

Reward Stability Determines the Contribution of Orbitofrontal Cortex to Adaptive Behavior

Justin S. Riceberg and Matthew L. Shapiro

Fishberg Department of Neuroscience and Alfred B. and Gudrun J. Kastor Neurobiology of Aging Laboratories, Mount Sinai School of Medicine, New York, New York 10029-5674

Animals respond to changing contingencies to maximize reward. The orbitofrontal cortex (OFC) is important for flexible responding when established contingencies change, but the underlying cognitive mechanisms are debated. We tested rats with sham or OFC lesions in radial maze tasks that varied the frequency of contingency changes and measured both perseverative and non-perseverative errors. When contingencies were changed rarely, rats with sham lesions learned quickly and performed better than rats with OFC lesions. Rats with sham lesions made fewer non-perseverative errors, rarely entering non-rewarded arms, and more win-stay responses by returning to recently rewarded arms compared with rats with OFC lesions. When contingencies were changed rapidly, however, rats with sham lesions learned slower, made more non-perseverative errors and fewer lose-shift responses, and returned more often to non-rewarded arms than rats with OFC lesions. The results support the view that the OFC integrates reward history and suggest that the availability of outcome expectancy signals can either improve or impair adaptive responding depending on reward stability.

Introduction

Animals respond to changing contingencies by altering their behavior to maximize reward, tending to repeat responses that lead to reward and avoid non-rewarded or punished responses (Sutton and Barto, 1998). Contingencies can change at different and unpredictable rates, so that adaptive behavior requires a balance between persistence and flexibility. These different response tendencies are described ethologically as, e.g., “exploitation” and “exploration” (Cohen et al., 2007), and operationally as, e.g., “win-stay,” “win-shift,” and “lose-shift” strategies that are guided by reward history and expected outcomes. Rats learn readily to use these and other strategies with appropriate training (Packard et al., 1989).

The orbitofrontal cortex (OFC) is crucial for adapting optimally to changing contingencies. Reversal learning requires animals to respond to a previously unrewarded stimulus and stop responding to a stimulus that had been associated with reward. OFC damage impairs reversal learning in humans (Fellows and Farah, 2003), nonhuman primates (Dias et al., 1996), and rats (Schoenbaum et al., 2002), but the cognitive mechanisms supported by the OFC are debated. The impairment has been described as a consequence of disinhibition, an inability to inhibit previously rewarded responses (Mishkin, 1964), but OFC lesions also increase responses to irrelevant, not-recently-rewarded stimuli (Kim and Ragozzino, 2005; Walton et al., 2010). Two theories of OFC function propose different computational ac-

counts for both types of errors. A “flexible stimulus–reward mapping” hypothesis suggests that the OFC computes flexible stimulus–reward associations by adjusting responses to stimuli as rapidly as reward values change (Rolls, 2004). This view predicts that the OFC should be important for reversal learning independent of how rapidly reversals occur. An alternative “outcome expectancy” theory suggests that the OFC integrates the history of stimulus–reward associations and predicts expected reward outcome (Schoenbaum et al., 2009). This view predicts that the OFC should be especially important when well-established contingencies change unexpectedly but less important when contingencies change often and the integrated history of rewards is less informative.

The present experiment investigated the contribution of OFC to spatial reversal learning in a radial arm maze task that included several discriminative choices and varied the number of trials between reversals. The experiment was designed to test how the rate of contingency changes affected perseverative and non-perseverative responses during spatial reversal learning. If the OFC contributes to flexible mapping of places to rewards and increasing the frequency of contingency changes increases demand for such rapid remapping, then deficits produced by OFC lesions should covary with reversal frequency. If, however, the OFC contributes to integrating reward histories to compute outcome expectancies and increasing the frequency of contingency changes reduces the relevance of these computations to discriminative responses, then deficits produced by OFC lesions should vary inversely with reversal frequency.

Materials and Methods

Experimental design

To test the differential predictions of flexible reward mapping and outcome expectancy theories of OFC function, two experiments varied the frequency and order of contingency changes. In the first experiment, rats were given either neurotoxin or sham OFC lesions, acclimated to a radial maze, trained in an initial spatial discrimina-

Received Feb. 16, 2012; revised Sept. 13, 2012; accepted Sept. 18, 2012.

Author contributions: J.S.R. and M.L.S. designed research; J.S.R. performed research; J.S.R. and M.L.S. analyzed data; J.S.R. and M.L.S. wrote the paper.

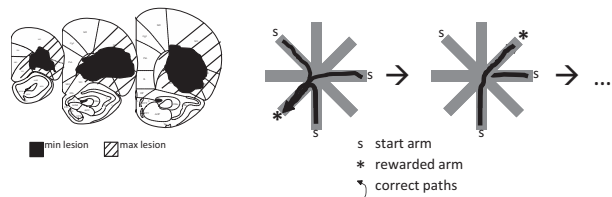
This work was supported by the Mount Sinai School of Medicine and the National Institutes of Health Grants MH065658 and MH073689 and Silvio O. Conte Award MH094263. We thank Maojuan Zhang for technical assistance and Erin Rich, Jake Young, Kevin Guise, and Mark Baxter for their comments on previous drafts of this manuscript.

Correspondence should be addressed to Dr. Matthew L. Shapiro, Mount Sinai School of Medicine, One Gustave Levy Place, Box 1065, New York, NY 10029. E-mail: matthew.shapiro@mssm.edu.

DOI:10.1523/JNEUROSCI.0776-12.2012

Copyright © 2012 the authors 0270-6474/12/3216402-08\$15.00/0

a) Maze acclimation → OFC/Sham Lesion → Initial training → Reversal Learning Tasks



b) Reversal Learning Tasks

Task Type	Trial/goal	Goals/session
Low Frequency Reversal - LFR	20+	3
Medium Frequency Reversal - MFR	10	5-6
High Frequency Reversal - HFR	3-7	9-13

Figure 1. Experimental design. *a*, All animals followed the same testing sequence. Coronal sections +4.2 mm, +3.7 mm, and +3.2 mm from bregma are reproduced (Paxinos and Watson, 1998), with markings indicating lesion size; black regions represent minimum lesion locations, and diagonal lines represent maximum lesion locations. All tasks took place on the eight-arm radial maze. Three arms on the radial maze served as start arms. Correct trajectories are shown as arrows to the rewarded arm (*). *b*, An outline of the three reversal tasks showing decreasing numbers of trials per goal before reversal from LFR to HFR.

tion, and tested in three tasks that varied in the number of trials given between reversals (Fig. 1). These tasks were intended to favor different types of response tendencies, or “strategies.” The “low-frequency reversal” (LFR) task was given after initial training and required each rat to perform at least 20 trials before contingencies were changed. The LFR task was designed to encourage the formation and use of stable outcome expectancies and to promote a win–stay strategy. A “medium-frequency reversal” (MFR) task gave each animal 10 trials between contingency changes. A “high-frequency reversal” (HFR) task gave each animal three to seven trials between contingency changes. The HFR task was designed to discourage the use of stable outcome expectancies and to favor a lose–shift strategy. The MFR task was designed as an intermediate. To determine whether the order of training (i.e., LFR, MFR, HFR) influenced the results in the first experiment, a second experiment evaluated a separate group of rats given either OFC or sham lesions and trained only in an HFR task during which contingencies changed after three consecutive correct trials.

Animals

Fifty-nine male Long–Evans rats weighing 275–350 g at the beginning of the experiment were housed individually in a colony room after a 12 h light/dark cycle. After acclimating to the colony room for at least 1 week, the rats were food restricted to no less than 85% of their *ad libitum* body weight and maintained on a restricted diet for the duration of the experiment. The rats were tested in seven cohorts given different subsets and sequences of the tasks, each with balanced numbers of rats given sham and OFC lesions. Three cohorts were tested in all three tasks in order (LFR, MFR, and then HFR); one cohort was trained only in the LFR, and three cohorts were trained first in HFR (see below, experiment 2). All procedures were performed in accordance with Institutional Animal Care and Use Committee guidelines and those established by the National Institutes of Health.

Apparatus

A radial maze elevated 45 cm above the floor had eight wooden arms (55 cm long, 11 cm wide) that met at 45° angles on an octagonal central platform (17.5 cm/side). A food well (4 cm diameter) at the end of each arm held reward on wire mesh. The arms were designated A–H. A waiting platform (30 cm wide, 35 cm high) stood next to the maze. The walls in the testing room displayed posters and paintings for peripheral stimuli.

Maze acclimation

Rats were handled for 3 d and acclimated to the testing environment by allowing them to forage for randomly scattered chocolate sprinkles on

the maze. Inaccessible food rewards were distributed around the maze beneath the wire mesh to minimize the use of odor cues to guide foraging.

Surgery

Rats were anesthetized with a continuous-flow isoflurane, mounted in a stereotaxic frame, and given a preoperative dose of flunixin–meoglumine (Banamine, 2 mg/kg) intramuscularly. The scalp was shaved, anesthetized (0.5 ml of lidocaine with epinephrine, s.c.), cleaned with betadine, and incised at the midline, and the skin and periosteum were retracted. Four holes were drilled from bregma in each hemisphere: 4 mm anterior, ±2.2 mm, and ±3.7 mm lateral; and 3 mm anterior, ±3.2 mm, and ±4.2 mm lateral. NMDA (20 μg/μl in a phosphate buffer vehicle; Sigma) or vehicle were infused at a rate of 0.1 μl/min using 30-gauge infusion cannulae connected by SILASTIC tubing to 1 μl Hamilton syringes (Hamilton) mounted on an infusion pump. At the anterior sites, cannulae were lowered 4.2 mm ventral to the skull surface, and infusions were given for 90 s. At the posterior sites, cannulae were lowered 5.2 mm ventral to the skull surface, and infusions were given for 60 s. After the infusions, cannulae were left in place for 5 min. The incision was closed with wound clips, and antibiotic ointment was applied. Buprenorphine (0.05 mg/kg, i.p.) was given for at least 3 d postoperatively. Rats were monitored during recovery for behavioral disturbances and signs of infection and allowed to recover for 5–10 d before beginning behavioral training.

Behavioral testing

Before training and during maze acclimation, the rats ate food in each of the eight arms. During training, three of the eight arms (A, D, and G) were used as start arms (Fig. 1), and chocolate sprinkles were placed in the food cup of one of the five remaining arms designated as potential goals. The order of testing each rat was pseudorandomized between days. Each trial began when a rat was placed in a predetermined pseudorandomly ordered start arm facing away from the center of the maze. No more than three consecutive trials used the same start arm. All eight arms were available as choices at all times, and entering one full body length into a goal arm defined a choice. If the rat chose the correct arm, it was allowed to consume the food and was placed on the waiting platform while food was replaced in the correct arm. If the rat chose the incorrect arm, it was returned to the waiting platform without reward unless self-correction was permitted, as described below. A block of trials was defined as a series of consecutive trials with the same goal arm. Each rat was tested in a block of trials, with an intertrial interval of 5–10 s. After the rat finished a block, the next block ensued. Errors were categorized operationally based on the correct response in the current and previous block. A choice was defined as a perseverative error if the rat chose the goal arm that had been rewarded during the previous block and a non-perseverative error if the rat chose any other unrewarded arm (see Fig. 4a, Table 1). A testing session included 3–13 blocks and thus 2–12 reversals. One session of the LFR and HFR tasks was given each day for 2 d; one session of the MFR task was given in 1 d only. During the LFR trials, each of the five designated goal arms was used as a goal for the first time. The arm that served as the goal arm during initial training also served as the first goal arm during both sessions of the LFR task, whereas the second and third goal arms of each session were novel. During MFR and HFR in experiment 1, all of the arms used as goals were repeated.

Experiment 1: LFR → MFR → HFR

Initial learning. Rats were initially trained in a spatial win–stay task. Training continued until the rat reached the criterion of eight consecutive correct trials [trials to criterion (TTC); see Fig. 2a], with a limit of 30

Table 1. Response categories after correct or error trials

Previous trial	Current choice	
	Repeat	Switch
Correct	Win–stay	Win–shift
Error	Lose–stay	Lose–shift
	Incorrect choice to the goal of the previous block?	
	Yes	No
Error type	Perseverative	Non-perseverative

trials on each training day. Rats were then tested for 24 trials per day until they performed >80% correct.

LFR task. In the LFR task, rats were trained to a criterion performance of eight consecutive correct choices and were then given an additional 12 trials to ensure a minimum of 20 trials per block. Reward was then moved to a different arm, and this reversal learning followed identical criterion and additional trials as the first block. Self-correction was permitted during the first four trials of each block. The relatively high number of consecutive trials per goal (20+) was intended to establish stable reward contingencies, favor a win–stay strategy, and encourage the use of history-driven reward expectancies.

MFR and HFR tasks. In the MFR and HFR tasks, the rats were given a fixed number of trials in each block, and no performance criterion was used. In MFR sessions, 10 trials were given in each block, and each session included five to six blocks. In HFR sessions, three to seven trials were given in each block, and each session included 9–13 blocks. Self-correction was permitted for the first two trials of each block during MFR and just the first trial of each block during HFR. The relatively low number of consecutive trials per goal was intended to discourage the use of stable reward contingencies, emphasize immediate contingencies, and encourage a lose–shift strategy.

Experiment 2: HFR only (HFR-initial)

A separate cohort of rats was given either sham ($n = 9$) or OFC ($n = 8$) lesions and trained from the outset in an HFR task. This experiment was designed to test the extent to which the order of training used in experiment 1 influenced the results of different reversal frequencies. In this HFR task, each block consisted of the number of trials required for each rat to reach a TTC of three consecutive correct. Once this criterion was reached, the reward location changed and a new block began. On day 1, rats were trained until four blocks were completed. On day 2, rats were trained until five blocks were completed. On day 3, rats were trained until six blocks were completed. Self-correction was permitted on all trials. In this experiment, each of the goal arms was used for the first time. The number of self-correction arm visits during error trials was assessed in all tasks and both experiments.

Data analyses

Performance was quantified by tabulating the probability of correct responses and the rate of perseverative and non-perseverative errors as defined operationally above. The statistics were calculated by dividing the total number of correct (or error) trials by the total number of trials. For the LFR task, the rate of acquisition was measured by both probability correct and TTC. For the HFR-initial task, the rate of acquisition was measured by TTC. To assess performance during reversals (see Fig. 2c), each block was aligned by the trial number after the contingency change. The mean probability correct for each trial number was computed for each rat, and these results were compiled to compute group performance means for each trial. The tendency to use a win–stay strategy was defined operationally as the probability of making a correct choice after a correct trial. Similarly, the tendency to use a lose–shift strategy was operationalized as the probability of making a correct choice after an error trial. ANOVA quantified the effects of treatment, task, error-type effects, and their interactions. *Post hoc* tests used Bonferroni's corrections. Pearson's r assessed the correlation between perseverative and non-perseverative error rates and the probability correct between the LFR and HFR tasks. OFC disruption typically produces a relatively transient effect on reversal

learning (Schoenbaum et al., 2002; Boulougouris et al., 2007; Rich and Shapiro, 2007; Young and Shapiro, 2009). Consistently, because the behavioral effects reported here were also transient, we report only the data collected during the first session for each task.

Histology

Rats were deeply anesthetized with isoflurane and pentobarbital (50 mg/kg, i.p.) and transcardially perfused with ice-cold PBS, followed by 10% Formalin solution. Brains were removed and postfixed in 10% Formalin for at least 24 h and then cryoprotected in 15% followed by 30% sucrose solution. Coronal sections (40 μ m) were stained with formol-thionin and compared with a standard brain atlas (Paxinos and Watson, 1998) to confirm lesion coordinates. Only animals with sufficient lesions to tissue dorsal to the rhinal fissure, mainly ventral orbital and lateral orbital cortices, as well as medial orbital frontal cortex, while preserving entirely prelimbic and infralimbic prefrontal cortices (PL/IL), were included (Fig. 1). Maximal lesions extended dorsally into the ventral frontal association cortex or the primary motor cortex and laterally into the agranular insular cortex, but performance measures were insensitive to lesion size.

Results

OFC lesions did not impair initial learning

Rats with OFC lesions typically learn new contingencies normally but are impaired when established contingencies are altered (Schoenbaum et al., 2002; Kim and Ragozzino, 2005). The present results confirmed these observations. Rats with OFC lesions learn to approach a single rewarded location as rapidly as sham-operated rats (controls) (Fig. 2a; TTC, sham vs lesion: $t_{(22)} = 0.64$, $p = 0.45$). All rats attained the >80% correct in 24 trials criterion within 2 d, with no differences between rats with sham and OFC lesions (days to 80% correct in 24 trials, sham vs lesion: $t_{(22)} = 0.37$, $p > 0.05$). Thus, the OFC was not required for rats to learn a spatial contingency.

Reversal frequency determined whether OFC lesions impaired or improved learning

In the LFR task, contingencies changed only after a performance criterion had been reached and rats approached a rewarded arm a minimum of 25 times. Control rats learned the modified contingencies more quickly than those with OFC lesions (probability correct sham vs lesion, treatment \times task interaction: $F_{(2,23)} = 17.27$, $p < 0.001$; Bonferroni's correction for sham vs lesion at LFR: $t = 4.43$, $p < 0.001$, Fig. 2b; and treatment: $F_{(1,18)} = 83.01$, $p < 0.001$, Figure 2c, left). Control rats performed significantly better than rats with OFC lesions beginning at the fifth trial after reversal (Bonferroni's correction for control vs lesion: $t = 3.28$, $p < 0.05$; Fig. 2c, left). Controls reach criterion faster than rats with OFC lesions [TTC (goal 2): control, 17.01 ± 1.1 vs lesion, 20.9 ± 1.4 , $t_{(19)} = 2.095$, $p < 0.05$; TTC (goal 3): control, 12.8 ± 0.63 vs lesion, 24.3 ± 2.47 , $t_{(19)} = 3.94$, $p < 0.001$]. In the MFR task, controls and rats with OFC lesions performed indistinguishably (Fig. 2b,c, middle). In the HFR task, the control rats learned more slowly than the rats with OFC lesions, beginning at the second trial after reversal (probability correct sham vs lesion, treatment \times task interaction: $F_{(2,23)} = 17.27$, $p < 0.001$; Bonferroni's correction for sham vs lesion at HFR: $t = 3.81$, $p < 0.001$, Fig. 2b; treatment: $F_{(1,5)} = 22.45$, $p < 0.001$, Fig. 2c; trial 2: Bonferroni's correction for sham vs lesion, $t = 2.82$, $p < 0.05$, Fig. 2c, right). Within the HFR task, the different learning rates between the groups did not vary with the length of the previous block (two-way ANOVA of treatment vs previous block length effect: $F_{(1,4)} = 1.122$, $p > 0.05$). Further, the number of self-correction arm visits per error trial did not distinguish the groups in any of

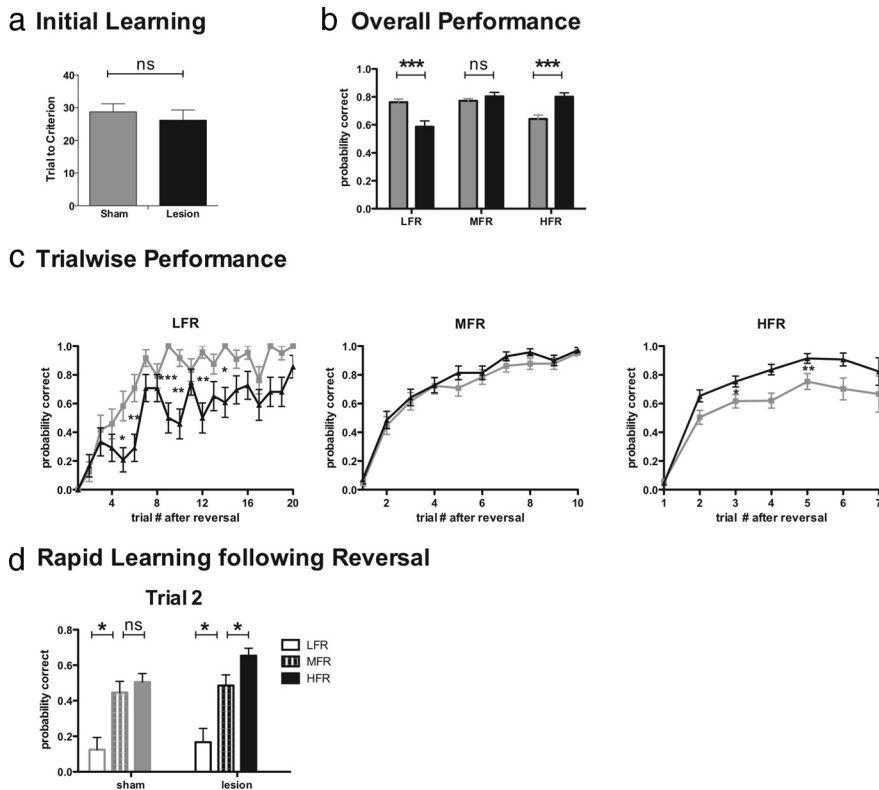


Figure 2. Reversal frequency determines the effect of OFC lesions on learning. *a*, Mean \pm SEM TTC was similar in sham (gray) and lesion (black) groups during initial learning. *b*, Overall performance (\pm SEM) was impaired by OFC lesions during LFR but improved during HFR. *c*, Sham rats performed significantly better than rats with OFC lesions during LFR. Learning was equivalent during MFR. Rats with OFC lesions performed significantly better than sham rats during HFR. *d*, Within-group analysis of performance on trial 2 during LFR (open bars), MFR (dashed bars), and HFR (filled bars) shows consistent improvement across tasks in rats with OFC lesions (black) but not sham rats (gray). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the tasks. These results show that OFC lesions affected reversal learning differently in accordance with specific task demands. When well-established contingencies changed, OFC lesions impaired reversal learning, but when contingencies changed rapidly, the same lesions facilitated reversal learning.

To assess acquisition rates across contingency changes within treatment, we compared performance on the second trial after reversal across tasks. Sham rats performed better in MFR compared with LFR (probability correct on trial 2 of LFR vs MFR: $t = 2.791$, $p < 0.05$; Fig. 2*d*, left) but equivalently in MFR and HFR (probability correct on trial 2 of MFR vs HFR: $t = 0.77$, $p > 0.05$, NS; Fig. 2*d*, left). Rats with OFC lesions performed better in LFR than MFR and improved again in HFR (probability correct on trial 2 of LFR vs MFR: $t = 2.805$, $p < 0.05$; probability correct on trial 2 of MFR vs HFR: $t = 2.359$, $p < 0.05$; Fig. 2*d*, right). Rats with OFC lesions learned faster as contingency changes occurred more frequently.

Reversal frequency and OFC lesions altered responses after correct or incorrect trials

OFC lesions can impair win–stay behavior (Berlin et al., 2004; Clarke et al., 2008; Rudebeck and Murray, 2008; Tsuchida et al., 2010), perhaps because they reduce the influence of positive feedback on behavioral choice. We investigated whether OFC lesions altered sensitivity to positive or negative feedback by evaluating performance on trials after correct or error trials. If the rats were equally sensitive to positive and negative feedback, then the probability of a correct choice should be influenced equally by preceding errors and correct choices (Table 1). If the animals were more

sensitive to positive than to negative feedback, then the probability of a correct choice should increase differentially after a correct trial; conversely, if the rats were more sensitive to negative than to positive feedback, then the probability of a correct choice should increase after an error. Controls performed significantly better than rats with OFC lesions after correct trials during LFR (treatment \times task interaction effect: $F_{(2,63)} = 9.30$, $p < 0.001$; Bonferroni's correction for sham vs lesion at LFR: $p < 0.01$; Fig. 3*a*) but not during MFR or HFR. Conversely, rats with OFC lesions performed significantly better than controls after error trials during HFR (treatment \times task interaction: $F_{(2,63)} = 6.17$, $p < 0.01$; Bonferroni's correction for sham vs lesion at HFR: $p < 0.01$; Fig. 3*b*) but not in LFR or MFR. The probability of a lose–shift response was primarily consistent across trials within a block in each group [mean \pm SEM probability correct after an error trial: sham, 0.43 ± 0.05 ; OFC, 0.66 ± 0.02 ; ANOVA (trials): $F_{(4,1)} = 0.54$, $p = 0.71$].

The results suggest that the influence of positive and negative feedback depends on reward history and that the OFC helps to maintain recently rewarded responses. When an environment is associated with stable expected outcomes, the OFC may facilitate flexible responding by promoting win–stay behavior. In environments with unstable expected outcomes, OFC may limit flexible responding by interfering with lose–shift behavior.

OFC lesions increased non-perseverative errors during LFR and reduced such errors during HFR

The OFC has been proposed to contribute to adaptive behavior by inhibiting previously rewarded responses, so that OFC lesions increase perseverative errors (Jones and Mishkin, 1972). More recent work in both rats and monkeys with OFC lesions emphasizes its effects on associating precise reward value across many responses, manifested behaviorally as increased exploratory or “shifting” behavior during changing contingencies (Kim and Ragozzino, 2005; Walton et al., 2010). We therefore analyzed the relative number of perseverative and non-perseverative errors during reversal learning (Fig. 4*a*, Table 1). As described in Materials and Methods, perseverative errors were defined operationally by a reentry into the arm rewarded in the previous block, whereas non-perseverative errors were defined as entering any other non-rewarded arm. During the standard reversal learning LFR task, controls made fewer perseverative errors ($t_{(22)} = 2.60$, $p < 0.05$) and far fewer non-perseverative errors than rats with OFC lesions ($t_{(22)} = 9.50$, $p < 0.001$; Fig. 4*b*). In contrast, when contingencies changed rapidly in the HFR task, control rats made significantly more non-perseverative errors than rats with OFC lesions ($t_{(20)} = 3.56$, $p < 0.01$; Fig. 4*b*), and the number of perseverative errors was equivalent between groups. Moreover, although the perseverative error rate (errors/total trials) was identical between groups in each task (data not shown), the number of non-perseverative errors accounted for the most variance in performance

between the groups (task × treatment interaction: perseverative errors, $F_{(2,63)} = 3.985, p = 0.0235$; non-perseverative errors, $F_{(2,63)} = 33.72, p < 0.001$). Although the self-correction procedure during the LFR may have influenced the preponderance of non-perseverative, as opposed to perseverative, errors compared with previous experiments that forbid self-correction (Kim and Ragozzino, 2005), the procedure was equally available to rats with sham or OFC lesions and thus cannot account for performance differences across treatment. The results suggest that the OFC bidirectionally alters behavioral flexibility by primarily influencing non-perseverative, not perseverative, responses in contexts with stable and unstable outcome expectancies.

Error types correlated with performances across tasks

The results so far show that the frequency of contingency changes in the radial maze determined the effects of OFC lesions on reversal learning and suggest that the same neuropsychological mechanisms might guide responding across tasks. For example, if OFC lesions increased the probability of adopting a lose–shift, rather than a win–stay, strategy, then individual rats’ non-perseverative errors in the LFR should predict better performance in the HFR.

To test this possibility, we analyzed how different types of errors in the LFR task predicted performance in the HFR task using Pearson’s *r*. Two complementary patterns emerged. First, rapid learning in the LFR predicted perseverative, but not non-perseverative, errors in the HFR (perseverative: $r = 0.75, p = 0.0075$; non-perseverative: $r = 0.09, p > 0.05$; Fig. 5*b*). Controls performed well in the LFR and were more likely than rats with OFC lesions to return to the arm rewarded in the previous trial. Thus, controls were more sensitive to positive than negative feedback, consistent with a win–stay strategy (Fig. 3*a*). In contrast, the control rats performed relatively poorly in the HFR, were less likely to choose the correct response after an error trial, and showed a reduced tendency to lose–shift (Fig. 3*b*). The controls’ adaptive tendency to win–stay in the LFR task predicted their tendency to perseverate in the HFR task. Second, rapid learning in the HFR task retrodicted non-perseverative, but not perseverative, errors in the LFR task (non-perseverative: $r = 0.66, p = 0.025$; perseverative: $r = 0.37, p > 0.05$; Fig. 5*a*). Rats with OFC lesions performed well in the HFR and were less likely than control rats to return to a non-rewarded arm from the previous trial, thereby showing an increased sensitivity to negative feedback and increased tendency to lose–shift (Fig. 3*b*). Rats that were more likely to make non-perseverative responses during LFR were more likely to perform well during HFR, in which a lose–shift strategy was adaptive.

OFC lesions improved performance when initial training consisted of rapid contingency changes

The results described so far suggest that the effects of OFC lesions vary with reversal frequency. However, because the different reversal frequencies were given in order, from slowest to fastest, the

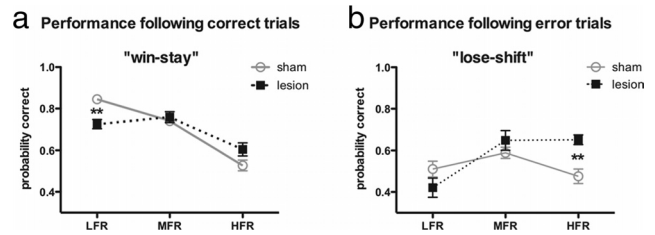
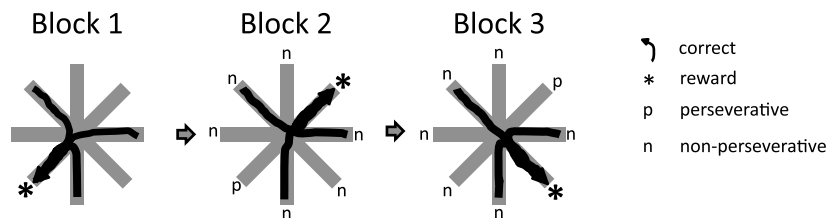


Figure 3. Reversal frequency alters the effects of OFC lesions on performance after correct or error trials. *a*, The sham rats (gray circles) performed better than the rats with OFC lesions (black squares) on the trial after a correct trial during LFR (** $p < 0.01$). *b*, Rats with OFC lesions (black squares) performed better than sham rats (gray circles) on the trial after errors during HFR (** $p < 0.01$).

a Categorizing Perseverative or Non-Perseverative Errors



b Non-perseverative Errors Account for Performance Differences

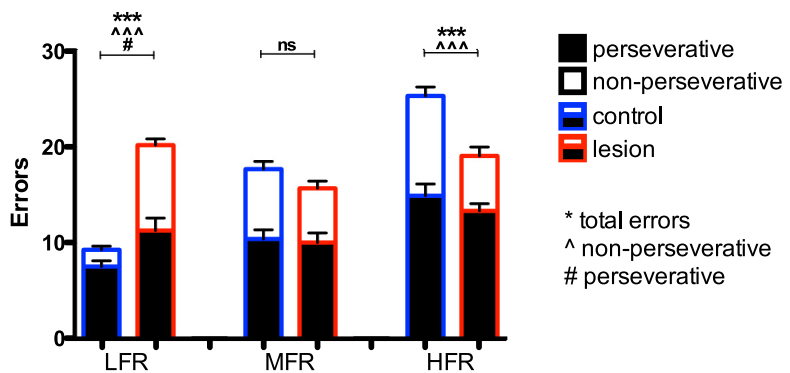
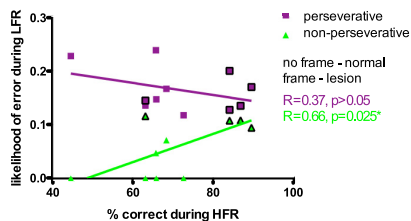


Figure 4. Non-perseverative, not perseverative, errors account for performance differences between groups. *a*, Entering the previously rewarded arm defined a perseverative error. Entering any other non-rewarded arm defined a non-perseverative error. *b*, During LFR, sham rats (blue frame) made fewer total (** $p < 0.001$), non-perseverative (open bars, *** $p < 0.001$), and perseverative (filled bars, * $p < 0.05$) errors than rats with OFC lesions (red frame). During HFR, sham rats made more total and non-perseverative errors (** $p < 0.001$ for both) than rats with OFC lesions.

results could also be influenced by training sequence. For example, the rats given sham lesions may have adopted a win–stay strategy because they were trained first on LFR, and their poor performance relative to rats with OFC lesions during subsequent HFR could reflect this previous training. From this view, the better performance by rats with OFC compared with sham lesions during the HFR task does not reveal enhanced lose–shift learning as much as diminished win–stay learning during LFR training. If such is the case, then intact rats should learn faster than rats with OFC lesions when both groups are trained initially in an HFR task. Alternatively, if OFC lesions facilitate learning when contingencies change rapidly, then rats with OFC lesions should learn HFR faster than intact animals independent of training history.

To distinguish these possibilities, we trained a separate cohort of rats on a modified version of the HFR (HFR-initial) task with-

a LFR non-persistent errors predict HFR performance



b LFR performance predicts HFR perseverative errors

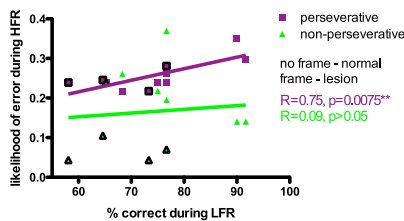
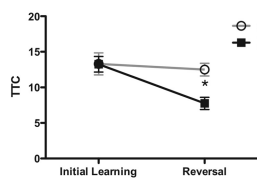


Figure 5. Error types differentially correlate with performance across tasks. **a**, Likelihood of non-persistent errors during LFR predicted better performance during HFR (green dots). Likelihood of perseverative errors during LFR did not predict HFR performance (purple dots). **b**, Better performance during LFR predicted the likelihood of making perseverative errors (purple dots), but not non-persistent errors (green dots), during HFR. Dots with frames are individual rats with OFC lesions, and dots without frames are individual rats with sham lesions.

a HFR-initial TTC Performance



b Performance following error or correct trials

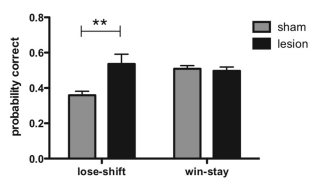


Figure 6. OFC lesions improve performance when initial training is in an HFR task. **a**, During initial learning, rats with sham (gray) and OFC (black) lesion performed indistinguishably, but during the reversal learning, rats with OFC lesions performed better than controls (* $p < 0.05$). **b**, Rats with OFC lesions performed better than sham rats on trials after errors (** $p < 0.01$, left) but equally on trials after correct trials (right).

out previous training in either LFR or MFR. All rats learned to approach the initial rewarded arm at similar rates (control vs lesion during initial learning: $t_{(16)} = 0.026, p = 0.98$; Fig. 6a, left). When contingencies were modified, however, rats with OFC lesions learned more quickly than those with sham lesions (sham vs OFC: treatment \times test interaction, $F_{(1,16)} = 4.2, p < 0.05$; Bonferroni's correction for reversal, $p < 0.05$; Fig. 6a, right). The rats with OFC lesions continued to perform better than controls during the second day (control vs lesion: treatment, $F_{(1,16)} = 7.45, p < 0.01$), but performance was similar in both groups by the third day (control vs lesion: treatment, $F_{(1,16)} = 1.498, p = 0.22$). As in the first experiment, rats with OFC lesions performed better than sham controls on trials after incorrect responses (probability correct sham vs lesion: $t_{(16)} = 3.24, p < 0.01$; Fig. 6b, left) but not on trials after correct responses (probability correct sham vs lesion: $t_{(16)} = 0.43, p = 0.67$; Fig. 6b, right). OFC lesions improved learning when contingencies changed rapidly independent of training sequence.

Discussion

The OFC is crucial for adapting normally to changing contingencies, especially in reversal learning, but its precise contribution to cognition is unclear. Some experiments report that OFC lesions increase perseveration by disinhibiting responses to previously rewarded stimuli, whereas others report increased exploration, disinhibiting approaches to stimuli that have not been associated previously with reward. One theory that attempts to explain these paradoxical findings proposes that the OFC integrates reward history to generate outcome expectancies associated with stimuli or responses (Schoenbaum et al., 2009). The present experiment tested this theory by

changing the frequency with which contingencies changed. As reported previously, OFC lesions impaired reversal learning when stable contingencies were well established before they were changed (LFR). In LFR, control animals tended to follow a win-stay strategy, returning to previously rewarded arms more often than rats with OFC lesions. In contrast, OFC lesions improved reversal learning when contingencies changed rapidly (HFR). In HFR, rats with OFC lesions tended to follow a lose-shift strategy by avoiding previously non-rewarded arms. OFC lesions had no effect on performance when contingencies were changed at an intermediate rate (MFR). Although OFC lesions slightly increased perseverative errors

in LFR, non-persistent errors accounted for most of the behavioral differences between groups. Rats with sham and OFC lesions tended to follow distinct win-stay and lose-shift strategies, so that non-persistent errors in the LFR predicted rapid learning in the HFR, whereas perseverative errors in the HFR retrodicted rapid learning in the LFR. Together, the results show that OFC lesions alter sensitivity to positive and negative feedback and are consistent with the claim that the OFC computes reward expectancies based on integrated reward history. The availability of these computations may influence response strategies and either improve or impair learning depending on the stability of contingencies.

Reward history signaling and multiple memory systems

The integrated reward history theory of OFC proposes that OFC signals the reward expectancies of stimuli by integrating their history of reward and computing reward likelihood (Schoenbaum et al., 2009; Schoenbaum and Esber, 2010). When contingencies change, expected outcome signals are contrasted with actual outcome signals, and the difference drives learning (Sutton and Barto, 1998). During reversal learning of well-established contingencies, the integrated reward expectancy coded by OFC should support large error signals and facilitate the encoding of new associations. Thus, the present LFR task established strong and stable reward associations before contingencies were reversed, so that large-magnitude OFC expectancy signals could help normal animals learn quickly (Schoenbaum et al., 2002; Kim and Ragozzino, 2005; Young and Shapiro, 2009). Without those signals, rats with OFC lesions learned the LFR task more slowly. The same computations account for the relatively slow learning by normal animals in the HFR task. In this case, the history of reward includes many contingency changes, so that the magnitude of the expectancy and its corresponding error signal should be relatively small. Furthermore, because outcome expectancy derives from the history of reward integrated over time, OFC signals and their contribution to behavior would lag responses based on immediate contingency changes. In the HFR task, expectancy signals based on integrated reward history could interfere with performance guided by other brain structures that form more rapid associations with unconditioned stimuli. The pattern of results is analogous to the effects of PL/IL disruption on repeated strategy switches. During the first few switches, memory for the most recently useful strategy guided behavior, and PL/IL activity was required for successful performance. After many switches, however, rats switched normally between the highly

familiar tasks even when PL/IL was dysfunctional. The result suggested that rats learned to use immediate contingencies, rather than memory, to guide responses, a tactic that did not require the PL/IL (Rich and Shapiro, 2007). These and the present results also suggest that rats with OFC lesions undergoing repeated LFR training may eventually outperform controls, as shown in similar reversal learning tasks (Schoenbaum et al., 2002; Boulougouris et al., 2007). In these experiments, rats adapted to changing contingencies by modifying behavioral tactics and did so using brain systems that did not require prefrontal activity. The present results also implicate OFC in the overtraining reversal effect, in which animals learn reversals more efficiently if the initial discrimination is overtrained (Reid, 1953; Capaldi, 1963). Encoding of expected outcomes in OFC, and thus a higher-magnitude error signal, in overtrained animals may contribute to both effects.

The reward expectancy view explains both the facilitated (LFR) and impaired (HFR) learning by normal animals compared with those with OFC lesions. With the OFC and its history-based outcome expectancy signals, actions “gain momentum” so that behavior is sustained despite occasional reward lapses. Without the OFC, brain structures that respond more rapidly to instantaneous contingencies may gain more influence on behavior (Pasupathy and Miller, 2005). An analogous situation may occur when lesions of the amygdala rescue reversal learning deficits in rats with OFC lesions (Baxter and Browning, 2007; Stalnaker et al., 2007). Without OFC activity, basolateral amygdala neurons respond more slowly to contingency changes (Saddoris et al., 2005), and this inflexibility is associated with slower reversal learning. Lesions to basolateral amygdala neurons may remove inaccurate signals and thereby improve reversal performance supported by common downstream structures, such as the dorsal striatum (Stalnaker et al., 2007). Similarly, ventromedial PFC lesions that facilitate reversal learning have been interpreted as “removing a brake on learning by subcortical areas” (Graybeal et al., 2011).

Perseverative versus non-perseverative errors and response tendencies across tasks

The present findings revealed increases in both perseverative and non-perseverative “exploratory” errors after OFC lesions but suggest that OFC lesions altered the patterns of errors that occurred depending on the frequency of contingency changes. The results suggest that, beyond outcome expectancy, OFC lesions alter responses to rewarding and non-rewarding events in patterns that cannot be explained in terms of perseveration. Monkeys with orbitofrontal damage that fail to respond optimally to changing contingencies exhibit an increase in switching behavior, not perseveration (Walton et al., 2010). Here, compared with controls, rats with OFC lesions made more non-perseverative errors when contingencies changed slowly but fewer non-perseverative errors when contingencies changed rapidly (Fig. 4). Controls were more likely to reenter a recently chosen rewarded arm than rats with OFC lesions in the LFR task but were less likely to avoid a recently entered, non-rewarded arm in the HFR task. OFC lesions thus altered the relative influence of reward and non-reward on subsequent behavior and may have shifted response strategies toward lose–shift. This interpretation is consistent with the correlation between the different error types across tasks. Good performance in the LFR task correlated with perseverative errors in the HFR task, whereas good performance in the HFR task correlated with non-perseverative errors in the LFR task. In other words, normal rats followed a win–stay strategy that improved performance when contingencies were relatively

stable but impaired performance when contingencies changed quickly. In contrast, rats with OFC lesions followed a lose–shift strategy that impaired performance when contingencies were stable but improved performance when contingencies changed rapidly. The rats with OFC lesions outperformed sham rats during the HFR not only because controls performed worse but also because rats with OFC lesions performed better, as early as the second trial after contingencies changed (Fig. 2*d*).

OFC and spatial encoding

Across mammalian species, OFC dysfunction impairs learning about non-reward (Tait and Brown, 2007; Burke et al., 2009), credit assignment (Walton et al., 2010), win–stay behavior (Tsuchida et al., 2010) and increases responses to relevant (Noonan et al., 2010) and irrelevant (Kim and Ragozzino, 2005) stimulus options. The role of the OFC in selecting among multiple discriminative options has been evaluated in tasks in which animals learn to flexibly discriminate and associate discrete stimuli, such as odors (Kim and Ragozzino, 2005) or images (Noonan et al., 2010), with reward. Indeed, most models of OFC function do not describe its potential role in flexibly linking stimulus–stimulus associations (McDonald and White, 1993), such as places, with reward. Orbitofrontal neurons exhibit spatial correlates, but these reflect distinct spatially directed responses (Feierstein et al., 2006) or associations between a place and a reward (Lipton et al., 1999), consistent with a role in signaling expected outcomes. OFC neurons fire throughout one or more entire maze arm, with fivefold less spatial resolution than hippocampal place fields in the same maze (Young and Shapiro, 2011*b*). Primate OFC neurons rarely encode spatial aspects of tasks, and this discrepancy could reflect differences in tasks, anatomy (e.g., rodents lack the rostral cortex of primates), or spatial processing among species (Zald, 2006). In contrast to rodent experiments, for example, few recording tasks in nonhuman primates allow the animals to move through space. Thus, the extent to which a “location” can be considered an integrated stimulus and thereby offer a suitable experimental setting to test Rolls’ flexible stimulus–reward mapping theory may depend on task design, species, and the subregion of OFC under investigation (Butter et al., 1969; Meunier et al., 1997; Boulougouris et al., 2007; Kennerley and Wallis, 2009; Mar et al., 2011). Nonetheless, the present results extend previous findings (Young and Shapiro, 2009, 2011*b*) that suggest that OFC has a similar role in guiding appropriate action selection even when discriminative stimuli are kept constant, and only the spatial location of rewards change. This result is consistent with a role for OFC in representing extended action sequences and plans (Tsujimoto et al., 2011) and in reward-expectancy guided memory retrieval (Young and Shapiro, 2011*a*).

Conclusions

OFC lesions can improve or impair spatial reversal learning depending on the stability of rewarded associations. Future studies will assess interactions between the OFC and other structures required for these tasks, including the hippocampus and neostriatum, by recording multisite neural activity simultaneously while the frequency of contingency changes is varied.

References

- Baxter MG, Browning PG (2007) Two wrongs make a right: deficits in reversal learning after orbitofrontal damage are improved by amygdala ablation. *Neuron* 54:1–3.

- Berlin HA, Rolls ET, Kischka U (2004) Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. *Brain* 127:1108–1126.
- Boulougouris V, Dalley JW, Robbins TW (2007) Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat. *Behav Brain Res* 179:219–228.
- Burke KA, Takahashi YK, Correll J, Brown PL, Schoenbaum G (2009) Orbitofrontal inactivation impairs reversal of Pavlovian learning by interfering with “disinhibition” of responding for previously unrewarded cues. *Eur J Neurosci* 30:1941–1946.
- Butter CM, McDonald JA, Snyder DR (1969) Orality, preference behavior, and reinforcement value of nonfood object in monkeys with orbital frontal lesions. *Science* 164:1306–1307.
- Capaldi EJ (1963) Overlearning reversal effect in a spatial discrimination task. *Percept Mot Skills* 16:335–336.
- Clarke HF, Robbins TW, Roberts AC (2008) Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *J Neurosci* 28:10972–10982.
- Cohen JD, McClure SM, Yu AJ (2007) Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos Trans R Soc Lond B Biol Sci* 362:933–942.
- Dias R, Robbins TW, Roberts AC (1996) Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* 380:69–72.
- Feierstein CE, Quirk MC, Uchida N, Sosulski DL, Mainen ZF (2006) Representation of spatial goals in rat orbitofrontal cortex. *Neuron* 51:495–507.
- Fellows LK, Farah MJ (2003) Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain* 126:1830–1837.
- Graybeal C, Feyder M, Schulman E, Saksida LM, Bussey TJ, Brigman JL, Holmes A (2011) Paradoxical reversal learning enhancement by stress or prefrontal cortical damage: rescue with BDNF. *Nat Neurosci* 14:1507–1509.
- Jones B, Mishkin M (1972) Limbic lesions and the problem of stimulus–reinforcement associations. *Exp Neurol* 36:362–377.
- Kennerley SW, Wallis JD (2009) Encoding of reward and space during a working memory task in the orbitofrontal cortex and anterior cingulate sulcus. *J Neurophysiol* 102:3352–3364.
- Kim J, Ragozzino ME (2005) The involvement of the orbitofrontal cortex in learning under changing task contingencies. *Neurobiol Learn Mem* 83:125–133.
- Lipton PA, Alvarez P, Eichenbaum H (1999) Crossmodal associative memory representations in rodent orbitofrontal cortex. *Neuron* 22:349–359.
- Mar AC, Walker AL, Theobald DE, Eagle DM, Robbins TW (2011) Dissociable effects of lesions to orbitofrontal cortex subregions on impulsive choice in the rat. *J Neurosci* 31:6398–6404.
- McDonald RJ, White NM (1993) A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. *Behav Neurosci* 107:3–22.
- Meunier M, Bachevalier J, Mishkin M (1997) Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys. *Neuropsychologia* 35:999–1015.
- Mishkin M (1964) Perseveration of central sets after frontal lesions in monkeys. In: *The frontal granular cortex and behavior* (Warren JM, Akert K, eds), pp 219–241. New York: McGraw-Hill.
- Noonan MP, Walton ME, Behrens TE, Sallet J, Buckley MJ, Rushworth MF (2010) Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc Natl Acad Sci U S A* 107:20547–20552.
- Packard MG, Hirsh R, White NM (1989) Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. *J Neurosci* 9:1465–1472.
- Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873–876.
- Paxinos G, Watson C (1998) *The rat brain in stereotaxic coordinates*. San Diego: Academic.
- Reid LS (1953) The development of noncontinuity behavior through continuity learning. *J Exp Psychol* 46:107–112.
- Rich EL, Shapiro ML (2007) Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. *J Neurosci* 27:4747–4755.
- Rolls ET (2000) The orbitofrontal cortex and reward. *Cereb Cortex* 10:284–294.
- Rolls ET (2004) The functions of the orbitofrontal cortex. *Brain Cogn* 55:11–29.
- Rudebeck PH, Murray EA (2008) Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *J Neurosci* 28:8338–8343.
- Saddoris MP, Gallagher M, Schoenbaum G (2005) Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. *Neuron* 46:321–331.
- Schoenbaum G, Esber GR (2010) How do you (estimate you will) like them apples? Integration as a defining trait of orbitofrontal function. *Curr Opin Neurobiol* 20:205–211.
- Schoenbaum G, Nugent SL, Saddoris MP, Setlow B (2002) Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *Neuroreport* 13:885–890.
- Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK (2009) A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nat Rev Neurosci* 10:885–892.
- Stalnaker TA, Franz TM, Singh T, Schoenbaum G (2007) Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron* 54:51–58.
- Sutton RS, Barto AG (1998) *Reinforcement learning: an introduction*. Cambridge, MA: Massachusetts Institute of Technology.
- Tait DS, Brown VJ (2007) Difficulty overcoming learned non-reward during reversal learning in rats with ibotenic acid lesions of orbital prefrontal cortex. *Ann N Y Acad Sci* 1121:407–420.
- Tsuhida A, Doll BB, Fellows LK (2010) Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *J Neurosci* 30:16868–16875.
- Tsujimoto S, Genovesio A, Wise SP (2011) Comparison of strategy signals in the dorsolateral and orbital prefrontal cortex. *J Neurosci* 31:4583–4592.
- Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* 65:927–939.
- Young JJ, Shapiro ML (2009) Double dissociation and hierarchical organization of strategy switches and reversals in the rat PFC. *Behav Neurosci* 123:1028–1035.
- Young JJ, Shapiro ML (2011a) Orbitofrontal cortex and response selection. *Ann N Y Acad Sci* 1239:25–32.
- Young JJ, Shapiro ML (2011b) Dynamic coding of goal-directed paths by orbital prefrontal cortex. *J Neurosci* 31:5989–6000.
- Zald DH (2006) The rodent orbitofrontal cortex gets time and direction. *Neuron* 51:395–397.