Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making

Jung Hoon Sul, Hoseok Kim, Namjung Huh, Daeyeol Lee, and Min Whan Jung

1Neuroscience Laboratory, Institute for Medical Sciences, Ajou University School of Medicine, Suwon 443-721, Korea
2Department of Neurobiology, Yale University School of Medicine, New Haven, CT 06510, USA

SUMMARY

We investigated how different sub-regions of rodent prefrontal cortex contribute to value-based decision making, by comparing neural signals related to animal’s choice, its outcome, and action value in orbitofrontal cortex (OFC) and medial prefrontal cortex (mPFC) of rats performing a dynamic two-armed bandit task. Neural signals for upcoming action selection arose in the mPFC, including the anterior cingulate cortex, only immediately before the behavioral manifestation of animal’s choice, suggesting that rodent prefrontal cortex is not involved in advanced action planning. Both OFC and mPFC conveyed signals related to the animal’s past choices and their outcomes over multiple trials, but neural signals for chosen value and reward prediction error were more prevalent in the OFC. Our results suggest that rodent OFC and mPFC serve distinct roles in value-based decision making, and that the OFC plays a prominent role in updating the values of outcomes expected from chosen actions.

INTRODUCTION

Decisions are influenced by the expectations about the costs and benefits of alternative choices. In an uncertain and dynamically changing environment, therefore, it would be advantageous for a decision maker to continually update such expectations based on the outcomes of previous choices. In reinforcement learning (RL) theory, the estimates for future rewards expected from different actions are referred to as action values, and they are updated according to the errors in predicting the reward obtained at each time step (reward prediction error or RPE; Sutton and Barto, 1998). Decision makers can therefore optimize long-term consequences of their actions simply by choosing the actions with maximum action values. Previous studies have shown that the prefrontal cortex (PFC) is an important brain structure for RL (Buckley et al., 2009; Lee et al., 2007; Rushworth and Behrens, 2008; Rushworth et al., 2007b; Seo and Lee, 2008). Considering that the PFC comprises a large area of the frontal lobe consisting of many heterogeneous regions, it would be important to identify the distinct computational processes served by different regions within the PFC to understand neural mechanisms underlying value-based decision making.
Recently, numerous behavioral, physiological, and neuroimaging studies have found dissociable effects of lesions and regional specialization of neural and blood oxygenation level-dependent (BOLD) signals across different areas of the PFC and anterior cingulate cortex (ACC) while humans or monkeys were performing various decision-making tasks (Buckley et al., 2009; Daw et al., 2006; Lee, 2006; Lee et al., 2007; Rushworth and Behrens, 2008; Rushworth et al., 2007b; Seo and Lee, 2008; Wallis, 2007). As a result, principles for functional specialization within the PFC and ACC are beginning to emerge. For example, the orbitofrontal cortex (OFC) and ACC appear to play particularly important roles in encoding and updating values of expected outcomes (Rushworth et al., 2007a), whereas the lateral PFC appears to be necessary for maintaining in working memory the state representation necessary for identifying optimal choices in a given environment (Buckley et al., 2009; Lee et al., 2007). On the other hand, similar types of neural signals, such as values and animal’s choice, are often found redundantly across different regions of the PFC (Lee, 2006, 2008), but few studies have compared different areas of the PFC under the same experimental condition (c.f., Buckley et al., 2009). Thus, it is still unclear how different regions of the PFC work together for optimal decision making in a dynamic environment.

Compared to the primate PFC, less is known about functional differentiation across different regions of the rodent PFC involved in value-based decision making. In the present study, we investigated this issue by examining temporal dynamics of neural signals related to animal’s choices, their outcomes and action values across different regions within the rat PFC. In particular, we addressed the following issues. First, we tested whether value processing is a function implemented across all regions within the rat PFC. A number of studies have found value-related signals in the OFC of humans, monkeys and rats (Mainen and Kepecs, 2009; O’Doherty, 2004; Rolls and Grabenhorst, 2008; Schoenbaum et al., 2006; Wallis, 2007), indicating its preferential involvement in value processing. However, other parts of the primate PFC, including dorsolateral and dorsomedial PFC, and ACC have been shown to convey value signals as well (Kennerley et al., 2009; Kennerley and Wallis, 2009; Kim et al., 2008; Lee and Seo, 2007; Rushworth and Behrens, 2008; Rushworth et al., 2007b; Rushworth et al., 2009; Seo and Lee, 2007, 2008, 2009), raising the possibility that value processing might be a general characteristic of the PFC-ACC network. To our knowledge, value processing in the PFC other than the OFC has not been investigated in rats. Second, we examined whether the rat PFC conveys neural signals related to RPE, which is the difference between the actual and expected rewards (Sutton and Barto, 1998). Brain imaging studies in humans have found RPE signals in the OFC (O’Doherty et al., 2003). In monkeys, on the other hand, RPE signals have been found in the ACC (Amiez et al., 2005; Matsumoto et al., 2007; Seo and Lee, 2007), but not in the OFC (Kennerley et al., 2009). A recent study in rats also failed to find RPE signals in the OFC (Takahashi et al., 2009). Thus, whether and which part of the rat PFC conveys RPE signals remains unclear. Third, we investigated which part of the rat PFC conveys neural signals predictive of animal’s upcoming actions. To our knowledge, neural signals for future choice of action have not been clearly shown in the rat PFC. To address these issues, we analyzed the activity of neurons recorded from the medial wall of the rat PFC, which consists of the dorsal ACC, prelimbic cortex (PLC) and infralimbic cortex (ILC), and the lateral OFC in rats performing a dynamic two-armed bandit task in which the animals faced two probabilistically-reinforced choices in each trial.

RESULTS

Choice and locomotive behavior

Six rats each performed a total of 17–30 sessions of a dynamic two-armed bandit task (Huh et al., 2009; Kim et al., 2009; Fig. 1A). The animals were allowed to choose freely between two goals that delivered water reward with different probabilities (see Experimental Procedures).
Reward probability for each goal was constant within a block of 35–45 trials, but changed across four blocks without any sensory cues so that relative reward probabilities could be discovered only by trial and error. The animals started to choose the goal associated with a higher reward probability more frequently within 10–20 trials after a block transition, indicating that they quickly detected the changes in relative reward probabilities and adjusted their behaviors accordingly (Fig. 1B). A logistic regression analysis (Huh et al., 2009; Lau and Glimcher, 2005) revealed that a reward obtained from a given choice tended to encourage the animal to make the same choice (Fig. 1C). Thus, animal’s choice behavior was influenced by the history of animal’s choices and their outcomes, with more recent choice outcomes having greater effects.

The animals showed stereotyped movement trajectories on the maze (Fig. 2B). We divided animal’s behavior into five distinct behavioral stages as in our previous study (Kim et al., 2009). They were delay (2.0±0.0 s), go (1.0±0.1 s), approach (1.0±0.3 s), reward (7.6±2.3 s) and return (1.9±0.7 s, mean±SD) stages, with the beginning of a new trial defined as the onset of the delay stage (Fig. 2A). The beginning of the approach stage was determined separately for each session as the time when the animal’s movement trajectory began to diverge depending on the upcoming goal choice of the animal (Fig. 2C–D). Therefore, animal’s movement trajectory was independent of the upcoming choice of the animal during the time period between the reward and the next approach stages (Fig. 2B). On the other hand, the trajectory was different in the early delay stage (1.1±0.6 s, mean±SD) depending on the previous goal choice of the animal (Fig. 2B).

**Neural signals for choice and its outcome**

Single-unit activity was recorded from a total of 730, 751 and 1,148 neurons (≥ 500 spikes during each recording session) in the dorsal ACC, PLC/ILC, and lateral OFC, respectively (Fig. 1D), of six rats. The number of neurons recorded from ILC was small (n=187), and therefore, the data obtained from the PLC and ILC were combined since the results from these two areas were similar (see Supplemental Figure S1). We analyzed neural signals related to various internal and external variables using multiple regression (Corrado and Doya, 2007). We first analyzed neural signals for animal’s choice, its outcome (i.e., reward) and their interaction in the current and previous trials (eq. 1, see Experimental Procedures). Fractions of neurons conveying these neural signals are shown in the left, middle and right panel, respectively, in Fig. 3A. In all three areas, neural signals for animal’s choice \([C(t)]\) were weak before choice was made (in the delay and go stages) but increased steeply after animal’s trajectory diverged to reflect its choice (approach, reward and return stages). The analysis based on a 100-ms time window advancing in 50-ms time steps showed that the choice signal arose approximately 100 (ACC) or 50 (PLC/ILC) ms before the onset of the approach stage only in the mPFC, whereas such signals became significant after the onset of the approach stage in the OFC (see Experimental Procedures for the determination of choice signal onset; Fig. 3B). The choice signal then persisted throughout the following trial so that significant levels of previous choice signal \([C(t-1)]\) were found in the delay, go, approach and even in the reward stage (large circles in Fig. 3A). Overall, both current and previous choice signals were stronger in the ACC than in the other areas (indicated by triangles in Fig. 3A).

Neural signals related to choice outcome \([R(t)]\) increased steeply in all three areas after the outcome of animal’s choice was revealed in the reward stage, and decreased gradually during the subsequent stages. Similar to the signals related to animal’s choice, choice outcome signals also persisted during the next trial so that the previous choice outcome signal \([R(t-1)]\) stayed above chance level in the delay stage. In the OFC, these previous outcome signals were further elevated during the approach and reward stages (Fig. 3A). The neural signal for choice outcome two trials before \([R(t-2)]\) was overall weak but above chance level and somewhat elevated.
during the approach and reward stages in the OFC. Therefore, choice outcome signals were stronger in the OFC than in the other areas.

Signals related to animal’s choice or its outcome alone are not sufficient, and they must be combined to specify how the decision-making strategy should be updated. Indeed, significant choice×outcome interactions \(X(t)\) were observed in the reward stage in all three areas (Fig. 3A), indicating that many PFC neurons conveyed conjunctive choice and outcome signals. Thus, all regions represented the information necessary for evaluating the consequences of choices. The neural signal for previous choice×previous outcome interaction \([X(t-1)]\) was overall weak (Fig. 3A).

**Neural signals for value**

We then examined neural signals related to action values that were estimated using a simple RL model (Sutton and Barto, 1998). Animal’s actual choices were well accounted for by this model (Fig. 1B; see Supplemental Table S1 and S2), suggesting that this model estimated animal’s subjective values well. Using action values estimated with this RL model, therefore, we examined neural signals related to decision value \(\Delta Q(t)\), which was defined as the difference between the left and right action values, and chosen value \(Qc(t)\) which is the value of the chosen action in the current trial. Whereas action values represent estimates of future rewards expected from different actions (left and right goal choices), decision value represents their difference, hence the relative desirability of left vs. right goal choices. Neural signals for decision value, therefore, would be useful for action selection. Neural signals for chosen value, on the other hand, would be useful to update chosen value according to the outcome of animal’s choice, because the information on chosen value (estimated future reward) and choice outcome (actual reward) can be combined to compute RPE. Chosen value can be increased or decreased depending on whether the actual outcome is better (positive RPE) or worse (negative RPE) than itself (Sutton and Barto, 1998).

We ran a multiple regression analysis that included these value terms in addition to previous choice and previous outcome, which were correlated with decision value and chosen value, respectively (Supplemental Figure S2). This model also included autoregressive terms to control for autocorrelations in spike counts (Supplemental Table S4), and current choice, current outcome and their interaction as explanatory variables (eq. 2). Signals related to decision value were weak, but nevertheless significant (large circles in Fig. 4A) with no significant variation across regions before the animal committed to choose a particular goal (delay and go stages). For example, during the last 1 s of the delay stage, significant numbers of ACC (76, 10.4%, binomial test, \(p<0.001\), PLC/ILC, (61, 8.1%, \(p<0.001\), and OFC (95, 8.3%, \(p<0.001\)) neurons modulated their activity according to the decision value. Similar results were obtained when the neural data were analyzed with the decision value replaced with action values in the regression (eq. 2). Action value signals were weak, but significant before animal’s choice of action with no significant variation across regions (Fig. 4B). Neural signals related to chosen value were weak before animal’s goal choice (delay and go stages), but increased during the approach and early reward stages and gradually decayed thereafter (Fig. 4). Although significant fractions of neurons modulated their activity according to the chosen value in all three areas in the approach and reward stages (binomial test, large circles in Fig. 4), the fraction was significantly greater in the OFC (\(\chi^2\)-test, triangles in Fig. 4).

The above analyses showed that all the components necessary for updating chosen value, namely the information about animal’s choice, its outcome and chosen value, converge in multiple areas of the PFC during the reward stage. The time courses of these neural signals are shown at a higher temporal resolution in Fig. 5A. Neural signals for animal’s choice and chosen value arose at least 0.5 s before the outcome of animal’s choice was revealed and slowly decayed during the reward stage, and this was particularly noticeable in the OFC. Choice
outcome signals then appeared at the onset of the reward stage so that signals related to choice, its outcome and chosen value temporally overlapped during the initial phase of the reward stage. Thus, chosen value signal was uploaded before the choice outcome signal arrived in the OFC and they overlapped briefly in the early reward stage (an example is shown in Fig. 5B). The overall pattern was similar in the ACC and PLC/ILC, but neural signals for chosen value were much weaker in these areas compared to the OFC. During the first 1 s of the reward stage, a larger fraction of OFC neurons (235 out of 1,148, 20.5%) significantly modulated their activity according to chosen value than ACC (96 out of 730, 13.2%, $\chi^2$-test, $p<0.001$) and PLC/ILC (101 out of 751, 13.4%, $p<0.001$) neurons (eq. 2). Similarly, the fraction of neurons encoding both chosen value and choice outcome during the first 1 s of the reward stage was significantly higher in the OFC (11.6%, 133 out of 1,148 neurons) compared to the ACC (5.3%, 39 out of 730, $\chi^2$-test, $p<0.001$) and PLC/ILC (6.3%, 47 out of 751, $p<0.001$). We therefore limit the following description of neural activity related to RPE and updated chosen value to the OFC (for the results from the mPFC, see Supplemental Table S3 and Supplemental Figure S3).

**Neural signals for RPE and updated chosen value**

For the majority of the 133 OFC neurons that significantly modulated their activity according to chosen value as well as choice outcome, activity was better accounted for by the model containing RPE (eq. 3) than that containing updated chosen value (eq. 4; 83 out of 133, 62.4%; $\chi^2$-tests, $p<0.001$), indicating that OFC neurons largely encoded RPE during the initial phase of the reward stage. Because RPE and updated chosen value are computed by the difference between and weighted sum of choice outcome and chosen value, respectively $[\text{RPE} = R(t) – Qc(t)]$ and updated chosen value $= R(t) + (1-\alpha)Qc(t)$, where $\alpha$ is learning rate and $0<\alpha<1$; see Experimental Procedures], those neurons with the opposite signs of the coefficients for chosen value and choice outcome are expected to modulate their activity according to RPE. Conversely, those neurons with the same signs of their coefficients are expected to modulate their activity according to updated chosen value. Indeed, the 133 OFC neurons, except one, modulated their activity according to RPE or updated chosen value as predicted by the relative signs of the regression coefficients associated with chosen value and choice outcome during the first 1 s of the reward stage (Fig. 6). Temporal profiles of the activity of these OFC neurons around the time of reward delivery are shown in Fig. 7A according to the signs of the coefficients for chosen value and choice outcome (see Supplemental Figure S4 for individual examples).

Mean normalized activity of these OFC neurons during the first 1 s of the reward stage showed largely linear relationships with RPE (positive RPE-slope neurons: $n=44$, $r=0.979$, $p<0.001$; negative RPE-slope neurons: $n=40$, $r=-0.951$, $p<0.001$) or updated chosen value (positive updated value-slope neurons: $n=14$, $r=0.978$, $p<0.001$; negative updated value-slope neurons: $n=35$, $r=-0.9804$, $p<0.001$; Fig. 7B).

To examine whether individual OFC neurons encode positive and negative values of RPE consistently, we divided trials into rewarded and unrewarded trials (corresponding to positive and negative RPE, respectively) and separately calculated coefficients for RPE using spike counts during the first 1 s of the reward stage (eq. 3). It should be noted that the regression coefficients for RPE and chosen value are the same when rewarded and unrewarded trials are separately analyzed. The slope of the linear regression between the coefficients for positive and negative RPE was close to 1 ($0.926$, $n=84$, $p<0.001$; Fig. 8) for the neurons that significantly modulated their activity according to chosen value and choice outcome with opposite signs of coefficients (eq. 2), indicating that these OFC neurons encode RPE symmetrically across positive and negative domains.

Neuron. Author manuscript; available in PMC 2011 May 13.
Although these results were based on a simple RL model, additional analyses confirmed that our findings were independent of the use of a specific RL model. First, we analyzed the same neural data using several different RL models, and obtained similar results. Second, analyses using only steady-state neural data (last 20 trials in each block) yielded similar results. Third, a model-free analysis in which trial-by-trial action values were replaced with block reward probabilities also yielded consistent results (Supplemental Tables S1–S3 and Supplemental Figure S3). Finally, responses of OFC neurons to the same choice outcome were modulated by the previous choice outcome in a manner consistent with RPE signals (Supplemental Figure S5).

**DISCUSSION**

We compared temporal dynamics of neural signals related to values and animal’s choices in the lateral OFC and mPFC of rats performing a dynamic two-armed bandit task. Before the animal chose its action, all regions conveyed significant signals for decision value, namely the relative difference in the rewards expected from the two options. However, discrete neural signals predicting the upcoming choice were weak and arose immediately before the behavioral manifestation of animal’s choice only in the mPFC, suggesting that rodent PFC may not be involved in advanced movement planning. Once the animal’s locomotive trajectory diverged, neural signals for chosen value arose strongly in the OFC, and they were combined with signals related to animal’s choice and its outcome. As a result, many neurons in the OFC changed their activity according to RPE as well as updated chosen value, suggesting a major role of the OFC in updating action values. On the other hand, although mPFC neurons conveyed significant neural signals related to animal’s choice, its outcome and their interaction over multiple trials, they carried relatively weak chosen value signals, suggesting that the mPFC plays a relatively minor role in updating action values and these signals might be used primarily for other purposes than updating values.

**Role of PFC in action selection**

Neural signals related to animal’s upcoming choice were weak in both OFC and mPFC, suggesting that rodent PFC does not make a significant contribution to advanced movement planning and action selection. This was unexpected, because significant neural signals for upcoming choice of action during a free-choice task have been found in the primate PFC (Seo and Lee, 2008), supplementary eye field (Coe et al., 2002) and ACC (Seo and Lee, 2007). We cannot exclude the possibility that action selection is served by only a small subset of neurons in the rodent PFC. In addition, behavioral tasks used in our study were substantially different from those used in previous monkey studies. Whereas rats navigated towards a branching point before committing their goal choices in our study, monkeys in previous neurophysiological studies were simply required to direct their gazes or move a handle to register their choices. Previous studies on striatal neuronal activity in rats performing a free-choice task with lever press (Ito and Doya, 2009) or spatial navigation (Kim et al., 2009) found similar patterns of choice-, outcome-, and value-related neural signals, suggesting that the motor response to register animal’s choice may not be an important factor in determining neuronal activity related to these choice-related variables. Nevertheless, experimental settings are quite different between monkey and rat studies, and we cannot rule out the possibility that the mode of choice expression has a greater effect on PFC compared to striatal activity. For example, if rats are allowed to make a choice without a need for navigation, then neural signals for value and choice might appear earlier. Finally, it is possible that the ACC and PLC/ILC are directly involved in action selection, but the animals chose future actions only immediately before behaviorally revealing their choices. We consider this last scenario unlikely, however, because compared to the PFC, signals related to the upcoming action tend to arise earlier in the dorsomedial striatum (Kim et al., 2009) and rostral medial agranular cortex (RMAC; Sul et al., unpublished...)
observation), which has been proposed to be a rodent homolog of primate supplementary motor area (Donoghue and Wise, 1982; Neafsey et al., 1986; Reep et al., 1990; Sanderson et al., 1984). It will be important in the future to compare temporal dynamics of action selection signals across different areas including the PFC, RMAC, basal ganglia (Ito and Doya, 2009; Kim et al., 2009; Kim et al., 2007), and superior colliculus (Felsen and Mainen, 2008) in rats under the same experimental condition in order to reveal the brain regions directly responsible for future action selection.

Whereas clear preparatory signals for action selection were not found in the rat PFC, significant decision value signals were found in all areas examined in the present study. Thus, the rodent PFC might influence action selection indirectly by conveying relative values of potential choices. However, signals related to decision value identified in the present study were largely due to the changes in neural activity across blocks (see Supplemental Figure S6). Thus, the evidence for neural signals closely tracking decision value signals on a trial-by-trial basis before the animal’s selection of its action is relatively weak. It remains to be determined whether the rodent PFC would still encode decision value signals when they change more dynamically than in the present study.

Role of OFC in updating chosen value

The chosen value signal was particularly strong in the OFC and it overlapped temporally with neural signals for animal’s choice and its outcome following the outcome of animal’s choice was revealed. Thus, all the signals necessary to update chosen value converged in the OFC, suggesting a prominent role of the OFC in updating the value of chosen action. Our results are consistent with a large number of studies demonstrating OFC activity related to expected outcomes in various animal species (O’Doherty, 2007; Rolls, 2000; Schoenbaum et al., 2006; Wallis, 2007). In addition, a large body of behavioral studies implicated the OFC in adaptive modification of choice behavior according to the changing values of expected outcomes, such as in reversal learning and reinforcer devaluation (reviewed in Rushworth et al., 2007a; Schoenbaum et al., 2006). Inability to update values associated with potential choices would prevent the animals from making optimal choices when values associated with potential choices change. The results from our study suggest that in addition to encoding expected outcomes (Takahashi et al., 2009), the OFC might be also actively involved in updating chosen value. The role of rodent OFC in evaluating choice outcomes has previously been proposed based on the finding that some OFC neurons jointly encode choice and outcome signals (Feierstein et al., 2006; but see Furuyashiki et al., 2008). However, the exact nature of chosen value signals observed in the present study remains to be determined. The OFC, especially the lateral division, has been proposed to represent stimulus-specific values, but not action-specific values in different animal species (O’Doherty, 2007; Ostlund and Balleine, 2007; Rushworth et al., 2009). The chosen value signal found in the present study might represent either the value of chosen action or the value associated with sensory stimuli the animal encountered at each goal location.

OFC activity related to reward prediction error

In the present study, both model-based and model-free analyses yielded converging evidence for bidirectional RPE signals in the OFC. Although positive vs. negative RPE is confounded with the choice outcome, consistent results were obtained even in the analysis in which the current choice outcome was fixed (Supplemental Figure S5). Although previous studies have found weak or no RPE signals in the OFC (Kennerley et al., 2009; Rosenkilde et al., 1981; Takahashi et al., 2009; Thorpe et al., 1983), the behavioral tasks employed in those studies were not ideal for detecting signals related to RPE. For example, early recordings in the primate OFC have reported a small percentage (<2%) of neurons with burst of activity related to unexpected negative outcomes during the reversal learning of go/no-go visual discrimination.
Thorpe et al., 1983) or the delayed response task (Rosenkilde et al., 1981). Similarly, more recent studies in rodents (Takahashi et al., 2009) and primates (Kennerley et al., 2009) have failed to find the evidence for RPE signals in the OFC. It should be noted, however, that the reversal task is not particularly sensitive for studying the neural activity related to RPE, since for well-trained animals, large RPE occurs only immediately after the reversal and therefore infrequently. In addition, the animals tested by Kennerley et al. (2009) did not need to update the values associated with different stimuli during the task, since the parameters (e.g., magnitude and probability) of the rewards associated with different stimuli were fixed and familiar to the animal.

It is notable that midbrain dopamine neurons, but not OFC neurons, in rats showed RPE-related activity in the same reversal task (Takahashi et al., 2009). Dopamine neurons might encode RPE with a higher signal-to-noise ratio than OFC neurons so that RPE-related neuronal activity is more readily detectable in dopamine neurons in a reversal paradigm. Alternatively, OFC neurons might change their activity very rapidly after reversal so that analyzing neural activity averaged across several trials after block transition (Takahashi et al., 2009) might dilute RPE-related neural activity in the OFC. Either way, these results collectively suggest that midbrain dopamine neurons and OFC neurons convey different types of RPE signals.

In many theories of learning (Rescola and Wagner, 1972; Sutton and Barto, 1998), RPE is a scalar quantity that can be negative (actual outcome is worse than expected) or positive (actual outcome is better than expected), and can be used to both facilitate and suppress a subsequent choice behavior. However, neurons quantitatively encoding RPE across positive and negative domains have been reported only in the rodent striatum (Kim et al., 2009), although neuronal activity resembling RPE has been reported in several different regions of the primate brain (Amiez et al., 2005; Belova et al., 2007; Hong and Hikosaka, 2008; Matsumoto et al., 2007; Schultz, 1998; Seo and Lee, 2007). Our results show quantitative and bidirectional encoding of RPE by individual neurons in the brain structure (OFC) that is likely to play a key role in representing and updating values (Mainen and Kepecs, 2009; Rolls and Grabenhorst, 2008; Rushworth et al., 2007a; Schoenbaum et al., 2009). Therefore, the brain might update values bidirectionally based on a common neural process as postulated by numerous learning theories instead of relying on separate neural processes for increasing and decreasing values. This could be implemented through activity-dependent synaptic plasticity (Bear, 1995), for example, by changing synaptic weights for down-stream value-coding neurons in different directions depending on the strength of the activation caused by inputs from RPE-coding OFC neurons.

It has long been thought that dopamine neurons broadcast RPE signals to widespread areas of the brain so that error-based learning can take place in dopaminceptive areas such as the frontal cortex and basal ganglia (Schultz, 1998, 2006). However, because dopamine neurons are limited in conveying negative RPE signals (Bayer and Glimcher, 2005; Morris et al., 2004), bidirectional RPE signals found in the OFC cannot be fully explained by dopaminergic inputs, which might also serve other functions such as incentive salience (Berridge, 2007; Matsumoto and Hikosaka, 2009). Moreover, the fact that chosen value signals were available in the OFC before choice outcome was revealed is more consistent with the possibility that RPE is computed de novo in the OFC through the convergence of signals related to chosen value and choice outcomes. We have shown previously that the dorsomedial striatum conveys stronger chosen value and RPE signals than the ventral striatum (VS) in rats (Kim et al., 2009). Anatomical studies have shown that striatal projections of the OFC (excluding agranular insular areas) are directed mostly to the dorsal striatum (DS) in rats (Berendse et al., 1992; Schilman et al., 2008), and behavioral studies have shown that inactivation/lesions of either the OFC or dorsomedial striatum impair reversal learning in rats (Ragozzino, 2007). Together with the present findings, these results suggest that the OFC-DS network might play a central role in updating values in rats.
Past choice and past outcome signals in the mPFC

In RL theory, values of alternative actions are computed based on the history of animal’s choices and their outcomes. Because the learning rate ($\alpha$) of the animals was relatively high in the present study (see Experimental Procedures), action values in a given trial were largely dominated by the animal’s choice and its outcome in the previous trial. The ACC and PLC/ILC conveyed strong signals for the animal’s choice and its outcome in the previous trial. Thus, the ingredient signals to compute chosen value were available in the mPFC. Nevertheless, compared to the OFC, chosen value signals were weaker in the mPFC. These results suggest that the previous choice and previous outcome signals might be stored temporarily in the mPFC and contributed indirectly to computing values. Our results might appear at variance with previous physiological studies that reported RPE-related neural activity in the primate ACC (Amiez et al., 2005; Matsumoto et al., 2007). However, only a small fraction of ACC neurons encoded quantitative RPE (graded responses to different levels of RPE) in those studies (17 out of 372, 4.6% in Amiez et al., 2005; 16 out of 351, 4.6% in Matsumoto et al., 2007), which is comparable to the present finding in the rat ACC (25 out of 730, 3.4%; Supplemental Figure S3 and Table S3).

The previous choice signal might reflect spatial working memory functions of the rat mPFC. Lesions to the mPFC, especially the PLC, are known to impair spatial working memory in rats (Vertes, 2004, 2006). The previous choice signal, which was particularly strong in the ACC, might be also used to bridge the temporal gap between animal’s choice of action and its outcome. Because there is often a substantial delay between commitment of an action and delivery of a reward, the choice of action has to be remembered until its outcome is revealed to causally relate them, which is often referred to as the temporal credit assignment problem (Sutton and Barto, 1998; Curtis and Lee, 2010), although in the present study the memory of previous action was no longer required after the reward stage. Finally, previous choice and previous outcome signals might be used to quickly capture the relationship between the history of animal’s choices and a reward (i.e., task rules). Our task did not require the animal to find such a pattern, but in the real world, a particular sequence of actions/rewards is more likely to yield a reward than others. The rat mPFC is well known to play a critical role in encoding task rules (Birrell and Brown, 2000; Jung et al., 2008; Kesner and Rogers, 2004; Ragozzino, 2007). In summary, the previous choice and previous outcome signals in the ACC and PLC/ILC might reflect other computational processes than computing values.

EXPERIMENTAL PROCEDURES

Behavioral task

Six young male Sprague-Dawley rats (approximately 9–11 weeks old, 330–350 g) were water-deprived (maintained > 80% of ad libitum body weight) and trained in a dynamic two-armed bandit task (Fig. 1A) as previously described (Kim et al., 2009) except that the duration of the delay stage was 2 instead of 3 s. Animal’s choice in each trial was rewarded stochastically with a probability that was constant within a block of trials, but changed across blocks. Each animal was tested for a total of 17 to 30 sessions, and each session consisted of four blocks of trials. The number of trials in each block was 35 plus a random number drawn from a geometric distribution with a mean of 5, with the maximum set at 45. The following four combinations of reward probabilities were used in each session: 0.72:0.12, 0.63:0.21, 0.21:0.63 and 0.12:0.72. The sequence was determined randomly with the constraint that the richer alternative always changed its location at the beginning of a new block. Details of the behavioral task including the maze control and behavioral stages are described in our previous study (Kim et al., 2009). The experimental protocol was approved by the Ethics Review Committee for Animal Experimentation of the Ajou University School of Medicine.
Neurophysiological recordings

Single-neuron activity was recorded with tetrodes from the dorsal ACC, PLC, ILC and lateral OFC. In three animals, all 12 tetrodes were implanted in the mPFC (2.7 mm anterior and 0.7 mm lateral to bregma). For the other three animals, six tetrodes were implanted in the mPFC and the other six were implanted in the lateral OFC (3.2 mm anterior and 3.3 mm lateral to bregma; Fig. 1D). After completion of a daily recording session, tetrodes were advanced by 35–75 µm (mPFC) or 30–40 µm (OFC). Units were recorded the next day without further moving the tetrodes. In all six animals, when the electrodes implanted in the mPFC were presumably near the border between the ACC and PLC judging from the history of electrode advancement, they were advanced by 0.6 mm before resuming unit recording in the PLC to facilitate the separation between ACC vs. PLC units. Similarly, the electrodes were advanced by 0.6 mm near the presumable border between the PLC and ILC to facilitate the separation between PLC vs. ILC units. Unit signals were amplified ×10,000, filtered between 0.6–6 KHz, digitized at 32 KHz and stored on a personal computer using a Cheetah data acquisition system (Bozemann, MT, USA). The animal’s head position was also monitored by tracking light-emitting diodes mounted on the headstage at 60 Hz. When recordings were completed, small marking lesions were made by passing an electrolytic current (50 µA, 30 s, cathodal) through one channel of each tetrode and recording locations were verified histologically as previously described (Baeg et al., 2001; Fig. 1D).

Reinforcement leaning model

Action values \([Q_{\text{action}}(t)]\) were computed in each trial according to the Rescola-Wagner rule (or Q-learning model; Rescola and Wagner, 1972; Sutton and Barto, 1998) as the following:

\[
\begin{align*}
\text{if } \text{action} &= \alpha(t) \quad RPE &= R(t) - Q_{\text{action}}(t) \\
Q_{\text{action}}(t+1) &= Q_{\text{action}}(t) + \alpha RPE \\
\text{else} \quad Q_{\text{action}}(t+1) &= Q_{\text{action}}(t)
\end{align*}
\]

where \(\alpha\) is the learning rate, \(R(t)\) represents the reward in trial \(t\) (1 if rewarded and 0 otherwise), and \(\alpha(t)\) indicates animal’s choice of action in trial \(t\) (left or right goal choice). Decision value \([\Delta Q(t)]\) was the difference between two action values \([Q_L(t) - Q_R(t)]\) and chosen value was the action value chosen in a given trial. Actions were chosen according the softmax action selection rule in which choice probability varied as a graded function of the decision value (Kim et al., 2009). Thus, the probability for selecting the left goal \([P_L(t)]\) was defined as:

\[
P_L(t) = \frac{1}{1 + \exp(-\beta \Delta Q)}
\]

where \(\beta\) is the inverse temperature that defines the degree of exploration in action selection. The parameters \(\alpha\) (0.364–0.742; 0.550±0.149) and \(\beta\) (3.12–4.29; 3.71±0.42, mean±SD) were estimated for the entire dataset from each animal using a maximum likelihood procedure (Seo et al., 2009).

Isolation and classification of neurons

Single units were isolated by examining various two-dimensional projections of spike waveform parameters as previously described (Baeg et al., 2003; Supplemental Figure S7). The identity of unit signals was determined based on the clustering pattern of spike waveform parameters, averaged spike waveforms, baseline discharge frequencies, auto-correlograms, and inter-spike interval histograms (Baeg et al., 2007). For those units that were recorded for two or more days, the session in which the units were most clearly isolated from background...
noise and other unit signals was used for analysis. Recorded neuronal signals in all areas were classified into broad-spiking neurons (putative pyramidal cells) and narrow-spiking neurons (putative interneurons; Feierstein et al., 2006; Supplemental Figure S7). The majority of the analyzed units were broad-spiking neurons (ACC: \( n = 655, 89.7\% \); PLC/ILC: \( n = 686, 91.3\% \); OFC: \( n = 1,073, 93.4\% \)). Although both types of neurons were included in the analyses, essentially the same results were obtained when narrow-spiking neurons were excluded from the analyses (data not shown).

**Multiple regression analysis**

Neural signals related to animal’s choices and their outcomes were estimated using the following regression model:

\[
S(t) = a_0 + a_1 C(t) + a_2 R(t) + a_3 X(t) + a_4 Q(t) + a_5 Q_c(t) + a_6 C(t-1) + a_7 R(t-1) + A(t) + \varepsilon(t),
\]

where \( S(t) \) indicates spike discharge rate, \( C(t) \), \( R(t) \) and \( X(t) \) represent animal’s choice (left or right), its outcome (0 or 1) and their interaction, respectively, in trial \( t \), \( \varepsilon(t) \) is the error term, and \( a_0 \sim a_9 \) are the regression coefficients. To estimate the latency of the neural signals related to animal’s upcoming choice for a given cortical area, we repeated this regression analysis using a 100-ms sliding window advancing in 50-ms steps. Then, the latency was determined as the time difference between the onset of the approach stage and the first time when the fraction of the neurons showing significant effects of animal’s upcoming choice became and remained significantly higher than chance level (binomial test, \( p < 0.05 \)) for a minimum of 250 ms (5 bins).

Neural signals related to values were examined using the following model:

\[
S(t) = a_0 + a_1 C(t) + a_2 R(t) + a_3 X(t) + a_4 \Delta Q(t) + a_5 Q_c(t) + a_6 C(t-1) + a_7 R(t-1) + A(t) + \varepsilon(t),
\]

where \( \Delta Q(t) \) and \( Q_c(t) \) denote decision value (the difference between left and right action values) and chosen value, respectively. \( A(t) \) stands for an autoregressive term that consisted of spike discharge rates in the previous three trials as the following:

\[
A(t) = a_8 S(t-1) + a_9 S(t-2) + a_{10} S(t-3),
\]

where \( a_8 \sim a_{10} \) are regression coefficients. It was included in all regression models to control for spike autocorrelation (Supplemental Table S4) that tends to inflate value-related neural signals (Supplemental Figure S6). Animal’s choice and its outcome in the previous trial were included in the model because they were substantially correlated with \( \Delta Q(t) \) and \( Q_c(t) \), respectively (Supplemental Figure S2).

The following models were used to determine whether neuronal activity is more correlated with RPE or updated chosen value \( [upQ_c(t)] \):

\[
S(t) = a_0 + a_1 C(t) + a_2 \Delta Q(t) + a_3 C(t-1) + a_4 R(t-1) + a_5 RPE + A(t) + \varepsilon(t)
\]

where \( upQ_c(t) = Q_c(t) + aRPE \).

\[
S(t) = a_0 + a_1 C(t) + a_2 \Delta Q(t) + a_3 C(t-1) + a_4 R(t-1) + a_5 QuPQ_c(t) + A(t) + \varepsilon(t)
\]
Statistical analysis

Significance of a regression coefficient was tested based on a t-test, and significance of the fraction of neurons for a given variable was tested with a binomial test. A significant difference in the fraction of neurons among different brain regions was tested with a χ²-test. P value < 0.05 was used as the criterion for a significant statistical difference unless noted otherwise. Data are expressed as mean±SEM unless noted otherwise.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a grant from Brain Research Center of the 21st Century Frontier Research Program (M.W.J.) and the National Institute of Health (D.L.).

REFERENCES


Groenewegen, HJ.; Berendse, HW. Anatomical relationships between the prefrontal cortex and the basal ganglia in the rat. In: Thierry, AM.; Goldman-Rakic, PS.; Christen, Y., editors. Motor and cognitive functions of the prefrontal cortex. Berlin and Heidelberg: Spring-Verlag; 1994. p. 51-77.


Vertes RP. Interactions among the medial prefrontal cortex, hippocampus and midline thalamus in emotional and cognitive processing in the rat. Neurosci 2006;142:1–20.

Figure 1.
Behavioral task, choice behavior and recording sites. (A) Behavioral task. Rats were allowed to choose freely between two goal locations (yellow discs) on a modified figure 8-shaped maze. Green arrows indicate photobeam detectors. Scale bar: 10 cm. (B) Animal’s choice behavior during one example recording session. The probability to choose the left goal (P_L) is plotted in moving average of 10 trials. Gray vertical lines indicate block transitions. Numbers at the top indicate mean reward probabilities associated with left and right goal choices. Tick marks denote trial-by-trial choices of the animal (top: left choice; bottom: right choice; long: rewarded trial; short: unrewarded trial). Gray line: actual choice of the animal. Black line: probability given by action values of the RL model. (C) The effect of past choice outcomes on the current goal choice in one animal. The influence of past choice outcomes (up to 10 trials) on the current choice was estimated with a logistic regression model. Positive regression coefficients indicate that positive choice outcomes for a given goal increased the probability of choosing the same goal subsequently. The error bars denote 95% confidence intervals. (D) Recording sites. The photomicrographs show coronal sections of the brain that contain marking lesions (yellow arrows). The damage on the dorsal cortex was produced in the process of removing the microdrive array at the end of recordings. Left, mPFC; Right, lateral OFC. Scale bar, 1 mm.
Figure 2.
Behavioral stages and animal’s locomotive behavior. (A) Behavioral stages. The behavioral task was divided into delay (D), go (G), approach to reward (A), reward consumption (Rw) and return (Rt) stages. Dotted lines denote transition points between stages. Onset of the delay stage marked the beginning of a trial (blue dotted line). The black arrows indicate alternative movement directions of the animals. Scale bar, 10 cm. (B) Animal’s movement trajectories from an example session. Starting from the reward stage, trials were divided into two groups depending on the upcoming goal choice of the animal (blue, left choice; red, right choice). Trials were decimated (3 to1) to enhance visibility. (C–D) Determination of the onset of the approach stage. The beginning of the approach stage was determined as the time when the animal’s movement trajectory began to diverge depending on the upcoming goal choice of the animal. The graphs show the time course of X-coordinates of animal’s position data near the onset of the approach stage during one recording session. Green dotted line (0 ms) corresponds to the time when the animal reached a particular vertical position [near “A” in (A)] determined by visual inspection to show clear separation in the animal’s X positions according to its choice, whereas the gray line (onset of the approach stage) corresponds to the time when the difference in the X positions of the left- and right-choice trials first became statistically significant ($t$-test, Neuron. Author manuscript; available in PMC 2011 May 13.
$p<0.05$). (C) X-coordinates of all trials. (D) Mean X-coordinates of the left-choice and right-choice trials.
Figure 3.
Neural activity related to animal’s choice and its outcome in the current and previous trials. 
(A) Fractions of neurons encoding animal’s choice (C), its outcome (R), or their interaction 
(X) in the current (t) and previous two trials (t-1 and t-2) were plotted in non-overlapping 0.5 s time windows across different behavioral stages (pre-Delay, last 1 s of the return stage; pre-Appr, last 1 s of the go stage; Appr, approach stage). The results from the go stage are shown twice aligning trials to the onset as well as the end of the go stage (pre-Appr). The vertical lines indicate the onset of a behavioral stage. Triangles indicate significant variation across regions ($\chi^2$-test, $p<0.05$). (B) The fraction of neurons that were significantly modulated by animal’s choice was plotted using a sliding window of 100 ms that was advanced in 50-ms time steps. In all plots, large circles indicate those fractions that are significantly higher than the significance level used (binomial test, $p<0.05$). The shading indicates the mean of the minimum

*Neuron*. Author manuscript; available in PMC 2011 May 13.
fractions significantly above chance which are slightly different across the ACC, PLC/ILC and OFC.
Figure 4.
Neural signals related to different types of value. (A) Fractions of neurons that significantly modulated their activity according to decision value ($\Delta Q$) or chosen value ($Q_c$). (B) The same regression analysis was performed with decision value replaced with left and right action values [$Q_L(t)$ and $Q_R(t)$, respectively]. The results for the other variables (eq. 2) were similar to those shown in Figure 3 and are not shown. The analysis time windows and symbols are as in Fig. 2.
Figure 5.
Convergence of neural signals related to animal’s choice, its outcome and chosen value. (A) Fractions of neurons modulating their activity according to the current choice \( C(t) \), green, its outcome \( R(t) \), blue or chosen value \( Qc(t) \), orange are shown in a 500 ms sliding window that was advanced in 100 ms steps for the 2-s time periods before and after the onset of the reward stage (time 0). The shading indicates the minimum fraction significantly above chance (binomial test, \( p=0.05 \)). (B) An example OFC neuron that modulated its activity according to animal’s choice, its outcome as well as chosen value in the current trial. Spike density functions (Gaussian kernel with \( \sigma = 100 \) ms) were constructed separately according to animal’s choice (left or right), its outcome (rewarded or unrewarded), or 4 different intervals of chosen value.
Figure 6.
Distribution of the standardized regression coefficients (SRC) for choice outcome and chosen value. Saturated colors indicate those OFC neurons that significantly modulated their activity according to both choice outcome and chosen value, and light colors indicate those that encoded either choice outcome or chosen value only. The remaining neurons are indicated in gray. Red and blue indicate those neurons in which activity was more correlated with RPE- or updated chosen value, respectively (eq 3 and 4).
Figure 7.
OFC population activity related to RPE or updated chosen value. Activity of OFC neurons encoding both choice outcome \([R(t)]\) and chosen value \([Q_c(t)]\) during the first 1 s of the reward stage were analyzed \((n=133)\). (A) The 133 OFC neurons were divided into 4 groups according to the signs of the regression coefficients for choice outcome and chosen value, and normalized spike density functions (divided by the peak value of each neuron’s spike density function averaged for all trials) were plotted according to 8 equally-divided ranges of RPE or updated chosen value \((upQc)\). The number of neurons in each category is shown above the corresponding plot. (B) Normalized activity averaged across the same groups of neurons during the first 1 s of the reward stage was plotted as a function of RPE or updated chosen value using the set of ranges as in (A). Activity of each neuron was normalized using the mean and SD for all trials. Error bars: SEM.
Figure 8.
Bidirectional encoding of RPE in the lateral OFC. The graph shows the relationship between the coefficients for RPE in rewarded (positive RPE) and unrewarded (negative RPE) trials that were estimated for the activity during the first 1 s of the reward stage. RPE (red circles), neurons encoding both chosen value and choice outcome with opposite signs; upQc (blue circles), neurons encoding both chosen value and choice outcome with same signs; Qc only (green circles), neurons encoding chosen value only; Others (open circles), the remaining neurons. The line was determined by a linear regression for RPE-coding neurons (red circles).