

# Amygdala Response to Facial Expressions Reflects Emotional Learning

Christine I. Hooker, Laura T. Germine, Robert T. Knight, and Mark D'Esposito

Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, California 94720-3190

The functional role of the human amygdala in the evaluation of emotional facial expressions is unclear. Previous animal and human research shows that the amygdala participates in processing positive and negative reinforcement as well as in learning predictive associations between stimuli and subsequent reinforcement. Thus, amygdala response to facial expressions could reflect the processing of primary reinforcement or emotional learning. Here, using functional magnetic resonance imaging, we tested the hypothesis that amygdala response to facial expressions is driven by emotional association learning. We show that the amygdala is more responsive to learning object–emotion associations from happy and fearful facial expressions than it is to the presentation of happy and fearful facial expressions alone. The results provide evidence that the amygdala uses social signals to rapidly and flexibly learn threatening and rewarding associations that ultimately serve to enhance survival.

**Key words:** amygdala; classical conditioning; emotion; fear; reward; fMRI

## Introduction

Although it is well established that the human amygdala is engaged in the perception of facial expressions of emotion (Zald, 2003), the exact function of the amygdala in this process is debated. One influential hypothesis, derived from the role of the amygdala in pavlovian conditioning, suggests that emotional facial expressions act as a predictive cue that a biologically relevant (e.g., threatening or rewarding) stimulus is in the environment (Whalen, 1998). Animal and human research shows that the amygdala is responsive to aversive and appetitive primary reinforcement (i.e., an unconditioned stimulus), such as shock or food, but it is more critical for learning the association of a predictive cue (e.g., tone) and subsequent reinforcement (e.g., shock) (Gallagher and Holland, 1994; LeDoux, 2000; Holland and Gallagher, 2004). In this context, the facial expression of fear or joy could act as a signal that an aversive or appetitive stimulus is in the immediate vicinity. However, facial expressions can act as a primary reinforcer as well as a predictive cue (Canli et al., 2002; Blair, 2003; O'Doherty et al., 2003b). Therefore, separating the predictive value from the inherent reinforcement value requires a direct comparison of facial expressions used for association learning compared with facial expressions presented without learning.

Previous research investigating the response of the amygdala

to social signals in association learning has focused on the assumed level of ambiguity in the object–emotion association and the pattern of habituation in response to the social signal. Data suggest that if the object–emotion association is ambiguous, more amygdala activity is initially engaged to learn the association (Whalen et al., 2001; Adams et al., 2003; Hooker et al., 2003). In addition, if the ambiguous association is prolonged without learning, than amygdala activity habituates (Fischer et al., 2003), presumably because, without apparent consequences, learning the cause of emotional response is not as crucial to survival as it initially appeared. However, these data provide only indirect evidence that the amygdala is analyzing faces for the purpose of association learning.

Here, we directly test the hypothesis that the amygdala is responsive to facial expressions for emotional association learning by using functional magnetic resonance imaging (fMRI) to compare neural activity when subjects are engaged in learning object–emotion associations from facial expressions compared with perceiving facial expressions without any learning requirement.

In the association learning (AL) task, subjects saw a visual display of two novel objects on either side of a woman's face. A cue indicated one object. Subjects predicted whether the woman would react with a fearful versus neutral or happy versus neutral facial expression. After the prediction, the woman turned and reacted to the object. This object–emotion association was maintained for a short number of trials and then reversed. Subjects used facial expression prediction outcomes to learn the object association. Previous research shows that amygdala activity is highest during initial learning and decreases responding after emotional learning contingencies are well established (Buchel et al., 1999). Therefore, we maximally engage association learning mechanisms by rapidly reversing emotional associations at unpredictable intervals, creating associations that are temporarily

Received July 22, 2005; revised July 12, 2006; accepted July 17, 2006.

This work was supported by a University of California at Davis Medical Investigation of Neurodevelopmental Disorders Institute Postdoctoral Fellowship (C.I.H.), National Institutes of Health Grants MH71746 (C.I.H.), 066737 (R.T.K.), NS21135 (R.T.K.), and MH63901 (M.D.), and the James S. McDonnell Foundation. We thank Christina Kramlich and Tony Hooker for help with stimuli production and Sara Verosky and Asako Miyakawa for assistance with data analysis.

Correspondence should be addressed to Christine I. Hooker, Helen Wills Neuroscience Institute, 132 Barker Hall, University of California at Berkeley, Berkeley, CA 94720-3190. E-mail: chooker@berkeley.edu.

DOI:10.1523/JNEUROSCI.3048-05.2006

Copyright © 2006 Society for Neuroscience 0270-6474/06/268915-08\$15.00/0

maintained but never fully established, thereby preventing amygdala habituation.

In the expression-only (EO) task, subjects saw the same woman's neutral face on the screen and predicted whether she would become fearful versus neutral or happy versus neutral. After the prediction, the woman turned and displayed a facial expression, but there were no objects to form an association. Fear and happy object–emotion learning and expression-only perception were investigated in separate fMRI runs.

## Materials and Methods

**Participants.** Twelve healthy, right-handed adults (seven females; mean age, 25 years old; range, 19–36 years old) volunteered and were paid for their participation. All subjects gave written, informed consent before participation in accordance with the guidelines of the Committee for Protection of Human Subjects at the University of California, Berkeley. Subjects were screened for MR compatibility as well as neurological and psychiatric illness.

**Behavioral task.** Subjects performed two behavioral tasks while being scanned: the AL task and the EO task.

The AL task is a visual discrimination, reversal learning task. Subjects saw a visual display of two unrecognizable, neutral objects on either side of a woman's neutral face and predicted whether the woman would react with a fearful versus neutral or a happy versus neutral facial expression to the cued object. After the prediction, the woman turned to the object and reacted with an emotional (fear/happy) or neutral expression. This association remained constant for an unpredictable number of trials, ranging from one to nine trials (with an average of 4.6) and then reversed. There were 10 reversal learning trials per run (i.e., the first learning trial and nine reversals). Thus, subjects always made a forced-choice decision with two options: in fear association learning runs, subjects predicted a fearful or neutral response; in happy association learning runs, subjects predicted a happy or neutral response. Fear association learning and happy association learning were investigated in separate fMRI runs to allow the comparison of object–emotion and object–neutral association learning.

For example, in AL fear run 1, the first trial started with the visual display of a woman's neutral face, with object A on the right and object B on the left. Object A was indicated with a fixation cross and the subject predicted whether the woman would react fearfully or neutrally to the object. After the prediction, the woman turned and responded with a fearful expression. This constituted an initial learning trial in which object A is associated with fear. This association of object A to fear and object B to neutral remained constant for four trials, with two object A (fear) and two object B (neutral) association trials, appearing on the right or left side of the screen. On the fifth trial, the association reversed, so that object A became associated with neutral and object B became associated with fear.

In the EO task, subjects saw the same woman's neutral face on the screen and predicted whether the woman would become fearful versus neutral or happy versus neutral. After the prediction, the woman turned and displayed a facial expression, but there were no objects to form an association.

The face stimuli were exactly the same in the AL and EO tasks and consisted of three different expressions (fearful, happy, and neutral) from one woman. To create the stimuli, a rapid series of photographs were taken while a professional actor turned and responded to a neutral object (to her right and left) with a fearful, happy, and neutral expression. In the task, the outcome portion of the trial consisted of 10 photographs, presented for 200 ms each, of the head turn and emotional response, such that it seemed like a short (2 s) video to the subject. Each of the three facial expressions was coded by a trained psychologist using the facial action coding system to verify the authenticity of the emotional expression (Ekman and Friesen, 1978). In addition, each of the four AL runs used two new objects for a total of eight objects, which were downloaded from Michael Tarr's (Brown University, Providence, RI) Web site of objects (<http://alpha.cog.brown.edu:8200/stimuli/novel-objects/fribbles.zip>/view).

**Prescanning practice and instruction.** Subjects practiced the AL task

outside of the scanner with a block of 10 association learning trials (with no reversals). They were told that, when doing the task in the scanner, the object–emotion associations would reverse multiple times and they should use information from the woman's facial expression to update the object association and make an accurate prediction on the next trial. In addition, they were told that there would be another task in which they will only see the woman's face and will have to predict whether she will have an emotional or neutral reaction. They did not practice reversal trials or expression-only trials before entering the scanner. Subjects were instructed to keep their gaze fixated on the woman's face, which was centrally located in the picture display.

**Experimental design.** The tasks were divided into four different run types: (1) association learning fearful or neutral (AL F/N); (2) association learning happy or neutral (AL H/N); (3) expression-only fearful or neutral (EO F/N); and (4) expression-only happy or neutral (EO H/N).

Each subject was scanned on four association learning runs (two AL F/N and two AL H/N) and two expression-only runs (EO F/N and EO H/N), resulting in six experimental runs total. Each run had 56, 4 s trials, consisting of a 2 s cue (neutral face and objects display with target object indicated by fixation cross) and a 2 s outcome (face turns, looks at the target object, and has a neutral or emotional expression). Subjects made their prediction during the cue phase of the trial. There was a 2, 4, or 6 s jittered intertrial interval (black screen with a white fixation cross). Each run had 10 reversal learning trials (the first learning trial and nine reversals). In half of the reversal learning trials, subjects learned an object–emotion association (five trials), and, in the other half, subjects learned an object–neutral association (five trials). The total number of object–emotion and object–neutral association maintenance trials was balanced in each run (23 object–emotion and 23 object–neutral). Each object was displayed an equal number of times on the right and left.

The experimental design and trial structure of the EO task was equivalent to the AL task. An equal number of emotional (28 trials) and neutral (28 trials) were presented in each run.

**Image acquisition.** All images were acquired at 4 tesla using a Varian (Palo Alto, CA) INOVA MR scanner that was equipped with echo-planar imaging (EPI). For all experiments, a standard radiofrequency head coil was used, and a memory foam pillow comfortably restricted head motion. E-Prime software (Psychology Software Tools, Pittsburgh, PA) controlled the stimulus display and recorded subject responses via a magnetic-compatible fiber-optic keypad. A liquid crystal display projector (Epson, Long Beach, CA) projected stimuli onto a backlit projection screen (Stewart, Torrance, CA) within the magnet bore, which the subject viewed via a mirror mounted in the head coil.

Functional images were acquired during six runs. Each run included four dummy scans (with no data acquisition) and four scans at the beginning of the run, which were subsequently dropped from analysis to ensure steady-state magnetization for all analyzed data, resulting in 212 whole-brain volumes per experimental run and 1272 whole-brain volumes per subject for an entire session. Images were acquired with a set of parameters used to optimize signal in regions susceptible to dropout attributable to magnetic field inhomogeneity (Deichmann et al., 2003). Each volume acquisition included 20 3.5-mm-thick oblique axial slices (angled at  $\sim 18^\circ$  from the anterior commissure–posterior commissure line) with a 1 mm interslice gap, acquired in an interleaved manner. A two-shot T2\*-weighted EPI sequence (repetition time, 2000 ms; echo time, 28 ms; field of view, 22.4 cm<sup>2</sup>; matrix size, 64 × 64) was used to acquire blood oxygen level-dependent (BOLD) signal. EPI voxel size at acquisition is 3.5 × 3.5 × 4 mm. All EPI runs were preceded by a preparatory Z-shim to reduce signal dropout from dephasing in the slice direction (Deichmann et al., 2003). A high-resolution three-dimensional T1-weighted structural scan [magnetization-prepared fast low-angle shot (MPFLASH) sequence] and an in-plane low-resolution T2-weighted structural scan [gradient echo multislice (GEMS)] were acquired for anatomical localization.

**Image processing.** MRI data were processed and analyzed using SPM2 software (<http://www.fil.ion.ucl.ac.uk/spm>). Each EPI volume was realigned in space to the first scan, using a six-parameter, rigid-body, least-squares transformation algorithm. All subjects who showed >3 mm of movement across the session were dropped from analyses. After realign-

ment, EPI data were coregistered to the individual subject's coplanar (GEMS) and high-resolution (MPFLASH) anatomical images, normalized to Montreal Neurological Institute (MNI) atlas space, resliced to  $2 \times 2 \times 2$  isotropic voxels, and smoothed (8 mm full-width half-maximum).

**Data analysis.** Event-related BOLD responses were analyzed using a modified general linear model with SPM2 software. Our aim was to investigate brain activity in response to facial expressions occurring at the outcome portion of the trial, which provides information used for association learning. Therefore, we defined our trial types and general linear model according to the facial expression outcome. The cue portion of the trial remained constant throughout each task.

We defined our covariates of interest as follows: (1) AL reversal fear; (2) AL maintenance fear; (3) AL reversal neutral (in fear run); (4) AL maintenance neutral (in fear run); (5) AL reversal happy; (6) AL maintenance happy; (7) AL reversal neutral (in happy run); (8) AL maintenance neutral (in happy run); (9) EO fear; (10) EO neutral (in EO fear run); (11) EO happy; and (12) EO neutral (in EO happy run). The "AL reversal" condition includes the facial expression outcome that cues the new association in the initial learning trial as well as the new association in reversal learning trials. For example, "AL reversal fear" is a fearful facial expression that cues that the object–emotion association is reversing from neutral to fearful.

We then convolved the canonical hemodynamic response function with brain activity at the onset of the outcome type. Brain activity was high-pass filtered at 128 s, scaled by the global mean, and corrected for serial autocorrelation. We computed the difference in neural activity between two outcome types and then computed whether this difference was significant across subjects by entering the contrast value into a one-sample  $t$  test. We report all brain activity that exceeds the statistical threshold of  $t_{(11)} = 4.02$ ,  $p < 0.001$ , uncorrected for multiple comparisons. We subsequently corrected for multiple comparisons within the amygdala, anatomically defined by an MNI template, by using the small-volume correction (SVC) tool in SPM2 (Worsley et al., 1996). Peak activity is reported in MNI ( $x, y, z$ ) coordinates as is provided by SPM2.

We conducted additional analyses to investigate emotion effects in the EO task. An explanation of the methods and the findings are available in the supplementary data (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

## Results

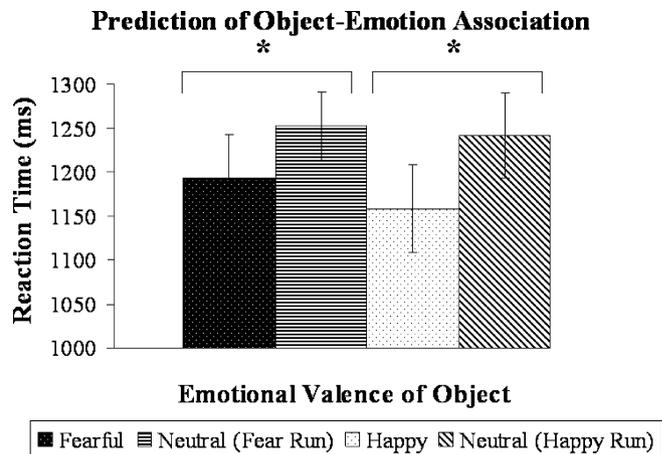
### Behavioral results

Because it was impossible to accurately predict when a reversal would occur, reversal trials (including the initial learning trial of each run) were not counted as errors in the analysis of behavioral data. Subjects performed well on the AL task (mean  $\pm$  SD; AL fear accuracy,  $95 \pm 4\%$ ; AL happy accuracy,  $94 \pm 3\%$ ). As expected, subjects were unable to predict emotional response in the EO task (EO fear accuracy,  $50 \pm 7\%$ ; EO happy accuracy,  $55 \pm 5\%$ ). Reaction time data indicated more efficient object–emotion association learning compared with object–neutral association learning. Subjects were significantly faster to predict that the object was associated with a happy or fearful reaction than a neutral reaction [fearful (mean of 1193 ms) versus neutral (mean of 1252 ms),  $t_{(10)} = -2.6$ ,  $p < 0.05$ ; happy (mean of 1158 ms) versus neutral (mean of 1241 ms),  $t_{(10)} = -3.7$ ,  $p < 0.005$ ; two-tailed paired  $t$  tests] (Fig. 1). There was no significant difference in accuracy when predicting a happy or fearful compared with neutral response.

### Imaging results: overview of analysis

First, we investigated our primary question regarding the role of the amygdala in learning object–emotion associations from emotional faces (happy and fearful) as compared to perceiving emotional faces (happy and fearful) without learning by comparing neural activity in the AL task with the EO task.

Second, we examined emotion effects in the AL versus EO



**Figure 1.** Average reaction time for subjects' prediction of the object–emotion association in maintenance trials in the AL task. Subjects responded significantly faster when predicting an object–emotion association compared with an object–neutral association in both fear and reward learning. \* $p < 0.05$ .

comparison by investigating each type of emotional learning separately as well as emotion-specific learning. The main goal for this analysis was to identify whether the amygdala responded to one type of emotional learning more than another.

Third, in the AL task, we investigated neural regions involved in the new learning of an object–emotion association at the reversal cue compared with the maintenance of an object–emotion association. The main goal for this analysis was to identify whether the amygdala was most responsive during initial learning.

Fourth, we investigated whether the amygdala was responsive to general learning. For this analysis, we examined object–neutral learning versus neutral face perception. We further investigated the general effects of learning by comparing the neural regions involved in learning associations from all faces compared with perceiving faces. The main goal for this analysis was to identify whether amygdala activity was specific for object–emotion learning or whether it was responsive to learning more generally.

### Learning object–emotion associations versus perceiving emotional faces: AL emotion versus EO emotion

To investigate the hypothesis that the amygdala is engaged in the analysis of facial expressions for the purpose of emotional learning, we computed the difference in neural activity between learning object–emotion associations from emotional facial expressions and perceiving emotional facial expressions without learning. In this analysis, we combined reversal and maintenance trials in the AL task and fearful and happy trials for both tasks, resulting in the following contrast: AL emotion (fear + happy) versus EO emotion (fear + happy).

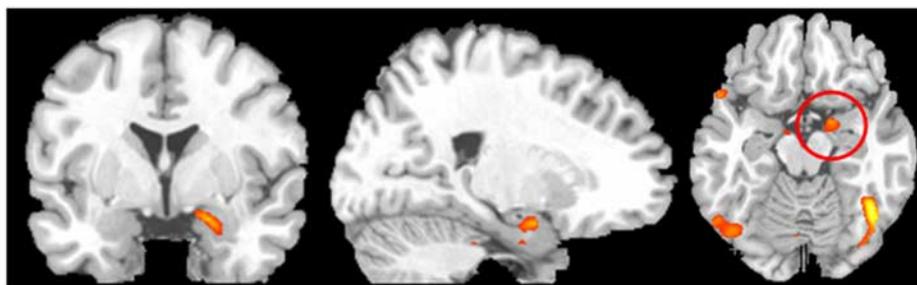
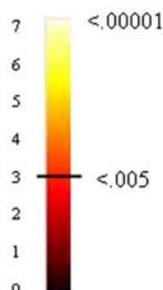
The right amygdala was significantly more active when learning object–emotion associations from facial expressions than it was to the perception of those same facial expressions alone. This amygdala activity is significant when correcting for multiple comparisons within the amygdala volume using small volume corrections (SVC tool in SPM2). Figure 2 shows the trial types used in the primary task comparison and a group activation map.

To confirm that the greater amygdala activity for AL emotion compared with EO emotion (shown in the above analysis) is the result of enhancement of activity for emotional learning in the AL task and not the absence of activity during the EO task, we calculated a one-sample  $t$  test for the percentage signal change from baseline for each condition within the amygdala region function-

### a) Association Learning (Fear+Happy) versus Expression Only (Fear+Happy)



### b) T statistic



**Figure 2.** *a* shows an example of the stimuli used in the primary analysis. Subjects were required to learn whether the woman would respond with a fearful versus neutral or happy versus neutral expression to the novel object. Fear and reward learning were investigated in separate fMRI runs to allow the comparison between object–emotion and object–neutral association learning. *b* shows greater amygdala activity for learning object–emotion associations compared with emotional faces presented without learning. The analysis includes all trials (reversal and maintenance) in which the object was associated with a fearful or happy expression. There was no significant amygdala activity (in the left or right hemisphere) for object–neutral association learning versus neutral faces or for fear learning versus reward learning. Positive activations shown with threshold  $t_{(11)} = 3.1, p < 0.005$  on coronal ( $y = 0$ ), sagittal ( $x = 22$ ), and axial ( $z = -16$ ) slices.

**Table 1.** Brain regions involved in object–emotion learning from emotional expressions compared with the perception of emotional expressions: AL emotion (fearful + happy) versus EO emotion (fearful + happy)

Brain region	Brodmann's area	MNI $x, y, z$ (peak)	$t$ statistic
AL emotion > EO emotion			
Middle frontal gyrus, posterior, right	9	50, 16, 42	5.9
Inferior frontal gyrus, right	44, 45	60, 30, 10	5.3
Fusiform gyrus, right	37	42, -68, -22	7.0
Fusiform gyrus, left	37	-54, -70, -6	4.5
Superior temporal sulcus, right	22	52, -42, 2	4.0
Amygdala, right		22, 0, -18	4.7
Hippocampus, left		-30, -8, -22	5.1
Thalamus, pulvinar region, right		8, -32, 4	5.9
Thalamus, pulvinar region, left		-14, -30, 4	4.4
EO emotion > AL emotion			
Middle frontal gyrus, left	9, 10	-32, 38, 26	5.3
Precentral gyrus, left	4, 6	-52, 2, 34	9.8
Post central gyrus, right	3	38, -24, 46	5.6
Post central gyrus, left	3	-36, -14, 46	5.1
Middle cingulate, right	23, 24	8, -12, 46	5.3
Middle cingulate, left	24	-10, 0, 46	4.9
Anterior insula, right	48	44, 20, 8	5.4
Anterior insula, left	48	-34, 20, 12	4.5
Posterior insula, right	48	38, -26, 12	6.8
Posterior insula, left	48	-42, -28, -12	5.0
Cuneus, right	23, 31	24, -56, 20	8.3
Cuneus, left	23	-14, -62, 20	4.3

AL includes both reversal and maintenance trials.

ally defined by the contrast. This analysis shows that the amygdala was significantly active compared with baseline for both conditions. The average  $\pm$  SD percentage signal change from baseline for AL emotion is  $0.27 \pm 19$  ( $t_{(11)} = 4.9, p < 0.001$ ) and for EO emotion is  $0.19 \pm 0.24$  ( $t_{(11)} = 2.7; p < 0.01$ ; one-sample  $t$  tests).

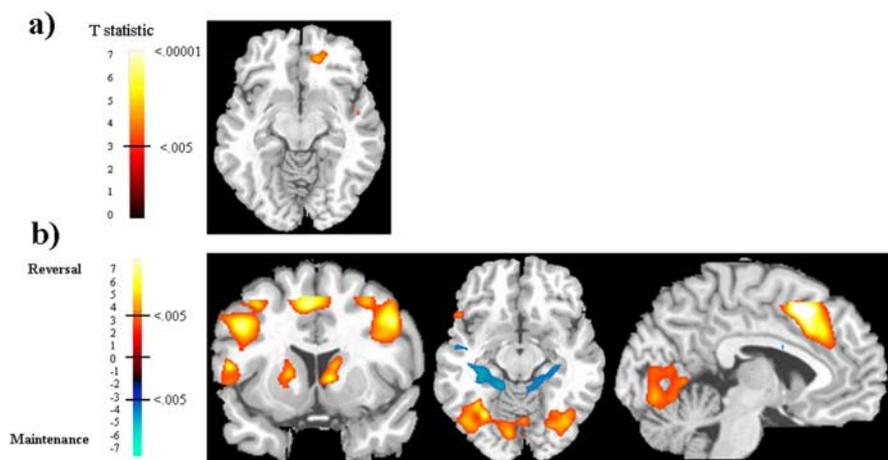
Table 1 lists all activations for this contrast [AL emotion (fear + happy) vs EO emotion (fear + happy)] that survive the threshold of  $t_{(11)} = 4.02, p < 0.001$ . In particular, there was more activity for object–emotion association learning compared with

the perception of emotional expressions in regions associated with learning, such as the right dorsolateral prefrontal cortex (DLPFC) and right ventrolateral prefrontal cortex (VLPFC), as well as regions associated with face perception, such as the right superior temporal sulcus (STS) and bilateral fusiform gyrus.

#### Emotion effects

*Object–fear association learning and object–happy association learning investigated separately.* We investigated object–emotion association learning compared with emotion perception for each emotion separately: object–fear association learning versus fear face perception and object–happy association learning versus happy face perception. We found that the amygdala was more active in the learning condition, at a slightly subthreshold level, for both object–fear and object–happy learning when compared with fear and happy face perception. There was greater activity for AL fear versus EO fear in the right amygdala with peak activity located at  $(x, y, z)$  (22, -6, -8) ( $t_{(11)} = 2.75; p < 0.01$ ) and the left amygdala with peak activity at  $(-24, -8, -14)$  ( $t_{(11)} = 3.1; p < 0.005$ ). There was greater activity for AL happy versus EO happy in the right amygdala with peak activity located at  $(24, -2, -20)$  ( $t_{(11)} = 2.5; p < 0.01$ ). These results suggest that the enhancement of amygdala activity for learning compared with perception in the AL versus EO comparison is not driven by one emotion more than the other. Rather, the amygdala shows a consistent pattern of greater activity for object–emotion learning compared with emotion perception for both fearful and happy conditions.

*Object–fear association learning compared with object–happy association learning.* To identify neural regions that are preferentially responsive to a specific type of emotional learning, we performed the following interactions: (AL fear – EO fear) versus (AL happy – EO happy) to identify object–fear-specific learning; and (AL happy – EO happy) versus (AL fear – EO fear) to identify object–happy-specific learning. We found no preferential activity for either AL fear or AL happy in the amygdala. However, the medial orbitofrontal cortex (OFC) was preferentially responsive to object–happy association learning compared with object–fear learning, even while controlling for the effect of



**Figure 3.** *a* shows medial OFC activity ( $z = -10$ ) for object–happy association learning compared with object–fear association learning while controlling for the effects of happy and fearful faces presented alone: (AL happy – EO happy) versus (AL fear – EO fear). *b* shows brain regions active for new object–emotion association learning compared with the maintenance of object–emotion associations in the contrast: AL reversal (fear + happy) versus AL maintenance (fear + happy). Neural activity in response to facial expressions signaling reversal of associations (i.e., new learning) is shown in red, and neural activity in response to facial expressions signaling maintenance of associations is shown in blue. The coronal slice ( $y = 10$ ) shows reversal-related activity in the DLPFC, VLPFC, and dorsal (caudate) and ventral (putamen) striatum. The axial slice ( $z = -12$ ) shows reversal-related activity in lateral OFC and fusiform gyrus and maintenance-related activity in the hippocampus. The sagittal slice ( $x = -4$ ) shows reversal-related activity in the anterior cingulate. Activations shown with threshold  $t_{(11)} = 3.1, p < 0.005$ .

**Table 2. Brain activity during initial object–emotion association learning occurring at the reversal cue and object–emotion association learning when the association is maintained: AL emotion reversal versus maintenance**

Brain region	Brodmann's area	MNI $x, y, z$ (peak)	$t$ statistic
<b>AL emotion reversal &gt; maintenance</b>			
Middle frontal gyrus, anterior, right	10, 9	28, 42, 26	7.3
Middle frontal gyrus, anterior, left	10	–38, 50, 16	8.0
Middle frontal gyrus, posterior, right	9, 46	50, 16, 32	9.2
Middle frontal gyrus, posterior, left	9, 46	–48, 4, 40	7.4
Inferior frontal gyrus, lateral orbital gyrus, right	45, 47	42, 22, –4	4.9
Inferior frontal gyrus, lateral orbital gyrus, left	44, 45, 47	–54, 18, 6	11.0
Anterior cingulate cortex/ supplementary motor area, bilateral	32, 8	–4, 16, 50	8.5
Fusiform gyrus, right	37	40, –70, –12	4.7
Fusiform gyrus, left	37	–44, –66, 0	7.2
Superior temporal sulcus, right	22	64, –46, 4	4.2
Intraparietal sulcus, full extent, right	7	36, –64, 40	8.2
Intraparietal sulcus, full extent, left	7	–36, –42, 40	10.5
Caudate, right		10, 10, 4	5.5
Caudate, left		–18, 16, 0	6.9
Thalamus, right		8, –12, 14	4.5
Visual cortex, calcarine, right	17	4, –58, 4	6.4
Visual cortex, calcarine, left	17	–8, –78, –8	5.1
<b>AL emotion maintenance &gt; reversal</b>			
Posterior cingulate, left	23, 26	–10, –48, 18	6.9
Posterior insula, right	48	46, –10, 22	6.6
Posterior insula, left	48	–38, –12, 26	6.2
Posterior hippocampus, right	37	30, –34, –2	6.2
Posterior hippocampus, left	37	–32, –36, –4	5.2

happy and fearful faces presented alone: (AL happy – EO happy) versus (AL fear – EO fear). The peak activity for this medial OFC activity is located at (18, 42, –8) ( $t_{(11)} = 5.1$ ) and is shown in Figure 3*a*. Other brain regions that were more active for object–happy association learning include the following: right middle frontal gyrus, anterior portion located at Brodmann's area (BA

10, (46, 40, 22),  $t_{(11)} = 5.2$ ; posterior portion located at BA 9, (36, 24, 34),  $t_{(11)} = 6.5$ ; and the right intraparietal sulcus (IPS) located at (30, –40, 46),  $t_{(11)} = 6.2$ . There were no brain regions that were more responsive to object–fear learning compared with object–happy learning when controlling for the emotion-only faces.

*New object–emotion association learning compared with maintenance of object–emotion associations*

To investigate the neural activity involved in the initial acquisition of an object–emotion association compared with the maintenance of that object–emotion association, we compared AL emotion [reversal (fear + happy)] versus AL emotion [maintenance (fear + happy)]. There was no significant difference in amygdala activity between reversal trials, indicating a new object–emotion association, and maintenance trials, indicating the maintenance of the object–emotion association. However, there were robust activations in other brain regions, and these are listed in Table 2 and shown in Figure 3*b*. Most notably,

reversal trials preferentially activated the ventral and dorsal PFC as well as the ventral and dorsal striatum (i.e., ventral putamen extending to dorsal caudate) and the supplementary motor area (SMA). This has been shown previously in reversal learning tasks (Cools et al., 2002; O'Doherty et al., 2003a). In addition, we found greater activity for reversal learning compared with maintenance in the bilateral IPS, bilateral fusiform gyrus, and right STS. There was greater activity in the posterior hippocampus and posterior insula for the maintenance of object–emotion associations.

*General learning effects*

*Neutral association learning compared with perception of neutral faces.* Based on previous research, we predicted that the amygdala would be specifically involved in emotional association learning rather than neutral association learning more generally. To test this, we compared object–neutral association learning with neutral faces presented alone (AL neutral vs EO neutral). As predicted, we found no significant amygdala activity in this contrast. This analysis confirms amygdala specificity in object–emotion association learning. However, there were other brain regions responsive to neutral association learning. These are listed in Table 3 and include the DLPFC, IPS, fusiform gyrus, STS, hippocampus, and dorsal striatum (putamen). It is also noteworthy that we did not find VLPFC or thalamic activity for neutral association learning, although these regions were active for emotional learning.

*Association learning (all) versus expression only (all).* In addition, we compared all association learning trials (AL fear + AL happy + AL neutral) to all expression-only trials (EO fear + EO happy + EO neutral). This analysis shows the general effect of learning from faces while subtracting the effect of face perception. This contrast reveals the same brain regions as the contrast of AL emotion versus EO emotion with several specific exceptions. Most importantly, there was no significant amygdala or ventrolateral PFC activity in the contrast AL all versus EO all, suggesting that the addition of neutral association learning weak-

**Table 3. Brain regions involved in object association learning from neutral expressions compared with the perception of neutral expressions: AL neutral versus EO neutral**

Brain region	Brodman's area	MNI <i>x, y, z</i> (peak)	<i>t</i> statistic
<b>AL neutral &gt; EO neutral</b>			
Middle frontal gyrus, posterior, right	9	50, 14, 44	4.1
Intraparietal sulcus, right	7	56, -36, 46	4.7
Fusiform gyrus, right	37	44, -66, -22	5.4
Superior temporal sulcus, right	22	54, -56, 14	4.7
Hippocampus, left		-32, -22, -2	4.2
Putamen, left		-26, 0, 2	4.6
<b>EO neutral &gt; AL neutral</b>			
Precentral gyrus, left	4, 6	-48, 4, 28	7.2
Postcentral gyrus, right	3	58, -14, 28	4.3
Postcentral gyrus, left	3	-36, -12, 48	4.2
Middle cingulate, right	23, 24	8, -12, 46	4.6
Middle cingulate, left	24	-6, -22, 46	5.6
Posterior insula, right	48	38, -24, -12	5.6
Posterior insula, left	48	-34, -24, 16	4.6
Cuneus, calcarine sulcus, right	23, 31	24, -56, 18	5.4
Cuneus, calcarine sulcus, left	23	-22, -58, 22	4.5

ened activity in these regions. Also, the addition of neutral trials revealed significant activity for AL all in the right IPS and the left putamen. These regions were also active for neutral association learning (AL neutral vs EO neutral) but were not apparent in the contrast AL emotion versus EO emotion. The full list of brain activations for the general association learning in the contrast AL all versus EO all is listed in supplemental Table 1 (available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material).

## Discussion

We tested the hypothesis that the amygdala is engaged in analyzing emotional facial expressions for the purpose of learning emotional associations. We found significantly greater right amygdala activity when subjects learned object–emotion associations from fearful and happy expressions compared with the perception of fearful and happy expressions without any learning requirement. There was no evidence that this enhanced amygdala activity was driven by one type of object–emotion learning more than the other, supporting amygdala engagement when learning associations about both threat and reward value (Gottfried et al., 2002; Holland and Gallagher, 2004). Importantly, the amygdala was not preferentially active for learning object–neutral associations compared with perceiving neutral faces presented alone, confirming amygdala involvement in object–emotion association learning specifically and not neutral association learning more generally. This specificity in emotional learning is also supported by behavioral data showing faster predictions of happy and fearful emotional reactions compared with neutral reactions to the object.

These data provide direct evidence that the amygdala is engaged in analyzing facial expressions for the purpose of learning an association regarding a potentially threatening or rewarding stimulus. This supports the notion that facial expressions have a functional role for the observer (Keltner and Kring, 1998) and that the amygdala is the primary brain structure that uses informative social signals to facilitate goal-directed behavior and ultimately enhance survival. This interpretation is consistent with evidence showing that the amygdala is more involved in learning stimulus–reinforcement contingencies than processing primary reinforcement. For example, amygdala lesions impair the acqui-

sition of a conditioned response to a predictive cue but do not impair physiological response to the unconditioned stimuli (LaBar et al., 1995). We extend these findings by showing greater amygdala activity when learning the predictive relationship between an object and an emotional expression than for the perception of the emotional expression presented alone.

Although it has never been demonstrated previously, association learning has been discussed as an explanation of differential amygdala involvement in processing emotional faces presented alone (Whalen, 1998). For example, the amygdala responds more to a fearful face than an angry face. The interpretation is that an angry expression signals the presence of danger (the angry person) as well as its source (the observer), whereas a fearful expression is more ambiguous because it signals the presence of danger but not the source. Therefore, more amygdala activity is recruited to learn the association between the fearful expression and its source (Whalen et al., 2001). Furthermore, amygdala habituation after prolonged or repeated exposure of the same emotional face can be explained as initial activity engaged to learn an association between the emotion and provoking stimulus, but, without the opportunity to create the association, amygdala response decreases (Wright et al., 2001; Fischer et al., 2003).

Our paradigm uses a new approach and reveals data consistent with these previous interpretations. Rather than comparing different levels of assumed ambiguity about the source of threat or reward, we compare different levels of learning about the source. Furthermore, habituation patterns as well as face processing research suggest that amygdala response to multiple fearful faces presented alone is most likely a reflection of learning activity engaged by the novelty of a different face identity presented on every trial (Schwartz et al., 2003; Glascher et al., 2004). In our paradigm, we isolate learning and control for confounding variables by keeping face identity and emotional content constant but changing the emotional value (i.e., the association) of the provoking stimulus. Subjects have to use facial expressions to learn about the threat and reward value of the object. We compare this with a condition in which all aspects of the facial stimuli are identical to the learning condition, but there is no opportunity to use facial expressions to learn. The comparison reveals more amygdala activity for learning object–emotion associations compared with perceiving emotional faces, supporting the idea that amygdala response to facial expressions is driven by association learning.

However, the exact neurophysiological mechanism the amygdala uses to accomplish emotional association learning (from facial expressions or primary reinforcement) is still not clear. One view is that amygdala activity specifically binds a cue with subsequent reinforcement and thus is the site for acquisition of emotional associations (LeDoux, 1993). An alternative view is that the amygdala responds to signals, such as facial expressions, that a biologically relevant stimulus is in the environment, and this amygdala activity increases physiological vigilance (i.e., lowers neuronal firing thresholds), which enhances object–emotion association coding in other brain regions, such as the OFC and the dorsal and ventral striatum (Gallagher and Holland, 1994; Whalen, 1998; Holland and Gallagher, 1999; Davis and Whalen, 2001). These views are not mutually exclusive, because evidence suggests that each process may be mediated through different nuclei within the amygdala (Holland and Gallagher, 1999).

Our data are consistent with the idea that the amygdala facilitates learning object–emotion associations by using the emotional signal to increase vigilance and prepare learning networks to acquire new associations. Greater activity for the reversal cue,

which is the initial coding of the emotional association, versus the maintenance cue would provide evidence that amygdala neurons are specifically coding the new association. Instead, we found greater amygdala activity in the combination of reversal and maintenance object–emotion learning trials versus emotional faces alone, but we did not find greater amygdala activity for reversal versus maintenance trials. This is likely because the rapid reversals in our design required subjects to anticipate possible changes in reinforcement contingencies while continuing to strengthen associations during maintenance trials. In support of this interpretation, we found enhanced amygdala activity for all object–emotion learning trials, suggesting that the amygdala remained engaged during maintenance to strengthen emotional associations and/or prepare for impending reversals. In addition, the hippocampus, important for the declarative knowledge of stimulus–reinforcement contingencies (Phelps, 2004; Carter et al., 2006), was active during the maintenance period, suggesting that newly acquired associations were still being solidified. Furthermore, neural activity revealed for reversal trials replicates a network of regions previously identified as important in reversing reinforcement-related contingencies (Cools et al., 2002; O'Doherty et al., 2003a, 2004), including the ventral and dorsal PFC, lateral OFC, ventral and dorsal striatum, and SMA. Our findings that the amygdala is involved in learning but not specifically active for reversals is consistent with data showing that the OFC is more critically important for reversing reinforcement associations (Rolls et al., 1994; Schoenbaum et al., 2003b; Hornak et al., 2004) and that the flexible representation of reinforcement value depends on the interaction of the OFC and amygdala (Schoenbaum et al., 2003a).

However, although our data support the vigilance hypothesis (Whalen, 1998), our task was not optimized to investigate different phases of emotional learning, and therefore we cannot rule out the idea that the amygdala is specifically involved in the acquisition object–emotion relationships. Namely, to identify an enhancement of amygdala activity during acquisition at the reversal cue, the object association during the maintenance period must be so established that it no longer requires active learning. This was not the case in our task.

Interestingly, none of the previous reversal learning studies with human subjects identify greater amygdala activity for reversal compared with maintenance trials (Cools et al., 2002; Kringelbach and Rolls, 2003; O'Doherty et al., 2003a; Morris and Dolan, 2004). Nonetheless, different nuclei within the amygdala may mediate these different aspects of emotional learning (Buchel et al., 1998, 1999; Morris and Dolan, 2004; Phelps et al., 2004), and spatial resolution of fMRI is not fine enough to consistently distinguish between regions. Alternatively, in the context of reversal learning, the amygdala may remain active in anticipation of a change in reinforcement contingencies.

Regardless of which specific process the amygdala uses to learn object–emotion associations, our study shows that the amygdala is engaged in analyzing facial expressions for the purpose of learning and is not simply responding to the reinforcement value of a positive or negative face. This role of the amygdala in emotional learning during the analysis of facial expressions may explain inconsistencies in the patient literature. Other data show that amygdala lesion patients have normal expression of emotion and have inconsistent deficits in the recognition of emotional facial expressions in others (Anderson and Phelps, 2000). Not all amygdala lesion patients have problems recognizing emotional expressions; however, if they do show deficits, it is most likely an impairment in understanding the import

of a fearful facial expression (Adolphs et al., 1994). This finding is consistent within the context of emotional learning, because fear is a salient predictor of an impending threat. Furthermore, emotion recognition impairment in patients is most likely to occur after a congenital or early acquired lesion (Hamann et al., 1996; Meletti et al., 2003). One explanation of these findings is that these patients did not have a healthy amygdala to help them learn the predictive value of emotional expressions.

Although we found the effect of object–emotion learning in the right amygdala, it is unlikely that emotional learning is a lateralized function, because previous research has shown that unilateral left temporal lobectomy patients have fear learning deficits (LaBar et al., 1995). Nonetheless, our findings are consistent with fMRI studies showing greater right than left amygdala activity in fear learning (LaBar et al., 1998; Pine et al., 2001; Pegna et al., 2005) and neuropsychological studies showing greater emotion recognition deficits after right amygdala lesions (Anderson et al., 2000; Meletti et al., 2003; Benuzzi et al., 2004). Amygdala laterality may be influenced by task stimuli, with the left amygdala more active in emotional learning tasks using verbal instruction (Phelps et al., 2001) or verbal context (Kim et al., 2004), so our use of nonverbal facial stimuli may have preferentially involved the right amygdala. Another possibility, consistent with our data, is that the right amygdala may be more engaged when trying to learn the association causing a facial expression (Kim et al., 2003) as opposed to having that association already determined (Kim et al., 2004).

## References

- Adams RBJ, Gordon HL, Baird AA, Ambady N, Kleck RE (2003) Effects of gaze on amygdala sensitivity to anger and fear faces. *Science* 300:1536.
- Adolphs R, Tranel D, Damasio H, Damasio A (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372:669–672.
- Anderson AK, Phelps EA (2000) Expression without recognition: contributions of the human amygdala to emotional communication. *Psychol Sci* 11:106–111.
- Anderson AK, Spencer DD, Fulbright RK, Phelps EA (2000) Contribution of the anteromedial temporal lobes to the evaluation of facial emotion. *Neuropsychology* 14:526–536.
- Benuzzi F, Meletti S, Zamboni G, Calandra-Buonaura G, Serafini M, Lui F, Baraldi P, Rubboli G, Tassinari CA, Nichelli P (2004) Impaired fear processing in right mesial temporal sclerosis: a fMRI study. *Brain Res Bull* 63:269–281.
- Blair RJ (2003) Facial expressions, their communicatory functions and neuro-cognitive substrates. *Philos Trans R Soc Lond B Biol Sci* 358:561–572.
- Buchel C, Morris J, Dolan RJ, Friston KJ (1998) Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron* 20:947–957.
- Buchel C, Dolan RJ, Armony JL, Friston KJ (1999) Amygdala–hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. *J Neurosci* 19:10869–10876.
- Canli T, Sivers H, Whitfield SL, Gotlib IH, Gabrieli JD (2002) Amygdala response to happy faces as a function of extraversion. *Science* 296:2191.
- Carter RM, O'Doherty JP, Seymour B, Koch C, Dolan RJ (2006) Contingency awareness in human aversive conditioning involves the middle frontal gyrus. *NeuroImage* 29:1007–1012.
- Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J Neurosci* 22:4563–4567.
- Davis M, Whalen PJ (2001) The amygdala: vigilance and emotion. *Mol Psychiatry* 6:13–34.
- Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage* 19:430–441.
- Ekman P, Friesen WV (1978) The facial action coding system (FACS): a technique for the measurement of facial action. Palo Alto, CA: Consulting Psychologists.

- Fischer H, Wright CI, Whalen PJ, McInerney SC, Shin LM, Rauch SL (2003) Brain habituation during repeated exposure to fearful and neutral faces: a functional MRI study. *Brain Res Bull* 59:387–392.
- Gallagher M, Holland PC (1994) The amygdala complex: multiple roles in associative learning and attention. *Proc Natl Acad Sci USA* 91:11771–11776.
- Glascher J, Tuscher O, Weiller C, Büchel C (2004) Elevated responses to constant facial emotions in different faces in the human amygdala: an fMRI study of facial identity and expression. *BMC Neurosci* 5:45.
- Gottfried JA, O'Doherty J, Dolan RJ (2002) Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J Neurosci* 22:10829–10837.
- Hamann SB, Stefanacci L, Squire LR, Adolphs R, Tranel D, Damasio H, Damasio (1996) A recognizing facial emotion. *Nature* 379:497.
- Holland PC, Gallagher M (1999) Amygdala circuitry in attentional and representational processes. *Trends Cogn Sci* 3:65–73.
- Holland PC, Gallagher M (2004) Amygdala-frontal interactions and reward expectancy. *Curr Opin Neurobiol* 14:148–155.
- Hooker CI, Paller KA, Gitelman DR, Parrish TB, Mesulam MM, Reber PJ (2003) Brain networks for analyzing eye gaze. *Brain Res Cogn Brain Res* 17:406–418.
- Hornak J, O'Doherty J, Bramham J, Rolls ET, Morris RG, Bullock PR, Polkey CE (2004) Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *J Cogn Neurosci* 16:463–478.
- Keltner D, Kring A (1998) Emotion, social function, and psychopathology. *Rev Gen Psychol* 2:320–342.
- Kim H, Somerville LH, Johnstone T, Alexander AL, Whalen PJ (2003) Inverse amygdala and medial prefrontal cortex responses to surprised faces. *NeuroReport* 14:2317–2322.
- Kim H, Somerville LH, Johnstone T, Polis S, Alexander AL, Shin LM, Whalen PJ (2004) Contextual modulation of amygdala responsivity to surprised faces. *J Cogn Neurosci* 16:1730–1745.
- Kringelbach ML, Rolls ET (2003) Neural correlates of rapid reversal learning in a simple model of human social interaction. *NeuroImage* 20:1371–1383.
- LaBar KS, LeDoux JE, Spencer DD, Phelps EA (1995) Impaired fear conditioning following unilateral temporal lobectomy in humans. *J Neurosci* 15:6846–6855.
- LaBar KS, Gatenby JC, Gore JC, LeDoux JE, Phelps EA (1998) Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* 20:937–945.
- LeDoux JE (1993) Emotional memory systems in the brain. *Behav Brain Res* 58:69–79.
- LeDoux JE (2000) Emotion circuits in the brain. *Annu Rev Neurosci* 23:155–184.
- Meletti S, Benuzzi F, Nichelli P, Tassinaria CA (2003) Damage to the right hippocampal-amygdala formation during early infancy and recognition of fearful faces: neuropsychological and fMRI evidence in subjects with temporal lobe epilepsy. *Ann NY Acad Sci* 1000:385–388.
- Morris JS, Dolan RJ (2004) Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. *NeuroImage* 22:372–380.
- O'Doherty J, Critchley H, Deichmann R, Dolan RJ (2003a) Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J Neurosci* 23:7931–7939.
- O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ (2003b) Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41:147–155.
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
- Pegna AJ, Khateb A, Lazeyras F, Seghier ML (2005) Discriminating emotional faces without primary visual cortices involves the right amygdala. *Nat Neurosci* 8:24–25.
- Phelps EA (2004) Human emotion and memory: interactions of the amygdala and hippocampal complex. *Curr Opin Neurobiol* 14:198–202.
- Phelps EA, O'Connor KJ, Gatenby JC, Gore JC, Grillon C, Davis M (2001) Activation of the left amygdala to a cognitive representation of fear. *Nat Neurosci* 4:437–441.
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004) Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43:897–905.
- Pine DS, Fyer A, Grun J, Phelps EA, Szeszko PR, Koda V, Li W, Ardekani B, Maguire EA, Burgess N, Bilder RM (2001) Methods for developmental studies of fear conditioning circuitry. *Biol Psychiatry* 50:225–228.
- Rolls ET, Hornak J, Wade D, McGrath J (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J Neurol Neurosurg Psychiatry* 57:1518–1524.
- Schoenbaum G, Setlow B, Saddoris MP, Gallagher M (2003a) Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* 39:855–867.
- Schoenbaum G, Setlow B, Nugent SL, Saddoris MP, Gallagher M (2003b) Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. *Learn Mem* 10:129–140.
- Schwartz CE, Wright CI, Shin LM, Kagan J, Whalen PJ, McMullin KG, Rauch SL (2003) Differential amygdalar response to novel versus newly familiar neutral faces: a functional MRI probe developed for studying inhibited temperament. *Biol Psychiatry* 53:854–862.
- Whalen PJ (1998) Fear, vigilance, and ambiguity: initial neuroimaging studies of the human amygdala. *Curr Dir Psychol Sci* 7:177–188.
- Whalen PJ, Shin LM, McInerney SC, Fischer H, Wright CI, Rauch SL (2001) A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion* 1:70–83.
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996) A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4:58–73.
- Wright CI, Fischer H, Whalen PJ, McInerney SC, Shin LM, Rauch SL (2001) Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. *NeuroReport* 12:379–383.
- Zald DH (2003) The human amygdala and the emotional evaluation of sensory stimuli. *Brain Res Brain Res Rev* 41:88–123.