Differential brain activity during emotional versus nonemotional reversal learning

Differential Brain Activity During Emotional vs. Non-emotional Reversal Learning

Kaoru Nashiro, Michiko Sakaki, Lin Nga, and Mara Mather

University of Southern California

Author Note

This work was supported by grants from the National Institute on Aging (R01AG025340, K02AG032309 and 5T32AG000037).

Correspondence concerning this article should be addressed to Kaoru Nashiro, University of Southern California, 3715 McClintock Avenue, Los Angeles, CA 90089-0191. E-mail: nashiro@usc.edu

Abstract

The ability to change an established stimulus-behavior association based on feedback is critical for adaptive social behaviors. This ability has been examined in reversal learning tasks, where participants first learn a stimulus-response association (e.g., select a particular object to get a reward), and then need to alter their response when reinforcement contingencies change. While substantial evidence demonstrates that the orbitofrontal cortex (OFC) is a critical region for reversal learning, previous studies have not distinguished reversal learning for emotional associations from neutral associations. The current study examined whether OFC plays similar roles in emotional vs. neutral reversal learning. The OFC showed greater activity during reversals of stimulus-outcome associations for negative outcomes than for neutral outcomes. Similar OFC activity was also observed during reversals involving positive outcomes. Furthermore, OFC activity is more inversely correlated with amygdala activity during negative reversals than during neutral reversals. Overall, our results indicate that the OFC is more activated by emotional than neutral reversal learning and that OFC's interactions with the amygdala are greater for negative than neutral reversal learning.

Reversal learning is the ability to alter a behavior when reinforcement contingencies change. In a typical reversal learning task, one first learns stimulus-reward contingencies (e.g., selecting a particular object yields a monetary reward, or choosing the face that will show a happier expression). Once one has learned the initial association, the contingencies are reversed (e.g., the object that once yielded the reward no longer does so) at which point one needs to respond to the previously unrewarded stimulus to obtain a reward. Impairments in reversal learning are related to social abnormality and psychiatric disorders, such as obsessive compulsive disorder (Remijnse et al., 2006), major depressive disorder (Remijnse et al., 2009), psychopathy (Blair, Colledge, & Mitchell, 2001; Budhani, Richell, & Blair, 2006; Mitchell, Colledge, Leonard, & Blair, 2002), and intermittent explosive disorder (Best, Williams, & Coccaro, 2002); thus, reversal learning is a skill related to social and behavioral adaptation.

Previous research has identified the orbitofrontal cortex (OFC) as a critical region for reversal learning (Ghahremani, Monterosso, Jentsch, Bilder, & Poldrack, 2010; Kringelbach & Rolls, 2003; Rolls & Grabenhorst, 2008; Tsuchida, Doll, & Fellows, 2010). The OFC plays a key role in reversal learning of various associations, such as object-points (Budhani, Marsh, Pine, & Blair, 2007; Ghahremani et al., 2010), card-money (Fellows & Farah, 2003; Tsuchida et al., 2010) and face-expression contingencies (Kringelbach & Rolls, 2003; Rolls & Grabenhorst, 2008). The critical role of OFC in reversal learning was also found in animal models (Bissonette et al., 2008; Man, Clarke, & Roberts, 2009; Rudebeck et al., 2008).

However, it remains unclear whether the OFC is essential for reversal learning of emotional associations or reversal learning in general, irrespective of the emotional valence of associations. For example, one recent study (Nahum, Simon, Sander, Lazeyras, & Schnider, 2011) compared neural activity when the associations-to-be-reversed had negative valence (e.g., a spider) and when the associations-to-be-reversed had neutral valence (e.g., a disk). In this

study, participants were instructed to choose which of two faces would appear with a target (either a disk or a spider) on its nose. Over time, the face associated with the target was switched, and participants had to choose the previously incorrect face to see a target. The results revealed similar levels of activity in the OFC when reversing face-spider associations and face-disk associations, suggesting that OFC is important for reversal learning of previous associations irrespective of their emotionality.

In this study, however, reversal trials in the neutral condition involved emotional components as well. In the neutral condition where a disk was a target stimulus, a spider appeared on the nose of the previously correct face to indicate a reversal of face-disk associations. Thus, the cue to signal reversal in the neutral condition had negative valence (spider), which makes it unclear whether the observed OFC activity was as a result of avoiding to choose a previously correct face that is now associated with a spider (emotional associations) or in response to learning new associations between a correct face and a disk (neutral associations). To elucidate this, the current study introduced a novel neutral condition where outcome cues were always neutral even on reversal trials. In addition, we had two emotion conditions (positive and negative) to examine whether the different valence of the outcomes would produce different patterns of OFC activity during reversal learning. Using this paradigm, the current study examined whether OFC activity differs during reversal learning of emotional associations from that of neutral associations.

Recent studies have demonstrated another important aspect of the role of OFC in reversal learning. One study (Stalnaker, Franz, Singh, & Schoenbaum, 2007) using an operant reversal learning task of order-solution associations demonstrated that reversal learning was impaired in the OFC lesioned group but was not affected in the amygdala lesioned group. However, a more striking finding was that damage to both OFC and amygdala did not impair reversal learning

compared to a control group without any lesions. The results together suggest that the interactions between the OFC and the amygdala are critical for reversal learning rather than OFC activity alone, suggesting that the OFC has a modulating effect on the amygdala that protects old emotional representations. Similar effects of OFC and amygdala lesions were found for macaque monkeys' instrumental extinction learning, which also required memory updating of old emotional associations (Izquierdo & Murray, 2005).

Given the evidence that the OFC interacts with the amygdala to update old representations (Izquierdo & Murray, 2005; Stalnaker et al., 2007) and that the amygdala is more critical for emotional than neutral memory regardless of emotional valence (Hamann, Ely, Grafton, & Kilts, 1999), it seems possible that emotional reversal learning requires greater OFC activity to counteract the amygdala than does neutral reversal learning. Thus, we hypothesized that: 1) the OFC will show greater activity during emotional reversal learning than neutral reversal learning, and 2) OFC activity will be more negatively correlated with the amygdala during emotional reversal learning than neutral reversal learning.

Methods

Participants

Twenty undergraduates (M_{age} = 25.35, 12 males, 8 females, age range 19-35) participated in the study. They provided written informed consent approved by the University of Southern California (USC) Institutional Review Board and were paid for their participation. Prospective participants were screened and excluded for any medical, neurological, or psychiatric illness. Two participants were excluded from all analyses due to very poor task performance (their number of errors or number of no responses was greater than 3 standard deviations above the mean). One participant was excluded from all analyses due to excessive motion during the scan.

Materials

The face stimuli were color images obtained from the FACES database developed at the Max Planck Institute for Human Development (Ebner, Riediger, & Lindenberger, 2010), which included young, middle-aged and older adults' female and male faces.

Thirty individuals' faces, which had neutral, happy, angry, and eyeglasses versions, were used in the main experiment. These faces were grouped into fifteen pairs of two faces from the same age group (i.e., five pairs of younger faces, five pairs of middle-aged faces, and five pairs of older faces), and the gender of each pair was always the same (i.e., male-male, female-female pairs). One out of five pairs in each age category was randomly selected and assigned to each participant, resulting in three pairs from different age groups being used for each participant. Which of the three pairs were used for which of the three conditions was randomly determined for each participant. Gender of face pairs were counterbalanced across participants, such that half of the participants saw two female pairs and one male pair while the other half saw one female pair and two male pairs. Each of the faces in a pair randomly appeared on the left or right side of the screen on each trial.

Behavioral Procedures

Before the main experiment began, participants completed two shorter practice blocks outside the scanner. The procedure in the practice session was the same as the main task described below, except that it was shorter and had a different categorization rule. During practice, participants were asked to identify the person who had a baseball cap and then who was sad. We used two pairs of faces that were not used in the main experiment.

The main experiment consisted of positive, negative and neutral blocks, the order of which was randomized across the participants. At the beginning of each block, a prompt appeared; "Who is happy?" "Who is angry?" or "Who wears glasses?" in the positive, negative or neutral conditions respectively. Each trial lasted for 6 seconds and began with the

presentation of two neutral faces with a white background (see Figure 1). Participants were asked to select one face with the target characteristics (happy, angry, or eyeglasses) by pressing a key corresponding to the left or right side of the screen. Immediately after their response, feedback was presented for 1 second on a gray background. If the response was correct, the selected face changed (into a happy face, angry face, or face with eyeglasses), while the other face remained neutral. If the response was incorrect, both of the faces remained neutral. When the participant did not respond within 4 seconds, the warning "please respond faster" was displayed. The trial ended with a fixation cross for the remainder of the 6 seconds. After three to six consecutive correct responses, the correct face was reversed. Participants were asked to keep track of the correct face and change their answers as soon as they noticed the switch.

Trial Modeling

Each trial was categorized as one of three trial types: reversal, acquisition and other. 'Reversal' described individual trials where the participant selected the previously correct person, but this led to a neutral face expression indicating that the response was incorrect. Reversal trials were defined so that they always followed by a response shift in the next trial; thus, trials where the participant selected the previously correct person, but did not change their response in a subsequent trial were not included. This categorization allowed us to capture brain activity when the participant made a final error immediately before switching their response. It should be noted that there were no differences in terms of the perceptual properties or the stimulus emotionality across positive, negative and neutral conditions during the reversal trials since participants viewed two neutral faces during reversal in all conditions. 'Acquisition' included series of trials where the subject's correct choices of a particular person led to a change in the face (i.e., happy face, angry face, or face appearing with eyeglasses). The first trial of each condition was modeled as 'other' (regardless of whether the subject made a correct or incorrect

choice), as these trials required subjects to guess and do not reflect learning (or failure of learning) of previous associations. The rest of the trials, which did not fall into the categories of reversal or acquisition trials, were also aggregated as 'other.' For example, 'other' includes trials where the participant chose incorrect faces before reaching the criterion (three to six consecutive correct responses) or trials where the participant failed to respond within 4 seconds.

Functional MRI Data Acquisition and Preprocessing

Imaging was conducted with a 3 T Siemens MAGNETOM Trio scanner with a 12-channel matrix head coil at the University of Southern California Dana and David Dornsife Neuroimaging Center. The imaging parameters were repetition time (TR) = 2000 ms, echo time (TE) = 25 ms, slice thickness = 3 mm, interslice gap = 0 mm, flip angle (FA) = 90°, and field of view (FOV) = 192 mm x 192 mm. Data preprocessing were performed using FMRIB's Software Library (FSL; www.fmrib.ox.ac.uk/fsl), which included motion correction with MCFLIRT, spatial smoothing with a Gaussian kernel of full-width half-maximum 5 mm, high-pass temporal filtering equivalent to 100 seconds, and skull stripping of structural images with BET. MELODIC ICA (Beckmann & Smith, 2004) was used to remove noise components. Registration was performed with FLIRT; each functional image was registered to both the participant's high-resolution brain-extracted structural image and the standard Montreal Neurological Institute (MNI) 2-mm brain.

FMRI Data Analyses.

Whole-brain analysis. For each reversal trial for each participant, stimulus-dependent changes in BOLD signal were modeled with regressors for feedback and fixation events. Signal from the feedback and fixation periods were averaged for each valence condition. The selection period (the initial presentation of two neutral faces) was modeled as the baseline level of activity and therefore, was not included as a regressor. In addition, motion regressors were included to

adjust for volumes with sharp movement. 'Acquisition' and 'other' trials were also modeled. The regressors were convolved with a double-gamma hemodynamic response function and temporal filtering was applied as well. Temporal derivatives of each the regressors were also included.

Whole-brain analyses were conducted using FSL FEAT v. 5.98 (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). *Z* (Gaussianised T/F) statistic images were thresholded at the whole-brain level using clusters determined by $Z>2.3$ and a (corrected) cluster significance threshold of $p=0.05$ (Worsley, 2001) unless otherwise noted. Locations reported by FSL were converted into Talairach coordinates by the MNI-to-Talairach transformation algorithm (Lancaster et al., 2007). These coordinates were used to determine the nearest gray matter using the Talairach Daemon version 2.4.2 (Lancaster et al., 2000).

Regions-of-interest (ROI) analyses. Given previous findings that the lateral OFC, in particular, plays an important role in reversal learning (Hampshire & Owen, 2006; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001), we performed ROI analyses to examine whether this OFC sub-region shows different activities in reversal learning across the conditions. The left and right lateral OFC were structurally defined using UCLA's Laboratory of Neuro Imaging LPBA40 atlas (Shattuck et al., 2008), set at a 0.5 probabilistic threshold.

Given past findings that the amygdala also plays a role in reversal learning in interaction with the OFC (Izquierdo & Murray, 2005; Stalnaker et al., 2007), we performed ROI analyses for the left and right amygdala. The amygdala were segmented from each participant's high resolution structural scan using FreeSurfer (surfer.nmr.mgh.harvard.edu) and FSL FAST (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). For each participant, the amygdala from the segmenting software judged as more accurate was selected for further manual correction. Next, manual correction of this selected ROI was carried out using FSLView and involved

removing erroneous voxels in non-amygdala regions (e.g., hippocampus, white matter). For both ROI analyses, FSL Featquery was used to extract percent signal change values.

Functional connectivity analyses. To examine functional connectivity, we applied a beta series correlation analysis (Gazzaley, Cooney, Rissman, & D'Esposito, 2005; Rissman, Gazzaley, & D'Esposito, 2004). This allowed us to use trial-to-trial variability to characterize dynamic inter-regional interactions. The left lateral OFC, which served as the seed region, was functionally defined based on shared voxels from activation clusters (contrasting the positive and negative conditions, respectively, to the neutral) voxel-thresholded at a z=2.3 in the whole brain analysis.

First, a new GLM design file was constructed where each reversal trial was coded as a unique covariate, resulting in up to 39 independent variables (the maximum number of reversal trials achieved by participants across all three conditions). To reduce the confounding effects of the global signal change, the mean signal level over all brain voxels was calculated for each time point and was used as a covariate. The model also involved additional nuisance regressors for acquisition and 'other' trials. Second, the least squares solution of the GLM yielded a beta value for each reversal trial for each individual participant. These beta values were then sorted by conditions. Third, mean activity (i.e., mean parameter estimates) was extracted for each individual reversal trial from a seed region. Fourth, for each condition, we computed correlations between the seed's beta series and the beta series of all other voxels in the brain, thus generating condition-specific seed correlation maps. Correlation magnitudes were converted into *z*-scores using the Fisher's *r*-to-*z* transformation. Condition-dependent changes in functional connectivity were assessed using random-effects analyses, which were thresholded at the whole-brain level using clusters determined by $Z>2.3$ and a (corrected) cluster significance threshold of $p=0.05$.

Results

Behavioral Results

The errors made in the first trial of each condition were excluded, as those were guessing errors and were not due to failure of learning previous associations. The rest of the errors were divided into two types: reversal and other. The total number of reversal errors was calculated for each condition. A one-way repeated-measures ANOVA (conditions: positive, negative, neutral) revealed no significant difference between conditions ($M_{positive}$ = 10.41, SE = 0.47; $M_{negative}$ = 10.82, SE = 0.38, $M_{neutral}$ = 10.94, SE = 0.47), F (2, 32) = 0.90, MSE = 1.46, p = .42, η_p^2 = .05, suggesting that participants performed similarly across conditions. The total number of other errors was also calculated for each condition; however, no significant differences across conditions were found, F (2, 32) = 0.77, MSE = 0.91, p = .47, η_p^2 = .05.

FMRI Results

First, we contrasted brain activity during reversal and acquisition in order to examine whether the OFC is more important for reversal learning than acquisition. For the rest of the analyses, we contrasted brain activity during the reversal trials across conditions in which there were no differences in the perceptual properties or the stimulus emotionality (Figure 1B).

Brain regions showing greater activity during reversal than acquisition. When collapsed across the three valence conditions, reversal compared with acquisition trials produced increased activity in OFC/insula (BA 47/13), dorsolateral PFC (BA 9), frontopolar area (BA 10), and anterior cingulate cortex (BA 24 and 32). Furthermore, secondary motor cortex (BA 6), somatosensory association cortex (BA 7), V3 (BA 19), superior temporal gyrus (BA 22), and supramarginal gyrus part of Wernicke's area (BA 40) showed increased activity in reversal than acquisition trials. Thus, consistent with previous research (Ghahremani, Monterosso, Jentsch, Bilder, & Poldrack, 2010; Kringelbach & Rolls, 2003; Rolls & Grabenhorst, 2008; Tsuchida,

Doll, & Fellows, 2010), the OFC showed greater activity during reversal than acquisition trials, indicating a critical role of the OFC in reversal learning.

Brain regions showing different activity during emotional vs. neutral reversal learning. We examined our hypothesis that the positive and negative emotion conditions produce different patterns of brain activity than the neutral condition during reversal learning. The whole-brain analysis revealed greater activity in the negative than neutral conditions in inferior frontal gyrus/OFC (BA 47), precentral gyrus (BA 9), frontal pole (BA 10), anterior cingulate (BA 24, 32), and insula (BA 13). Other regions showing significant differences in the negative-neutral contrast are reported in Table 1. There were no significant findings in other contrasts (negative-positive, positive-negative, positive-neutral, neutral-positive, neutral-negative). However, when we used a lower threshold (a voxel-threshold of $z = 2.3$), the positive-neutral contrast yielded similar results to the ones in the negative-neutral contrast. When compared with the neutral condition, the positive condition produced greater activity in inferior frontal gyrus/OFC (BA 47; Figure 2), precentral gyrus (BA 9), frontal pole (BA 10), anterior cingulate (BA 24) and insula (BA 13). Although these results based on use of a lower threshold should be interpreted with caution, they provide useful information about the similarities between the positive and negative conditions in contrast with the neutral condition. Next, we combined the positive and negative conditions (together called the emotion condition) and contrasted them against the neutral condition. The emotion condition yielded greater activity in areas including inferior frontal gyrus/OFC (BA 47), precentral gyrus (BA 9), insula (BA 13) and anterior cingulate (BA 24) than did the neutral condition, whereas the reverse contrast showed no significant findings (Table 2; Figure 2). The results suggest that the OFC is more important for emotional than for neutral reversal learning. Although not hypothesized, other regions, such as insula, also seem more involved in emotional reversal learning than in neutral reversal learning.

ROI analysis for the lateral OFC. One-way ANOVAs (comparing positive, negative, and neutral conditions) were performed on the percent signal change from the left and right lateral OFC. There was a significant effect of condition in the left lateral OFC, $F(2, 32) = 6.55$, $MSE = 0.05$, $p < .01$, $\eta_p^2 = .29$, but not in the right lateral OFC ($p = .21$). Post-hoc t-tests suggest that the left lateral OFC showed significantly greater activity in the negative than the neutral conditions, $t(16) = 3.40$, $p = .004$, and in the positive than the neutral conditions, $t(16) = 2.22$, $p = .04$, whereas there was no significant difference between the negative and the positive conditions ($p = .18$; see Figure 3). These results suggest that the left lateral OFC is more involved in emotional reversal learning than in neutral reversal learning, regardless of valence. However, it remains unclear why this region showed reduced activity during neutral reversals than baseline, and additional investigation is needed to address this point.

ROI analysis for the amygdala.

One-way ANOVAs (comparing positive, negative, and neutral conditions) were performed on the percent signal change from the left and right amygdala. There was a marginally significant effect of condition in the left amygdala, $F(2, 32) = 2.95$, $MSE = 0.13$, $p = .067$, $\eta_p^2 = .16$, and a significant effect of condition in the right amygdala, $F(2, 32) = 7.44$, $MSE = 0.08$, $p = .002$, $\eta_p^2 = .32$. A post-hoc t-test suggests that the left amygdala showed significantly greater activity in the negative than the neutral conditions, $t(16) = 2.93$, $p = .01$, and the same pattern was seen in the right amygdala, $t(16) = 3.99$, $p = .001$ (Figure 4). The right amygdala also showed significantly greater activity in the positive than the neutral conditions, $t(16) = -2.59$, $p = .020$. There were no other significant findings.

Functional connectivity analysis with the left lateral OFC as a seed region. The whole brain connectivity analysis comparing the negative and neutral conditions revealed that the negative condition produced a significantly greater negative correlation between the left

lateral OFC and the left parahippocampal gyrus/amygdala than did the neutral condition (Figure 5; Table 3). We did not find greater negative correlations between the left lateral OFC and the amygdala in any other contrasts.

Discussion

While many previous studies suggested that OFC is important for reversal learning, they did not indicate whether the OFC is more involved in reversal learning of emotional associations or equally involved in reversal learning regardless of the valence of associations. To investigate this, we introduced a novel condition where feedback was always neutral, enabling us to examine the differences in neural activity during neutral vs. emotional reversal learning.

In line with our first hypothesis, we found that OFC is more involved in emotional reversal learning than neutral reversal learning. The whole-brain and ROI results revealed that the OFC produced greater activity during reversal learning of negative associations than of neutral associations. Although relatively weaker (and non significant) OFC activity was found in the positive-neutral contrast compared with the negative-neutral contrast, the positive and negative conditions showed a similar pattern of OFC activity during reversal trials (as compared with the neutral condition). In addition, the ROI analysis indicated that the left lateral OFC showed significantly greater activity in the negative and positive conditions than in the neutral condition, with no significant differences between the positive and negative conditions. These results largely supported our first hypothesis that OFC plays a more critical role in emotional than neutral reversal learning. We also found that OFC has greater inverse correlations with parahippocampal gyrus/amygdala during reversal learning in the negative condition than in the neutral condition. Although we did not find similar patterns in the positive–neutral contrast, these results are in line with our second hypothesis and suggest that OFC down-regulates amygdala to allow for flexible reversal learning.

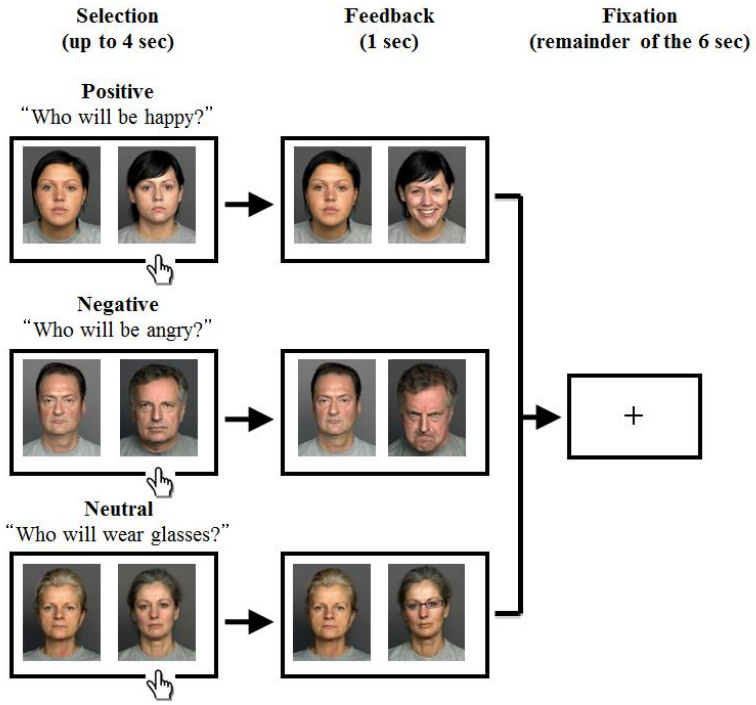
The negative correlations between the OFC regions and the amygdala have also been implicated in previous studies using different learning tasks that have reversal learning components. One study used an extinction learning paradigm where initial object-point associations were reversed in the extinction phase so that participants had to learn to respond to previously punishing objects and avoid responding to previously rewarding objects (Finger, Mitchell, Jones & Blair, 2008). During successful extinction, frontopolar OFC activity showed significant negative correlations with activity in the right and left amygdala. Similarly, a recent study on memory updating using a long-term memory paradigm (Sakaki, Niki, & Mather, 2011) found that the frontal pole had negative correlations with the amygdala when people learned new associations to old emotional items. These findings are consistent with the idea that the frontopolar OFC helps update old associations by countering amygdala's protection of previous representations (Schoenbaum, Saddoris, & Stalnaker, 2007; Stalnaker et al., 2007). By including a novel neutral condition, the current study further demonstrated that there were greater negative correlations between the OFC and amygdala during reversal learning of negative associations than that of neutral associations, consistent with the notion that OFC-amygdala interactions are particularly important for reversal learning of emotional associations.

The question remains as to why we did not observe greater negative correlations between the OFC and the amygdala in the positive than the neutral conditions. One possible explanation is that positive reversal learning did not evoke as strong an emotional response as did negative reversal learning; hence, reversals of positive associations required less OFC involvement to modulate old representations in the amygdala than did reversals of negative associations. In fact, our ROI results suggest that both the left lateral OFC and bilateral amygdala showed less activity during positive than negative reversal learning (albeit the differences between the positive and negative conditions were not significant), suggesting that positive reversal learning may require

less OFC resources than does negative reversal learning. Related to these findings, previous research suggests that negative reversal learning is more difficult or effortful than positive reversal learning. A recent ERP study (Willis, Palermo, Burke, Atkinson, & McArthur, 2010) found that people performed worse at switching associations formed with angry expressions than with happy expressions. In addition, they found that P3s amplitude was reduced and P3b latency was delayed during negative compared to positive reversal learning, suggesting that old negative representations may be more resistant to modification than old positive representations. Taken together, our findings suggest that OFC is involved in both positive and negative reversal learning; however, there might be differences between the two conditions with respect to task difficulty and the timing of neural activity. Further investigation is needed to test these possibilities.

In conclusion, the current study provides important new information about the role of OFC in reversal learning. Our results suggest that the OFC is more critical for emotional than neutral reversal learning and that OFC's interactions with the amygdala are greater for negative than neutral reversal learning. Future research should investigate more precise roles of the OFC during positive and negative reversal learning by using various levels of stimulus intensity and task difficulty.

A) Acquisition Trials



B) Reversal Learning Trials

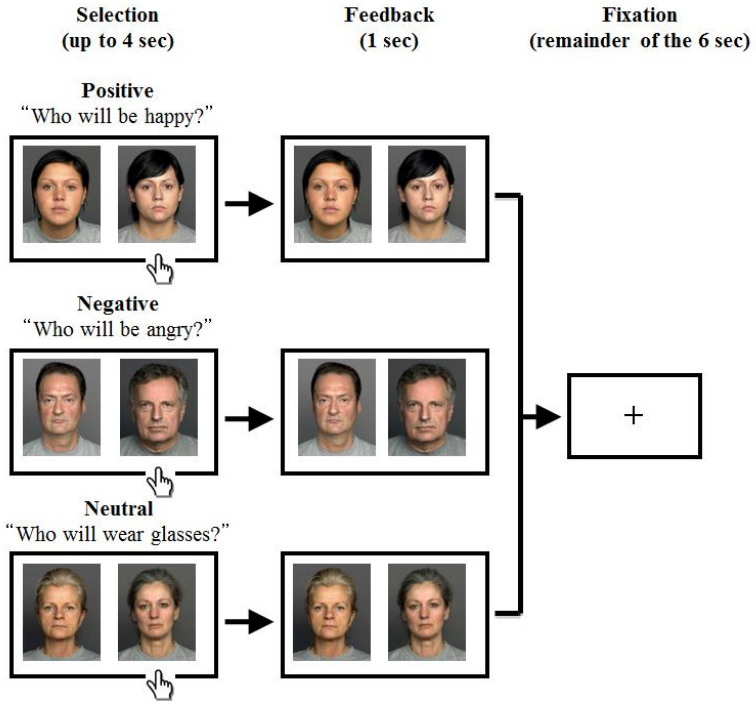


Figure 1. Experimental Procedure. The positive (top), negative (middle) or neutral blocks (bottom) were assigned to the participant in a random order. The two people were randomly assigned to the right or the left of the screen. The trial began with a presentation of two people

displaying neutral expressions during which the participant had to select one person by pressing a key. Feedback was presented for 1 sec, which was followed by a fixation cross for the remainder of the 6 sec. A) In Acquisition Trials where the response was correct, the selected face changed (into a happy face, angry face, or face with eyeglasses respectively), while the other face remained neutral. B) In Reversal Learning Trials where the response was incorrect, both of the faces remained neutral. Across conditions, the task for the subject was to keep track of the correct person because it switched mid-game. The correct person changed after between three and six consecutive correct trials; the number of trials before the change was unknown to the subject.

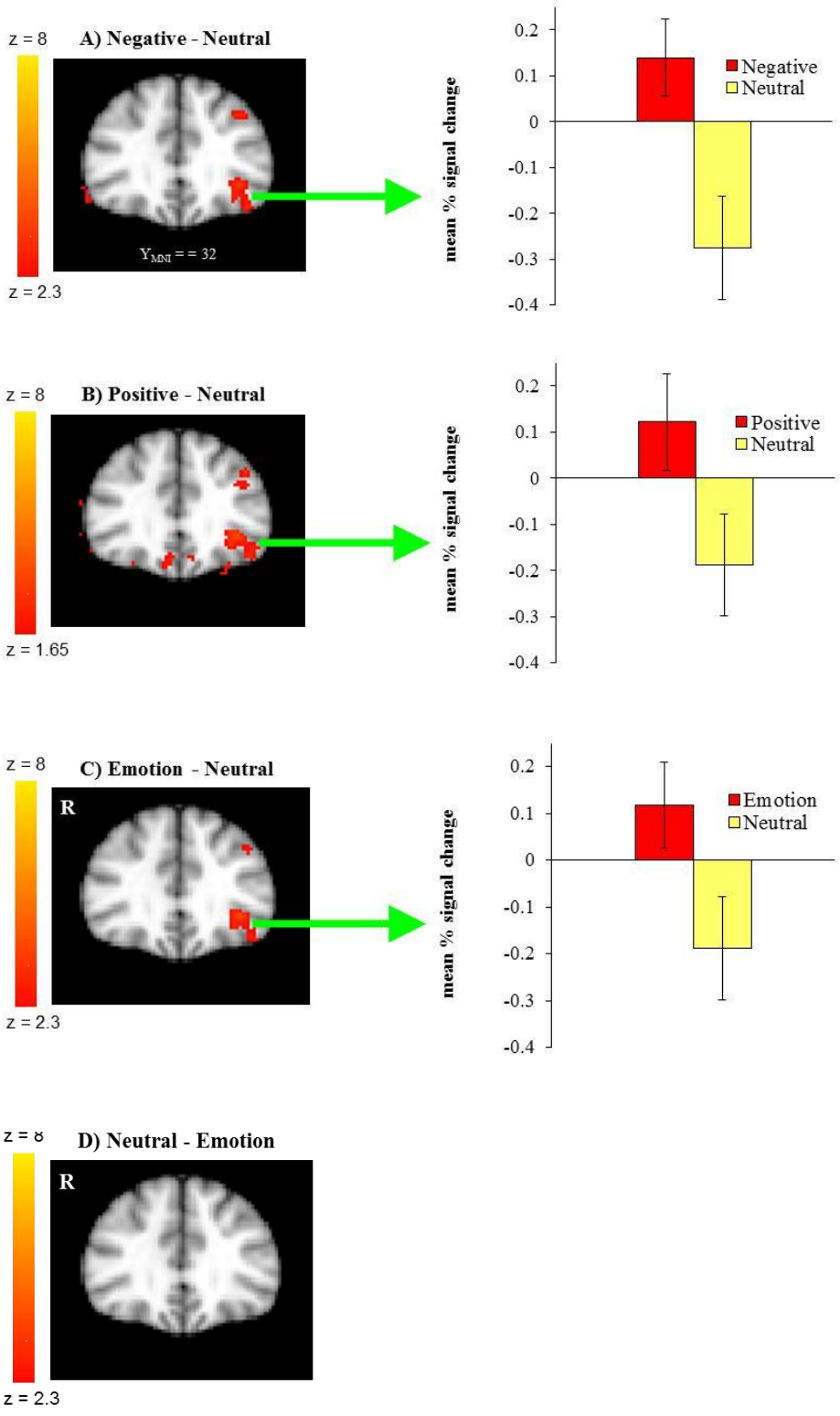


Figure 2. A) The OFC showed greater activity when participants reversed negative associations than neutral associations. B) The positive-neutral contrast also showed a similar pattern of left lateral OFC activity (as compared with the neutral condition) when the voxel threshold was lowered to z = 1.65 for image B. Although the low-threshold map should be interpreted with caution, it provides useful information about the similarities between the positive and negative

conditions in contrast to the neutral condition. C) When positive and negative conditions were combined, the emotion condition showed greater activity in the left lateral OFC than did the neutral condition, D) whereas the reverse contrast showed no significant findings. The images were threshholded at the whole-brain level using clusters determined by $z > 2.3$ and a (corrected) cluster significance threshold of $p = 0.05$, except for image B. The bar graphs show the mean % signal change within a sphere of 3-mm radius centered at the peak voxel in the left lateral OFC for each contrast (A [x,y,z] = -42, 32, -16; B [x,y,z] = -42, 26, -10; C [x,y,z] = -42, 26, -10).

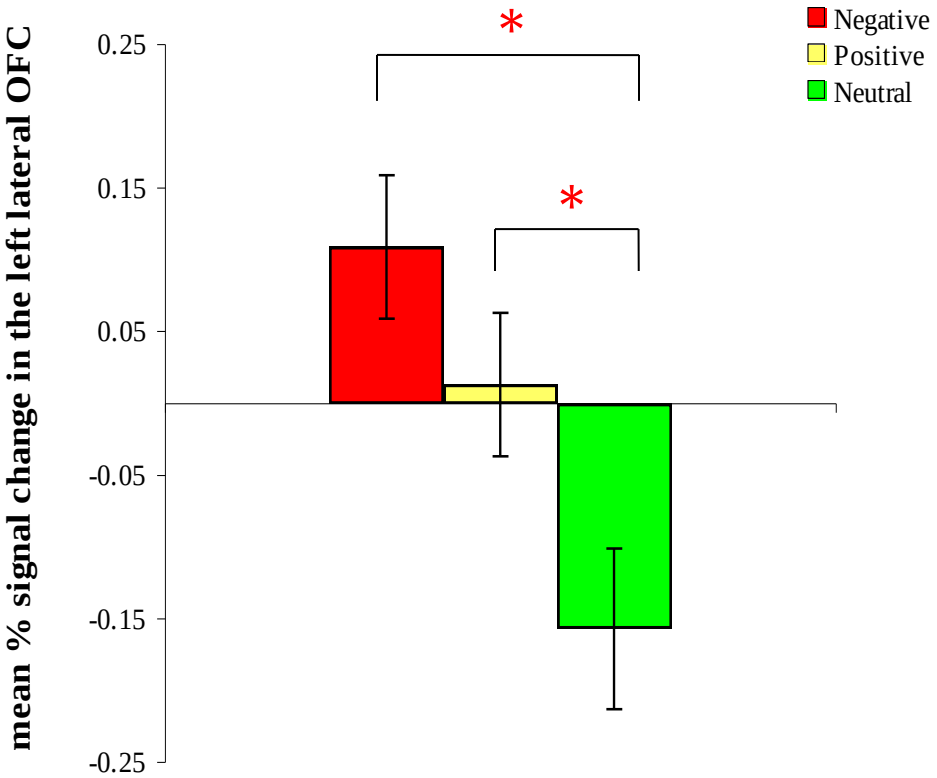


Figure 3. The left lateral OFC activity during reversal learning across conditions. The left lateral OFC showed significantly greater activity in the negative than neutral conditions and in the positive than neutral conditions (*ps* < .05).

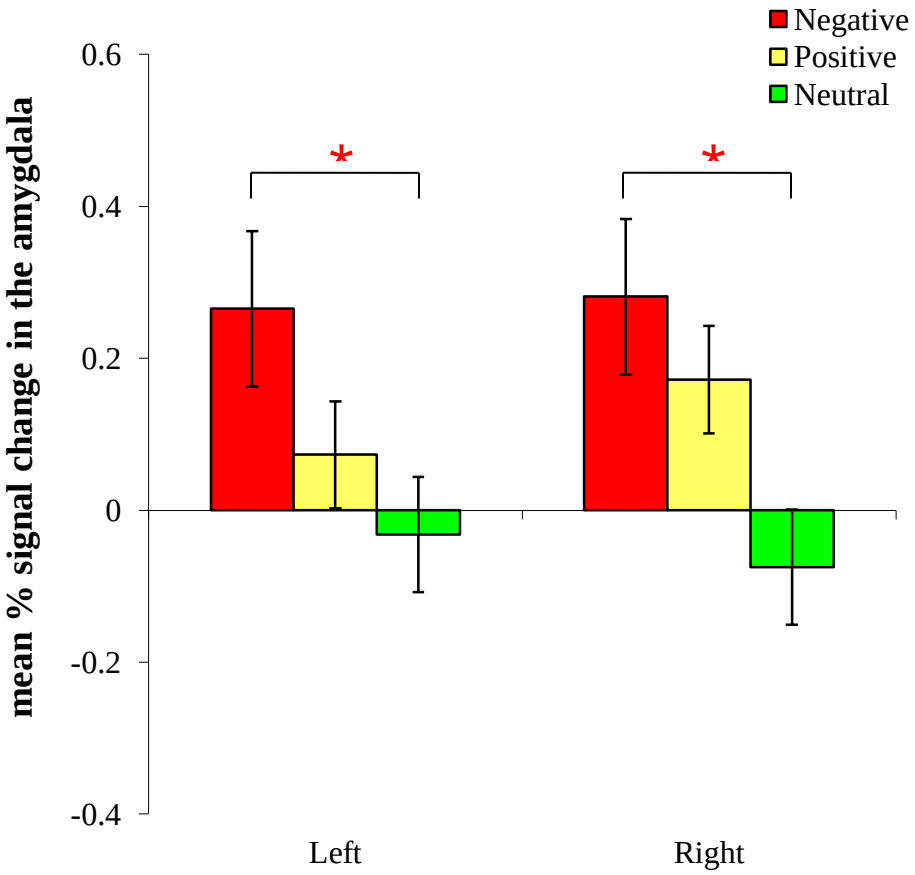


Figure 4. The amygdala activity during reversal learning across conditions. Both the left and right amygdala showed significantly greater activity in the negative than neutral conditions ($p <$ .05).

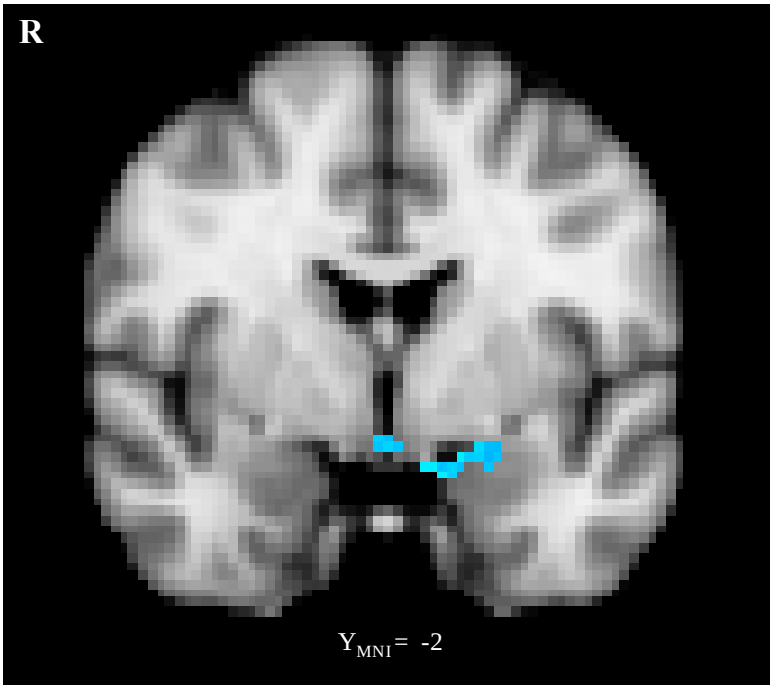


Figure 5. The left lateral OFC cluster showed more negative functional connectivity with the left parahippocampal gyrus/amygdala in the negative condition than in the neutral condition. The image was threshholded at the whole-brain level using clusters determined by $z > 2.3$ and a (corrected) cluster significance threshold of $p = .05$.

Table 1. Brain Activity Showing Significant Differences between Conditions during Reversal Learning.

			MNI			*Talairach*			
Area	*H*	*BA*	*x*	*y*	*z*	*x*	*y*	*z*	*Z-max*
Negative > Neutral									
Precuneus	L	7	0	-40	52	-5	-44	47	3.34
Postcentral Gyrus	L	3	-18	-30	74	-19	-36	67	3.27
Anterior Cingulate	L	24	2	-14	36	0	-18	35	3.19
Thalamus	R		22	-32	6	19	-32	7	3.64
Superior Temporal Gyrus	R	22	58	10	-14	53	8	-7	3.45
Parahippocampal Gyrus	R	34	26	2	-16	23	1	-10	3.31
Insula	L	13	-50	-48	22	-48	-48	19	3.58
Fusiform Gyrus	L	37	-48	-54	-2	-46	-52	-3	3.56
Fusiform Gyrus	L	37	-58	-56	-2	-55	-53	-4	3.48
Lentiform Nucleus	L		-32	-20	-2	-31	-20	0	3.40
Superior Temporal Gyrus	L	41	-38	-30	6	-36	-30	6	3.33
Caudate	L		-38	-30	2	-36	-30	3	3.26
Inferior Frontal Gyrus	L	47	-44	26	-30	-41	25	-21	3.19
Inferior Frontal Gyrus	L	47	-36	32	-4	-34	29	3	3.18
Inferior Frontal Gyrus	L	47	-52	22	-8	-49	20	-2	3.16
Postcentral Gyrus	R	31	6	-52	30	4	-53	26	3.47
Cuneus	L	7	0	-64	36	-2	-65	31	3.00
Cuneus	L	7	0	-74	40	-2	-74	33	2.90
Precentral Gyrus	L	9	-40	28	38	-38	21	40	3.44
Inferior Frontal Gyrus	L	9	-54	14	30	-51	9	31	3.20
Precentral Gyrus	L	6	-44	6	44	-42	0	43	3.05
Superior Temporal Gyrus	R	22	58	-32	2	53	-32	4	3.19
Insula	R	13	52	-34	26	47	-36	25	3.14
Superior Temporal Gyrus	R	42	66	-30	16	60	-31	17	3.10
Lentiform Nucleus	L		-20	10	8	-20	7	12	3.18
Claustrum	L		-34	2	6	-33	0	9	3.17
Claustrum	L		-34	-2	6	-33	-4	9	3.10
Anterior Cingulate	L	32	-4	42	-14	-4	39	-5	3.65
Frontal pole	R	10	2	66	-12	1	61	-1	3.03
Frontal pole	L	10	-8	66	-14	-8	61	-3	3.00
Middle Occipital Gyrus	L	19	-40	-82	10	-38	-79	5	3.63
Middle Occipital Gyrus	L	18	-26	-94	14	-25	-90	8	3.60
Cuneus	L	17	-22	-94	12	-22	-90	6	3.60
Positive > Neutral									
No significant results									
Negative > Positive									
No significant results									
Positive > Negative									
No significant results									
Neutral > Negative									
No significant results									

Neutral > Positive
No significant results

Table 2. Brain Activity Showing Significant Differences between Emotion vs. Neutral Conditions during Reversal Learning

			MNI			*Talairach*			
Area	*H*	*BA*	*x*	*y*	*z*	*x*	*y*	*z*	*Z-max*
Emotion > Neutral									
Insula	L	13	-50	-48	22	-48	-48	19	3.66
Fusiform Gyrus	L	37	-58	-56	6	-55	-54	4	3.50
Fusiform Gyrus	L	37	-48	-54	-2	-46	-52	-3	3.44
Anterior Cingulate	L	24	0	-16	36	-1	-20	35	3.26
Posterior Cingulate	L	31	2	-26	50	0	-30	46	3.24
Posterior Cingulate	L	23	-2	-10	32	-3	-14	32	3.20
Insula	R	13	52	-34	26	47	-36	25	3.32
Insula	R	13	60	-32	18	54	-33	18	3.30
Superior Temporal Gyrus	R	22	64	-40	10	58	-40	10	3.21
Transverse Temporal Gyrus	L	41	-34	-30	10	-33	-30	10	3.32
Lentiform Nucleus	L		-32	-20	-4	-31	-20	-2	3.19
Superior Temporal Gyrus	L	41	-40	-30	2	-38	-30	3	2.92
Precentral Gyrus	L	9	-40	28	36	-38	21	38	3.48
Inferior Frontal Gyrus	L	9	-54	14	30	-51	9	31	3.06
Inferior Frontal Gyrus	L	9	-50	12	28	-48	7	29	2.99
Inferior Frontal Gyrus	L	47	-36	32	-4	-34	29	3	3.47
Inferior Frontal Gyrus	L	47	-46	22	-22	-43	21	-14	3.19
Inferior Frontal Gyrus	L	47	-52	20	-14	-49	18	-7	3.09
Middle Occipital Gyrus	L	18	-22	-94	14	-22	-90	8	3.67
Middle Occipital Gyrus	L	19	-36	-88	14	-35	-85	8	3.46
Neutral > Emotion									
No significant results									

Table 3. Brain Regions Showing Differential Negative Connectivity with the Left Lateral OFC across Conditions

			MNI			*Talairach*			
Area	*H*	*BA*	*x*	*y*	*z*	*x*	*y*	*z*	*Z-max*
Negative > Neutral									
Parahippocampal gyrus/Amygdala	L	34	-10	0	-13	-10	-1	-8	4.14
Anterior Cingulate	L	25	1	5	-9	0	4	-3	3.44
Caudate	L		-12	24	-1	-12	21	5	3.30
Hypothalamus			1	2	-9	0	1	-4	3.24
Inferior Frontal Gyrus	L	47	-16	17	-10	-16	15	-4	3.19
Positive > Neutral									
No significant results									
Negative > Positive									
No significant results									
Positive > Negative									
Supramarginal Gyrus	R	40	48	-45	30	48	-45	30	3.50
Inferior Parietal Lobule	R	40	50	-43	51	50	-43	51	3.48
Precuneus	L	31	-9	-67	22	-9	-67	22	3.64
Cingulate Gyrus	L	31	-1	-41	39	-1	-41	39	3.79
Neutral > Negative									
Cingulate Gyrus	R	31	10	-44	42	8	-46	38	3.74
Middle Temporal Gyrus	L	39	-54	-73	11	-52	-71	6	5.05
Middle Temporal Gyrus	L	22	-54	-31	6	-51	-31	6	4.30
Superior Temporal Gyrus	R	22	40	-53	13	36	-53	11	3.82
Inferior Parietal Lobule	L	40	-60	-37	33	-57	-39	29	3.66
Neutral > Positive									
Superior Temporal Gyrus	L	42	-66	-32	19	-62	-33	17	4.22
Middle Temporal Gyrus	L	39	-54	-74	13	-52	-71	8	3.93
Anterior Cingulate	L	32	-16	24	34	-16	18	36	4.36
Insula	L	13	-45	-3	6	-42	-5	9	3.66

References

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging, 23*(2), 137-152. doi: 10.1109/TMI.2003.822821

Best, M., Williams, J. M., & Coccaro, E. F. (2002). Evidence for a dysfunctional prefrontal circuit in patients with an impulsive aggressive disorder. *Proceedings of the National Academy of Sciences, U.S.A, 99*(12), 8448-8453. doi: 10.1073/pnas.112604099

Bissonette, G. B., Martins, G. J., Franz, T. M., Harper, E. S., Schoenbaum, G., & Powell, E. M. (2008). Double dissociation of the effects of medial and orbital prefrontal cortical lesions on attentional and affective shifts in mice. *Journal of Neuroscience, 28*(44), 11124-11130. doi: 10.1523/JNEUROSCI.2820-08.2008

Blair, R. J., Colledge, E., & Mitchell, D. G. (2001). Somatic markers and response reversal: is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? *Journal of Abnormal Child Psychology*, *29*(6), 499-511.

Budhani, S., Marsh, A. A., Pine, D. S., & Blair, R. J. (2007). Neural correlates of response reversal: considering acquisition. *Neuroimage, 34*(4), 1754-1765. doi: 10.1016/j.neuroimage.2006.08.060

Budhani, S., Richell, R. A., & Blair, R. J. (2006). Impaired reversal but intact acquisition: probabilistic response reversal deficits in adult individuals with psychopathy. *Journal of Abnormal Child Psychology, 115*(3), 552-558. doi: 10.1037/0021-843X.115.3.552

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES--a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behavior Research Methods, 42*(1), 351-362. doi: 10.3758/BRM.42.1.351

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain, 126*(Pt 8), 1830-1837. doi: 10.1093/brain/awg180

Finger, E. C., Mitchell, D. G. V., Jones, M., & Blair, R. J. R.(2008). Dissociable roles of medial orbito-frontal cortex in human operant extinction learning. *Neuroimage, 43,*748–755. doi: 10.1016/j.neuroimage.2008.08.021

Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neuroscience, 8*(10), 1298-1300. doi: 10.1038/nn1543

Ghahremani, D. G., Monterosso, J., Jentsch, J. D., Bilder, R. M., & Poldrack, R. A. (2010). Neural components underlying behavioral flexibility in human reversal learning. *Cerebral Cortex, 20*(8), 1843-1852. doi: 10.1093/cercor/bhp247

Hamann, S. B., Ely, T. D., Grafton, S. T., & Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience, 2*(3), 289-293. doi: 10.1038/6404

Hampshire, A., & Owen, A. M. (2006). Fractionating attentional control using event-related fMRI. *Cerebral Cortex, 16*(12), 1679-1689. doi: 10.1093/cercor/bhj116

Izquierdo, A., & Murray, E. A. (2005). Opposing effects of amygdala and orbital prefrontal cortex lesions on the extinction of instrumental responding in macaque monkeys. *European Journal of Neuroscience, 22*(9), 2341-2346. doi: 10.1111/j.1460-9568.2005.04434.x

Kringelbach, M. L., & Rolls, E. T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction. *Neuroimage, 20*(2), 1371-1383. doi: 10.1016/S1053-8119(03)00393-8

Lancaster, J. L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., . . . Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBN-152 brain template. *Human Brain Mapping, 28*(11), 1194-1205. doi: 10.1002/hbm.20345

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., . . . Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping, 10*(3), 120-131. doi: 10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8

Man, M. S., Clarke, H. F., & Roberts, A. C. (2009). The role of the orbitofrontal cortex and medial striatum in the regulation of prepotent responses to food rewards. *Cerebral Cortex, 19*(4), 899-906. doi: 10.1093/cercor/bhn137

Mitchell, D. G., Colledge, E., Leonard, A., & Blair, R. J. (2002). Risky decisions and response reversal: is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? *Neuropsychologia, 40*(12), 2013-2022.

Nahum, L., Simon, S. R., Sander, D., Lazeyras, F., & Schnider, A. (2011). Neural response to the behaviorally relevant absence of anticipated outcomes and the presentation of potentially harmful stimuli: A human fMRI study. *Cortex, 47*(2), 191-201. doi: 10.1016/j.cortex.2009.11.007

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4*(1), 95-102. doi: 10.1038/82959

Remijnse, P. L., Nielen, M. M., van Balkom, A. J., Cath, D. C., van Oppen, P., Uylings, H. B., & Veltman, D. J. (2006). Reduced orbitofrontal-striatal activity on a reversal learning task

in obsessive-compulsive disorder. *Archives of General Psychiatry, 63*(11), 1225-1236. doi: 10.1001/archpsyc.63.11.1225

Remijnse, P. L., Nielen, M. M., van Balkom, A. J., Hendriks, G. J., Hoogendijk, W. J., Uylings, H. B., & Veltman, D. J. (2009). Differential frontal-striatal and paralimbic activity during reversal learning in major depressive disorder and obsessive-compulsive disorder. *Psychological Medicine, 39*(9), 1503-1518. doi: 10.1017/S0033291708005072

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage, 23*(2), 752-763. doi: 10.1016/j.neuroimage.2004.06.035

Rolls, E. T. (2004). The functions of the orbitofrontal cortex. *Brain and Cognition, 55*(1), 11-29. doi: 10.1016/S0278-2626(03)00277-X

Rolls, E. T., & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making. *Progress in Neurobiology, 86*(3), 216-244. doi: 10.1016/j.pneurobio.2008.09.001

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., & Rushworth, M. F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *Journal of Neuroscience, 28*(51), 13775-13785. doi: 10.1523/JNEUROSCI.3541-08.2008

Sakaki, M., Niki, K., & Mather, M. (2011). Updating Existing Emotional Memories Involves the Frontopolar/Orbito-frontal Cortex in Ways that Acquiring New Emotional Memories Does Not. *Journal of Cognitive Neuroscience.* doi: 10.1162/jocn_a_00057

Schoenbaum, G., Saddoris, M. P., & Stalnaker, T. A. (2007). Reconciling the roles of orbitofrontal cortex in reversal learning and the encoding of outcome expectancies.

Annals of the New York Academy of Sciences, 1121, 320–335. doi: 10.1196/annals.1401.001

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., . . . Toga, A. W. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage, 39*(3), 1064-1080. doi: 10.1016/j.neuroimage.2007.09.031

Stalnaker, T. A., Franz, T. M., Singh, T., & Schoenbaum, G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron, 54*(1), 51-58. doi: 10.1016/j.neuron.2007.02.014

Tsuchida, A., Doll, B. B., & Fellows, L. K. (2010). Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *Journal of Neuroscience, 30*(50), 16868-16875. doi: 10.1523/JNEUROSCI.1958-10.2010

Willis, M. L., Palermo, R., Burke, D., Atkinson, C. M., & McArthur, G. (2010). Switching associations between facial identity and emotional expression: A behavioural and ERP study. *Neuromage, 50*(1), 329-339. doi: 10.1016/j.neuroimage.2009.11.071

Worsley, K.J., Statistical analysis of activation images, in Functional MRI: An Introduction to Methods, P. Jezzard, P.M. Matthews, and S.M. Smith, Editors. 2001, Oxford University Press: New York.