Article

Neuron

Medial Orbitofrontal Cortex Mediates Outcome Retrieval in Partially Observable Task Situations

Highlights

- Rats with an intact mOFC can retrieve unobservable action outcomes to guide choice
- Damage to mOFC causes deficits in choice when task variables are partially observable
- mOFC damage does not cause deficits in choice when task variables are observable
- The mOFC retrieves outcome-specific information to guide goal-directed action

Authors

Laura A. Bradfield, Amir Dezfouli, Mieke van Holstein, Billy Chieng, Bernard W. Balleine

Correspondence

bernard.balleine@sydney.edu.au

In Brief

Choice between actions often requires the ability to retrieve action consequences in circumstances where they are only partially observable. Here, Bradfield et al. demonstrate that this critical determinant of decision-making depends specifically on the medial orbitofrontal cortex in rats.

Bradfield et al., 2015, Neuron *88*, 1–13
December 16, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.neuron.2015.10.044

CellPress

Please cite this article in press as: Bradfield et al., Medial Orbitofrontal Cortex Mediates Outcome Retrieval in Partially Observable Task Situations, Neuron (2015), http://dx.doi.org/10.1016/j.neuron.2015.10.044

Neuron
Article

Medial Orbitofrontal Cortex Mediates Outcome Retrieval in Partially Observable Task Situations

Laura A. Bradfield,[1] Amir Dezfouli,[1] Mieke van Holstein,[1,2] Billy Chieng,[1] and Bernard W. Balleine[1,*]
[1]Brain & Mind Centre, University of Sydney, 94 Mallett Street, Camperdown, NSW 2050, Australia
[2]Donders Institute for Brain Cognition and Behaviour, Radboud University, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands
*Correspondence: bernard.balleine@sydney.edu.au
http://dx.doi.org/10.1016/j.neuron.2015.10.044

SUMMARY

Choice between actions often requires the ability to retrieve action consequences in circumstances where they are only partially observable. This capacity has recently been argued to depend on orbitofrontal cortex; however, no direct evidence for this hypothesis has been reported. Here, we examined whether activity in the medial orbitofrontal cortex (mOFC) underlies this critical determinant of decision-making in rats. First, we simulated predictions from this hypothesis for various tests of goal-directed action by removing the assumption that rats could retrieve partially observable outcomes and then tested those predictions experimentally using manipulations of the mOFC. The results closely followed predictions; consistent deficits only emerged when action consequences had to be retrieved. Finally, we put action selection based on observable and unobservable outcomes into conflict and found that whereas intact rats selected actions based on the value of retrieved outcomes, mOFC rats relied solely on the value of observable outcomes.

INTRODUCTION

To choose appropriately between competing courses of action, an agent must be able to assign values to actions based on their consequences, whether those consequences are present in the immediate environment or not. As is becoming better recognized, however, assigning values to actions in the absence of their specific outcomes requires the ability to retrieve variables from memory that were observed, but that are currently only partially observable (Daw et al., 2005; Noonan et al., 2012; Wilson et al., 2014; Stalnaker et al., 2015). For example, in rodents, commonly used tasks such as outcome-specific devaluation (Balleine and Dickinson, 1998) or Pavlovian-instrumental transfer (Colwill and Rescorla, 1990; Corbit et al., 2001) require animals to choose between available actions based on the retrieval of a specific outcome, whether that retrieval is driven by the action itself or the presence of specific predictive stimuli.

A number of recent studies have connected this ability to retrieve or to infer partially observable task states with the function of the orbitofrontal cortex (OFC). In humans, multiple authors (Bechara et al., 1994; Schnider et al., 2005; Schnider, 2013) have described patients with brain damage encompassing the medial OFC (mOFC) as being unable to "think through" the consequences of their actions, relying instead on ingrained habits or immediate information to guide their actions. Similarly, individuals with anti-social personality disorder, which is characterized by an inability to foresee action outcomes, have been shown to display low levels of OFC activation in fMRI studies (see Decety et al., 2014; Liu et al., 2014). Other studies in both humans and other animals have found that damage to the mOFC creates deficits in probability and risk assessment (see Mar et al., 2011; Clark et al., 2008). Conversely, hyperactivation of mOFC has been reported in people suffering obsessive-compulsive disorder (OCD) (Lagemann et al., 2012; Saxena et al., 1998), characterized by compulsions driven by obsessively imagining outcomes that are not present or unobservable (e.g., germs). Finally, damage to areas homologous to mOFC in primates has been reported to induce deficits specifically in instrumental actions and not Pavlovian conditioned responses (Rudebeck and Murray, 2011), and Bouret and Richmond (2010) demonstrated that neuronal activity was greatest in the mOFC during actions that were self-initiated compared to those elicited by a cue. Interestingly, this activity was reduced by sating the animal on the outcome associated with the self-initiated action.

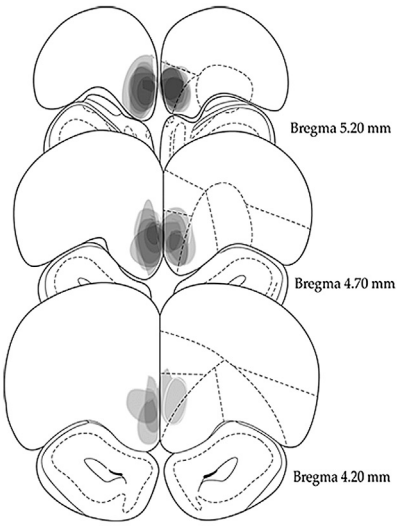
More recently, these and other observations have been developed into a formal theory of OFC function, suggesting that it provides a cognitive map of "task state space" (Wilson et al., 2014; Stalnaker et al., 2015). Specifically, reinforcement learning (RL) models posit that animals represent the structure of tasks through sets of "states" connected to each other through actions. States are generally signaled by environmental cues, such as a light or the sensory properties of the outcome. They can, however, be inferred internally based on previous cues that are not currently present and Wilson et al. (2014) contend that the OFC is involved in state representation in this latter case (i.e., when states are not explicitly cued). Although Wilson et al. (2014) did not assign this function specifically to medial or to ventral/lateral OFC, a number of findings suggest that ventral and lateral OFC play little if any role in goal-directed action (Ostlund and Balleine, 2007b; Balleine et al., 2011; Fellows, 2011). Together with the issues discussed above, among others (Rich and Wallis, 2014), this led us to evaluate this hypothesis by focusing on a potential role for the medial division of the OFC in the representation of the state space controlling goal-directed decision-making. To test this suggestion, we first established a

A

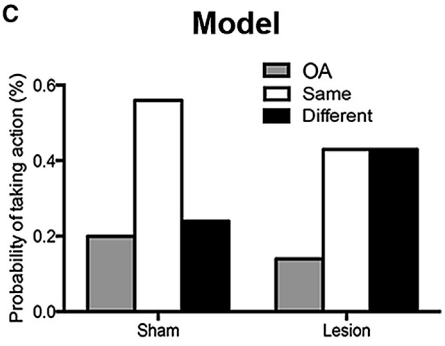
Pavlovian-instrumental-transfer

Pavlovian	Instrumental	Test
S1 – O1 S2 – O2	A1 – O1 A2 – O2	S1: A1 vs. A2 S2: A1 vs. A2

B



C



D

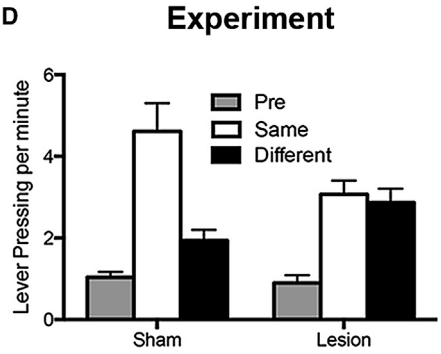


Figure 1. mOFC Lesions Impair Specific-PIT

(A) In the Pavlovian phase, rats were trained to associate two stimuli, S1 and S2, with pellet and sucrose outcomes (O1 and O2, counterbalanced). In the instrumental phase, the same rats were trained to associate two lever press responses (A1 and A2) with the same pellet and sucrose outcomes (A1-O1 and A2-O2). On test, each stimulus was presented separately and rats were given a choice between levers (A1 versus A2). The lever presses were recorded, but no outcomes were delivered.

(B) Representation of lesion placements showing each overlapping cytotoxic lesion.

(C) Simulated predictions for the responding of sham/control and lesion animals in specific-PIT. The error bars are present, but negligible on this, and all the simulation figures (OA = other action; i.e., any action other than the target actions).

(D) Mean lever pressing per minute (±1 SEM) during specific-PIT test for sham and lesioned animals.

series of hypotheses based on the suggested involvement of the mOFC in tasks for which states signaled by the outcome either were not observable, in outcome-specific devaluation and Pavlovian-instrumental transfer tests, or were fully observable, using contingency degradation and outcome-specific reinstatement tests. We simulated these hypotheses using the RL modeling approach proposed by Wilson et al. (2014) and then tested these hypotheses directly first using lesions of mOFC and then using the hM4Di designer receptor exclusively activated by designer drug (DREADDs) approach to generate more temporally specific inactivation.

These tests provided very clear support for the hypothesis, revealing performance deficits in rats with the mOFC inactivated when they were forced to rely on only partially observable task information. To provide direct evidence for a bias in information processing in mOFC animals, however, we developed a novel three stage decision-making task in which we aimed to put observable and unobservable outcome information into conflict, predicting that, whereas rats with an intact mOFC should select actions based on retrieved outcome values, those with mOFC lesions should eschew unobservable outcomes and select actions based on the value of observable outcomes alone.

RESULTS

See Supplemental Information for the full methods of the simulations and the experiments.

The mOFC Is Necessary for the Effects of Predictive Learning and Outcome Devaluation on Choice in a Partially Observable Task Environment

Predictions

We assessed the role of the mOFC in action selection in tasks with partially observable task information in two ways. First, we developed predictions computationally by simulating the role of mOFC in specific Pavlovian-instrumental transfer (specific-PIT) and in outcome devaluation using the RL modeling approach described by Wilson et al. (2014). Subsequently, we tested the predictions directly using cytotoxic lesions and chemogenetic-induced inactivation of the mOFC in rats. Full details of the simulations are provided in the Supplemental Information. First, in specific-PIT, the design of which is presented in Figure 1A, a stimulus associated with a specific outcome typically biases choice toward actions that earn that same outcome. To model this, we assumed that the presence of a Pavlovian state adds a bias (D) to the probability of earning the corresponding outcome by taking the corresponding action (Cartoni et al., 2013): i.e., for the intact agent, S_1:$p(O_1/A_1) \rightarrow p(O_1/A_1)+D$ and S_2: $p(O_2/A_2) \rightarrow p(O_2/A_2)+D$. Note that the specific-PIT test takes place in extinction, i.e., in the absence of an outcome. As such, in the case of mOFC dysfunction, we replaced states O_1 and O_2 with the single state O_{12} to represent the fact that, when the outcome is not present, animals cannot make specific predictions about the identity of the future outcomes: i.e.,S_1,S_2:$p(O_{12}/A_1) \rightarrow p(O_{12}/A_1)+D$ and S_1,S_2:$p(O_{12}/A_2) \rightarrow p(O_{12}/A_2)+D$. The simulation environments for each phase are presented in Figure S1.

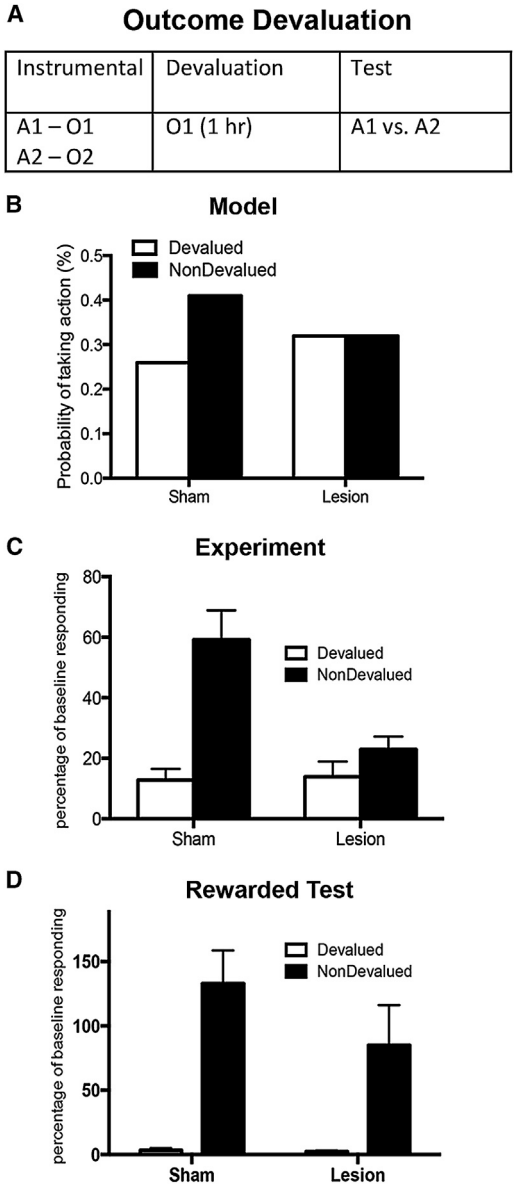


Figure 2. mOFC Lesions Impair Instrumental Outcome Devaluation Performance

(A) The instrumental phase is identical to Figure 1A (A1-O1 and A2-O2). For the devaluation, the rats were sated for 1 hr on one outcome, i.e., O1 (1 hr) prior to a choice test, A1 versus A2.

(B) Simulated predictions for the responding of sham/control animals in outcome devaluation (OA = 0.33 for Sham and OA = 0.36 for Lesioned animals) (data not shown).

(C) Mean percentage of baseline responding (±1 SEM) during outcome devaluation (in extinction) for sham and lesioned animals.

(D) Mean lever pressing per minute (±1 SEM) during outcome devaluation for sham and lesioned animals on a rewarded test.

Results of the simulations are depicted in Figure 1C. As the figure shows, for animals with an intact mOFC, the presentation of S_1 and S_2 results in increased performance of the action delivering the outcome predicted by the stimulus relative to the other action. Pavlovian contingencies are assumed to be intact in animals with mOFC lesions and so the stimuli should add bias to performance. However, because the rats cannot make specific predictions, each stimulus adds the bias to both action-outcome contingencies resulting in the general elevation of both actions.

For outcome devaluation (see Figure 2A), the performance of intact animals was modeled using the procedure described by Daw et al. (2005). During the devaluation phase, the reward value of one outcome decreases and so, in a subsequent extinction test, animals with an intact mOFC should recall the learned contingencies to select the action that earns the most highly valued outcome (the simulation environments for devaluation are presented in Figure S2). As the simulation in Figure 2B shows, in the case of intact animals, the action associated with the nondevalued outcome is selected more often, consistent with previous reports (Balleine and Dickinson, 1998). For animals with mOFC dysfunction, because testing in extinction ensures that information about which action leads to each outcome is unobservable, specific outcome predictions cannot be made. Therefore, the actions will be predicted to deliver both outcomes, and as the value of one of these outcomes is decreased by specific satiety, the simulation predicts a general decrement in the predicted value of both actions.

Testing Predictions: Lesions of the mOFC

This account predicts that both specific-PIT and outcome devaluation will be abolished in animals lacking a functional mOFC. We tested these predictions by comparing rats that had received permanent mOFC lesions by infusing cytotoxic concentrations of N-methyl-D-aspartate (NMDA) (n = 8) to sham controls (n = 9). Representations of lesion placements based on the stereotaxic atlas of the rat brain by Paxinos and Watson (1998) are shown in Figure 1B. mOFC lesions targeted the most rostral extent of the mOFC and produced substantial cell loss and shrinkage in this region. The caudal extent was almost never affected. A strict criterion was applied to lesion placements so that any lesions that affected a substantial part (>25%) of the prelimbic cortex were rejected. Therefore, 12 rats were excluded from the experiment because of incorrect lesion placement or size. There were 17 rats that were included in the analysis (Sham, n = 9 and Lesion, n = 8).

Specific-PIT

The lesions did not affect Pavlovian training or instrumental training (see Supplemental Information; Figure S1). In the specific-PIT, test rats received separate presentations of the stimuli (S1 and S2) and choice between the levers was assessed in extinction. As is clear from Figure 1D, only the sham group showed a robust specific-PIT effect, selectively increasing responding on the lever that had been paired with the same outcome as each of the stimuli during training. In contrast, the lesioned group responded nonselectively, increasing performance on both levers. There was a group × stimulus identity (same versus different) interaction, $F(1,15) = 7.06$, $p = 0.018$, supported by a significant simple effect for the sham group (same > different), $F(1,15) = 17.53$, $p = 0.001$, but no such effect for the lesion group (same = different), $F < 1$. Post-training chemogenetic inactivation of the mOFC using hM4Di DREADDs also produced a deficit in specific-PIT performance (see Supplemental Information; Figure S1M).

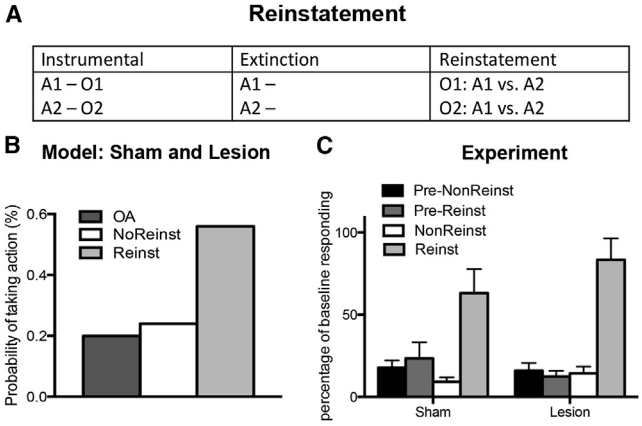


Figure 3. mOFC Lesions Spare Outcome-Selective Performance

(A) For outcome-selective reinstatement: the instrumental phase is identical to Figure 1A. The rats were then given 15 min extinction on both levers (A1- and A2-), followed by separate presentations of each outcome, after which the lever presses were recorded, but not rewarded.

(B) Simulated predictions for the responding of both sham and lesioned animals in reinstatement.

(C) Mean percentage of baseline responding (±1 SEM) during reinstatement.

Outcome Devaluation

Like specific-PIT, mOFC lesions produced impairment in outcome devaluation performance compared to shams (Figure 2C). For this test, the same rats used in the specific-PIT experiment were given 1 day of retraining on the action-outcome contingencies, followed 24 hr later by 1 hr of unrestricted access to one of the two outcomes to reduce its value through specific satiety. This was followed by a choice extinction test in which both levers were available, but no outcomes delivered. For the lesion study, we found a group × devaluation interaction, $F = (1,15) = 10.5$, $p = 0.005$, supported by a significant simple effect for the sham group (nondevalued > devalued), $F (1,15) = 34.39$, $p = 0.00$, but not the lesion group (nondevalued = devalued), $F (1,15) = 1.16$, $p = 0.298$. The effect of post-training chemogenetic inactivation of the mOFC was also assessed using hM4Di DREADDs in a separate group of rats. This treatment was found to produce a similar deficit in outcome devaluation performance (see Supplemental Information; Figure S2C), suggesting that the mOFC is specifically required during the choice test.

Taken together, the results of this first series of simulations and experiments demonstrate that the mOFC plays a general role in inferring task states in partially observable situations; in this case, based on outcome retrieval. Indeed, the simulations of this hypothesis predicted a pattern of results that was replicated by the actual performance of rats with lesions of the mOFC. Given the emphasis on unobservable outcomes in the role of mOFC, however, these effects should not emerge when outcomes are observable. We assessed this prediction in the next series of experiments.

The mOFC Is Not Required when Outcomes Are Observable

We conducted three separate tests of the role of mOFC when outcomes were delivered within the test (i.e., were observable): a rewarded devaluation test, an outcome-specific reinstatement test, and a contingency degradation test. As in the previous experiments, we first assessed predictions from the hypothesis for these specific tasks using the RL model (see Supplemental Information). For these simulations, it was assumed that, with outcomes delivered on test, each state was fully observable and trained contingencies could be recognized on test.

Rewarded Devaluation Test

Given this assumption, simulations suggested that no deficit in devaluation should be predicted for the rewarded test in rats lacking the mOFC (modeled identically to the result for intact rats in Figure 2B). This prediction was tested using the rats from the previous study. After retraining, devaluation was conducted as in the extinction test (i.e., a choice test between levers) except that the outcomes were delivered. We now found that outcome devaluation performance was intact for both groups (Figure 2D). There was a main effect of devaluation, $F (1,10) = 27.75$, $p = 0.00$, and no group × devaluation interaction, $F (1,10) = 1.36$, $p = 0.27$. Similarly, post-training chemogenetic mOFC inactivation using hM4Di DREADDs spared outcome devaluation performance in a rewarded test (see Supplemental Information; Figure S2D).

Outcome-Specific Reinstatement

The design of the outcome-specific reinstatement test is presented in Figure 3A. To model this effect, we followed the logic described for specific-PIT; i.e., the presence of an outcome during the test signaled that the outcome could be earned by performing its corresponding action, implying the action will be executed. Here, because the outcome is present during the test, we assumed that both intact and mOFC animals could correctly recall which outcome was produced by each action. In the first stage of the simulation, which was identical to the instrumental phase in specific-PIT, animals learned the contingencies between actions and outcomes. We assumed that rats without an intact mOFC were able to hold the outcome identity in working memory long enough to drive responding on the correct action without requiring access to a stored representation of the outcome. Therefore, intact and mOFC lesioned rats should both demonstrate intact performance. During the test, the effect of the presence of the outcome is modeled by adding a bias to the contingency of the corresponding action: i.e., O_1:$p(O_1/A_1) \rightarrow p(O_1/A_1)+D$ and O_2: $p(O_2/A_2) \rightarrow p(O_2/A_2)+D$ (the simulation environments for each phase of selective reinstatement are presented in Figure S3). As the simulation results indicate (Figure 3B), the presence of an outcome increases the likelihood that its corresponding action will be selected and in a similar fashion in animals with and without a functional mOFC.

To test this prediction, rats were first trained on two lever press actions to earn distinct outcomes, after which responding on both levers was extinguished for 15 min prior to two presentations of each outcome each separated by 7 min. Lever presses were compared for the period 2 min prior to and following each outcome delivery. Outcome specific reinstatement reflects an increase in the performance of the lever press action that, in training, was associated with the freely delivered outcome and this is the effect we observed in both the sham and mOFC lesioned rats (Figure 3C). Specifically, there was a main effect of reinstatement post-outcome delivery (reinstated > nonreinstated lever), $F (1,15) = 32.817$, $p = 0.00$, and no group × reinstatement interaction, $F < 1$.

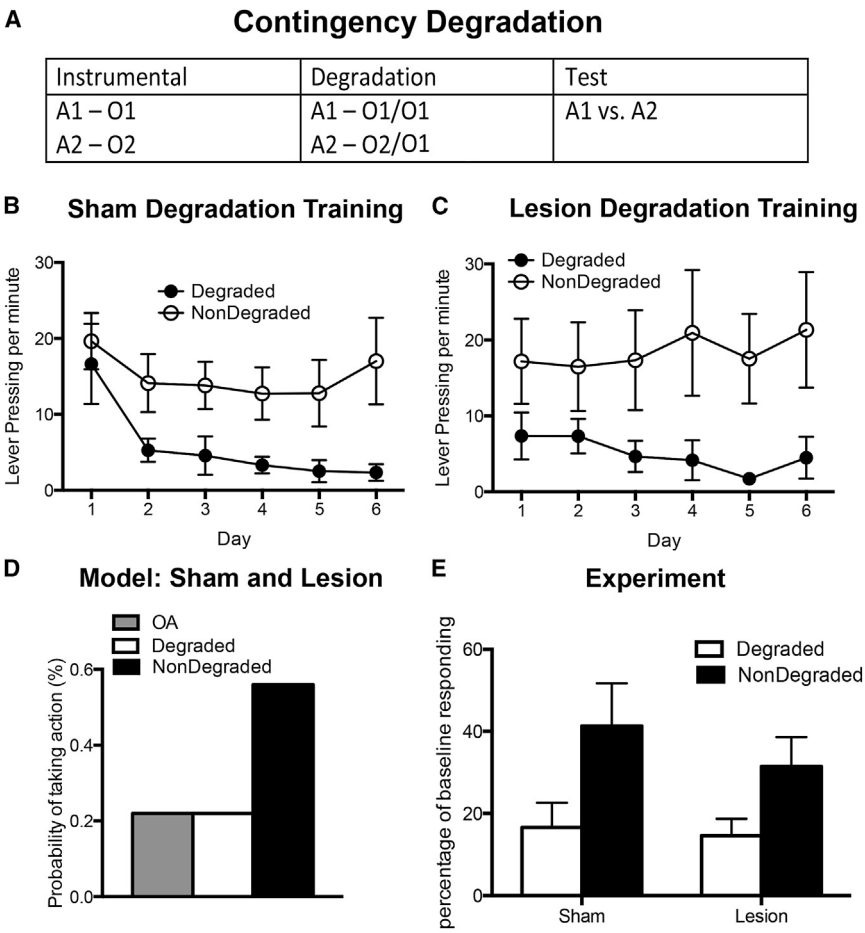


Figure 4. mOFC Lesions Spare Contingency Degradation

(A) For contingency degradation: the instrumental phase is identical to Figure 1A. During the degradation, one of the outcomes was delivered outside of the A-O contingency. The rats were then given a choice test, A1 versus A2.

(B) Mean lever pressing per minute (±1 SEM) for sham animals during contingency degradation.

(C) Mean lever pressing per minute (±1 SEM) for lesioned animals during contingency degradation.

(D) Simulated predictions for the responding of sham and lesioned animals during the contingency degradation test.

(E) Mean percentage of baseline responding (±1 SEM) of sham and lesioned animals during the contingency degradation test.

Contingency Degradation

Although intact performance in the rewarded devaluation and outcome-selective reinstatement tests suggests that mOFC lesioned rats encoded the appropriate action-outcome associations during training, it is possible that other contingencies mediated these effects. In the rewarded devaluation test, for example, sensitivity to punishment could be sufficient to rapidly rebias choice on test using stimulus-response (S-R) associations alone (Balleine et al., 2003), just as the formation of an (albeit very specific) S-R association could result in selective reinstatement (Ostlund and Balleine, 2007a). The clearest evidence of action-outcome encoding comes from a contingency degradation test (Dickinson and Mulatero, 1989; Balleine and Dickinson, 1998) and so we assessed performance in this test next.

The design of the contingency degradation assessment is presented in Figure 4A. As this figure shows, contingency degradation training was conducted in the presence of the earned outcomes, however, the test was not, and so it is critical in this study that we first establish predictions as to how rats with and without an intact mOFC will respond in this test based on our central hypothesis. To model degradation, the first stage is the same as previously described instrumental training. Degradation training typically arranges that the outcome (O1) earned by one action is delivered with an equal probability when the action is performed and when it is not performed, whereas the outcome (O2) of the other action depends solely on performing that action. Typically, rats decrease the performance of the degraded action relative to the nondegraded one (see Balleine and Dickinson, 1998). To model this, we assumed that the subject has a preference for actions that have a higher contingency (i.e., causal power denoted by ΔP). Contingency can be quantified by calculating the difference between the probability of gaining an outcome by performing an action relative to alternative actions, viz: $\Delta P(A_1) = P(O_1/A_1) - P(O_1/OA) - P(O_1/A_2) = 0$ versus $\Delta P(A_2) = P(O_2/A_2) - P(O_2/OA) - P(O_2/A_2) = 0.02$ (the simulation environments for each phase of contingency degradation are presented in Figure S4).

Simulation of the degradation test is depicted in Figure 4D (see Supplemental Information for more detail). Note that the representation of the environment (i.e., of the contingencies) will be the same in both groups because the outcome is present throughout training. As a consequence, in both groups the performance of the degraded action should decrease during degradation training because of a decline in ΔP that, having declined, will remain low during the extinction test. Therefore, even in the absence of the outcomes in that test, groups with and without a functional mOFC should similarly transfer degraded performance from training.

To test these predictions, we first retrained the rats on the same action-outcome associations and then introduced the change in contingency such that one of the outcomes was now also delivered independently of its lever press response. This degraded the contingency on that lever from having a positive value to zero. Importantly, as shown in Figures 4B and 4C, the acquisition of degradation was intact for both sham and mOFC lesioned rats (nondegraded > degraded). There was a main effect of degradation $F(1,15) = 9.677$, $p = 0.007$, and no group × degradation interaction, $F < 1$. Moreover, as predicted, the mOFC lesions did not affect performance in a choice

test conducted after the final day of degradation for which both levers were extended, but no outcomes delivered (Figure 4E). Again, there was a main effect of degradation, $F(1,15) = 9.334$, $p = 0.008$, but no group × degradation interaction, $F < 1$, nor a day × group × degradation interaction, $F < 1$. As is clear, therefore, the degradation effect observed was in accord with the predictions derived from the simulation illustrated in Figure 4D. Clearly, rats with damage to the mOFC were as able as the sham group to detect and to encode changes in the instrumental contingency and to transfer that contingency information to an extinction test. This is in contrast to previous results suggesting that, in the absence of the instrumental outcome, they were unable to transfer changes in performance based on predictive cues or changes in the value of the outcome. Therefore, in mOFC animals, information about the causal status of an *action* transferred from training to test, whereas, in other tests, information about the specific *outcome* associated with an action did not. This finding confirms, therefore, that mOFC critically mediates the influence of unobservable outcome-related information on decision-making.

mOFC Allows Rats to Infer Unrewarded States and so Choose Adaptively in the Absence of Specific Outcome Information

Our prior experiments provide consistent evidence that the mOFC plays a role in representing task states when these are based on unobservable action outcomes. However, this evidence was largely derived from a loss of function. In a final experiment, we sought to provide more direct evidence for this role of the mOFC by demonstrating a change in strategy after mOFC damage rather than merely the loss of one or other specific strategy. To achieve this, we conducted an experiment divided into three stages (see Figure 5A). For each stage, we could derive specific predictions regarding the performance of rats with an intact and a lesioned mOFC (Figure S5). In Stage 1, we trained naive rats to expect the absence of a specific outcome based on the performance of a specific lever press action. To achieve this, we first gave the rats Pavlovian pretraining associating distinct Pavlovian cues, S1 and S2, with each outcome, O1 and O2. Next, in alternating sessions, rats were trained to press the distinct levers (A1 and A2) for S1 and S2, but with the outcomes omitted. The aim of this intermixed training was to establish actions A1 and A2 as inhibitors of O1 and O2; i.e., whereas S1 alone predicted O1, A1 → S1 predicted the absence of O1 leaving A1 a specific inhibitor of O1. Similarly, whereas S2 alone predicted O2, A2 → S2 predicted the absence of O2 establishing A2 a specific inhibitor of O2. The simulation environment for this experiment is represented in Figure S5. Representations of lesion placements based on the stereotaxic atlas of the rat brain by Paxinos and Watson (1998) are shown in Figure 5B. mOFC lesions targeted the most rostral extent of the mOFC and produced substantial cell loss and shrinkage in this region. The caudal extent was never affected and any rats with substantial (>25%) prelimbic cortex (PL) damage were excluded. There were four rats that were excluded from the experiment because of incorrect lesion placement or size. Thus, 16 rats were included in the analysis (Sham, n = 8 and Lesion, n = 8).

Pavlovian Pretraining and Stage 1 Performance

We anticipated that rats with and without an intact mOFC could learn the initial Pavlovian S-O contingencies during pretraining, but that only those with an intact mOFC could infer the inhibitory A → noO relationships in Stage 1. Indeed, groups did not differ in Pavlovian pretraining; there was no main effect of group, $F < 1$, a main effect of stimulus period (Pre-CS, versus CS), $F(1,14) = 221.962$, $p = 0.00$, and of day, $F(1,14) = 16.998$, $p = 0.001$, but no group × stimulus period × day interaction, $F(1,14) = 1.989$, $p = 0.179$, Figure 5C. During Stage 1, however, we anticipated that lesioned rats would reduce their Pavlovian conditioned responses as they failed to learn the inhibitory contingencies (trained in alternating sessions), instead attributing A → S → no O and the S-O contingencies to a single state in which all stimuli are partially reinforced. In contrast, the responding of sham rats during Pavlovian conditioning should continue unaltered as they attribute the inhibitory A → S → no O contingencies to a different state (see Supplemental Information for further discussion). As expected, there was a trend toward the CS > PreCS difference becoming smaller over days for the lesion, but not for the sham group (marginal group × stimulus period × day interaction ($F(1,14) = 4.269$, $p = 0.058$; Figure 5D), but no difference in overall number of magazine entries (no main effect of group $F < 1$). Although inspection of Figure 5C suggests that this difference might have begun to emerge prior to Stage 1, it is notable that the group × stimulus period interaction was not significant at the end of Pavlovian pretraining $F(1,14) = 1.71$, $p = 0.212$ (Figure S5I), only began to emerge at the start of Stage 1 Pavlovian conditioning, $F = (1,14) = 3.621$, $p = 0.078$ (Figure S5J), and was significant by the end of Stage 1, $F(1,14) = 6.34$, $p = 0.025$, (Figure S5K).

This account further predicts that lesioned animals should make more magazine entries during the A → S → no O sessions themselves, as the failure to learn inhibitory contingencies ensures they continue to expect outcome presentations. Consistent with this, lesioned rats performed more overall magazine entries than sham rats during Stage 1 lever press sessions (main effect of group, $F(1,14) = 20.263$, $p = 0.00$) (Figure 5F). Further, there was a significant linear × group interaction, $F(1,14) = 5.009$, $p = 0.042$, suggesting that the lesion group linearly increased magazine entries across days of inhibitory lever press training, whereas the sham group did not. However, lever pressing itself did not differ between groups during Stage 1 (no main effect of group, $F(1,14) = 2.122$, $p = 0.167$, no group × day interaction, $F(1,14) = 1.85$, $p = 0.185$; Figure 5E), possibly because of low overall press rates.

Stage 2 Performance

To further assess this inferred inhibitory state, we challenged the rats to distinguish internal and external information about the presence/absence of the outcome by now delivering the outcomes where they had previously been omitted. That is, whereas previously A1 → S1 → no O1 and A2 → S2 → no O2, we now presented A1 → S1 → O1 and A2 → S2 → O2 and assessed the rate at which the new associations on the levers were acquired (effectively a retardation test). Sham rats were slower than lesioned rats to learn these excitatory associations, showing evidence of retardation (main effect of group, $F(1,14) = 10.52$, $p = 0.006$, and a group × day interaction, $F(1,14) = 11.29$, $p = 0.004$), as might

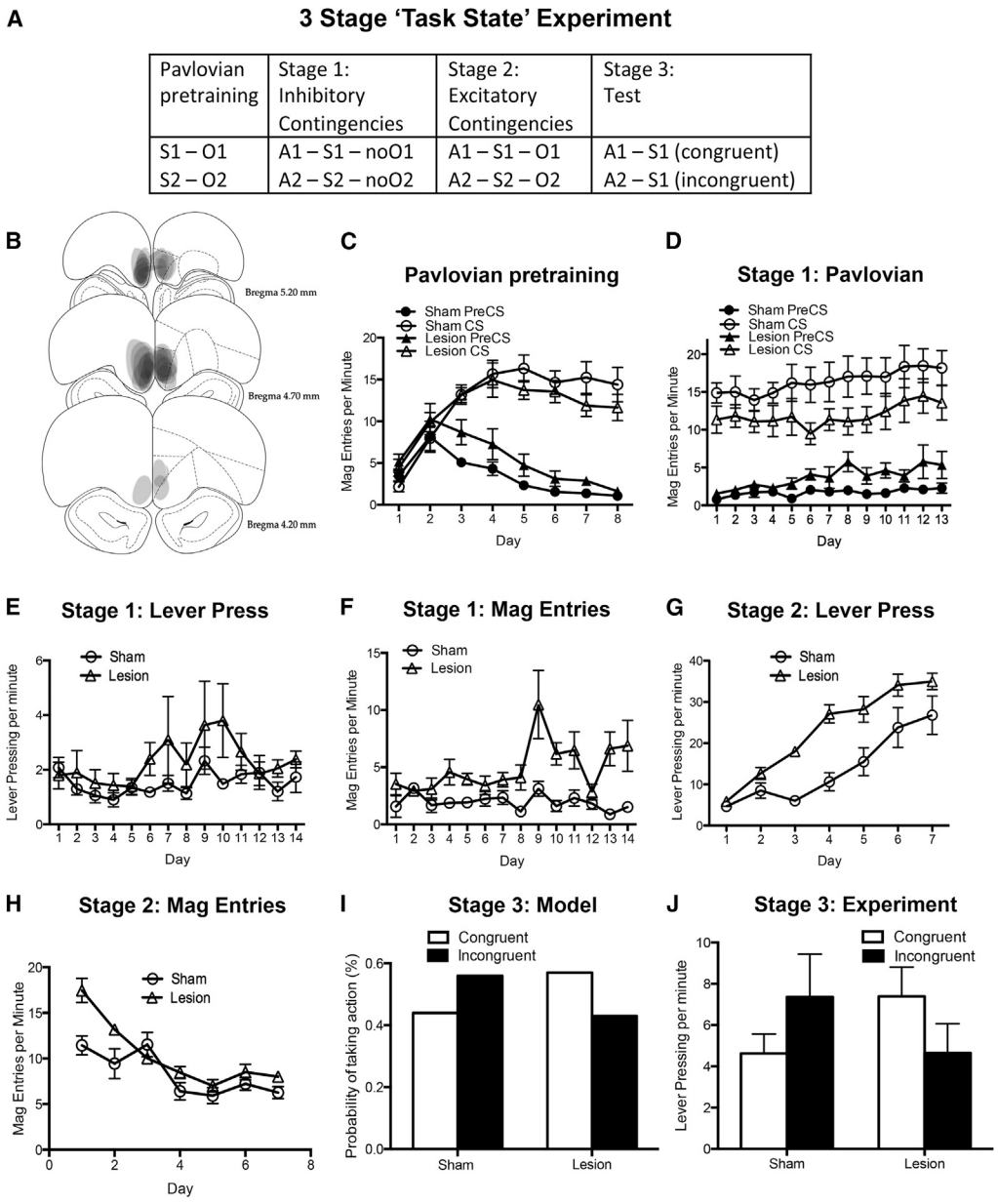


Figure 5. mOFC Lesions Impair the Ability to Identify States Defined by Inhibitory Action-Outcome Contingencies

(A) Rats received Pavlovian pretraining, during which they were trained to associate two stimuli, S1 and S2, with pellet and sucrose outcomes (O1 and O2, counterbalanced). Stage 1 then consisted of inhibitory contingency training: each time A1 action was taken S1 was presented, but the expected outcome O1 was not (A1-S1-noO1), and A2-S2 similarly predicted the absence of O2 (A2-S2-noO2). In Stage 2, rats were trained on the opposing excitatory associations (A1-S1-O1 and A2-S2-O2). On test, following 15 min of extinction on both levers, both actions earned S1 (i.e., A1-S1 and A2-S1). This stimulus was congruent with the trained contingencies for action A1 and incongruent for A2. Responding according to excitatory contingencies should be directed to A1 (i.e., congruent > incongruent), but responding according to inhibitory contingencies should be directed away from A1 and toward A2 (i.e., incongruent > congruent). The counterbalanced test was carried out 24 hr later.

(B) Representation of lesion placements showing each overlapping cytotoxic lesion.

(C) Mean magazine entries per minute (±1 SEM) during Pavlovian pretraining.

(D) Mean magazine entries per minute (±1 SEM) during Stage 1 Pavlovian conditioning (that alternated with lever press sessions).

(E) Mean lever pressing per minute (±1 SEM) during Stage 1 training of inhibitory contingencies.

(F) Mean magazine entries per minute (±1 SEM) during Stage 1 training of inhibitory contingencies.

(G) Mean lever pressing per minute (±1 SEM) during Stage 2 training of excitatory contingencies.

(H) Mean magazine entries per minute (±1 SEM) during Stage 2 training of excitatory contingencies.

(I) Simulated test results for Stage 3.

(J) Mean lever pressing per minute (±1 SEM) during the Stage 3 test.

be expected if sham, but not lesioned rats, encoded the prior inhibitory A-S-no O associations (thus inferring the "no outcome" state) and these competed with the current A→S→O associations (and the "outcome" state). Importantly, these differences were no longer significant by the end of excitatory lever press training in this second phase, $F(1,14) = 2.601$, $p = 0.13$. Further, although the group difference in magazine entries appeared to decrease during Stage 2 (Figure 5H), a main effect of group remained, $F(1,14) = 4.605$, $p = 0.05$ (although there was no group × day interaction $F(1,14) = 2.948$, $p = 0.108$).

Stage 3 Test

Finally, we conducted the critical test on the two levers over 2 days in which, after a period of extinction, both levers began to earn presentations of one of the stimuli, but no outcomes. On the first day of testing, both left and right levers earned S1 (A1→S1, A2→S1), whereas, for the second test day both levers earned S2 (A1→S2, A2→S2: for simplicity, only the first day of testing is shown in the design in Figure 5A). On both days, the stimulus earned by one of the levers was congruent with the trained associations (A1-S1 on day 1 and A2-S2 on day 2), whereas for the other it was incongruent (A1-S2 on day 1 and A2-S1 on day 2). In the absence of food outcomes, it was expected that sham rats would infer the "no outcome" state and act in accordance with the previous inhibitory associations; i.e., they should direct responding away from the congruent A1-S1 (day 1) and A2-S2 (day 2) combinations that previously predicted the absence of a specific outcome and toward the incongruent combinations A2-S1 (day 1) and A1-S2 (day 2) for which no inhibitory information had previously been provided. Thus, based on our simulations (Figure 5I; see Supplemental Information), we predicted that sham rats would respond more on the incongruent than the congruent lever. In contrast, we predicted that the mOFC lesioned rats would act in accordance with the only outcomes of lever pressing observable on test, the stimuli S1 and S2. Because these stimuli hold value as a result of prior Pavlovian conditioning, in the absence of inferring the inhibitory state mOFC lesioned animals should act in accordance with those previously trained contingencies and respond more for the congruent A→S relationships (i.e., A1→S1 and A2→S2) than for the incongruent ones (i.e., A1→S2 and A2→S1).

Responding during the initial extinction period was unremarkable, with no main effect of group ($F(1,14) = 3.188$, $p = 0.096$) and no group × min interaction, $F < 1$ (Figure S5L). More importantly, during the test period, and in line with our predictions and the results of our simulations (Figure 5I), we found no main effect of group ($F < 1$), or of lever ($F < 1$) but found a group × lever interaction, $F(1,14) = 10.465$, $p = 0.006$ (Figure 5J). Follow up analyses revealed significant simple effects of similar magnitude for both the sham, $F(1,14) = 5.233$, $p = 0.038$ and the mOFC lesion group, $F(1,14) = 5.233$, $p = 0.038$, although, the direction of the effect was different for the sham group (incongruent > congruent) relative to the lesion group (congruent > incongruent). Importantly, this interaction cannot be understood in terms of a simple deficit in extinction in the lesion group; a failure to extinguish would predict a nonspecific increase in responding on both levers. Similarly, extinction alone cannot account for the specificity of responding in the sham group.

The results of this experiment provide, therefore, three direct sources of evidence that rats with mOFC lesions adopt a distinctly different strategy to intact animals. Specifically, unlike shams, they rely on currently observable stimulus and action predictions rather than adjusting performance based on the inferred states derived from the predicted delivery of unobservable outcomes.

DISCUSSION

The results of the current series of experiments suggest that the mOFC mediates the retrieval of outcome identity when that information is necessary for choice between different goal-directed actions and where such outcomes are otherwise unobservable.

We began by testing recent accounts of OFC function that suggest its role is to retrieve task variables from memory that were previously observed, but that are currently unobservable or only partially observable (see Wilson et al., 2014; Stalnaker et al., 2015). To assess this suggestion, we first generated a series of hypotheses using computational RL modeling to derive specific predictions regarding mOFC involvement in a number of behavioral tasks. Simulation results suggested that mOFC dysfunction should impair performance in outcome devaluation and specific-PIT, both of which require the recall of specific outcome identities in extinction conditions. Our subsequent experiments confirmed these predictions; i.e., rats lacking a functioning mOFC showed clear deficits in each task. In contrast, the simulations and experimental results for tasks in which the outcome was fully observable—i.e., rewarded outcome devaluation, outcome-selective reinstatement, and contingency degradation—showed no difference between sham and mOFC lesioned animals. The failure to observe a deficit in contingency degradation was particularly instructive, demonstrating that mOFC involvement is specific to retrieving the outcome representation and that the retrieval of a previously encoded contingency between action and outcome (i.e., the altered action values encoded during the contingency degradation training phase) was not dependent on the mOFC.

Although consistent with the hypothesis, the evidence derived from these experiments was generated by deficits in function. In a final experiment, we developed a novel task in order to track the actual task representation in rats with and without a functional mOFC by providing an alternative strategy for the latter group. This experiment examined whether mOFC dysfunction prevented rats using outcome information to identify current states when these states were defined by either inhibitory or excitatory contingencies; i.e., states in which outcomes were absent and thus not observable (i.e., the inhibitory state) or in which they were present and so observable (the excitatory state). Predictions based on our RL model suggested that only animals with an intact mOFC should encode both states, whereas animals without an intact mOFC should encode only the state defined by excitatory contingencies (i.e., the state in which the outcomes were presented). Importantly, this was reflected in the experimental data; sham animals were able to identify the conditions of test (i.e., nonreinforcement), infer the inhibitory state, and act accordingly. The performance of the mOFC animals, by contrast, was consistent with the inference of a single excitatory state.

The Functional Role of mOFC Is Specific to Action-Dependent Outcome Retrieval

Together, these results suggest that the mOFC governs the retrieval of specific outcome representations for action selection in situations in which those outcomes are unobservable. This role for mOFC appears to be quite specific to the retrieval of stored outcome representations because mOFC inactivation spared the ability to hold information about recently experienced outcomes in short-term working memory. Specifically, the same animals that could not perform outcome devaluation or specific-PIT were able to respond accurately on the lever when partial reinforcement schedules enforced delays between outcome delivery, or during the test phase conducted after a single outcome delivery in the outcome-selective reinstatement test. Thus, although working memory mechanisms are often attributed to prefrontal cortical areas, our study suggests that they do not involve the mOFC, at least for working action-outcome associations, which must therefore depend on other regions (Wilson et al., 2014; Ragozzino et al., 2002).

The proposed role for mOFC also appears to be specific to the retrieval of action-dependent, as opposed to stimulus-dependent, outcome representations. This is consistent with findings that damage to the homologous region in macaque monkeys, i.e., Walker's area 14, did not affect stimulus-driven (Pavlovian) devaluation performance, but did affect rates of instrumental responding over days (Rudebeck and Murray, 2011). It is further consistent with the finding that instrumental outcome devaluation is impaired in Rhesus monkeys with lesions to their entire OFC (Rhodes and Murray, 2013). Furthermore, an fMRI study in humans revealed that the BOLD response in the mOFC was larger when participants chose an action associated with a valued outcome compared with an action associated with a devalued outcome (Valentin et al., 2007). Current results do appear to be at odds, however, with those of Gourley et al. (2010), who found no deficit in outcome devaluation performance in mice with (post-training) mOFC lesions. This difference is likely because the devaluation test employed by Gourley et al. (2010) was fundamentally different to the one performed here and involved comparing groups on a single score that consisted of lever pressing after being sated on the outcome that was normalized to lever pressing under deprivation conditions on the following day. Thus, their procedure appears to be more a test of general motivation than of sensory-specific outcome devaluation. The Gourley et al. (2010) finding is, however, also at odds with a prior finding that large OFC lesions encompassing mOFC impaired performance on a task similar to theirs (Butter et al., 1963).

Generally, therefore, the mOFC appears to be consistently involved in action-dependent outcome retrieval. Although the medial and lateral portions of the OFC appear to differ with regards to their role in instrumental outcome devaluation, there is some overlap in other deficits caused by damage to these regions. For example, similar to the results observed here, post-training lesions of lOFC impair the expression of specific-PIT (Ostlund and Balleine, 2007b) and combined lesions of the ventral/lateral OFC given either post- or pretraining impaired specific-PIT (Balleine et al., 2011). Unlike the current results, however, these vOFC/lOFC-induced deficits appear to be due to dysfunction in the processing of Pavlovian stimuli on choice; indeed these manipulations were found to impair Pavlovian contingency degradation (Ostlund and Balleine, 2007b) and previous evidence suggests that they also impair Pavlovian outcome devaluation (see Pickens et al., 2005). Although rat, primate, and human homologies must be treated with caution, taken together with the current results, we suggest that the medial and lateral OFC play a similar role, but in distinct functional systems: whereas the mOFC mediates action-dependent outcome retrieval, the v/lOFC appears to mediate stimulus-dependent outcome retrieval.

The mOFC Plays a Unique Role in Goal-Directed Action Relative to Other Prefrontal Cortex Structures

Another nearby structure that mediates responding in tasks similar to those in the current study is the PL. Specifically, like mOFC dysfunction, lesions of the PL have been found to abolish outcome devaluation performance (Corbit and Balleine, 2003; Killcross and Coutureau, 2003). There are three findings, however, that distinguish the pattern of deficits produced by PL dysfunction from current results: (1) PL damage leaves specific-PIT performance intact (Corbit and Balleine, 2003), (2) PL damage abolishes instrumental contingency degradation learning (Balleine and Dickinson, 1998; Corbit and Balleine, 2003), and (3) post-training PL lesions have been found to spare outcome devaluation performance (Ostlund and Balleine, 2005; Tran-Tu-Yen et al., 2009). Thus, the pattern of deficits observed after PL damage is consistent with a role in encoding the action-outcome contingency and this differs from the effect of mOFC damage, which leaves contingency degradation intact. In RL terms, this "contingency-encoding" function for PL is described as P(O|A). The role of mOFC can be further distinguished from that of other structures that play a role in instrumental conditioning, such as the basolateral amygdala (BLA), which is thought to encode incentive value, or in RL terms, the learning of the reward value of each state (see Balleine et al., 2003; Parkes and Balleine, 2013). Specifically, although lesions of the BLA also attenuate outcome devaluation performance, unlike mOFC lesions, BLA lesions attenuate devaluation performance in both extinction and rewarded tests.

Therefore, the observed pattern of deficits resulting from mOFC dysfunction is unique and, in conjunction with the results of the three stage decision-making task, strongly suggests that the mOFC functions to establish task states based on the retrieval of currently unobservable (action-dependent) outcome information. Although it might be surprising, given their intrinsic similarity, that damage to different regions of prefrontal cortex (PFC) should produce such unique patterns of deficits, it is strongly supported by current and prior research findings as well as compelling evidence that the anatomical networks to which each structure contributes are also unique (Hoover and Vertes, 2011; Reep et al., 1996).

mOFC Involvement in Circuits Mediating Goal-Directed Learning and Performance

The mOFC must, of course, carry out its function in concert with many of these structures to form the broader circuit mediating goal-directed action. Tract tracing studies have revealed that the mOFC is widely connected with regions important for the

encoding and initiation of goal-directed action, including the PL, posterior dorsomedial striatum (pDMS), nucleus accumbens (NAC) core, and BLA, as well as the insular cortex (IC), and mediodorsal thalamus (MD) (Reep et al., 1996; Hoover and Vertes, 2011). However, current results suggest that the mOFC input is more critical for the performance of goal-directed actions; i.e., in the way information is used to select actions in choice situations rather than in encoding specific associations between actions and outcomes. This suggests that projections arising from regions such as the IC and BLA to mOFC, as well as projections from mOFC to the medial portion of the striatum (particularly the NAC core), might be particularly important to our observed effects because these are the structures that have been specifically implicated in the *performance* of goal-directed actions (Corbit et al., 2001; Parkes and Balleine, 2013; Parkes et al., 2015; Hart et al., 2014). Nevertheless, a possible mechanism that could require mOFC input during action-outcome learning is that of inhibition and particularly inhibitory action-outcome associations, the importance of which have only recently begun to be recognized (Laurent and Balleine, 2015). This could be achieved via projections from mOFC to structures involved in action-outcome encoding such as PL, pDMS, and MD (Corbit and Balleine, 2003; Corbit et al., 2003; Yin et al., 2005). The idea that the mOFC mediates inhibitory action-outcome encoding could also account for other findings regarding OFC involvement in reflecting on unchosen actions in studies of counterfactual reasoning (Steiner and Redish, 2012) and regret (Steiner and Redish, 2014).

Similarly, many of the findings regarding the role of the mOFC in probability and/or risk estimation can be reinterpreted within the current framework. Probability estimations require the ability to retrieve abstract outcome representations to determine how often an outcome has been present versus absent and, therefore, any deficit in this ability might also impair the accuracy of such estimations. Stopper et al. (2014) examined the effects of reversible inactivation induced by intra-mOFC baclofen-muscimol infusions during a probabilistic discounting task and found that inactivating the mOFC increased the number of times animals chose the "large/risky" lever relative to the "small/certain" lever. Closer analysis revealed that this was the result of a tendency to "win-stay" following the receipt of a large, but risky reward relative to controls, which is precisely what should be predicted if an animal were relying on its most recent experience of reinforcement instead of estimating the overall likelihood of receiving an outcome. Similarly, Mar et al. (2011) found that mOFC lesions produced deficits in delay discounting in rats in a situation in which they had to choose between one lever that earned a single pellet immediately, and another lever that earned four pellets at increasing delays. Rats with mOFC lesions did not differ from controls at 0 s, but at 10 and 20 s delays, persisted more than control rats in choosing the lever associated with a large reward. Again, this result is to be expected if these relative delays were sufficient to challenge outcome retrieval processes in the lesioned rats, resulting in a reliance on more immediate reward. These differences dissipated at longer (40 and 60 s) delays, however, possibly when the limits of working memory had been reached in the intact controls. Human patients with vmPFC/mOFC damage have also been found to display riskier behavior than healthy controls in gambling tasks (Clark et al., 2008).

It is important to note that the role we have outlined for mOFC is not simply an "action-focused" version of the Schoenbaum et al. (2009) stimulus-guided "outcome expectancy" account of OFC function, despite some obvious overlap. This is mainly because, unlike the current account, Schoenbaum et al. (2009) do not distinguish between outcome expectancies that are formed in the presence versus the absence of the outcome. Thus, on their account, mOFC inactivation should impair both. Other accounts of OFC function have posited that vmPFC/mOFC might govern the assignment of value or attention to relevant choices (see Rudebeck and Murray, 2011). However, neither account can explain why mOFC lesioned rats should favor congruent versus incongruent responding in our final experiment. In contrast, a failure accurately to assign value to, or attend to, relevant options should be expected to generate a nonspecific deficit in responding. The mOFC could, however, play a related role in matching incentive value information to the specific identity of the outcome in the absence of the outcome itself. In a similar manner, the current study could be interpreted as broadly consistent with an alternate "neuroeconomic" theory of OFC function (Levy and Glimcher, 2012), that suggests the OFC is responsible for translating information about different types of reward into a "neural common currency" to influence decision-making. Levy and Glimcher (2012) based their theory on a meta-analysis of human fMRI studies that found activity in the vmPFC/OFC to be reflective of specific information about a variety of outcomes including money, food, water, novelty items, pain, and social subjective values. In addition, the calculation of a risk aversion parameter for one reward type (e.g., food) that could predict risk aversion for another reward type (e.g., money) was provided as evidence that vmPFC/OFC activity reflects the conversion of different outcome-specific information to a common scale. The success of such calculations must, however, necessarily depend on an underlying comparison of the features of outcomes that belong to different reward types, and this must almost certainly occur in the abstract; that is, when one or more of those outcomes are not directly observable. It is conceivable, therefore, that the currently proposed function for mOFC could underlie these kinds of value-based comparisons.

Conclusions

The findings reported here reveal a new role for mOFC function. Although outcome devaluation and specific-PIT are thought to have dissociable roles in action selection, both require the retrieval of specific outcome representations when those outcomes are unobservable and both are impaired by mOFC dysfunction. By contrast, outcome-selective reinstatement and contingency degradation assessments are conducted in the presence of outcomes and are unaffected by mOFC dysfunction. These experiments suggest, therefore, that the ability to think through the consequences of actions depends on the mOFC, and that mOFC dysfunction can impair decisions about which action to take when those consequences must be inferred. Furthermore, our final experiment demonstrated that mOFC dysfunction also impairs the ability to identify current

states of the world based on retrieval of specific outcome information, something that clearly altered the animals' strategy toward selecting actions based solely on observable information. Taken together, these results suggest that the mOFC mediates the retrieval of action-dependent outcome representations in situations where such outcomes are unobservable, a capacity that is critical for animals accurately to identify the state of the world and so select the optimal goal-directed action.

EXPERIMENTAL PROCEDURES

Full details of the experimental procedures and all simulations are provided in the Supplemental Information.

Subjects

For each lesion study, male Long-Evans rats, weighing between 300–400 g at the beginning of the experiment were used as subjects. During behavioral training and testing, rats were maintained at ≈85% of their free-feeding body weight by restricting their food intake to 10 g of their maintenance diet per day. All procedures were approved by the University of Sydney Animal Ethics Committee.

Behavioral Procedures: PIT, Outcome Devaluation, Outcome-Selective Reinstatement, and Contingency Degradation

Pavlovian Training

For the first 8 days, rats were placed in operant chambers for 60 min during which they received eight 2 min presentations of two conditioned stimuli (CS; white noise or clicker) paired with one of two outcomes (pellets or sucrose) that were presented on a random time 30 s schedule throughout the CS. Each CS was presented 4 times, with a variable intertrial interval (ITI) that averaged to 5 min. CS-outcome pairings were counterbalanced. Magazine entries throughout the session were recorded and separated into a CS period and an interval before CS presentations of equal length (PreCS; 2 min).

Instrumental Training

For the following 8 days, rats were trained to lever press on random ratio schedules of reinforcement. Each session lasted for 50 min and consisted of two 10 min sessions on each lever (i.e., 20 min on left lever and 20 min on right lever in total) separated by a 2.5 min time-out period in which the levers were retracted and the houselight was turned off. The order of presentation of each lever was pseudorandom and counterbalanced, as were the specific lever press-outcome contingencies. For the first 2 days, lever pressing was continuously reinforced. Rats were shifted to a random ratio (RR)-5 schedule for the next 3 days (i.e., each action delivered an outcome with a probability of 0.2), then to an RR-10 schedule (or a probability of 0.1) for the final 3 days.

Specific PIT

Following the last day of instrumental training, responding on both levers was first extinguished for 8 min to reduce baseline performance. Subsequently, each CS was presented four times over the next 40 min in the following order: clicker-noise-noise-clicker-noise-clicker-clicker-noise. Each CS lasted 2 min and had a fixed ITI of 3 min. Magazine entries and lever pressing rates were recorded throughout the session and responses were separated into PreCS and CS periods (2 min each). Lever presses were recorded, but not reinforced.

Devaluation Extinction Tests

The following day, rats were given 1 day of instrumental retraining on RR-10 in the manner previously described. On the following day, rats were given free access to either the pellets (25 g placed in a bowl) or the sucrose solution (100 ml in a drinking bottle) for 1 hr. The aim of this prefeeding procedure was to satiate the animal specifically on the prefed outcome, thereby reducing its value relative to the nonprefed outcome (cf. Balleine and Dickinson, 1998). Rats were then placed in the operant chamber for a 5 min choice extinction test. During this test, both levers were extended and lever presses recorded, but no outcomes were delivered. The next day, a second devaluation test was administered with the opposite outcome. Rats were then placed back into the operant chambers for a second 5 min choice extinction test.

Outcome-Devaluation Rewarded Test

Rats were devalued on either pellets or sucrose in the manner described previously. Rats were then placed in the operant chamber for 15 min in which both levers were extended, lever presses recorded, and outcomes were delivered.

Outcome-Induced Reinstatement Test

Subsequent to devaluation testing, rats received instrumental retraining on an RR-10 schedule for 1 day. The next day, rats received a 15 min period of extinction to lower the rate of responding. They then received four reinstatement trials separated by 7 min each. Each reinstatement trial consisted of a single delivery of either the sucrose solution or the grain pellet. All rats received the same trial order: sucrose, pellet, pellet, and sucrose. Responding was measured during the 2 min periods immediately before (pre) and after (post) each delivery.

Contingency Degradation Procedure

Subsequent to the reinstatement test, rats again received 1 day of instrumental retraining on an RR-10 schedule. Contingency degradation training occurred over the following 6 days. Rats continued to receive these same action-outcome pairings on an RR-20 schedule. In addition, one of the two outcomes (either pellets or sucrose) was delivered outside of the lever press-outcome contingency, i.e., in each second that no lever pressing occurred, either sucrose or pellets were delivered with the same probability (p[outcome/no action] = 0.05) that a lever press earned that outcome. As a result, the probability of earning one of the two outcomes was the same whether the animal pressed the lever or not. The other action-outcome contingency was nondegraded because the rat was still required to press the lever to receive that outcome. For half of the animals, the lever press-pellet contingency was degraded, and the lever press-sucrose contingency remained intact. The remaining animals received the opposite arrangement. Rats were given two 20 min training sessions each day, one on each lever.

Contingency Degradation Extinction Test

After the final day of contingency training, rats in both groups received a 5 min choice extinction test. During this test, both levers were extended and lever presses recorded, but no outcomes were delivered.

Behavioral Procedures: "Three Stage Task State" Experiment

Stage 1: Inhibitory Response-Stimulus-No Outcome Associations

Schedule of Pavlovian and Instrumental Training. Rats received only Pavlovian training for the first 8 days. Rats then received 14 alternating sessions of Pavlovian and inhibitory instrumental training.

Pavlovian Training

Rats were placed in operant chambers for ≈15 min, during which they received 12 10 s presentations of two conditioned stimuli (CS; houselight or tone) paired with one of two outcomes (pellets or sucrose) on a random time 5 s schedule throughout the CS. CS presentations were pseudorandom with a variable ITI that averaged to 1 min. CS-outcome contingencies were counterbalanced. Magazine entries throughout the session were recorded and separated into a CS period and an interval before CS presentations of equal length (PreCS: 10 s).

Instrumental Training

Rats were trained to lever press on a RR-2 schedule of reinforcement. Each session lasted for 12.5 min and consisted of two 5 min sessions on each lever separated by a 2.5 min time-out period in which the levers were retracted. The order of presentation of each lever was pseudorandom. For half of the animals in each group, the left lever earned a 2 s houselight presentation and the right lever earned a 2 s tone presentation. The remaining animals were trained on the opposite action-stimulus contingencies. No food outcomes were delivered during the instrumental training sessions.

Stage 2: Excitatory Response-Stimulus-Outcome Associations

Instrumental Training

Rats were trained to lever press on RR schedules of reinforcement. Each session lasted for 22.5 min and consisted of two 10 min sessions on each lever separated by a 2.5 min time-out period in which the levers were retracted. The order of presentation of each lever was pseudorandom. Excitatory associations were trained in a manner that directly opposed prior inhibitory associations (see Supplemental Information for details). Each CS presentation was 2 s, and CSs and food outcomes were presented concurrently for that 2 s.

For the first day, lever pressing was continuously reinforced. Rats were shifted to a RR-5 schedule for the next 2 days (i.e., each action delivered an outcome with a probability of 0.2), then to an RR-10 schedule (or a probability of 0.1) for the final 4 days.

Stage 3: Test

Rats were tested over the 2 days following the final day of excitatory instrumental training. Each test session began with a 15 min period of extinction to lower the rats' rate of responding on both levers. After 15 min, lever presses on both levers started earning one of the two stimuli (houselight or tone, counterbalanced) for 5 min. The response-stimulus (R-S) associations were therefore congruent with training for one of the lever press-stimuli combinations and incongruent for the other. During this test, no outcomes were delivered, but lever presses were recorded. The next day, a second test was administered with the opposite stimulus. That is, if both levers had previously earned houselight presentations, they now earned tone, and if both levers previously earned the tone, they now earned houselight presentations.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2015.10.044.

AUTHOR CONTRIBUTIONS

L.A.B. and B.W.B. conceived and designed the experiments; L.A.B. and M.v.H. performed surgery and conducted the behavioral experiments; L.A.B. analyzed the data from those experiments; B.C. conducted the electrophysiology assessment; A.D. performed model-based simulations and A.D., L.A.B., and B.W.B. developed the predictions based on those simulations; L.A.B. and B.W.B. developed the theoretical interpretation of the data; and all authors contributed to writing the manuscript.

ACKNOWLEDGMENTS

The research reported in the manuscript was supported by grants from the Australian Research Council, grants FL0992409 and DP150104878, and by a Senior Principal Research Fellowship from the National Health & Medical Research Council of Australia to B.W.B., grant #APP1079561.

Received: May 5, 2015
Revised: July 29, 2015
Accepted: October 13, 2015
Published: November 25, 2015

REFERENCES

Balleine, B.W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology *37*, 407–419.

Balleine, B.W., Killcross, A.S., and Dickinson, A. (2003). The effect of lesions of the basolateral amygdala on instrumental conditioning. J. Neurosci. *23*, 666–675.

Balleine, B.W., Leung, B.K., and Ostlund, S.B. (2011). The orbitofrontal cortex, predicted value, and choice. Ann. N Y Acad. Sci. *1239*, 43–50.

Bechara, A., Damasio, A.R., Damasio, H., and Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. Cognition *50*, 7–15.

Bouret, S., and Richmond, B.J. (2010). Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. J. Neurosci. *30*, 8591–8601.

Butter, C.M., Mishkin, M., and Rosvold, H.E. (1963). Conditioning and extinction of a food-rewarded response after selective ablations of frontal cortex in rhesus monkeys. Exp. Neurol. *7*, 65–75.

Cartoni, E., Puglisi-Allegra, S., and Baldassarre, G. (2013). The three principles of action: a Pavlovian-instrumental transfer hypothesis. Front. Behav. Neurosci. *7*, 153.

Clark, L., Bechara, A., Damasio, H., Aitken, M.R.F., Sahakian, B.J., and Robbins, T.W. (2008). Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. Brain *131*, 1311–1322.

Colwill, R.M., and Rescorla, R.A. (1990). Effect of reinforcer devaluation on discriminative control of instrumental behavior. J. Exp. Psychol. Anim. Behav. Process. *16*, 40–47.

Corbit, L.H., and Balleine, B.W. (2003). The role of prelimbic cortex in instrumental conditioning. Behav. Brain Res. *146*, 145–157.

Corbit, L.H., Muir, J.L., and Balleine, B.W. (2001). The role of the nucleus accumbens in instrumental conditioning: Evidence of a functional dissociation between accumbens core and shell. J. Neurosci. *21*, 3251–3260.

Corbit, L.H., Muir, J.L., and Balleine, B.W. (2003). Lesions of mediodorsal thalamus and anterior thalamic nuclei produce dissociable effects on instrumental conditioning in rats. Eur. J. Neurosci. *18*, 1286–1294.

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

Decety, J., Skelly, L., Yoder, K.J., and Kiehl, K.A. (2014). Neural processing of dynamic emotional facial expressions in psychopaths. Soc. Neurosci. *9*, 36–49.

Dickinson, A., and Mulatero, C.W. (1989). Reinforcer specificity of the suppression of instrumental performance on a non-contingent schedule. Behav. Processes *19*, 167–180.

Fellows, L.K. (2011). Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. Ann. N Y Acad. Sci. *1239*, 51–58.

Gourley, S.L., Lee, A.S., Howell, J.L., Pittenger, C., and Taylor, J.R. (2010). Dissociable regulation of instrumental action within mouse prefrontal cortex. Eur. J. Neurosci. *32*, 1726–1734.

Hart, G., Leung, B.K., and Balleine, B.W. (2014). Dorsal and ventral streams: the distinct role of striatal subregions in the acquisition and performance of goal-directed actions. Neurobiol. Learn. Mem. *108*, 104–118.

Hoover, W.B., and Vertes, R.P. (2011). Projections of the medial orbital and ventral orbital cortex in the rat. J. Comp. Neurol. *519*, 3766–3801.

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

Lagemann, T., Rentzsch, J., Montag, C., Gallinat, J., Jockers-Scherübl, M., Winter, C., and Reischies, F.M. (2012). Early orbitofrontal hyperactivation in obsessive-compulsive disorder. Psychiatry Res. *202*, 257–263.

Laurent, V., and Balleine, B.W. (2015). Factual and counterfactual action-outcome mappings control choice between goal-directed actions in rats. Curr. Biol. *25*, 1074–1079.

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

Liu, H., Liao, J., Jiang, W., and Wang, W. (2014). Changes in low-frequency fluctuations in patients with antisocial personality disorder revealed by resting-state functional MRI. PLoS ONE *9*, e89790.

Mar, A.C., Walker, A.L., Theobald, D.E., Eagle, D.M., and Robbins, T.W. (2011). Dissociable effects of lesions to orbitofrontal cortex subregions on impulsive choice in the rat. J. Neurosci. *31*, 6398–6404.

Noonan, M.P., Kolling, N., Walton, M.E., and Rushworth, M.F. (2012). Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. Eur. J. Neurosci. *35*, 997–1010.

Ostlund, S.B., and Balleine, B.W. (2005). Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. J. Neurosci. *25*, 7763–7770.

Ostlund, S.B., and Balleine, B.W. (2007a). Selective reinstatement of instrumental performance depends on the discriminative stimulus properties of the mediating outcome. Learn. Behav. *35*, 43–52.

Ostlund, S.B., and Balleine, B.W. (2007b). Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. J. Neurosci. *27*, 4819–4825.

Parkes, S.L., and Balleine, B.W. (2013). Incentive memory: evidence the basolateral amygdala encodes and the insular cortex retrieves outcome values to guide choice between goal-directed actions. J. Neurosci. *33*, 8753–8763.

Parkes, S.L., Bradfield, L.A., and Balleine, B.W. (2015). Interaction of insular cortex and ventral striatum mediates the effect of incentive memory on choice between goal-directed actions. J. Neurosci. *35*, 6464–6471.

Paxinos, G., and Watson, C. (1998). The Rat Brain in Stereotaxic Coordinates, Fourth Edition (San Diego, CA: Academic Press).

Pickens, C.L., Saddoris, M.P., Gallagher, M., and Holland, P.C. (2005). Orbitofrontal lesions impair use of cue-outcome associations in a devaluation task. Behav. Neurosci. *119*, 317–322.

Ragozzino, M.E., Detrick, S., and Kesner, R.P. (2002). The effects of prelimbic and infralimbic lesions on working memory for visual objects in rats. Neurobiol. Learn. Mem. *77*, 29–43.

Reep, R.L., Corwin, J.V., and King, V. (1996). Neuronal connections of orbital cortex in rats: topography of cortical and thalamic afferents. Exp. Brain Res. *111*, 215–232.

Rhodes, S.E., and Murray, E.A. (2013). Differential effects of amygdala, orbital prefrontal cortex, and prelimbic cortex lesions on goal-directed behavior in rhesus macaques. J. Neurosci. *33*, 3380–3389.

Rich, E.L., and Wallis, J.D. (2014). Medial-lateral organization of the orbitofrontal cortex. J. Cogn. Neurosci. *26*, 1347–1362.

Rudebeck, P.H., and Murray, E.A. (2011). Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior. J. Neurosci. *31*, 10569–10578.

Saxena, S., Brody, A.L., Schwartz, J.M., and Baxter, L.R. (1998). Neuroimaging and frontal-subcortical circuitry in obsessive-compulsive disorder. Br. J. Psychiatry Suppl. *35*, 26–37.

Schnider, A. (2013). Orbitofrontal reality filtering. Front. Behav. Neurosci. *7*, 67.

Schnider, A., Bonvallat, J., Emond, H., and Leemann, B. (2005). Reality confusion in spontaneous confabulation. Neurology *65*, 1117–1119.

Schoenbaum, G., Roesch, M.R., Stalnaker, T.A., and Takahashi, Y.K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nat. Rev. Neurosci. *10*, 885–892.

Stalnaker, T.A., Cooch, N.K., and Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. Nat. Neurosci. *18*, 620–627.

Steiner, A.P., and Redish, A.D. (2012). The road not taken: neural correlates of decision making in orbitofrontal cortex. Front. Neurosci. *6*, 131.

Steiner, A.P., and Redish, A.D. (2014). Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. Nat. Neurosci. *17*, 995–1002.

Stopper, C.M., Green, E.B., and Floresco, S.B. (2014). Selective involvement by the medial orbitofrontal cortex in biasing risky, but not impulsive, choice. Cereb. Cortex *24*, 154–162.

Tran-Tu-Yen, D.A., Marchand, A.R., Pape, J.R., Di Scala, G., and Coutureau, E. (2009). Transient role of the rat prelimbic cortex in goal-directed behaviour. Eur. J. Neurosci. *30*, 464–471.

Valentin, V.V., Dickinson, A., and O'Doherty, J.P. (2007). Determining the neural substrates of goal-directed learning in the human brain. J. Neurosci. *27*, 4019–4026.

Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., and Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. Neuron *81*, 267–279.

Yin, H.H., Ostlund, S.B., Knowlton, B.J., and Balleine, B.W. (2005). The role of the dorsomedial striatum in instrumental conditioning. Eur. J. Neurosci. *22*, 513–523.