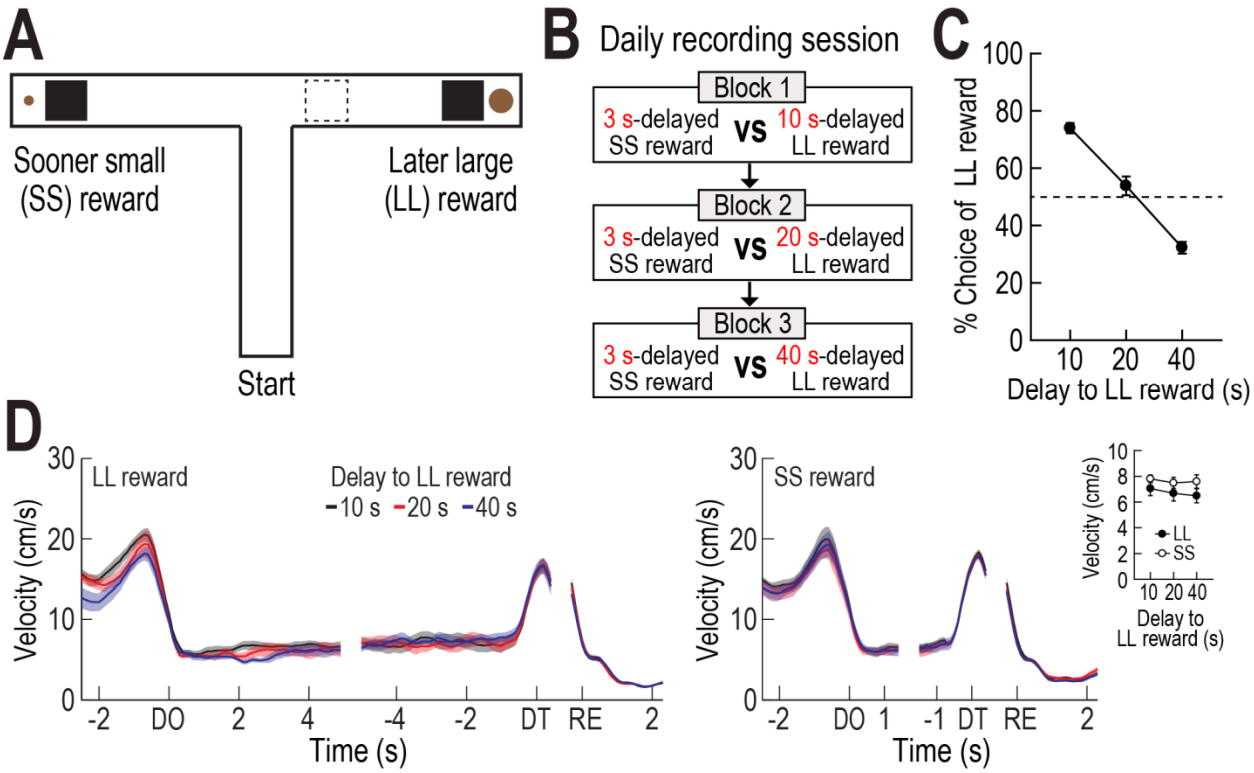

Figure 4.1. Choice performance on a delay-based decision making task. (**A**) Illustration of the T-maze. LL and SS rewards were baited at the end of two opposite goal arms. A rectangular wooden barrier (black square) was placed before each food cup to control animals' access to rewards. When rats chose a goal arm associated with LL reward, an additional barrier (indicated by the dashed rectangle) was placed at its entrance to prevent the animals from exiting the goal arm during the delay. (**B**) Daily experimental procedures. Three different lengths of delay to LL reward were randomly ordered and tested in separate blocks of trials. The delay to SS reward remained unchanged. Each block consisted of forced-choice trials, followed by free-choice trials. (**C**) Choice preference for LL reward as a function of delay to LL reward. (**D**) Changes in instantaneous velocity around waiting periods. Data are aligned on delay onset (DO), delay termination (DT), and reward (RE). The inset plot shows average velocities during the entire waiting periods. Shaded areas and error bars indicate SEM.

When a 10-s delay was imposed before LL reward, animals showed a strong preference for the LL reward (73.9%; Fig. 4.1C). The preference was weakened to near chance levels (53.8%) as the LL delay was extended to 20 s. Finally, a 40-s delay to LL reward reversed choice behavior (33.2%), such that rats chose the SS reward more often. A Pearson's correlation test revealed a significant negative correlation between choice performance and delay to LL reward ($r = -0.96$, $p < 0.001$). This result indicated that rats discounted delayed rewards.

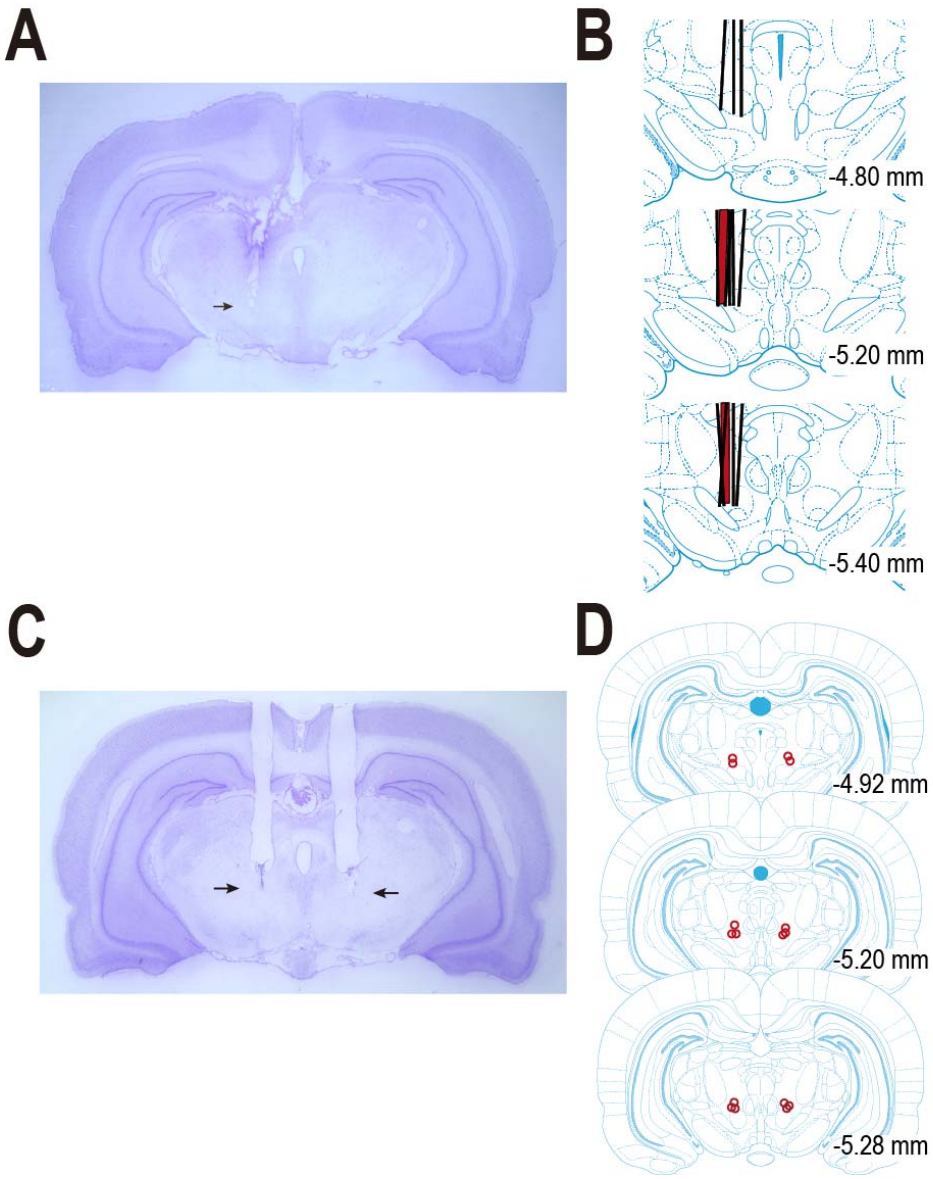

Figure 4.2. Histological verification of tetrode locations and cannula positions. (**A**) Nissl-stained section showing the final location of a tetrode tip in the MRF (indicated by the arrows). (**B**) Reconstruction of all tetrode tracks. MRF neurons in the delay discounting and reward discrimination tasks were recorded from the tetrodes in black and red, respectively. (**C**) Representative photomicrograph for bilateral cannula placements in the MRF. The arrows indicate the tips of injection cannulae. (**D**) Illustration of all microinjection sites.

Delay-excited activity in the delay discounting task

While rats performed the task, a total of 348 cells were recorded from electrodes located in the MRF (Fig. 4.2A, B). Their basal firing rates ranged from 0.3 to 95.74 Hz. Of these MRF cells, 117 cells (33.6%) were significantly excited during at least one of four different lengths of delay relative to their mean firing within the corresponding blocks of trials. For example, a representative delay-excited cell gradually ramped up its firing during delays and reached its peak around the end of the delays when the barriers were removed (Fig. 4.3A). Then the ramping activity quickly declined as the rat approached the food cup to obtain reward. Delay-excited responses were not attributable to delay-dependent changes in behavior such as the decreased velocity during waiting periods, since average velocities during the four different delays were not significantly correlated with average delay-excited firing rates normalized to the mean firing of the corresponding blocks (Pearson's correlation, absolute r values < 0.13, p values > 0.15; Fig. 4.3C). Instead, delay-excited activity was modulated by the discounted value of delayed rewards (Fig. 4.3B). Specifically, delay-excited activity prior to LL reward decreased with longer delays. Individual delay-excited cells exhibited significant inverse correlations (Spearman's rank correlation per each time bin) across three lengths of delay to LL reward, especially during the first 5-s period. Such inverse correlations were not evident during the delays before SS reward except for a short period (200 ms) in the middle of the delays.

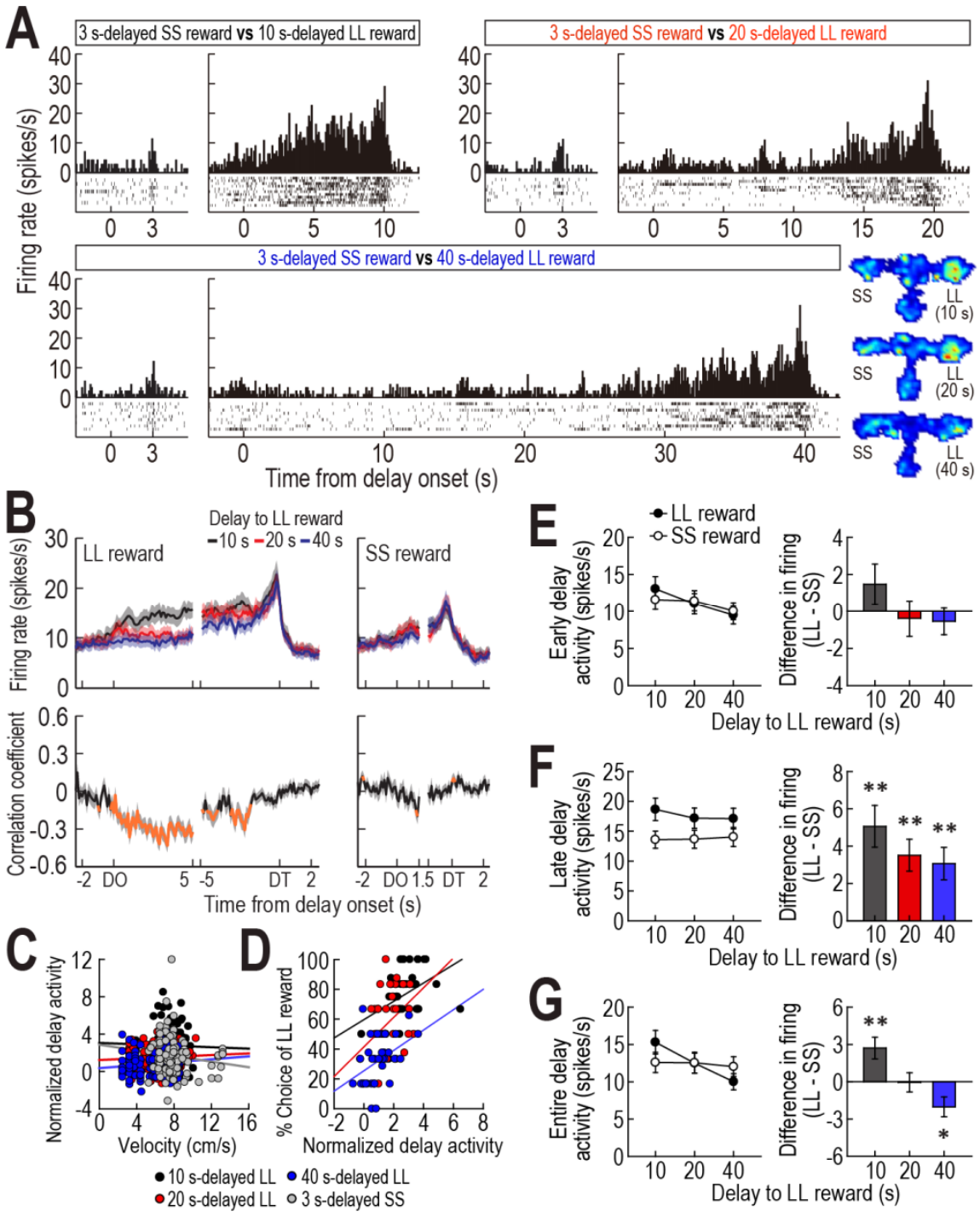

Figure 4.3. Delay-excited activity in the MRF. (**A**) A representative delay-excited cell. All trials in the histograms (bin width, 100 ms) were aligned to delay onset. Spatial firing rate maps in three blocks of trials showed that the delay-excited cell mainly fired during delays preceding LL reward. (**B**) Population responses of all delay-excited cells. Data are aligned on delay onset and

termination. Correlation coefficients for individual neuronal responses across three blocks of trials were calculated per each time bin. Orange data points that fell outside the 99% confidence interval obtained from a permutation test for at least two consecutive bins were considered significantly correlated. (**C**) Correlations between delay-excited activity and average velocity during delays. Delay-excited activity was normalized to the mean firing rate within each block of trials. (**D**) Correlations between normalized delay-excited activity and choice preference for LL reward. (**E**) Comparison of the early delay-excited activity that was measured in the first 1.5 s of delays, using the population histograms in (B). Differential firing rates between LL and SS reward trials within blocks were measured across individual neurons. (**F**) Late delay-excited responses during the last 1.5 s of delays. (**G**) Delay-excited activity during the entire delays. Shaded areas and error bars represent SEM. *$p < 0.05$, **$p < 0.01$; *t*-test.

In addition, comparisons of delay-excited activity between the two reward options within blocks also supported the view of reward value coding by delay-excited cells. During an early phase of waiting periods (the 1.5-s period after the delay onset), these cells showed no differential firing between SS and LL rewards (Fig. 4.3E). A repeated measures ANOVA revealed no significant effect of reward size ($F_{(1,116)} = 0.05$, $p = 0.82$), whereas a significant effect of block ($F_{(2,232)} = 10.69$, $p < 0.001$) and a significant interaction between reward size and block ($F_{(2,232)} = 3.28$, $p < 0.001$) were observed due to the decreased delay-excited activity as a function of delay to LL reward. However, delay-excited cells displayed different levels of peak firing depending on the magnitude of upcoming rewards during a late phase of delays (the 1.5-s period before the delay termination; Fig, 4.3F). A repeated measures ANOVA confirmed that the final ramping activity prior to LL reward was significantly higher than that before SS reward ($F_{(1,116)} = 24.99$, $p < 0.001$), but the final ramping responses within the same reward conditions were not different as indicated no significant effect of block ($F_{(2,232)} = 0.76$, $p = 0.47$) and no interaction between the factors ($F_{(2,232)} = 2.39$, $p = 0.09$).

When average delay-excited activity was measured for the entire delays (Fig. 4.3G), an ANOVA with repeated measures found a significant effect of block ($F_{(2,232)} = 14.28$, $p < 0.001$) and a significant interaction ($F_{(2,232)} = 17.16$, $p < 0.001$), without an effect of reward size ($F_{(1,116)} = 0.1$, $p = 0.75$). Interestingly, *post hoc* comparisons revealed that delay-excited activity was significantly greater in the LL reward condition than in the alternative SS reward condition ($p = 0.002$) in the blocks where rats preferred a 10 s-delayed LL reward. Delay-excited responses between SS and LL reward conditions were not different ($p = 0.95$) as the animals became indifferent between a SS and a 20 s-delayed LL reward. Delay-excited activity was significantly lower in the LL reward condition ($p = 0.01$) when behavioral choices were biased toward SS

reward. These firing patterns indicated that the average delay-excited activity during the entire delays reflected the relative value of expected rewards. Moreover, a linear relationship between delay-excited activity and behavioral performance was determined using normalized delay-excited firing during the entire delays. If multiple delay-excited cells existed in a given recording session, their average firing was used for this analysis. Pearson's correlation tests demonstrated that the more elevated the delay-excited cells were relative to their mean firing during all three delays to LL reward, the more strongly the rats preferred LL reward (r values > 0.39, p values < 0.05; Fig. 4.3D).

Delay-inhibited activity in the delay discounting task

A different group of MRF cells (41/348, 11.8%) was significantly inhibited during at least one of four delays. A representative cell exhibited delay-specific inhibition, followed by short-lasting excitation when rats approached the food cup (Fig. 4.4A). These firing patterns closely resembled the changes in velocity around the time of waiting periods (Fig. 4.1D). Indeed, normalized delay-inhibited responses were significantly correlated with average velocities during the three delays prior to LL reward (Pearson's correlation, r values > 0.34, p values < 0.05; Fig. 4.4C), even though such a positive correlation was marginally significant during the delay to SS reward ($r = 0.29$, $p = 0.058$). As seen from the population activity (Fig. 4.4B), delay-inhibited cells showed no differential firing across blocks within the same reward conditions, but the average firing rates in in LL reward conditions were lower than those in SS reward conditions. A repeated measures ANOVA demonstrated a significant effect of reward size ($F_{(1,40)} = 8.79$, $p = 0.005$; Fig. 4.4E), but no effect of block ($F_{(2,80)} = 0.26$, $p = 0.77$) and no interaction between the factors ($F_{(2,80)} = 0.71$, $p = 0.49$). The greater inhibition in LL reward condition was also in line

with the significantly lower velocity during waiting periods for LL reward than for SS reward (Fig. 4.1D, inset). Thus, these results indicated that delay-inhibited cells represented the animals' movement in the task. It appeared that the delay-inhibited cells had no direct relationship with behavioral performance, since any significant correlations were not observed between delay-inhibited responses and choice preference for LL reward in three blocks of trials (absolute r values < 0.35, p values > 0.09; Fig. 4.4D),

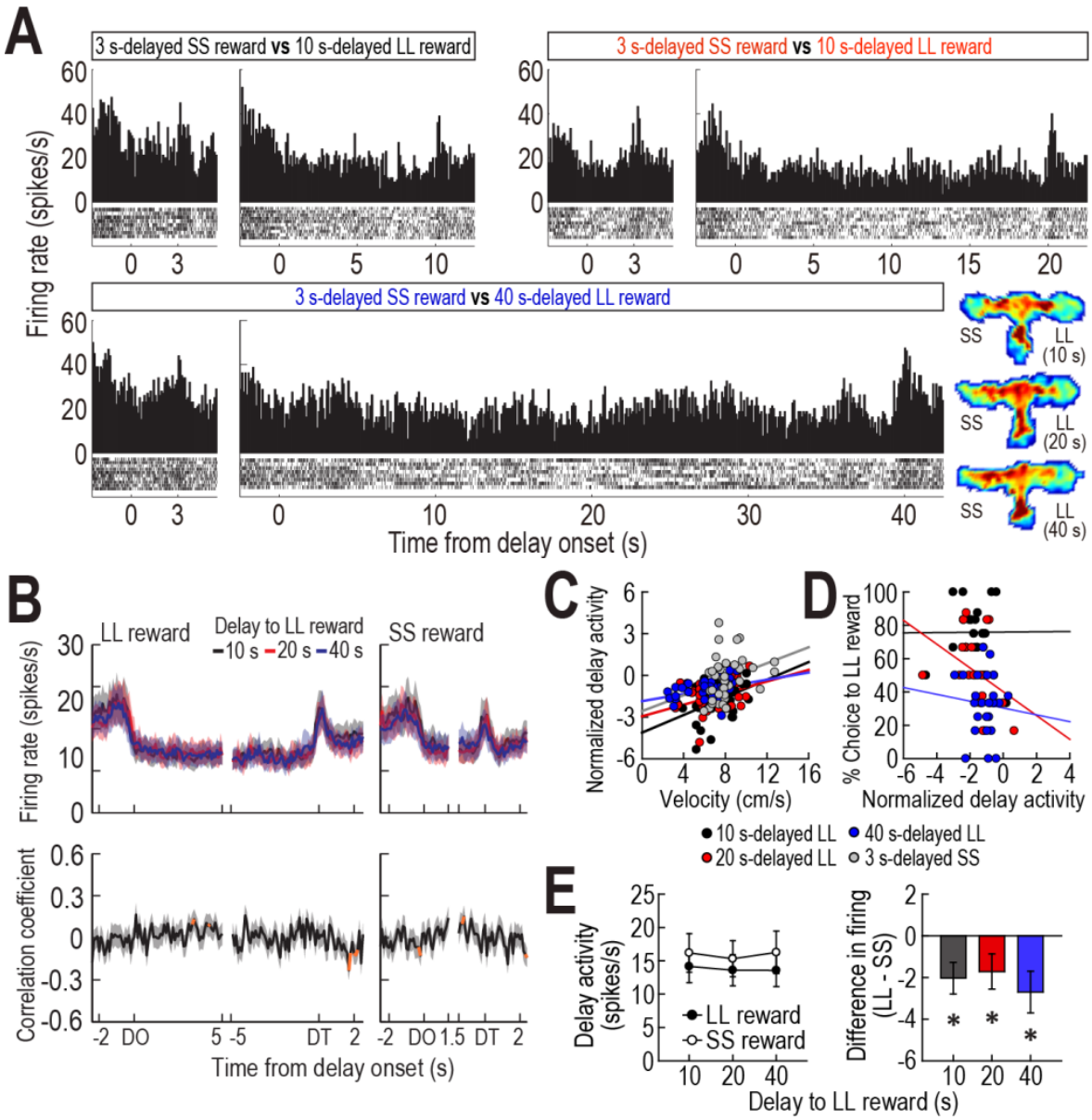

Figure 4.4. Delay-inhibited activity in the MRF. (**A**) A representative delay-inhibited cell in three blocks of trials. (**B**) Population histograms (bin width, 100 ms) of all delay-inhibited cells and correlation coefficients over the course of delays. Significant correlation coefficients for more than two consecutive bins were depicted in orange. (**C**) Correlations between normalized delay-inhibited response and average velocity during delays. (**D**) Correlations between normalized delay-inhibited activity and choice preference for LL reward. (**E**) Delay-excited activity during the entire delays and differential responses between LL and SS reward trials within blocks (*$p < 0.05$; *t*-test). Shaded areas and error bars show SEM.

Reward activity in the delay discounting task

It has been reported that MRF cells exhibit phasic responses to reward (Puryear and Mizumori, 2008). In the current study, 39 cells (11.2%) were briefly excited at the time of reward encounters. As seen from a representative neuron and population responses (Fig. 4.5A, B), reward responses were modulated by the magnitude of rewards regardless of different delays preceding them. An ANOVA with repeated measures revealed a significant effect of reward size ($F_{(1,38)} = 9.08$, $p = 0.005$; Fig. 4.5C), whereas no effect of block ($F_{(2,76)} = 0.26$, $p = 0.78$) and no interaction between the variables ($F_{(2,76)} = 1.19$, $p = 0.31$) were found. These results suggest that reward-responsive cells encode the value of obtained rewards.

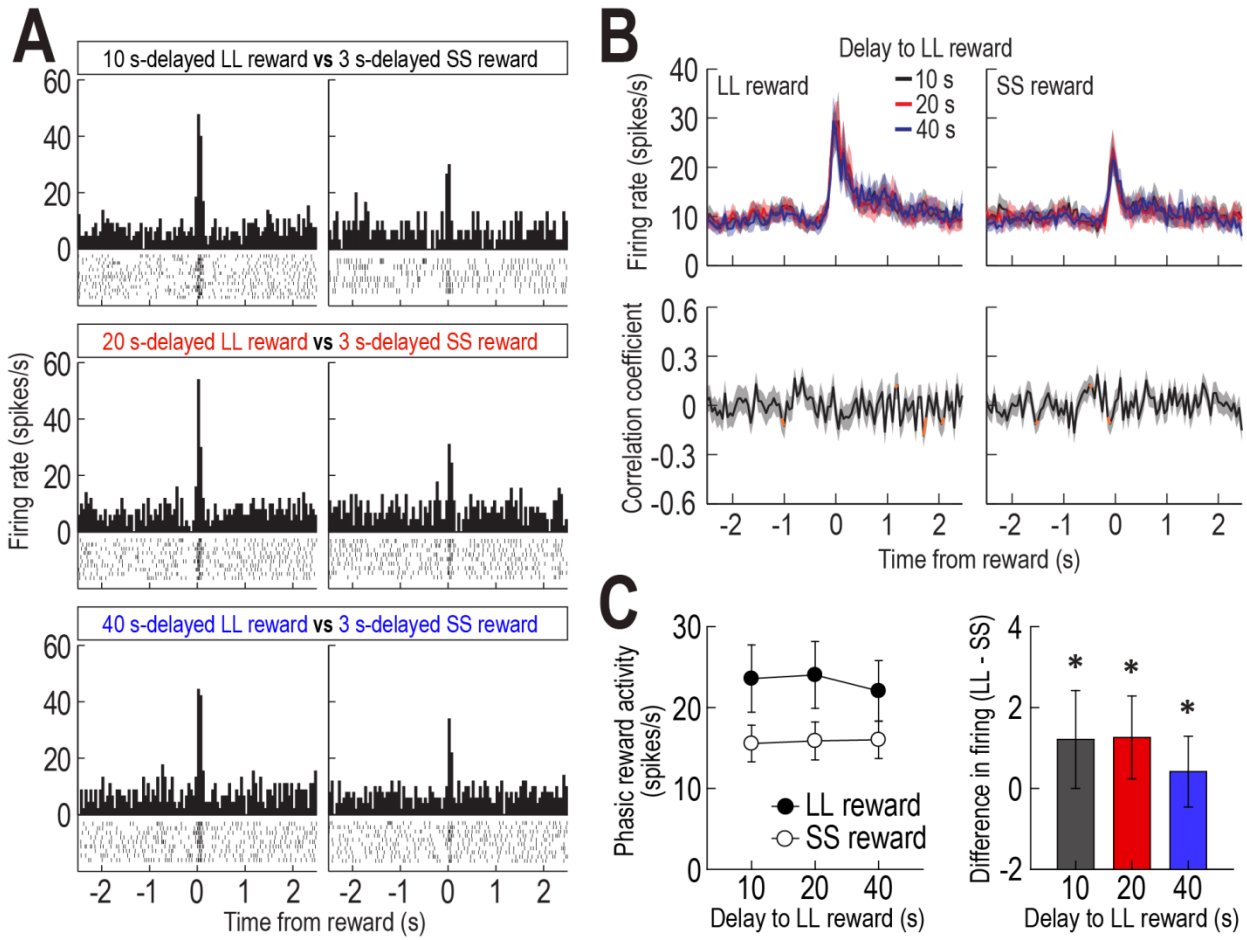

Figure 4.5. Reward activity in the MRF. (**A**) A representative cell showing reward responses (bin width, 50 ms) in three blocks of trials. (**B**) Population reward responses and correlation coefficients around the time of obtaining rewards. Orange data points indicate significant correlation coefficients (**C**) Average reward activity and differential responses between LL and SS rewards within blocks (*$p < 0.05$; *t*-test). Shaded areas and error bars indicate SEM.

Responses of MRF cells in the reward discrimination task

To further investigate the impact of reward size on the three groups of MRF cells when delays were held constant, 2 additional rats implanted with a single bundle of multiple tetrodes (Fig. 4.2B) were trained in the reward discrimination task. They had to choose between large and small rewards that were equally delayed by 5 or 10 s in separate blocks of trials (Fig. 4.6A). During 33 behavioral recording sessions, the actual delays were 5.17 ± 0.17, and 10.18 ± 0.3 (mean ± SD). The rats showed strong choice biases toward large rewards in both blocks (Fig. 4.6B). A paired *t*-test found no statistical difference in behavioral performance between the blocks ($t_{(1)} = 1$, $p = 0.5$). The preference for large rewards was also reflected in the animals' movement. As shown in Figure 4.6C, they moved faster toward the end of the goal arm associated with large rewards. However, no differences in velocity were observed during waiting periods for both reward sizes as indicated by no effects of reward size ($F_{(1,1)} = 0.22$, $p = 0.72$) and block ($F_{(1,1)} = 0.71$, $p = 0.55$) as well as no interaction between the factors ($F_{(1,1)} = 9.04$, $p = 0.2$; Fig. 4.6D).

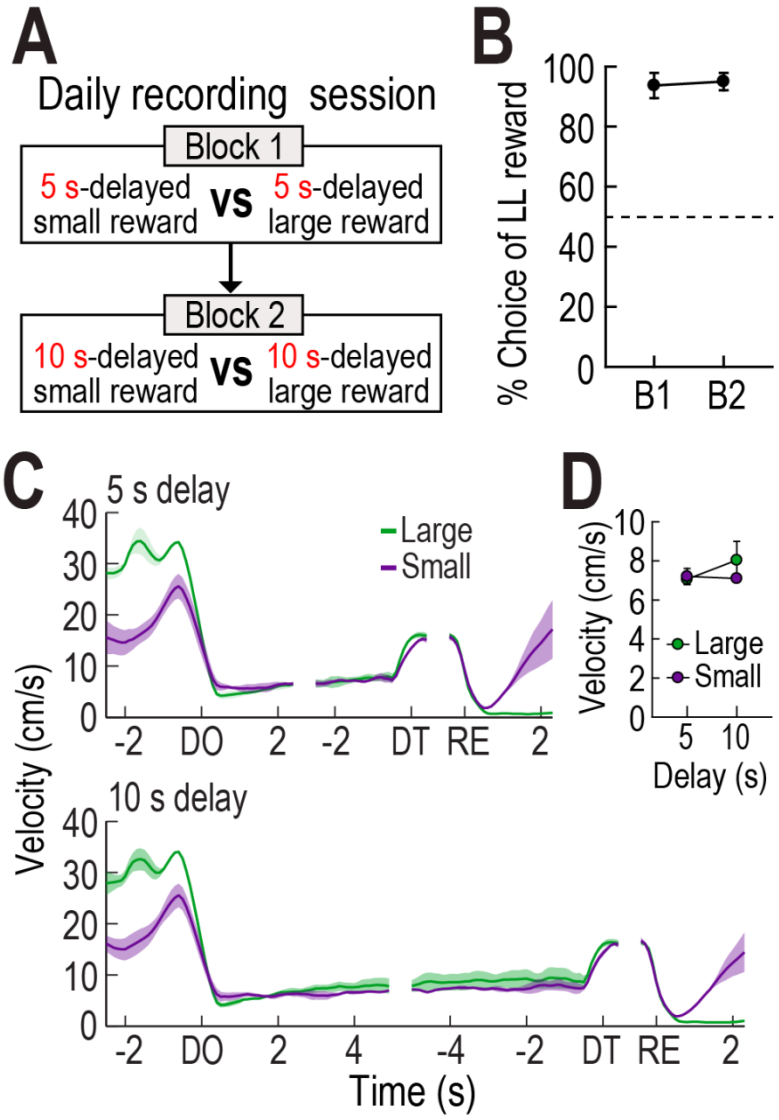

Figure 4.6. Behavioral performance in a reward discrimination task. (**A**) Daily experimental procedures. The goal arms associated small and large rewards were randomly selected on each day. The rewards were equally delayed, but two different lengths of delay were randomly used in separate blocks of trials. (**B**) Choice preference for LL reward in two blocks. (**C**) Velocity traces around waiting periods. Data are aligned on delay onset (DO), delay termination (DT), and reward (RE). (**D**) Average velocities during the entire waiting periods. Shaded areas and error bars show SEM.

Out of 350 cells recorded in the reward discrimination task, 79 delay-excited cells (22.6%) gradually increased their firing during waiting periods for both large and small rewards, but the magnitude of the ramping activity was strikingly different depending on expected reward values (Fig. 4.7A, B). A repeated measures ANOVA demonstrated that average delay-excited responses throughout the entire delays were significantly higher in anticipation of larger rewards ($F_{(1,78)} = 60.92$, $p < 0.001$; Fig. 4.7C). A significant effect of block ($F_{(1,78)} = 7.98$, $p = 0.006$) was also found without any interaction between the variables ($F_{(1,78)} = 1.06$, $p = 0.31$), which indicated that delay-excited cells were elevated more when the same amounts of reward were expected to be available sooner (i.e., 5 s) rather than later (i.e., 10 s). Interestingly, when delay-excited responses were analyzed during the late phase of delays (the 1.5-s period before the delay termination), delay-excited cells exhibited comparable levels of firing in the same reward conditions irrespective of delay lengths. This observation was confirmed by a significant effect only for reward size (repeated measures ANOVA, $F_{(1,78)} = 39.99$, $p < 0.001$; Fig. 4.7D), but not for block ($F_{(1,78)} = 0.09$, $p = 0.77$). There was no interaction between the factors. These results suggest that delay-excited cells in the MRF encode the information of expected rewards at the current moment over the course of delays.

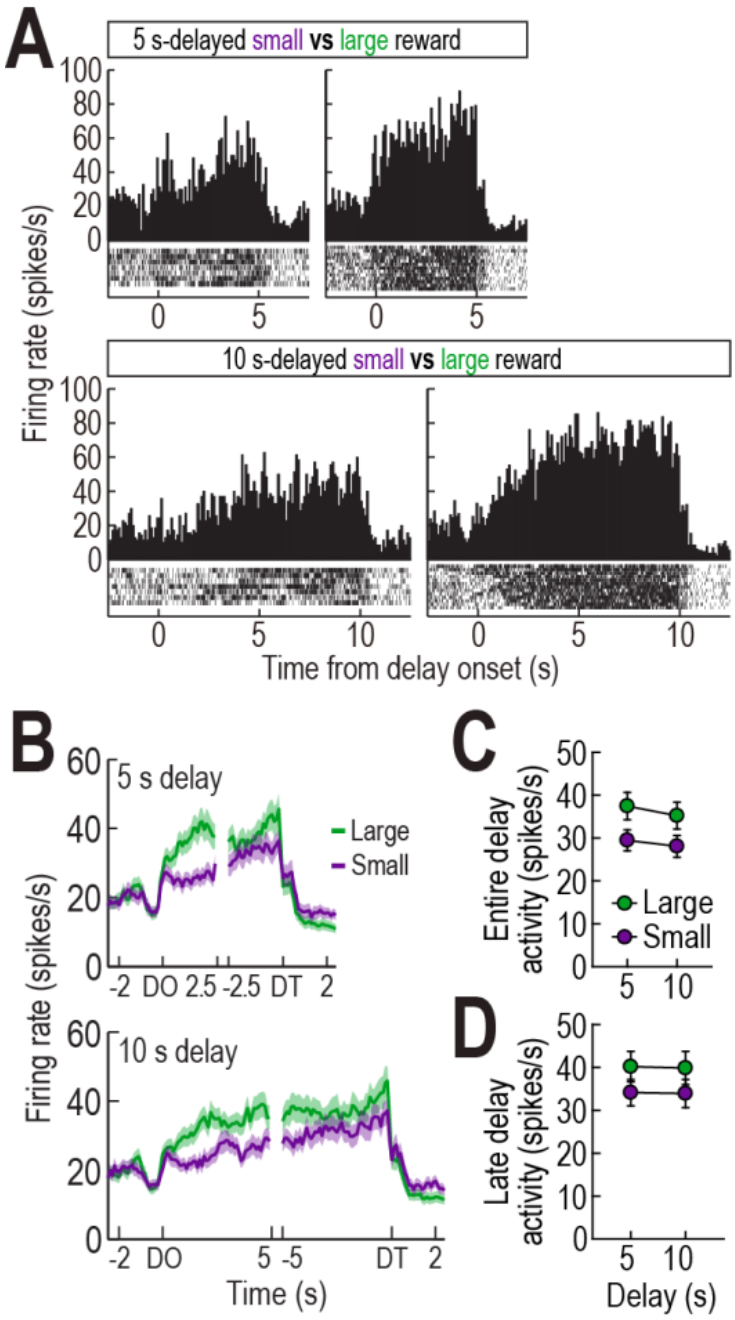

Figure 4.7. Delay-excited activity in the reward discrimination task. (**A**) A representative delay-excited cell exhibiting ramping activity during delays (bin width, 100 ms). (**B**) Population responses of all delay-excited cells. Histograms are aligned on delay onset (DO) and termination (DT). (**C**) Average delay-excited responses during the entire delays. (**D**) Late delay-excited activity during the last 1.5 s of delays. Shaded areas and error bars represent SEM.

Seventeen delay-inhibited (4.9%) and twelve reward-responsive cells (3.4%) were identified in the reward discrimination task. The former group of cells (Fig. 4.8A) continuously fired in parallel with the animals' velocity (Fig. 4.6C). While the rats were moving at different velocities before and after waiting periods, delay-inhibited cells showed differential responses. When the velocities decreased to similarly low levels during delays (Fig. 4.6D), there were no differences in delay-inhibited activity between blocks as well as between trials for large and small rewards. Accordingly, a repeated measures ANOVA comparing average responses during the entire delays found no effects of reward size ($F_{(1,16)} = 0.26$, $p = 0.62$; Fig. 4.8B) or block ($F_{(1,16)} = 1.18$, $p = 0.29$) with no interaction between the factors ($F_{(1,16)} = 0.03$, $p = 0.86$). Reward-responsive cells also consistently signaled different amounts of reward by showing higher phasic responses to larger reward (Fig. 4.8C). Since this group of cells was not influenced by delay length, no distinct changes in reward activity were observed between blocks. A repeated measures ANOVA demonstrated a significant effect of reward size ($F_{(1,11)} = 9.89$, $p = 0.009$; Fig. 4.8D), but no effect of block ($F_{(1,11)} = 0.62$, $p = 0.45$) and no interaction between the variables ($F_{(1,11)} = 0.62$, $p = 0.45$). These results corroborated the previous results that delay-inhibited cells represented movement information and reward-responsive cells signaled the value of encountered rewards.

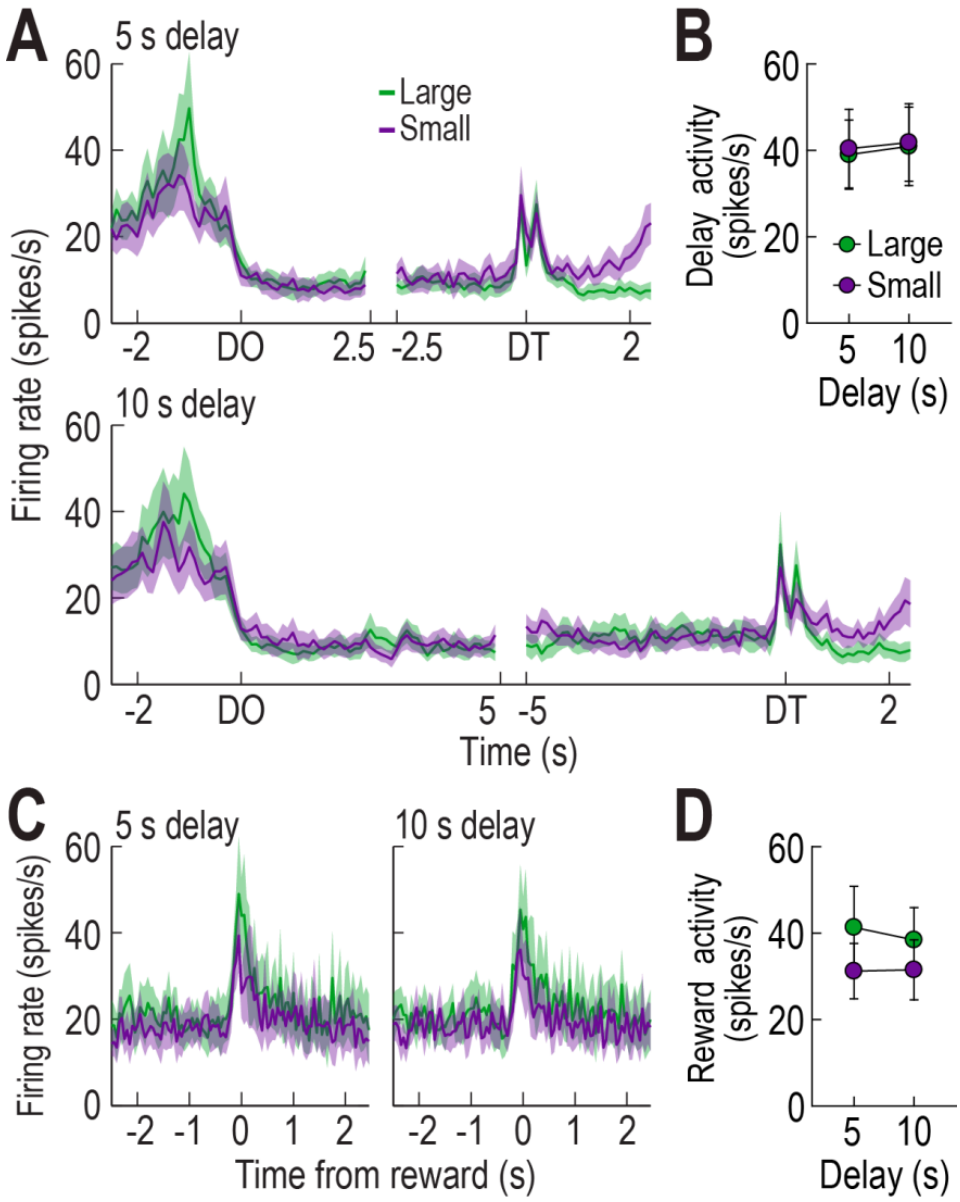

Figure 4.8. Delay-inhibited and reward-responsive cells in the reward discrimination task. (**A**) Population histograms (bin width, 100 ms) of all delay-inhibited cells. Firing rates are aligned on delay onset (DO) and termination (DT). (**B**) Average delay-inhibited responses during the entire lengths of delay (bin width, 50 ms). (**C**) Population responses of all reward-responsive cells. (**D**) Average reward responses in two blocks of trials. All graphs show mean ± SEM.

Effect of MRF inactivation on delay-discounting behavior

Because the current recording area contained a proportionally larger number of delay-excited cells that encoded the information about expected reward during delay, it was predicted that MRF inactivation would alter choice preference for LL reward due to the loss of reward expectancy signals. To address this hypothesis, 8 rats with bilateral cannulae aimed at the MRF (Fig. 4.2C, D) were tested in a modified version of the delay discounting task, in which a 3-s delay was included in addition to the original three delays to LL reward (Fig. 4.9A). SAL and MUS were used for MRF manipulations (Jo et al., 2007) and each drug was microinfused on two consecutive days to test choice behavior with both ascending and descending sequences of four delays prior to LL reward. These behavioral data were combined to minimize possible effects of presentation order.

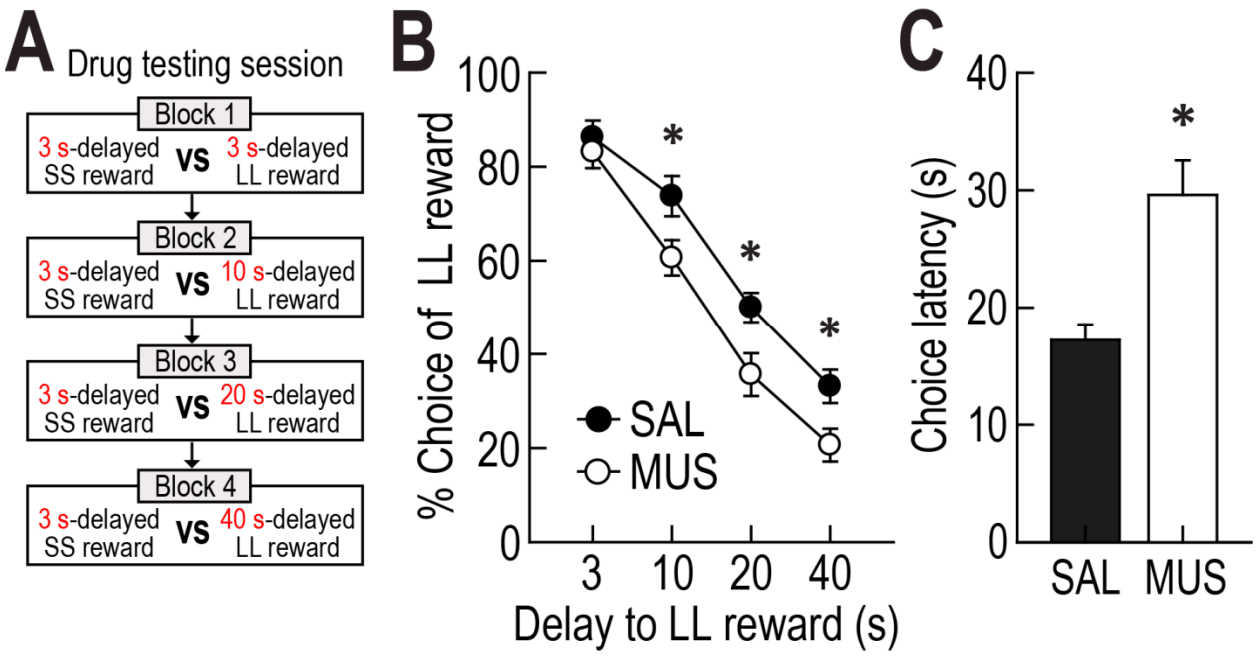

Figure 4.9. Effects of bilateral inactivation of the MRF on choice performance. (**A**) Drug testing procedures. Four different lengths of delay prior to large rewards were used in separate blocks of trials. Each drug was tested with both ascending and descending sequences of the four delays. (**B**) Behavioral performance after either SAL or MUS injection into the MRF. MRF inactivation significantly reduced choice preference for LL reward. (**C**) Average choice latencies from the start position to the entrance into chosen goal arms. MUS injection into the MRF significantly increased the choice latency, compared to SAL injection. All graphs show mean ± SEM.

All rats in both drug-injected conditions decreased their preference for LL reward as the delays to LL reward increased (Fig. 4.9B). More importantly, MUS injections into the MRF reduced the animals' choice biases toward LL reward compared to SAL injections. An ANOVA with repeated measures demonstrated significant effects of block ($F_{(3,21)} = 140.39$, $p < 0.001$) and drug ($F_{(1,7)} = 25.74$, $p = 0.001$), with no interaction between block and drug ($F_{(3,21)} = 2.35$, $p = 0.1$). Moreover, planned comparisons of drug effects within each block (Bonferroni's *t*-test) found that the preference for LL reward in MUS-injected conditions was significantly lower across all blocks (p values < 0.006) except for one block in which both SS and LL rewards were equally delayed by 3 s ($p = 0.55$). These results indicate that the MRF is critical for delay-based decision making, but not for discriminating the locations associated with different reward values.

It should be noted that MRF inactivation also affected other aspects of the animals' behavior on the maze. Specifically, when average latencies from leaving the start position to entering a chosen goal arm were measured on drug-testing days, MUS infusions significantly increased the times taken for the rats to make choices than SAL infusions (paired *t*-test, $t_{(7)} = 6.16$, $p < 0.001$; Fig. 4.9C). The increased choice latency appeared to result from alternations in movement-related activity by delay-inhibited cells and/or reductions in motivational value for delayed rewards after inactivating delay-excited cells. However, MUS-injected rats consumed all rewards available on food cups, indicating no effect of MRP inactivation on consummatory behavior.

Discussion

The RF has long been implicated in motor (Siegel and McGinty, 1977; Fabre et al., 1983; Peterson, 1984), cognitive (Pragay et al., 1978; Kinomura et al., 1996), and motivational

processes (Olds et al., 1969; Phillips and Olds, 1969; Puryear and Mizumori, 2008). Consistent with this view, three groups of MRF neurons were correlated with such functions in the present decision making tasks for variously delayed rewards. Delay-inhibited cells represented locomotor activity by tightly firing in parallel with the rats' velocity on the maze, while reward-responsive cells encoded the magnitude of obtained rewards. In addition, a proportionally larger number of delay-excited cells signaled information about expected rewards during waiting periods. Their ramping responses were initially elevated with different slopes in anticipation of differently delayed rewards. However, their peak firing at the end of the delays reached similar levels when equally sized rewards were expected regardless of the different delay lengths preceding the rewards. Average delay-excited activity during the entire delays signaled the discounted value of expected rewards. Accordingly, when delay-excited cells showed stronger responses during delays to LL reward, rats tended to choose the reward more often. These firing properties suggest that delay-excited cells encode motivational value of future reinforcing outcomes. In agreement with the electrophysiological results, bilateral inactivations of the MRF altered the animals' movement on the maze and decreased their choice preference for LL reward.

It is still possible that delay-excited cells reflected other aspects of behavior, emotion, or cognition that occurred during delays, instead of the current motivational value. For instance, the delay-excited activity might reflect an efference copy of the motor command for upcoming approach behavior after the termination of delays (Miles and Evarts, 1979). This idea predicts that neural representations regarding the motor planning are identical during the late phase of delays prior to both SS and LL rewards, since the rats needed to travel the same distance to the food cups regardless of the size of expected rewards after the wooden barriers were removed. Indeed, the approaching velocities were not different between the two reward conditions (Fig.

4.1D). However, the final delay-excited responses were graded by reward size, which makes the efference copy idea unlikely. Alternatively, delay-excited cells may signal negative arousal (e.g, frustration and anxiety) elevated during reluctant waiting periods (Moruzzi and Magoun, 1949), but this explanation cannot account for the positive linear relationship between the delay-excited activity and choice preference for LL reward (Fig. 4.3D). Another possibility is that the delay-excited activity represents the animals' attention to the wooden barriers during delays. This notion can account for delay-specific increases in firing, but it fails to explain stronger delay-excited responses in anticipation for larger rewards (Fig. 4.7B). Overall, delay-excited activity cannot be understood without the integration of value information. Among the multiple functions of the MRF, the motivational value account best describes the firing patterns of delay-excited cells.

Making a choice between outcomes that are available at different times requires the decision maker to compare the discounted values of the two outcomes. During the subsequent temporal delay, the expected value of the chose outcome should be continuously represented in order to maintain willingness to wait. Previous studies using Pavlovian and instrumental learning paradigms suggest that the relative value of delayed rewards is computed at the time of choice in multiple brain structures (McClure et al., 2004; Kable and Glimcher, 2007; Ballard and Knutson, 2009) such as prefrontal cortex (Roesch et al., 2006; Kim et al., 2008), striatum (Apicella et al., 1992; Roesch et al., 2009; Cai et al., 2011; Day et al., 2011), and midbrain DA neurons (Roesch et al., 2007; Fiorillo et al., 2008; Kobayashi and Schultz, 2008; Gan et al., 2010). Expectancy-related signals during waiting periods are suggested to be maintained in the orbitofrontal cortex (OFC), a subregion of the prefrontal cortex (Roesch et al., 2006), striatum (Apicella et al., 1992; Day et al., 2011), and GABAergic neurons in the ventral tegmental area (VTA) (Cohen et al.,

2012). In addition, the current study provides compelling evidence that the MRF is also one component of the valuation systems that represent expected reward values. The major feature which distinguishes the MRF from the other valuation systems is that its neural activity can represent both absolute and relative values of expected rewards. The final peaks of delay-excited responses reliably indicated the amounts of upcoming rewards, while average delay-excited responses throughout the entire waiting periods reflected the relative desirability of delayed rewards (Fig. 4.3B and 4.7B). By contrast, when OFC cells and putative GABAergic neurons in the VTA were examined in the same delay discounting task, their ramping activity during delays mainly reflected the relative value of differently delayed rewards over the course of delays including the late phase of the delays.

It appeared that the MRF was not involved in value comparisons at the time of choice. As shown in the spatial firing maps of a representative cell (Fig. 4.3A), no delay-excited cells displayed significant changes in firing at critical places for decision making (e.g., the start arm or the junction of the three T-maze arms). Therefore, it is highly likely that the OFC and the striatum play a critical role in comparing decision values before selecting a goal arm and then the chosen reward information may be transmitted into the MRF. Supporting evidence can be found in the anatomical literature showing reciprocal connections between the two areas (OFC and VTA) and the MRF (Jones and Yang, 1985; Heimer et al., 1991; Vertes, 2004; Hoover and Vertes, 2011).

However, midbrain DA neurons unlikely contribute to delay-excited activity in the MRF, because the DA neurons recorded in the same task signaled the relative value of delayed rewards at the time of delay termination, rather than before delay onset. Conversely, it is possible that the preceding delay-excited activity in the MRF may influence the DA responses at the end of delays.

DA phasic activity has been thought to report errors in reward prediction (Schultz et al., 1997; Roesch et al., 2007). To compute prediction errors, dopaminergic neurons need expected reward signals. Given that the MRF directly sends glutamatergic projections to midbrain dopaminergic neurons (Geisler et al., 2007; Watabe-Uchida et al., 2012), delay-excited cells in the MRF may convey value information to be used by dopaminergic neurons for computing prediction error signals. To elucidate how the MRF interacts with midbrain dopaminergic systems and other valuation systems (e.g, the OFC and the striatum), further studies are necessary.

Chapter 5: General discussion

The studies presented in the current dissertation provide some insights into the prefrontal contribution to signaling prediction errors by DA cells in the VTA. Among multiple prefrontal subregions in rodents, the mPFC and the OFC were of interest because their single neurons show reward expectancy-related firing (Pratt and Mizumori, 2001; Miyazaki et al., 2004; Roesch et al., 2006; Sul et al., 2010; Takahashi et al., 2013). In the following section, it is discussed what information the two structures may provide to DA cells in navigation-based decision tasks.

mPFC modulation of DAergic prediction errors

To examine the mPFC's contribution to prediction errors, firing patterns of DA cells in the VTA were compared before and after microinfusing muscimol, a $GABA_A$ receptor agonist, into the mPFC while rats performed a spatial working memory task on a radial 8-arm maze (Jo et al., 2013). The task required the animals to keep track of previously visited spatial locations and choose unvisited maze arms in order to obtain small or large amounts of chocolate milk. Performance in the task has been demonstrated to depend on the mPFC and some of its single neurons represent the magnitude of expected rewards while rats approached the rewards in the task (Pratt and Mizumori, 2001). Consistent with prediction error signals observed in Pavlovian conditioning paradigms (Schultz et al., 1997; Pan et al., 2005; Clark et al., 2010), DA cells exhibited a response shift from the time of obtaining rewards to approximately 100 ms before the acquisition of the rewards over the course of behavioral recording sessions. The relatively short shift indicated that odors emanating from the reward itself and/or visual identification of its presence could have served as reword-predicting cues. In addition, as it was reported that

GABAergic neurons in the VTA encode the value of expected outcomes (Cohen et al., 2012), some of non-DA neurons in the task exhibited an elevation in firing during approaches to rewards and their increased responses were graded by the size of expected rewards. Temporary inactivation of the mPFC induced a severe reduction in firing of the non-DA cells during the approach and a significant increase in DA activity in response to the predictive sensory cues. Thus, mPFC dysfunction increased DAergic prediction errors triggered by the presence of rewards as rats approached reward locations with less expectation. When DA cells responded after obtaining rewards during the first several recording sessions in which the sensory cues had not fully acquired their predictive values, mPFC inactivation led to no alternation in reward activity by DA cells. These findings suggest that the mPFC provides value signals to midbrain DA systems selectively when future outcomes are expected in the spatial working memory task.

It was further examined whether mPFC dysfunction consistently disrupted value signals in the VTA while rats were engaged in other behavioral tasks that required a different mPFC function than outcome prediction. For instance, in a delay discounting task where rats are faced with a choice between two future outcomes that are available at different times, the mPFC is implicated in processing time intervals, rather than comparing the relative value of two delayed outcomes (Dietrich and Allen, 1998; Narayanan et al., 2006; Rudebeck et al., 2006; Kim et al., 2013). To address this idea, individual VTA neurons were monitored with mPFC inactivation from rats performing a delay discounting task in which two goal arms of a T-maze were associated with differently delayed rewards (Chapter 2). In a given trial, the animals had to choose one goal arm and wait for a certain temporal delay until a wooden barrier in front of the reward was removed. DA cells in this task signaled prediction errors by decreasing phasic responses to expected rewards as behavioral recording sessions progressed. Instead, their

responses were transferred to the end of waiting periods, which indicated that the removal of the wooden barrier served as the reward-predicting cue in the delay discounting task. The barrier itself was not a predictive stimulus because DA cells displayed no phasic responses when rats first arrived in front of it at the time of delay onset. A group of non-DA cells that represented the value of delayed rewards were also identified. These cells continuously increased firing rates during waiting periods. Although briefly inhibited at the time of delay termination, the ramping activity resumed and peaked at the time of obtaining actual rewards. Muscimol injections into the mPFC significantly increased DA responses to both the predictive cue at the time of delay termination and reward acquisition relative to their baseline activity before the drug. However, the same mPFC manipulation led to no changes in firing of the non-DA cells. These results suggest that the mPFC is not a source of value information in the delay-based decision making task.

The elevated DAergic prediction errors were not attributed to the general disinhibition after reduced inputs from the mPFC because mPFC inactivation lowered spontaneous burst firing of DA cells that was measured prior to the behavioral testing. Instead, it is highly likely that, as suggested by the literature (Kim et al., 2013), the mPFC may convey temporal information to DA cells in the task. It is well known that DAergic prediction errors are influenced by precise temporal relationships between reward-predicting cues and reward delivery (Montague et al., 1996; Fiorillo et al., 2008; Kobayashi and Schultz, 2008). Specifically, DA cells exhibit phasic responses to a well-expected reward if it is delivered at a different time than scheduled. Therefore, if mPFC-inactivated rats perceived the lengths of delay to be shorter than before due to an inability to keep track of the elapsed time during the waiting periods, DAergic prediction errors at the times of delay termination and reward would increase. Overall, it appears that the

mPFC regulates signaling of prediction errors by providing different information depending on task demands.

OFC modulation of DAergic prediction errors

The OFC has long been known to encode the value of expected reward (Tremblay and Schultz, 1999; O'Doherty et al., 2001; Gottfried et al., 2003; Izquierdo et al., 2004; Padoa-Schioppa and Assad, 2006; Roesch et al., 2006). Thus, it was hypothesized that value signals in the OFC were fed to the VTA in the previous delay discounting task. To test this notion, the same animals previously tested for mPFC inactivation were again used to inactivate the OFC while DA and non-DA cells were recorded (Chapter 2). Indeed, OFC inactivation strikingly reduced expected value signals by the non-DA cells during waiting periods for delayed rewards. The disrupted reward prediction induced a significant decrease in phasic activation of DA cells at the time of delay termination and a sudden increase at the time of reward encounters. These alternations of DAergic prediction errors provide compelling evidence that the OFC provides model-based value to midbrain DA systems in the delay-based decision making task.

In contrast to the findings, Takahashi et al (2011) reached a different conclusion about the OFC's contribution to DAergic prediction error signaling. They reported that damage to the OFC disrupted a sudden change in prediction error when rewards were unexpectedly increased or decreased in value. Without intact OFC function, prediction errors triggered by predictive cues were also unable to reflect the value of chosen rewards when the reward options were selected freely, but not forcefully. Subsequent computational modeling on the basis of TD framework demonstrated that such altered DA responses could be explained if the OFC was

responsible for conveying not value signals per se, but rather state information about the current choices and their likely outcomes (Takahashi et al., 2011; Wilson et al., 2014).

To determine whether the OFC represented value signals or various states of the ongoing task, single unit activity in the OFC was recording in a modified version of the delay discounting task in which rats were required to choose between a sooner small (SS) reward and a later large (LL) reward on the same T-maze (Chapter 3). Three different delays (10, 20, and 40 s) were imposed prior to LL reward across blocks of trials, whereas the delay to SS reward was held constant at 3 s throughout the experiments. Different groups of OFC cells showed excited responses to a series of task-relevant events and periods, such as delay onset, waiting period, delay termination, reward, or post-reward period after consumption. Individual neurons responding in each state signaled a preferred reward condition by firing more strongly than in the other three conditions. However, the population activity of these cells reflected outcome values estimated in each state of the task. Specifically, when future rewards were evaluated at the time of delay onset and during waiting periods, population responses represented the relative value of the differently delayed rewards. When the rewards were subsequently evaluated shortly before (i.e., at the time of delay termination), during, and shortly after reward consumption, population activity of different OFC cells encoded the magnitude of the delayed rewards, irrespective of different delays preceding them. These results provide compelling evidence that the OFC represents both specific outcome information and their relative values at the individual and population levels, respectively. The value signals within each state seem to be modulated by changing the proportion of individual cells preferring a more valuable outcome relative to other cells preferring less valuable options. Therefore, the OFC is a strong candidate to convey both

state representations and value signals to midbrain DA systems during delay-based decision making.

Electrophysiological techniques in combination with chemical inactivation or lesions have contributed to understanding how the PFC modulates neural activity in midbrain DA systems of rats freely behaving in various reinforcement learning and decision making tasks (Takahashi et al., 2011; Jo et al., 2013). The results indicate that normal coding of DAergic prediction errors requires reward expectancy represented in the PFC (Pratt and Mizumori, 2001; Miyazaki et al., 2004; Roesch et al., 2006; Sul et al., 2010; Takahashi et al., 2013; Stalnaker et al., 2014). Different subregions in the PFC are selectively involved in providing expected reward values to the DA systems depending on cognitive demands in different behavioral tasks (e.g., the mPFC in a spatial working memory task and the OFC in a delay discounting task).

Although PFC inactivation significantly decreased expected value signals by non-DA, presumably GABAergic, neurons in the VTA, they still exhibited elevated firing as the acquisition of expected rewards became imminent. Even after PFC dysfunction, DAergic prediction errors in response to predictive cues were also graded by the value of expected rewards, although decreased in amplitude. Thus, it seems likely that the ventral striatum independently convey model-based values to the DA systems (Apicella et al., 1992; Suri and Schultz, 2001; O'Doherty et al., 2004; Roesch et al., 2009; Cai et al., 2011; Day et al., 2011). It is also possible that the MRF is another source of value signals, given that the MRF directly sends glutamatergic projections to DA cells (Geisler et al., 2007; Watabe-Uchida et al., 2012).

Reference

Apicella P, Scarnati E, Ljungberg T, Schultz W (1992) Neuronal activity in monkey striatum related to the expectation of predictable environmental events. J Neurophysiol 68:945-960.

Ballard K, Knutson B (2009) Dissociable neural representations of future reward magnitude and delay during temporal discounting. Neuroimage 45:143-150.

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129-141.

Cai X, Kim S, Lee D (2011) Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. Neuron 69:170-182.

Carr DB, Sesack SR (2000) Projections from the rat prefrontal cortex to the ventral tegmental area: target specificity in the synaptic associations with mesoaccumbens and mesocortical neurons. J Neurosci 20:3864-3873.

Clark JJ, Hollon NG, Phillips PE (2012) Pavlovian valuation systems in learning and decision making. Curr Opin Neurobiol 22:1054-1061.

Clark JJ, Sandberg SG, Wanat MJ, Gan JO, Horne EA, Hart AS, Akers CA, Parker JG, Willuhn I, Martinez V, Evans SB, Stella N, Phillips PE (2010) Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. Nat Methods 7:126-129.

Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482:85-88.

Davis KL, Kahn RS, Ko G, Davidson M (1991) Dopamine in schizophrenia: a review and reconceptualization. Am J Psychiatry 148:1474-1486.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704-1711.

Day JJ, Jones JL, Carelli RM (2011) Nucleus accumbens neurons encode predicted and ongoing reward costs in rats. Eur J Neurosci 33:308-321.

Day JJ, Roitman MF, Wightman RM, Carelli RM (2007) Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. Nat Neurosci 10:1020-1028.

Dietrich A, Allen JD (1998) Functional dissociation of the prefrontal cortex and the hippocampus in timing behavior. Behav Neurosci 112:1043-1047.

Doya K (2002) Metalearning and neuromodulation. Neural Netw 15:495-506.

Fabre M, Rolls ET, Ashton JP, Williams G (1983) Activity of neurons in the ventral tegmental region of the behaving monkey. Behav Brain Res 9:213-235.

Feierstein CE, Quirk MC, Uchida N, Sosulski DL, Mainen ZF (2006) Representation of spatial goals in rat orbitofrontal cortex. Neuron 51:495-507.

Fields HL, Hjelmstad GO, Margolis EB, Nicola SM (2007) Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. Annu Rev Neurosci 30:289-316.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898-1902.

Fiorillo CD, Newsome WT, Schultz W (2008) The temporal precision of reward prediction in dopamine neurons. Nat Neurosci.

Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PE, Akil H (2011) A selective role for dopamine in stimulus-reward learning. Nature 469:53-57.

Floresco SB, West AR, Ash B, Moore H, Grace AA (2003) Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. Nat Neurosci 6:968-973.

French SJ, Totterdell S (2002) Hippocampal and prefrontal cortical inputs monosynaptically converge with individual projection neurons of the nucleus accumbens. J Comp Neurol 446:151-165.

Furuyashiki T, Holland PC, Gallagher M (2008) Rat orbitofrontal cortex separately encodes response and outcome information during performance of goal-directed behavior. J Neurosci 28:5127-5138.

Futami T, Takakusaki K, Kitai ST (1995) Glutamatergic and cholinergic inputs from the pedunculopontine tegmental nucleus to dopamine neurons in the substantia nigra pars compacta. Neurosci Res 21:331-342.

Gallagher M, McMahan RW, Schoenbaum G (1999) Orbitofrontal cortex and representation of incentive value in associative learning. J Neurosci 19:6610-6614.

Gan JO, Walton ME, Phillips PE (2010) Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci 13:25-27.

Gariano RF, Groves PM (1988) Burst firing induced in midbrain dopamine neurons by stimulation of the medial prefrontal and anterior cingulate cortices. Brain Res 462:194-198.

Geisler S, Wise RA (2008) Functional implications of glutamatergic projections to the ventral tegmental area. Rev Neurosci 19:227-244.

Geisler S, Derst C, Veh RW, Zahm DS (2007) Glutamatergic afferents of the ventral tegmental area in the rat. J Neurosci 27:5730-5743.

Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585-595.

Gottfried JA, O'Doherty J, Dolan RJ (2003) Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 301:1104-1107.

Grace AA (2000) Gating of information flow within the limbic system and the pathophysiology of schizophrenia. Brain Res Brain Res Rev 31:330-341.

Grace AA, Bunney BS (1984) The control of firing pattern in nigral dopamine neurons: burst firing. J Neurosci 4:2877-2890.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J Neurosci 28:5623-5630.

Heimer L, Zahm DS, Churchill L, Kalivas PW, Wohltmann C (1991) Specificity in the projection patterns of accumbal core and shell in the rat. Neuroscience 41:89-125.

Hoover WB, Vertes RP (2011) Projections of the medial orbital and ventral orbital cortex in the rat. J Comp Neurol 519:3766-3801.

Hornak J, O'Doherty J, Bramham J, Rolls ET, Morris RG, Bullock PR, Polkey CE (2004) Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. J Cogn Neurosci 16:463-478.

Izquierdo A, Suda RK, Murray EA (2004) Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. J Neurosci 24:7540-7548.

Jackson ME, Frost AS, Moghaddam B (2001) Stimulation of prefrontal cortex at physiologically relevant frequencies inhibits dopamine release in the nucleus accumbens. J Neurochem 78:920-923.

Jo YS, Lee I (2010) Disconnection of the hippocampal-perirhinal cortical circuits severely disrupts object-place paired associative memory. J Neurosci 30:9850-9858.

Jo YS, Lee J, Mizumori SJ (2013) Effects of prefrontal cortical inactivation on neural activity in the ventral tegmental area. J Neurosci 33:8159-8171.

Jo YS, Park EH, Kim IH, Park SK, Kim H, Kim HT, Choi JS (2007) The medial prefrontal cortex is involved in spatial memory retrieval under partial-cue conditions. J Neurosci 27:13567-13578.

Joel D, Niv Y, Ruppin E (2002) Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw 15:535-547.

Jones BE, Yang TZ (1985) The efferent projections from the reticular formation and the locus coeruleus studied by anterograde and retrograde axonal transport in the rat. J Comp Neurol 242:56-92.

Jones JL, Esber GR, McDannald MA, Gruber AJ, Hernandez A, Mirenzi A, Schoenbaum G (2012) Orbitofrontal cortex supports behavior and learning using inferred but not cached values. Science 338:953-956.

Jung MW, Baeg EH, Kim MJ, Kim YB, Kim JJ (2008) Plasticity and memory in the prefrontal cortex. Rev Neurosci 19:29-46.

Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. Nat Neurosci 10:1625-1633.

Kamin LJ (1969) Predictability, surprise, attention, and conditioning. In: Punishment and Aversive Behavior (Campbell BA, Church RM, eds), pp 279–296. New York, NY: Appleton-Century-Crofts.

Kawato M, Samejima K (2007) Efficient reinforcement learning: computational theories, neuroscience and robotics. Curr Opin Neurobiol 17:205-212.

Kennerley SW, Wallis JD (2009) Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. Eur J Neurosci 29:2061-2073.

Kennerley SW, Behrens TE, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat Neurosci 14:1581-1589.

Kim J, Ghim JW, Lee JH, Jung MW (2013) Neural correlates of interval timing in rodent prefrontal cortex. J Neurosci 33:13834-13847.

Kim S, Hwang J, Lee D (2008) Prefrontal coding of temporally discounted values during intertemporal choice. Neuron 59:161-172.

Kinomura S, Larsson J, Gulyas B, Roland PE (1996) Activation by attention of the human reticular formation and thalamic intralaminar nuclei. Science 271:512-515.

Kobayashi S, Schultz W (2008) Influence of reward delays on responses of dopamine neurons. J Neurosci 28:7837-7846.

Lodge DJ (2011) The medial prefrontal and orbitofrontal cortices differentially regulate dopamine system function. Neuropsychopharmacology 36:1227-1236.

McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. Science 306:503-507.

McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G (2011) Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. J Neurosci 31:2700-2705.

Meyer-Lindenberg A, Miletich RS, Kohn PD, Esposito G, Carson RE, Quarantelli M, Weinberger DR, Berman KF (2002) Reduced prefrontal activity predicts exaggerated striatal dopaminergic function in schizophrenia. Nat Neurosci 5:267-271.

Miles FA, Evarts EV (1979) Concepts of motor organization. Annu Rev Psychol 30:327-362.

Miyazaki K, Miyazaki KW, Matsumoto G (2004) Different representation of forthcoming reward in nucleus accumbens and medial prefrontal cortex. Neuroreport 15:721-726.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936-1947.

Morita K, Morishima M, Sakai K, Kawaguchi Y (2012) Reinforcement learning: computing the temporal difference of values via distinct corticostriatal pathways. Trends Neurosci 35:457-467.

Moruzzi G, Magoun HW (1949) Brain stem reticular formation and activation of the EEG. Electroencephalogr Clin Neurophysiol 1:455-473.

Narayanan NS, Horst NK, Laubach M (2006) Reversible inactivations of rat medial prefrontal cortex impair the ability to wait for a stimulus. Neuroscience 139:865-876.

Niv Y, Joel D, Dayan P (2006) A normative perspective on motivation. Trends Cogn Sci 10:375-381.

Norton AB, Jo YS, Clark EW, Taylor CA, Mizumori SJ (2011) Independent neural coding of reward and movement by pedunculopontine tegmental nucleus neurons in freely navigating rats. Eur J Neurosci.

O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 4:95-102.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452-454.

Okada K, Toyama K, Inoue Y, Isa T, Kobayashi Y (2009) Different pedunculopontine tegmental neurons signal predicted and actual task rewards. J Neurosci 29:4858-4870.

Olds J, Mink WD, Best PJ (1969) Single unit patterns during anticipatory behavior. Electroencephalogr Clin Neurophysiol 26:144-158.

Omelchenko N, Sesack SR (2009) Ultrastructural analysis of local collaterals of rat ventral tegmental area neurons: GABA phenotype and synapses onto dopamine and GABA cells. Synapse 63:895-906.

Overton PG, Tong ZY, Clark D (1996) A pharmacological analysis of the burst events induced in midbrain dopaminergic neurons by electrical stimulation of the prefrontal cortex in the rat. J Neural Transm 103:523-540.

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223-226.

Padoa-Schioppa C, Cai X (2011) The orbitofrontal cortex and the computation of subjective value: consolidated concepts and new perspectives. Ann N Y Acad Sci 1239:130-137.

Pan WX, Hyland BI (2005) Pedunculopontine tegmental nucleus controls conditioned responses of midbrain dopamine neurons in behaving rats. J Neurosci 25:4725-4732.

Pan WX, Schmidt R, Wickens JR, Hyland BI (2005) Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J Neurosci 25:6235-6242.

Peterson BW (1984) The reticulospinal system and its role in the control of movement. New York: Academic Press.

Phillips MI, Olds J (1969) Unit activity: motiviation-dependent responses from midbrain neurons. Science 165:1269-1271.

Pragay EB, Mirsky AF, Ray CL, Turner DF, Mirsky CV (1978) Neuronal activity in the brain stem reticular formation during performance of a "go-no go" visual attention task in the monkey. Exp Neurol 60:83-95.

Pratt WE, Mizumori SJ (2001) Neurons in rat medial prefrontal cortex show anticipatory rate changes to predictable differential rewards in a spatial memory task. Behav Brain Res 123:165-183.

Puryear CB, Mizumori SJ (2008) Reward prediction error signals by reticular formation neurons. Learn Mem 15:895-898.

Ray CL, Mirsky AF, Pragay EB (1982) Functional analysis of attention-related unit activity in the reticular formaton of the monkey. Exp Neurol 77:544-562.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical conditioning II: current research and theory (Black AH, Prokasky WF, eds), pp 64–99. New York, NY: Appleton-Century-Crofts.

Roesch MR, Taylor AR, Schoenbaum G (2006) Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. Neuron 51:509-520.

Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 10:1615-1624.

Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G (2009) Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. J Neurosci 29:13365-13376.

Rolls ET, Baylis LL (1994) Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex. J Neurosci 14:5437-5452.

Rolls ET, Critchley HD, Mason R, Wakeman EA (1996) Orbitofrontal cortex neurons: role in olfactory and visual association learning. J Neurophysiol 75:1970-1981.

Rudebeck PH, Walton ME, Smyth AN, Bannerman DM, Rushworth MF (2006) Separate neural pathways process different decision costs. Nat Neurosci 9:1161-1168.

Schoenbaum G, Chiba AA, Gallagher M (1999) Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. J Neurosci 19:1876-1884.

Schoenbaum G, Takahashi Y, Liu TL, McDannald MA (2011) Does the orbitofrontal cortex signal value? Ann N Y Acad Sci 1239:87-99.

Schoenbaum G, Setlow B, Nugent SL, Saddoris MP, Gallagher M (2003) Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. Learn Mem 10:129-140.

Schultz W (2002) Getting formal with dopamine and reward. Neuron 36:241-263.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599.

Sesack SR, Carr DB, Omelchenko N, Pinto A (2003) Anatomical substrates for glutamate-dopamine interactions: evidence for specificity of connections and extrasynaptic actions. Ann N Y Acad Sci 1003:36-52.

Siegel JM, McGinty DJ (1977) Pontine reticular formation neurons: relationship of discharge to motor activity. Science 196:678-680.

Smittenaar P, FitzGerald TH, Romei V, Wright ND, Dolan RJ (2013) Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. Neuron 80:914-919.

Stalnaker TA, Cooch NK, McDannald MA, Liu TL, Wied H, Schoenbaum G (2014) Orbitofrontal neurons infer the value and identity of predicted outcomes. Nat Commun 5:3926.

Stefani MR, Moghaddam B (2005) Systemic and prefrontal cortical NMDA receptor blockade differentially affect discrimination learning and set-shift ability in rats. Behav Neurosci 119:420-428.

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16:966-973.

Steriade M, McCarley RW (1990) Brainstem control of wakefulness and sleep. New York: Plenum Press.

Sul JH, Kim H, Huh N, Lee D, Jung MW (2010) Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron 66:449-460.

Suri RE, Schultz W (2001) Temporal difference model reproduces anticipatory neural activity. Neural Comput 13:841-862.

Sutton RS, Barto AG (1990) Time-derivative models of pavlovian reinforcement. In: Learning and Computational Neuroscience: Foundations of Adaptive Networks (Gabriel M, Moore J, eds), pp 497–537. Cambrideg, MA: MIT Press.

Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, Schoenbaum G (2011) Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. Nat Neurosci 14:1590-1597.

Takahashi YK, Roesch MR, Stalnaker TA, Haney RZ, Calu DJ, Taylor AR, Burke KA, Schoenbaum G (2009) The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. Neuron 62:269-280.

Takahashi YK, Chang CY, Lucantonio F, Haney RZ, Berg BA, Yau HJ, Bonci A, Schoenbaum G (2013) Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. Neuron 80:507-518.

Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. Science 307:1642-1645.

Tong ZY, Overton PG, Clark D (1996) Stimulation of the prefrontal cortex in the rat induces patterns of activity in midbrain dopaminergic neurons which resemble natural burst events. Synapse 22:195-208.

Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. Nature 398:704-708.

van Duuren E, Escamez FA, Joosten RN, Visser R, Mulder AB, Pennartz CM (2007) Neural coding of reward magnitude in the orbitofrontal cortex of the rat during a five-odor olfactory discrimination task. Learn Mem 14:446-456.

Vazquez-Borsetti P, Cortes R, Artigas F (2009) Pyramidal neurons in rat prefrontal cortex projecting to ventral tegmental area and dorsal raphe nucleus express 5-HT2A receptors. Cereb Cortex 19:1678-1686.

Vertes RP (2004) Differential projections of the infralimbic and prelimbic cortex in the rat. Synapse 51:32-58.

Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43-48.

Wallis JD (2012) Cross-species studies of orbitofrontal cortex and value-based decision-making. Nat Neurosci 15:13-19.

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

Weinberger DR (1987) Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 44:660-669.

Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. Neuron 81:267-279.

Wise RA (2004) Dopamine, learning and motivation. Nat Rev Neurosci 5:483-494.

Woolf NJ (1991) Cholinergic systems in mammalian brain and spinal cord. Prog Neurobiol 37:475-524.

Worgotter F, Porr B (2005) Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. Neural Comput 17:245-319.

Zweifel LS, Parker JG, Lobb CJ, Rainwater A, Wall VZ, Fadok JP, Darvas M, Kim MJ, Mizumori SJ, Paladini CA, Phillips PE, Palmiter RD (2009) Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. Proc Natl Acad Sci U S A 106:7281-7288.

Curriculum Vitae

Eucation

Ph.D., Behavioral Neuroscience (2014)

Department of Psychology, University of Washington, Seattle, WA (2009-2014)

Advisor: Sheri J. Y. Mizumori, Ph.D.

M.A., Biopsychology (2007)

Department of Psychology, Korea University, Seoul, Korea (2005-2007)

Advisor: June-Seek Choi, Ph.D.

Thesis title: Involvement of the medial prefrontal cortex in spatial memory retrieval under partial-cue conditions.

B.A., Psychology (2005)

Department of Psychology, Korea University, Seoul, Korea (1998-2005)

Research experience

• Research Assistant, Department of Psychology, University of Washington, WA (2009-2014)

Advisor: Sheri J.Y. Mizumori, Ph.D.

Research field: Prefrontal modulation of midbrain dopamine systems in navigation-based decision making tasks

• Research Assistant, Department of Psychology, University of Iowa, IA (2008-2009)

Advisor: Inah Lee, Ph.D.

Research field: Interactions between the hippocampus and perirhinal cortex in an object-place paired association task

• Graduate Student/Research Assistant, Department of Psychology, Korea University, Seoul, Korea (2005-2007)

Advisor: June-Seek Choi, Ph.D.

Research fields

1) Effects of aging and early-adulthood stress on fear memory and declarative memory

2) Roles of the medial prefrontal cortex in fear memory acquisition and extinction

• Undergraduate Research Assistant/Animal Caretaker, Department of Psychology, Korea University, Seoul, Korea (2003-2004)

Advisor: Hyun Taek Kim, Ph.D.

Research field: Effects of antidepressant drugs, rTMS, and ECT in an animal model of depression

Teaching experience

• Teaching Assistant, Neurobehavioral Lab (Fall 2011, Spring 2012 and 2014, Winter 2014)

• Teaching Assistant, Biopsychology (Spring 2011 and 2013, Fall 2012 and 2013, Winter 2013)

• Teaching Assistant, Introduction to Psychology (Winter 2012)

Awards

• Hunt Fellowship, Department of Psychology, University of Washington, 2014

• Alcor Fellowship, Department of Psychology, University of Washington, 2013

• Top Scholar Award, Graduate School, University of Washington, 2010

• Dissertation Excellence Award, Graduate School, Korea University, 2008

• Brain Korea 21 Fellowships, Department of Psychology, Korea University, 2006

Publications

Redila, V., Kinzel, C., **Jo, Y.S.**, Puryear, C.B., and Mizumori, S.J.Y. (2014). A Role for the Lateral Dorsal Tegmentum in Memory and Decision Neural Circuitry. *Neurobiology of Learning and Memory* (in press).

Jo, Y.S. and Choi, J.-S. (2014). Memory retrieval in response to partial cues requires NMDA receptor-dependent neurotransmission in the medial prefrontal cortex. *Neurobiology of Learning and Memory* 109:20-26.

Mizumori, S.J.Y. and **Jo, Y.S.** (2013). Homeostatic regulation of memory systems and adaptive decisions. *Hippocampus* 23:1103-1124.

Jo, Y.S., Lee, J., and Mizumori, S.J.Y. (2013). Effects of prefrontal cortical inactivation on neural activity in the ventral tegmental area. *Journal of Neuroscience* 33:8159-8171. (Selected as featured article)

Yoon, K.-J., Lee, H.-R., **Jo, Y.S.**, An, K., Jung, S.-Y., Jeong, M.-W., Kwon, S.-K., Kim, N.-S., Jeong, H.-W., An, S.-H., Kim, K.-T., Lee, K., Kim, E., Kim, J.-H., Choi, J.-S., Kaang, B.-K,

and Kong, Y.-Y. (2012). Mind bomb-1 is an essential modulator of long-term memory and synaptic plasticity via the Notch signaling pathway. *Molecular Brain* 5:40.

Norton, A.B.W., **Jo, Y.S.**, Clark, E.W., Taylor, C.A., and Mizumori, S.J.Y. (2011). Independent neural coding of reward and movement by pedunculopontine tegmental nucleus neurons in freely navigating rats. *European Journal of Neuroscience* 33:1885-1896.

Jo, Y.S. and Lee, I. (2010). Disconnection of the hippocampal-perirhinal cortical circuits severely disrupts object-place paired associative memory. *Journal of Neuroscience* 30:9850-9858.

Jo, Y.S. and Lee, I. (2010). Perirhinal cortex is necessary for acquiring, but not for retrieving object-place paired association. *Learning & memory* 17:97-103.

Lee, J., **Jo, Y.S.**, Sung, Y.H., Hwang, I.K., Kim, K., Kim, S.-Y., Yi, S.S., Choi, J.-S., Sun, W., Seong, J.K., and Lee, H.-W. (2010). Telomerase deficiency affects normal brain functions in mice. *Neurochemical Research* 35:211-218.

Kim, S.C., **Jo, Y.S.**, Kim, I.H., Kim, H., and Choi, J.-S. (2010). Lack of the medial prefrontal cortex activation underlies the immediate extinction deficit. *Journal of Neuroscience* 30:832-837.

Kim, I.H., Park, S.K., Hong, S.T., **Jo, Y.S.**, Kim, E.J., Park, E.H., Han, S.B., Shin, H.-S., Sun, W., Kim, H.T., Soderling, S.H., and Kim, H. (2009). Inositol 1,4,5-trisphosphate 3-kinase A functions as a scaffold for synaptic Rac signaling. *Journal of Neuroscience* 29:14039-14049.

Jo, Y.S., Park, E.H., Kim, I.H., Park, S.K., Kim, H., Kim, H.T., and Choi, J.-S. (2007). The medial prefrontal cortex is involved in spatial memory retrieval under partial-cue conditions. *Journal of Neuroscience* 27:13567-13578.

Research talks

Jo, Y.S. and Mizumori, S.J.Y. (2013). Delay-dependent activity by midbrain reticular formation cells during delay-based decision making. *NIDA center seminar, University of Washington, Seattle, WA.*

Jo, Y.S., Lee, J., and Mizumori, S.J.Y. (2011). Differential effects of prefrontal cortex inactivation on reward signals of dopamine neurons in the ventral tegmental area. *NIDA center seminar, University of Washington, Seattle, WA.*

Jo, Y.S. and Mizumori, S.J.Y. (2011). Roles of the ventral tegmental area neurons in a delay-based decision making task. *35th Annual Winter Conference on Neurobiology of Learning & Memory, Park City, UT.*

Jo, Y.S. and Mizumori, S.J.Y. (2010). Delay responses of the ventral tegmental area neurons during a delay-based decision making task. *39th Annual Psychology Research Festival, Department of Psychology, University of Washington, Seattle, WA.*

Conference presentations

Jo, Y.S., Fobbs, W., and Mizumori, S.J.Y. (2013). Effects of orbitofrontal inactivation on dopamine cell activity during a delay-based decision task. *Society for Neuroscience Abstract, San Diego, CA.*

Jo, Y.S. and Mizumori, S.J.Y. (2012). Delay activity by mesencephalic reticular formation neurons during intertemporal decision making. *Society for Neuroscience Abstract, New Orleans, LA.*

Jo, Y.S., Lee, J., and Mizumori, S.J.Y. (2011). Differential effects of prefrontal cortex inactivation on reward response and anticipation in dopamine neurons of the ventral tegmental area. *Society for Neuroscience Abstract, Washington, DC.*

Jo, Y.S. and Lee, I. (2009). Contralateral, but not ipsilateral inactivations of the hippocampus and perirhinal cortex severely disrupt object-place paired associative memory. *Society for Neuroscience Abstract, Chicago, IL.*

Lee, I. and **Jo, Y.S.** (2009). Effects of perirhinal cortical lesions are temporary on retrieval but relatively permanent on the acquisition for discriminating similar object-place paired associations. *Society for Neuroscience Abstract, Chicago, IL.*

Yoon, K.-J., Ahn, K.-M., **Jo, Y.S.**, Joung, S.-Y., Lee, H.-R., Kaang, B.-K., Choi, J.-S., Kim, J.-H., and Kong, Y.-Y. (2009). Mind bomb-1 is an essential modulator of long-term memory and synaptic plasticity via Notch signaling pathway. *Society for Neuroscience Abstract, Chicago, IL.*

Park, J., Kwon, J., **Jo, Y.S.**, Kim, I.H., Park, E.-H., Kim, H.-T., and Choi, J.-S. (2008). Infralimbic cortex stimulation paired with CS applied under anesthesia facilitates fear extinction. *Society for Neuroscience Abstract, Washington, DC.*

Kim, S.C., **Jo**, **Y.S.**, Kim, I., Choi, E., Kim, E., and Choi, J.-S. (2008). Differential expression of immediate early gene in medial prefrontal cortex underlies extinction memory recall following immediate and delayed extinction. *Society for Neuroscience Abstract, Washington, DC.*

Jo, Y.S., Park, E.H., Seol, G.H., Kim, H.T., Choi, J.-S. (2007). Role of the medial prefrontal cortex in spatial memory retrieval under partial-cue conditions. *Society for Neuroscience Abstract, San Diego, CA.*

Park, E.H., **Jo, Y.S.**, Lee, K.J., Kim, J.-W., Rhyu, I.J., Choi, J.-S., and Kim, H.T. (2007). Effect of early-adulthood stress on hippocampal synapse in aged rats. *Society for Neuroscience Abstract, San Diego, CA.*

Jo Y.S., Park E.H., Kim H.T. and Choi J.-S. (2007). The medial prefrontal cortex is critical for spatial memory retrieval from degraded context, but not spatial working memory. *International Brain Research Organization Abstract, Melbourne, Australia.*

Park, E.H., **Jo, Y.S.**, Hong, S.B., Kim, E.J., Choi, J.-S., and Kim, H.T. (2006). Effects of early-life stress exposure and aging in the acquisition and retrieval of fear memory. *Federation of European Neuroscience Societies Abstract, Vienna, Austria.*

Park, E.H., **Jo, Y.S.**, Park, E.H., Kim, E.J., Hong, S.B., Choi, J.-S., and Kim, H.T. (2005). Early stress exposure exacerbates aging-related cognitive deficits in rats. *Society for Neuroscience Abstract, Washington, DC.*