Published in final edited form as:

Brain Res. 2008 August 28; 1227: 110–119. doi:10.1016/j.brainres.2008.06.033.

Distinguishing expected negative outcomes from preparatory control in the human orbitofrontal cortex

Stefan Ursu¹, Kristi A. Clark², V. Andrew Stenger³, and Cameron S. Carter¹

¹Department of Psychiatry, University of California-Davis, Sacramento, California 95817

²Department of Neurology, 1506D Gonda Center, UCLA, Los Angeles, California 90095

³Department of Medicine, University of Hawaii, Honolulu, Hawaii 96813

Abstract

The human orbitofrontal cortex (OFC) plays a critical role in adapting behavior according to the context provided by expected outcomes of actions. However, several aspects of this function are still poorly understood. In particular, it is unclear to what degree any subdivisions of the OFC are specifically engaged when negatively valenced outcomes are expected, and to what extent such areas might be involved in preparatory active control of behavior. We examined these issues in two complementary functional magnetic resonance imaging (fMRI) studies in which we simultaneously and independently manipulated monetary incentives for correct performance, and demands for active preparation of cognitive control. In both experiments, preparation for performance was associated with lateral PFC activity in response to high incentives, regardless of their valence, as well as in response to increased task demands. In contrast, areas of the OFC centered around the lateral orbital sulcus responded maximally to negatively perceived prospects, even when such prospects were associated with decreases in preparatory cognitive control. These results provide direct support for theoretical models which posit that the OFC contributes to behavioral regulation by representing the value of anticipated outcomes, but does not implement active control aimed at avoiding or pursuing outcomes. Furthermore, they provide additional converging evidence that the lateral OFC is involved in representing specifically the affective impact of anticipated negative outcomes.

Keywords

reward; cognitive control; prefrontal cortex; inhibition; incentives

1. Introduction

While it is widely accepted that the human prefrontal cortex is critically involved in executive control functions (Fuster, 1997), the principles governing these functions and the sub-specialization of various prefrontal areas are still a matter of intense debate (Miller and Cohen, 2001). As a subdivision of the prefrontal cortex (PFC), the orbitofrontal cortex (OFC) has specific cytoarchitecture and connectivity and has been shown to be critically involved in goal-directed behavior (Rolls, 1998). Furthermore dysfunction of this cortical region has been

Corresponding author: Stefan Ursu, M.D., Ph.D., UC Davis Imaging Research Center, 4701 X St., Sacramento, CA 95817, Tel: (916) 734 4149, Fax: (916) 734 8750, E-mail: stefan.ursu@ucdmc.ucdavis.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

implicated in several neuropsychiatric disorders, including anxiety (Rauch et al., 1997), major depression (Liotti et al., 2002) and drug addiction (Goldstein and Volkow, 2002; London et al., 2004). Substantial evidence from both non-human primate (Dias et al., 1996; Roesch and Olson, 2004; Tremblay and Schultz, 2000a) and human studies (Bechara et al., 1998; Berlin et al., 2004; O'Doherty et al., 2001) suggests that the OFC is critical for adaptive behavior associated with changes in the outcomes of actions, such as the quality, magnitude and probability of rewards and punishments experienced during performance of cognitive tasks. However, the precise mechanisms through which motivational processes mediated by the OFC contribute to behavioral control remain the subject of active research (see Kringelbach and Rolls (2004) for a review).

In particular, two fundamental questions with regards to the OFC function are outstanding. One is the degree of regional specificity for positive vs. negative outcomes, particularly during the preparation/expectation phase of task performance. The second is the relationship between representations of expected outcomes and task performance. That is, to what extent are these representations involved in the active modification of behavior through implementation of active, top-down cognitive control processes? In most human studies conducted to date, the evaluation of outcome predictors is rapidly followed by a response, making difficult to definitively separate the contribution of outcome expectation from feedback processing (O'Doherty et al., 2001; O'Doherty et al., 2003). Many of these studies focused on indices of OFC activity in response to rewards (Berns et al., 2001; Dreher et al., 2006; Elliott et al., 2003; Gottfried et al., 2003; O'Doherty et al., 2002; Rogers et al., 1999), while others showed that foci in the OFC react specifically to probabilistic expectation and/or presentation of aversive stimuli (Breiter et al., 2001; O'Doherty et al., 2001; Seymour et al., 2005). While these findings provide overall support for the existence of OFC representations subserving avoidance behaviors (Beer et al., 2006), they were generated using paradigms which did not explicitly make differential predictions if the OFC would be involved in engagement of active control aimed at avoiding outcomes with negative affective impact. Therefore, it is still unclear to what degree the OFC is involved in representing affective impact of outcomes per se, versus the engagement of cognitive functions necessary to pursue or avoid certain outcomes.

Indeed, changes in expected outcomes (e.g. potential rewards or punishments) frequently correlate with changes in efficacy of top-down control processes, either because subjects engage control in order to obtain rewards or avoid the punishments, or simply because oftentimes conditions leading to rewards or punishments carry increased demands for controlled performance (Maunsell, 2004). In order to uncouple the tight association between outcome representations and implementation of control, we used fMRI in a slow event-related design which independently manipulated monetary incentives for correct performance and the demands for cognitive control in the preparatory phase of a task requiring various levels of suppression of prepotent responses. Specific task parameters (such as response speed pressure) were manipulated to create experimental conditions in which the two hypothesized OFC-related mechanisms (implementation of preparatory cognitive control vs. representations of expected outcomes) predicted distinct patterns of activity in the OFC.

2. Results

Experiment 1 - Behavioral results

In this version of the task, subjects committed errors in approximately 10% of trials, and error rates were not differentially affected by either task [$F(1,18) = 0.97$, $p = 0.34$] or incentive level [$F(1,18) = 0.20$, $p = 0.66$]. An ANOVA of the mean reaction times in correct trials confirmed the two effects of interest for our hypotheses. Collapsing across levels of Incentives, the HC task was associated with increased demands for control relative to LC (simple main effect of task, [$F(1,18) = 23.34$, $p < 0.001$], with longer latencies in the HC trials, see Figure 2a).

Conversely, the simple main effect of incentives was also significant, confirming that subjects engagement of cognitive control was more effective in reward trials (reaction times were faster for reward trials than for non-rewarded ones [$F(1,18) = 11.50, p = 0.003$]). There was no interaction between incentives and task [$F(1,18) = 0.01, p = 0.92$].

Experiment 1 - Functional imaging results

Given the focus of these studies on the expectancy phase of task performance, most analyses were conducted on the 8 scans covering the preparatory sample-target interval (scans S1–S8). However, some relevant confirmatory tests (see *Task effects* and Discussion sections) were also conducted on the 8 scans which sampled the post-target delay (i.e. post-response, scans S9–S16). All analyses were conducted as analyses of variance of the BOLD signal over time, and therefore did not use assumptions regarding the shape of the BOLD response. Activations were discussed if statistical maps were significant at a threshold of $p < 0.05$ corrected for multiple comparisons, and the directionality of effects in each region of interest discussed was verified, aside from direct inspection of the timeseries, through a planned t test (see Methods). As mentioned before, the close relationship between expected outcomes and cognitive control makes examination of either incentive or task effects (Incentive by Scan or Task by Scan interaction, respectively) relevant for examining the neural substrates of both outcome representation and cognitive control. Nevertheless, distinguishing between the two types of processes is possible by examining the relationship between brain activation in various experimental conditions and specific predictions made by each mechanism in those particular conditions. Since one of the aims of the study was to distinguish representations of outcomes from associated top-down cognitive control mechanisms, we focused on identifying activation foci in the OFC and in more dorsal areas of the PFC, since the latter are thought to be specifically involved in implementation of active control processes. The comprehensive list of all forebrain regions of activations identified by our analyses are provided as supplementary material in tables S1 and S2.

Incentive effects—In this experiment, the lateral orbitofrontal cortex showed more activity during preparation for Non-reward than for Reward trials (left BA 11/47, see Figure 3A, bottom graph). In contrast, in more superior aspects of the PFC activity was higher after Reward than after No-Reward cues (e.g. left BA8/9, Figure 3A, top graph).

Task effects—Increased activation to HC relative to LC trials was noted in lateral aspects of the PFC (BA10/46 and 10/47, see Figure 3B). In contrast, in this experiment the HC trials did not increase the activity in the orbital PFC. To ensure that this lack of OFC activity in response to manipulation of demands for control was not a false negative result, we performed a series of additional analyses. Firstly, we relaxed the threshold used for examining the statistical maps and verified that no orbitofrontal areas showed increased activity during HC trials. Secondly, after identifying the left OFC cluster of voxels sensitive to changes in incentives (see Incentive effects above), we also performed a confirmatory Task $\times$ Incentive $\times$ Scan ANOVA on the signal change of those voxels and found that the three-way interaction, the simple main effect of Task and the Task $\times$ Scan interaction were all highly non-significant. Thirdly, we performed an omnibus test of the Task $\times$ Incentive $\times$ Scan interaction for the preparatory phase, and confirmed that no areas of activation were identified in the OFC. Finally, we sought to confirm that the behavioral task placed significant demands on cognitive inhibition processes at the time of responses to targets. For this, we verified that an omnibus test of the Task $\times$ Scan interaction following the probe presentation (scans 9–16) identified cortical areas previously involved in suppression of interference from competing response sets (Bunge et al., 2002; Konishi et al., 1998) and in response inhibition (Aron et al., 2004; Bunge et al., 2002; Waldvogel et al., 2000). Such areas showing higher activity after HC relative to

LC targets included inferior lateral prefrontal, motor and premotor cortices (see table S1 for details).

Experiment 2 – Results

The results of the study described above were consistent with the hypothesis that the lateral OFC represents the expected affective impact of the least favorable outcomes, as suggested by the increased activity to No-reward trials. This interpretation is consistent with behavioral economic models which point out that affective value of outcomes is determined relative to the range of possible outcomes of a given decision or action (Mellers et al., 1997; Ursu and Carter, 2005). In order to obtain additional, direct support for the specific involvement of the lateral OFC in representing expected negative outcomes, we conducted a second study which added a negative incentive condition (Penalty). In the Penalty condition, subjects performed in order to avoid losing a small amount of their winnings as punishment for responding incorrectly or too slowly to the probe (see Methods).

Experiment 2 - Behavioral results

In Experiment 2, Task $\times$ Incentive ANOVAs of mean reaction times and error rates showed that, as in Experiment 1, performance was improved by the presence of incentives (Figure 2b). The effect was apparent in latency of correct trials [$F(2,16) = 14.78$, $p < 0.001$], as well as in accuracy [$F(2,16) = 3.62$, $p = 0.04$]. These measures showed equal improvement from positive and negative incentives, relative to neutral trials (t-tests of Penalty vs. Reward trials yielded p values of 0.3 and 0.5 respectively). The simple main effect of task was also significant both in reaction times (slower in HC trials, $F(1,16) = 17.72$, $p = 0.001$) and overall accuracy (higher error rates in HC trials, $F(1,16) = 13.63$, $p = 0.002$). Similar to Experiment 1, there was no interaction between Incentives and Task in either latency or accuracy of responding (all p values greater than 0.35).

Experiment 2 - Functional imaging results

Incentive effects—As can be seen in Figure 4A, the incentive analysis of the preparatory interval isolated a left lateral OFC set of voxels (BA 11/47, bottom graph). The signal change in this area was strongest during Penalty trials relative to either Reward or No-Reward trials. In contrast, the activity in dorsolateral regions of the PFC such as BA9/8 was equally increased during preparation for either Reward or Penalty trials relative to No-Reward trials (Figure 4A, top graph), while in medial aspects of the PFC was strongest for Reward trials (table S2).

Task effects—The Task by Scan interaction revealed that preparation for HC trials was associated with higher activity relative to LC trials in both ventral and dorsal PFC (Figure 4B). Specifically, as can be seen in the top graph of Figure 4B, HC preparatory activity was increased in the left dorsolateral PFC, including Brodmann's areas 45, 9 and 46. This omnibus contrast also isolated a left lateral OFC cluster of HC related activity (BA 47/11, see Figure 4B). This HC-related activity increase in the OFC did not reach significance in the *a priori* test of the peak voxel. However, when we conducted the analogous test on the signal changes averaged across all voxels of the cluster, the increase in the HC trials was significant [$t(16) = 5.2$, $p < 0.04$].

Since the trend for task effects in the preparatory OFC activity described above could be due to either task-related outcome representations or to task-related implementation of control we conducted additional analyses to distinguish between the two mechanisms. Firstly, we conducted a three-way ANOVA of Task by Incentive by Scan in the preparatory interval. This analysis did not isolate any activation areas in the OFC. Secondly, we also tested the Task $\times$ Scan interaction after correct responses to targets (to avoid feedback-related effects). In this analysis, the most ventral focus of activation was located in the left inferior frontal gyrus (BA

47), but no areas of activation were found in the OFC (see table S2 for details). Thus, there was a consistent failure to isolate OFC activity in contrasts in which the influence of anticipated or experienced negative feedback was eliminated.

3. Discussion

In both experiments presented above the fMRI results were consistent with the hypothesis that lateral OFC areas (centered around the lateral orbital sulcus) activated most strongly under expectation of the potential outcome associated with negative affect, regardless of whether this expectation was incentive-induced (both experiments) or task-dependent (the frequent negative outcomes associated with a more demanding task, as in experiment 2). Furthermore, in Experiment 1 OFC activity to incentive-induced negative outcomes manifested as greater activity for unrewarded trials than rewarded ones. This suggested that the lateral OFC represents negative outcomes in a relative manner (i.e. the least favorable outcome available, since No-Reward trials were, in this experiment, missed opportunities for rewards). Importantly, this result is inconsistent with direct OFC involvement in control processes, as the No-Reward trials were associated with poorer performance (slower responses) than rewarded ones. Furthermore, the results of the analysis by task were also inconsistent with direct OFC involvement in control processes, as its activity was not increased in the high control condition when error rates (i.e. task-dependent outcomes) were matched across the high and low control tasks.

In Experiment 2, the effects of incentive-induced potential negative outcomes on the lateral OFC were confirmed by identifying lateral OFC activity that was more sustained during preparation for Penalty trials relative to Reward or No-Reward. Since both behavioral and imaging results suggested similar engagement of control processes during Penalty and Reward trials, the Penalty-specific OFC activity is inconsistent with incentive-induced control being supported by the OFC. Finally, the analysis of task effects in Experiment 2 revealed that OFC activity was greater during high control trials, which were also more frequently associated with negative outcomes (errors). This replicated many existing results from experimental designs in which increased demands for control coincide with expectations of poorer outcomes:

In sum, these complementary experiments help answer an important outstanding question regarding OFC function in behavioral regulation. Previous results have indirectly suggested a specific OFC involvement in outcome representation by isolating OFC activation in conditions which do not explicitly require active control mechanisms (Kim et al., 2006; Seymour et al., 2005). These findings are extended by our results from Experiment 1 in which the lateral OFC activity ran counter to the pattern predicted by involvement in implementation of active control. Thus, our results provide direct evidence against lateral OFC involvement in active cognitive control per se, and support theories involving the OFC in online representations of rewards and punishments (Frank and Claus, 2006; Rolls, 2000). Furthermore, in the four analyses presented here (incentive-related and task-related, in each experiment), OFC activity consistently followed the predictions generated by the expected negative outcomes hypothesis. In contrast, increased preparatory cognitive control was associated with lateral OFC activity only when the experimental manipulation also led to higher error rates, and thus to higher frequency of missed rewards or losses. Therefore, they also suggest that previous results of OFC involvement in various conditions requiring behavioral control or inhibition may actually reflect task-related outcome representations, such as the likelihood of negatively valenced affect associated with errors or other negative outcomes.

Such a framework is in agreement with many previous studies investigating the role of various prefrontal subdivisions in executive control. For instance, several studies have suggested that patients with damage of the OFC are impaired in processing rewards and penalties. The

behavioral deficits seen in these studies are consistent with a lack of motivational impact from possible negative consequences (Bechara et al., 1994; Bechara et al., 2000; Camille et al., 2004; Rolls et al., 1994). Furthermore, electrophysiology and lesion experiments in non-human primates showed that the OFC is involved in stimulus-reward associations and their reversals (Clarke et al., 2004; Rolls et al., 1996), and that in the PFC there is a dorsal-ventral decreasing gradient in the proportion neurons that maintain rule- versus goal-relevant representations (Hikosaka and Watanabe, 2000; Wallis and Miller, 2003; Watanabe et al., 2002). One would expect such processes to be involved in computations whose net result was integration of reward or punishment-related information into a higher order function of representing the potential outcomes of actions. Similarly, non-human primate studies have generated mixed results: while some OFC neurons increase their firing in response to reward-predicting stimuli (Roesch and Olson, 2004; Rolls et al., 1996; Tremblay and Schultz, 1999; Tremblay and Schultz, 2000b), others respond specifically to non-preferred foods (Hikosaka and Watanabe, 2000) or to stimuli predicting lack of reward (Rolls et al., 1996; Tremblay and Schultz, 2000a), and no topographic distributions of the two types of neurons have been described. On the other hand, the OFC has been found to be critical in conditions when demands for various forms of behavioral inhibition are high (Casey et al., 1997; Elliott and Dolan, 1999; Rolls, 1998). Our results help integrate these findings by suggesting that the lateral OFC contributes to the motivational context for behavioral inhibition through representations of expected affective impact of outcomes, such as those resulting from inappropriate response tendencies or changes in reinforcer contingencies. Coupled with the patterns of activity noted in the rest of the PFC, these results also suggest that the active implementation of cognitive control necessary to overcome prepotent response tendencies seems to be limited to the lateral aspects of the PFC (Garavan et al., 2002; MacDonald et al., 2000).

An issue related to the role of the OFC in representing negative outcomes is the degree of segregation of the neural circuits subserving approach versus avoidance-related behaviors. Previous human neuroimaging studies have produced evidence consistent with the presence of such a dichotomy within the OFC at the stage of response/feedback of task performance (O'Doherty et al., 2001; Remijnse et al., ; Small et al., 2001) or in conditions of anticipation of negatively valenced stimuli which did not explicitly preclude engagement of preparatory control processes (Seymour et al., 2005). In both of our studies, the pattern of activity in the lateral OFC during preparation suggested specificity to anticipation of negative outcomes. In contrast, responses to anticipation of rewards appeared stronger in areas spanning from the medial OFC (gyrus rectus) to the medial wall (see Tables S1 and S2). This pattern of results is consistent with the presumed bias of various prefrontal regions for processing of reward vs. punishment-related information (Remijnse et al., 2005). Further research should be focused on precisely characterizing the neural substrates of approach and avoidance behaviors, especially since they have the potential to refine models of neuropsychiatric disorders. For instance, our findings of increased OFC activity in response to potential negative outcomes provide a potential basis for interpreting a large body of literature which implicates the OFC in the pathogenesis of anxiety disorders (Grachev and Apkarian, 2000; Saxena et al., 1998), for which excessive concerns with future negative consequences of actions are a hallmark. At the same time, they may prove important for refining existing models of addictive behaviors in which the OFC features a prominent role (Goldstein and Volkow, 2002; London et al., 2000).

An intriguing finding of these studies was that the patterns of Reward-related activity in the lateral OFC changed between the two experiments. Interestingly, these differences in activity during anticipation of rewards were consistent with predictions made by recent theories of decision-making (Mellers et al., 1997), which are described in detail elsewhere (Ursu and Carter, 2005). Briefly, these theories (Mellers et al., 1997; Mellers, 2001) posit that alternative outcomes play an important role in shaping the expected affective impact of outcomes through expectations of disappointment or regret (Breiter et al., 2001; Camille et al., 2004) when

missing an opportunity for reward is likely. This characteristic of anticipated outcome representations may help further resolve apparent discrepancies in the literature. For instance, it could account for OFC activations noted during the preparatory interval of trials in which rewards are delivered according to probabilistic rules and thus carry significant “risk” of experiencing negative affect in response to missing expected rewards (Breiter et al., 2001; Dreher et al., 2006; Kim et al., 2006; Ramnani and Miall, 2003) or in which the alternative outcome of a trial is a loss (Taylor et al., 2004). It is also possible that differences across experiments in Reward-related activity may have resulted, at least in part, from any differences in the neural substrates of approach vs. avoidance-related motivation. It has been recently shown that, in certain situations, the “general” motivational context of a task (i.e. the “situational focus” of approach vs. avoidance) may interact with the “local” motivational context (i.e. the “reward structure” of individual trials, such as gains for good performance or losses for poor performance) and modulate behavioral performance (Maddox et al., 2006). In both our experiments subjects performed with the goal of gaining a monetary bonus, though it is reasonable to assume that the addition of prospects of losses (Penalty trials) in Experiment 2 may have affected this general motivational context. Unfortunately, the lack of behavioral differences between Reward and Penalty trials prevents a targeted analysis of the interactions between “situational focus” and “reward structure” of each trial type, and how they may have influenced the representations of outcome in the lateral OFC. Thus, future studies will have to address these questions with more sensitive manipulations of the general vs. local motivational context, and systematic examinations of outcome-related activity in the OFC.

A possible alternative account for the current findings is that the OFC activation noted in our two studies could be a correlate of other types of control processes that have been previously linked to the OFC, such as active reversals in affective associations with specific stimuli (Clarke et al., 2004; Dias et al., 1996; Fellows and Farah, 2003). However in our design, for any given subject, the colors predicting each type of incentive did not change. Furthermore, since switching to a task more frequently associated with punishments (i.e. HC condition) may be perceived as an affective reversal, such reversals were virtually eliminated by the blocked execution of the two tasks. Thus, it is unlikely that such processes played a significant role in our experiments. Nevertheless, this issue remains to be addressed more directly in future studies.

Finally, it is important to note that in addition to DLPFC task-related activation in the inferior convexity (Talairach coordinates 47, 41, -7, corresponding to BA47 or Walker’s lateral area 12 in non-human primates, see Figure 3A) was observed in both experiments. This is not surprising, as the cytoarchitecture and connectivity of BA 47 are markedly different from that of the orbital cortical surface and more similar to that of the dorsolateral PFC (Barbas, 1992; Cavada et al., 2000). For example, the connectivity of the inferior convexity is more heavily biased towards premotor cortices (Barbas, 1992) whereas the OFC has relatively more connections with basal forebrain and limbic structures (Ghashghaei and Barbas, 2001). This suggests that areas of the inferior convexity shown to serve goal selection functions (Arana et al., 2003; O’Doherty et al., 2003) are likely to part of the same functional complex as the DLPFC and highlight the importance of distinguishing lateral convexity from OFC in future human brain mapping studies of incentive processing and decision making. It is possible that some of the conflicting findings in the literature reflect a lack of a distinction, in imaging or lesion studies, between the closely juxtaposed lateral convexity and OFC proper.

In conclusion, the results of these two complementary fMRI studies suggest that the OFC contributes to goal-directed behavior through the representation of anticipated affective impact of outcomes, rather than through top-down control of behavior, which appears to be limited to more dorsal aspects of the PFC. Comparing the anticipatory incentive-related activity across the two studies provided evidence that the lateral aspects of the OFC specifically represent

potential negative outcomes of actions, even when such representations result from anticipated missed opportunities for reward. These findings should help to constrain hypotheses of future studies investigating the neural substrates of motivational processes, and to advance models of neuropsychiatric disorders in which these processes are thought to be disturbed.

4. Experimental Procedure

Thirty-six adult right-handed subjects participated in two separate fMRI experiments (19 and 17, of whom 9 and 8 were female, respectively), after providing written informed consent. Subjects were screened for history of neurological or psychiatric disorders. The age range of the two groups was 18–35 and 18–38, respectively. The experimental protocols were approved by the university Institutional Review Board.

Experiment 1

We used a modified delayed discrimination task in which subjects were cued at the beginning of each trial as to the type of monetary incentive that could be expected for speeded correct responding. The stimuli used were four variants of a star obtained using the Wingding font available in MS Word software, presented (in each trial) in one of two possible colors, which coded for various levels of monetary incentives (see Figure 1). Each trial consisted of 1 second presentation of a sample stimulus (a single star), an 11 second sample-target interval, and a 1 second presentation of the target stimulus (two stars presented side by side, which subtended approximately 5 degrees of visual angle). In the target stimulus, the location of the sample star (left or right) was randomized from trial to trial, and the subjects' task was to respond either in the direction or opposite to the spatial bias established by the location of the sample star (see "task demands manipulation" below). Immediately after the offset of the target the performance feedback was presented for 0.5 seconds, followed by 10.5 seconds of inter-trial interval.

Manipulation of control demands—At the beginning of each block of 24 trials, subjects were instructed to use the index fingers to respond to targets according to one of two rules (i.e. tasks). One task required them to identify the location of the sample star in the target stimulus and respond with the index finger of the hand on the same side of the sample. This constituted the low cognitive control condition (LC) since responding in the direction of the spatial bias established by the location of the sample star in the target requires little cognitive control. The other task required subjects to identify the location of the sample star and respond with the index finger of the hand on the opposite side of the sample. Thus, this trial type required more cognitive control (high cognitive control condition, HC) in order to overcome the tendency to respond on the same side with the sample stimulus (Kornblum, 1994; Simon, 1969).

Manipulation of incentives—Stimuli were displayed (randomly from trial to trial) in one of two colors, coding for Reward (\$0.5 reward for a correct response) or No-Reward. Subjects were also told that they had only one second to respond to the target, and that late responses would count as errors. The order of the two types of tasks (2 LC blocks and 2 HC blocks) was pseudo-randomized across subjects. Immediately preceding scanning, all subjects practiced 12 trials of each task with a 5 second inter-stimulus interval. The mapping of the colors onto incentive levels was counterbalanced across subjects.

Experiment 2

The task structure was similar to Experiment 1, except that responses were performed with the index and middle fingers of the right hand and that in each trial stimuli could be presented in one of three colors (coding for Reward trials, i.e. \$1 gain for a correct response, Penalty, i.e. no reward would be won for correct responses, but incorrect or slow responses would result in a \$1 loss from the winnings, and No-Reward, i.e. no monetary outcome). Subjects were

instructed that the response deadline was brief and therefore they should be as fast and as accurate as possible in order to maximize winnings, but that negative total scores will not result in actual losses. In order to eliminate potential differential speed-accuracy trade-offs in the Reward vs. Penalty conditions, unbeknownst to the subjects, the initial response window (750ms) was shortened by 10% after streaks of 5 correct and fast responses to incentive trials. Conversely, two consecutive errors or no responses in incentive trials increased the response deadline by 10%, as did every late but correct response. Subjects received 4 blocks of practice trials (with 5 seconds inter-stimulus interval) immediately previously to scanning. During scanning, three blocks of each task (HC and LC) were administered, in pseudorandomized order. All subjects except one ended the task with positive scores and won an average of \$14.68 (range -5 – 31, SD 9.1).

fMRI methods—Scanning was performed using a 3T GE Signa scanner with a standard head coil. Functional scans were acquired in the same location as the immediately preceding set of structural images: thirty-four 3.2 mm thick slices (aligned to the AC-PC plane in Exp 1) and twenty-eight 5mm slices (coronal, perpendicular to the AC-PC in Exp 2). Images were acquired using a “reversed” spiral scanning protocol (TR = 1500ms, TE = 25ms, flip = 70°, FOV = 200mm), which substantially improved the signal-to-noise-ratio in areas of the brain susceptible to inhomogeneity artifacts (Glover and Law, 2001).

Functional data were movement corrected, and post-correction parameters revealed mean scan-to-scan movement of less than 1mm in all subjects (Woods et al., 1998a). In the timeseries of each voxel, linear trends within each scanning run (i.e. block of trials) were removed. Data were then co-registered to a common reference structural scan (the Montreal Neurological Institute reference brain) using a 2nd order (60 parameters) non-linear warping algorithm (Woods et al., 1998b). Finally, data were smoothed with a 6mm FWHM three-dimensional Gaussian filter.

Since existing data suggest that the shape of the hemodynamic response function (HRF) in the OFC may be significantly more variable than in other cortical areas (O'Doherty et al., 2001; Ursu and Carter, 2005), omnibus statistical analysis was carried out using a mixed effects (subject as a random factor) ANOVA model (NIS software, Laboratory for Clinical Cognitive Neuroscience, University of Pittsburgh, and the Neuroscience of Cognitive Control Laboratory, Princeton University), which does not require assumptions regarding the shape of the hemodynamic response function (MacDonald et al., 2000; Ursu et al., 2003). These analyses used MR signal intensity as dependent variable and scan, incentive level or task as within subject, repeated measures factors. In the resultant F maps, prefrontal clusters of active voxels were identified based on voxel-wise alpha values and cluster thresholds determined through Monte Carlo simulations of fMRI data (Forman et al., 1995). After estimation of the spatial autocorrelation of the statistical maps, these criteria ensured an image-wise protection against type I error of 0.05 in each experiment. The directionality of condition differences in voxel clusters with significant Condition $\times$ Time interaction was confirmed by conducting one planned paired t test at the peak signal change difference between conditions ($p < 0.05$, two tailed). All imaging analyses included only trials resulting in correct responses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Authors thank Christopher May, Vincent van Veen and Susan Ravizza for helpful input and Kate Fissell for continuous technical support. This work was supported by an Independent Scientist Career Development Award (MH64190, C.S.C.) and a Burrows-Wellcome Translational Clinical Scientist award (#1002274, C.S.C.).

References

- Arana FS, Parkinson JA, Hinton E, Holland AJ, Owen AM, Roberts AC. Dissociable Contributions of the Human Amygdala and Orbitofrontal Cortex to Incentive Motivation and Goal Selection. *J. Neurosci* 2003;23:9632–9638. [PubMed: 14573543]
- Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex. *Trends Cogn Sci* 2004;8:170–177. [PubMed: 15050513]
- Barbas H. Architecture and cortical connections of the prefrontal cortex in the rhesus monkey. *Adv Neurol* 1992;5791–5115.
- Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 1994;50:7–15. [PubMed: 8039375]
- Bechara A, Damasio H, Tranel D, Anderson SW. Dissociation Of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience* 1998;18:428–437. [PubMed: 9412519]
- Bechara A, Tranel D, Damasio H. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain* 2000;123:2189–2202. [PubMed: 11050020]
- Beer JS, John OP, Scabini D, Knight RT. Orbitofrontal Cortex and Social Behavior: Integrating Self-monitoring and Emotion-Cognition Interactions. *J. Cogn. Neurosci* 2006;18:871–871. [PubMed: 16839295]
- Berlin HA, Rolls ET, Kischka U. Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. *Brain* 2004;127:1108–1126. [PubMed: 14985269]
- Berns GS, McClure SM, Pagnoni G, Montague PR. Predictability modulates human brain response to reward. *J Neurosci* 2001;21:2793–2798. [PubMed: 11306631]
- Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 2001;30:619–639. [PubMed: 11395019]
- Bunge SA, Dudukovic NM, Thomason ME, Vaidya CJ, Gabrieli JD. Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. *Neuron* 2002;33:301–311. [PubMed: 11804576]
- Camille N, Coricelli G, Sallet J, Pradat-Diehl P, Duhamel J-R, Sirigu A. The Involvement of the Orbitofrontal Cortex in the Experience of Regret. *Science* 2004;304:1167–1170. [PubMed: 15155951]
- Casey BJ, Trainor RJ, Orendi JL, Schubert AB, Nystrom LE, Giedd JN, Castellanos FX, Haxby JV, Noll DC, Cohen JD, Forman SD, Dahl RE, Rapoport JL. A developmental functional MRI study of prefrontal activation during performance of a Go-No-Go task. *Journal of Cognitive Neuroscience* 1997;9:835–847.
- Cavada C, Company T, Tejedor J, Cruz-Rizzolo RJ, Reinoso-Suarez F. The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cereb Cortex* 2000;10:220–242.
- Clarke HF, Dalley JW, Crofts HS, Robbins TW, Roberts AC. Cognitive Inflexibility After Prefrontal Serotonin Depletion. *Science* 2004;304:878–880. [PubMed: 15131308]
- Dias R, Robbins TW, Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. *Nature* 1996;380:69–72. [PubMed: 8598908]
- Dreher JC, Kohn P, Berman KF. Neural coding of distinct statistical properties of reward information in humans. *Cereb Cortex* 2006;16:561–573. [PubMed: 16033924]
- Elliott R, Dolan RJ. Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. *Journal of Neuroscience* 1999;19:5066–5073. [PubMed: 10366639]
- Elliott R, Newman JL, Longe OA, Deakin JFW. Differential Response Patterns in the Striatum and Orbitofrontal Cortex to Financial Reward in Humans: A Parametric Functional Magnetic Resonance Imaging Study. *J. Neurosci* 2003;23:303–307. [PubMed: 12514228]
- Fellows LK, Farah MJ. Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain* 2003;126:1830–1837. [PubMed: 12821528]
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magnetic Resonance in Medicine* 1995;33:636–647. [PubMed: 7596267]

- Frank MJ, Claus ED. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 2006;113:300–326. [PubMed: 16637763]
- Fuster, JM. The prefrontal cortex : anatomy, physiology, and neuropsychology of the frontal lobe. Vol. Philadelphia: Lippincott-Raven; 1997.
- Garavan H, Ross TJ, Murphy K, Roche RA, Stein EA. Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *Neuroimage* 2002;17:1820–1829. [PubMed: 12498755]
- Ghashghaei HT, Barbas H. Neural interaction between the basal forebrain and functionally distinct prefrontal cortices in the rhesus monkey. *Neuroscience* 2001;103:593–614. [PubMed: 11274781]
- Glover GH, Law CS. Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magn Reson Med* 2001;46:515–522. [PubMed: 11550244]
- Goldstein RZ, Volkow ND. Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the frontal cortex. *Am J Psychiatry* 2002;159:1642–1652. [PubMed: 12359667]
- Gottfried JA, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* 2003;301:1104–1107. [PubMed: 12934011]
- Grachev ID, Apkarian AV. Anxiety in healthy humans is associated with orbital frontal chemistry. *Mol Psychiatry* 2000;5:482–488. [PubMed: 11032381]
- Hikosaka K, Watanabe M. Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cereb Cortex* 2000;10:263–271. [PubMed: 10731221]
- Kim H, Shimojo S, O'Doherty JP. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* 2006;4:1453–1461.
- Konishi S, Nakajima K, Uchida I, Sekihara K, Miyashita Y. No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *Eur J Neurosci* 1998;10:1209–1213. [PubMed: 9753190]
- Kornblum S. The way irrelevant dimensions are processed depends on what they overlap with: the case of Stroop- and Simon-like stimuli. *Psychol Res* 1994;56:130–135. [PubMed: 8008775]
- Kringelbach ML, Rolls ET. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Prog Neurobiol* 2004;72:341–372. [PubMed: 15157726]
- Liotti M, Mayberg HS, McGinnis S, Brannan SL, Jerabek P. Unmasking disease-specific cerebral blood flow abnormalities: mood challenge in patients with remitted unipolar depression. *Am J Psychiatry* 2002;159:1830–1840. [PubMed: 12411216]
- London ED, Ernst M, Grant S, Bonson K, Weinstein A. Orbitofrontal cortex and human drug abuse: functional imaging. *Cereb Cortex* 2000;10:334–342. [PubMed: 10731228]
- London ED, Simon SL, Berman SM, Mandelkern MA, Lichtman AM, Bramen J, Shinn AK, Miotto K, Learn J, Dong Y, Matochik JA, Kurian V, Newton T, Woods R, Rawson R, Ling W. Mood disturbances and regional cerebral metabolic abnormalities in recently abstinent methamphetamine abusers. *Arch Gen Psychiatry* 2004;61:73–84. [PubMed: 14706946]
- MacDonald AW 3rd, Cohen JD, Stenger VA, Carter CS. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 2000;288:1835–1838. [PubMed: 10846167]
- Maddox WT, Baldwin GC, Markman AB. A test of the regulatory fit hypothesis in perceptual classification learning. *Mem Cognit* 2006;34:1377–1397.
- Maunsell JHR. Neuronal representations of cognitive state: reward or attention? *Trends in Cognitive Sciences* 2004;8:261–265. [PubMed: 15165551]
- Mellers BA, Schwartz A, Ho K, Ritov I. Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science* 1997;8:423–429.
- Mellers BA. Anticipated Emotions as Guides to Choice. *Current Directions in Psychol Sci* 2001;10:210–214.
- Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 2001;24:167–202.

- O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* 2001;4:95–102. [PubMed: 11135651]
- O'Doherty J, Critchley H, Deichmann R, Dolan RJ. Dissociating Valence of Outcome from Behavioral Control in Human Orbital and Ventral Prefrontal Cortices. *J. Neurosci* 2003;23:7931–7939. [PubMed: 12944524]
- O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ. Neural responses during anticipation of a primary taste reward. *Neuron* 2002;33:815–826. [PubMed: 11879657]
- Ramnani N, Miall RC. Instructed delay activity in the human prefrontal cortex is modulated by monetary reward expectation. *Cereb Cortex* 2003;13:318–327. [PubMed: 12571121]
- Rauch SL, Savage CR, Alpert NM, Fischman AJ, Jenike MA. The functional neuroanatomy of anxiety: a study of three disorders using positron emission tomography and symptom provocation. *Biological Psychiatry* 1997;42:446–452. [PubMed: 9285080]
- Remijne PL, Nielen MM, Uylings HB, Veltman DJ. Neural correlates of a reversal learning task with an affectively neutral baseline: An event-related fMRI study. *Neuroimage* 2005;26:609–618. [PubMed: 15907318]
- Roesch MR, Olson CR. Neuronal Activity Related to Reward Value and Motivation in Primate Frontal Cortex. *Science* 2004;304:307–310. [PubMed: 15073380]
- Rogers RD, Owen AM, Middleton HC, Williams EJ, Pickard JD, Sahakian BJ, Robbins TW. Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex. *J Neurosci* 1999;19:9029–9038. [PubMed: 10516320]
- Rolls ET, Hornak J, Wade D, McGrath J. Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery and Psychiatry* 1994;57:1518–1524.
- Rolls ET, Critchley HD, Mason R, Wakeman EA. Orbitofrontal cortex neurons: Role in olfactory and visual association learning. *Journal of Neurophysiology* 1996;75:1970–1981. [PubMed: 8734596]
- Rolls, ET. The orbitofrontal cortex. In: Roberts, AC.; Robbins, TW.; W. L., editors. *The prefrontal cortex: executive and cognitive functions*. Vol. New York: Oxford University Press; 1998. p. 67–86. ed. ^eds
- Rolls ET. The orbitofrontal cortex and reward. *Cerebral Cortex* 2000;10:284–294. [PubMed: 10731223]
- Saxena S, Brody AL, Schwartz JM, Baxter LR. Neuroimaging and frontal-subcortical circuitry in obsessive-compulsive disorder. *British Journal of Psychiatry - Supplement* 1998;26–37. [PubMed: 9829024]
- Seymour B, O'Doherty JP, Koltzenburg M, Wiech K, Frackowiak R, Friston K, Dolan R. Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat Neurosci* 2005;8:1234–1240. [PubMed: 16116445]
- Simon JR. Reactions toward the source of stimulation. *J Exp Psychol* 1969;81:174–176. [PubMed: 5812172]
- Small DM, Zatorre RJ, Dagher A, Evans AC, Jones-Gotman M. Changes in brain activity related to eating chocolate: from pleasure to aversion. *Brain* 2001;124:1720–1733. [PubMed: 11522575]
- Taylor SF, Welsh RC, Wager TD, Luan Phan K, Fitzgerald KD, Gehring WJ. A functional neuroimaging study of motivation and executive function. *Neuroimage* 2004;21:1045–1054. [PubMed: 15006672]
- Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. *Nature* 1999;398:704–708. [PubMed: 10227292]
- Tremblay L, Schultz W. Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. *Journal of Neurophysiology* 2000a;83:1877–1885. [PubMed: 10758099]
- Tremblay L, Schultz W. Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. *Journal of Neurophysiology* 2000b;83:1864–1876. [PubMed: 10758098]
- Ursu S, Stenger VA, Shear MK, Jones MR, Carter CS. Overactive action monitoring in obsessive-compulsive disorder: evidence from functional magnetic resonance imaging. *Psychol Sci* 2003;14:347–353. [PubMed: 12807408]
- Ursu S, Carter CS. Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: implications for neuroimaging studies of decision-making. *Brain Res Cogn Brain Res* 2005;23:51–60. [PubMed: 15795133]

- Waldvogel D, van Gelderen P, Muellbacher W, Ziemann U, Immisch I, Hallett M. The relative metabolic demand of inhibition and excitation. *Nature* 2000;406:995–998. [PubMed: 10984053]
- Wallis JD, Miller EK. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur J Neurosci* 2003;18:2069–2081. [PubMed: 14622240]
- Watanabe M, Hikosaka K, Sakagami M, Shirakawa S. Coding and monitoring of motivational context in the primate prefrontal cortex. *J Neurosci* 2002;22:2391–2400. [PubMed: 11896178]
- Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC. Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography* 1998a;22:139–152. [PubMed: 9448779]
- Woods RP, Grafton ST, Watson JD, Sicotte NL, Mazziotta JC. Automated image registration: II. Intersubject validation of linear and nonlinear models. *J Comput Assist Tomogr* 1998b;22:153–165. [PubMed: 9448780]

Figure 1. Behavioral tasks used in the two experiments

Both studies used a slow event-related design (the timing of the 16 scans sampling each trial is depicted in parallel with the events of a trial). Each trial consisted of a brief presentation of a sample stimulus, a sample-target interval, and a brief presentation of the target stimulus. Immediately after target offset, the feedback screen was presented (transparent rectangles in the figure), followed by the inter-trial interval. At the beginning of each block of trials, subjects were given the response rule for that block: identify the location of the sample star in the target and respond on the same side (low control trial, LC), or to identify the location of the sample star but respond on the opposite side (high control condition, HC, due to the need to suppress the spatial response bias). The color of the stimuli changed randomly from trial to trial, indicating the monetary consequences of the response to the target. In Experiment 1, two colors were used, coding for Reward or No-Reward trials. Subjects only had one second to respond to the target (i.e. late responses counted as errors). In Experiment 2, three colors coded for Reward, No-Reward, and potential Penalties (i.e. incorrect or slow responses resulted in a loss). Unbeknownst to the subjects, the response deadline in Experiment 2 was individually titrated in order to avoid differential speed-accuracy tradeoffs in the incentive trials (see Methods and Behavioral results).

Figure 2. Behavioral results of the two studies

Effects of task are presented in the left column graphs while effects of incentives are on the right. **Panel A.** In Experiment 1, mean reaction time of correct trials was shorter for Reward than No- Reward trials, and for LC task relative to HC. In this version of the task, response accuracy was similar across conditions. **Panel B.** In Experiment 2, subjects' responses were faster (top graph) and more accurate (bottom graph) for LC than HC, and for Reward or Penalty relative to No-Reward. No accuracy or reaction time difference was present between Reward and Penalty.

Figure 3. Effects of Incentives and Task during the preparation phase in Experiment 1
 3D renderings of brain areas with differential dynamics of activity (i.e. significant condition by scan interactions). All graphed activations were significant in planned t tests. To summarize the directionality of effects, the graphs display the average signal change during the 6 seconds around the peak activity of each condition. **Panel A:** The contrast of Incentives showed that the left lateral OFC (bottom graph, left BA 11/47, Talairach coordinates -24mm, 36mm, -15mm) was activated during expectation of least favorable outcomes, i.e. preparation for No-Reward trials relative to Reward. This was also direct evidence against a role in incentive-induced control, since No-Reward trials were associated with less effective task execution. Dorsal PFC areas (top graph) showed increased activity during preparation for Reward trials.

Panel B: Task contrast, controlled for task-related outcomes (i.e. counterbalanced incentive levels and equal accuracy rates to HC vs. LC trials): Only the dorsal PFC and inferior convexity (BA10/46 and 10/47, Talairach coordinates of peaks -36, 53, -4 and -47, 41, -7) were active during preparation for HC trials.

Figure 4. Effects of Incentives and Task during the preparation phase in Experiment 2
 3D renderings of brain areas with differential dynamics of activity (the graphs display the average signal change during the 6 seconds around the peak activity of each condition). **Panel A:** Comparing incentive effects (with counterbalanced HC vs LC trials) showed lateral OFC activity (peak Talairach coordinates -31, 29, -10) to expectation of negative incentive-related outcomes (i.e. Penalty trials). In contrast, activity in the dorsolateral PFC (top graph, peak Talairach coordinates -44, 40, 30) was equally increased by Reward or Penalty. **Panel B:** The HC vs LC task contrast when the HC task was associated with negative outcomes (i.e. increased error rates). Similar to Experiment 1, preparation for HC trials elicited increased activity in the dorsolateral PFC (peak Talairach coordinates -53, 20, 18). In addition, a similar albeit weaker

effect was present in the lateral OFC region (BA 11/47, peak Talairach coordinates -34, 35, -10) activated by anticipation of negative incentives in both experiments (see Results).