frontiers in Psychology

REVIEW
published: 16 July 2015
doi: 10.3389/fpsyg.2015.00995

Dissociable functions of reward inference in the lateral prefrontal cortex and the striatum

Shingo Tanaka [1], Xiaochuan Pan [1,2], Mineki Oguchi [1], Jessica E. Taylor [1,3] and Masamichi Sakagami [1,3]*

[1] *Brain Science Institute, Tamagawa University, Machida, Japan,* [2] *Institute for Cognitive Neurodynamics, East China University of Science and Technology, Shanghai, China,* [3] *Graduate School of Brain Sciences, Tamagawa University, Machida, Japan*

In a complex and uncertain world, how do we select appropriate behavior? One possibility is that we choose actions that are highly reinforced by their probabilistic consequences (model-free processing). However, we may instead plan actions prior to their actual execution by predicting their consequences (model-based processing). It has been suggested that the brain contains multiple yet distinct systems involved in reward prediction. Several studies have tried to allocate model-free and model-based systems to the striatum and the lateral prefrontal cortex (LPFC), respectively. Although there is much support for this hypothesis, recent research has revealed discrepancies. To understand the nature of the reward prediction systems in the LPFC and the striatum, a series of single-unit recording experiments were conducted. LPFC neurons were found to infer the reward associated with the stimuli even when the monkeys had not yet learned the stimulus-reward (SR) associations directly. Striatal neurons seemed to predict the reward for each stimulus only after directly experiencing the SR contingency. However, the one exception was "Exclusive Or" situations in which striatal neurons could predict the reward without direct experience. Previous single-unit studies in monkeys have reported that neurons in the LPFC encode category information, and represent reward information specific to a group of stimuli. Here, as an extension of these, we review recent evidence that a group of LPFC neurons can predict reward specific to a category of visual stimuli defined by relevant behavioral responses. We suggest that the functional difference in reward prediction between the LPFC and the striatum is that while LPFC neurons can utilize abstract code, striatal neurons can code individual associations between stimuli and reward but cannot utilize abstract code.

Keywords: lateral prefrontal cortex, striatum, reward inference, model-free learning, model-based learning

OPEN ACCESS

Edited by:
Mitsuhiro Okada, Keio University, Japan

Reviewed by:
V. S. C. Pammi, University of Allahabad, India
Anthony J. Porcelli, Marquette University, USA

****Correspondence:***
Xiaochuan Pan, Institute for Cognitive Neurodynamics, East China University of Science and Technology, Meilong Road 130, Shanghai 200237, China
pxc@ecust.edu.cn

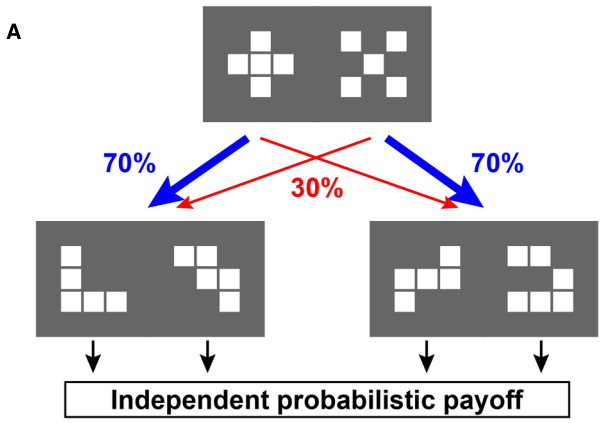
Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Psychology

Received: *11 March 2015*
Accepted: *30 June 2015*
Published: *16 July 2015*

Citation:
Tanaka S, Pan X, Oguchi M, Taylor JE and Sakagami M (2015) Dissociable functions of reward inference in the lateral prefrontal cortex and the striatum. Front. Psychol. 6:995. doi: 10.3389/fpsyg.2015.00995

Introduction

Reward prediction is paramount for learning behavior (Sutton and Barto, 1998; Schultz, 2006) and for decision-making processes in the brain (Rangel et al., 2008). Much research has shown that many brain areas are involved in reward prediction (Yamada et al., 2004; Knutson and Cooper, 2005; Padoa-Schioppa and Assad, 2006; Paton et al., 2006; Behrens et al., 2007, 2008; Hare et al., 2008; Hayden et al., 2008; Hikosaka et al., 2008; Haber and Knutson, 2010; Rushworth et al., 2011; Levy and Glimcher, 2012; Garrison et al., 2013). The basal ganglia and multiple sub-areas in the prefrontal cortex especially play important but different roles in the reward prediction process



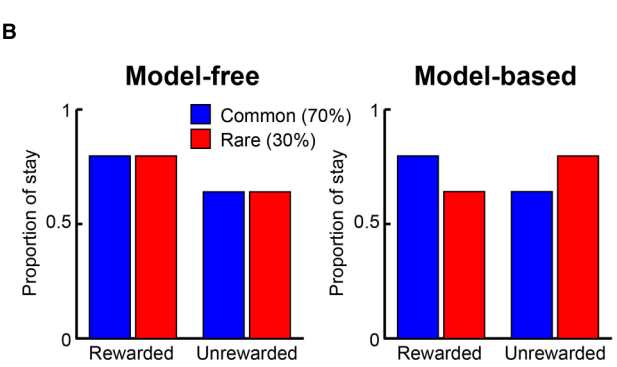


FIGURE 1 | A strategy to examine model-free and model-based learning processes. (A) The two-stage Markov decision task described in Daw et al. (2011). The choice of the first stage led predominantly (70%) to one of the two pairs of options and rarely (30%) to the other pair. The probability of payoff associated with the four second-stage options was changed independently according to a Gaussian random walk protocol. **(B)** A model-free (simple reinforcement learning) strategy predicts higher stay probability in the first stage choice on the next trial after a rewarded present trial than after an unrewarded present trial, regardless of whether the present first stage choice led to a common or rare transition. On the other hand, a model-based strategy predicts higher stay probability in the first stage choice on the next choice after a rewarded present trial than after an unrewarded present trial in a common transition case, and after an unrewarded present trial in a rare transition case.

(Watanabe, 1996; Hollerman et al., 1998; O'Doherty et al., 2003; Roesch and Olson, 2003; Samejima et al., 2005; Hare et al., 2008; Hikosaka and Isoda, 2010; Diekhof et al., 2012). Several fMRI studies have demonstrated the importance of both the lateral prefrontal cortex (LPFC) and the striatum in the basal ganglia for reward prediction and have compared the functional difference in reward prediction between them (McClure et al., 2004; Tanaka et al., 2006; Kahnt et al., 2011). Some studies in monkeys have also directly examined neuronal activities in the LPFC and striatum, providing results that suggest that both areas are involved in the learning of stimulus-reward (SR) associations and that both represent positive and negative reward prediction (Pasupathy and Miller, 2005; Kobayashi et al., 2007; Histed et al., 2009; Asaad and Eskandard, 2011; Pan et al., 2014). Because of the neuroanatomical and pharmacological differences between the cerebral cortex and basal ganglia, the likely functional differences between the LPFC and the striatum can be predicted. For example, numerous discussions have been performed about the functional differences between the prefrontal cortex and the striatum for the learning of behavior in the frameworks of goal-directed/habit learning (Balleine and Dickinson, 1998; Killcross and Coutureau, 2003; O'Doherty et al., 2004; Tricomi et al., 2009; Balleine and O'Doherty, 2010; McNamee et al., 2015).

Recently, from the viewpoint of computational theory, vigorous discussion has emerged about the functional differences in the learning of behavior between the LPFC and the striatum. In particular, the hypothesis of Daw et al. (2005) which relates the difference between "model-based vs. model-free" processes to the difference in functions of the LPFC and striatum, is supported by the results of studies on humans and primates (Joel et al., 2002; Bunge et al., 2003; Doya, 2008; Maia, 2009; Rygula et al., 2010; Beierholm et al., 2011). According to this hypothesis, while the model-free process allows reward prediction to be achieved directly by reinforcement learning without internal models, the model-based process generates in the brain an internal model of the environment (such as cognitive map; Tolman, 1948), grasps the relationship among states in the environment, and predicts rewards depending on these relationships (Glascher et al., 2010). To segregate the model-free and model-based processes, state transition tasks were used in several studies (Glascher et al., 2010; Daw et al., 2011; Doll et al., 2012; Lee et al., 2014; Deserno et al., 2015). For example, in a task with a structure that has a SR relationship as shown in **Figure 1A**, we may predict different responses depending on whether the model-free strategy (**Figure 1B**, left) or the model-based strategy (**Figure 1B**, right) is adopted. In the state transition task, state prediction error (SPE) can be calculated for each choice because transitions between choices are determined stochastically. Glascher et al. (2010) showed that reward prediction error (RPE) calculated from the model-free process is represented in the striatum and that SPE is represented in the LPFC. Because SPE cannot be calculated from the model-free process alone but requires the model-based process instead, it seems the state transition task is useful for separating the model-free and model-based processes in the striatum and LPFC. Daw et al. (2011) also showed that a RPE calculated from the model-free process was represented in the striatum with a state transition task in **Figure 1**. They additionally showed a RPE calculated from the model-based process which happened to be represented in the striatum rather than the LPFC. Based on these results, Daw suggested that this task can separate the model-free and model-based processes but cannot separate the striatal function and LPFC functions. This proposal has been further supported by several studies (Wunderlich et al., 2012; Lee et al., 2014; Walsh and Anderson, 2014; Deserno et al., 2015). Therefore, the hypothesis that the differences in the function of reward prediction in these two areas corresponds to the difference between reward prediction using model-free and model-based processes appears dubious.

Here, we attempt to dissociate the reward prediction functions of the striatum and the LPFC by focusing on two elements. The first is the existence/non-existence of information on transition among environmental states. This has long been discussed in relation to rule-based behavior and higher-order conditioning and many studies have shown the importance of the LPFC for such types of behavior (Sakagami and Niki, 1994; White and Wise, 1999; Hoshi et al., 2000; Wallis et al., 2001; Amemori and Sawaguchi, 2006; Han et al., 2009; Vallentin et al., 2012). The second element we consider is whether and how the LPFC and striatum use subjects' experience to predict reward. This has a history of debate in relation to syllogism, transitive inference, categorical inference, and so forth (McGonigle and Chalmers, 1977; Blaisdell et al., 2006; Murphy et al., 2008). It has been argued that the model-free reinforcement learning process, which is believed to be performed in the nigro-striatal circuit, requires direct experience of obtaining reward for reward prediction (Schultz et al., 1997; Daw et al., 2005; Doya, 2008). On the contrary, it has been argued that the LPFC may integrate several pieces of fragmentary information to associate a stimulus with a future event without learning their interrelations directly (Miller, 2000; Duncan, 2001; Pan et al., 2008, 2014; Pan and Sakagami, 2012). However, the amount of research that has attempted to directly investigate these ideas in regards to striatum and LPFC function remains meager. Therefore, we shall discuss here whether experimental results and the difference in the functions of the prefrontal cortex and striatum are consistent with the two elements proposed above.

The Sequential Paired-Association Task with an Asymmetric Reward Schedule

To clarify the difference in the reward predictive functions in the LPFC and striatum, Pan et al. (2008) developed a reward inference task. In this task, the monkey subjects first learned two stimulus–stimulus association sequences (here denoted: A1→B1→C1 and A2→B2→C2, where A1, B1, C1, A2, B2, and C2 were six different visual stimuli; **Figure 2A**). These were learned in sequential paired-association trials (SPATs) with a symmetrical reward schedule (Pan et al., 2008). After having mastered the task, the monkeys were then taught an asymmetric reward schedule using reward instruction trials (RITs), in which one stimulus (C1 or C2) was paired with a large reward (0.4 ml of water) and the other stimulus (C2 or C1) with a small reward (0.2 ml of water). In behavioral and single-unit recording sessions, RITs were followed by SPATs within each block (**Figure 2B**). In the SPATs, the amount of reward received at the end of correct trials was consistent with that of the RITs from the same block: if C1 had been paired with the large reward, and C2 with the small reward in RITs, then in the subsequent SPATs the sequence A1→B1→C1 would lead to the larger reward, while the sequence A2→B2→C2 would lead to the smaller reward, and *vice versa*. Because in each block, subjects were taught in RITs whether C1 or C2 would be paired with the larger reward in the following SPATs, and because these associations changed randomly between blocks, experience from the previous block could not be effectively used to predict reward in the SPATs of the current block.

Pan et al. (2008) investigated whether monkeys would be able to transfer the reward information associated with C1 and C2 in the RITs to the first visual stimuli, A1 and A2, in the SPATs. Stimuli A1 and A2 were not directly paired with the different amounts of reward. However, if the monkeys could use both the SR relations (C with reward amount), and the stimulus–stimulus (A→B→C) associations, then after the RITs they should be able to predict reward amount at the time of the first stimulus presentation of A1 or A2 in a SPAT. On the contrary, if the monkeys just depended on the experience of SR relations from the previous block, their reward prediction should not necessarily be correct, particularly at the time of the first presentation of A1 or A2 in the first SPAT. Behaviorally, Pan et al. (2008) confirmed a significant decrease in performance in the first choice of the first SPAT (selection of B on the basis of A) when the trial used a sequence leading to smaller amount of reward (see Figure 1D in Pan et al., 2008). This shows that despite not yet receiving the reward itself, performance differed right from the first stimulus presentation depending on reward size (and therefore probably motivation). This indicates that monkeys were able to infer which reward condition they were currently experiencing right from the first stimulus presentation after reward instruction of C1 and C2.

Two different neuronal response patterns in the sequential paired-association task with an asymmetric reward schedule can be predicted based on the model-based and model-free learning processes. Using model-based learning processes (Daw et al., 2005), relevant brain areas should represent stimulus–stimulus associations acquired in the task in a tree-search manner, i.e., A1→B1→C1 and A2→B2→C2. Once C1 has been paired with the large reward in RITs and A1 is then presented in the SPAT as the first cue, the model-based system would search the tree-structure from A1 to B1, from B1 to C1 and from C1 to the large reward, and thereby predict that A1 would be associated with the large reward. In contrast, the model-free system does not store stimulus–stimulus associations; instead it saves a "cached" reward value associated with each stimulus (Daw et al., 2005). For example, when the brain experiences that in the current block the A1-sequence is paired with the large reward and the other sequence with the small reward, a larger value (e.g., 1) is assigned to the stimuli A1, B1, and C1, and a smaller value (e.g., 0) to the stimuli A2, B2, and C2. In the next block, the brain then learns that the stimulus C1 is paired with the small reward in RITs and the value of C1 is changed from 1 to 0, however the values of A1 and A2 remain yet unchanged. When A1 is then presented in SPATs, the model-free system would predict that it would lead to the large reward because A1 is still associated with the larger reward value. Overall, the model-based system is expected to predict reward information for the first stimulus (A1 or A2) on the basis of associations between the stimulus (C1 or C2) and reward acquired in RITs in the current block, while the model-free system is expected to predict reward information for the first stimulus on the basis of experience from the previous block. By considering the neuronal activity recorded from the LPFC and the striatum in the sequential paired-association task with the asymmetric reward schedule, we can verify whether the neurons in these areas use the model-based learning process or the model-free learning process.

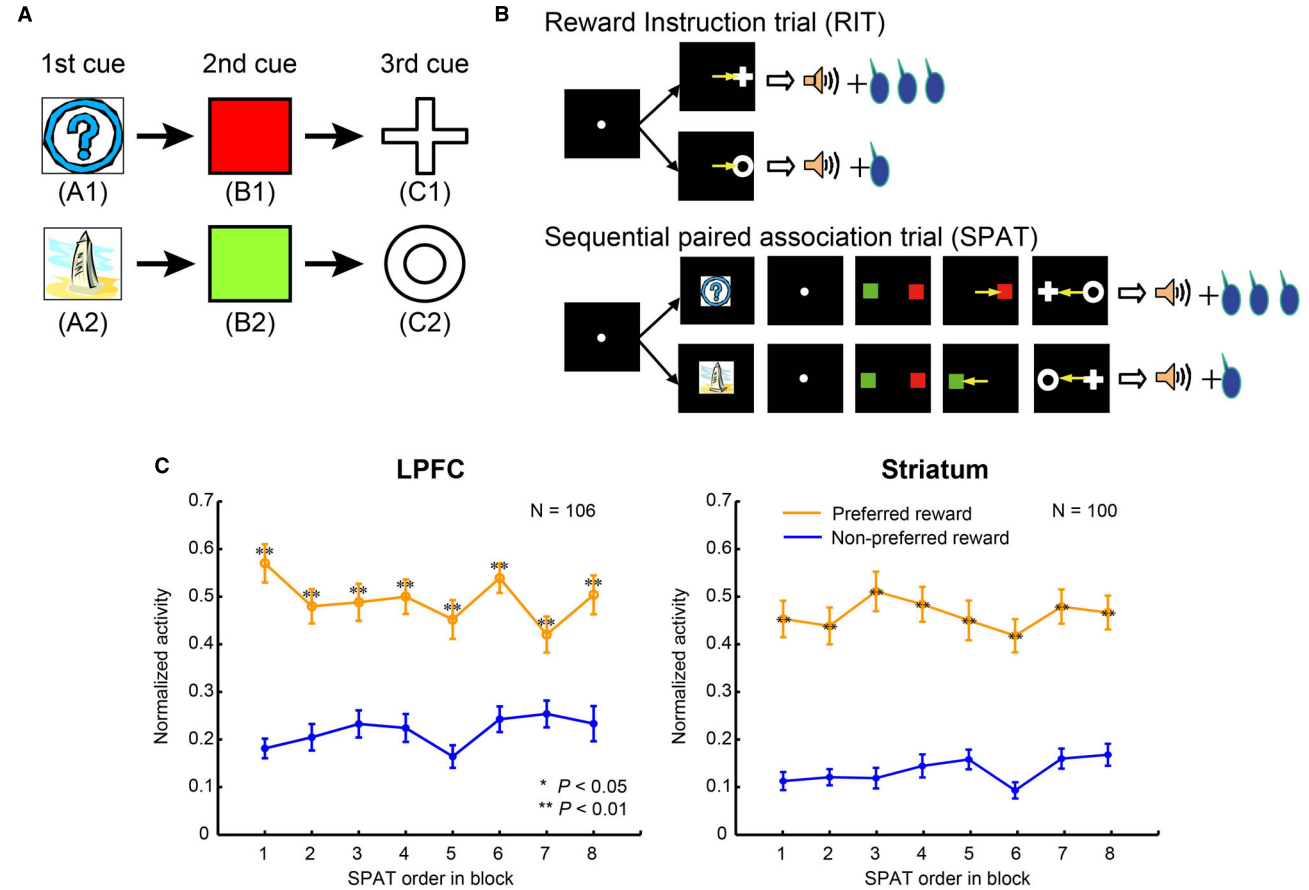


FIGURE 2 | The sequential paired-association task and neuronal activities in the LPFC and striatum. (A) Two stimulus–stimulus associative sequences were learned by monkeys. These six stimuli were denoted as "old stimuli"; the monkeys extensively experienced them with different amounts of rewards. **(B)** An asymmetric reward schedule was used in each block in which RITs were presented first and then followed by SPATs. The stimulus-reward contingency was the same in RITs and SPATs within each block, but this was randomized between blocks. The yellow arrows indicate saccadic eye movements and were not actually shown in the experiment. **(C)** Population activity of reward neurons to the old stimuli (A1 and A2) when they were presented as the first cue in SPATs. The left column shows neuronal activity in the LPFC, and the right column shows the activity in the striatum. The x-axis in the figure indicates SPAT order in the block directly after RITs. The normalized activity of each neuron was sorted into the preferred reward condition (orange curves) and the non-preferred reward condition (blue curves), and this was averaged across the population of reward neurons. Statistical significance was determined by Mann–Whitney U test, $^{**}P < 0.01$. Error bars indicate SEM.

Ability to Use State Transition to Predict Reward

Pan et al. (2008) recorded single-unit activity in the LPFC and striatum of monkeys performing this task (for recording sites, see Figures 3 and 4 in Pan et al., 2014). The majority of reward neurons, which were defined as showing differential averaged activity for stimuli indicating different amounts of reward, modulated their activity at the time of the first stimulus presentation in a SPAT (229/546 recorded neurons in the LPFC and 188/366 in the striatum). Results showed that LPFC neurons discriminated the large reward condition from the small reward condition right from the first SPATs (**Figure 2C**, left), indicating that LPFC neurons performed in a model-based manner. In addition to this, striatal neurons also distinguished the two reward conditions from the first SPATs in one block (**Figure 2C**, right), inconsistent with the predicted response pattern from model-free process. The findings that even striatal neurons could correctly predict rewards right from the first SPAT immediately after RITs indicate that striatal neurons also possess some information about state transition of stimuli in the SPAT task. Therefore, it is reasonable to suggest that the striatal neurons, in addition to the LPFC neurons, perform reward prediction in a model-based manner.

Model-based signals in the striatum have been found in both the state transition task and the sequential paired-association task (Daw et al., 2011; Lee et al., 2014; Pan et al., 2014), suggesting that the striatum may not simply use mode-free learning rules to predict reward. Neither of these two tasks could dissociate reward prediction functions in the LPFC and striatum. However, when we examined these two tasks carefully, we found subjects to have extensively experienced state transition, stimulus–stimulus and SR associations. Therefore, it is possible that striatal neurons simply utilize memorized relations (experiences) to predict

reward. The behavior and neuronal activity patterns that can be predicted based on memorized experiences should be similar to results based on the model-based strategy. For example, the sequential paired-association task with an asymmetric reward schedule was repeatedly performed using six fixed stimuli: A1, B1, C1, A2, B2, and C2. There were four conditioned SR associations in the task, (1) C1→LR (large reward), A1→LR and A2→SR (small reward); (2) C1→SR, A1→SR and A2→LR; (3) C2→LR, A1→SR and A2→LR; (4) C2→SR, A1→LR and A2→SR. The monkeys extensively experienced each of these associations. If the monkeys and neurons memorized each of the conditioned SR associations, it would be easy for them to determine which stimulus (A1 or A2) would be paired with a large reward after reward instruction with C1 or C2. In that case, the responses of the LPFC and striatal neurons could be explained by the reward prediction based on memorized experiences. Some studies have shown that the activity of dopamine neurons could be modulated by memorized experiences of state transitions and SR associations to represent RPE and play a role in reward learning in the striatum (Nakahara et al., 2004; Bromberg-Martin et al., 2010; Enomoto et al., 2011). These results together suggest that in order to disassociate reward prediction functions in the LPFC and striatum, it is important to consider the effect of experience with reward-stimulus associations in the task. To investigate this issue, Pan et al. (2014) conducted a study where the monkeys were required to predict reward for a stimulus without any direct reward experience.

Ability to Predict Reward Without Experience

Pan et al. (2014) conducted an experiment where session-unique stimuli were introduced into the above task and tested whether the monkey's behavior and LPFC and striatal neurons could carry out similar reward prediction with newly introduced stimuli. The monkeys were trained to learn new stimulus associations in a delayed matching-to-sample task with a symmetric reward schedule (**Figure 3A**). The new stimuli were learned to be associated with one of the two color patches (B1 or B2). These newly learned stimuli are referred to as "new stimuli," while the stimuli A1, B1, C1, A2, B2, and C2 are referred to as "old stimuli." In total, the monkeys learned 924 new stimuli (462 new stimulus pairs) to be associated with the color patches (B1 or B2). For ease of explanation, the new stimulus that was randomly selected from each pair to be presented in the very first SPAT of the relevant block shall be referred to as N1, and the second new stimulus of each pair shall be referred to as N2; however it is important to note that there is not simply one N1 stimulus but 462 N1 stimuli, and the same for N2 stimuli ($N1_1$–$N1_{462}$ and $N2_1$–$N2_{462}$). The newly learned stimuli were classified into two groups according to the old stimuli that they were associated with. The new stimuli associated with B1 were classified into the A1-group and the new stimuli associated with B2 were classified into the A2-group (A1, B1, and C1 belonged to A1 group and A2, B2, and C2 to A2 group; **Figure 3A**). Up to this point, the monkeys had experienced no direct associations between new stimuli and C1 or C2, and also no information about the asymmetric reward schedule with respect to the new stimuli (**Figure 3B**, upper panel). After having fully acquired the new associations, the monkeys performed the reward instructed sequential paired-association task with the new stimuli (**Figure 3B**, middle panel). This was identical to the reward instructed SPATs with old stimuli except that in these SPATs a newly learned stimulus was presented as the first cue instead of an old stimulus (A1 or A2). Behaviorally, the monkeys showed significantly higher performance when first choosing from new stimuli in the large than in the small reward trials (see Figure 2 in Pan et al., 2014). This indicates that the monkeys correctly predicted the reward information for the first new stimulus (N1 or N2) that was presented in SPATs based on the reward information associated with C1 or C2 in RITs, without the requirement of direct associations between the reward information and the new stimuli.

Reward-related neurons from the LPFC and striatum were recorded while monkeys performed the reward instructed SPAT task with new stimuli. Almost all the reward-related neurons in the LPFC and striatum, which showed reward differential activity for A1 and A2 in SPATs with old stimuli, showed similar reward differential activity with new stimuli (at least on average). When Pan et al. (2008) concentrated on the single-unit activity to the first new stimulus (N1; independent of the stimulus group it belonged to, **Figure 3B**, bottom panel) in the very first SPAT just after the RITs, reward-related neurons in the striatum did not show differential activity regardless of whether N1 predicated large or small reward (**Figure 3C**, right), however reward-related neurons in the LPFC did (**Figure 3C**, left). The striatal reward-related neurons did discriminate the two reward conditions from the second presentation of N1. These results seem to show that the striatum needs direct experience of SR associations to predict reward. However, when Pan et al. (2008) looked at neuronal activity to the second new stimulus (N2) when it was first presented in the SPAT (after one or two presentations of N1), even the striatal neurons could show reward differential activity (**Figure 3D**). This result indicates that in a new stimulus pair (N1 and N2) striatal neurons are able to infer the reward amount of N2 (after receiving reward information about N1) without direct experience of which reward amount would follow N2. This type of function is called a disjunctive inference (Johnson-Laird et al., 1992).

Lateral prefrontal cortex neurons could predict the reward amount of a new stimulus from the very first SPAT (just after the RITs). This result cannot be explained by the reward prediction based on memorized experiences because at this point the monkeys had no past experience of the appropriate SR assignment. Therefore, this result indicates that neurons in the LPFC had the ability to combine the results of two associations to predict future outcomes. This ability is called a transitive inference. In this experiment LPFC neurons combined the association of the new stimuli with C (through B), with the association of C with reward size, to predict the reward amount. Striatal neurons were unable to combine these stimulus–stimulus and SR associations to predict reward. Instead striatal neurons could predict the reward amount (e.g., small) of the second new stimulus (N2) from a pair after directly experiencing the alternative stimulus (N1) with the alternative amount of reward

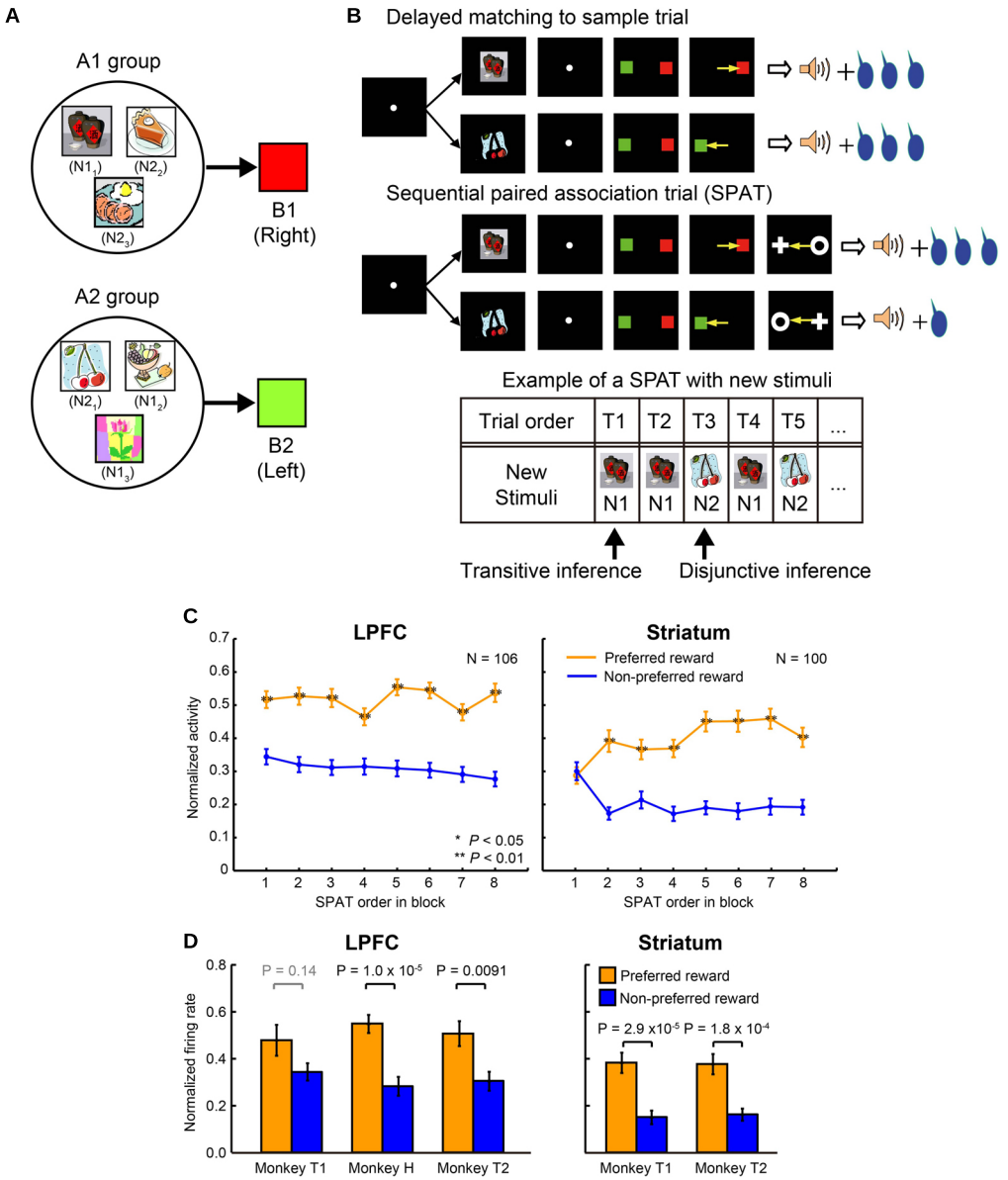


FIGURE 3 | Transitive and disjunctive reward inferences in the LPFC and striatum. (A) Examples of new stimuli associated with the two color patches (B1 and B2). New stimuli that were associated with B1 were classified into the A1-group, and new stimuli that were associated with B2 were classified into the A2-group. **(B)** Reward inference in the sequential paired-association task with new stimuli. The monkeys learned new associations in a delayed matching-to-sample task in which the positions of B1 and B2 were fixed (e.g., B1 was always presented on the right side of the fixation spot, while B2 was always presented on the left side). The procedure of the asymmetric reward schedule is almost same as those shown in **Figure 2B**, except that a new stimulus were presented as the first cue instead of an old stimulus (A1 or A2) in SPATs. In the very first SPAT after reward instruction trials, the monkeys and neurons had no direct experience of new stimuli and asymmetric reward associations. The bottom panel shows an example new stimulus pair presented in a block. After RITs, one of the paired stimuli (N1) was presented as the first cue in the first (T1) and second (T2) SPATs. The other new stimulus of the pair (N2) was presented in the third SPAT (T3), and so on. If neurons could predict reward information for N1 in the very first SPAT (T1), one could say that these neurons have the ability of transitive reward inference (C1→reward and N1→B1→C1, so N1→reward). If neurons cannot predict reward information for N1 in T1 trials, but can infer reward information for N2 presented T2/T3 trials after directly experiencing the N1-reward contingency, one could say that these neurons have the ability of disjunctive reward inference. **(C)** Population activities of reward neurons to new stimuli as a function of SPAT order in the first block where new stimuli were presented for the first time. LPFC neurons discriminated the two reward conditions (preferred vs. non-preferred) from the very first SPAT (T1) after reward instruction trials (left column). Striatal neurons did not distinguish the preferred reward condition from the non-preferred reward condition in T1 trials. The orange curves represent the preferred reward condition, and the blue curves represent the non-preferred reward condition. Statistical significance was determined by Mann–Whitney U tests, $^{*}P < 0.05$, $^{**}P < 0.01$. Error bars indicate SEM. **(D)** Both the LPFC (left column) and striatal (right column) activities to the second new stimulus (N2) discriminated the reward conditions in T2/T3 trials after direct experience with the first new stimulus (N1) and with the reward contingency. The orange bars indicate the preferred reward condition and the blue bars indicate the non-preferred reward condition. Statistical significance was calculated by Wilcoxon rank sum test. Error bars indicate SEM.

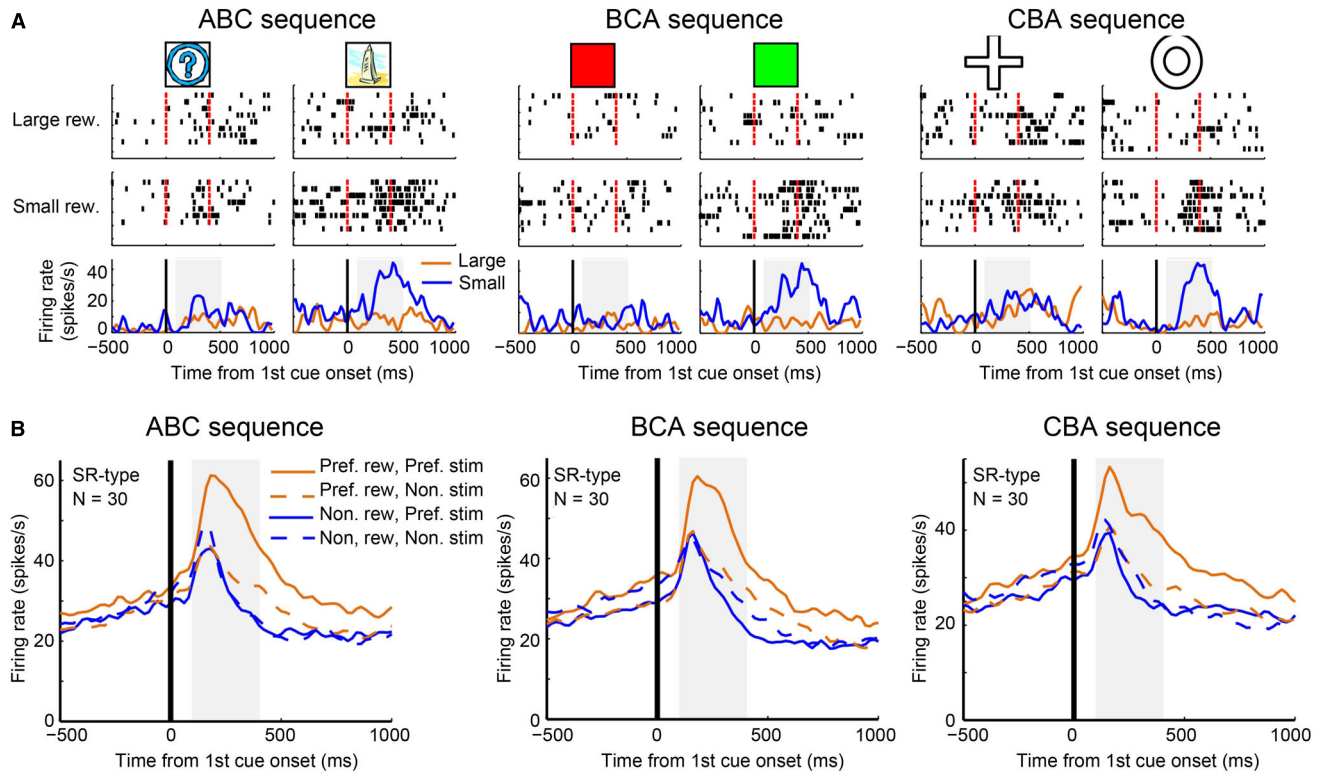


FIGURE 4 | Example activity of a stimulus-reward neuron and population activities. (A) Single-unit activity of a stimulus-reward neuron for three sequences: ABC (A1 vs. A2), BCA (B1 vs. B2), and CAB (C1 vs. C2) sequences. In each sequence, all trials were sorted by four conditions—the first cue stimulus (i.e., A1 vs. A2) and the two reward conditions (larger reward vs. small reward)—and aligned with the first cue onset. In the raster-grams, red ticks mark the onset and offset of the first cue. In the histograms, the orange curves represent data from large-reward trials and the blue curves represent data from small-reward trials. **(B)** The population histograms of stimulus-reward neurons for three sequences. Trials of each stimulus-reward type cell were sorted by four conditions: preferred reward condition and preferred stimulus (solid orange curves), preferred reward condition and non-preferred stimulus (dashed orange curves), non-preferred reward condition and preferred stimulus (solid blue curves), and non-preferred reward condition and non-preferred stimulus (dashed blue curves).

(e.g., large). This result indicates that the striatal neurons can perform disjunctive inferences. Congruent with this finding, Bromberg-Martin et al. (2010) reported that dopamine neurons, which have a close relationship with reward prediction in the striatum, are able to use disjunctive inferences to generate RPE signals. Furthermore, disjunctive inference is similar to a key function used for establishing the model-based process in the state transition task performed by Daw et al. (2011). The evidence that the nigro-striatal network is involved in reward prediction via disjunctive inference indicates that the model-based vs. model-free process distinction is not simply equivalent to dissociation in LPFC and striatal functions.

Reward Inference by Abstract Neural Code

We further investigated why it may be that LPFC neurons can perform transitive reward inference while striatal neurons cannot. Pan et al. (2008) found a subgroup of reward neurons (SR type) in the LPFC that showed differential reward activity for only one of the first stimuli (A1 or A2). Originally, Pan et al. (2008) used an ABC sequence (A1→B1→C1 and A2→B2→C2) for recording and looked at the activity pattern to stimulus A (A1 or A2). However, by investigating only this sequence, they could not tell whether SR type neurons reflect categorical information or whether they simply reflect visual properties of the first cues. To address this, Pan et al recorded activity of SR neurons in the LPFC with another two sequential associations: the BCA sequence (B1→C1→A1 and B2→C2→A2) and the CAB sequence (C1→A1→B1 and C2→A2→B2). The majority of SR neurons showed reward-differential activity only for a group of relevant visual stimuli (e.g., A1, B1 and C1, or A2, B2 and C2; **Figure 4A**). This tendency was confirmed by the population activity of SR neurons (**Figure 4B**). Therefore, these neurons (hereby referred to as "category-reward" neurons) likely coded both the category information of visual stimuli (either A1 or A2 group), and reward information (either large or small), simultaneously.

In related literature, many studies have reported that neurons in the LPFC code categorical information (Freedman et al., 2001; Shima et al., 2007; Meyers et al., 2008; Cromer et al., 2010; Seger and Miller, 2010). Sakagami and Tsutsui (1999) trained monkeys to make a go response to, for example green and purple colors, and a no-go response to, for example red and yellow colors (Sakagami and Tsutsui, 1999). Many neurons in the ventrolateral PFC showed go/no-go differential activity based on

color (color go/no-go neurons). A majority of them also showed color grouping. For example if a neuron showed differentially increased activity to green, then the same neuron tended to show the same activity pattern to purple, while if a neuron increased activity to red, then the same neuron tended to show the same activity pattern to yellow. Activity of these neurons did not simply reflect go/no-go discrimination because the same neurons did not show differential activity when the monkeys performed the same go/no-go task with motion cues. As the LPFC is located between sensory output areas and motor execution areas, the task of the LPFC is likely related to the conversion of sensory information to motor information (Sakagami and Pan, 2007). If the area has several hierarchical stages for this process, it is natural that some neurons in the LPFC should code both sensory and motor information in a manner similar to the "abstract" coding seen in both reward-category neurons and in the color go/no-go neurons. In support of this idea, Shima et al. (2007) found "motor-category" neurons in the LPFC.

It is an interesting question to ask whether or not striatal neurons encode category information. Evidence against this was found in a recent study (Antzoulatos and Miller, 2011). In this study, Antzoulatos and Miller (2011, 2014) compared neuronal activity patterns for the LPFC and striatum during a dot-based shape category learning task in which the monkeys learned to associate stimuli from one category with a right saccade and stimuli from the other category with a left saccade. In each recorded session, two novel dot-pattern prototypes (two categories) were introduced. In the first block, a single stimulus per category was presented and the monkeys learned the relevant stimulus-response associations. In following blocks, more and more new stimuli were added. The monkeys were unable to learn each stimulus-response association; instead, they learned category-response associations. It was found that striatal neurons represented stimulus-response association in the early learning stage. In the late learning or category performing stage, LPFC neurons encoded category-response associations but striatal neurons did not represent such category information. These results suggest that striatal neurons do not classify new category members into a group or represent their category information.

In the sequential paired-association task of Pan et al. (2008) the monkeys might have, through extensive training, grouped stimuli requiring the same response together as a functional category according to intended behavior. Some LPFC neurons appeared to represent category information of those associated stimuli. It is known that LPFC neurons also receive reward information from the OFC and subcortical areas, such as the striatum, amygdala, and dopaminergic neurons in the midbrain (Schultz, 2000; Wallis and Miller, 2003; Sakagami and Watanabe, 2007). Therefore, some LPFC neurons involved in categorization might also receive reward information and thereby function as category-reward neurons. In the sequential paired-association task with new stimuli, the monkeys were unable to rely on rote memory to predict the amount of reward for the new stimuli because they had not yet been directly taught associations between the new stimuli and reward. Effectively, the monkeys had to integrate independently acquired associations to infer the reward value of new stimuli. The category-reward neurons may be what the brain uses to fulfill this integration function. Each member from the preferred category was found to evoke similar response patterns in category-reward neurons; this processing may be the way in which relations between reward and each member in the same category were established. If a newly introduced stimulus is a functional member of a given category, then the category information of the new stimulus and the reward information acquired in RITs could together activate the category-reward neurons. This activity of the category-reward neurons may allow reward neurons in the LPFC to infer the reward information of the new stimuli. The striatal neurons were unable to predict reward for the first new stimuli presented in the first SPATs. As shown above, the striatal neurons were not able to use this categorical code approach. Instead they likely used memorized experiences to know which reward N2 must be associated with after directly experiencing the alternative reward in association with N1. Our suggestion that the LPFC but not the striatum may be capable of categorical coding is reinforced by the finding that LPFC neurons showed categorical related activity whereas striatal neurons showed response related activity in a category learning task (Antzoulatos and Miller, 2011, 2014). Overall, reward prediction neurons in the LPFC, but not those in the striatum, were able to predict the reward amount of newly introduced stimuli at their first presentation in SPATs possibly because the LPFC is capable of categorical coding whereas the striatum is not (**Figure 5**).

Discussion and Conclusion

Here, we suggest that LPFC neurons can perform the categorization process. The categorization process is regarded as the process utilized to determine which things belong together. In a related study three types of categories were proposed: the "perceptual category," the "relational category," and the "association category" (Zentall et al., 2002). The categorization process we discussed here is likely to have been of the associative type, in which shared functions of the stimuli are important instead of the physical properties of them. The LPFC encodes associative categorization (Pan et al., 2008), and may utilize it as a model to simulate or predict reward information for both well learned old stimuli and newly introduced stimuli. This process does not require the animals and neurons to directly experience associations between stimuli and reward. The striatum did not represent the category information as a model; instead, it might have stored paired stimuli-reward information to predict reward after directly experiencing the association between one stimulus of a pair and its associated reward. The functional difference between the LPFC and striatum might not simply rely on model-based vs. model-free learning rules. It might instead rely on whether and how the two areas integrate experiences with current task state information and use different strategies to predict reward.

The probabilistic state transition task performed by Daw et al. (2011) and the sequential paired-association task with old stimuli by Pan et al. (2008) both consist of several sequences of stimuli or actions. In both tasks, to obtain reward at the end of trial, it is

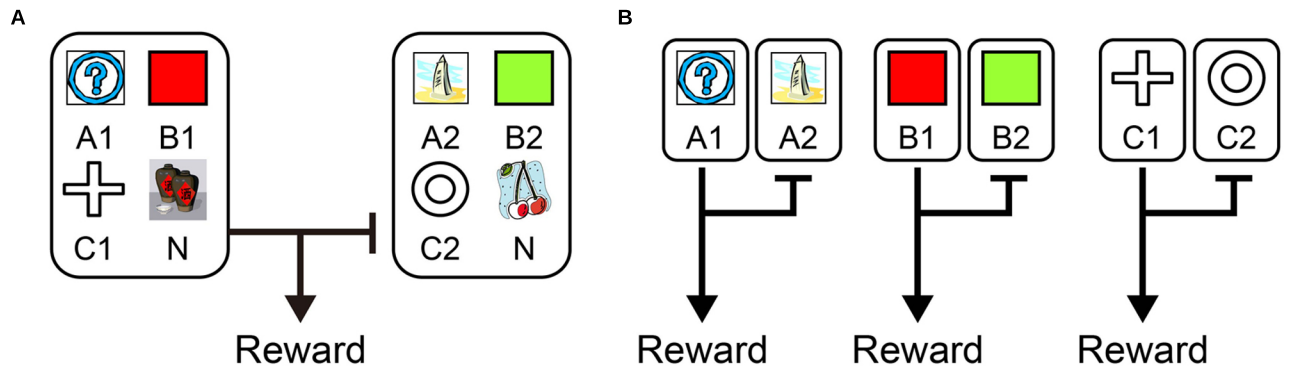


FIGURE 5 | Schematic drawing of the function of LPFC and striatal neurons. (A) LPFC neurons can use knowledge of categorical relationships to predict reward. **(B)** Striatal neurons can code the individual associations between stimuli and reward.

necessary to assign the potential reward to early stimuli or actions properly. This credit assignment program has been discussed within the framework of explicit/implicit learning processes (Fu and Anderson, 2008). Fu and Anderson (2008) performed a state transition task similar to the task performed by Daw et al. (2011) and showed that the subjects displayed faster state learning when explicit memory was dominantly relied upon and that implicit reinforcement learning required state information. Therefore, these explicit/implicit learning processes are presumably related to the model-based/model-free learning processes. On the other hand, in Pan et al.'s (2008) sequential paired-association task with new stimuli, the monkeys' behaviors and LPFC activity showed reward related activities without the direct experience of obtaining reward after presentations of these new stimuli. This result indicates that reward value can be assigned to the stimuli without experience. It is possible to explain this by suggesting that reward value is assigned to abstract information, such as the functional category, rather than to representation of the individual stimulus. Therefore, experience of reward with any member of a category causes reward to be assigned to this whole category, thereby making it possible for one to be able predict reward for another member of the same category without direct experience.

Where could the model-free learning process be performed in the brain? We believe that the model-free reinforcement learning process is the basis of the reward related learning. Furthermore, a lot of studies support the idea of the existence of a circuit for reinforcement learning in the nigro-striatal circuit (Schultz et al., 1997; Daw et al., 2005; Doya, 2008). Therefore, the model-free learning process is likely performed in parts of the striatum. However, even if the nigro-striatal circuit simply performs the model-free reinforce learning process, when it receives signals calculated in the prefrontal cortex with the model-based process, its activity then appears to perform the model-based learning process. In other words, the learning process in the striatum depends on the signal sent to the striatum. Recently, it was proposed that the LPFC works as an arbitrator of model-based and model-free strategies (Lee et al., 2014). Therefore, it is possible that the LPFC controls the signal sent to the striatum and allocates the degree of control over behavior determined by model-based and model-free systems. Nonetheless, the degree to which we can understand the learning process performed in the striatum based simply on subjects' behaviors and neural activities may be limited. To precisely understand the learning process performed in the striatum, it would be necessary to reveal the signal processing mechanism in the striatum using methods which can examine the circuit mechanism directly, i.e., optogenetics and the DREADD (Designer Receptors Exclusively Activated by Designer Drugs) system.

There remains no doubt about the existence of the model-based learning process in the LPFC. Here, we extend this idea by specifically proposing the existence of abstraction or categorization in the LPFC. While this can explain the SR associations found in the SPAT task, more precise neurophysiological research is required to explain the model-based learning process in the state transition task. Furthermore, it is suggested that the LPFC also contributes to other complex cognitive processes which are involved in the model-based system (Yoshida et al., 2010; Donoso et al., 2014). Further study is necessary to extend understanding of the specific function of the LPFC in model-based learning processes.

In conclusion, to clarify the functional differences in reward prediction between the LPFC and striatum in the monkey brain, we compared activity patterns of neurons in these two areas mainly from studies using a sequential paired association task with an asymmetric reward schedule. To predict reward, both LPFC and striatal neurons were able to use knowledge about state transitions. LPFC neurons could predict reward via transitive inference and striatal neurons could predict reward via disjunctive inference even in previously unexperienced situations. These results suggest the existence of a model-based system in both the LPFC and the striatum. These results also indicate that the model-based vs. model-free hypothesis is not sufficient to explain the functional difference between the LPFC and striatum. Instead, the difference seems to be that the LPFC neurons can utilize abstract code (in this case stimulus categorization; **Figure 5A**) to associate a stimulus with a reward, whereas while the striatal neurons can code the individual associations between a stimulus and a reward (or sequence of stimuli-reward, **Figure 5B**), they cannot use abstract code.

Acknowledgments

This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas (4303) from MEXT (Ministry of Education, Culture, Sports, Science and Technology) of Japan (http://decisions.naist.jp/index.html) and partially supported by the Strategic Research Program for Brain Sciences from Japan Agency for Medical Research and development, AMED. XP is supported by National Natural Science Foundation of China (No. 11232005, No. 11472104) and sponsored by Shanghai Pujiang Program (No. 13PJ1402000). We thank Sanae Harada for her help on writing the manuscript.

References

Amemori, K., and Sawaguchi, T. (2006). Rule-dependent shifting of sensorimotor representation in the primate prefrontal cortex. *Eur. J. Neurosci.* 23, 1895–1909. doi: 10.1111/j.1460-9568.2006.04702.x

Antzoulatos, E. G., and Miller, E. K. (2011). Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. *Neuron* 71, 243–249. doi: 10.1016/j.neuron.2011.05.040

Antzoulatos, E. G., and Miller, E. K. (2014). Increases in functional connectivity between prefrontal cortex and striatum during category learning. *Neuron* 83, 216–225. doi: 10.1016/j.neuron.2014.05.005

Asaad, W. F., and Eskandard, E. N. (2011). Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. *J. Neurosci.* 31, 17772–17787. doi: 10.1523/JNEUROSCI.3793-11.2011

Balleine, B. W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37, 407–419. doi: 10.1016/S0028-3908(98)00033-1

Balleine, B. W., and O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69. doi: 10.1038/npp.2009.131

Behrens, T. E., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221. doi: 10.1038/nn1954

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. S. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Beierholm, U. R., Anen, C., Quartz, S., and Bossaerts, P. (2011). Separate encoding of model-based and model-free valuations in the human brain. *Neuroimage* 58, 955–962. doi: 10.1016/j.neuroimage.2011.06.071

Blaisdell, A. P., Sawa, K., Leising, K. J., and Waldmann, M. R. (2006). Causal reasoning in rats. *Science* 311, 1020–1022. doi: 10.1126/science.1121872

Bromberg-Martin, E. S., Matsumoto, M., Nakahara, H., and Hikosaka, O. (2010). Multiple timescales of memory in lateral habenula and dopamine neurons. *Neuron* 67, 499–510. doi: 10.1016/j.neuron.2010.06.031

Bunge, S. A., Kahnm, I., Wallis, J. D., Miller, E. K., and Wagner, A. D. (2003). Neural circuits subserving the retrieval and maintenance of abstract rules. *J. Neurophysiol.* 90, 3419–3428. doi: 10.1152/jn.00910.2002

Cromer, J. A., Roy, J. E., and Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron* 66, 796–807. doi: 10.1016/j.neuron.2010.05.005

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215. doi: 10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711. doi: 10.1038/nn1560

Deserno, L., Huys, Q. J., Boehme, R., Buchert, R., Heinze, H. J., Grace, A. A., et al. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1595–1600. doi: 10.1073/pnas.1417219112

Diekhof, E. K., Kaps, L., Falkai, P., and Gruber, O. (2012). The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude—an activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. *Neuropsychologia* 50, 1252–1266. doi: 10.1016/j.neuropsychologia.2012.02.007

Doll, B. B., Simon, D. A., and Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.* 22, 1075–1081. doi: 10.1016/j.conb.2012.08.003

Donoso, M., Collins, A. G., and Koechlin, E. (2014). Human cognition. Foundations of human reasoning in the prefrontal cortex. *Science* 344, 1481–1486. doi: 10.1126/science.1252254

Doya, K. (2008). Modulators of decision making. *Nat. Neurosci.* 11, 410–416. doi: 10.1038/nn2077

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nat. Rev. Neurosci.* 2, 820–829. doi: 10.1038/35097575

Enomoto, K., Matsumoto, N., Nakai, S., Satoh, T., Sato, T. K., Ueda, Y., et al. (2011). Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proc. Natl. Acad. Sci. U.S.A.* 108, 15462–15467. doi: 10.1073/pnas.1014457108

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316. doi: 10.1126/science.291.5502.312

Fu, W. T., and Anderson, J. R. (2008). Solving the credit assignment problem: explicit and implicit learning of action sequences with probabilistic outcomes. *Psychol. Res.* 72, 321–330. doi: 10.1007/s00426-007-0113-7

Garrison, J., Erdeniz, B., and Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neurosci. Biobehav. Rev.* 37, 1297–1310. doi: 10.1016/j.neubiorev.2013.03.023

Glascher, J., Daw, N., Dayan, P., and O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585–595. doi: 10.1016/j.neuron.2010.04.016

Haber, S. N., and Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35, 4–26. doi: 10.1038/npp.2009.129

Han, S., Huettel, S. A., and Dobbins, I. G. (2009). Rule-dependent prefrontal cortex activity across episodic and perceptual decisions: an fMRI investigation of the criterial classification account. *J. Cogn. Neurosci.* 21, 922–937. doi: 10.1162/jocn.2009.21060

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., and 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. doi: 10.1523/JNEUROSCI.1309-08.2008

Hayden, B. Y., Nair, A. C., McCoy, A. N., and Platt, M. L. (2008). Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron* 60, 19–25. doi: 10.1016/j.neuron.2008.09.012

Hikosaka, O., Bromberg-Martin, E., Hong, S., and Matsumoto, M. (2008). New insights on the subcortical representation of reward. *Curr. Opin. Neurobiol.* 18, 203–208. doi: 10.1016/j.conb.2008.07.002

Hikosaka, O., and Isoda, M. (2010). Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. *Trends Cogn. Sci.* 14, 154–161. doi: 10.1016/j.tics.2010.01.006

Histed, M. H., Pasupathy, A., and Miller, E. K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63, 244–253. doi: 10.1016/j.neuron.2009.06.019

Hollerman, J. R., Tremblay, L., and Schultz, W. (1998). Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J. Neurophysiol.* 80, 947–963.

Hoshi, E., Shima, K., and Tanji, J. (2000). Neuronal activity in the primate prefrontal cortex in the process of motor selection based on two behavioral rules. *J. Neurophysiol.* 83, 2355–2373.

Joel, D., Niv, Y., and Ruppin, E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw.* 15, 535–547. doi: 10.1016/S0893-6080(02)00047-3

Johnson-Laird, P. N., Byrne, R. M., and Schaeken, W. (1992). Propositional reasoning by model. *Psychol. Rev.* 99, 418–439. doi: 10.1037/0033-295X.99.3.418

Kahnt, T., Heinzle, J., Park, S. Q., and Haynes, J. D. (2011). Decoding the formation of reward predictions across learning. *J. Neurosci.* 31, 14624–14630. doi: 10.1523/JNEUROSCI.3412-11.2011

Killcross, S., and Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex* 13, 400–408. doi: 10.1093/cercor/13.4.400

Knutson, B., and Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* 18, 411–417. doi: 10.1097/01.wco.0000173463.24758.f6

Kobayashi, S., Kawagoe, R., Takikawa, Y., Koizumi, M., Sakagami, M., and Hikosaka, O. (2007). Functional differences between macaque prefrontal cortex and caudate nucleus during eye movements with and without reward. *Exp. Brain Res.* 176, 341–355. doi: 10.1007/s00221-006-0622-4

Lee, S. W., Shimojo, S., and O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 687–699. doi: 10.1016/j.neuron.2013.11.028

Levy, D. J., and Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Curr. Opin. Neurobiol.* 22, 1027–1038. doi: 10.1016/j.conb.2012.06.001

Maia, T. V. (2009). Reinforcement learning, conditioning, and the brain: successes and challenges. *Cogn. Affect. Behav. Neurosci.* 9, 343–364. doi: 10.3758/CABN.9.4.343

McClure, S. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science* 306, 503–507. doi: 10.1126/science.1100907

McGonigle, B. O., and Chalmers, M. (1977). Are monkeys logical? *Nature* 267, 694–696. doi: 10.1038/267694a0

McNamee, D., Liljeholm, M., Zika, O., and O'Doherty, J. P. (2015). Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate FMRI study. *J. Neurosci.* 35, 3764–3771. doi: 10.1523/JNEUROSCI.4677-14.2015

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., and Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* 100, 1407–1419. doi: 10.1152/jn.90248.2008

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nat. Rev. Neurosci.* 1, 59–65. doi: 10.1038/35036228

Murphy, R. A., Mondragón, E., and Murphy, V. A. (2008). Rule learning by rats. *Science* 319, 1849–1851. doi: 10.1126/science.1151564

Nakahara, H., Itoh, H., Kawagoe, R., Takikawa, Y., and Hikosaka, O. (2004). Dopamine neurons can represent context-dependent prediction error. *Neuron* 22, 269–280. doi: 10.1016/S0896-6273(03)00869-9

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454. doi: 10.1126/science.1094285

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337. doi: 10.1016/S0896-6273(03)00169-7

Padoa-Schioppa, C., and Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226. doi: 10.1038/nature04676

Pasupathy, A., and Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433, 873–876. doi: 10.1038/nature03287

Pan, X., Fan, H., Sawa, K., Tsuda, I., Tsukada, M., and Sakagami, M. (2014). Reward inference by primate prefrontal and striatal neurons. *J. Neurosci.* 34, 1380–1396. doi: 10.1523/JNEUROSCI.2263-13.2014

Pan, X., and Sakagami, M. (2012). Category representation and generalization in the prefrontal cortex. *Eur. J. Neurosci.* 35, 1083–1091. doi: 10.1111/j.1460-9568.2011.07981.x

Pan, X., Sawa, K., Tsuda, I., Tsukada, M., and Sakagami, M. (2008). Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. *Nat. Neurosci.* 11, 703–712. doi: 10.1038/nn.2128

Paton, J. J., Belova, M. A., Morrison, S. E., and Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* 439, 865–870. doi: 10.1038/nature04490

Rangel, A., Camerer, C., and Montague, R. (2008). A framework for studying the neurobiology ofvalue-based decision-making. *Nat. Rev. Neurosci.* 9, 545–556. doi: 10.1038/nrn2357

Roesch, M. R., and Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J. Neurophysiol.* 90, 1766–1789. doi: 10.1152/jn.00019.2003

Rushworth, M., Noonan, M., Boorman, E., Walton, M., and Behrens, T. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron* 70, 1054–1069. doi: 10.1016/j.neuron.2011.05.014

Rygula, R., Walker, S. C., Clarke, H. F., Robbins, T. W., and Roberts, A. C. (2010). Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning. *J. Neurosci.* 30, 14552–14559. doi: 10.1523/JNEUROSCI.2631-10.2010

Sakagami, M., and Niki, H. (1994). Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: relation to relevant task conditions. *Exp. Brain Res.* 97, 423–436. doi: 10.1007/BF00241536

Sakagami, M., and Pan, X. (2007). Functional role of the ventrolateral prefrontal cortex in decision making. *Curr. Opin. Neurobiol.* 17, 228–233. doi: 10.1016/j.conb.2007.02.008

Sakagami, M., and Tsutsui, K. (1999). The hierarchical organization of decision making in the primate prefrontal cortex. *Neurosci. Res.* 34, 79–89. doi: 10.1016/S0168-0102(99)00038-3

Sakagami, M., and Watanabe, M. (2007). Integration of cognitive and motivational information in the primate lateral prefrontal cortex. *Ann. N. Y. Acad. Sci.* 1104, 89–107. doi: 10.1196/annals.1390.010

Samejima, K., Ueda, Y., Doya, K., and Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science* 310, 1337–1340. doi: 10.1126/science.1115270

Schultz, W. (2000). Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1, 199–207. doi: 10.1038/35044563

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* 57, 87–115. doi: 10.1146/annurev.psych.56.091103.070229

Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599. doi: 10.1126/science.275.5306.1593

Seger, C. A., and Miller, E. K. (2010). Category learning in the brain. *Annu. Rev. Neurosci.* 33, 203–219. doi: 10.1146/annurev.neuro.051508.135546

Shima, K., Isoda, M., Mushiake, H., and Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature* 45, 315–318. doi: 10.1038/nature05470

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tanaka, S. C., Samejima, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S., et al. (2006). Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. *Neural Netw.* 19, 1233–1241. doi: 10.1016/j.neunet.2006.05.039

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208. doi: 10.1037/h0061626

Tricomi, E., Balleine, B. W., and O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *Eur. J. Neurosci.* 29, 2225–2232. doi: 10.1111/j.1460-9568.2009.06796.x

Vallentin, D., Bongard, S., and Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. *J. Neurosci.* 32, 6621–6630. doi: 10.1523/JNEUROSCI.5071-11.2012

Wallis, J. D., Anderson, K. C., and Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature* 411, 953–956. doi: 10.1038/35082081

Wallis, J. D., and Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* 18, 2069–2081. doi: 10.1046/j.1460-9568.2003.02922.x

Walsh, M. M., and Anderson, J. R. (2014). Navigating complex decision spaces: problems and paradigms in sequential choice. *Psychol. Bull.* 140, 466–486. doi: 10.1037/a0033455

Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature* 382, 629–632. doi: 10.1038/382629a0

White, I. M., and Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Exp. Brain Res.* 126, 315–335. doi: 10.1007/s002210050740

Wunderlich, K., Dayan, P., and Dolan, R. J. (2012). Mapping value based planning and extensively trained choice in the human brain. *Nat. Neurosci.* 11, 786–791. doi: 10.1038/nn.3068

Yamada, H., Matsumoto, N., and Kimura, M. (2004). Tonically active neurons in the primate caudate nucleus and putamen differentially encode instructed motivational outcomes of action. *J. Neurosci.* 24, 3500–3510. doi: 10.1523/JNEUROSCI.0068-04.2004

Yoshida, W., Seymour, B., Friston, K. J., and Dolan, R. J. (2010). Neural mechanisms of belief inference during cooperative games. *J. Neurosci.* 30, 10744–10751. doi: 10.1523/JNEUROSCI.5895-09.2010

Zentall, T. R., Galizio, M., and Critchfield, T. S. (2002). Categorization, concept learning and behavior analysis: an introduction. *J. Exp. Anal. Behav.* 78, 237–248. doi: 10.1901/jeab.2002.78-237

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Tanaka, Pan, Oguchi, Taylor and Sakagami. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.