doi:10.1093/scan/nsu085 SCAN (2015) 10, 537–544

Prediction errors to emotional expressions: the roles of the amygdala in social referencing

Harma Meffert,[1] Sarah J. Brislin,[2] Stuart F. White,[1] and James R. Blair[1]

[1]Section of Affective and Cognitive Neuroscience, National Institutes of Health, Bethesda, MD 20892, and [2]Clinical Psychology Program, Florida State University, Tallahassee, FL 32306

Social referencing paradigms in humans and observational learning paradigms in animals suggest that emotional expressions are important for communicating valence. It has been proposed that these expressions initiate stimulus-reinforcement learning. Relatively little is known about the role of emotional expressions in reinforcement learning, particularly in the context of social referencing. In this study, we examined object valence learning in the context of a social referencing paradigm. Participants viewed objects and faces that turned toward the objects and displayed a fearful, happy or neutral reaction to them, while judging the gender of these faces. Notably, amygdala activation was larger when the expressions following an object were less expected. Moreover, when asked, participants were both more likely to want to approach, and showed stronger amygdala responses to, objects associated with happy relative to objects associated with fearful expressions. This suggests that the amygdala plays two roles in social referencing: (i) initiating learning regarding the valence of an object as a function of prediction errors to expressions displayed toward this object and (ii) orchestrating an emotional response to the object when value judgments are being made regarding this object.

Keywords: social referencing; amygdala; fMRI; emotion; neuroscience; learning

INTRODUCTION

Emotional expressions play a critical role in the transmission of valence information between conspecifics. In humans, this is seen in the context of social referencing paradigms (Klinnert *et al.*, 1986; Aktar *et al.*, 2013), where emotional expressions of the caregiver to novel objects influence the child's approach/avoidance responses to these objects (i.e. if the caregivers smile toward the object, children are more likely to approach, while if they show fear, children are more likely to avoid the object). Moreover, it has been suggested that social referencing contributes to the early learning of anxiety, as infants of anxious parents get more exposure to expressions of parental anxiety (Muris *et al.*, 1996; Hudson and Rapee, 2001; Fisak and Grills-Taquechel, 2007; Murray *et al.*, 2009). In animal work, this is seen in the context of observational learning paradigms (Mineka and Cook, 1993), where fearful displays by the mother result in the infant monkey displaying fear to the object.

The suggestion is that emotional expressions can initiate stimulus-reinforcement learning (Blair, 2003), where the emotional expression is associated with the novel object. The amygdala has been implicated in aversive and appetitive stimulus-reinforcement learning (e.g. Everitt *et al.*, 2000; LeDoux, 2000; Murray, 2007; Tye *et al.*, 2010). This suggests that the amygdala might be importantly involved in stimulus-reinforcement learning in response to emotional expressions such as fear and happiness (cf. Hooker *et al.*, 2006) and sadness (Blair, 2003). Certainly, work indicates that particularly fearful (e.g. Adolphs, 2010) but also happy expressions (e.g. Fusar-Poli *et al.*, 2009; N'Diaye *et al.*, 2009) elicit amygdala activity. Moreover, recent animal work has shown that amygdala lesions block the acquisition and expression of observational fear (Jeon *et al.*, 2010).

However, work on the reinforcing value of emotional expressions remains in its infancy. Hooker *et al.* (2006) showed increased amygdala responses to both happy and fearful expressions when eye gaze was directed toward objects rather than toward empty space, consistent with a role of this structure in learning object valence from expression information. Moreover, two instrumental learning studies showed (i) that striatal activity was greater to happy relative to sad expressions, when these acted as response reinforcers (Scott-Van Zeeland *et al.*, 2010) and (ii) that striatal activity to happy expressions was modulated by prediction error (PE; Lin *et al.*, 2012). PEs reflect the difference between received and expected reinforcement, and signal an update in stimulus value (Rescorla and Wagner, 1972).

The goal of the current study was to use model-based functional magnetic resonance imaging (fMRI; cf. O'Doherty *et al.*, 2007) to determine regions sensitive to PE and stimulus value information regarding emotional expressions, during a social referencing paradigm. Given previous work, we predicted amygdala responsiveness to fear (and possibly happy) expression PE signaling and striatal responsiveness to happy expression PE signaling. Given findings of ventromedial prefrontal cortex (vmPFC) responsiveness to stimulus value (association with a happy expression, Lin *et al.*, 2012), we predicted vmPFC responsiveness to the stimulus value of objects according to their association with happy (and possibly fearful) expressions.

METHODS

Participants

Thirty healthy adult volunteers were recruited from the community through newspaper ads and fliers. Subjects were in good physical health as confirmed by a complete physical exam, with no history of any psychiatric illness as assessed by the Diagnostic and Statistical Manual for Mental Disorders IV (DSM-IV, Association, 1994) criteria based on the Structural Clinical Interview for DSM-IV Axis I disorders (First *et al.*, 1997). All subjects gave written informed consent to participate in the study, which was approved by the National Institute of Mental Health Institutional Review Board. Two subjects were excluded because of problems with preprocessing the MRI data. Five subjects were excluded because of their below average behavioral performance. First, one participant failed to respond to 95% of the trials and was removed from further analysis. Of the remaining participants, two were outliers (>2 s.d. above the group mean) in number of missed

Received 30 August 2013; Revised 1 May 2014; Accepted 13 June 2014
Advance Access publication 17 June 2014

This work was supported by the Intramural Research Program of the National Institute of Mental Health, National Institutes of Health under grant number 1-ZIA-MH002860-08.

Correspondence should be addressed to Harma Meffert, National Institutes of Health, National Institute of Mental Health, Section of Affective and Cognitive Neuroscience, 9000 Rockville Pike, Building 15k, Room 300-E, MSC 2670, Bethesda, MD 20814, USA. E-mail: harma.meffert@nih.gov.

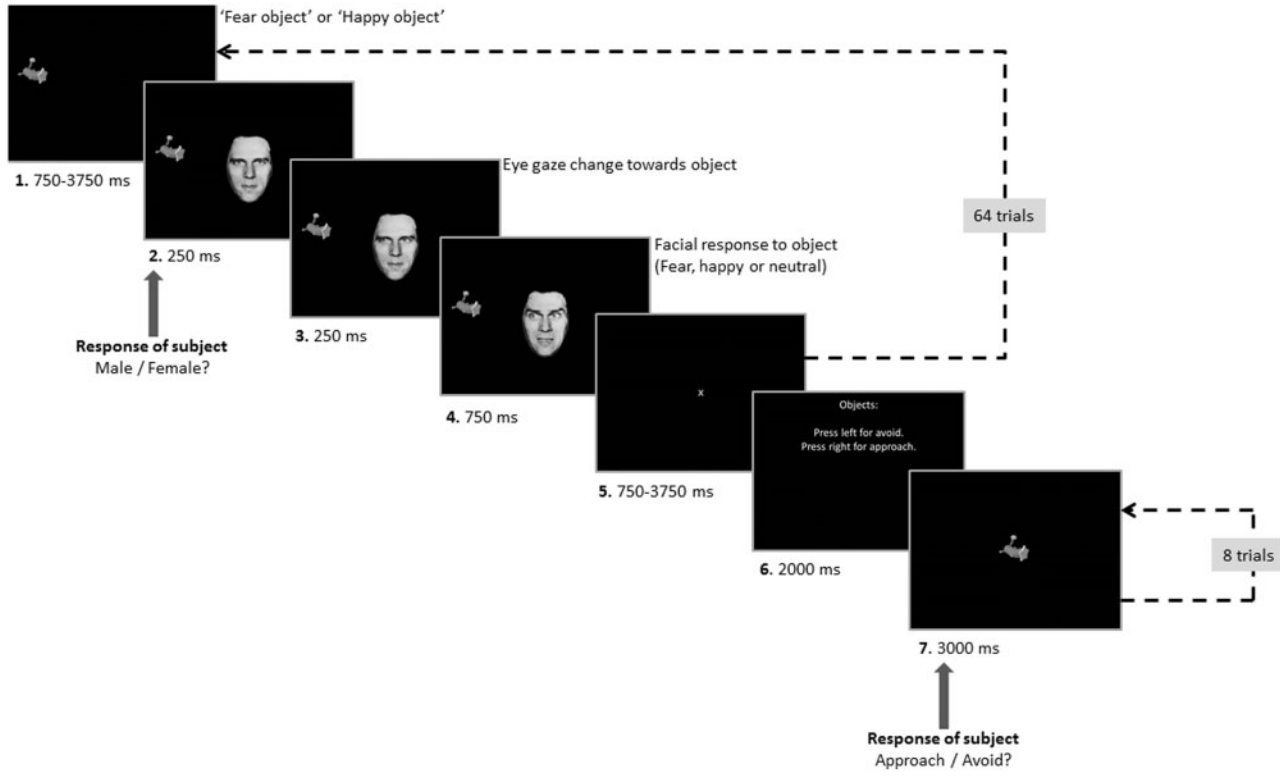

Fig. 1 Task design. During the first part of each run, participants performed gender judgments on Ekman faces (64 trials). Facial expressions were probabilistically related to each object. During the second part of each run, participants evaluated whether they would approach or avoid each of the presented objects (each object was presented twice).

responses, and two were outliers in terms of inaccuracies in gender judgment (>2 s.d. above the group mean). As such the data from 23 subjects were analyzed (48% female; average age 26.91 ± 4.69 years). IQ was assessed with the Wechsler Abbreviated Scale of Intelligence (two-subtest form, Wechsler, 1999); average IQ = 118.96 (s.d. = 10.53).

Experimental design

Each trial began with the presentation of an object on the left or right side of the screen (see Figure 1). After a jittered interval (750–3750 ms), a neutral face with an eye gaze oriented toward the participant appeared in the middle of the screen. After 250 ms, the eye gaze of the displayed face shifted toward the object. Another 250 ms later, the facial expression changed to fearful or happy or remained neutral. After 750 ms, the object and the face were removed from display, and a fixation cross appeared in the middle of the screen. There was then a jittered interval (750–3750 ms) before the next trial began. Participants were instructed to judge the gender of the face as quickly and accurately as possible. Participants had to respond within 750 ms after onset of the face, otherwise the trial was recorded as a missed trial.

Eight objects were selected from http://www.tarrlab.org/ (stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University) and converted to gray scale using Adobe Photoshop. In addition, four faces from the Ekman test battery were selected (MO, NR, PE and WF), two male and two female. Eye gaze in these pictures was manipulated using Photoshop. Stimuli were presented using E-Prime.

Participants completed four task runs in total, each containing 64 trials. Each run was presented twice and used four of the eight objects and three of the Ekman faces. During each run, two of the objects could elicit fearful expressions (i.e. 'fear' objects), while the other two could elicit happy facial expressions (i.e. 'happy' objects). Every actor was used an equal number of times for 'happy' and 'fear' objects. However, the facial response to an object was probabilistic. Two of the objects (one for fearful and the other happy) elicited an emotional reaction on 75% of the trials (high 'fear' and high 'happy' objects), while the other two objects elicited an emotional reaction on 25% of the trials (low 'fear' and low 'happy' objects). On the remaining trials, objects generated a neutral response. The final part of each run involved the presentation of each of the four objects shown during that run twice. During this testing phase of the run, the participant was asked to state, via button press, whether he or she would approach or avoid the object. Each presentation was for 3000 ms, and there were an additional four fixation trials of 1000 ms. Thus, over the course of the study, the participant judged 16 'fear' and 16 'happy' objects.

On the basis of the participant's reinforcement history, a learning curve was modeled establishing expected values (EVs) and PEs for each object. The EV of an object represents the expectation a participant has that this object will be followed by a certain facial response and ranges from 0 (maximal expectation of a neutral face) to 1 (maximal expectation of an emotional face). On individual trials, the EV associated with objects lies somewhere between 0 and 1. The PE for an object represents the deviation from that expectation and ranges from −1 (maximal negative deviation, i.e. getting a neutral face when there was maximal expectation of getting an emotional face) to 1 (maximal positive deviation, i.e. getting an emotional face when there was maximal expectation of getting a neutral face).

The EV for the first trial of each object was 0. PE was then calculated based on the feedback (F), which was coded 1 (e.g. 'happy' object

eliciting a happy expression) or 0 (e.g. 'happy' object eliciting a neutral expression) with the formula:

$$PE_{(t)}=F_{(t)}-EV_{(t)}$$

where the PE for the current trial (t) equaled the feedback value for the current trial minus the EV for the current trial. EV was calculated via the following formula:

$$EV_{(t)}= EV_{(t-1)}+(\alpha^{*}PE_{(t-1)})$$

where the EV of the current trial (t) equals the EV of the previous trial (t−1) plus the PE of the previous trial multiplied by the learning rate (α). The overall learning rate α was calculated as follows. First, model-based end-of-run EVs were calculated per object as a function of learning rate, where learning rate was iterated from 0.001 to 0.999 in steps of 0.001. Second, participant-based end-of-run 'EVs' were calculated per object based on their choice data, by dividing the number of avoid selections for that object by the total number of selections. Third, participant-based 'EVs' were correlated with model-based EVs for each learning rate. These correlations were squared, and the highest R^2 was determined in order to find the best learning rate for that particular object and run. These learning rates were then averaged across objects, which yielded an average learning rate of 0.474.

Image acquisition

A total of 152 functional images per run were taken with a gradient echo planar imaging (EPI) sequence (repetition time = 2900 ms; echo time = 27 ms; 64 × 64 matrix; 90° flip angle; 24 cm field of view). Whole-brain coverage was obtained with 46 axial slices (thickness, 2.5 mm with 0.5 mm spacing; in-plane resolution, 3.75 mm × 3.75 mm) using 3.0 T GE Signa Scanner. A high-resolution anatomical scan (three-dimensional spoiled gradient recalled acquisition in a steady state; 3.0 T: repetition time = 7 ms; echo time = 2.984 ms; 24 cm field of view; 12° flip angle; 128 axial slices; thickness, 1.2 mm; 256 × 192 matrix) in register with the EPI dataset was obtained covering the whole brain.

Image processing

Data were analyzed within the framework of the general linear model using Analysis of Functional Neuroimages (AFNI; Cox, 1996). Both individual- and group-level analyses were conducted. The first five volumes in each scan series, collected before equilibrium magnetization was reached, were discarded. Motion correction was performed by registering all volumes in the EPI dataset to a volume collected close to acquisition of the high-resolution anatomical dataset.

The EPI datasets for each participant were spatially smoothed (isotropic 6 mm kernel) to reduce variability among individuals and generate group maps. Next, the time series data were normalized by dividing the signal intensity of a voxel at each point by the mean signal intensity of that voxel for each run and multiplying the result by 100. This means that the resultant regression coefficients at the stage of implementing the linear model will represent a percentage of signal change from the mean. The participants' anatomical scans were individually registered to the Talairach and Tournoux atlas (Talairach and Tournoux, 1988). The individuals' functional EPI data were then registered to their Talairach anatomical scan within AFNI.

Following this, four indicator regressors were generated for objects that could elicit fear, objects that could elicit happiness, expression feedback for fear objects and expression feedback for happy objects. Four additional regressors were created by parametrically modulating the first two indicator regressors by the EV for the trial and the second two indicator regressors by the PE for the trial. Two additional regressors were created by parametrically modulating the second two indicator regressors (for expression feedback) by the reaction time of the gender judgment. Finally, two indicator regressors were created for the onset of rating the approach/avoidance to 'fear' objects and 'happy' objects. All regressors were created by convolving the train of stimulus events with a gamma variate hemodynamic response function to account for the slow hemodynamic response. None of the regressors were orthogonalized with respect to another regressor. Linear regression modeling was performed using the 12 regressors described above plus regressors to model a first-order baseline drift function. This produced a β coefficient and associated t statistic for each voxel and regressor.

fMRI data analysis

Group analysis was performed on the modulated contrasts using t-tests (AFNI's 3dttest++). Two one-sample t-tests were conducted to assess regions that showed a modulation of the Blood Oxygenation Level Dependent (BOLD) response by PE, separately for facial expressions following 'fear' and 'happy' objects, followed by a two-sample paired t-test to examine for which regions the PE-modulated BOLD response differed for facial expressions following 'fear' and 'happy' objects. In addition, a conjunction analysis was conducted to assess which regions were similarly modulated by PE for both the facial expressions following 'fear' and 'happy' objects. Finally, a two-sample paired t-test was conducted to examine for which regions the PE-modulated BOLD response differed for facial expressions following 'fear' compared with 'happy' objects.

Two one-sample t-tests were then conducted to assess regions that showed a modulation of the BOLD response by EV, separately for objects that could elicit fearfulness and objects that could elicit happiness. This was followed by a two-sample paired t-test to examine for which regions the EV-modulated BOLD response differed for 'fear' object trials compared with 'happy' object trials.

Finally, two one-sample t-tests were conducted to assess for which regions the BOLD was different compared with baseline during the testing phase, separately for rating 'fear' objects or 'happy' objects. A two-sample paired t-test was then conducted to examine for which regions the BOLD response differed during the testing phase between rating 'fear' objects compared with 'happy' objects.

Group-level analyses were masked using a whole-brain mask created in AFNI based on the mean normalized anatomical images of all subjects. Statistical maps were created for each analysis by thresholding at a single-voxel P-value of $P < 0.005$. To correct for multiple comparisons, we performed a spatial clustering operation using ClustSim with 10 000 Monte Carlo simulations taking into account the EPI matrix covering the gray matter. This procedure yielded a minimum cluster size (10 voxels) with a map-wise false-positive probability of $P < 0.05$, corrected for multiple comparisons. Given our a priori hypotheses, regions of interest (ROIs), were obtained for the amygdala, caudate and nucleus accumbens using AFNI software's Desai anatomical maps (Desikan *et al.*, 2006) for the caudate and nucleus accumbens and the CA_PM_18_MNIA anatomical maps for the amygdala (Amunts *et al.*, 2005; Eickhoff *et al.*, 2005). A small volume-corrected ROI analysis via ClustSim was used on this regions (initial threshold: $P < 0.005$ with minimum cluster sizes identified for each ROI at a corrected $P < 0.02$).

RESULTS

Behavioral results

Three two-sample paired t-tests compared 'fear' object trials with 'happy' object trials on gender judgment reaction times (RTs), number of missed trials and number of errors. Participants were significantly slower to judge gender on 'fear' relative to 'happy' object trials [$t(22) = 2.405$, $P = 0.025$; mean ('fear') = 687.51 ± 110.8 ms;

mean ('happy') = 678.4 ± 109.6 ms]. There were no differences in number of missed trials [$t(22) = 1.340$, $P = 0.194$] or error rate for 'fear' relative to 'happy' object trials [$t(22) = 0.241$, $P = 0.812$].

At the end of each run, subjects rated whether they wanted to approach or avoid each object that had been presented during the run. A ratio was calculated per object category (objects that could elicit fearful or happy expressions), which indicated the proportion of avoidance selections relative to the number of evaluations per object. A two-sample paired *t*-test revealed that participants were significantly more likely to avoid objects that could elicit fearful expressions (mean = 0.630 ± 0.27) than objects that could elicit happy expressions [mean = 0.34 ± 0.23); $t(22) = 3.397$, $P = 0.003$].

Neural correlates of PEs

The first goal of this study was to identify brain regions modulated by PE associated with fearful and happy expressions in the context of a social referencing task. An initial *t*-test revealed significant positive modulation of the BOLD response by the PE for fearful expressions within bilateral fusiform gyrus and right superior temporal sulcus (see Figure 2) and negative modulation in the dorsal premotor cortex [Table 1(a)]. A small volume correction (SVC) showed significant modulation of the BOLD response by the PE for fearful expressions within bilateral amygdala (see Figure 3) but not in striatum. In addition, a correlation analysis indicated that the modulated BOLD response for fearful expressions in the bilateral amygdala did not significantly correlate with the proportion of avoidance selections for 'fear' objects, respectively, at the individual level [left ($r = 0.147$, $P = 0.503$); right ($r = 0.100$, $P = 0.648$)].

With respect to happy expressions, there was a significant modulation of the BOLD response by the PE within left amygdala (see Figure 3), bilateral fusiform gyrus (see Figure 2), bilateral posterior temporal sulcus, bilateral dorsal premotor cortex, bilateral ventral lateral prefrontal cortex and dorsal medial prefrontal cortex [see Table 1(b)]. In all cases, modulation was positive; the more unexpected the happy expression, the greater the activity within these regions. An SVC showed significant modulation of the BOLD response by the PE for happy expressions within right amygdala as well (see Figure 3) but not in striatum. A correlation analysis indicated that the modulated BOLD response for happy expressions in one region within the left amygdala [Talairach coordinates (TC) = −16.5; −1.5; −9.5] correlated with the proportion of approach selections for 'happy' objects, respectively, at the individual level ($r = 0.514$, $P = 0.012$). The correlations in two other regions in left (TC = −28.5; −4.5; −15.5) and right amygdala (TC = 25.5; −1.5; −18.5) did not reach significance [left ($r = -0.029$, $P = 0.896$); right ($r = -0.034$, $P = 0.878$)].

Following this, a conjunction analysis was conducted to determine regions showing a common modulation to fearful and happy expressions as a function of PE. This revealed significant overlap within bilateral fusiform gyrus [see Table 1(c) and Figure 3 for the overlap between the red and blue regions].

Finally, a two-sample paired *t*-test was conducted to determine regions showing significantly differential sensitivity to PE for fearful and happy expressions. This revealed significantly stronger modulation of the BOLD response by PE for happy facial expressions relative to fearful facial expressions in bilateral middle occipital gyrus, bilateral lingual gyrus and the right dorsal premotor cortex [Table 1(d)].

Neural correlates of EV

The second goal of this study was to identify brain regions in which the BOLD signal correlated with the EV for objects. A *t*-test identified regions showing modulation of the BOLD response by the expectation that the object would engender a fearful expression. This revealed modulation of the BOLD response by the 'fearful EV' in bilateral middle temporal gyrus, see Table 2a. With respect to happy expressions, a *t*-test showed that regions in bilateral fusiform gyrus showed modulation of the BOLD response by 'happy EV', see Table 2b. In addition, a two-sample *t*-test revealed that left fusiform gyrus and right lingual gyrus were modulated stronger by 'happy EV' compared with 'fear EV', see Table 2c.

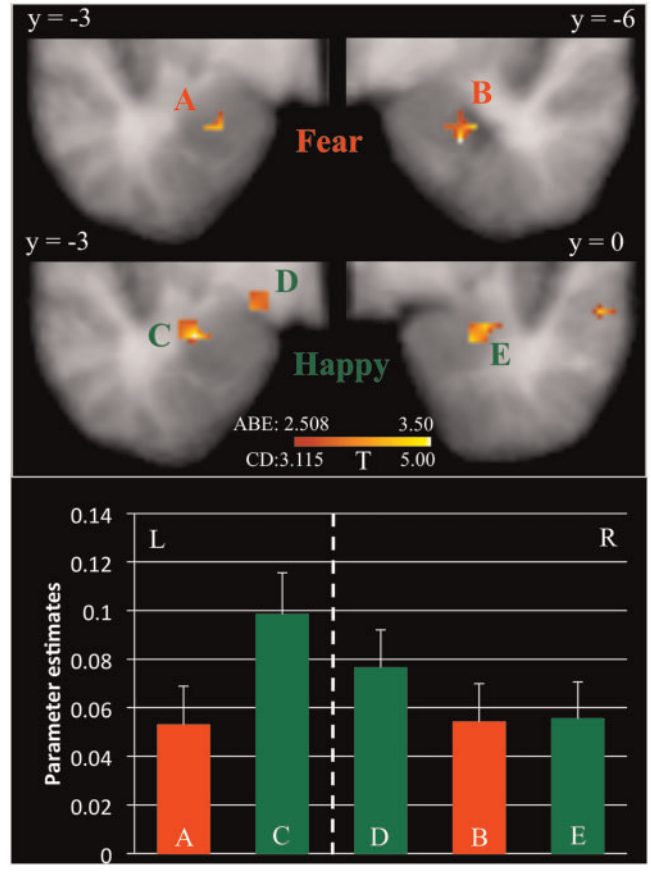

Fig. 2 BOLD response modulated by PE in bilateral fusiform gyrus. Results are depicted separately for the BOLD response modulated by PE for fearful facial expressions (red) and for happy facial expressions (blue). Regions in red are slightly transparent to illustrate the overlap between happy and fearful PE modulation. L = left; R = Right.

Neural correlates of object rating

Because of our interest in the involvement of the amygdala, the third objective of this paper was to assess its involvement in rating the objects. Using SVC, a *t*-test showed that the left amygdala was recruited while participants were rating 'happy' objects (TC = −25.5; −4.5; −9.5). The amygdala was not significantly recruited while participants rated 'fear' objects.

Following this, we correlated the BOLD response in left amygdala during 'happy' object rating against the proportion of avoidance selections for 'happy' objects, but this correlation failed to reach significance ($r = 0.340$; $P = 0.112$).

DISCUSSION

The present study investigated PE and EV signaling engendered by fearful and happy expressions, in the context of a variant of a social referencing task. There were five main findings: first, there was significant positive modulation of the BOLD response by PE for fearful expressions and by PE for happy expressions within bilateral amygdala.

Table 1 Brain regions modulated by PE (a) for fearful expressions, (b) for happy expressions, (c) for happy and fearful expressions and (d) for happy *vs* fearful expressions

Region[a]	Hemisphere	Brodmann's area	*t*	Coordinates of peak significance (*x y z*)			Voxels	Post hoc
(a) PE for fearful expressions								
Amygdala*	Right		3.744	25.5	−7.5	−21.5	8	+
Amygdala*	Left		3.248	−22.5	−4.5	−18.5	4	+
Fusiform gyrus—inferior temporal gyrus	Right	37/19	5.529	49.5	−64.5	−0.5	252	+
Fusiform gyrus	Left	37	4.277	−37.5	−46.5	−15.5	23	+
Inferior temporal gyrus	Left	37	5.114	−49.5	−70.5	5.5	76	+
Superior temporal sulcus	Right	22	4.018	52.5	−40.5	5.5	13	+
Precentral gyrus	Right	4	−4.315	37.5	−16.5	50.5	21	−
(b) PE for happy expressions								
Amygdala	Left		5.230	−28.5	−4.5	−15.5	19	+
Amygdala	Left		5.169	−16.5	−1.5	−9.5	9	+
Amygdala*	Right		3.374	25.5	−1.5	−18.5	7	+
Fusiform gyrus—middle occipital gyrus	Left	37/18	9.517	−25.5	−79.5	8.5	785	+
Fusiform gyrus—middle occipital gyrus	Right	37/18	8.408	49.5	−64.5	2.5	741	+
Superior temporal sulcus	Left	21	3.794	−52.5	−34.5	−0.5	15	+
Superior temporal sulcus	Right	41	5.463	58.5	−43.5	11.5	39	+
Precentral gyrus	Left	6	4.766	−40.5	−4.5	50.5	78	+
Precentral gyrus	Right	6	4.851	49.5	1.5	32.5	45	+
Postcentral gyrus	Right	1	4.075	46.5	−16.5	47.5	42	+
Inferior frontal gyrus	Right	44	4.350	37.5	10.5	26.5	25	+
Inferior frontal gyrus	Left	44	4.127	−40.5	10.5	23.5	17	+
Dorsal medial prefrontal cortex	Right	24	4.161	7.5	−4.5	47.5	14	+
Superior parietal lobe	Left	7	4.908	−28.5	−58.5	47.5	16	+
Temporal pole	Right	38	4.766	55.5	4.5	−12.5	15	+
(c) PE for fearful and happy expressions[b]								
Fusiform gyrus—inferior temporal gyrus	Right	37		44.7	−60.5	−3.4	130	n/a
Fusiform gyrus	Left	37		−36	−50.1	−14.2	10	n/a
Inferior temporal gyrus	Left	37		−43.5	−66.9	1.9	29	n/a
(d) PE for fearful *vs* happy expressions								
Middle occipital gyrus—cuneus	Left	18/19	−5.560	−13.5	−88.5	8.5	99	H>F
Middle occipital gyrus—cuneus	Right	19	−5.406	25.5	−82.5	17.5	53	H>F
Lingual gyrus	Left	18	−6.151	−19.5	−70.5	−9.5	69	H>F
Lingual gyrus	Left	19	−3.998	−7.5	−85.5	−3.5	10	H>F
Precentral gyrus—postcentral gyrus	Right	3/4	−4.815	40.5	−16.5	56.5	62	H>F

All results are thresholded at $P = 0.005$ uncorrected and a cluster extent threshold of 10 voxels (corresponding to map-wise false-positive probability of $P < 0.05$). Clusters marked with * survive an SVC at $P < 0.02$. ± indicates activation/deactivation compared with baseline. H>F indicates Happy larger Fear.
[a]The regions are according to the Talairach Daemon atlas (http://www.nitrc.org/projects/tal-daemon/).
[b]Coordinates are based on center of mass.

Second, the strength by which the BOLD response was modulated by the PE for happy expressions was associated with the proportion of approach selections for 'happy' objects. Third, there was significant modulation of the BOLD response by PE for both expressions in bilateral fusiform gyrus. Fourth, left amygdala was activated while participants evaluated whether they would approach or avoid objects associated with happy facial expressions. Fifth, participants were significantly more likely to avoid objects that could elicit fearful expressions relative to objects that could elicit happy expressions.

Social referencing paradigms in humans and observational fear paradigms in animals involve observers learning the valence of novel objects on the basis of other individual's emotional reactions to these objects (Klinnert *et al.*, 1986; Mineka and Cook, 1993; Bayliss *et al.*, 2007; Aktar *et al.*, 2013). The goal of the current study was to develop a form of social referencing paradigm suitable for fMRI. The current data indicate success in this goal; participants were significantly more likely to avoid objects that could elicit fearful expressions relative to objects that could elicit happy expressions.

Considerable literature suggests that the amygdala is responsive to fearful (e.g. Adolphs, 2010) and, to a lesser extent, happy expressions (e.g. Fusar-Poli *et al.*, 2009; N'Diaye *et al.*, 2009). In addition, considerable literature stresses the importance of the amygdala for stimulus-reinforcement learning (Baxter and Murray, 2002; Everitt *et al.*, 2003). It has been argued that the responsiveness of the amygdala to fearful and happy expressions reflects a role for this structure in learning the value of stimuli on the basis of these emotional expressions (Blair, 2003). Previous fMRI work had indicated that the amygdala plays a role in this process; amygdala responsiveness was greater to faces showing fearful and happy expressions toward objects, rather than empty space (Hooker *et al.*, 2006). Moreover, recent animal work has shown that the amygdala is critical for observational fear (Jeon *et al.*, 2010). However, these studies did not reveal the computational processes that the amygdala might be particularly sensitive to in its role in learning on the basis of emotional expressions. The current study indicates that the amygdala responds to the PE associated with a facial expression (at least a happy or fearful expression), displayed toward a novel object. Thus, the less the observer expects a particular emotional reaction to the object, the greater the PE after receiving such an emotional expression and the greater the amygdala response. The size of the PE is thought to determine speed of learning (Rescorla and Wagner, 1972). We assume that the greater the PE, the greater the learning by the amygdala of the valence associated with the novel object as a function of the social referencing context.

It is notable that there was a significant modulation of the BOLD response within the amygdala by the PE for both fearful and happy expressions. This is consistent with previous work showing amygdala involvement for negative as well as positive stimuli (e.g. Murphy *et al.*, 2003; Viding, 2012). Previous work has implicated the amygdala in both aversive and appetitive conditioning (e.g. Everitt *et al.*, 2000; LeDoux, 2000; Murray, 2007; Tye *et al.*, 2010). The current study

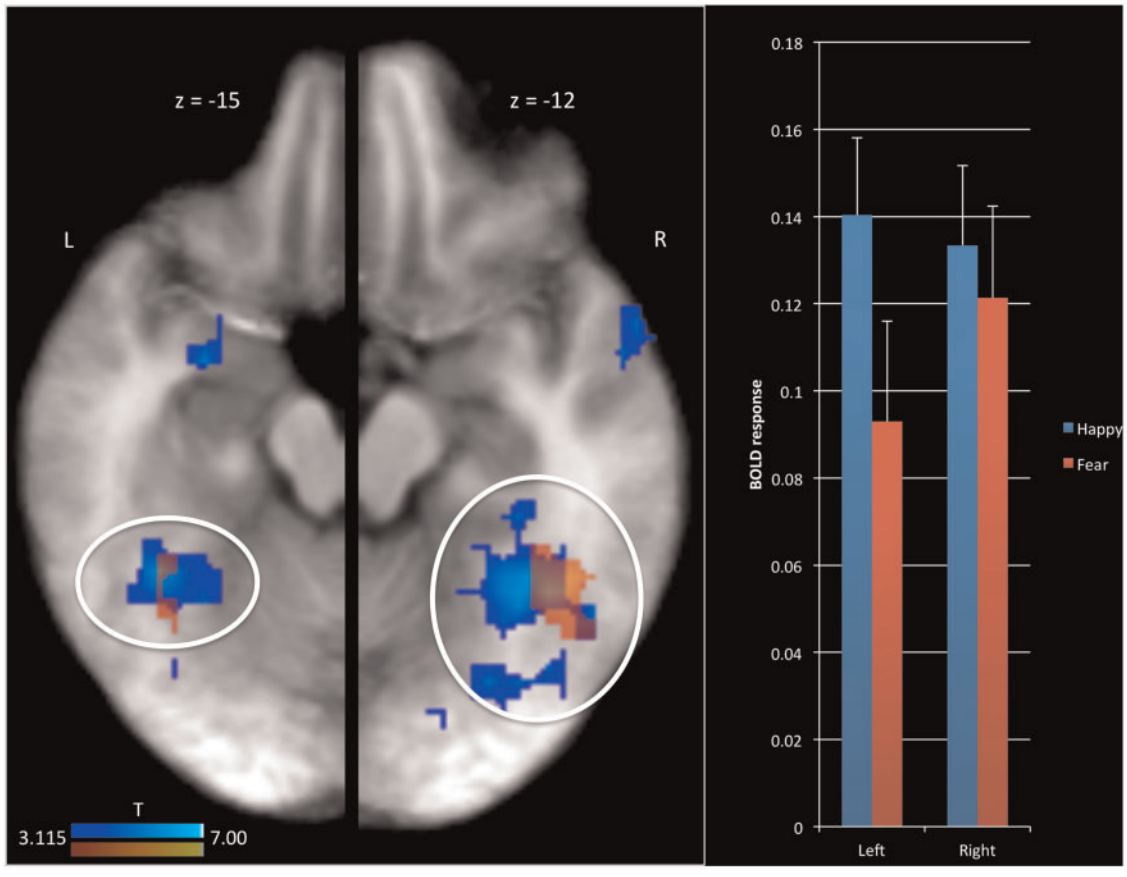

Fig. 3 BOLD response modulated by PE in the amygdala. Results are depicted separately for the BOLD response modulated by PE for fearful facial expressions (Clusters A and B, in red) and for happy facial expressions (Clusters C, D and E, in green). L = left; R = Right.

Table 2 Brain regions modulated by EV (a) for 'fear' objects, (b) for 'happy' objects, (c) for 'happy' *vs* 'fear' objects

Region[a]	Hemisphere	Brodmann's area	*t*	Coordinates of peak significance (*x y z*)			Voxels	Post hoc
(a) EV for 'fear' objects								
Middle temporal gyrus	Left	21	3.799	−58.5	−43.5	−0.5	19	+
Middle temporal gyrus	Left	21	4.001	−49.5	−28.5	−0.5	16	+
Middle temporal gyrus	Right	37	4.311	49.5	−61.5	−6.5	13	+
(b) EV for 'happy' objects								
Fusiform gyrus	Right	37	4.839	28.5	−55.5	−9.5	51	+
Fusiform gyrus	Left	37	5.104	−28.5	−52.5	−9.5	19	+
(c) EV for 'fear' *vs* 'happy' objects								
Fusiform gyrus	Left	37	−5.555	−28.5	−49.5	−9.5	10	H>F
Lingual gyrus	Right	17	−4.842	10.5	−82.5	2.5	21	H>F

All results are thresholded at $P = 0.005$ uncorrected and a cluster extent threshold of 10 voxels (corresponding to map-wise false-positive probability of $P < 0.05$). ± indicates activation/deactivation compared with baseline. H > F indicates Happy larger Fear.
[a]The regions are according to the Talairach Daemon atlas (http://www.nitrc.org/projects/tal-daemon/).

likewise suggests that the amygdala is involved in both aversive and appetitive social referencing, a form of socially induced conditioning (Mineka and Cook, 1993), and in this context, is sensitive to social PE information. Some of the literature suggests that the amygdala codes an unsigned PE, meaning that the recruitment of the amygdala is strong when the deviation from the expected award is either positive or negative (e.g. Roesch *et al.*, 2012; Metereau and Dreher, 2013). Our observation of positive modulation by both fearful and happy facial expressions is consistent with this suggestion.

Our study indicated that the BOLD response in large portions of the posterior fusiform gyrus was positively modulated by PE for both fearful and happy expressions, albeit stronger for happy facial expressions compared with fearful facial expressions. This was confirmed by a conjunction analysis, which showed significant overlap in bilateral fusiform gyrus for fear and happy PE-modulated BOLD responses. Considerable work has shown that fusiform gyrus is implicated in face processing in what has been termed the fusiform face area (Kanwisher *et al.*, 1997). Recent work has extended this view by

indicating that the fusiform gyrus, as well as regions in inferior occipital gyrus and the posterior superior temporal sulcus, may contain multiple face selective regions (Weiner and Grill-Spector, 2013). The current data suggest that activity within some of these regions can be modulated by the PE in response to the emotion displayed by the face. The temporal cortex is reciprocally connected with the amygdala (Freese and Amaral, 2009). It is suggested that the sight of an object, previously associated with a facial expression, triggers a representation of this expression in the fusiform gyrus (FFG) through its connections with the amygdala, thereby enhancing attention to these faces (Pessoa and Ungerleider, 2004; Mitchell *et al.*, 2006; Blair and Mitchell, 2009). Indeed, a recent connectivity study showed that voxels in the fusiform face area had 'characteristic patterns of connectivity' with bilateral amygdala, which could predict activation in fusiform face area (FFA) upon seeing faces (Saygin *et al.*, 2012). In short, the modulation of activity within fusiform by expression PE might reflect a secondary consequence of sensitivity to PE expression information within the amygdala and the connectivity of this region with fusiform cortex.

In contrast to predictions, striatal areas were not significantly modulated by PE of BOLD responses for either fearful or happy expressions. This is despite considerable work indicating responsiveness to PE information within the striatum (McClure *et al.*, 2003; O'Doherty *et al.*, 2003; Seymour *et al.*, 2007) and findings that (primarily dorsal) striatum is responsive to happy expressions (Scott-Van Zeeland *et al.*, 2010). It should be noted though that this last study observing (primarily dorsal) striatal responsiveness to happy expressions, involved instrumental learning paradigms; the happy face was the 'reward' for committing a particular action (Scott-Van Zeeland *et al.*, 2010, although see Lin *et al.*, 2012). The striatum is involved in organizing motor responses (Brunia and van Boxtel, 2000; Watanabe and Munoz, 2010), and increased activation in striatum leading up to an action has been demonstrated in non-human primates (Takikawa *et al.*, 2002; Itoh *et al.*, 2003) and in humans (Watanabe and Munoz, 2010). Indeed, Kable and Glimcher (2009) noted that value signal in dorsal and ventral striatum is related to actions. In contrast, the facial expressions in the current study did not serve as a reward for an action but rather provided positive valence information to a novel object. Studies employing Pavlovian-type learning paradigms have reported PE-modulated activity within striatum; however, the region of striatum has typically been ventral (nucleus accumbens) (Abler *et al.*, 2006; Bray and O'Doherty, 2007; Seymour *et al.*, 2007). Modulation by PE within ventral striatum might have been expected in the current study. However, this region is subject to signal dropout (e.g. Nikolova *et al.*, 2012)—our failure to observe PE-modulated ventral striatal activity may reflect scanner coverage issues. Alternatively, the requirement to make a 'gender judgment' may account for the absence of ventral striatal activity in the current study (suggestion offered by anonymous reviewer).

Participants were more likely to avoid objects that were associated with fearful expressions compared with objects that were associated with happy expressions. We also observed an association between learning about 'happy' objects and the BOLD response modulated by the unexpectedness of happy expressions. In addition, the left amygdala was recruited when participants were rating 'happy' objects. These results suggest that the amygdala is involved in learning about object valence in the context of a social referencing task. It is true that there was not a similar association between learning about 'fear' objects and the PE-modulated BOLD response for fearful facial expressions. This would have been expected especially because fearful expressions are most consistently associated with amygdala activation (Phan *et al.*, 2002; Murphy *et al.*, 2003; Costafreda *et al.*, 2008). However, the amygdala is clearly recruited by happy facial expressions (Phan *et al.*, 2002; Zald, 2003; Costafreda *et al.*, 2008; Sergerie *et al.*, 2008; Fusar-Poli *et al.*, 2009). Moreover, when effect sizes of amygdala activation are considered, a review of the literature suggests an overall slightly stronger effect size for happy compared with fearful expressions in the amygdala (Sergerie *et al.*, 2008). In short, we expect that future studies will observe associations between learning and PE-modulated activity for both fearful and happy expressions.

Two limitations should be mentioned with respect to the current study. First, learning could not be directly indexed in the current study on a trial-by-trial basis. Instead, the data were modeled using an average learning rate set at 0.65 based on the object rating after each run. It is worth noting that it is not untypical to model BOLD data using the average learning rate of the group of participants (e.g. Glascher and Buchel, 2005; Pessiglione *et al.*, 2006; Seymour *et al.*, 2007; Rodriguez, 2009; van der Heiden *et al.*, 2013) and that 0.65 is within the range of the learning rate implemented in these studies. Second, we were unable to find any association between our measure of learning for 'fear' objects and the BOLD response in the amygdala modulated by the PE for fearful expressions. This may reflect a lack of variability due to our crude measure of avoidance ratings; at the end of each session, participants viewed each presented object twice and for each object participants evaluated whether they would either avoid or approach the object. Alternatively, this may reflect a Type II error due to a lack of sufficient power. It may also reflect the fact that the PE-modulated brain responses were overall slightly stronger for happy facial expressions compared with fearful facial expressions.

In conclusion, the current study suggests that the amygdala plays two roles in social referencing: (i) initiating learning regarding the valence of an object as a function of PEs to expressions displayed toward this object and (ii) orchestrating an emotional response to the object when value judgments are being made regarding this object. Moreover, the amygdala's role is seen for both aversive, fearful and appetitive, happy reinforcers. Fusiform gyrus was also sensitive to PE information for both emotional expressions, though whether this reflects secondary effects as a consequence of amygdala input is unclear.

Conflict of Interest

None declared.

REFERENCES

Abler, B., Walter, H., Erk, S., Kammerer, H., Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage*, *31*(2), 790–5.

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences*, *1191*(1), 42–61.

Aktar, E., Majdandzic, M., de Vente, W., Bogels, S.M. (2013). The interplay between expressed parental anxiety and infant behavioural inhibition predicts infant avoidance in a social referencing paradigm. *Journal of Child Psychology Psychiatry*, *54*(2), 144–56.

Amunts, K., Kedo, O., Kindler, M., et al. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anatomy and Embryology*, *210*(5), 343–52.

Association, A.P. (1994). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. Washington, DC: American Psychiatric Association.

Baxter, M.G., Murray, E.A. (2002). The amygdala and reward. *Nature Reviews Neuroscience*, *3*(7), 563–73.

Bayliss, A.P., Frischen, A., Fenske, M.J., Tipper, S.P. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition*, *104*(3), 644–53.

Blair, R.J. (2003). Facial expressions, their communicatory functions and neuro-cognitive substrates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 561–72.

Blair, R.J., Mitchell, D.G. (2009). Psychopathy, attention and emotion. *Psychological Medicine*, *39*(4), 543–55.

Bray, S., O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *Journal of Neurophysiology*, *97*(4), 3036–45.

Brunia, C.H.M., van Boxtel, G.J.M. (2000). Motor preparation. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G., editors. *Handbook of Psychophysiology*. New York: Cambridge University Press, pp. 507–32.

Costafreda, S.G., Brammer, M.J., David, A.S., Fu, C.H. (2008). Predictors of amygdala activation during the processing of emotional stimuli: a meta-analysis of 385 PET and fMRI studies. *Brain Research Reviews*, *58*(1), 57–70.

Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–73.

Desikan, R.S., Segonne, F., Fischl, B., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–80.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–35.

Everitt, B.J., Cardinal, R.N., Hall, J., Parkinson, J.A., Robbins, T.W. (2000). Differential involvement of amygdala subsystems in appetitive conditioning and drug addiction. In: Aggleton, J.P., editor. *The Amygdala: A Functional Analysis*. Oxford: Oxford University Press, pp. 353–90.

Everitt, B.J., Cardinal, R.N., Parkinson, J.A., Robbins, T.W. (2003). Appetitive behavior: impact of amygdala-dependent mechanisms of emotional learning. *Annals of the New York Academy of Sciences*, *985*, 233–50.

First, M.B., Gibbon, R.L., Williams, J.B.W., Benjamin, L.S. (1997). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II)*. Washington, DC: American Psychiatric Press, Inc.

Fisak, B., Jr, Grills-Taquechel, A.E. (2007). Parental modeling, reinforcement, and information transfer: risk factors in the development of child anxiety? *Clinical Child and Family Psychology Review*, *10*(3), 213–31.

Freese, J., Amaral, D. (2009). Neuroanatomy of the primate amygdala. In: Whalen, P., Phelps, E., editors. *The human amygdala*. New York: Guilford, pp. 3–42.

Fusar-Poli, P., Placentino, A., Carletti, F., et al. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry and Neuroscience*, *34*(6), 418–32.

Glascher, J., Buchel, C. (2005). Formal learning theory dissociates brain regions with different temporal integration. *Neuron*, *47*(2), 295–306.

Hooker, C.I., Germine, L.T., Knight, R.T., D'Esposito, M. (2006). Amygdala response to facial expressions reflects emotional learning. *Journal of Neuroscience*, *26*(35), 8915–22.

Hudson, J.L., Rapee, R.M. (2001). Parent-child interactions and anxiety disorders: an observational study. *Behaviour Research and Therapy*, *39*(12), 1411–27.

Itoh, H., Nakahara, H., Hikosaka, O., Kawagoe, R., Takikawa, Y., Aihara, K. (2003). Correlation of primate caudate neural activity and saccade parameters in reward-oriented behavior. *Journal of Neurophysiology*, *89*(4), 1774–83.

Jeon, D., Kim, S., Chetana, M., et al. (2010). Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. *Nature Neuroscience*, *13*(4), 482–8.

Kable, J.W., Glimcher, P.W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, *63*(6), 733–45.

Kanwisher, N., McDermott, J., Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–11.

Klinnert, M.D., Emde, R.N., Butterfield, P., Campos, J.J. (1986). Social referencing: the infant's use of emotional signals from a friendly adult with mother present. *Developmental Psychology*, *22*(4), 427–32.

LeDoux, J.E. (2000). Emotion circuits in the brain. *The Annual Review of Neuroscience*, *23*, 155–84.

Lin, A., Adolphs, R., Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*, *7*(3), 274–81.

McClure, S.M., Berns, G.S., Montague, P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339–46.

Metereau, E., Dreher, J.-C. (2013). Cerebral correlates of salient prediction error for different rewards and punishments. *Cerebral Cortex*, *23*(2), 477–87.

Mineka, S., Cook, M. (1993). Mechanisms involved in the observational conditioning of fear. *Journal of Experimental Psychology General*, *122*(1), 23–38.

Mitchell, D.G., Richell, R.A., Leonard, A., Blair, R.J. (2006). Emotion at the expense of cognition: psychopathic individuals outperform controls on an operant response task. *Journal of Abnormal Psychology*, *115*(3), 559–66.

Muris, P., Steerneman, P., Merckelbach, H., Meesters, C. (1996). The role of parental fearfulness and modeling in children's fear. *Behaviour Research of Therapy*, *34*(3), 265–8.

Murphy, F.C., Nimmo-Smith, I., Lawrence, A.D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cognitive Affective and Behavioral Neuroscience*, *3*(3), 207–33.

Murray, E.A. (2007). The amygdala, reward and emotion. *Trends in Cognitive Sciences*, *11*(11), 489–97.

Murray, L., Creswell, C., Cooper, P.J. (2009). The development of anxiety disorders in childhood: an integrative review. *Psychological Medicine*, *39*(9), 1413–23.

N'Diaye, K., Sander, D., Vuilleumier, P. (2009). Self-relevance processing in the human amygdala: gaze direction, facial expression, and emotion intensity. *Emotion*, *9*(6), 798–806.

Nikolova, Y.S., Bogdan, R., Brigidi, B.D., Hariri, A.R. (2012). Ventral striatum reactivity to reward and recent life stress interact to predict positive affect. *Biological Psychiatry*, *72*(2), 157–63.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–37.

O'Doherty, J.P., Hampton, A., Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy Science*, *1104*, 35–53.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–5.

Pessoa, L., Ungerleider, L.G. (2004). Neuroimaging studies of attention and the processing of emotion-laden stimuli. *Progress in Brain Research*, *144*, 171–82.

Phan, K.L., Wager, T., Taylor, S.F., Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI'. *Neuroimage*, *16*(2), 331–48.

Rescorla, L.A., Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.H., Prokasy, W.F., editors. *Classical Conditioning II: Current Theory and Research*. New York: Appleton-Century-Crofts, pp. 64–99.

Rodriguez, P.F. (2009). Stimulus-outcome learnability differentially activates anterior cingulate and hippocampus at feedback processing. *Learning and Memory*, *16*(5), 324–31.

Roesch, M.R., Esber, G.R., Li, J., Daw, N.D., Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce–Hall and Rescorla–Wagner coexist within the brain. *European Journal of Neuroscience*, *35*(7), 1190–200.

Saygin, Z.M., Osher, D.E., Koldewyn, K., Reynolds, G., Gabrieli, J.D., Saxe, R.R. (2012). Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature Neuroscience*, *15*(2), 321–7.

Scott-Van Zeeland, A.A., Dapretto, M., Ghahremani, D.G., Poldrack, R.A., Bookheimer, S.Y. (2010). Reward processing in autism. *Autism Research*, *3*(2), 53–67.

Sergerie, K., Chochol, C., Armony, J.L. (2008). The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *32*(4), 811–30.

Seymour, B., Daw, N., Dayan, P., Singer, T., Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience*, *27*(18), 4826–31.

Takikawa, Y., Kawagoe, R., Hikosaka, O. (2002). Reward-dependent spatial selectivity of anticipatory activity in monkey caudate neurons. *Journal of Neurophysiology*, *87*(1), 508–15.

Talairach, J., Tournoux, P. (1988). *Co-Planar Stereotaxix Atlas of the Human Brain*. Stuttgart: Thieme.

Tye, K.M., Cone, J.J., Schairer, W.W., Janak, P.H. (2010). Amygdala neural encoding of the absence of reward during extinction. *Journal of Neuroscience*, *30*(1), 116–25.

van der Heiden, L., Scherpiet, S., Konicar, L., Birbaumer, N., Veit, R. (2013). Inter-individual differences in successful perspective taking during pain perception mediates emotional responsiveness in self and others: an fMRI study. *NeuroImage*, *65*(0), 387–94.

Viding, E. (2012). Amygdala response to preattentive masked fear in children with conduct problems: the role of callous-unemotional traits. *The American Journal of Psychiatry*, *169*(10), 1109–16.

Watanabe, M., Munoz, D.P. (2010). Presetting basal ganglia for volitional actions. *Journal of Neuroscience*, *30*(30), 10144–57.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio: TX: Pearson.

Weiner, K.S., Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychological Research*, *77*(1), 74–97.

Zald, D.H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews*, *41*(1), 88–123.