Neuron
Article

Separable Learning Systems in the Macaque Brain and the Role of Orbitofrontal Cortex in Contingent Learning

Mark E. Walton,[1,2,*] Timothy E.J. Behrens,[1,2,*] Mark J. Buckley,[1] Peter H. Rudebeck,[1] and Matthew F.S. Rushworth[1]
[1]Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK
[2]These authors contributed equally to this work
*Correspondence: mark.walton@psy.ox.ac.uk (M.E.W.), behrens@fmrib.ox.ac.uk (T.E.J.B.)
DOI 10.1016/j.neuron.2010.02.027

SUMMARY

Orbitofrontal cortex (OFC) is widely held to be critical for flexibility in decision-making when established choice values change. OFC's role in such decision making was investigated in macaques performing dynamically changing three-armed bandit tasks. After selective OFC lesions, animals were impaired at discovering the identity of the highest value stimulus following reversals. However, this was not caused either by diminished behavioral flexibility or by insensitivity to reinforcement changes, but instead by paradoxical increases in switching between all stimuli. This pattern of choice behavior could be explained by a causal role for OFC in appropriate contingent learning, the process by which causal responsibility for a particular reward is assigned to a particular choice. After OFC lesions, animals' choice behavior no longer reflected the history of precise conjoint relationships between particular choices and particular rewards. Nonetheless, OFC-lesioned animals could still approximate choice-outcome associations using a recency-weighted history of choices and rewards.

INTRODUCTION

Learning, tracking, and updating the predictive value associated with environmental stimuli is essential to advantageous decision making. One region consistently implicated in the guidance of such adaptive choice behavior is the orbitofrontal cortex (OFC). It has been suggested that OFC is crucial for processing negative outcomes (Fellows, 2007; Kringelbach and Rolls, 2004) or for the ability to inhibit previously chosen actions (Chudasama and Robbins, 2003; Clarke et al., 2008; Dias et al., 1997; Elliott et al., 2000; Jones and Mishkin, 1972). Deficits in flexible adjustments of decision-making that are witnessed after OFC lesions are therefore often assumed to result either from either an insensitivity to the absence of rewards or a perseveration of choice.

To date, the cardinal tests of flexible reward-guided decision making have been two-option deterministic reversal learning tasks. It has been frequently demonstrated that OFC lesions impair performance following such reversals, even though the initial stimulus-outcome discrimination learning appears unaffected (Butter, 1969; Clarke et al., 2008; Dias et al., 1997; Fellows and Farah, 2003; Iversen and Mishkin, 1970; Izquierdo et al., 2004; Jones and Mishkin, 1972; Rolls et al., 1994; Schoenbaum et al., 2002). Both single-neuron and BOLD responses in this region also show rapid changes to reflect new associations when stimulus-reinforcement contingencies change (O'Doherty et al., 2003; Stalnaker et al., 2006; Tremblay and Schultz, 2000; Walton et al., 2004). Nonetheless, the precise role that OFC plays in this type of flexible decision-making is not clear. This is in part because such reversal tasks are limiting as they can often be solved using a simple rule-based strategy and do not require animals to continuously track the value of several independent alternatives to decide what to do.

Therefore, the present study was designed to reexamine the causal function of OFC in decision making in the context of changing reward values (Figure 1). Macaque monkeys performed different versions of a three-armed bandit task (Figures 1B–1E). In the first conditions, the reward associations of the three options could change both gradually and independently of one another meaning that fluctuations in outcome assignments occurred both with and without reversals in the identity of the most highly rewarding option (Figure 1D). To explore the role of OFC even in unchanging probabilistic environments, in the remaining three conditions, reward assignments remained stable although the average reward rate of the task environment was manipulated so that animals had to integrate across more trials in some conditions than others to discover the identity of the most highly rewarding option (Figure 1E).

Two key sets of findings were made. First, despite replicating the observation that selective OFC lesions impaired decision-making following reversal in the identity of the best-rewarded option, finer-grained analyses of trial-by-trial choice behavior demonstrated that this was not due to a failure to inhibit previously rewarding responses or insensitivity to negative outcomes as lesioned animals were also just as able as controls to respond to local changes in reward likelihood when the identity of the best stimulus remained the same.

The second set of findings demonstrate that the overall pattern of impairments can be explained by considering the OFC as critically concerned with specific contingent learning, the process by which the credit for an outcome becomes assigned to the

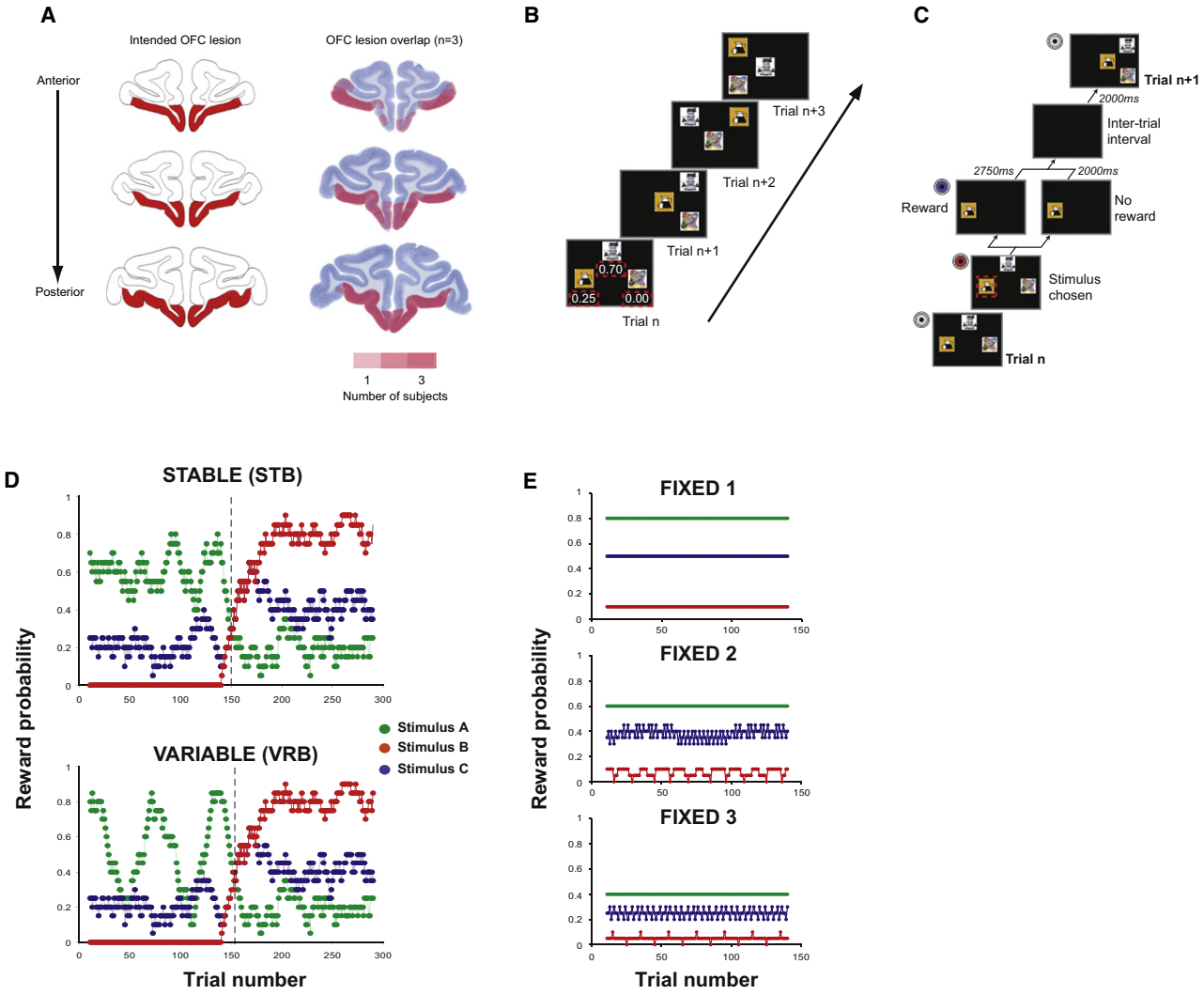

Figure 1. OFC Lesion Location and Task Schematic

(A) Diagram of intended (left) and actual (right) OFC lesion locations. Redness of shading on the actual lesion diagram represents the number of animals (1–3) showing overlap at each location.

(B and C) Schematic of trial-by-trial (B) and within-trial (C) task structure. On each trial, monkeys were presented with three clipart stimuli in one of four possible locations on a touchscreen (trials n to n+4). Each stimulus was associated with different outcome probabilities (example probabilities in red dashed boxes on trial n are shown for illustrative purposes only). On each trial, selecting one stimulus caused the other two options to extinguish and reward to be delivered according to the reward schedule. Gray, blue, and red circles = different 250 ms tones.

(D and E) Predetermined reward schedules used in the changeable (D) and fixed (E) conditions. The schedules determined whether or not reward was delivered for selecting a stimulus (stimulus A–C) on a particular trial. Dashed black lines in (D) represent the reversal point in the schedule when the identity of the highest value stimulus changes.

appropriate previous choice. This process is particularly taxed at times such as when reward contingencies change, where multiple different stimuli might be chosen and different outcomes witnessed (Seo and Lee, 2008).

It has long been known that, even during normal behavior, a given outcome can reinforce not just the choice that led to its delivery but also other responses that were made close in time, both preceding and even following this outcome (Thorndike, 1933). This "spread-of-effect" was visible in our control animals, though it was dwarfed by the tendency to associate an outcome with its correct, causal choice (i.e., appropriate credit assignment). By contrast, monkeys with OFC lesions exhibited a specific deficit in the ability to associate an outcome with its correct choice and, in doing so, unmasked their tendency to instead relate outcomes with choices that occurred close in time. This meant that their choice behavior was now predominantly driven by the association between the recent history of outcomes and recent history of rewards.

Such a facility to learn using recent choice and reward histories would allow animals to make accurate approximations of

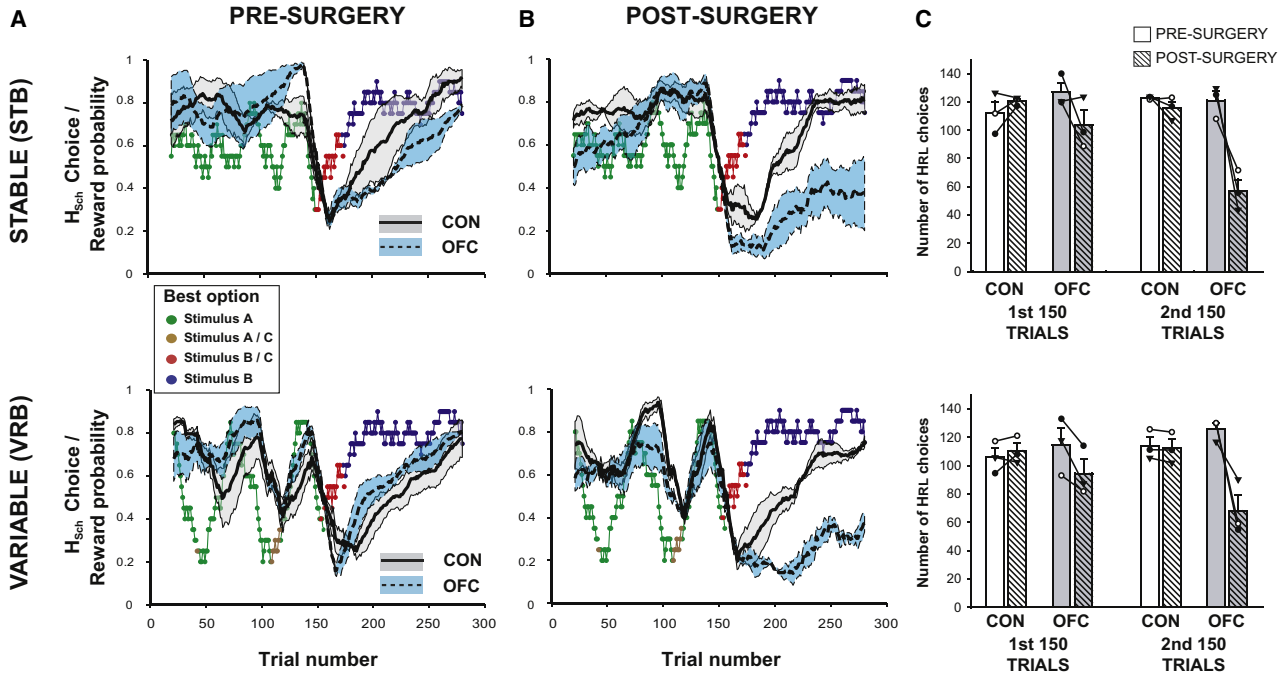

Figure 2. Likelihood of Choosing H_{sch} in STB (Upper Panels) and VRB (Lower Panels)

(A and B) Average pre- (A) and postsurgery (B) choice behavior in the control (solid black line) and OFC groups (dashed black line). SEMs are filled gray and blue areas respectively for the two groups. Colored points represent the reward probability and identity of H_{sch} (stimulus A–C).

(C) Average number of choices during the first or second 150 trials that were congruent with H_{RL} (the subjectively highest value option as defined by a reinforcement learning model). Controls, white bars; OFCs, gray bars. Symbols and connecting lines represent data for individual animals.

specific stimulus-outcome associations during periods when the same choice is made repeatedly for a constant rate of reward, but learning would be severely compromised when the pattern of choices and outcomes is variable, such as following a reversal. We demonstrate finally a key implication of this idea: OFC lesions impair learning even in environments where contingencies never change (Figure 1E), so long as initial credit assignment is made difficult by placing animals in a context where the average reward rates are lower, guiding animals to have a mixed history of choices between the available options. We therefore argue that a crucial role of the OFC is in learning and updating predictive contingent relationships between particular choices and consequent outcomes and that a failure in this faculty can account for the pattern of impaired decision-making seen after lesions to parts of primate OFC.

RESULTS

Changeable Three-Armed Bandit Schedules: Reversal Behavior

To probe the specific function of the OFC during flexible decision making, we tested six macaques—three controls and three animals given selective OFC lesions after presurgical testing (Figure 1A and see Supplemental Information available online)—on two types of continuously varying three-armed bandit tasks where animals had to choose their responses based on stimulus-reward probabilities (Figures 1B–1D). At the start of each testing session, animals were presented with three novel stimuli, meaning that they had no information other than the reinforcement delivered following a choice to guide their estimates of the expected values associated with that option. Whether or not reward was received for a particular stimulus choice was controlled by pre-determined outcome schedules. In the first set of experiments, two comparable outcome schedules were used—"Stable" (STB) and "Variable" (VRB)—in which the likelihoods of each alternative leading to reward varied continuously over the session, with identity of the most rewarding option reversing half-way through (Figure 1D, right of dashed line). Trial-by-trial reward probabilities were fixed according to these schedules and were identical for each animal and in each testing session using a particular schedule.

We analyzed the data based on both the "objective" value associated with each stimulus (based on a ±10 trial running average of the programmed reward probability, where the objectively highest value stimulus at any point in time was referred to as H_{sch}) and on estimates each animal's "subjective" value (the experienced reward probabilities obtained using a simple Rescola-Wagner model with a Boltzmann action selection rule [Behrens et al., 2007; Sutton and Barto, 1998]), where the subjectively highest value stimulus at any point in time was referred to as H_{RL}(see Experimental Procedures). Preoperatively, all animals rapidly learned to find the option with the highest probability of reward on both schedules (Figure 2A), and following the reversal of the identity of the H_{sch} option at around trial

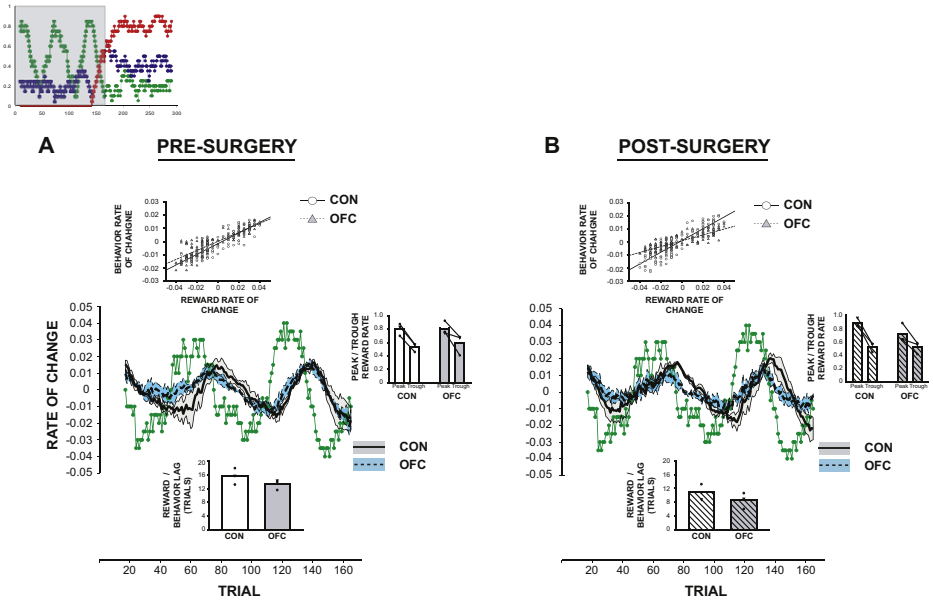

Figure 3. Tracking Value during the First 150 Trials of the Variable Schedule

Responsiveness of choice behavior to changes in reward likelihood of the highest value stimulus during the first 150 trials of VRB schedule (shaded area in upper inset) both before (A) and after (B) surgery. Main figure depicts rate of change of reward likelihood (green points) along with rate of change of behavior in controls (solid black line; gray shading = SEM) and OFCs (dashed black line; blue shading = SEM). Inset graphs show the average peak and lowest rates of choosing the highest value stimulus (right panel), the lag between changes in reward likelihood and behavior (lower panel), and the relationship between the rate of change of reward likelihood and of delagged choice behavior (upper panel). Controls, white bars; OFCs, gray bars.

150, animals then altered their pattern of choices to discover the new H_{sch}. There was no difference in the rates of selection of the highest value option in the two groups defined either by H_{sch} or H_{RL} (all $p > 0.14$). Importantly, too, there was no effect of testing session (all $p > 0.2$), suggesting that animals did not develop a model of the underlying task structure.

Following surgery, there was a dramatic change in choice patterns, with the OFC-lesioned group failing to find and persist with the new H_{sch} option following reversal on both schedules (Figure 2B). When the data were divided up into the initial learning and tracking phase (first 150 trials) and the reversal phase (second 150 trials), there was a significant three-way Lesion Group × Surgery × Phase interaction in the H_{RL} data ($F_{1,4} = 25.8$, $p = 0.007$), which post hoc tests showed was driven by the fact that the OFC group was only significantly worse at choosing the H_{RL} option than the controls postoperatively during the reversal phase in both conditions ($p = 0.001$) but not during any other period of testing (this was also true for H_{sch}: $p = 0.008$ for difference between the groups during the postoperative reversal phase, $p > 0.2$ otherwise; Figure 2C).

Changeable Three-Armed Bandit Schedules: Initial Learning, Value Tracking, and Choice Alteration

The above analyses demonstrated that there was no statistically-evident alteration in choice behavior during the first 150 trials of either schedule when OFC-lesioned animals initially had to learn and track the highest value stimulus in either condition. A further analysis investigating the average number of trials to reach a criterion of >65% H_{sch} choices prior to the reversal also found no significant differences in the rate of learning pre- or postoperatively (Mann-Whitney test: $p > 0.12$ in both conditions). This is comparable to several previous lesion studies to have reported effects following reversals along with intact discrimination learning (Clarke et al., 2008; Izquierdo et al., 2004; Schoenbaum et al., 2002).

Especially notable is that choice performance of the OFC-lesioned group is comparable to that of the control animals even in the VRB schedule when the local likelihood of H_{sch} resulting in reward is fluctuating markedly. This rapid behavioral response to local changes in the rate of both negative and positive feedback would appear to contradict several accounts of OFC function during flexible decision-making that have suggested that this region is important for detecting negative feedback in order to subsequently adjust behavior (Fellows, 2007; Kringelbach and Rolls, 2004).

To explore this, we examined three measures of stimulus-outcome sensitivity and flexible performance during this first phase of VRB: (1) the lag in trials between the H_{sch} reward rate fluctuating and H_{sch} choice performance changing (lower inset panel in Figures 3A and 3B), (2) the relationship between change in H_{sch} choice performance and H_{sch} reward rate fluctuations (adjusted for the above average lag in performance; upper inset panel in Figures 3A and 3B), and (3) the difference in the average highest and lowest proportion of H_{sch} choices during fluctuations in H_{sch} likelihood (right-hand inset panel in Figures 3A and 3B). These analyses together probe the degree to which OFC-lesioned animals are able to respond to changes, particularly decrements, in the local reward rate.

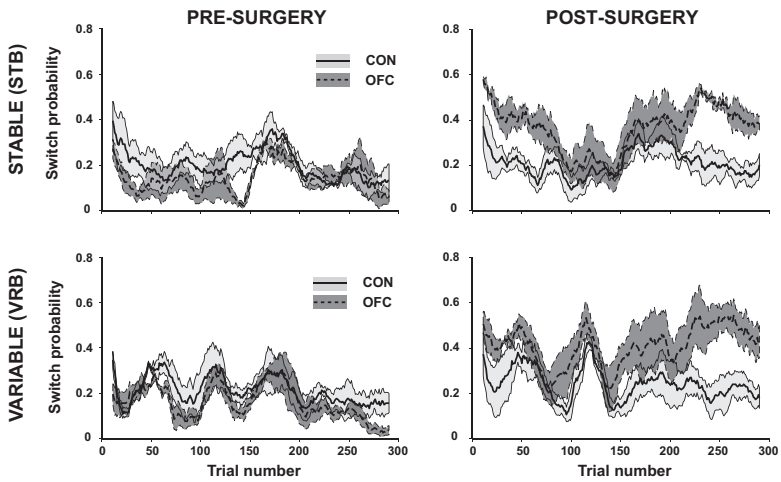

Figure 4. Rates of Switching Behavior during the Changeable Schedules

Pre- and postsurgery average trial-by-trial switching likelihood across STB and VRB in control (solid black line, light gray shading = SEM) and OFC (dashed black line; dark gray shading = SEM) animals. See also Figure S1.

There was no difference in how quickly the OFC-lesioned group on average responded to a local change in H_{sch} following surgery (measure a: Lesion Group × Surgery: $F_{1,4} = 0.01$, $p = 0.98$), and there was also no significant reduction in the sensitivity of the relationship between H_{sch} choice performance and H_{sch} reward rate postoperatively in the OFC group (as indexed by the slope relating these two parameters; measure b: Lesion Group × Surgery: $F_{1,4} = 1.95$, $p = 0.26$). Similarly, there was no significant reduction in the average range of H_{sch} choices in the two groups (measure c: Lesion Group × Surgery: $F_{1,4} = 1.65$, $p = 0.27$). Taken together, this demonstrates that during the first phase of VRB, the OFC-lesioned monkeys could track local changes in reward rates of the currently selected option, both when there was an increase in negative or positive feedback for selecting the best option, militating against any theory emphasizing the role of OFC in simply responding to negative feedback.

Such flexible behavior would also appear to rule out the notion that OFC lesions cause inflexible or perseverative responding in the face of changes in reinforcement (Elliott et al., 2000). This conclusion is bolstered by analyses of the trial-by-trial patterns of choice alternation behavior in the two groups. The point when OFC-lesioned animals were exhibiting impairments during the reversal phase on both schedules was actually associated with a local *increase* in the rate of switching between the alternatives (Figure 4). Overall, the lesioned monkeys were on average 1.6–4.7 times more likely to change their stimulus selection compared to the previous trial than prior to the lesion across the testing schedules (Lesion Group × Surgery: $F_{1,4} = 8.66$, $p = 0.042$; Figure 4). This was even the case examining just the 50 trials immediately postreversal in the two schedules ($p = 0.032$), underlining that any deficit here could not be caused by perseveration. Moreover, further analyses demonstrated that the OFC-lesioned animals' increased rate of switching was not a consequence of the reversal deficit causing these animals to receive less frequent rewards and was not modulated by receipt or absence of reward (Figure S1).

To summarize, the findings replicate previous studies demonstrating reversal deficits following a switch in the identity of the H_{sch} in OFC-lesioned animals accompanied by a largely intact ability to make appropriate choices when initially learning the values of the options. However, these same lesioned animals were able to track local changes in value of the currently chosen option. Rather than being caused by insensitivity to negative feedback or an inability to update response strategies, the OFC group's impairment was the result of an increased propensity to alternate between the different available options.

Specific Contingent Learning in a Changeable, Multioption Environment

OFC lesions cause profound deficits in flexible alterations of behavior. While such flexible learning must indeed be reliant on reward processing, it also has a determinant that is perhaps even more fundamental: the understanding of the *causal* relationship between a particular choice and its contingent outcome.

It has long been known that choices closely followed by reward are more likely to be repeated on subsequent occasions whereas those followed by aversive consequences become likely to be avoided ("Law-of-Effect," Thorndike, 1911). However, it is frequently overlooked that rewards do not just reinforce the choices that lead to them but also reinforce other choices made contiguously, either in the recent past or even those closely *following* on subsequent occasions. Such choices, even though they are just temporally contiguous with reward, rather than causally responsible for reward, are often repeated ("Spread-of-Effect," Thorndike, 1933; see also White, 1989). In runs of repeated choices, it is possible that such a mechanism could drive learning even in the absence of any direct association between choice and outcome. However, such a mechanism would be particularly inflexible in situations where choices or reward contingencies changed over time as ambiguities would exist as to which stimulus had caused which outcome.

It is therefore possible that the characteristic reversal deficit associated with OFC lesions is caused by an inability to associate a particular choice with a particular outcome. Both lesion and single-unit studies have suggested that OFC might carry a representation of the choice that was made when outcomes are received (Meunier et al., 1997; Tsujimoto et al., 2009). In order to test this idea, we ran a multiple logistic regression analysis (Barraclough et al., 2004; Lau and Glimcher, 2005) to see which combination of factors best explained animal choices. We included as regressors in the analysis all of the possible combinations of choice and outcome in the recent past (trials n-1 to n-5), along with a confound regressor for trial n-6 to capture

longer term choice/reward trends (Supplemental Information; Figure 5A). This allowed us to investigate the influence of specific choice-outcome associations on current behavior (red crosses, Figure 5A). Importantly, it was also possible to extract information about potential false associations as the value of an outcome is assigned backward based on choices made in previous trials (green area, Figure 5A) and as the value of an outcome spreads forward to choices made in subsequent trials (blue area, Figure 5A). In order to have adequate data to get accurate estimates of the strength of influence of these factors, we included data from both STB and VRB and two other analogous three-armed bandit schedules (Rudebeck et al., 2008; Figure S2).

Preoperatively, all animals' choices were strongly influenced by the stimuli they had recently selected and by the outcomes received for each of those choices, an effect that diminished with increasing separation from the current trial (Figures 5B, 5C, and S3). Prior to surgery, therefore, animals were able to associate specific choices with resulting outcomes. However, there was also a smaller influence of both the interaction between the previous reward and choice history (Figure 5D) and, for a few trials into the past, between the previous choice and reward history (Figure 5E). Hence, preoperative animals exhibited "Spread-of-Effect," being likely to associate outcomes with unrelated choices made near in time.

Following surgery, the influence of specific choice-outcome associations on behavior was profoundly reduced postoperatively, an effect that was particularly prominent on trials near to the current one (Lesion Group × Surgery × Past Trial: $F_{4,16} = 3.37$, $p = 0.035$; Figures 5B and 5C). OFC-lesioned animals therefore demonstrated a significant impairment in the ability to use the direct association between a specific choice and its resultant outcome to guide choice behavior. Unlike these specific associations, OFC lesions caused no effect on the degree to which animals associated the previous outcome with the choices made in the past (Figure 5D; interactions including Lesion Group × Surgery: $F < 2.37$, $p > 0.12$) or the degree to which they associated the past rewards with the previous choice (Figure 5E; interactions including Lesion Group × Surgery: $F < 0.93$, $p > 0.62$). Individual analyses of the postoperative data showed that there was a significant influence on current choices in both groups of associations between both the latest outcome and choice history and between the previous choice and reward history (Controls: both $F_{1,2} > 263.99$, $p < 0.005$; OFCs: both $F_{1,2} > 20.03$, $p < 0.047$).

This logistic regression analysis implies that OFC-lesioned animals are able to process the outcomes of choices but show a profound impairment in associating these outcomes with the relevant preceding choice on which they were contingent, instead forming an association between their overall integrated history of choices and an overall integrated history of outcomes. This theory makes explicit predictions of situations in which OFC lesioned animals should exhibit counterintuitive and counterproductive behavior. In the following sections, we examine these situations in detail. In brief, the theory predicts that OFC animals will perform like controls in situations where the integrated recent history of choices is strongly predictive of each individual choice, and the integrated history of rewards is strongly predictive of each individual reward.

If OFC-lesioned animals are using their history of choices and outcomes, rather than particular conjoint choice-reward associations, to update their value estimates for each option, this group should also then exhibit a particular pattern of deficits when a new stimulus is chosen (for example, option B) after long history of choice on another stimulus (i.e., option A), as is the case in reversal learning. (Note that options "A," "B," and "C" do not necessarily directly refer to stimuli A, B and C, as depicted in Figure 1, but instead to sequences of similar choices). To investigate this hypothesis, we examined the effect of an outcome—reward or no reward—on a newly chosen stimulus, after various different histories of choices. If credit is correctly assigned, animals should always be more likely to reselect B on the following trial (n) if its choice on the previous trial (n-1) was rewarded than if it did not result in reward. By corollary, they should be less likely to switch back to A after B's that are rewarded than those that are not. Moreover, this effect should of course be *independent* of choice history if all credit is properly assigned to the new choice, B. By contrast, if the credit for the new outcome is assigned not to the choice that causes the outcome, but instead to the integrated history of choices, then reinforcement for a reward for choosing option B will be assigned partly based on previous choices of option A. Importantly, as the length of previous choice history on A increases, a reward on B should make the animals *less likely to choose B and more likely to choose A*, as the credit for this outcome is falsely attributed to A.

Figure 6A (and Figure S4) shows just such an effect. We plotted the difference between trials following a reinforced B and those following an unreinforced B (trial n-1) in the likelihood that an animal will switch back to A (trial n). (Note that the likelihood of a reward on B being preceded by a rewarded A choice is not affected by the length of the previous choice history and is no different between the groups [both $p > 0.27$].) As predicted, preoperatively the effect of reward on B is to make them less likely to switch back to A (Figure 6A) and more likely to reselect B (Figure S4). However, the OFC-lesioned animals exhibited a very different pattern of responses (Lesion Group × Surgery × $Trial^{n-1}$ Reward: $F_{1,4} = 7.69$, $p = 0.050$), with these animals postsurgery showing both a significantly increased propensity to switch back to option A after a reward on B and a decreased tendency to switch back to A after *not* receiving a reward on B (post hoc tests, both $p < 0.05$; Figure S4). Analyzing just the postoperative data, there was also a three-way interaction between Lesion Group × $Trial^{n-1}$ Reward × Choice History ($F_{2,8} = 5.68$, $p = 0.029$), caused by the fact that this effect became more prominent the longer the previous choice history on A.

Such behavior—a tendency to link the reinforcement received on the previous trial to the recent history of choices—was not simply caused by an increase in random choices in the OFC-lesioned animals. After any history of A choices, a reward for choosing option B made a subsequent C selection *less likely in both groups* (main effect of $Trial^{n-1}$ Reward: $F_{1,4} = 18.79$, $p = 0.012$; no interactions between Lesion Group × Surgery, all $F < 1.1$, $p > 0.37$; Figure S4). Taken together, this demonstrates that the OFC-lesioned animals are updating stimulus-value representations based not on the specific association between a choice and its outcome but instead partially based on the history of recent choices and an outcome.

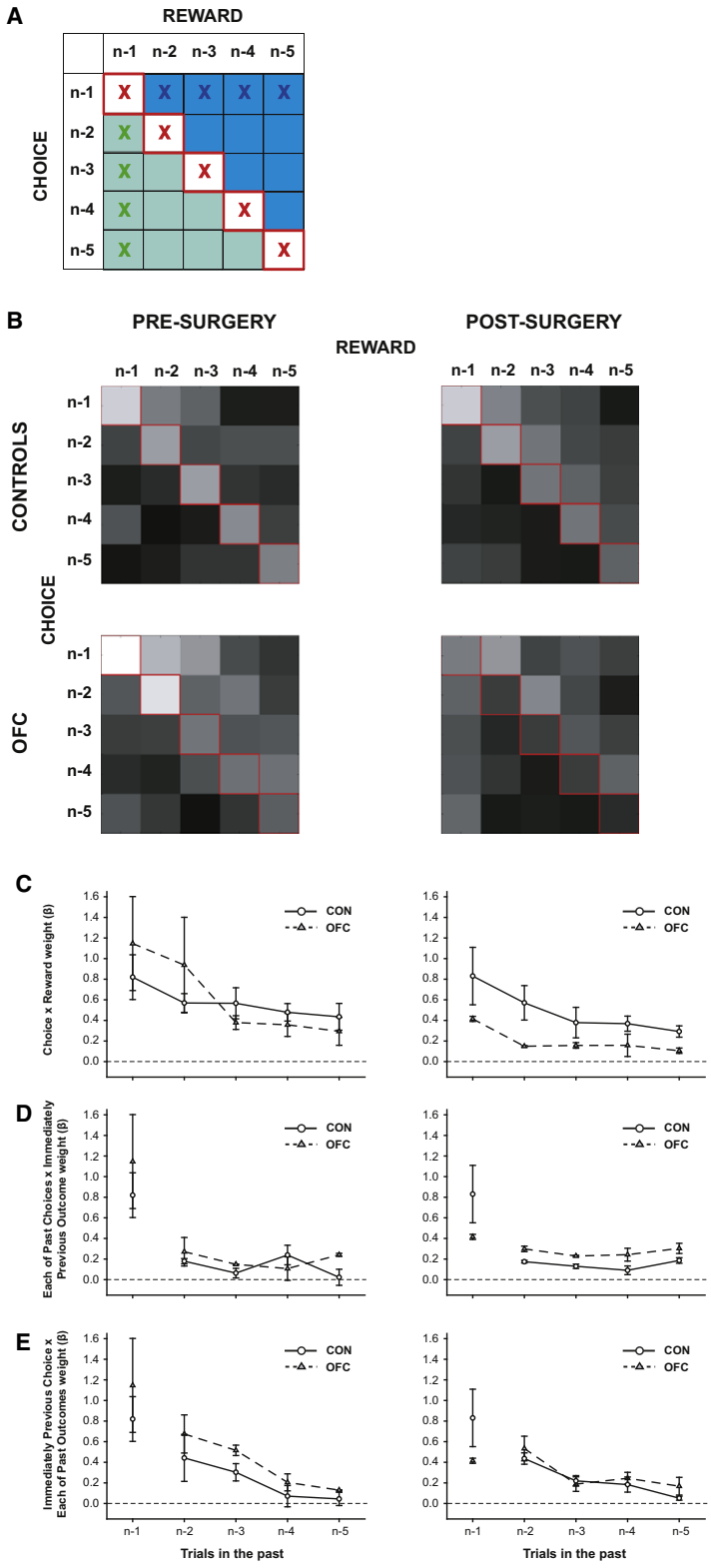

Figure 5. Influence of Recent Choices and Recent Outcomes on Current Behavior

(A) Matrix of components included in logistic regression. Red (i), green (ii), and blue (iii) X's respectively mark elements representing the influence of: (i) recent choices and their specific outcomes; (ii) the previous choice and each recent past outcome, and (iii) the previous outcome and each recent past choice, on current behavior. Green area represents influence of associations between choices and rewards received in the past; blue area represents the influence of associations between past rewards and choices made in the subsequent trials.

(B) Regression weights for this matrix for each group pre- and post-operatively, log-transformed for ease of visualization (bright pixels = larger regression weights).

(C–E) Plots of influence of X-marked components in (A). The data for the first trial in the past in (C)–(E) are identical. Symbols and bars show mean and SEM values for controls (black circles, solid black lines) and OFCs (gray triangles, dashed gray lines).

See also Figures S2 and S3.

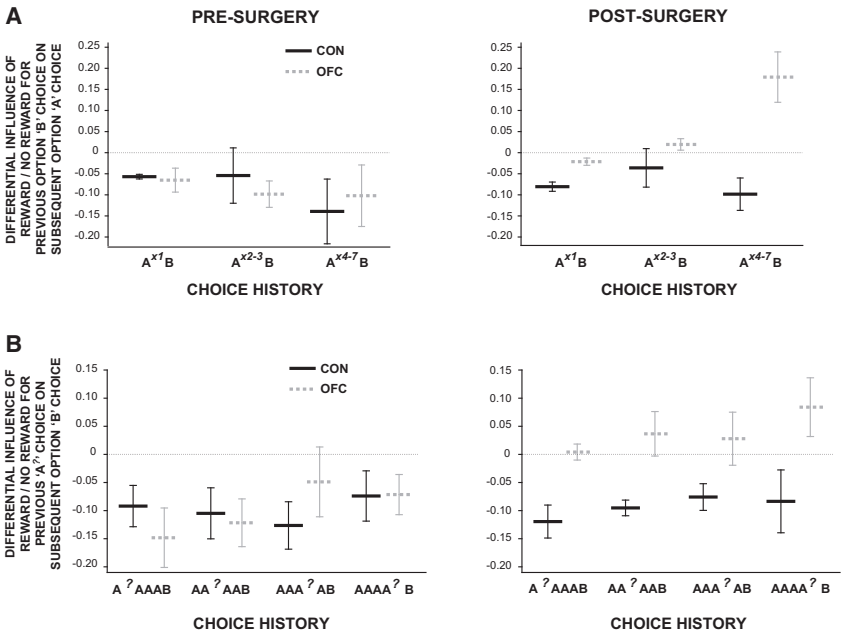

Figure 6. Influence of Past Choices (A) and Rewards (B) on Current Choice in Changeable Three-Armed Bandit Tasks

(A) Difference in likelihood of choosing option A on trial n after previously selecting option B on trial n-1 as a function of whether or not reward was received for this choice. Data is plotted based on the length of choice history on A (1 previous choice of A, left plots; 2–3 previous choices of A, middle plots; 4–7 previous choices of A, right plots). See also Figure S4.

(B) Difference in likelihood of choosing option B on trial n after previously selecting option A on trials n-2 to n-5 and option B on the previous trial (n-1), as a function of whether a particular previous A choice ($A^?$) was or was not rewarded. Bars show mean and SEM values for controls (solid black lines) and OFCs (dashed gray lines). See also Figure S5.

Similar predictions can be made if the OFC-lesioned animals are acting on an integrated history of *rewards* as opposed to updating associations on the basis of the most recent reward. For example, reward delivered at trial n-2 in the lesioned animals should be associated not only with the choice made at n-2, but also with the subsequent choice made at trial n-1(compare, in Figures 5B and 5E, the influence of an association between the outcome on trial n-2 with the choice on trial n-1 with specific choice-outcome associations). This leads to the counterintuitive prediction that a new "B" choice should be more likely to be repeated if a previous "A" choice were rewarded than if not. Precisely this effect can be observed in Figures 6B and S5. We extracted patterns of choices where animal selected the same option (e.g., A) on at least 4 occasions (trial n-2 to $\geq$ n-5) and then changed to a new option (e.g., B; at trial n-1). We then examined the probability that the animal would reselect the new option (B; at trial n) as a function of reward being delivered on one of the previous A choices.

Control animals were always more likely to select A, and less likely to select B or C, at trial n if a previous A choice had sometime been rewarded than if not, regardless on which previous trial a reward was delivered on A. The credit for a reward delivered after an A choice was predominantly being correctly assigned to stimulus A. Moreover, this effect was cumulative such that their likelihood of returning to an A choice at trial n increased as more of the previous A trials (n-2 to n-5) were rewarded. By contrast, the OFC-group showed a markedly different effect postsurgery such that, compared to when no reward was delivered for a recent A choice, a recent reward on A *decreased* the likelihood of choosing A and significantly *increased* the likelihood of persisting with option B than before the lesion (Lesion Group × Surgery × Option A Reinforcement: $F_{1,4} = 18.64$, $p = 0.012$). The reinforcement after the A choice was strongly affecting the value of a subsequent B choice despite the fact that the reinforcement occurred *before* the B choice (Figure 6B). Similarly, these animals' choices on trial n of options A or B (though, importantly, not of C) were significantly less influenced by the frequency of rewards for previous A choices than controls (e.g., switch to A: Lesion Group × Surgery × Past Option A Reinforcement Frequency: $F_{2,8} = 6.43$, $p = 0.022$).

When this effect was broken down, we found that it was mainly driven by a decrease in the OFC-lesioned animals' likelihood of choosing option B again following a recent choice of A that was not reinforced (post hoc test: $p = 0.003$), although there was also a trend in this group for an increase in the likelihood of persisting with B following a recent reward for A (post hoc test: $p = 0.061$; Figure S5). Inspection of Figure 6B shows that the effect was particularly prominent the closer in time the reward on A occurs to the option B choice. Nonetheless, there was no significant interaction between Lesion Group and the number of trials elapsed since the Past Reward Trial as inspection of Figure 6B shows that there is a similar, if less marked effect also in the controls.

Comparable examples of where the OFC-lesioned animals' choices either resembled or predictably diverged from control groups' choices as a function of recent reward, and choice history could also be observed when examining alternation behavior. OFC-lesioned animals were significantly more likely to switch than controls after only 1–2 choices of the same option ($p < 0.05$), but rates of persistence became indistinguishable following 3 or more choices of one stimulus (Supplemental Results; Figure S6).

Taken together, the evidence suggests that the choices OFC-lesioned animals make are strongly influenced by their recency- and frequency-weighted history of past choices and of previous rewards. Although OFC-lesioned animals appear unable to make and update specific associations between the stimulus they

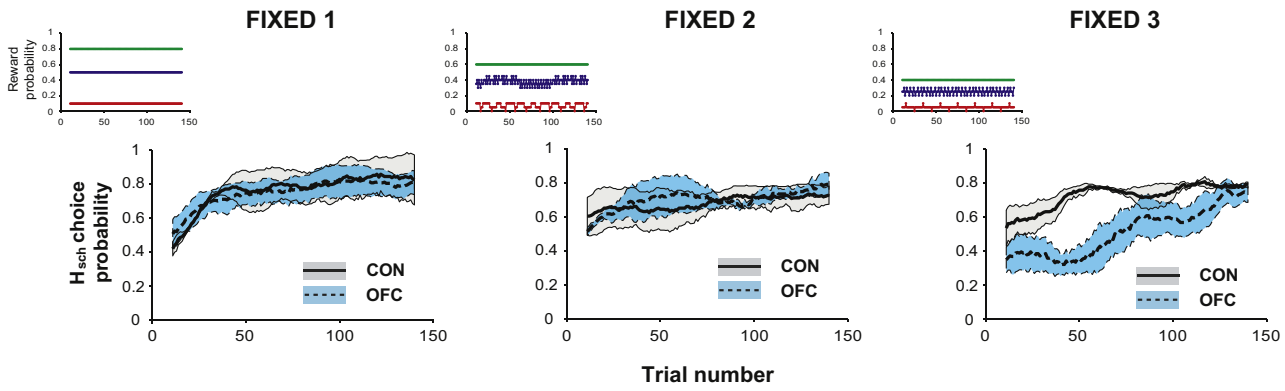

Figure 7. Likelihood of Choosing H_{sch} in the Fixed Three-Armed Bandit Schedules
Controls, solid black lines (gray shading = SEM); OFCs, dashed gray lines (blue shading = SEM). Inset panels depict each predetermined reward schedule.

chose and the outcome they received, they are just as able to use the contiguity between recent choices and rewards to select what responses to make. Such stimulus-outcome approximations would result in the animals learning accurate value representations when either their reward and/or choice histories are relatively constant, such as during the first 150 trials of STB and VRB, but will lead to aberrant learning when reward and choice histories are mixed and changeable, as exemplified by reversal situations where the values of stimuli alter such that the identity of the highest value stimulus changes.

Fixed Three-Armed Bandit Schedules

While the emphasis in many theories of OFC function has been on guiding choices in changeable environments, the evidence presented here indicates that this may be the result of a critical role for this structure in guiding specific contingent learning between stimulus-based choices and their outcomes which is severely taxed during reversal learning. Contrary to previous examples of intact discrimination learning in OFC-lesioned animals, it should therefore be possible to observe deficits following OFC lesions even using *fixed* schedules of reinforcement where there are never any reversals if it is made more difficult to determine the best option by lowering the reward likelihood of H_{sch}, therefore requiring animals to integrate across more choices to determine which option was best. To investigate this, we tested the control and OFC-lesioned monkeys on three new fixed three-armed bandit schedules (Figure 1E). The ratio of likelihoods of the three options was the same in each condition, but the overall yield differed, with the options in FIXED 2 or FIXED 3 schedules rewarding at 0.75 and 0.5 times the rate as in FIXED 1 (Figure 7).

Using comparable analyses to those employed with STB and VRB, we investigated the speed of learning to choose H_{sch} on $\geq$65% of trials. While OFC-lesioned animals determined the identity of H_{sch} in a similar number of trials in FIXED 1 and 2 (Mann-Whitney test: both $p > 0.8$), they were significantly impaired at finding this in FIXED 3 which had the lowest rate of reinforcement, even though the values associated with the stimuli were constant throughout (Mann-Whitney test: $p = 0.05$). However, although slower at learning, the OFC group did eventually reach a comparable level of performance as controls. When the session was divided up into 2 halves (first and second 75 trials) the OFC group made fewer H_{sch} responses in FIXED 3 during the first half of the session, when learning the identity of the best stimulus (main effect of group: $F_{1,4} = 7.97$, $p = 0.048$), but they were no different from the control group during the second half (main effect of group: $F_{1,4} = 1.64$, $p = 0.27$). Also as in the changeable schedules, this deficit manifested itself in increased patterns of choice alteration that were particularly evident in FIXED 3 (Condition × Group: $F_{2,8} = 5.64$, $p = 0.030$, post hoc tests showed significant difference between switching in FIXED 3 and the other schedules).

DISCUSSION

OFC has long been associated with enabling animals to alter their behavior in response to changes in reinforcement, particularly during reversals in stimulus-outcome associations (Fellows, 2007; Murray et al., 2007; Schoenbaum et al., 2007). The current study replicated this finding using a changeable, stochastic three-armed bandit paradigm, with OFC-lesioned animals being markedly slower to update their choices than controls when the identity of the highest value option reversed (Figure 2). However, in contrast to several accounts (Dias et al., 1997; Fellows, 2007; Kringelbach and Rolls, 2004), this deficit was neither caused by insensitivity to positive or negative feedback and accompanying changes in reward rate (Figures 3 and S1), nor was it due to perseverative response selection (Figure 4).

If OFC lesions do not cause difficulties in reward monitoring or inhibiting previous responses, the obvious question arises as to what specific role the OFC plays in guiding adaptive decision making in a changeable environment. Our findings imply that its function is to guide contingent learning, a mechanism that allows rewards received for a particular choice among several alternatives to be correctly credited to that option alone (Figures 5 and 6).

Such an impairment in appropriate contingent learning may seem to be immediately contradicted by the fact that the large majority of previous studies have implicated OFC as only important following reversals in outcome associations but not during

initial discrimination learning (Clarke et al., 2008; Fellows and Farah, 2003; Izquierdo et al., 2004; Schoenbaum et al., 2002) and that the OFC-lesioned animals' choice behavior in the current study was also largely unimpaired prior to the reversal. However, a second conclusion from our data is that OFC-lesioned animals are still able to rely upon alternative learning mechanisms which use representations of choice and reward history to approximate a link between the stimuli selected and outcomes received. This reliance results in the reinforcing effects of reward being assigned both backward based on the recency- and frequency-weighted history of choices (Figures 5A and S4) and also forward to choices made after an outcome is received (Figures 5B and S5).

A method of estimating stimulus value based on recent choice and reward histories, if fed into a response selection algorithm, would result in largely appropriate decision making in any situation when reward and choice histories are uniform; in other words when much of the recent choice history comprises of only taking one option and much of the recent reward history comprises of only one type of outcome (either reward or nonreward; Figure S1). Such uniform histories will be more likely when few alternatives are available or when the expected values of the options are far apart. It is just such conditions that prevail in the majority of studies of OFC in adaptive decision-making, the first half of the changeable three-armed bandit schedules (Figure 2), and in two of the three Fixed schedules tested here (Figure 7). However, when these conditions are not met, such as at a time of reversal (Figure 2) or, in the Fixed three-armed bandit schedule, where the absolute values were the lowest (Figure 7) then accurate learning is compromised. At such times, this method of approximating stimulus-reward associations will also often cause the values of the options to be closer together, leading to increased patterns of switching which, in turn, exacerbates the impairment (Figures 4 and S1).

Several lines of evidence support the idea that OFC impairment reflects the disruption of a mechanism for forming precise associations between particular choices and their resultant rewards, in the presence of an intact and simpler learning mechanism only capable of estimating such associations. First, data from logistic regression analyses demonstrated that the influence of precise paired associations between stimulus choices and reward outcomes, which was a strong predictor of choice behavior in control animals, was significantly reduced following OFC lesions (Figures 5B and 5C). By contrast, two other determinants of behavior remained, now unchecked, to influence decisions: first, recency-weighted associations between a received outcome and choices made before that outcome (Figure 5D), and second, recency-weighted associations between a chosen stimulus and rewards received before that choice (Figure 5E). While the overall weight of these factors was weaker in control animals than that of recent specific stimulus-outcome associations, both of these factors contributed to choices to an almost equal extent pre- and postsurgery in both the control and OFC groups. Therefore, it appears as if the approximation of stimulus-outcome associative learning used by the OFC group postoperatively is not a novel learning strategy to compensate for their deficit but instead is present in all animals. However, in normal animals, this "Spread-of-Effect" (Thorndike, 1933; White, 1989) will typically be dwarfed by the influences of knowledge of specific choice-outcome contingencies.

If it is the case that, without the guidance of specific stimulus-outcome associations, choice behavior is increasingly determined by the unmasked influence of choice history and recent reward, then this should have implications for the types of choices that are made. In controls and, prior to surgery, in the OFC group, a reward always increased the likelihood that animals would reselect the reinforced choice at the expense of others on later trials. However, following surgery in the OFC group, credit for a reinforcer was assigned not only to the choice that caused the reinforcement, but also to choices that had been made in surrounding trials (Figure 6). This effect was so pronounced that, after a long history of one choice (for example, option A), a new choice (i.e., option B) was *less likely to be reselected* if rewarded than if not (Figures 6A and S4); but also *more likely to be reselected* if the preceding A was rewarded than if not (Figures 6B and S5). In both cases, therefore, the credit for the reward was predominantly assigned to the wrong choice. However, this was done in a predictable fashion such that, rather than being assigned randomly, reinforcement was specifically distributed between choices that preceded and followed the reward.

Crucially, as in the controls, no credit was assigned to the third option (in this example, option C) that was absent in the local choice pattern. Reward for choosing either A or B made a future selection of C less likely in all circumstances (Figures S4 and S5). The specificity of the effect in relation to the recent choice history rules out any explanation based on undifferentiated increases in stimulus similarity or generalization after the lesions.

The claim that the OFC is essential for representing the conjunction of a particular reward with a particular choice having been made is consistent with several strands of evidence from previous studies. First, while there is an extensive literature demonstrating that OFC activity reflects information about anticipated and received reward value (Gottfried et al., 2003; Hare et al., 2008; Hikosaka and Watanabe, 2000; Schoenbaum et al., 2003; Tremblay and Schultz, 2000; van Duuren et al., 2008; Wallis and Miller, 2003) there are also data to show that OFC neurons dynamically encode information both about preceding and upcoming rewards (Simmons and Richmond, 2008). Second, as well as encoding outcomes, OFC maintains representations of currently relevant stimuli and choices over time or reactivates them at the time that reward is received (Lara et al., 2009; Meunier et al., 1997; Tsujimoto et al., 2009; Wallis and Miller, 2003). An extended role in contingent learning might also underlie an OFC contribution to maintaining expectations of specific outcomes and of future rewards across a delay. OFC lesions impair the ability of animals to alter their behavior toward a cue predicting a particular outcome if it has been devalued either by prefeeding or previously pairing it with nausea (Izquierdo et al., 2004; Ostlund and Balleine, 2007). Such a result might be expected if there were a deficit in representing the contingent link between different stimuli and their conjoint outcomes. OFC-lesioned animals fail to update associations correctly in situations when the outcome associated with one of two distinct stimulus-outcome pairings starts to be delivered non-contingently (Ostlund and Balleine, 2007). Similarly,

changes in delay-based decision making (Rudebeck et al., 2006) may also result from a failure to generate at the choice point a representation of a future large reward contingent upon tolerating a delay as well as from incorrectly updating value representations when the contingent choice and outcome are separated in time or by other choices (as may be the case in the "credit assignment problem").

OFC is well placed anatomically to mediate specific stimulus-reward learning. OFC, particularly lateral OFC which was the focus of the lesion in the current study (Figure 1A), is the recipient of afferents from high-level sensory areas in temporal and perirhinal cortex as well as of reinforcement information from limbic structures such as the amygdala (Carmichael and Price, 1995a, 1995b; Croxson et al., 2005; Morecraft et al., 1992). It is also one site of termination of dopamine fibers (Lewis, 1992) which could provide another source of information about expected value and deviations from such expectations (Schultz, 2007). OFC projects back to parts of the temporal lobe and amygdala, thus potentially allowing it to influence associative learning processes in these regions (Liu et al., 2000; Saddoris et al., 2005; Yanike et al., 2009). It may be important to understand the role that the OFC plays in contingent learning in the context of its relative specialization, in both the rodent and primate, for learning relationships between stimuli and outcomes rather than between responses and outcomes (Ostlund and Balleine, 2007; Rudebeck et al., 2008).

The conclusion that OFC is crucial for an aspect of stimulus-outcome learning and that this drives its role in reversal learning is consistent with the emphasis placed on OFC in associative learning by other researchers (Schoenbaum et al., 2007; Takahashi et al., 2009). The present study, however, emphasizes that more than one mechanism might mediate the association of stimuli with choices: an OFC-centered system for learning specific contingent stimulus-outcome pairings and at least one other more temporally-imprecise mechanism based on recent choices and outcomes that is spared in the OFC-lesioned animals. The notion that OFC-mediated reversal deficits are partly caused by the way remaining learning systems consequently function was previously implied by a study by Stalnaker and colleagues (2007) in which an OFC impairment was ameliorated following lesions to the amygdala, a structure known to be involved in aspects of associative learning.

While we and others have emphasized that there are many similarities between the structure and function of rodent and primate OFC (Price, 2007; Rushworth et al., 2007; Schoenbaum et al., 2006), it is nonetheless important to consider that cytoarchitectural studies indicate that primate OFC has expanded to include areas of granular and dysgranular cortex (Price, 2007; Wise, 2008), such as the relatively lateral OFC areas 11 and 13 that were the focus of the lesions in the current study. It is likely that other OFC regions, including more lateral, ventromedial or posterior agranular areas, may play subtly different roles in guiding stimulus-based learning and decision making (Butter, 1969; Fellows and Farah, 2003; Iversen and Mishkin, 1970).

In order for OFC-lesioned animals to be able to approximate associative learning based on recent choices and rewards, it is necessary that these elements should be represented in structures other than the OFC. Such signals have in fact proved to be relatively widespread and present in several brain areas, such as anterior cingulate cortex and striatum (Lau and Glimcher, 2007; Luk and Wallis, 2009; Seo and Lee, 2007, 2009). However, it is notable that these areas do not necessarily contain information about the conjoint history of rewards received in the context of particular choices, which may instead be a function of dorsomedial and lateral prefrontal cortex and lateral intraparietal cortex (Seo et al., 2007, 2009; Seo and Lee, 2009; Uchida et al., 2007). There is already a wealth of evidence for a multiplicity of learning systems in the brain (Balleine and Dickinson, 1998; Rangel et al., 2008; Weiskrantz, 1990). Our data provide evidence for a distinction between an OFC-based system for learning specific stimulus-reward contingencies and at least one additional extra-OFC system for reinforcement-based learning that incorporates recent choice history and the temporal contiguity of reward with subsequent choices.

In most studies to date, the contingencies between choices and reward have been straightforward, with limited available options and possible outcomes and with the associations between the two remaining stable. However, in many situations outside the laboratory, when there are frequently multiple alternatives and also delays between the consequences of responses and their causal antecedents, it is not straightforward to form appropriate associations. Learning in such complex situations can follow two distinct strategies: either through monitoring integrated choice and reward histories (for example, using eligibility traces: Sutton and Barto, 1998), or, where possible, through keeping track of individual choices and inferring precise associations between particular choices and rewards (Bogacz et al., 2007; Seo and Lee, 2008). This latter process likely requires an explicit encoding of the pertinent cues that have been encountered and the choices that have been made in order for reinforcement to update the appropriate predictors of eventual success or failure (Fu and Anderson, 2008). In the case of stimulus-reward learning, our findings suggest that the OFC may be crucial for the latter, but not the former, of these two strategies.

EXPERIMENTAL PROCEDURES

Animals

Six adult male rhesus macaque monkeys (*Macaca mulatta*), aged between 4 and 10 years and weighing between 7 and 13 kg were used in these experiments. Three animals acted as unoperated controls, whereas the other three received bilateral aspiration OFC lesions following training and presurgical testing. All animals were maintained on a 12 hr light/dark cycle and had 24 hr ad libitum access to water, apart from when testing. All experiments were conducted in accordance with the United Kingdom Animals Scientific Procedures Act (1986).

Behavioral Testing and Analysis

Prior to the start of experiments reported here, all monkeys had previous experience of using touchscreens and of other three-armed bandit tasks (see Rudebeck et al., 2008), though they had never performed with these particular schedules. On each testing session, animals were presented with three novel stimuli which they had never previously encountered, assigned to one of the three options (A–C). Stimuli could be presented in one of four spatial configurations and each stimulus could occupy any of the three positions specified by the configuration (Figure 1B). Configuration and stimulus position was determined randomly on each trial meaning that animals were required to use stimulus identity rather than action- or spatially based values to guide

their choices. Stimulus presentation, experimental contingencies, and reward delivery was controlled by custom-written software (Figure 1C).

Reward was delivered stochastically on each option according to five predefined schedules: STB and VRB (changeable schedules) or FIXED 1–3 (fixed schedules; Figures 1D and 1E). The likelihood of reward for any option and of H_{sch} and H_{RL} choices was calculated using a moving 20 trial window (±10 trials). Whether or not reward was delivered for selecting one option was entirely independent of the other two alternatives. Available rewards on unchosen alternatives were not held over for subsequent trials. Each animal completed five sessions under each schedule, tested on different days with novel stimuli each time. For STB and VRB, the sessions were interleaved (i.e., day 1, STB1; day 2, VRB1; day 3, STB2; day4, VRB2; etc.) and data were collected both pre- and postoperatively. For the FIXED conditions, schedules were run as consecutive sessions, starting with the five sessions of FIXED 1, then five sessions of FIXED 2, and finally five of FIXED 3 and only postoperative data were collected. The changeable schedules comprised of 300 trials per session and the fixed schedules of 150 trials per session.

The data from STB and VRB were analyzed both as a function of H_{sch} (the objectively highest value stimulus available) and of H_{RL} (the subjectively highest value stimulus given the animals' choices as derived using a standard Rescola-Wagner learning model with a Boltzmann action selection rule). The reward learning rate (α) was fitted individually to each animal's pre-surgery data using standard nonlinear minimization procedures.

Where appropriate, data from all tasks are reported using parametric repeated-measures ANOVA (see Supplemental Information).

To establish the contribution of choices recently made and rewards recently received on subsequent choices, we performed three separate logistic regression analyses, one for each potential stimulus (A, B, C). This gave us three sets of regression weights, $\widehat{\beta}_A, \widehat{\beta}_B, \widehat{\beta}_C$ and three sets of covariances, $\widehat{C}_A, \widehat{C}_B, \widehat{C}_C$. We proceeded to combine the regression weights into a single weight vector using the variance-weighted mean:

$$\widehat{\beta} = \left(\widehat{C}_A^{-1} + \widehat{C}_B^{-1} + \widehat{C}_C^{-1}\right)^{-1}\left(\widehat{C}_A^{-1}\widehat{\beta}_A + \widehat{C}_B^{-1}\widehat{\beta}_B + \widehat{C}_C^{-1}\widehat{\beta}_C\right)$$

Surgery

Surgical procedures and histology for these animals have been described in detail elsewhere (Rudebeck et al., 2008). In brief, animals were given aspiration lesions to the OFC using a combination of electrocautery and suction under isoflurane general anesthesia. The lesions were comparable to those reported in Izquierdo et al. (2004), taking the tissue medial to the lateral orbital sulcus up to the gyrus rectus on the medial surface (Figure 1A; see Supplemental Information).

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, Supplemental Results, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.neuron.2010.02.027.

ACKNOWLEDGMENTS

This work was funded by the Medical Research Council, UK, and the Wellcome Trust (M.E.W.). We would like to thank Mark Baxter for assistance with anesthesia, Greg Daubney for the histology, and the Biomedical Services Team for excellent animal husbandry, as well as Peter Dayan for constructive advice and Erie Boorman for helpful comments on the manuscript.

Accepted: February 24, 2010
Published: March 24, 2010

REFERENCES

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

Barraclough, D.J., Conroy, M.L., and Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. Nat. Neurosci. *7*, 404–410.

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

Bogacz, R., McClure, S.M., Li, J., Cohen, J.D., and Montague, P.R. (2007). Short-term memory traces for action bias in human reinforcement learning. Brain Res. *1153*, 111–121.

Butter, C.M. (1969). Perseveration in extinction and in discrimination reversal tasks following selective prefrontal ablations in Macaca mulatta. Physiol. Behav. *4*, 163–171.

Carmichael, S.T., and Price, J.L. (1995a). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. J. Comp. Neurol. *363*, 615–641.

Carmichael, S.T., and Price, J.L. (1995b). Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. J. Comp. Neurol. *363*, 642–664.

Chudasama, Y., and Robbins, T.W. (2003). Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. J. Neurosci. *23*, 8771–8780.

Clarke, H.F., Robbins, T.W., and Roberts, A.C. (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.

Croxson, P.L., Johansen-Berg, H., Behrens, T.E., Robson, M.D., Pinsk, M.A., Gross, C.G., Richter, W., Richter, M.C., Kastner, S., and Rushworth, M.F. (2005). Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. J. Neurosci. *25*, 8854–8866.

Dias, R., Robbins, T.W., and Roberts, A.C. (1997). Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: restriction to novel situations and independence from "on-line" processing. J. Neurosci. *17*, 9285–9297.

Elliott, R., Dolan, R.J., and Frith, C.D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. Cereb. Cortex *10*, 308–317.

Fellows, L.K. (2007). The role of orbitofrontal cortex in decision making: a component process account. Ann. N Y Acad. Sci. *1121*, 421–430.

Fellows, L.K., and Farah, M.J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. Brain *126*, 1830–1837.

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

Gottfried, J.A., O'Doherty, J., and Dolan, R.J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science *301*, 1104–1107.

Hare, T.A., O'Doherty, J., Camerer, C.F., Schultz, W., and Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J. Neurosci. *28*, 5623–5630.

Hikosaka, K., and Watanabe, M. (2000). Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. Cereb. Cortex *10*, 263–271.

Iversen, S.D., and Mishkin, M. (1970). Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. Exp. Brain Res. *11*, 376–386.

Izquierdo, A., Suda, R.K., and Murray, E.A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. J. Neurosci. *24*, 7540–7548.

Jones, B., and Mishkin, M. (1972). Limbic lesions and the problem of stimulus—reinforcement associations. Exp. Neurol. *36*, 362–377.

Kringelbach, M.L., and Rolls, E.T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog. Neurobiol. *72*, 341–372.

Lara, A.H., Kennerley, S.W., and Wallis, J.D. (2009). Encoding of gustatory working memory by orbitofrontal neurons. J. Neurosci. *29*, 765–774.

Lau, B., and Glimcher, P.W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. J. Exp. Anal. Behav. *84*, 555–579.

Lau, B., and Glimcher, P.W. (2007). Action and outcome encoding in the primate caudate nucleus. J. Neurosci. *27*, 14502–14514.

Lewis, D.A. (1992). The catecholaminergic innervation of primate prefrontal cortex. J. Neural Transm. Suppl. *36*, 179–200.

Liu, Z., Murray, E.A., and Richmond, B.J. (2000). Learning motivational significance of visual cues for reward schedules requires rhinal cortex. Nat. Neurosci. *3*, 1307–1315.

Luk, C.H., and Wallis, J.D. (2009). Dynamic encoding of responses and outcomes by neurons in medial prefrontal cortex. J. Neurosci. *29*, 7526–7539.

Meunier, M., Bachevalier, J., and Mishkin, M. (1997). Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys. Neuropsychologia *35*, 999–1015.

Morecraft, R.J., Geula, C., and Mesulam, M.M. (1992). Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey. J. Comp. Neurol. *323*, 341–358.

Murray, E.A., O'Doherty, J.P., and Schoenbaum, G. (2007). What we know and do not know about the functions of the orbitofrontal cortex after 20 years of cross-species studies. J. Neurosci. *27*, 8166–8169.

O'Doherty, J., Critchley, H., Deichmann, R., and Dolan, R.J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. J. Neurosci. *23*, 7931–7939.

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

Price, J.L. (2007). Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Ann. N Y Acad. Sci. *1121*, 54–71.

Rangel, A., Camerer, C., and Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. *9*, 545–556.

Rolls, E.T., Hornak, J., Wade, D., and McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. J. Neurol. Neurosurg. Psychiatry *57*, 1518–1524.

Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., and Rushworth, M.F. (2006). Separate neural pathways process different decision costs. Nat. Neurosci. *9*, 1161–1168.

Rudebeck, P.H., Behrens, T.E., Kennerley, S.W., Baxter, M.G., Buckley, M.J., Walton, M.E., and Rushworth, M.F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. J. Neurosci. *28*, 13775–13785.

Rushworth, M.F., Behrens, T.E., Rudebeck, P.H., and Walton, M.E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends Cogn. Sci. *11*, 168–176.

Saddoris, M.P., Gallagher, M., and Schoenbaum, G. (2005). Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. Neuron *46*, 321–331.

Schoenbaum, G., Nugent, S.L., Saddoris, M.P., and Setlow, B. (2002). Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. Neuroreport *13*, 885–890.

Schoenbaum, G., Setlow, B., Saddoris, M.P., and Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron *39*, 855–867.

Schoenbaum, G., Roesch, M.R., and Stalnaker, T.A. (2006). Orbitofrontal cortex, decision-making and drug addiction. Trends Neurosci. *29*, 116–124.

Schoenbaum, G., Saddoris, M.P., and Stalnaker, T.A. (2007). Reconciling the roles of orbitofrontal cortex in reversal learning and the encoding of outcome expectancies. Ann. N Y Acad. Sci. *1121*, 320–335.

Schultz, W. (2007). Multiple dopamine functions at different time courses. Annu. Rev. Neurosci. *30*, 259–288.

Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J. Neurosci. *27*, 8366–8377.

Seo, H., and Lee, D. (2008). Cortical mechanisms for reinforcement learning in competitive games. Philos. Trans. R. Soc. Lond. *363*, 3845–3857.

Seo, H., and Lee, D. (2009). Behavioral and neural changes after gains and losses of conditioned reinforcers. J. Neurosci. *29*, 3627–3641.

Seo, H., Barraclough, D.J., and Lee, D. (2007). Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. Cereb. Cortex *17* (*Suppl 1*), i110–i117.

Seo, H., Barraclough, D.J., and Lee, D. (2009). Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. J. Neurosci. *29*, 7278–7289.

Simmons, J.M., and Richmond, B.J. (2008). Dynamic changes in representations of preceding and upcoming reward in monkey orbitofrontal cortex. Cereb. Cortex *18*, 93–103.

Stalnaker, T.A., Roesch, M.R., Franz, T.M., Burke, K.A., and Schoenbaum, G. (2006). Abnormal associative encoding in orbitofrontal neurons in cocaine-experienced rats during decision-making. Eur. J. Neurosci. *24*, 2643–2653.

Stalnaker, T.A., Franz, T.M., Singh, T., and Schoenbaum, G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. Neuron *54*, 51–58.

Sutton, R.S., and Barto, A.C. (1998). Reinforcement Learning: An introduction (London: MIT Press).

Takahashi, Y.K., Roesch, M.R., Stalnaker, T.A., Haney, R.Z., Calu, D.J., Taylor, A.R., Burke, K.A., and Schoenbaum, G. (2009). The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. Neuron *62*, 269–280.

Thorndike, E.L. (1911). Animal Intelligence: Experimental Studies (New York: Macmillan).

Thorndike, E.L. (1933). A proof of the law of effect. Science *77*, 173–175.

Tremblay, L., and Schultz, W. (2000). Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. J. Neurophysiol. *83*, 1877–1885.

Tsujimoto, S., Genovesio, A., and Wise, S.P. (2009). Monkey orbitofrontal cortex encodes response choices near feedback time. J. Neurosci. *29*, 2569–2574.

Uchida, Y., Lu, X., Ohmae, S., Takahashi, T., and Kitazawa, S. (2007). Neuronal activity related to reward size and rewarded target position in primate supplementary eye field. J. Neurosci. *27*, 13750–13755.

van Duuren, E., Lankelma, J., and Pennartz, C.M. (2008). Population coding of reward magnitude in the orbitofrontal cortex of the rat. J. Neurosci. *28*, 8590–8603.

Wallis, J.D., and Miller, E.K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur. J. Neurosci. *18*, 2069–2081.

Walton, M.E., Devlin, J.T., and Rushworth, M.F. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. Nat. Neurosci. *7*, 1259–1265.

Weiskrantz, L. (1990). Problems of learning and memory: one or multiple memory systems? Philos. Trans. R. Soc. Lond. *329*, 99–108.

White, N.M. (1989). Reward or reinforcement: what's the difference? Neurosci. Biobehav. Rev. *13*, 181–186.

Wise, S.P. (2008). Forward frontal fields: phylogeny and fundamental function. Trends Neurosci. *31*, 599–608.

Yanike, M., Wirth, S., Smith, A.C., Brown, E.N., and Suzuki, W.A. (2009). Comparison of associative learning-related signals in the macaque perirhinal cortex and hippocampus. Cereb. Cortex *19*, 1064–1078.

Neuron, Volume *65*

Supplemental Information

Separable Learning Systems in the Macaque Brain and the Role of Orbitofrontal Cortex in Contingent Learning

Mark E. Walton, Timothy E.J. Behrens, Mark J. Buckley, Peter H. Rudebeck, and Matthew F.S. Rushworth

1. Supplemental Data

Supplemental Results

Figure S1 (related to Figure 4).

Figure S2

Figure S3 (related to Figure 5)

Figure S4 (related to Figure 6A)

Figure S5 (related to Figure 6B)

Figure S6

2. Supplemental Experimental Procedures

3. Supplemental References

Supplemental Data

Choice alternation as a function of local reward rate

OFC-lesioned animals demonstrated raised rates of trial-by-trial switching behavior following surgery (Figure 4). However, it is possible that this was simply a consequence of the reversal deficit causing these animals to receive less frequent rewards than they had prior to the reversal. As can be seen in Figure S1A, whereas both groups of animals pre-operatively were increasingly likely to persist with choosing an option as the local reward rate increased, OFC-lesioned monkeys post-operatively did not display this pattern of increased persistence with increasing local reward rate except when the reward rate was at its highest (≥0.7 rewards/choice)(Lesion Group x Surgery x Reward Rate: $F_{8,32}$=3.05, p=0.011). This was particularly marked in the post-reversal phase of both conditions.

Importantly, when the data were divided up by whether or not a reward was delivered immediately before a switch, the OFC-lesioned animals displayed a comparable increased propensity to alternate between choices compared to controls following either a positive or negative outcome on the previous trial (Lesion Group x Surgery x Previous Reward and Lesion Group x Surgery x Previous Reward x Reward Rate: both Fs<2.5, both p>0.14) (Figure S1B). All these effects were replicated if instead rates of switching were investigated as a function of subjective stimulus values rather than local reward rates. This demonstrates that the OFC lesion did not cause a particular problem with monitoring and responding to negative reinforcement or with inhibiting responding to the previously most highly rewarding stimulus (Fellows, 2007; Kringelbach and Rolls, 2004).

Choice alternation as a function of recent reward- and choice-histories

An integrated history of recent rewards is most predictive of the current reward in two situations: when the recent reward rate is very low (as current rewards are very unlikely), and when the recent reward rate is very high (as current rewards are very likely). Data already presented (Figure S1) depicted the monkeys' alternation behavior as a function of local reward rate. At the lowest and highest reward rates, OFC and control patterns of switching are indistinguishable. By contrast, when the local reward rate was at intermediate levels (meaning that the current trial was equally likely to be rewarded or not and could not, therefore, be predicted using the integrated history of reward), the OFC group's switching behavior deviated significantly from that of the control group.

In this vein, we also re-examined whether the OFC-lesioned animals' patterns of response alternation in STB and VRB were being influenced by their recent history of choices by plotting trial-by-trial rates of switching as a function of the number of times prior to a switch that they had selected the same option. In order to obtain sufficient data for this, we collapsed across both phases of the STB and VRB schedules. While an equivalent pattern of significantly increased response alternation was observed in the OFC-lesioned animals following sequences of 1 or 2 choices of the same option (sequence of 1 response type: Lesion Group x Surgery x Value: $F_{4,16}=7.13$, $p=0.002$; sequence of 2 response types: Lesion Group x Surgery: $F_{1,4}=10.21$, $p=0.033$), as the sequences increased to 3-5 selections of the same option, the OFC group's likelihood of persisting increased and become indistinguishable from controls (Figure S6). Therefore, as the lesioned animals' choice history became more consistent, their pattern of choices also became more similar to control animals. This again implies that OFC-lesioned animals might be using reinforcement not to update the value of the immediately

preceding chosen option but instead to revalue all the options as a function of recent choice history.

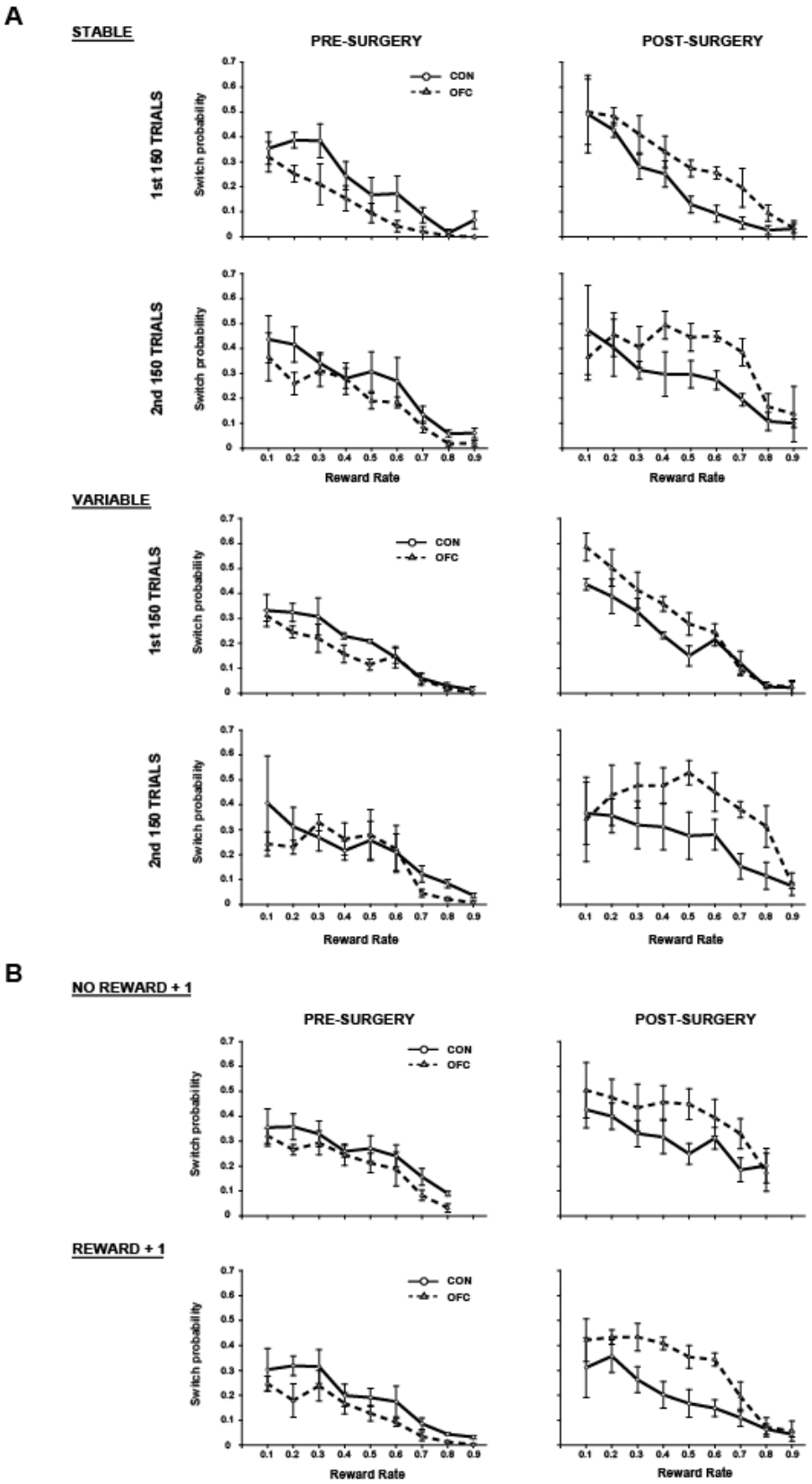
A
STABLE
PRE-SURGERY
POST-SURGERY
1st 150 TRIALS
2nd 150 TRIALS
Switch probability
CON
OFC
Reward Rate
VARIABLE
B
NO REWARD + 1
REWARD + 1

Figure S1 (related to Figure 4). Switching likelihood as a function of recent local reward rate (rewards / trial, averaged over the past 10 trials) divided up (A) by condition (STB, upper panels; VRB, lower panels), by surgery (pre-surgery, left-hand column; post-surgery, right-hand column) and by phase (1st 150 trials, pre-reversal, or 2nd 150 trials, post-reversal) and (B) by whether or not the previous trial was rewarded (no reward on previous trial, upper panels; reward on previous trial, lower panels). Controls = open circles and filled line; OFCs = gray triangle and dashed line.

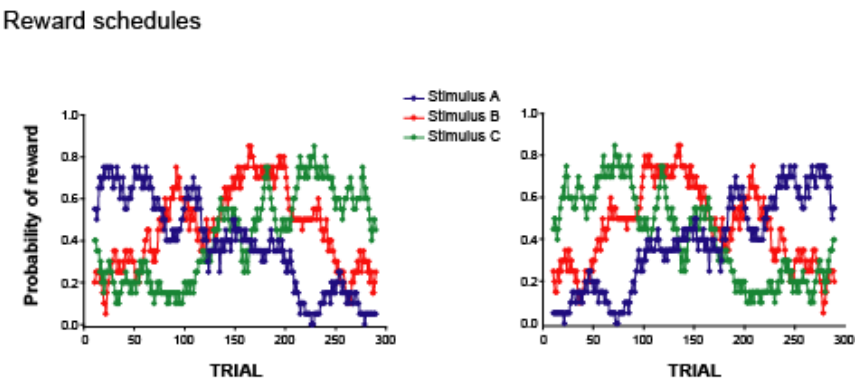

Figure S2. Predetermined reward schedules from two additional 3-armed bandit conditions (which are mirror images of each other, with, for instance, the likelihood of reward on trial 10 in the left-hand condition being identical to trial 290 in the right-hand one). Animals' choices from 5 sessions of testing on both schedules was included in the logistic regression (Figure 5) and reward-/choice-history analyses (Figure 6, S5-6) in order to provide sufficient trials to obtain adequate estimates of the effects of reward- and choice history. Choice behavior in one of the conditions (left-hand panel) has previously been reported in Rudebeck et al. (2008). As before, the schedules determined whether or not reward was delivered for selecting a stimulus (A-C) on a particular trial.

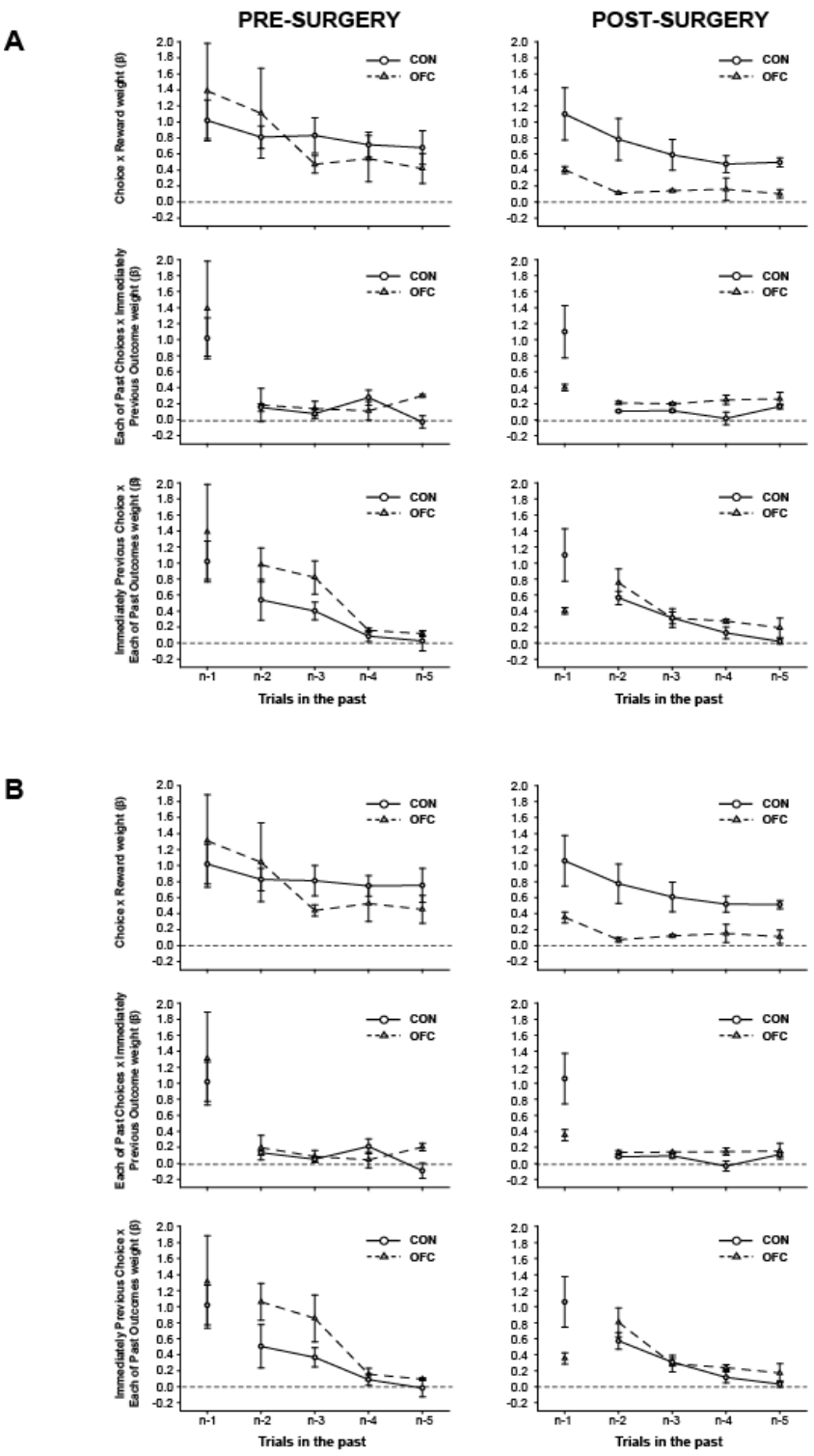

Figure S3 (related to Figure 5). Influence of recent choices and their specific outcomes, each past choice and the previous outcome, and each past outcome and the previous choice on current choice behavior as a function of (A) A choices, and (B) B or C choices only.

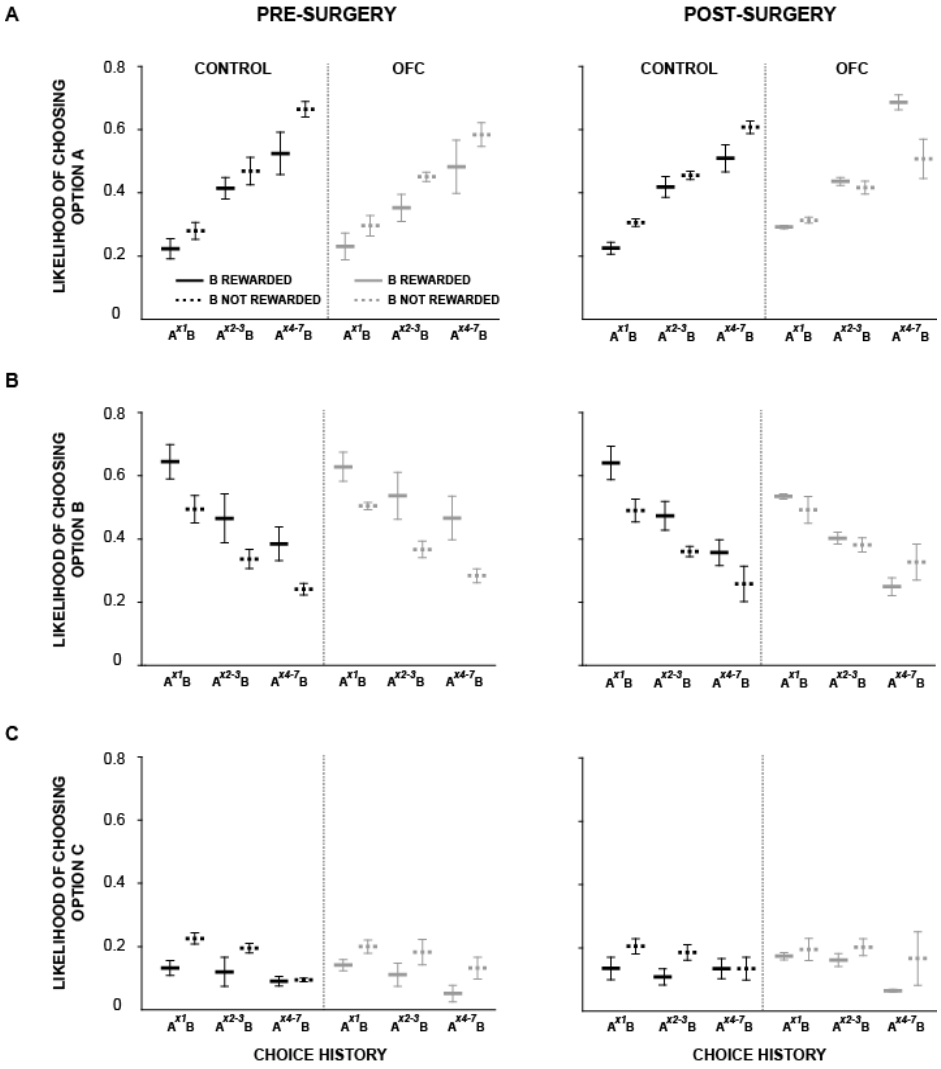

Figure S4 (related to Figure 6A). Influence of past choices of one option (A) on current choice behavior (trial n) in changeable 3-armed bandit tasks as a function of reward received for choosing option B on the previous trial (trial n-1). Note, as elsewhere, *options* A, B, and C do not necessarily refer to selection of *stimuli* A, B, and C but instead to similar patterns of choices (i.e., an "AAB" history can be made up of choices of stimulus AAB, AAC, BBC, BBA, CCB, or CCA). Top row: likelihood of choosing option A on trial n after either receiving a reward (filled line) or not receiving a reward (dashed line) for choosing option B on the previous trial (n-1). Middle row: likelihood of choosing option B on trial n. Bottom row: likelihood of choosing option C on trial n. The data in Figure 5 depicts the above data as the subtraction of (B rewarded - B not rewarded) for each choice history sequence.

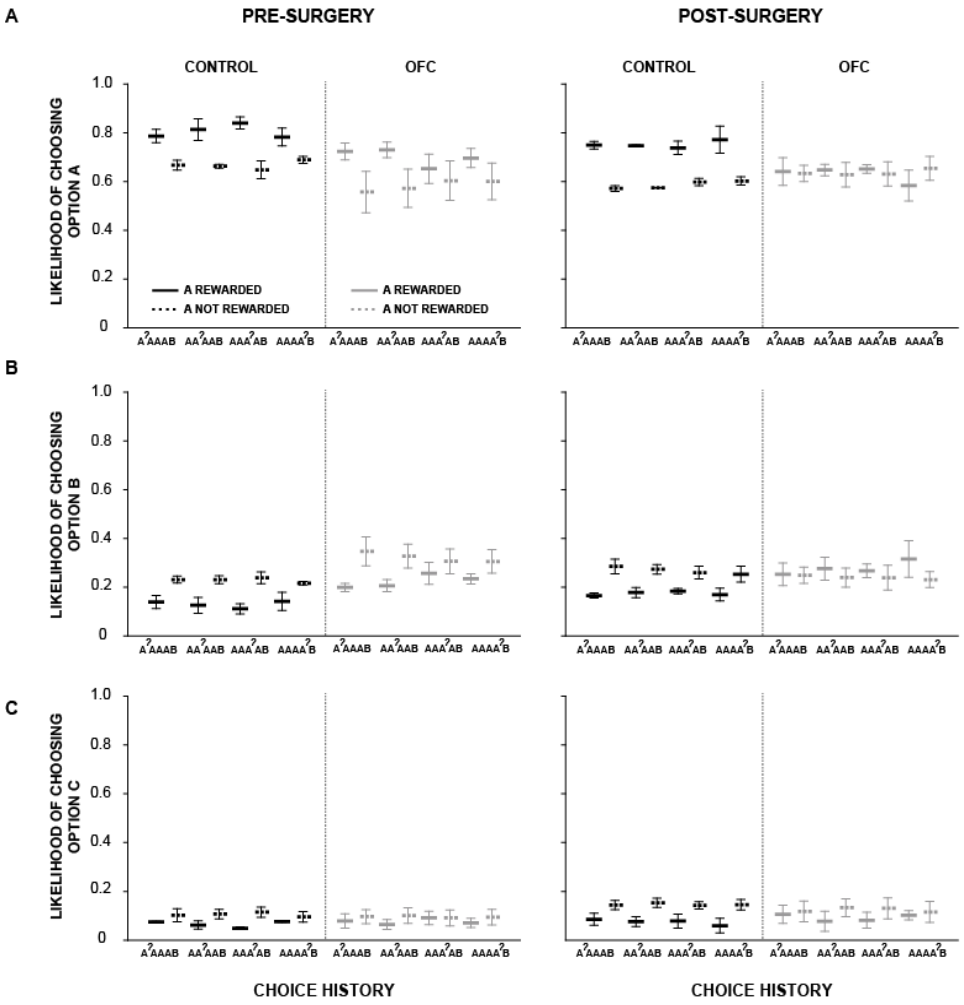

Figure S5 (related to Figure 6B). Likelihood of choosing a particular option on the current trial (n) after having chosen option A on 4 past trials (n-2 to n-5) and then option B on the previous trial (n-1), plotted as a function of reinforcement on one particular A option in the past (A?). Top row = likelihood of choosing option A on trial n when previous A choice (A?) was either rewarded (filled line) or not rewarded (dashed line); middle row = likelihood of choosing option B on trial n; bottom row = likelihood of choosing option C on trial n.

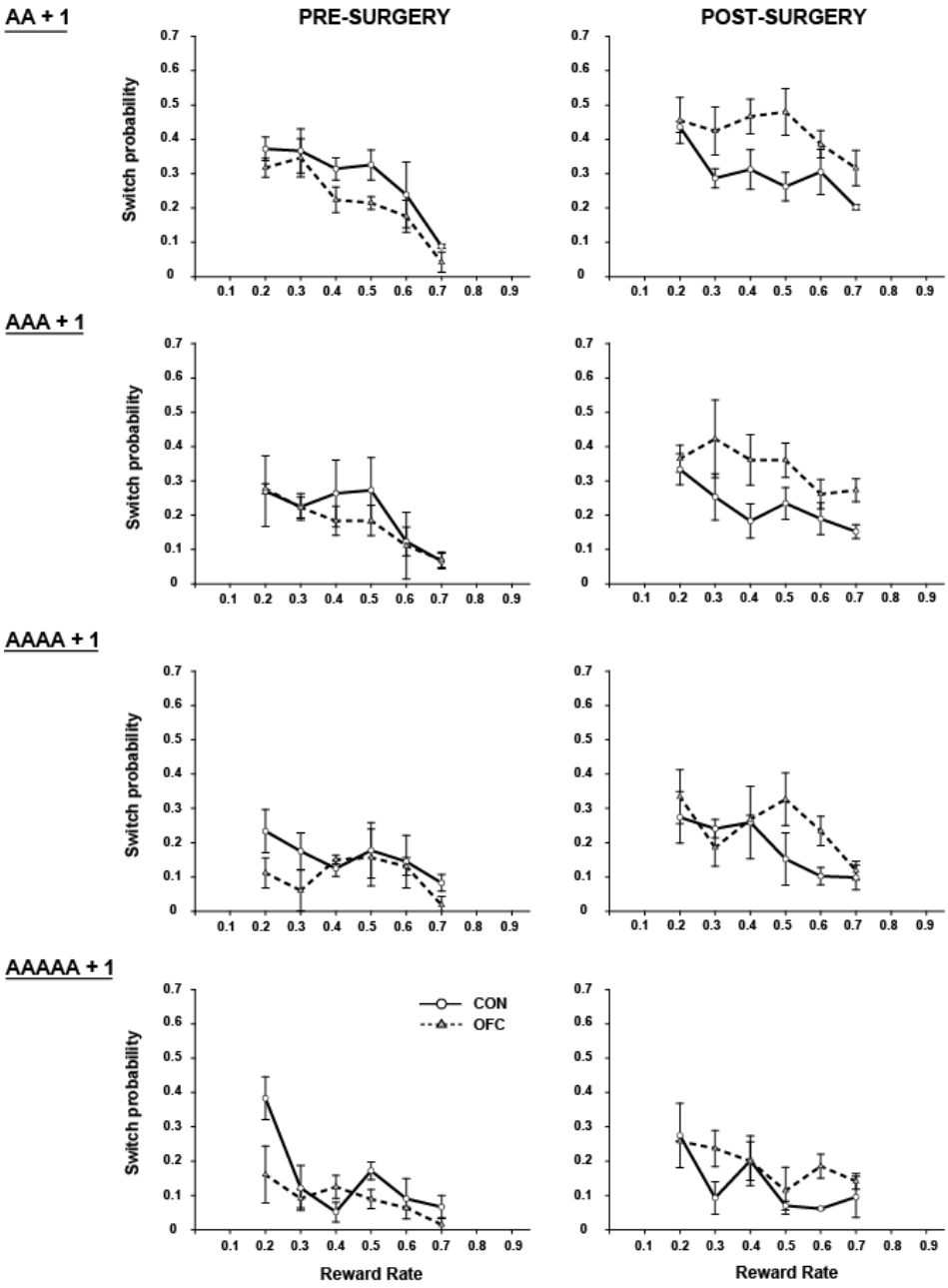

Figure S6. Switching likelihood across all trials of STB and VRB as a function of recent local reward rate (past 10 trials) divided up by uniformity of recent choice history. Top row = likelihood of switching on the trial after having made the same choice twice (AA+1); second row = likelihood of switching having made the same choice three times (AAA+1); third row = likelihood of switching having made the same choice four times (AAAA+1); bottom row = likelihood of switching having made the same choice five times (AAAAA+1). As elsewhere, "A" can refer to selection of stimulus A, B or C with the appropriate choice sequence. Controls = open circles and filled line; OFCs = gray triangle and dashed line.

Supplementary Experimental Procedures

Apparatus

Each monkey sat unrestrained in a wheeled transport cage placed 20cm from a touch-sensitive monitor (38cm wide x 28cm high) in a testing room on which visual stimuli could be presented (8 bit color clipart bitmap images, 128 x 128 pixels) and responses recorded. Rewards (190mg Noyes pellets) were delivered from a dispenser (MED Associates) into a food well immediately to the right of the touch screen. A large metal food box, situated to the left below the touch screen, contained each individual's daily food allowance (given in addition to the reward pellets) consisting of proprietary monkey food, fruit, peanuts and seeds, delivered immediately after testing each day. This was supplemented by a forage mix of seeds and grains given ~6 hours prior to testing in the home cage. Stimulus presentation, experimental contingencies, reward delivery and food box opening was controlled by a computer using in-house software.

Statistical Analyses

Where appropriate, data from STB and VRB are reported using parametric repeated-measures ANOVA, with within-subjects factors of Surgery (2 levels: Pre- or Post-Surgery), Condition (2 levels: STB or VRB), and Testing Session (5 levels: Session 1–5). Analyses of performance before and after reversal in identity of H_{sch} included the factor of Phase (2 levels: 1st or 2nd 150 trials in a session), and response alteration analyses included local reward rate – the average likelihood of reward per trial across the previous 10 trials – or subjective reward value (both 9 levels: 0.1–0.9). FIXED schedules were analyzed comparably, though without the factor of Surgery (as all testing occurred post-surgery). Performance criterion measures used geometric means of the number of trials

taken to choose the H_{sch} option on ≥65% trials over the 5 sessions to account for skew induced by days on which no criterion was reached (and so a maximum of 140 trials was logged). These were then compared with separate Mann-Whitney tests as to account for violations in normality in the data.

Logistic Regression

In order to ascertain the influence of specific choice-outcome associative learning and associations based on recent choice- and reward-histories, we performed three separate logistic regression analyses, one for each potential stimulus (A,B,C). This gave us three sets of regression weights , $\hat{\beta}_A, \hat{\beta}_B, \hat{\beta}_C$ and three sets of covariances $\hat{C}_A, \hat{C}_B, \hat{C}_C$. The regression weights into a single weight vector using a variance-weighted mean (Lindgren, 1993):

$$\hat{\beta} = \left(\hat{C}_A^{-1} + \hat{C}_B^{-1} + \hat{C}_C^{-1}\right)^{-1}\left(\hat{C}_A^{-1}\hat{\beta}_A + \hat{C}_B^{-1}\hat{\beta}_B + \hat{C}_C^{-1}\hat{\beta}_C\right)$$

However, results were essentially identical if we instead used the arithmetic mean:

$$\hat{\beta} = \frac{\left(\hat{\beta}_A + \hat{\beta}_B + \hat{\beta}_C\right)}{3}.$$

The remainder of this section will describe the analysis of only the "A" choices, and imply corollaries for B and C.

We used as the dependent variable a binary indicator variable which took the value 1 whenever the animal chose A and the value 0 whenever the animal did not choose A (i.e. when they chose B or C). We then formed independent variables (IVs) as based on all

possible combinations of recent past choices and recent past rewards (trials n-1, n-2,…,n-6)(Figure 5A). Each IV took the value 1 when, for the particular choice-outcome interaction, the animal chose A and was rewarded, the value -1 when the animal chose B or C and was rewarded, and the value 0 when there was no reward (Figure 5B). We then fit a standard logistic regression with these 36 IVs to give us estimates of $\hat{\beta}_A$ and $\hat{C}_A$.

The data depicted in Figure 5 are the influence on trials n-1 to n-5 when A was rewarded and Bs or Cs were unrewarded. However, the data were essentially unaffected when only A rewards or B,C rewards were included in the design matrix (Figure S5). As the 5th row and column is the only one in the matrix that contains variance from the choices and outcomes on trial n-5, it will therefore be sensitive to any longer-term choice/reward trends. To avoid this effect, we therefore included a 6th row/column in the matrix describing choices and outcomes n-6. These regressors were included as confound regressors for the 5th row and are therefore not shown.

Surgery and Histology

Surgical procedures in these animals have been described in detail elsewhere (Rudebeck et al., 2008). The lesions were intended to be comparable to those reported in Izquierdo et al. (2004), taking the tissue medial to the lateral orbital sulcus up to the gyrus rectus on the medial surface. The rostral and caudal boundaries were by imaginary perpendicular lines connecting, respectively, the rostral- and caudal-most points of the medial and lateral orbital sulci. Immediately following surgery and for ~5 days subsequently, animals were given non-steroidal anti-inflammatory analgesic (0.2 mg/kg meloxicam, orally) and antibiotic (8.75 mg/kg amoxicillin, orally), and were allowed at

least 3 weeks for full recovery prior to post-operative testing. Post-operative data collection for the experiments reported here started between 8-12 weeks after surgery.

Following completion of all testing, animals were deeply anesthetized with sodium pentobarbitone and perfused with 90% saline and 10% formalin, their brains removed and placed in 10% sucrose formalin until they sank. The brains were subsequently blocked in the coronal plane at the level of the most medial part of the central sulcus. Each brain was cut in 50 μm coronal sections, with every 10^{th} section retained and stained with cresyl violet for analysis of the extent of the lesion.

The extent of the OFC lesions has also been described in detail elsewhere (Rudebeck et al., 2008). In brief, the lesions were largely as intended, reliably destroying the tissue in Walker's areas 11 and 13 in all cases (Walker, 1940) (Figure 1A). On the lateral extent, area 12 was largely spared except for part of this region in the left hemisphere of one animal. The lesion was more variable in the extent to which area 14 on the medial surface was damaged, with anterior medial sections largely spared along with posterior parts of the gyrus rectus.

Supplemental References

Lindgren, B.W. (1993). Statistical Theory, 4th Edition edn (New York: Chapman & Hall).

Walker, A.E. (1940). A cytoarchitectural study of the prefrontal area of the macaque monkey. J Comp Neurol *73*, 59-86.