Different Reward Processing in the Primate Prefrontal Cortex and Striatum Using a Reward Inference Task
INTRODUCTION .................................................................................................................. 3

MATERIALS AND METHODS ............................................................................................... 5
  SUBJECTS ......................................................................................................................... 5
  TASK ................................................................................................................................. 6
    Sequential paired-association task ................................................................................. 6
    Reward instruction trials .............................................................................................. 7
    Reward instructed sequential paired-association task with old stimuli ................. 9
    Training of new stimuli pairings ................................................................................. 9
    Reward instructed sequential paired-association task with new stimuli ........... 10
  DATA ACQUISITION ........................................................................................................ 10
    The recording sites in the LPFC and striatum ......................................................... 11
  DATA ANALYSIS ............................................................................................................ 12
    Definition of stimulus neurons, reward neurons and stimulus-reward neurons ........ 12
    Reward index ............................................................................................................... 14
    Selection and classification of neurons ..................................................................... 14
    Visually evoked response latency ............................................................................. 15
    Inter-spike interval (ISI) distribution ......................................................................... 15
    Definition of early cue period and late cue period ............................................. 15
    Population peri-stimulus histograms ..................................................................... 16

RESULTS ............................................................................................................................. 16
  BEHAVIORAL RESULTS ............................................................................................... 16
  NEURAL DATABASE ......................................................................................................... 19
  REWARD NEURONS IN THE LPFC AND STRIATUM ..................................................... 25
  REWARD-MODULATED EFFECTS FOR OLD STIMULI AND NEW STIMULI IN THE LPFC
  AND STRIATUM .............................................................................................................. 33
  STIMULUS-REWARD NEURONS IN THE LPFC AND STRIATUM .................................. 35
  IDENTIFICATION OF BROAD-SPike AND NARROW-SPike NEURONS ..................... 39
  CLASSIFYING BS AND NS NEURONS INTO REWARD AND OTHER NEURONS ........ 43
  BS AND NS REWARD NEURONS ................................................................................... 45

DISCUSSION ......................................................................................................................... 50
  INFERENCE AND CATEGORY ....................................................................................... 50
  PREFRONTAL BS AND NS REWARD NEURONS ............................................................ 54

ACKNOWLEDGEMENTS .................................................................................................... 57

REFERENCES ...................................................................................................................... 58
Introduction

Inferences are very important cognitive abilities that generate new information from exciting experiences or knowledge. Principally, inferences can be classified as deduction, induction, and abduction (an inferring process that seeks a reason to explain an observed phenomenon). Studies have been conducted regarding subtypes of inferences in diverse species of animals. For instance, males fish (Astatotilapia burtoni) can transitively infer their hierarchical relationship (who are stronger/weaker than themselves) by merely observing the other fish's performance in territorial fights. Great apes perform disjunctive reasoning, and New Caledonian crows notice a potential threat in an abduction way.

Computationally, Bayesian inference would be used to explain the mechanism of inferences. Bayes' rule can explain why you are more likely to say somebody has got a cold rather than pneumonia or heartburn when observing his coughing. However, animals' cognitive abilities are far from Bayesian inference, more data needs to be collected in neuroscience. fMRI experiments have found lateral, medial, or orbital prefrontal cortex and hippocampus were related to inferences, however, due to fMRI's low temporal and spatial resolution we don't know how exactly neurons in the brain makes inferences possible. The purpose of this study is to reveal the mechanisms in which single neurons are involved in a task that required inferences and what kind of information a neuron could generate to complete such task. Accordingly, we employed single-unit recording and chose a subtype of deductive inferences named transitive inference, i.e., if A is B and B is C, then A is C, to explore the mechanism of inferences. Transitive inference task was initially introduced into experiments by Piaget to test the reasoning ability of the child. Later, transitive inference paradigm was applied on pigeons, jays, chimpanzees, as well as humans. Behaviorally the paradigm of transitive inference and higher order conditioning are almost identical, so from the perspective of behaviorism, one would think the process of transitive inferences is the same as that of higher ordering conditioning denying the possibility that monkeys could perform inferences in their minds. However, conditioning could only describe the relation between stimuli and responses on behavioral level, and it's not a theory explaining the neural basis of inferences.

We designed a transitive inference task so as to observe how monkeys perform inference in their brain. In our task, we use two sets of three stimuli, one is A1, B1 and C1, the other is A2, B2 and C2. Stimuli in each set are linked together, i.e., A1→B1→C1, A2→B2→C2. For example, the monkeys are required to select B1 rather
than B2 when A1 is shown. Since animals are sensitive to the amount of reward, they prefer larger rewards rather than smaller rewards\textsuperscript{15}, we used reward of juice as an indicator to find out if the monkeys could infer the reward by given a clue of stimulus A1 or A2. Specifically, we taught the monkeys that C1 was linked with a reward and C2 was not, but sometimes we reversed this reward contingency so that C1 was not rewarded and C2 was rewarded. The monkeys could infer that A1 would be rewarded if C1 had been linked with a reward. If we have altered the reward to be linked with C2 indicating C1 was not rewarded and taught this information to the monkeys, then they could infer that A1 would not be rewarded. So, the process of inference can be replaced into a reward prediction process that is easier to observe related neural signal.

The prefrontal cortex (PFC) encodes different types of abstract rules which reflect the structure or the meaning of the external world, e.g., matching or no-matching rules: cats or dogs; go or no-go: “greater than” or “less than”\textsuperscript{112,108,27}. The PFC cortex is also goal-directed, and can flexibly represent relevant information which is necessary to complete tasks\textsuperscript{63}. Further, the PFC can generalize an acquired rule to new environments to resolve novel problems\textsuperscript{19}. Therefore, the PFC should play an important role in inferences which require such higher cognitive abilities to predict reward outcomes. On the other hand, in addition to the PFC, the striatum is also involved in reward prediction processes\textsuperscript{110,38,85,70,92}. Anatomically, the PFC and the striatum are tightly connected\textsuperscript{1,117}. The lateral PFC (LPFC) has two types of connections with the striatum: a focal projection and a diffuse projection\textsuperscript{25,32}. In the focal projection, LPFC projects primarily to the head of the caudate and part of the rostral putamen. In the diffuse projection, clusters of terminal fibers from the LPFC are widely distributed throughout the striatum. The anatomical data suggest that there are dense convergent and divergent connections from the LPFC to the striatum\textsuperscript{1,117,32}.

However, the striatum is functionally deemed to have a different mechanism when predicting reward outcomes. The striatum learns action values directly by trial and error, without building explicit models of the environment and task (i.e. “model-free learning”)\textsuperscript{16,43}. For example, the striatum uses temporal-difference learning to guide behavior in order to maximize future reward by simply reinforcing successful actions\textsuperscript{67}. On the other hand, the PFC has been shown to play an important role in model-based learning\textsuperscript{16,88,74}, such as in the encoding of abstract rules\textsuperscript{112,108,27}, in planning behaviors\textsuperscript{89,87}, and in dynamic decision-making\textsuperscript{71,5}.

On the basis of these observations, we hypothesized that while the LPFC neurons were involved in a reward inference process that predicted reward value based on knowledge of the causal structure of the task, the striatal neurons predicted reward values directly by trial and error.
value through direct learning. To reveal how these two regions may have disparate strategies in reward prediction, we recorded single-unit activity from the LPFC and striatum of monkeys performing a sequential paired-association task under an asymmetric reward schedule\textsuperscript{75}. Although neurons in both regions predicted reward outcomes immediately after the first cue of A1 or A2, we found only neurons in LPFC, rather than those in the striatum, predicted reward outcomes in trials when inference was required.

Furthermore, neurons in the LPFC can be classified into two types: pyramidal cells and interneurons. Intracellular recordings have reported that pyramidal cells had broader spike waveforms than interneurons\textsuperscript{61}. Thus we classified the recorded neurons in these two groups based on the waveforms obtained in extracellular recordings, and associated each type of neurons with differential functions in the behavioral tasks\textsuperscript{3,44,41,103}. These two types of neurons may differ in morphology, neurotransmitter types, and electrophysiological properties\textsuperscript{12,61,116}. It has been reported that putative pyramidal cells and interneurons in the LPFC may have distinct functional roles in higher-order cognitive functions\textsuperscript{78,39,40}. So, as we have found that prefrontal neurons could infer the reward outcomes in our task, we asked two questions: 1) Whether do both putative pyramidal cells and interneurons encode reward information in the PFC, and 2) If the two classes of neurons encode reward information, what type of reward information do each type represent? We therefore analyzed distinct functional property of prefrontal neurons in our task separately depending on the types of neurons (pyramidal cells and interneurons) judged by spike shapes.

**Materials and Methods**

**Subjects**

Totally four male Japanese monkeys (*Macaca fuscata*) served as subjects in this study (Hop, 7.5kg, Tap, 6.5kg, Tom, 8.9kg and Zep, 8.5kg). We implanted a head-holder and two recording chambers (one in each hemisphere) for each monkey under aseptic techniques with ketamine (4.6–6.0 mg/kg by intramuscular injection) and sodium pentobarbital (Nembutal, 4.5–6.0 mg/kg by intravenous injection) anesthesia. The size of the chamber was 40 mm (length, anterior-posterior) × 30 mm (width, lateral-medial), and each was implanted with its center located at around the end of the principal sulcus, which allowed us to record neural activity in both the
LPFC and striatum simultaneously from the same chamber. All surgical and experimental protocols were approved by the Animal Care and Use Committees (H26-42) at Tamagawa University and were in accordance with the National Institutes of Health’s Guide for Care and Use of Laboratory Animals.

During each experimental session, the monkeys were seated in a primate chair (with their head fixed) inside a completely enclosed sound-attenuated and electrically shielded room. A 21-inch CRT display (FE220, NEC, Japan) with 60Hz refresh rate was set at a distance of 60.0 cm in front of the monkey for the presentation of visual stimuli. Eye movements were monitored by the Eyelink2 system (SR Research Ltd, Mississauga, Canada) with 500 Hz sample rate. All stimulus presentation and behavioral procedures were controlled by the TEMPO system (Reflective Computing, USA).

**Task**

*Sequential paired-association task*

In this study, the monkeys were required to learn two associative sequences (Fig. 1A) in a sequential paired-association task (Fig. 1B). Six discriminable icons were used as visual cues, counterbalanced across three subjects (a question mark and a sketch of tower was denoted as A1 and A2; red and green patches as B1 and B2; a cross and a circle as C1 and C2). The two correct sequential associations were: A1 → B1 → C1 and A2 → B2 → C2.

Firstly, we trained the monkeys to learn the association between A1 and B1, and the association between A2 and B2. After showing the monkeys A1 or A2 at the center of the monitor, B1 and B2 were displayed on the monitor simultaneously. If the monkeys selected the correct answer (i.e., B1 to A1, and B2 to A2), they would get a drop of juice as the reward. By trial and error the monkeys successfully acquired such associations. Secondly, by a same process we trained the monkeys to learn the association between B1 and C1, and the association between B2 and C2. At last, we combine the three stimuli of each set in one trial and trained the monkeys to do a two-step selection (e.g., select B1 or B2 after seeing A1 or A2, then in the same trial select C1 or C2 if the previous selection is correct).

Particularly, each sequential paired-association trial (SPAT, Fig. 1B) started with the onset of a white fixation spot (0.21° of visual angle) presented at the center of the monitor. The monkeys had to fixate on the spot for a random duration (800–1,200ms) within a virtual fixation window (3° x 3°). Subsequently the first stimulus
cue, for example A1, was presented for 400ms at the center of the monitor. After a variable delay period (700–1,200ms) the fixation spot disappeared, and at the same time the second cues, B1 and B2, were presented pseudo-randomly (using a modified Gellerman sequence) at the left and right positions to the center on the CRT (6° of visual angle from center). If the subject made a saccade to the target cue (e.g. B1): this was denoted as the first correct choice. Immediately after the correct first choice, the distracter (e.g., B2) disappeared, and the monkeys continued fixating on the target cue (B1) for another 600ms. After the disappearance of the target cue (B1) the third cues, C1 and C2, were simultaneously displayed pseudo-randomly to the left and right of where the target cue (B1) had been shown (5° of visual angle from the central position of B1). This instructed the monkeys to make a further saccadic eye movement to the correct target cue (e.g., C1) in the second choice. After two correct choices, the monkey received a drop of water as reward and an auditory tone of 1kHz at the end of the trial.

The saccade was judged to be correct if the eye position stayed at least 200ms within a virtual window (3° × 3° of visual angle) centered on the position of the target stimulus. The monkeys had to keep its fixation inside the virtual fixation window during the fixation, cue presentation, and delay periods. If the monkeys moved its fixation out of this window the trial was rejected as a fixation break. When the monkeys made a fixation break or an erroneous choice, the trial was aborted, and a high tone of 4kHz indicated an error. Then, after a longer inter-trial interval (ITI) (6s, the normal ITI after a correct trial was 3s), the same trial was repeated until the monkeys were able to complete it correctly. The repeated trials after error were referred to as correction trials. Through the SPAT training the monkeys learned the two correct associative sequences: A1→B1→C1 and A2→B2→C2. The monkeys also learned another two sequences: the BCA sequence (B1→C1→A1 and B2→C2→A2) and the CAB sequence (C1→A1→B1 and C2→A2→B2), however these two sequences were not used in current study.

**Reward instruction trials**

When the monkeys acquired the two sequential associations, we introduced an asymmetric reward schedule using reward instruction trials. During reward instruction trials, the monkeys were instructed as to which stimulus (C1 or C2) was paired with a large reward (Fig. 1D76). This was done by a visually guided saccade task, in which, after fixation, C1 or C2 was presented (skipping the A1, A2, and B1, B2 steps) and the monkeys had to saccade to whichever appeared. The different visual stimuli (C1 and C2) were paired with different amounts of water. In one case C1 was
followed by a large reward (0.4 ml) and C2 was paired with a small reward (0.1 ml); in the other case, C1 was followed by a small reward and C2 by a large reward.

**Figure 1.** Sequential paired-association task with an asymmetric reward schedule. 

A. Two stimuli sets (A1-B1-C1 and A2-B2-C2) were learned by the monkeys. These six icons were termed “old stimuli.”

B. Time events in the sequential paired-association trial. The monkeys made a choice by a saccadic eye movement, as indicated by small yellow arrows.

C. Examples of new stimuli associated with two color patches (B1 and B2). The monkeys learned each pair of new stimuli (i.e., NS1 and NS2) with B1 or B2 in a delayed matching-to-sample task with a symmetric reward schedule.

D. Asymmetric reward schedule in one block.

---

(76)
Reward instructed sequential paired-association task with old stimuli

After having fully acquired how to complete both the reward instruction and the sequential paired-association tasks, the monkeys were required to perform a combination of the two subtasks in one block. At the beginning of each block, three or four reward instruction trials were presented for the monkeys indicating which target (C1 or C2) was paired with the large reward and which with the small reward. Reward instruction trials were then followed by the SPATs (8–15 trials). The procedure of the SPATs was the same as described above, except that the amount of reward was asymmetric. The asymmetric reward rule was consistent between the two types of trials within one block. E.g. in a given block, if C1 was paired with the large reward and C2 with the small reward in reward instruction trials, then the A1→B1→C1 sequence would be followed by the large reward and the A2→B2→C2 sequence would be paired with the small reward in the SPATs. The asymmetric reward schedule was pseudo-randomly assigned between blocks.

Training of new stimuli pairings

The monkeys were trained to learn new stimulus associations in a delayed matching-to-sample task with a symmetric reward schedule. The new stimulus was learned to be associated with one of the two color patches (B1 or B2). In this task, the color patches were always presented on the same side of the screen (e.g. B1 on left, B2 on right) in order to speed learning. We downloaded 400 icons from the Internet, classified them into 200 new pairs randomly, and examined each pair visually to avoid reuse. These newly learned stimuli shall hereby be referred to as “new stimuli”, while the stimuli A1, B1, C1, A2, B2 and C2 shall hereby be referred to as “old stimuli”. 3 monkeys (Hop, Tom and Tap) were trained with new stimuli.

After learning one pair of new stimuli, the monkeys moved on to learn another pair of new stimuli, and so on. For this training, we assigned one daily session each week to train the monkeys to learn 4–5 pairs of associations between the new stimuli and B1 and B2 without recording. 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. Up to this point, the monkeys had received 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.

**Reward instructed sequential paired-association task with new stimuli**

After having fully acquired the new associations, the monkeys performed the reward instructed sequential paired-association task with new stimuli (Fig. 1D). 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 the old stimulus (A1 or A2), and the second cues, B1 and B2 (the two-color patches), were always shown in the same fixed positions as during training.

**Data acquisition**

Action potentials of single neurons were recorded extracellularly with tungsten electrodes (FHC, Bowdoinham, ME, 0.8–1.5MΩ) from the LPFCs of the monkeys (Fig. 3), and the caudate and putamen of the two monkeys (Tap and Tom, Fig. 4). Usually, two microelectrodes, one into the LPFC and the other into the striatum, were used to record single-unit activity. Each electrode was inserted via a guide tube through a grid system (holes: 0.6 mm wide and 1.0 mm apart from center to center; Nakazawa, Tokyo, Japan) into the cortical surface; then each electrode was advanced into the target position independently by means of the NAN-electrode-drive (NAN-instruments LTD, Israel). The Plexon system (Plexon Inc, Texas, USA) was used to amplify neuronal action potentials and discriminate individual spike waveforms online.

The system then saved the spike timing and selected waveforms (in an 800 µs window) on the Plexon PC together with the timing of task events. The eye movement data and spiking time were also saved on the behavioral control PC (TEMPO client PC). In the caudate and putamen, we recorded the spike activity of presumed projection neurons that showed phasic response to task events and had low spontaneous activity35, but not those of presumed interneurons that showed irregular tonic discharge2.

During recording sessions, once a neuron was isolated, its properties were tested in the reward instructed SPAT with the old stimuli (and here, the positions of the second cues B1 and B2 were pseudo-randomized). The first cue was A1 or A2. If the neuron's activity was found to be modulated by the amount of reward, then we continued recording the activity of that neuron while the monkey performed the
reward instructed SPAT with the new stimuli (the positions of the second cues B1 and B2 being fixed). Each pair of new stimuli was tested with only one neuron. A pair of new stimuli was repeatedly tested for at least two blocks for each neuron recorded.

The advantage of the above recording method is that it reduced the number of new stimuli associations which the monkeys had to learn in total (It’s quite hard for the monkeys to learn the new stimuli, and so the less the monkeys had to learn, the less time and effort required). The disadvantage was that we might have neglected neurons that would show reward-modulated activity to the new stimuli, but not to the old stimuli. However, in this study we were interested in studying neurons that were related to value coding. These value-coding neurons inherently must respond to the value of both old and new stimuli. The type of neurons which we potentially neglected to record could not be called “value-coding neurons” per se, as they would respond to only the value of the new stimuli, but not the value of the old stimuli. This type of neuron, if it exists, may be more related to “value learning”, a concept we did not aim to address in the current experiment.

**The recording sites in the LPFC and striatum**

Fig. 3 and Fig. 4 show the recorded tracks and distributions of reward neurons in the LPFC (both hemispheres for Tap, the left hemisphere for Hop, and the right hemisphere for Tom), and in the striatum (two hemispheres for Tap, and the right hemisphere for Tom), respectively. Confirmation of recording locations was based on magnetic resonance imaging (Siemens MRI, 3-T Trio. A Tim MRI scanner). To reconstruct the recording sites based on MR images, we took MR images of the monkeys’ brain with inserted markers. These markers were 21 small silicone tubes that had been filled with phantom liquid and then inserted into the recording grid. These tubes mimicked the recording electrodes and showed higher contrast in MR images (as indicated by the three short gray lines in Fig. 4A and Fig. 4D). The three long white lines in Fig. 4A and Fig. 4D were extensions of those tube lines. On the basis of the position of the first recorded neuron in the cortex, and the position of the first neuron recorded in the striatum, we were able to map neurons in the striatum.

In the prefrontal cortex, we mainly recorded single units from the ventro-lateral area (VLPFC, beneath the principle sulcus). Reward neurons tended to be dispersed widely in the VLPFC, i.e., there was no distinct spatial distribution. In the striatum, we predominately recorded neural activity from the anterior part, anterior to the anterior commissure (AC), including the caudate nucleus and the putamen. This area usually is referred to as the association striatum. Several studies involving monkeys
have reported that neurons in this association striatum are related to learning\textsuperscript{36,66,113}.

**Data analysis**

**Definition of stimulus neurons, reward neurons and stimulus-reward neurons**

Off-line analysis was carried out using custom-made Matlab programs on a PC. To calculate the behavioral performance (correct rate) of the first choice as a function of SPAT order in blocks after reward instruction, the data were averaged across all blocks performed by the monkeys. Using the correction method, we excluded the data in repeated trials (the trials directly after error trials) to calculate the correct rate. We also calculated the behavioral performance in response to new stimuli presented for the first time in SPAT blocks immediately after reward instruction trials. Trials in each of the first SPAT blocks were classified into two sequences: the *first trial-sequence* and the *second trial-sequence*, which shall be described below. These sequences shall hereby be referred to as N\textsubscript{1} and N\textsubscript{2} (standing for the first trial-sequence and the second trial-sequence, respectively). Additionally, when we wish to refer to particular trials within a trial sequence we shall put the trial number on the end, for example N\textsubscript{1,1} stands for trial 1 in the first trial-sequence, N\textsubscript{1,2} stands for trial 2 in the first trial-sequence, N\textsubscript{2,1} stands for trial 1 in the second trial-sequence, etc.

Each first trial-sequence consisted of the very first SPAT within the block (N\textsubscript{1,1}), and subsequent trials within this block in which the same new stimulus was presented (N\textsubscript{1,2}, N\textsubscript{1,3}, etc.; please note that these are not necessarily the second and third trials in the block, but rather the second and third trials in which this stimulus was presented within the block). The second trial-sequence included trials in which the alternative new stimulus (N\textsubscript{2,1}, N\textsubscript{2,2}, N\textsubscript{2,3}, etc.; see the schematic block in the bottom panel in Fig. 2\textsuperscript{76}) was presented as the first cue.

Overall, for each pair of new stimuli, there was one first trial-sequence and one second trial-sequence. To calculate the behavioral performance in response to the new stimuli in the first trial-sequence, we first determined the accuracy in each trial in each first trial-sequence, then averaged the correct rates across all first trial-sequences separately for the large reward condition and the small reward condition (see Fig. 2\textsuperscript{76}). The behavioral performance in response to the new stimuli in the second trial-sequences was calculated in the same way.
For the analysis of neuronal data, we used only correct trials. In this study, we concentrated on neuronal activity in two time epochs: the first cue period and the early delay period. The first cue period occurred from 100ms to 500ms after the first cue onset, and the early delay period occurred from 500ms to 900ms after the first cue onset.

The activity of each neuron was analyzed by a two-way ANOVA, with old stimuli (A1 vs. A2) and reward (large vs. small) as the main factors in the first cue period and in the early delay period, respectively. Depending on the results of the ANOVA (P<0.01), neurons were classified into three types: stimulus neurons, reward neurons and stimulus-reward neurons. Stimulus neurons showed a significant main effect of stimuli (P<0.01), but no a main effect of reward and no interaction between the two factors. Therefore, stimulus neurons only discriminated the visual properties of stimuli, regardless of reward conditions. Neurons that only showed a significant main effect of reward (P<0.01) were classified into reward type. This type of neurons predicted whether the amount of reward associated with a stimulus would be large or small, regardless of which stimulus was presented. Stimulus-reward neurons showed a significant interaction between stimuli and reward and/or significant main effects of both stimulus and reward. Some neurons showed reward-modulated activity in both the first cue and early delay periods. If a neuron was identified as reward type in the first cue period, it was excluded from the population in the early delay period so that the population neurons in the two periods would not overlap.

Next the activity of the neurons recorded with new stimuli was further analyzed by two-way ANOVAs (new stimulus × reward) in the first cue and early delay periods. Our aim was to examine whether or not reward neurons defined by old stimuli would keep the same response properties for new stimuli. In order to examine the interaction between stimulus familiarity and reward, we performed a two-way ANOVA using the activity of each neuron that had been recorded for both old and new stimuli. The two main factors were stimulus familiarity (old vs. new) and reward amount (large vs. small).

The spike density function of single neurons was generated by averaging activities in every 1ms-bin across trials, and then smoothed by a Gaussian envelope with σ=30ms. To generate the population histogram, we first calculated the firing rate of each trial in every 1ms bin. We then subtracted from this, the mean firing rate from the fixation period (500ms epoch prior to the first cue onset). Finally, we averaged the normalized firing rate in every bin across all trials for each neuron and across population neurons, and finally smoothed this data using a Gaussian envelope with σ=30ms.

For each reward neuron whose activity was recorded using the new stimuli, we
first determined its activity responding to the new stimuli in each trial in the first and second trial sequences, then normalized the activities by the maximal and minimal responses to the new stimuli in all trials. The normalized activity in each trial was averaged across the first trial sequences (or across the second trial sequences) separately for the preferred reward condition and the non-preferred reward conditions.

**Reward index**

To evaluate the strength of reward modulation of reward neurons, we calculated a normalized reward index for each neuron using the following equation:

\[
\text{Reward Index} = \frac{R_{\text{pref}} - R_{\text{non-pref}}}{R_{\text{pref}} + R_{\text{non-pref}}}
\]

where \(R_{\text{pref}}\) indicates the activity in the preferred reward condition, and \(R_{\text{non-pref}}\) in the non-preferred reward condition. For each neuron, this reward index was calculated for the old and new stimuli, respectively. All trials across blocks were included to calculate the reward index for the old stimuli in the LPFC and the striatum. To compute the reward index for the new stimuli, we again included all trials across blocks in the LPFC. However, we excluded the first trials in which the new stimuli were presented for the first time as the first cue in SPATs (i.e., \(N_{1,1}\) and \(N_{2,1}\)) in the caudate and putamen.

**Selection and classification of neurons**

Neurons in the LPFC were classified into putative pyramidal cells and interneurons based on the following criteria in further analysis: (1) each neuron was clearly isolated from the other neurons and from multiunit activity based on clear clustering in the principle components of waveforms (Offline Sorter, Plexon Inc.). (2) Each unit showed a clear spike refractory period. (3) The waveforms of each neuron exhibited a downward voltage deflecting (a trough, negative related to baseline) followed by an upward voltage deflecting with a clear peak. (4) LPFC neurons with deviated waveforms or saturated amplitude waveforms were excluded from analysis.

A function of \(k\)-means classifier (\(k=2\), squared Euclidean distance) in Matlab was used to classify neurons into two groups based on the shape of averaged spike waveforms. In order to reduce possible effects of different amplitude and timings of minimum of waveforms, we normalized the waveforms of each neuron by the differences between their peak and trough values and aligned them by their minimum. The \(k\)-means algorithm (\(k=2\)) aims to partition the spike waveforms into 2 clusters in which each waveform belongs to the cluster with the nearest mean
waveform, so as to minimize the variance of waveforms within the cluster\textsuperscript{56}. The k-means algorithm does not examine any statistical significance among k clusters or does not necessarily find global optimal clusters. Therefore, we calculated waveform durations (between the trough and peak in the averaged waveform) to further confirm that the classification by k-means (k=2) was appropriate. The advantage of k-means (k=2) classifier is that this algorithm objectively classifies spike waveforms into two clusters on the basis of their shapes, without the requirement of a pre-determined border between the two clusters. We referred to the neurons that had the smaller mean spike width as narrow-spike (NS) neurons, and to the neurons that had the larger mean spike width as broad-spike (BS) neurons.

**Visually evoked response latency**

To determine the visually evoked response latency for each NS or BS neuron, its averaged spike density histogram was derived with 1ms resolution, and smoothed by a Gaussian envelope with $\sigma=15$ms for all trials. We calculated the mean and standard deviation (SD) of firing rates during a 200ms-time-window prior to the first cue onset across all trials for each neuron, and set the threshold at the mean of baseline firing rates plus 3 SDs. If the neuronal activity exceeded the threshold for three consecutive time bins after the first cue onset, the response latency was defined as the duration from the first cue onset to the first of these time bins. If a neuron was unable to reach this criterion, it was excluded from the latency count.

**Inter-spike interval (ISI) distribution**

We used the FieldTrip open source Matlab toolbox\textsuperscript{73} to calculate the ISI distribution and local variation for each NS or BS neuron. The ISI distribution of each neuron was fitted by a Gamma distribution and its parameters such as the peak mode were estimated.

**Definition of early cue period and late cue period**

We observed both the BS and NS Reward neurons showed visually evoked response to the presentation of the first cue in the two reward conditions (see Fig 14A\textsuperscript{29}). This visually evoked response would take strong effect on comparison between the baseline activity and the cue period activity. In the current task design, we were unable to separate the evoked response from the reward-related signal. To avoid this
effect, the cue period was divided into two time epochs: the early cue period (100 – 300ms from the first cue onset) and the late cue period (300 – 500ms from the first cue onset). There might be more visually evoked activities in the early cue period, and less in the late cue period. We compared the activity in the fixation period with the activity in the late cue period to demonstrate the reward information represented in each Reward neurons. The activity in the delay period should not include the visually evoked response, and it was compared directly with the baseline activity.

Population peri-stimulus histograms

To generate the population histogram, we first calculated the firing rate in every 1ms bin within each trial, then averaged the firing rate in every bin across all trials for each neuron and across population neurons, and finally smoothed the data using a Gaussian envelope with $\sigma$=30ms.

Results

Three Japanese monkeys (Hop, Tap and Tom) were firstly required to perform the reward instructed sequential paired-association task with old stimuli (the first cue stimuli: A1 and A2. see Materials and Methods). In this task, the monkeys extensively experienced the stimulus-reward reversals block-by-block. Subsequently the monkeys were required to perform the task but with the new stimulus pairs which had not been previously presented with an asymmetrical reward schedule. In total, we tested 112 pairs of the new stimuli for Hop, 214 for Tap, and 136 for Tom. This allowed us to analyze the behavior and neuronal activity in the sequential paired-association trials (SPATs) in which a new stimulus was presented for the very first time.

Behavioral results

We found that the behavior of the three monkeys was systematically influenced by the amount of reward. The monkeys showed significantly higher correct rates on the first choice in the larger reward trials for both the old and new stimuli than in the smaller reward trials (Mann-Whitney U test, $P<0.01$).

We further examined the correct choice rates in response to the initial presentation of the new stimuli in the SPATs after reward instruction of C1 and C2
The SPATs in each of these blocks were separated into the first and second trial-sequences (see Materials and Methods).

We focused our analysis on the first-choice performance in the first SPATs in the first and second trial-sequences (i.e. the first choices in N₁,1 and N₂,1). In these specific trials, the new stimuli were presented for the first time in the SPATs, and therefore the monkeys didn’t have any opportunity to directly pair the new stimuli with a particular (large or small) amount of reward. Nevertheless, the monkeys showed significantly higher correct rates in the larger reward condition from the first trials in both trial-sequences (Fig. 2, Mann-Whitney U test, **: P<0.01, *: P<0.05), suggesting the possibility that the monkeys could correctly infer the reward. Behaviorally, however, we could not yet easily determine whether the monkeys actually predicted the amount of reward based on the first cues (the new stimuli), or if instead, they were simply relying on associations with the second cues (B₁ and B₂; old stimuli) to predict the reward.

Evidence for reward inference should ideally be sought in the neuronal data at the time of the first cue presentation, that is, before the monkeys can rely on a second cue so as to recall an already-established reward association (e.g., B₁→C₁→Reward).
Figure 2. The effect of reward amount on the behavioral performance of the three monkeys. A, B, The performance of monkey Hop showing the accuracy of the first choice (selection of B1 or B2 on the basis of new stimuli) in SPATs as a function of SPAT order in the first trial sequence (A), and the second trial sequence (B). C, D, Performance of monkey Tap in the first trial sequence (C) and in the second trial sequence (D). E, F, Performance of monkey Tom in the first trial sequence (E) and in the second trial sequence (F). (76)
Neural database

Accordingly, we recorded the activity of 546 neurons from the LPFCs of the three monkeys (Hop: 152; Tap: 217; Tom: 177, see Fig. 3) and the activity of 366 neurons from the caudate and putamen of two of the monkeys (Tap: caudate: 95, putamen: 160; Tom: caudate: 73, putamen: 38; see Fig. 4) while they performed the reward-instructed sequential paired-association task with the old stimuli (the first cues: A1 and A2). The activity of each neuron was analyzed using a two-way ANOVA: (stimulus (A1 or A2) × reward (large or small), P < 0.01) in the first cue and early delay periods that were prior to the second cues. Depending on the ANOVA results, the neurons were classified into three types: stimulus neurons, reward neurons and stimulus-reward neurons. Table 1, Table 2 and Table 3 show the number of each type of neurons in the LPFC, the caudate, and the putamen, respectively. In this part, we mainly focused on the reward neurons in the LPFC and striatum to investigate how these neurons processed reward information for the new stimuli, and additionally we looked at the activity of stimulus-reward neurons in the LPFC representing categorical information of the stimuli.

In the LPFC we found 92 reward neurons in the first cue period and 63 reward neurons in the early delay period. There were 61 caudate and 52 putamen reward neurons in the first cue period and 16 caudate and 28 putamen reward neurons in the early delay period. The proportion of reward neurons in the LPFC was significantly lower than those in the caudate and putamen (28.4% (155/546) in the LPFC: 45.8% (77/168) in the caudate, and 40.4% (80/198) in the putamen, χ²-test, Chi-square=21.682, df=2, P<0.01). The ratio of reward neurons did not differ significantly between the caudate and putamen (χ²-test, Chi-square=1.094, df=1, P=0.296). About half (53.5% (83/155)) of the LPFC reward neurons and three-quarters of the caudate (74% (57/77)) and putamen (77.5% (62/80)) reward neurons showed stronger activity in large vs. small reward trials. The ratio in the LPFC was significantly lower as compared to the ratio in the caudate (χ²-test, Chi-square=9.015, df=1, P<0.01). For the reward neurons in both the LPFC and striatum, we refer to the reward condition which produced higher activity as the "preferred reward condition" and the other reward condition as the "non-preferred reward condition".

Next, we examined how many neurons showed significant reward effects in both the first cue and early delay periods. Of the 92 LPFC neurons that showed reward-modulated activities to the old stimuli in the first cue period, 42 (45.7%) of them also showed reward-modulated activities in the early delay period. Of the 61 caudate neurons that showed reward-modulated activity to the old stimuli in the first
cue period, 39 (63.9%) of them also showed reward-modulated activities in the early delay period. Additionally, of the 52 putamen neurons that showed reward-modulated activity to the old stimuli in the first cue period, 27 (51.9%) of them also showed reward-modulated activities in the early delay period.

In this study, we also recorded the neural activity of these reward neurons with the new stimuli in the LPFC, the caudate and putamen. In the LPFC, 106 out of 155 reward neurons were recorded for both the old and new stimuli (see Table 176). In the striatum, 100 (50 neurons from the caudate and 50 neurons from the putamen) out of 157 reward neurons were tested with the new stimuli (see Table 276 and 376). Each of them was tested with at least one pair of the new stimuli (median number of pairs: 2; range: 1–6).

To examine how familiarity of stimuli affected reward information encoded by these reward neurons, the activity of each of them was analyzed using a two-way ANOVA with two factors (stimulus familiarity (old vs. new) and reward (large vs. small)). Of the 106 LPFC reward neurons, 38 (35.8%) neurons showed only a significant main effect of reward (P<0.01), and neither significant main effect of familiarity nor any interaction between the two factors (P>0.01), indicating that these neurons encoded reward information regardless of stimulus familiarity. About 38.7% (41/106) of LPFC reward neurons had significant main effects of reward and stimulus familiarity (P<0.01), but no significant interaction between them, indicating that these neurons represented the familiarity information and the reward information of stimuli independently. We found 19 (17.9%) LPFC neurons showed a significant interaction between the two factors of reward and stimulus familiarity (P<0.01). Additionally, four neurons proved only selective to old vs. new stimuli, and another four neurons had no effects of reward or stimulus familiarity and showed no interaction.

Of the 50 caudate reward neurons, 16 (32%) neurons showed a significant main effect of reward, with no significant main effect of stimulus familiarity and no interaction between the two factors. There were 17 (34%) neurons that had significant main effects of reward and familiarity, but displayed no interaction between them. 17 (34%) neurons showed a significant interaction between the two factors.

Of the 50 putamen reward neurons, 27 (54%) neurons showed a significant main effect of reward, with no significant main effect of stimulus familiarity and no interaction between the two factors. There were 12 (24%) neurons that had significant main effects of reward and familiarity, but displayed no interaction between them. 10 (20%) neurons showed a significant interaction between the two factors. One neuron proved only selective to the old vs. new stimulus. We found that the fractions of reward neurons whose activity was modulated by the new stimuli did not differ
significantly in the first cue and early delay periods either in the LPFC (in the first cue period: 40/63 (63.5%), in the early delay period: 25/43 (58.1%), χ²-test, Chi-square=0.309, df=1, P=0.5785), in the caudate (in the first cue period: 28/41 (68.3%); or in the early delay period: 6/9 (66.7%), χ²-test, Chi-square=0.009, df=1, P=0.924) or in the putamen (in the first cue period: 19/37 (51.4%); in the early delay period:3/13 (23.1%), χ²-test, Chi-square=3.121, df=1, P = 0.077). When the ratios within the three areas were compared, the putamen showed a smaller proportion of reward neurons whose activity was sensitive to stimulus familiarity than the LPFC or the caudate did (the putamen: 22/50 (44%), the caudate: 34/50 (68%), the LPFC: 65/106 (61.3%), χ²-test, Chi-square=6.543, df=2, P=0.038). The proportions in the LPFC and caudate did not differ significantly (χ²-test, Chi-square=0.654, df=1, P=0.419).

In the LPFC, we found 64 stimulus-reward neurons and 32 stimulus neurons in the first cue period. In the striatum, we found 24 stimulus-reward neurons and 8 stimulus neurons in the first cue period. There were few neurons that were identified as stimulus-reward or stimulus type only in the early delay period (Stimulus-reward: 10 neurons in the LPFC and 7 in the striatum; Stimulus: 7 neurons in the LPFC and 3 in the striatum). We did not analyze these early delay period stimulus-reward neurons and stimulus neurons further for this study because the activity of these neurons was not tested with new stimuli. The proportion of stimulus-reward neurons was higher in the LPFC than in the striatum (64/546 (11.7%) in the LPFC, 24/366 (6.5%) in the striatum; χ²-test, Chi-square=6.703, df=1, P<0.01). About half of stimulus-reward neurons (35/64 (54.7%)) in the LPFC showed higher activity in large reward trials, and 15/24 (62.5%) of stimulus-reward neurons in the striatum preferred large reward, but these did not differ significantly (χ²-test, Chi-square=0.434, df=1, P=0.510).
Table 1. Classification of LPFC neurons in the three monkeys

<table>
<thead>
<tr>
<th>Monkey</th>
<th>Recorded neurons</th>
<th>Reward neurons</th>
<th>Stimulus-reward neurons</th>
<th>Stimulus neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First cue period</td>
<td>Early delay period</td>
<td>First cue period</td>
</tr>
<tr>
<td>Hop</td>
<td>152</td>
<td>30 (26)</td>
<td>19 (13)</td>
<td>17 (11)</td>
</tr>
<tr>
<td>Tap</td>
<td>217</td>
<td>25 (16)</td>
<td>23 (17)</td>
<td>26 (19)</td>
</tr>
<tr>
<td>Tom</td>
<td>177</td>
<td>37 (21)</td>
<td>21 (13)</td>
<td>21 (16)</td>
</tr>
<tr>
<td>Total</td>
<td>546</td>
<td>92 (63)</td>
<td>63 (43)</td>
<td>64 (46)</td>
</tr>
</tbody>
</table>

Numbers between parentheses indicate the number of neurons tested with new stimuli. Each neuron is assigned to one of the two time periods to avoid double counting (see Materials and Methods).

Table 2. Classification of caudate neurons in the two monkeys

<table>
<thead>
<tr>
<th>Monkey</th>
<th>Recorded neurons</th>
<th>Reward neurons</th>
<th>Stimulus-reward neurons</th>
<th>Stimulus neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First cue period</td>
<td>Early delay period</td>
<td>First cue period</td>
</tr>
<tr>
<td>Tap</td>
<td>95</td>
<td>35 (24)</td>
<td>7 (2)</td>
<td>14 (4)</td>
</tr>
<tr>
<td>Tom</td>
<td>73</td>
<td>26 (17)</td>
<td>9 (2)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Total</td>
<td>168</td>
<td>61 (41)</td>
<td>16 (9)</td>
<td>18 (8)</td>
</tr>
</tbody>
</table>

Numbers between parentheses indicate the number of neurons tested with new stimuli. Each neuron is assigned to one of the two time periods to avoid double counting (see Materials and Methods).

Table 3. Classification of putamen neurons in the two monkeys

<table>
<thead>
<tr>
<th>Monkey</th>
<th>Recorded neurons</th>
<th>Reward neurons</th>
<th>Stimulus-reward neurons</th>
<th>Stimulus neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First cue period</td>
<td>Early delay period</td>
<td>First cue period</td>
</tr>
<tr>
<td>Tap</td>
<td>160</td>
<td>39 (27)</td>
<td>24 (9)</td>
<td>6 (5)</td>
</tr>
<tr>
<td>Tom</td>
<td>38</td>
<td>13 (10)</td>
<td>4 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Total</td>
<td>198</td>
<td>52 (37)</td>
<td>28 (13)</td>
<td>6 (5)</td>
</tr>
</tbody>
</table>

Numbers between parentheses indicate the number of neurons tested with new stimuli. Each neuron is assigned to one of the two time periods to avoid double counting (see Materials and Methods).
Figure 3. Anatomical location of recording sites in the LPFC of the monkeys. A-C, Distribution of recording sites and the distribution of reward neurons in the monkey Tap (A), the monkey Tom (B), and the monkey Hop (C). The empty black circles represent recorded sites at which no reward-related neurons were found. The filled red circles represent recording sites at which reward neurons were found. (76)
Figure 4. Recording sites in the striatum of the two monkeys (Tap and Tom) plotted on MRI coronal sections. To reconstruct the recording sites based on MR images, we took MR images of the monkeys’ brains with inserted markers (see Materials and Methods). A, Coronal MR image of Tap where the anterior commissure (AC) 0. The long white lines were extensions of those markers. B, C. Locations of recorded neurons in the right and left hemispheres of Tap, respectively. D, Coronal MR image of Tom where AC 0. The long white lines were extensions of these markers. E, Locations of recorded neurons in the right hemisphere of Tom. White circles represent recorded neurons that showed no reward-modulated activity. Red circles represent reward neurons. AC indicates the anterior commissure; the sections anterior and posterior to the AC are indicated by plus and minus numbers (distances in mm), respectively. (76)
Reward neurons in the LPFC and striatum

A typical reward neuron in the LPFC (Fig. 5A) showed significantly higher activity in large than in small reward trials (two-tailed t-test, P<0.01), and no differential activity between the two old stimuli (A1 versus A2) in the first cue period. This neuron showed a similar reward-modulated activity when the new stimuli were presented as the first cue (Fig. 5B). Thus, it appears that the reward neurons encoded reward-related information associated with the stimuli, regardless of their visual properties and regardless of the group to which the stimulus belonged, a type of neural activity which has been observed in previous reports.

We found that the majority of the 106 reward neurons in the LPFC (93/106, 87.7%) showed reward-type activity to the new stimuli in the first cue and/or in the early delay periods (two-way ANOVA (stimulus versus reward), P<0.05). The remaining 13 neurons showed no reward-modulated activity to the new stimuli tested. The population histogram of reward neurons in the LPFC shows that they distinguished the preferred from the non-preferred reward condition not only for the old stimuli, but also for the new stimuli in both the first cue period (Fig. 5C and D) and in the early delay period (Fig. 5E and F).

An example reward neuron in the caudate showed reward-modulated activity to both the old and new stimuli regardless of their visual properties (Fig. 6A and B). All 50 caudate and 90% (45/50) of putamen reward neurons showed reward-modulated activity to the new stimuli in the first cue and/or in the early delay period (two-way ANOVA (stimulus versus reward), P<0.05). The remaining 5 putamen neurons showed no reward-modulated activity to the new stimuli tested. The population histogram shows that the caudate and putamen reward neurons discriminated between the two reward conditions on the basis of both the old and new stimuli in the first cue period (Fig. 6C and E) and in the early delay period (Fig. 6D and F).
Figure 5. An example LPFC reward neuron and population histograms. In these figures, activity in the first and repetition trials with new stimuli was included in the analysis. A, B. Response of a LPFC reward neuron to old stimuli (A) or new stimuli (B). All trials were sorted into four conditions based on the first cue stimulus (A1 group vs A2 group) and on the two reward conditions (large vs small reward) and then aligned with the first cue onset. Red lines in the rastergrams mark the onset and offset of the first cue. The orange curves in the histograms represent data from large reward trials and the blue curves represent data from small reward trials. The gray areas indicate the first cue period for analysis of neuronal activity. Three pairs of new stimuli were tested for this neuron (indicated by the three different colors in B). C, D. Population histograms of LPFC reward neurons that showed significant reward modulation for old stimuli in the first cue period (C) and for new stimuli in the first cue period (D). The gray areas denote the first cue period. E, F. Population histograms of LPFC reward neurons that showed significant reward modulation in the early delay period (the gray area), but not in the first cue period for old stimuli (E) and for new stimuli (F). To make the histograms, trials for each cell were sorted by the preferred reward condition (orange curve) and the non-preferred reward condition (blue curve). In addition, the activity in each trial was normalized by the firing rate during the 500ms fixation epoch before the first cue onset (indicated by the thick black line). (76)
Figure 6. Histograms of activity of an example caudate reward neuron and the caudate and putamen reward neuron populations. In these figures, activity in the first and repetition trials of new stimuli were included for analysis. The formats are the same as those in Figure 5. A, B. Single-unit activity of a caudate reward neuron to old stimuli (A) or new stimuli (B) presented as the first cue in SPATs. This neuron was identified as reward type for both old and new stimuli. C. Population histograms of caudate reward neurons that showed significant reward-modulated activity to old stimuli in the first cue period. The left column represents the population histograms for old stimuli and the right column indicates the population histograms for new stimuli. D. Population histograms of caudate reward neurons that showed significant reward-modulated activity to old stimuli in the early delay period (the gray area), but not in the first cue period. The left column represents the population histogram for old stimuli and the right column indicates the population histograms for new stimuli. E. Population histograms of putamen reward neurons that showed significant reward-modulated activity to old stimuli in the first cue period. The left column represents the population histograms for old stimuli and the right column indicates the population histograms for new stimuli. F. Population histograms of putamen reward neurons that showed significant reward-modulated activity to old stimuli in the early delay period, but not in the first cue period. The left column represents the population histogram for old stimuli and the right column indicates the population histograms for new stimuli. (76)
We were interested in whether the reward neurons in these areas could predict the reward value of the first cue stimulus (particularly for the new stimuli) in SPATs after only experiencing reward instruction with C1 and C2. We focused on the neural activity in the first SPAT block in which the new stimuli or the old stimuli were presented for the first time to each recorded neuron (Fig. 7A and 8A). Notably, in these very first SPATs, the new stimuli were for the first time presented as the first stimulus cues. Therefore, before the end of these specific trials, the monkeys (as well as the neurons) had had no opportunity to directly associate the new stimulus with a particular amount of reward. They also had never experienced any direct pairing between the new stimuli and C1 and C2. As reference activity, we also calculated the activity of each reward neuron to the old stimuli in the first SPAT block after reward instruction. In these trials, the monkeys and neurons had already well experienced the old stimuli in the SPATs with asymmetric reward schedules.

We found that when an old stimulus was presented as the first cue, reward neurons in the LPFC, caudate and putamen discriminated the two reward conditions (large and small reward) right from the first SPAT in both the first trial-sequence (N1;1, Fig. 7A, Fig. 8A and 8E) and the second trial-sequence (N2;1, Fig. 7B, Fig. 8B and 8F). The results of two-way ANOVA (reward vs. trial order) showed that the neural activity to the old stimuli had a significant main effect of reward (P<0.001), but no significant effect of trial order (P>0.3) and no significant interaction between the two factors (P>0.3). Further post-hoc tests with Bonferroni correction revealed that the reward neurons in the three areas showed significantly higher activity in the preferred reward condition than in the non-preferred reward condition (P<0.01).

However, regional differences in response activity were found when the new stimuli were presented as the first cues. The LPFC reward neurons were able to predict the reward values of the new stimuli from the first SPAT after reward instruction in both the first and the second trial-sequences (i.e. in both N1;1 and N2;1; Fig. 7C and D) despite the fact that monkeys had never directly learned the new stimulus-reward contingency. The two-way ANOVA (reward vs. trial order) demonstrated that neural activity seen in Fig. 7C and 7D had a significant main effect of reward (P<0.001), but showed neither effect of trial order (P>0.5) nor interaction (P>0.4). Post-hoc tests with Bonferroni correction suggested that LPFC reward neurons had significantly higher activity in the preferred reward condition vs. the non-preferred reward condition from the first SPAT (P<0.01).
Figure 7. Population activities of LPFC neurons as a function of SPAT order in the first and second trial sequences. Here, the analysis was based solely on the first block of SPATs in which new stimuli (or old stimuli) were presented for the first time for each recorded neuron. A, B, Normalized activity of LPFC reward neurons to old stimuli in the first trial sequence (A) and in the second trial sequence (B). C, D, Normalized activity of the same population neurons to new stimuli in the first trial sequence (C) and in the second trial sequence (D). The normalized activity was sorted into the preferred reward condition (orange curves) and the non-preferred reward condition (blue curves). Statistical significance was determined by Mann–Whitney U test; *p < 0.05; **p < 0.01. Error bars indicate the SEM. (76)
In contrast to the LPFC reward neurons, during the first trial-sequence the reward neurons in the caudate and putamen did not distinguish the preferred from non-preferred reward conditions in the very first SPAT (i.e. in N1,1 trials; Fig. 8C and 8G). Using the two-way ANOVA, we could show the caudate neural activity (as seen in Fig. 8C), had a significant main effect of reward (P<0.001), and a significant interaction between reward and trial order (P=0.002), but no significant effect of trial order (P>0.9). Post-hoc tests with Bonferroni correction revealed that the caudate neurons had no reward-modulated activity in the first trials (P>0.05), but they did show significantly differential activity between the two reward conditions from the second trial in the first-trial sequence (i.e., from N1,2 trials, P<0.01). The activity of the putamen neurons in Fig. 8G showed a significant main effect of reward (P<0.001) and interaction of reward and trial order (P=0.014), but again no significant effect of trial order (P>0.3). A post-hoc test with Bonferroni correction revealed that the putamen neurons showed no reward-modulated activity in the first two trials (i.e., in N1,1 and N1,2 trials; P>0.05), but that the differential activity between the two reward conditions gradually increased from the third trial onwards, and reached a significant level in the fourth trial (P<0.01).

The same populations of neurons in the caudate and putamen predicted reward information for the new stimuli from the first SPATs in the second trial-sequence (i.e. in N2,1 trials; Fig. 8D and H), although the monkeys had never directly learned the new stimulus-reward contingency. The results of a two-way ANOVA showed that neural activity (shown both in Fig. 8D and Fig. 8H) showed significant main effects of reward (P<0.01), but no effect of trial order (P>0.3) as well as interaction (P>0.7) in the second trial-sequence. A post-hoc test with Bonferroni correction additionally showed that both the caudate and putamen neurons had significantly higher activity in the preferred reward condition as opposed to the non-preferred reward condition from the first trial in the second-trial sequence. These results suggest that the striatal reward neurons used the new stimulus-reward contingency presented in the first trial-sequence to predict the reward information for the other stimulus from each pair shown in the second trial-sequence. To examine the consistency of neural response patterns among the three monkeys, we separately calculated the response patterns of the LPFC and striatal reward neurons in the first and second trial-sequences for each of the animals. The results from individual monkeys were consistent with the population data from the three monkeys in the LPFC and the two monkeys in the caudate and putamen.

These results demonstrated that both the LPFC and striatal neurons could transfer reward information paired with C1 and C2 to well-experienced stimuli (A1 and A2) in SPATs. More importantly, the LPFC neurons which responded to reward
value for the old stimuli were also able to infer the reward value of the new stimuli on the basis of reward information associated with C1 and C2. Those neurons did not require direct experience of the new stimulus-old stimulus (C1 or C2) associations, or of the new stimulus-reward contingency, to do so. Striatal neurons which responded to reward value for the old stimuli, (both in caudate and putamen), however, appeared unable to transfer reward information associated with C1 and C2 to the new stimulus. Nevertheless, these striatal neurons could use the reward information associated with one stimulus in the new stimuli-pair to infer that of the other stimulus.
Average activity of caudate reward neurons to old stimuli

Average activity of caudate reward neurons to new stimuli

Average activity of putamen reward neurons to old stimuli

Average activity of putamen reward neurons to new stimuli

n = 50
**Figure 8.** Population activities of caudate and putamen neurons as a function of SPAT order in the first and second trial sequences. Here, the analysis was based solely on the first block of SPATs in which new or old stimuli were presented for the first time for each recorded neuron. *A, B*, Normalized activity of caudate reward neurons to old stimuli in the first trial sequence (*A*) and in the second trial sequence (*B*). *C, D*, Normalized activity of the same caudate neurons to new stimuli in the first trial sequence (*C*) and in the second trial sequence (*D*). *E, F*, Normalized activity of putamen reward neurons to old stimuli in the first trial sequence (*E*) and in the second trial sequence (*F*). *G, H*, Normalized activity of the same putamen neurons to new stimuli in the first trial sequence (*G*) and in the second trial sequence (*H*). (76)

**Reward-modulated effects for old stimuli and new stimuli in the LPFC and striatum**

To compare reward-modulated neuronal activity to the old stimuli with reward-modulated neuronal activity to the new stimuli, we calculated a normalized reward index for each reward neuron (see Materials and Methods). Fig. 9A shows the distribution of reward index values of 106 LPFC reward neurons. It was found that the reward index values were significantly larger for the old than for the new stimuli (Mann-Whitney U test, $P<0.1*10^{-8}$), indicating that reward neurons showed greater activity difference between the preferred and non-preferred reward conditions to the old than to the new stimuli. Even excluding the 13 neurons that showed no reward-modulated activity to the new stimuli, the remaining 93 neurons still had stronger reward effects for the old stimuli relative to the new stimuli (Mann-Whitney U test, $P<0.1*10^{-6}$). We found similar reward-modulated effects in the caudate and putamen. Caudate reward neurons showed significantly greater reward index values for the old stimuli than the new stimuli (Fig. 9B, Mann-Whitney U test, $P<0.01$). This was also true for putamen reward neurons (Fig. 9C, Mann-Whitney U test, $P<0.1*10^{-6}$). Excluding the 5 neurons whose activity was not modulated by reward amount for the new stimuli, the remaining 45 putamen neurons still showed larger reward indexes for the old stimuli than for the new stimuli (Mann-Whitney U test, $P<0.1*10^{-4}$).

We compared further the reward index values in the three brain regions separately for the old and new stimulus conditions. A non-parametric ANOVA test (Kruskal-Wallis test) showed that the reward index values (for both the old and new stimuli) were significantly dependent on the factor of brain area ($P<0.001$). A post-hoc test with Bonferroni correction revealed that the reward index values in the LPFC were significantly smaller than those in the caudate and putamen ($P<0.01$), but that the reward index values in the caudate and putamen did not significantly differ from each other ($P>0.05$). These results indicate that compared to the LPFC neurons, the striatal reward neurons can better discriminate between the preferred and non-preferred reward conditions for the old as well as the new stimuli.
Figure 9. Distributions of reward index values of LPFC, caudate, and putamen neurons. A, Reward index values of LPFC reward neurons for old stimuli (upper) and new stimuli (lower). B, Reward index values of caudate reward neurons for old stimuli (upper) and new stimuli (lower). C, Reward index values of putamen reward neurons for old stimuli (upper) and new stimuli (lower).
Stimulus-Reward neurons in the LPFC and Striatum

Additionally, we also analyzed the neural activity of stimulus-reward neurons with new stimuli, and examined the responsive characteristics of stimulus-reward neurons in the LPFC and the striatum. A representative LPFC stimulus-reward neuron (Fig. 10A) showed significantly higher activity in larger as opposed to smaller reward trials for one old stimulus (in this example, A1, two-tailed t-test, P<0.01), but no differential response was found to the other old stimulus (in this example, A2, two-tailed t-test, P=0.065). The illustrated neuron was tested with three pairs of new stimuli (Fig. 10B). When each of the three new stimuli from the A1-group was presented as the first cue, the activity was significantly modulated by the reward amount (two-tailed t-test, P<0.1*10^{-3}), but not for the three stimuli from the A2-group (two-tailed t-test, P=0.46). So, the neuron appeared to encode not only the reward value of the stimuli, but also the group membership of the stimuli. Here, stimuli from the A1-group are defined as the preferred stimuli for this neuron (recall that the old stimulus A1 is the preferred old stimulus and the three new stimuli associated with B1 are the preferred new stimuli), and stimuli from the A2-group as the non-preferred stimuli.

We also observed stimulus-reward type neural activity to old stimuli in the striatum (Fig. 11). An example neuron showed the reward-modulated activity to the old stimulus A2 presented as the first cue (Fig. 11A, two-tailed t-test, P<0.1*10^{-4}), but no such reward-modulated activity to the old stimulus A1 (two-tailed t-test, P=0.0653). So, for this neuron, A2 is the preferred stimulus and the smaller reward condition is the preferred reward condition. However, this neuron showed neither reward selectivity nor stimulus selectivity when new stimuli were presented as the first cues (Fig. 11B, two-way ANOVA: reward factor P=0.093; stimulus factor: P=0.49; interaction: P=0.88), and recovered the stimulus-reward type activity when old stimuli (A1 and A2) reappeared. This neuron seemed to encode a specific old stimulus-reward relation, and couldn’t generalize this relation to new stimuli even from the same group.

We created population histograms for stimulus-reward neurons sorted by two rewards (preferred reward vs. non-preferred reward) and two stimuli (preferred stimulus vs. non-preferred stimulus) conditions for old stimuli and new stimuli, respectively (Fig. 10C, 10D, 11C, 11D). The population histogram of these stimulus-reward neurons in the LPFC showed the highest activity for the preferred old stimuli in the preferred reward condition (orange solid curve in Fig. 10C). This activity pattern was maintained for the new stimuli from the same group in the same reward condition (the orange solid curve in Fig. 10D), confirming that LPFC
stimulus-reward neurons encode reward information specific to a group of old and new stimuli (the preferred group). The striatal stimulus-reward neurons showed the highest activity to the preferred old stimulus in the preferred reward condition (Fig. 11C); the activity pattern, however, disappeared on presentation of the new stimuli from the same preferred group in the same preferred reward condition (Fig. 11D), indicated that the stimulus-reward neurons in the striatum encoded a specific stimulus-reward association, and did not maintain stimulus and reward preferences across old and new stimuli from the same group.

To further quantify the group information encoded by stimulus-reward neurons in the LPFC and striatum, we also calculated a category index for each stimulus-reward neuron (Fig. 10E, 11E). The category index reflected each neuron’s averaged difference in activity to stimuli from the different groups versus the activity difference to stimuli from the same group. Positive category index values indicated greater activity difference to stimuli across the different groups than within each group, and negative index values indicated the opposite. The category index was computed for each stimulus-reward neuron in the preferred and non-preferred reward conditions. Fig. 10E displays the distribution of category index values in the LPFC. In the preferred reward condition, the category indexes were significantly greater than zero (Wilcoxon signed-rank test, $P<0.1*10^{-4}$), indicating that LPFC stimulus-reward neurons had group selectivity for both old and new stimuli. In the non-preferred reward condition, the neurons conveyed no group information (Wilcoxon signed-rank test, $P=0.98$). Fig. 11E presents the distribution of category index values in the striatum. As can be seen, the category index values did not differ significantly from zero, either in the preferred or in non-preferred reward conditions (Wilcoxon signed-rank test, $P=0.74$ in preferred reward condition and $P=0.64$ in the non-preferred reward condition), indicating that striatal stimulus-reward neurons did not represent any group information about the old and new stimuli in either reward condition.
Figure 10. An example LPFC stimulus-reward neuron and population histograms. A, B, Response of a LPFC reward neuron to old stimuli (A) or new stimuli (B). The formats are the same as those in Figure 5 and 6. A, B, Single-unit activity of a LPFC stimulus-reward neuron to old stimuli (A) or new stimuli (B) presented as the first cue in SPATs. This neuron was identified as stimulus-reward type for both old and new stimuli. C, D, Population histograms of LPFC stimulus-reward neurons that showed significant reward modulation for old stimuli in preferred-stimulus conditions, (C) and for new stimuli in preferred-stimulus conditions (D). E, displays the distribution of category index values in the LPFC. The category indexes were significantly greater than zero in the preferred reward condition.
**Figure 11.** An example Striatal stimulus-reward neuron and population histograms. A, B, Response of a Striatal stimulus-reward neuron to old stimuli (A) or new stimuli (B). A, B, Single-unit activity of a Striatal stimulus-reward neuron to old stimuli (A) or new stimuli (B) presented as the first cue in SPATs. This neuron was only identified as stimulus-reward type for old stimuli. C, D, Population histograms of Striatal stimulus-reward neurons that showed significant reward modulation for old stimuli in preferred-stimulus conditions, (C) but not for new stimuli in preferred-stimulus conditions (D). E, displays the distribution of category index values in the striatum. The category indexes did not differ significantly from zero, either in the preferred or in non-preferred reward conditions.
Identification of broad-spike and narrow-spike neurons

Since cortical neurons can be classified into pyramidal cells and interneurons, we were interested in whether the two types of cells could both encode reward information, then we separated the prefrontal neurons into these two types. This time we added the neural activity data of the forth monkey (Zep) which had been excluded in the previous analysis because we hadn't trained Zep with new stimuli. A total of 493 neurons were selected for analysis based on the four criteria (see Materials and Methods, 91 from Hop, 152 from Tap, 158 from Tom and 92 from Zep). To classify neurons based on their waveforms, the action potential waveforms of each neuron were normalized by the differences between their peak and trough values, aligned by their troughs, and then averaged. Across our sampled neurons, waveforms had similar biphasic shapes but varied in duration, defined here as the time between the trough and peak of the averaged waveform (Fig 12A).

All neurons were classified into two groups by using a clustering method of $k$-means($k=2$) based on their shapes of averaged waveforms (see Materials and Methods). Waveforms of the two classified neurons are separately presented in Fig 12B-E for each monkey. We found that there were 383 (77.7%) BS neurons and 110 (22.3%) NS neurons in the database.

The duration between the trough and peak in the averaged waveform was calculated for each of the neurons. Distributions of the durations were bimodal (Fig 12F-I, Hartigan’s dip statistic test, $P<0.01$ in each of the four monkeys). Durations of NS neurons (blue bars) were significantly shorter than durations of BS neurons (red bars) in each of the monkeys, respectively (for Hop, the median duration of NS neurons: 0.2ms, the median duration of BS neurons: 0.475ms; for Tap, 0.2ms, 0.45ms; for Tom, 0.2ms, 0.525ms; for Zep, 0.2ms, 0.475ms. Mann Whitney U test, $P<0.1*10^{-5}$).

Several different physiological properties were found between the BS and NS neurons. First, the NS neurons showed significantly higher discharges than did BS neurons in all task periods (Fig 13A-C, Mann Whitney $U$ test, $P<0.001$). Second, NS neurons showed faster visual responses than did BS neurons (Fig 13D, the median latencies: 177.0 ms, 145.5 ms for BS and NS neurons, respectively, Mann Whitney $U$ test, $P<0.001$). Third, we did not observe burst firing patterns in either the population of the BS neurons or in the population of the NS neurons to be characterized by local variation (LV) values that were less than one (the median LV values: 0.813, 0.534 for the BS and NS neurons). But the LV values significantly differed between cell types (Mann Whitney $U$ test, $P<0.001$). Fourth, in ISI distributions, peak mode times...
of the BS neurons were significantly longer than that of the NS neurons (the median model time: 32 ms, 15 ms for BS and NS cells, Mann Whitney U test, P<0.001). Overall, the observed differences in spike waveforms, durations, response properties and spike train statistics between the BS and NS neurons were consistent with the distinct characteristics of pyramidal cells and interneurons reported in previous studies, suggesting that most of the BS neurons were pyramidal cells and most of the NS neurons were interneurons in our database.
Fig 12. Identification of BS and NS neurons. A, Average spike waveforms of one BS neuron (red curve) and one NS neuron (blue curve) recorded in an electrode simultaneously. The duration was calculated between the trough and peak of the average waveform. B-E, Average spike waveforms of two groups of neurons classified by k-means (k = 2) classifier for the four monkeys, Hop (B), Tap (C), Tom (D) and Zep (E), respectively. The red waveforms indicate the BS neurons and the blue ones represent the NS neurons. F-I, Distributions of durations for the four monkeys, Hop (F), Tap (G), Tom (H) and Zep (I). The red bars indicate durations of the BS neurons and the blue bars indicate durations of the NS neurons. The distributions were not unimodal (Hartigan’s dip statistic test, P < 0.01).
Fig 13. **Characteristic responses of BS and NS neurons.** A-C, Population histograms of BS (red curves) and NS (blue curves) neurons aligned at the first cue onset (A), the second set of cues onset (B) and the third set of cues onset (C). The shaded areas around the curves indicate SEM. The two gray areas in (A) indicate the cue and delay periods, respectively. The NS neurons showed significantly higher firing rates in all task periods than did the BS neurons (Mann Whitney U test, P < 0.001). D, Cumulative curves of visual response latencies for BS neurons (red curve) and NS neurons (blue curve). NS cells showed faster visual responses than did BS neurons (Mann Whitney U test, P < 0.001). (22)
Classifying BS and NS neurons into Reward and other neurons

Although we have found reward neurons in PFC, when we classified those neurons into BS and NS neurons, it’s unclear whether both BS and NS neurons had reward neurons respectively, therefore we performed a two-way ANOVA (two factors: stimulus (A1 vs. A2) x reward condition (large vs. small)) again to analyze the activity in the cue (100-500ms from the first cue onset) and delay periods (500-900ms from the first cue onset), respectively. Based on results of the two-way ANOVA (P<0.01), both the BS and NS neurons could also be classified into Reward, Stimulus and Stimulus-reward neurons in the cue and delay periods.

In the cue period, there were 59 Reward neurons, 35 Stimulus neurons, and 37 Stimulus-reward neurons out of 383 BS neurons (Table 4). Within 110 NS neurons, 18 of them were Reward neurons, 13 were Stimulus neurons, and 24 were Stimulus-reward neurons (Table 5). We calculated the incidence of reward neurons in the BS and NS classes. The incidence of Reward neurons did not differ significantly between the two classes (BS Reward neuron: 15.4% (59/383), NS Reward neuron: 16.4% (18/110), $\chi^2$-test, F=0.06, df=1, P=0.807).

In the delay period, we found the 383 BS neurons to include 72 (29) Reward neurons, 17 (5) Stimulus neurons, and 24 (10) Stimulus-reward neurons (Table 4). The number in the parentheses indicated the number of neurons that were also identified as the same type in the cue period. The 110 NS neurons included 30 (11) Reward neurons, 2 (0) Stimulus neurons, and 6 (4) Stimulus-reward neurons (Table 5). The incidence of Reward neurons in the BS class did not significantly differ from that in the NS class ($\chi^2$-test, Chi-square=3.239, df=1, P=0.071). These results indicated that the NS neurons, as well as the BS neurons, played functional roles in encoding information of reward in the LPFC. We further analyzed response properties of reward neurons in the two cell classes to find out how these neurons encode reward information.
Table 4 Classification of LPFC broad-spike (BS) neurons in the four monkeys (22)

<table>
<thead>
<tr>
<th>Monkey</th>
<th>BS neurons</th>
<th>BS R-neurons</th>
<th>BS S-neurons</th>
<th>BS SR-neuron</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cue period</td>
<td>Delay period</td>
<td>Cue period</td>
<td>Delay period</td>
</tr>
<tr>
<td>Hop</td>
<td>71</td>
<td>9</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Tap</td>
<td>116</td>
<td>18</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>Tom</td>
<td>125</td>
<td>24</td>
<td>28</td>
<td>5</td>
</tr>
<tr>
<td>Zep</td>
<td>71</td>
<td>8</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>383</td>
<td>59</td>
<td>72</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 5 Classification of LPFC narrow-spike (NS) neurons in the four monkeys (22)

<table>
<thead>
<tr>
<th>Monkey</th>
<th>NS neurons</th>
<th>NS R-neurons</th>
<th>NS S-neurons</th>
<th>NS SR-neuron</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cue period</td>
<td>Delay period</td>
<td>Cue period</td>
<td>Delay period</td>
</tr>
<tr>
<td>Hop</td>
<td>20</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Tap</td>
<td>36</td>
<td>3</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Tom</td>
<td>33</td>
<td>9</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Zep</td>
<td>21</td>
<td>3</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>18</td>
<td>30</td>
<td>15</td>
</tr>
</tbody>
</table>
BS and NS Reward neurons

Population histograms of BS and NS Reward neurons

The reward neurons encoded reward information regardless of the visual properties of the stimulus. In our database, about half of the Reward neurons showed higher activity in larger reward trials, and the other half had higher activity in smaller reward trials. Fig 14A* shows population histograms of 59 BS neurons (red curves) and of 18 NS neurons (blue curves) that were identified as the Reward neurons in the cue period. Fig 14B* shows population histograms of 43 BS neurons and of 19 NS neurons that were identified as the Reward neurons in the delay period but not in the cue period. Both the BS and NS Reward neurons discriminated the preferred from non-preferred reward conditions in the cue (mean normalized activity for BS: 0.50 (preferred), 0.21 (non-preferred), Mann Whitney U test, P<0.001; for NS: 0.57 (preferred), 0.3 (non-preferred), Mann Whitney U test, P<0.001) and delay periods (mean normalized activity for BS: 0.47 (preferred), 0.16 (non-preferred), Mann Whitney U test, P<0.001; for NS: 0.54 (preferred) and 0.22 (non-preferred), Mann Whitney U test, P<0.001).

BS Reward neurons encode the preferred reward information

Next, we examined what type of reward information the BS Reward neurons encoded. To investigate this question, we compared the pre-cue activity (-300-0ms prior to the first cue onset) with the post-cue activity on the same trials (see Materials and Methods). From the population level, these 59 BS Reward neurons significantly increased their firing rates in the late cue period (300-500ms from the first cue onset) compared to firing rates in the fixation period under the preferred reward conditions (see the solid red curve in Fig 14A*, median baseline rate: 5.5 Hz, median discharge rate in the late cue period: 12.4 Hz, Wilcoxon signed rank test, P<0.001), but showed no activity changes under the non-preferred reward conditions (see the dashed red curve in Fig 14A*, median baseline rate: 4.8 Hz, median discharge rate in the late cue period: 4.1 Hz, Wilcoxon signed rank test, P=0.874). The activity of individual neurons in the fixation and late cue periods was shown in the preferred (Fig 14C*) and non-preferred (Fig 14D*) reward conditions, respectively. Consistently, about 80% of BS Reward neurons (47/59) showed significantly increased activity in the late cue period relative to the baseline activity (Mann Whitney U test, P<0.05) in the preferred reward condition (see filled circles above the diagonal line in Fig 14C*). In contrast, the majority of these neurons (72.3%, 45/59) had no significant activity changes between the two periods in the non-preferred reward condition (see open circles in Fig 14D*, P>0.05). The incidence of significant neurons was higher in the preferred reward condition (preferred: 79.7%, non-preferred: 23.7%, χ²-test, Chi-square=36.958,
The proportion of insignificant neurons was lower in the preferred reward condition (preferred: 15.5%, non-preferred: 79.7%, $\chi^2$-test, Chi-square=49.076, df=1, P<0.01).

The BS Reward neurons identified in the delay period had the same response pattern as did those BS Reward neurons in the cue period. We observed the population activity of these 72 BS Reward neurons were significantly higher in the delay period than in the fixation period under the preferred reward condition (baseline: 4.8 Hz, the delay period: 11.3 Hz, Wilcoxon signed rank test, P<0.001), and the activity in the two periods did not significantly differ in the non-preferred reward condition (baseline: 4.2 Hz, the delay period: 3.4 Hz, Wilcoxon signed rank test, P=0.189). In the preferred reward condition, 82% of individual neurons (59/72) had significantly greater activity in the delay period than in the fixation period (see filled triangles above the diagonal line in Fig 14E, Mann Whitney U test, P<0.05). In the non-preferred reward condition, however, over 70% of them (52/72) did not show activity differences in the two periods (see open triangles in Fig 14F, P>0.05). The incidence of significant neurons was higher and the incidence of insignificant neurons was lower in the preferred reward condition ($\chi^2$-test, Chi-square=42.653, df=1, P<0.01). The activity patterns of population and individual BS Reward neurons consistently suggested that the majority of the BS Reward neurons represented the preferred but not the non-preferred reward information in the late cue and delay periods.

**NS Reward neurons encode the non-preferred reward information**

The NS Reward neurons displayed a response pattern that differed from the response pattern of the BS Reward neurons. In the non-preferred reward condition, the population activity of the NS Reward neurons in the baseline period was significantly higher than the activity in both the late cue and delay periods (see dashed blue curves in Fig 14A and B, 18 NS Reward neurons in the cue period: median baseline rate: 27.2 Hz, median rate in the late cue period: 16.6 Hz, P<0.01; 30 NS Reward neurons in the delay period: baseline: 24.9 Hz, the delay period: 10.3 Hz, P<0.001, Wilcoxon signed rank test). In the preferred reward condition, the baseline activity of the NS Reward neurons was not significantly different from the activity either in the late cue period (see the solid blue curve in Fig 14A, baseline: 23.2 Hz, the late cue period: 32.9 Hz, Wilcoxon signed rank test, P=0.094) or in the delay period (see the solid blue curve in Fig 14B, baseline: 23.4 Hz, the delay period: 23.9 Hz, Wilcoxon signed rank test, P=0.271). For individual neurons, the majority of the NS Reward neurons showed significantly higher baseline activity relative to the activity in the late cue period (see filled circles under the diagonal line in Fig 14H, 77.8%, 14/18, Mann Whitney U test, P<0.05) or in the delay period (see filled triangles under
the diagonal line in Fig 14J, 86.7%, 26/30, P<0.05) under the non-preferred reward condition. In contrast, under the preferred reward condition, the baseline activity of the majority of the same population neurons were not significantly different from the activity either in late cue period (see open circles in Fig 14G, 72.2%, 13/18, Mann Whitney U test, P>0.05) or in the delay period (see open triangles in Fig 14I, 70%, 21/30, P>0.05). In the both late cue and delay periods, the incidences of significant neurons were lower and the incidences of insignificant neurons were higher in the preferred reward condition relative to that in the non-preferred reward condition (\(\chi^2\)-test, P<0.01: for the late cue periods: Chi-square=19.935, df=1; for the delay periods: Chi-square=9.942, df=1). The response patterns of the NS Reward neurons suggested that the majority of the NS Reward neurons represented the non-preferred but not the preferred reward information in the late cue and delay periods.
Spikes/Sec.  
Time from the first cue onset (ms)  
Activity in the baseline period (spike/sec.)  
Activity in the late cue period  
BS (n=59)  
preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
NS (n=18)  
preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
BS (n=43)  
non-preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
NS (n=19)  
 preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
BS (n=72)  
non-preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
NS (n=30)  
non-preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
Spikes/Sec.  
Time from the first cue onset (ms)  
Activity in the baseline period (spike/sec.)  
Activity in the late cue period  
BS (n=59)  
non-preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
NS (n=18)  
non-preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
BS (n=43)  
preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
NS (n=19)  
 preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
BS (n=72)  
non-preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period  
NS (n=30)  
non-preferred reward  
Activity in the baseline period (spike/sec.)  
Activity in the delay period
Fig 14. Response properties of the BS and NS R-neurons in the cue and delay periods. A, Population histograms of the BS (red curves) and NS (blue curves) R-neurons identified in the cue period. B, Population histograms of the BS and NS R-neurons identified only in the delay period not in the cue period. The activity of each neuron was sorted by the two reward conditions: the preferred reward condition (solid curves) and the non-preferred reward condition (dashed curves). The shaded areas around the curves indicate SEM. The gray area in (A) is the cue period, and the gray area in (B) is the delay period. C-D, Scatterplots of the baseline activity of the BS R-neurons against the late cue activity in the preferred (C) and non-preferred (D) reward conditions. E-F, Scatterplots of the baseline activity of the BS R-neurons against the delay activity in the preferred (E) and non-preferred reward conditions. G-H, The activity of the NS R-neurons in the late cue period against the baseline activity in the preferred (G) and non-preferred (H) reward conditions. I-J, The activity of the NS R-neurons in the delay period against the baseline activity in the preferred (I) and non-preferred (J) reward conditions. Filled circles and triangles indicate statistical significance (sig., Mann Whitney U test, P < 0.05) and open ones indicate no statistical significance (n.s., P > 0.05). (22)
Discussion

We observed reward neurons in the LPFC, caudate, and putamen. Reward neurons in these three areas encoded reward-related information independent of the visual properties and the group membership of stimuli, a type of neural activity that has previously been reported\(^{92,48,53}\). However, we additionally found that the LPFC reward neurons (defined by reward modulated responses to old stimuli) were able to infer reward values for new stimuli which were presented in the very first SPATs for both first and second trial-sequences. In contrast, the observed striatal reward neurons (again defined by reward modulated responses to old stimuli) could not predict reward values for new stimuli in the very first SPATs in the first trial-sequences, whereas in second trial-sequences, they could. These results suggest that the recorded neurons in the LPFC and striatum have different reward prediction mechanisms. Furthermore, when we classified LPFC neurons into BS (putative pyramidal cells) and NS (putative interneurons) groups, we found both the BS and NS neurons in the LPFC distinguished one reward condition from the other reward condition, indicating that both of them were involved in reward processing. The BS Reward neurons raised their firing rates to represent the preferred reward information, while the NS reward neurons reduced their discharge rates to represent the non-preferred reward information. The results suggest that BS and NS neurons encode the preferred and non-preferred information via distinct mechanisms in the task.

Inference and category

Throughout the reward-instructed SPATs with old stimuli (A1 and A2 as the first cues), the monkeys extensively experienced the stimulus-reward contingencies' reversals block-by-block. E.g. in one block, the A1-group was associated with the larger reward, and the A2-group with the smaller reward, but in other blocks, this schedule was sometimes reversed. In this paradigm, it was possible for the monkeys to apply a conditional discrimination strategy to predict the reward amount for old stimuli. The monkeys, for instance, might learn conditional stimulus-reward associations: if C1→LR (larger reward), then A1→LR, A2→SR (smaller reward), and so on, and memorize all of the conditional associations in a virtual look-up table. By searching through such a table, the monkeys could easily determine which stimulus (A1 or A2) would be paired with a larger reward after reward instruction trials with C1 or C2.
The key advantage of the current task design was the introduction of the new stimuli, which prevented any conditional discrimination strategy. The monkeys learned associations between the new stimuli and B1 or B2 in a symmetric reward paradigm, and in the training sessions the new stimuli were not directly paired with either an asymmetric amount of reward or the third cues (C1 and C2). Therefore, when the new stimuli were presented for the first time in SPATs as first cues, the monkeys could not retrieve the new stimuli-reward associations from the virtual look-up table. Thereby, the task with new stimuli ruled out the possibility that the monkeys simply relied on memory to predict the amount of reward. On the contrary, the monkeys had to integrate several independently acquired associations to infer the reward outcomes of new stimuli. The task with the new stimuli also demonstrated that although recorded LPFC and striatal neurons showed similar response patterns to the well-experienced old stimuli, they could be differentiated by their response patterns to the new stimuli.

In our task, the prefrontal stimulus-reward neurons showed their preference to stimuli pertained to the same group (A1 or A2 group), whereas the striatal stimulus-reward neurons didn’t. There were 2 possibilities of information that stimulus-reward neurons in the LPFC might represent, one was the relation between C1/C2 and LR/SR, which meant the first cue in the SPAT always reminded the monkeys of the final result (e.g. C1 was LR) of each trial, the other was the relation between C1 group/C2 group and LR/SR, which meant the first cue invited the monkey to ponder the stimulus group information and its reward contingency in each block (e.g. C1 stimulus group/red group was LR in current trial). The latter case is obviously categorical representation in the LPFC, the former is also a categorical representation since all the stimuli used as the first cues are equivalent to either C1 or C2 in the monkeys’ minds. The monkeys could use a prototype stimulus (e.g. C1) of a stimuli group (e.g. C1/Red stimulus group) to name such category, just as in Japan, people are using Hotchkiss which is merely a model number to generally refer to stapler. Our findings of categorical coding in the LPFC are in line with many other studies for instance, in Sakagami’s experiment, they found ventrolateral prefrontal neurons showed almost identical activities to both color stimulus green and purple if the two colors had been paired with the same action requirement (e.g., Go/No–go)90. Further, those neurons did not simply discriminate Go/No–go action because if they were using motion cues in Go/No–go task those neurons were no longer showing different activities between Go and No–go trials. In short, those VLPFC neurons grouped color stimuli based on their significance or meaning in the task rather than their physical appearances. The lack of category information in the striatum was also proved by other studies comparing neural activities between the PFC and the striatum in an
abstract category learning task. The researchers found in the late learning stage (category-performing stage), striatal neurons couldn't represent category information whereas prefrontal neurons did\textsuperscript{14}.

Besides, it's proved to be possible that hippocampus could concatenate 2 elements, which were not actually linked to each other (e.g., A-C), based on prior knowledge that both of the 2 elements had been linked with another element before (e.g., A-B, B-C)\textsuperscript{96,115}. In Shohamy's fMRI experiments, participants first learned that, e.g. stimulus $X_1$ is paired with $X_2$, stimulus $Y_1$ is paired with $Y_2$, then participants learned $X_1$ is paired with a reward and $Y_1$ is paired with nothing. At last participants were asked to choose between $X_2$ and $Y_2$, despite the fact that $X_2$ and $Y_2$ had never been paired with reward or non-reward directly, participants showed their preference to $X_2$. The association between $X_2$ and reward emerged in the learning phases rather than testing phases, and hippocampus supported the dynamic integration of overlapping stimuli ($X_1$ and reward were both linked with $X_2$). However, different with the prefrontal neurons, neurons in the hippocampus could predict C after seeing A without any direct experience that had linked A to C, however, no categorical information was found to be encoded in the hippocampus.

Albeit our task was a transitive inference task and the monkeys could predict the reward by a deductive approach that always recall the successive corresponding stimulus after seeing each cue (e.g., recall B1 after seeing N1, then recur C1 after recalling B1, and definitely come up with the information that C1 is LR/SR), the emergence of categorical information of the stimuli represented by stimulus-reward neurons provided the monkeys an alternative inductive approach to complete the task, i.e., generalized an abstract rule that all the stimuli $A_1$, $B_1$, $C_1$ and $N_1$ belonged to $C_1$ group/red group by repeatedly observing the relationships between those stimuli. Then after seeing the 1\textsuperscript{st} cue (e.g., $N_1$) the monkeys could recall that $N_1$ was an element of $C_1$ group, further by recalling the prior knowledge that $C_1$ group was LR/SR, the monkeys would predict the reward outcome of the current trial.

It's reasonable to consider such categorical information made it possible to perform transitive inference in the LPFC rather than in the striatum. Stimuli in the same group were bound together according to their functional equivalence\textsuperscript{119}, and were encoded by the neurons in the LPFC which could predict/infer reward outcomes as well. Pigeons were found to be able to use categorical information to guide their behavior. Vaughan trained pigeons to discriminate 2 photographic slides sets (Set A and Set B) consisted of photos of trees\textsuperscript{105}. Pigeons at first were trained to response to Set A rather than Set B, later, the response contingency reversed and then reveried again repeatedly. If the pigeons got familiar with the task, the first 4~5 trials were enough for the pigeons to response correctly afterwards when the stimulus-response
contingency was shifted. They could pass the feature (response or non-response) of some stimuli to the rest of the stimuli in the same category without experiencing all the stimuli in Set A and B by trial and error. Categorical information emerges from experiences and can in turn guide behavior when subjects are confronted with new circumstances in the form of inferences. Inferences don’t need to employ categorical information, but categorical information may facilitate inferences.

Category is an efficient and flexible way to store diverse information. Features of one category can be extended when new members containing new features are added to that category. Soon old members carry forward those new features. Exchanging features between new and old members is a process of inferences, and the results of inferences don’t need always to be true. Category can emerge from functional equivalence but don’t restrict equivalence to members which makes categorical representation flexible. In fact, individual neuron in the PFC can represent multiple categories (e.g. animals and cars) 14. In deep learning neural networks of a language model, the learned word vectors for Tuesday and Wednesday were found to be similar to the word vectors for Sweden and Norway50. That means both a node in an artificial neural network and a real neuron in a biological neural network can represent different categories. Extension to new features or new categories helps animals to generate new information in their minds.

So far, we know basal ganglia can encode value information of stimuli which could be used by subjects to infer reward outcomes of a stimulus. Specifically, different parts of the striatum can encode different types of value, such as flexible and stable values37. Disjunctive inference processes found in the striatum is intriguing compared with other recent studies. Value-based reward prediction is a common paradigm in investigating functions of basal ganglia. The basal ganglia can’t master a complicated model of the environment, e.g. categorical information of the stimuli, however it can grasp some basic rules of our experimental design (if one group of stimuli is linked to the larger reward, then the other must be linked to the smaller reward). The striatum does not need dopaminergic prediction error to learn the reward outcomes, rather it can provide predictive information to dopaminergic neurons. In an experiment similar to our disjunctive inference paradigm, if a reward outcome was inferable, despite the truth that a stimulus was linked to a large/small reward for the first time, dopamine neurons responded less astonished88. Hence, merely focusing on value representation of the striatum is not enough, in the future we have to dig deep into the mechanism that how the striatum learns the structures of complex external environment without prediction error signals. Although, anatomically there are dense connections from the LPFC to the striatum, it doesn’t seem the LPFC have sent all its information to the striatum, otherwise, the caudate
or the putamen could have predicted the reward in the very first trial with new stimulus. The results suggest that the reward information is not processed first in one area and then send to the others. Since the putamen needed more than one trial to learn the new stimulus-reward contingency compared to the caudate, these three areas might have processed reward information independently. Further investigation is needed to clarify what kind of information is transferred among these areas.

Prefrontal BS and NS reward neurons

LPFC neurons were classified into the BS and NS groups by k-means method based on their shapes of averaged spike waveforms. The observed differences in physiological responses between the two classes of neurons indicate that they resembled putative pyramidal cells and interneurons. First, the distributions of waveform durations were bimodal (see Fig12. F-I22). The NS neurons had shorter durations, and the BS neurons had longer durations. Intracellular and whole-cell recording studies with histological methods have demonstrated that pyramidal cells have broader waveform durations than do interneurons61,68,10,28. Therefore, most of the BS neurons in current study should be pyramidal cells and most of NS neurons should be interneurons. Second, the NS neurons showed significantly higher firing rates (see Fig13. A-C22), which is a typical characteristic of interneurons10,28,13. Intracellular studies have found that interneurons have higher firing rates than pyramidal cells when stimulated by current injection or visual stimuli62. The mean baseline activity of BS neurons (5.4 Hz) was also in line with the typical mean rate of 4-8Hz for pyramidal cells97,114. Third, intracellular and anatomical studies have estimated that roughly 70-80% of cortical neurons are pyramidal cells and that the remaining neurons are interneurons10,9,49,50. Consistent with these findings, we found that 77.7% of our sample neurons were BS neurons, and 22.3% of them were NS neurons. Fourth, the NS neurons showed shorter visually evoked latencies than did the BS neurons. Intracellular experiments have found that interneurons have significantly larger and faster EPSPs than pyramidal cells79. This may cause a lower threshold for spike generation and shorter latencies for EPSP-spike coupling in interneurons. Fifth, the BS neurons had significantly higher local variation values than did the NS neurons, indicating that the BS neurons showed a relatively burst-like pattern95. In addition, the BS neurons showed longer peak mode time in ISI distributions than did the NS neurons. These spike train statistical differences between the two classes of neurons were in agreement with statistical observations in pyramidal cells and interneurons79,57,69. Overall, the distinct properties in spike waveforms, response patterns and spike trains between the two classes of neurons
suggested that our sample of LPFC neurons were classified into putative pyramidal cells and putative interneurons correctly based on their spike waveforms.

Both the BS and NS Reward neurons discriminated the preferred from non-preferred reward conditions (see Fig 14A and B). Their population activities were stronger in the preferred than non-preferred reward conditions. To determine whether a neuron encodes reward information or not, we usually compare the neuronal activity on trials containing different rewards. If one neuron shows activity differences on different rewarded trials, we say this neuron encodes reward information; otherwise, this neuron does not encode reward information. This type of analysis has been often used in the literature of reward experiments, but it does not tell us what type of reward information the reward neuron encodes. To further investigate this issue, the activity of a reward neuron is compared in different time epochs on the same trials.

We compared the activity of BS and NS neurons in the fixation period with the activity in the late cue period and with the activity in the delay period. The neuronal activity at the fixation period did not reflect any reward information. After the first cue presentation, the monkeys were able to predict the reward amount in current trial, and therefore the neuronal activity could encode reward information during the cue and delay periods. Interestingly, the BS Reward neurons increased their discharge rates in the late cue and delay periods relative to the baseline rates in the preferred reward condition, while they did not change their activity in the non-preferred reward condition (see Fig 14C, D, G and H). This pattern of results indicated that the activity of BS Reward neurons encoded the preferred reward information but not the non-preferred reward information. On the contrary, in the non-preferred reward condition the NS Reward neurons decreased their firing rates relative to the baseline rates, but they did not change their activity in the preferred reward condition (see Fig 14E, F, I and J). This pattern of results suggested that the NS Reward neurons encoded the non-preferred reward information, but not the preferred reward information. Our findings indicated that the BS and NS Reward neurons complementarily encoded the reward information, with each type of neuron representing reward information via different mechanisms.

One can argue that the dichotomous activity properties we have found in these BS and NS reward neurons may depend on which period that we had chosen for baseline. We have also used a period from -600ms to -300ms prior to the first cue onset as baseline to compare the neural activities between preferred reward and non-preferred condition for the reward neurons. The NS reward neurons did not significantly decrease their firing rates in the late cue period compared to the baseline rates (P=0.14123) in non-preferred reward conditions, however the results of the other 3
conditions (BS preferred reward, BS non-preferred reward and NS preferred reward) in the late cue period and all the 4 conditions in the delay period were in line with our findings. This may be because we didn’t have enough data of NS reward neurons and such neurons in the late cue period might still contain a component of visually evoked response to the presentation of the first cue.

Some studies that examined responses of both types of neurons to preferred and non-preferred visual stimuli reported more complex response patterns of the BS and NS neurons. In a motion direction discrimination task, both the BS and NS neurons in the LPFC increased their firing rates to the preferred and anti-preferred direction motions relative to the baseline activity in the fixation period. There was no suppressive population activity to the anti-preferred direction motion in the NS neurons. In another numerical categorization task, the BS and NS neurons in the LPFC could encode various numerosities by their graded activity. Responses of both types of neurons to the least-preferred numerosity were depressed compared to the baseline activity. Those observations were inconsistent with our findings that both the BS and NS neurons utilized a binary strategy to encode the preferred and non-preferred information for reward. One possible explanation was that the monkey had to memorize and discriminate several visual stimuli simultaneously in those tasks, which enabled the BS and NS neurons in the LPFC to represent visual information in the complex way. In our task, there were two visual stimuli (A1 and A2) and two reward conditions (large and small reward). The BS and NS neurons were able to represent reward information via this binary method. This type of binary coding strategy was consistent with the task structure.
Acknowledgements

All the results above have been published as original papers in Journal of Neuroscience\textsuperscript{76} and PLoS ONE\textsuperscript{22}.

I really appreciate professional guidance from M. Sakagami and X. Pan.
I’m grateful to M. Koizumi for technical assistance, S. Tanaka, M. Oguchi, O. Yokoyama, N. Watanabe for their comments at lab meetings, M. Kimura, Y. Isomura, Y. Sakai, K. Samejima, T. Matsuda for their advices at seminar and retreats.
Thank you for all the other people met at Tamagawa university.
To my family, thank you for supporting me mentally and physically.
This dissertation is dedicated to whom I’ll always love.
References


18. Diester, Ilka, and Andreas Nieder. "Complementary contributions of prefrontal neuron classes in abstract


76. Pan, Xiaochuan, et al. "Reward inference by primate prefrontal and striatal neurons." Journal of Neuroscience


