doi:10.1093/scan/nsn033 SCAN (2008) 3, 303–312

The role of the amygdala in implicit evaluation of emotionally neutral faces

Alexander Todorov and Andrew D. Engell
Department of Psychology and Center for the Study of Brain, Mind, and Behavior, Princeton University, Princeton, NJ, USA

The amygdala is involved in the evaluation of novel stimuli, including faces. We examined whether the amygdala is engaged during the evaluation of emotionally neutral faces along trait-specific dimensions such as trustworthiness and attractiveness or along a general valence dimension. Using behavioral data from evaluation of faces on 14 trait dimensions and fMRI data from an implicit evaluation paradigm, we show that the extent to which the amygdala responds to variations of faces on specific dimensions is a function of the valence content of these dimensions. Variations on dimensions with clear valence connotations (e.g. trustworthiness) engaged the amygdala more strongly than variations on dimensions with less clear valence connotations (e.g. dominance). In addition to the amygdala, several other regions—right superior occipital gyrus, right middle temporal/occipital gyrus and bilateral fusiform gyri—were involved in valence evaluation of faces. However, the relation between these regions and face valence was accounted for by the amygdala's response to faces. The findings suggest that the amygdala (i) automatically evaluates novel faces along a general valence dimension; and (ii) modulates a face responsive network of regions in occipital and temporal cortices.

Keywords: face perception; amygdala; valence evaluation; social cognition

The amygdala, a subcortical brain region critical for the evaluation of novel stimuli (Davis and Whalen, 2001; Amaral, 2002; Phelps and LeDoux, 2005; Vuilleumier, 2005), is engaged during the evaluation of trustworthiness from emotionally neutral faces (Adolphs *et al.*, 1998; Winston *et al.*, 2002; Engell *et al.*, 2007; Todorov *et al.*, 2008). Functional magnetic resonance imaging (fMRI) studies show that the amygdala's response increases as trustworthiness decreases, even when participants are engaged in a task that does not demand explicit evaluation of faces (Winston *et al.*, 2002; Engell *et al.*, 2007; Todorov *et al.*, 2008). For example, in our prior study (Engell *et al.*, 2007), participants were presented with blocks of novel faces and asked after each block to indicate whether a test face was presented in the block. Although this task did not demand explicit evaluation of the faces, activation in the amygdala changed as a function of the trustworthiness of the faces, as assessed by judgments of a group of participants different from the participants in the fMRI study.

However, it is unclear whether the amygdala's response is due to trustworthiness *per se* or whether it reflects a general valence evaluation of faces. To date, the literature has focused almost exclusively on trustworthiness evaluation (e.g. Adolphs *et al.*, 1998; Winston *et al.*, 2002; Engell *et al.*, 2007; Todorov *et al.*, 2008). Although we have previously suggested that the amygdala's response may reflect valence evaluation (Engell *et al.*, 2007), there have been no formal tests of this proposal. It is a well-known finding in social psychology that valence evaluation is central for person impressions (Fiske *et al.*, 2007; Kim & Rosenberg, 1980; Rosenberg *et al.*, 1968; Wyer & Srull, 1979; cf. Osgood *et al.*, 1957). In fact, trustworthiness judgments from faces are highly correlated with other social judgments. For example, for the set of faces used in our prior study (Engell *et al.*, 2007), the correlations of trustworthiness judgments with judgments of attractiveness, aggressiveness and intelligence were 0.75, –0.76 and 0.63, respectively. This shared variance reflects general valence evaluation of faces (Oosterhof and Todorov, 2008; Todorov, 2008).

The first hypothesis of the current study was that the amygdala is involved in general valence evaluation of emotionally neutral faces rather than in evaluation of faces on specific trait dimensions such as trustworthiness. According to this hypothesis, face variations on any social dimension (e.g. trustworthiness, attractiveness, aggressiveness) should engage the amygdala to the extent that this dimension has a valence component. In other words, variations on dimensions with clear valence connotations (e.g. trustworthiness and meanness) should engage the amygdala more strongly than variations on dimensions with less clear valence connotations (e.g. dominance).

To test this hypothesis, we reanalyzed the fMRI data of our prior study (Engell *et al.*, 2007), in which participants were ostensibly engaged in a face memory task and did not explicitly evaluate the faces, using a large behavioral data set of trait judgments of the faces used in that study. First, we established

Received 20 June 2008; Accepted 8 August 2008
Advance Access publication 1 October 2008

We thank Nick Oosterhof, Chris Said and Sara Verosky for their comments on previous versions of this manuscript and Valerie Loehr for her help with the behavioral studies. National Science Foundation (BCS-0446846).

Correspondence should be addressed to Alexander Todorov, Department of Psychology, Princeton University, Princeton, NJ 08540, USA. E-mail: atodorov@princeton.edu.

mean ratings for the faces on 14 trait dimensions, which have been determined to be important in evaluation of novel faces (Oosterhof and Todorov, 2008). Second, we derived the amygdala's response to each of these faces from our prior fMRI study. Specifically, we extracted the amygdala activation to each face from face-responsive voxels within the bilateral amygdala for each participant, and then averaged this activation across participants. Thus, for each face we had measures of its evaluation on 14 trait dimensions and the amygdala activation to the face in an implicit evaluation paradigm. This allowed us to conduct an analysis at the level of the faces.

Typically, identifying brain regions that respond as a function of trait judgments of faces is done by regressing the hemodynamic response on these judgments (e.g. Winston *et al.*, 2002; Engell *et al.*, 2007). However, this strategy is susceptible to a serious collinearity problem if used for multiple judgments that are highly correlated. For example, the shared variance between attractiveness and trustworthiness judgments of the faces used here was 56%. The problem is further compounded with the introduction of additional regressors such as judgments of threat and emotional stability. Without good a priori reasons of which social judgments should be controlled for in the analysis, it is not clear how to resolve the collinearity problem.

In this article, we pursued an alternative strategy. First, we estimated the relation between the average amygdala activation to faces and multiple judgments of these faces obtained from a large group of participants separate from the fMRI participants. That is, we computed the correlations between the face ratings on each of the 14 trait dimensions with the strength of the amygdala response. As shown in our prior study (Engell *et al.*, 2007), the amygdala's response to face trustworthiness was better predicted by judgments of trustworthiness averaged across a large number of participants than by the individual judgments of fMRI participants. Partly, this result is due to the fact that judgments averaged across participants are more reliable than individual judgments. Further, consensual components of judgments, as measured by aggregated judgments, most likely reflect facial properties that have similar effects across perceivers in contrast to idiosyncratic components of judgments (Hönekopp, 2006; Engell *et al.*, 2007). Second, we submitted the trait judgments to a principal components analysis (PCA) to obtain estimates of the shared variance among judgments. As shown previously, the first principal component (PC) of trait judgments from faces reflects general valence evaluation (Oosterhof and Todorov, 2008; Todorov, 2008). Third, we tested whether there were significant relationships between the amygdala activation and specific trait judgments after controlling for the valence component.

As noted above, the valence hypothesis predicts that the degree of engagement of the amygdala in the processing of faces on social dimensions should be a function of the valence content of these dimensions (i.e. the variance of the dimension accounted for by the latent valence component). The greater the valence content of a trait dimension, the stronger the engagement of the amygdala should be. In other words, the variance accounted for by the valence component in trait judgments should predict the variance accounted for by these judgments in the amygdala. Further, if the amygdala is engaged in general valence evaluation of faces rather than in specific trait evaluation, once the valence content of trait judgments is controlled for, there should not be residual variance accounted for by specific trait judgments in the amygdala.

In addition to the main hypothesis that the amygdala is engaged in general valence evaluation of novel faces, we tested two other hypotheses. First, we tested whether other face responsive regions are also engaged in valence evaluation of faces. Specifically, we identified all face responsive regions and extracted the average activation to each face in these regions. Then, as in the case for the analysis of the amygdala activation, we tested (i) for significant relations between the average activation and trait judgments; and (ii) whether such relations could be accounted for by the valence content of the judgments.

Second, we tested whether the amygdala modulates the activation in other regions involved in the valence evaluation of faces. There is good anatomical evidence that the amygdala projects back to regions in both occipital and temporal cortex (Amaral *et al.*, 2003) and that it can amplify the response in these regions for emotionally salient stimuli (Vuilleumier, 2005). The face response in the amygdala should account for the relation between face valence and activation in these regions to the extent that the amygdala modulates the response to faces in regions involved in valence evaluation.

METHODS

Participants

Three hundred and forty-five undergraduate students participated in the trait judgments studies for partial course credit or payment, and 15 participants from the community in and around Princeton, NJ, participated in the fMRI study for payment.

Face stimuli

In both the behavioral and fMRI studies, we used the same set of 66 Caucasian standardized faces (Lundqvist *et al.*, 1998) with direct gaze and neutral expressions. These were photographs of amateur actors between 20 and 30 years of age with no facial hair, earrings, eyeglasses or visible make-up, all wearing gray T-shirts. All of these faces were categorized as emotionally neutral in an expression categorization task (Engell *et al.*, 2007).

Behavioral studies

The 66 faces were rated on 14 trait dimensions. Twelve of the dimensions were identified in a prior behavioral study ($n = 55$) as frequently used to characterize novel faces

(Oosterhof and Todorov, 2008). Two other trait dimensions—dominance and threat—were added because of their importance for person perception (Wiggins *et al.*, 1989; Bar *et al.*, 2006). The faces were rated by a separate group of participants on each of the trait dimensions. Participants were asked to rate only one of the traits using their 'gut instinct'. The detailed procedures are described in Oosterhof and Todorov (2008). Each trait was rated by at least 18 participants. The faces were presented three times in three separate blocks. Within each block, the order of faces was randomized for each participant. The response scale ranged from 1 [Not at all (trait term)] to 9 [Extremely (trait term)]. The faces were presented three times to increase the interrater agreement and, consequently, the reliability of judgments by reducing the measurement error for each participant. For each participant, we computed the average judgment for each of the 66 faces. The reliability of the trait judgments ($\alpha > 0.90$ for each trait judgment) was computed on the standardized average ratings of participants.

fMRI study

The detailed procedures of the study are described in Engell *et al.* (2007). Participants were told that the study was about face memory. The task consisted of two data acquisition runs. Each run contained six blocks of 11 face images—the 66 faces used in the behavior studies—presented in random order. All runs began with a 12 s presentation of a fixation cross. Within a block, each of the 11 faces was presented for 1 s in a jittered event-related fashion. Each ISI was randomly chosen from an exponential distribution with target mean ISI = 3.5 s and minimum ISI = 1.5 s. At the conclusion of each block, a red fixation cross appeared on a white screen until a predetermined time point (52 s from the beginning of the block) at which time another face (the test face) was presented for 1 s. The participant's task was to report whether the identity of the test face was the same as any of the faces in that block. Each of the six blocks was separated by a 12 s rest period in order to allow hemodynamic activity to return to baseline levels.

The blood oxygenation level-dependent (BOLD) signal was used as a measure of neural activation. Echo planar images (EPI) were acquired with a Siemens 3.0 Tesla Allegra head-dedicated scanner (Siemens, Erlangen, Germany) with a standard 'bird-cage' head coil (TR, 2000 ms, TE, 30 ms, flip angle, 90°, matrix size, 64×64). Near whole-brain coverage was achieved with 33 interleaved 3 mm axial slices. At the beginning of each scan session a high-resolution anatomical image (T1-MPRAGE, TR, 2500 ms, TE, 4.3 ms, flip angle, 8°, matrix size 256×256) was acquired for use in registering activity to each participant's anatomy and for spatially normalizing data across participants.

Data were analyzed with Analysis of Functional NeuroImages (AFNI; Cox, 1996) using standard preprocessing procedures. Participant motion was corrected using a 6 parameter 3D motion-correction algorithm following slice scan-time correction. The data were low-pass filtered with a frequency cut-off of 0.1 Hz subsequent to spatial smoothing with a 6 mm full width at half minimum (FWHM) Gaussian kernel. Finally, the signal was normalized to percent signal change from the mean.

For statistical analysis each stimulus time-series was convolved with a hemodynamic response function to create a regressor for face perception. In addition, regressors of no interest were included in the multiple regression model to factor out variance associated with mean, linear and quadratic trends in each run as well as participant head motion. The 9-parameter landmark method of Talairach and Tournoux (1988) was used in order to spatially normalize the activation maps across participants. A *t*-test was performed on the coefficients for face perception given by the multiple regression analysis for each participant to test the significance of coefficients across participants.

Face-responsive amygdala voxels were defined as those voxels located within the amygdala region (as identified by the Talairach-Tournoux Daemon atlas included in AFNI) and showing a significant response to faces relative to a fixation baseline ($P < 0.001$, *uncorrected*). The resulting mask comprised a 492 mm^3 cluster in the left amygdala and a 307 mm^3 cluster in the right amygdala. In order to obtain parameter estimates for each face, the multiple regression analysis was run a second time treating each event (i.e. each individual face) in the convolved time-series as a separate regressor. The response to each of the 66 faces was then averaged across the face-responsive amygdala voxels and averaged across all 15 participants.

We used a threshold of $P < 0.05$ (corrected) to identify face responsive voxels outside of the amygdala. Correction for multiple comparisons was done by using the AlphaSim program included in AFNI. A minimum cluster size of 270 mm^3 was used to achieve the corrected *P*-value as determined by a Monte Carlo simulation with voxel-wise threshold of $P < 0.0001$. In the identified face responsive regions, we averaged the response to each of the 66 faces across voxels and across participants.

To summarize, first, we collected ratings of the faces on 14 trait dimensions. Second, we identified face-responsive voxels and extracted the mean activation to each face in these voxels. This voxel selection was unbiased with regard to trait dimensions (face responsiveness was the only functional criteria for inclusion). Similarly, the task of the participants in the fMRI study was unbiased with regard to trait dimensions (participants did not attend to any particular dimension). Thus, we had measures for each face on its evaluation on social dimensions and brain activation in face-responsive regions. We used these measures to conduct analyses at the level of the stimulus faces.

RESULTS

Evaluation of faces and the amygdala

Consistent with the valence hypothesis, the amygdala activation correlated negatively with all judgments on positive traits (e.g. intelligent, caring) and positively with all judgments on negative traits (e.g. weird, mean) (Figure 1). That is, across trait dimensions, the amygdala responded more strongly to faces that were evaluated negatively. Although all trait judgments (except for dominance) correlated significantly with the amygdala's response, there was considerable variation in the magnitude of the correlations. As outlined in the 'Introduction' section, according to the valence hypothesis, this variation should be predicted by the valence content of the specific judgments.

To obtain a measure of general valence evaluation, we submitted the trait judgments to a PCA. As expected (Osgood *et al.*, 1957; Rosenberg *et al.*, 1968; Kim and Rosenberg, 1980), the first PC, which accounted for 62.9% of the variance, reflected valence evaluation of faces (Table 1). All judgments of positive traits had positive loadings and all judgments of negative traits had negative loadings on this component. The second PC, which accounted for 18.9% of the variance, could be interpreted as dominance evaluation. The judgments with highest loadings on this component were dominance, confidence, aggressiveness and threat (Table 1). The third PC accounted for <6% of the variance, with an eigenvalue <1, and did not have a clear interpretation. As shown by Oosterhof and Todorov (2008),

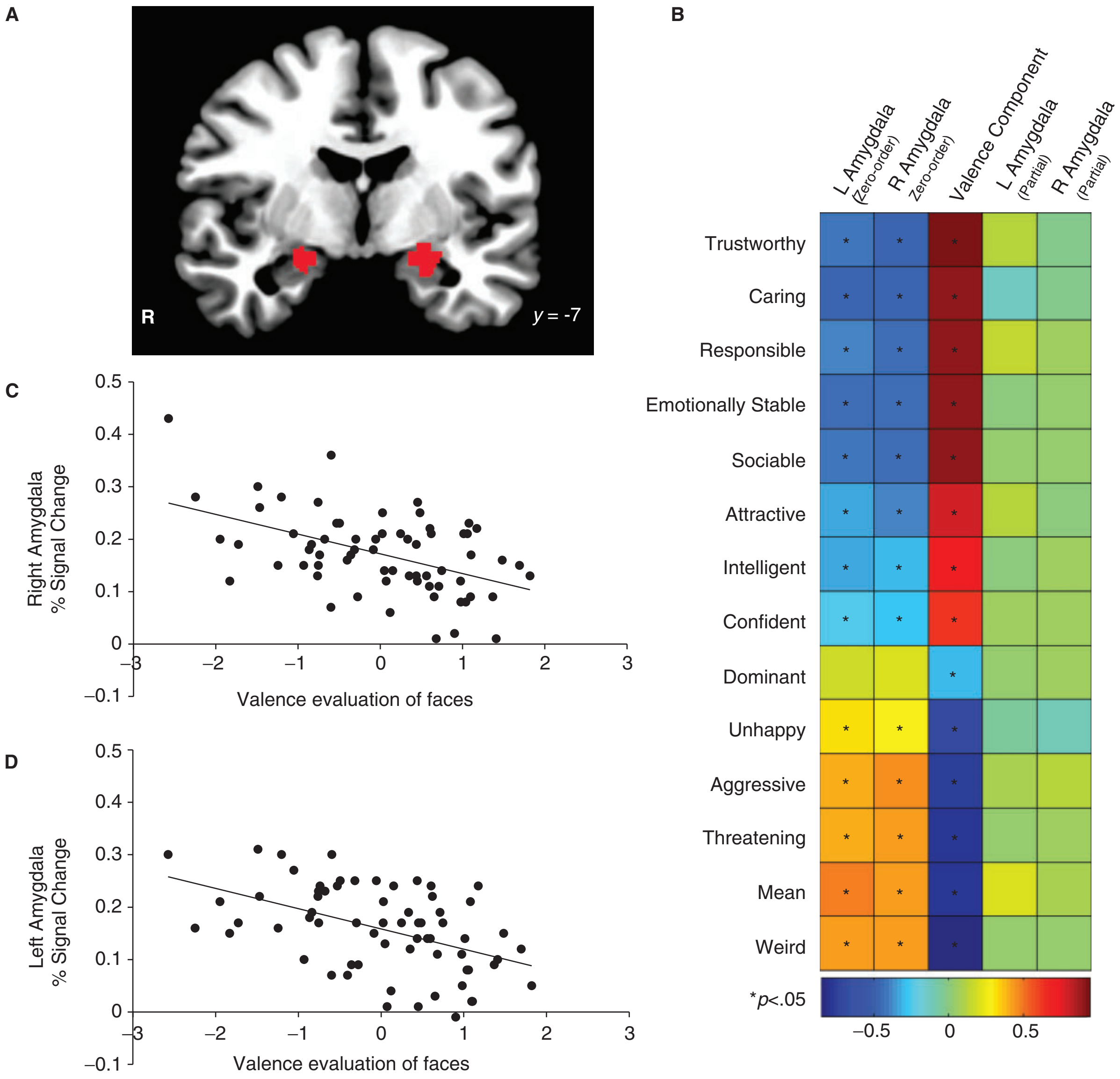

Fig. 1 The relation between the amygdala's response to emotionally neutral faces and variations of these faces on trait dimensions. A coronal brain slice showing face responsive voxels in bilateral amygdala (**A**). An intensity color plot showing correlations between the response in left and right amygdalae to faces and trait judgments of these faces (**B**). The first two columns show zero-order correlations and the fourth and fifth columns show partial correlations controlling for the valence content of the judgments. The third column shows the correlations between trait judgments and a valence component derived from a principal components analysis of the judgments. The traits are ordered according to their correlations with the valence component (Table 1). Scatter plots of the amygdala's response to faces (**C** for right and **D** for left) and their values on the valence component. Each point represents a face.

this two-dimensional solution is robust with respect to both face stimuli and trait judgments used to estimate the PCs.

The valence component, which is a linear combination of the 14 trait judgments, was correlated with both the response in the right and left amygdala ($r = -0.50$ and -0.48, respectively, $P < 0.001$, Figure 1C and D). Consistent with the findings from the trait judgments, more negative evaluations of faces evoked a stronger response in the amygdala. In contrast to the correlations with the first PC, the amygdala's response was uncorrelated with the second PC ($r = 0.06$ and 0.07, for right and left amygdala, respectively).

We used the variance accounted for by the valence component for each trait judgment as an estimate of the valence content of the trait dimension. For example, the valence component accounted for 90% of the variance of trustworthiness judgments, 61% of the variance of threat judgments and 9% of the variance of dominance judgments (Table 1). This variance was strongly correlated with the variance accounted for by each judgment in the amygdala's response to faces (Figure 2A and B, $r = 0.90$ and 0.79 for right and left amygdala, respectively, $P < 0.001$). That is, the stronger the association of a trait judgment with the valence component, the stronger this judgment engaged the amygdala.

Table 1 Loadings of trait judgments of emotionally neutral faces on the first two principal components of a principal components analysis

Trait judgment	Valence evaluation	Dominance evaluation
Trustworthy	0.95	−0.01
Caring	0.91	−0.23
Responsible	0.91	0.15
Emotionally stable	0.91	0.27
Sociable	0.90	0.25
Attractive	0.79	0.35
Intelligent	0.70	0.20
Confident	0.63	0.70
Dominant	−0.30	0.91
Unhappy	−0.70	−0.06
Aggressive	−0.75	0.62
Mean	−0.78	0.48
Threatening	−0.78	0.52
Weird	−0.85	−0.28
Explained variance	62.9%	18.9%

Note: The loadings represent the correlations of the trait judgments with the principal components. The third principal component had an eigenvalue <1.

Moreover, after controlling for the valence content of the trait judgments, there were no significant relationships between any of the judgments and the amygdala's response. Specifically, the range of partial correlations between judgments and the amygdala's response controlling for the valence component was from −0.13 to 0.19 ($M = 0.03$, s.d. = 0.06, for right; and $M = 0.05$, s.d. = 0.08 for left amygdala) and none of the correlations reached significance (Figure 1B).

Evaluation of faces and face-responsive regions other than the amygdala

In addition to the amygdala, a number of other regions showed a stronger response to faces than to a baseline (Table 2). These included right superior occipital gyrus (SOG), bilateral fusiform gyri (FG), right middle temporal/occipital gyrus (MTG/MOG) and two frontoparietal regions. All regions in occipital and temporal cortex showed significant correlations with trait judgments. For the frontoparietal regions, none of the correlations between the activation to faces and trait judgments reached significance.

As shown in Figure 3, the pattern of correlations between the activation to faces in right SOG, FG and right MTG/MOG and trait judgments of these faces was similar to the pattern of correlations observed for the amygdalae. The brain activation in these regions correlated positively with judgments on negative traits and negatively with judgments on positive traits. Thirty-three out of the 56 (regions—trait judgments) correlations were significant. As shown in Table 3, all of these regions except for the right SOG also correlated significantly with the valence component. The correlation for the right SOG was marginally significant.

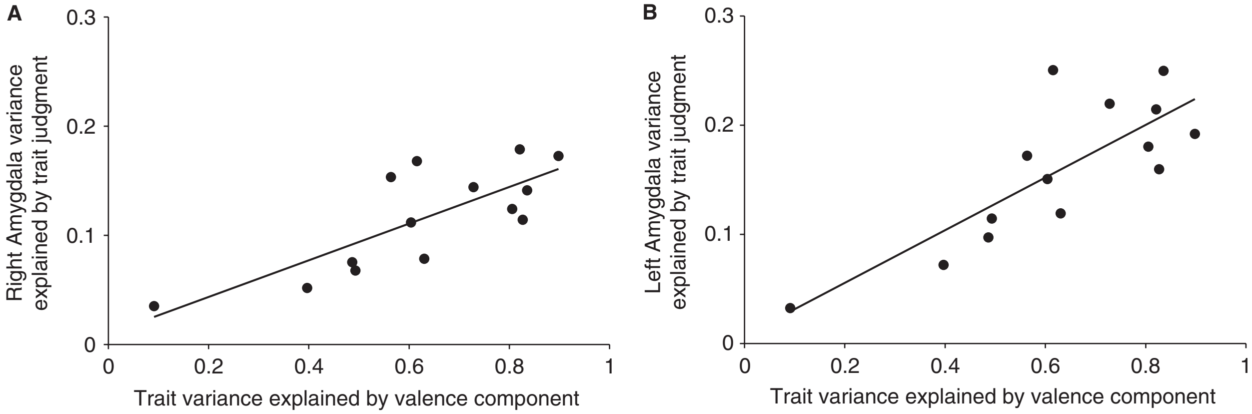

Fig. 2 Scatter plots of the variance accounted for by the valence component in each of 14 trait judgments and the variance accounted for by these judgments in the amygdala's response to emotionally neutral faces (**A** for right and **B** for left amygdala).

Table 2 Regions showing a significantly greater response to faces than to a baseline ($P < 0.05$, corrected for multiple comparisons)

Region	Cluster size (mm^3)	Coordinates			Max *t*
Right amygdala and adjacent regions[a]	3096	17	−5	−11	5.40
Left amygdala and adjacent regions[b]	812	−17	−5	−11	5.41
Right fusiform gyrus	1072	36	−44	−16	5.45
Left fusiform gyrus	378	−37	−38	−16	5.92
Right middle temporal/occipital gyrus	1355	50	−73	5	5.42
Right superior occipital gyrus	1381	19	−96	24	7.06
Left paracentral lobule	427	−6	−26	64	5.71
Right precentral gyrus	442	51	−5	45	5.38

[a] This cluster extends into the right uncus, hippocampus, parahippocampal gyrus and thalamus.
[b] This cluster extends into the left parahippocampal gyrus.

Faces that were evaluated negatively evoked a stronger response in all these regions than faces that were evaluated positively. None of the regions correlated significantly with the second PC. The two frontoparietal regions that were face responsive (Table 2) did not correlate significantly with either of the PCs.

As in the case of the amygdala, after controlling for the shared variance with the valence component, there were no significant relationships between any of the judgments and the activation in the right SOG, right MTG/MOG and FG (Figure 3). Thus, it appears that all these regions were engaged in general valence evaluation of faces.

Relationships between the amygdala and regions involved in the processing of face valence

Although the pattern of correlations between trait judgments and activation in the right SOG, FG and right MTG/MOG (Figure 3) was similar to the pattern of correlations for the amygdala, the magnitude of the correlations was smaller for these regions than for the amygdala. As shown in Table 4, the average variance accounted for by trait judgments in the amygdala was larger than the variance accounted for in any of the other four regions, $ts(13) > 4.15$, $P < 0.001$, for the right amygdala, and $ts(13) > 3.13$, $P < 0.008$, for the left amygdala.

Second, in addition to the weaker correlations, the magnitude of the correlations seemed to be moderated to a smaller extent by the valence content of the judgments than the magnitude of the correlations with the amygdala (Figure 2). Only the left fusiform gyrus and the right MTG/MOG showed significant associations between the valence content of the judgments and the extent to which these judgments engaged these regions (Table 4).

The findings are consistent with the hypothesis that the activity in these regions was modulated by the amygdala's response to the faces. If this is the case, the correlation between valence evaluation of faces and activation in face-responsive regions should be reduced when the amygdala's response is controlled for in the analysis. As shown in Table 3, there were no significant associations between the activation in the right SOG, FG, right MTG/MOG and the valence component after controlling for the activation to faces in left and right amygdala. In contrast, the correlation between amygdala activation and this component remained significant (–0.37, $P < 0.003$ and –0.33, $P < 0.010$, for right and left amygdala, respectively) after controlling for the activation in all four regions.

DISCUSSION

Consistent with the valence hypothesis, the findings suggest that the amygdala is involved in general valence evaluation of novel faces rather than in evaluation of faces on specific trait dimensions. The findings also suggest that the amygdala's engagement in general valence evaluation of novel faces is automatic, as the fMRI task did not demand explicit face evaluation. The extent to which the amygdala is engaged in tracking variations of faces on social dimensions is a function of the valence content of these dimensions. Therefore, it is not surprising that previous studies have found significant negative correlations between the amygdala's response and the perceived trustworthiness of emotionally neutral faces (Winston *et al*., 2002; Engell *et al*., 2007; Todorov *et al*., 2008) given the strong correlation between this trait and the valence component (0.95).

Though valence evaluation of faces encompasses evaluation on multiple social dimensions, it is best approximated by trustworthiness judgments (Oosterhof and Todorov, 2008). Thus, trustworthiness judgments may be sufficient to model how the valence of faces is evaluated in the brain, as practically it would often be unfeasible to collect multiple judgments of faces to measure their valence evaluation. However, it would be misleading to describe this evaluation as trustworthiness evaluation *per se* at the level of neural response, at least, in experimental contexts that do not provide information other than still images of faces. As shown here, this evaluation most likely reflects general valence evaluation of faces.

Our findings also showed that in addition to the amygdala, activation in a number of regions in occipital and temporal cortex varied as a function of the valence evaluation of faces. As in the case of the amygdala, the activation to faces in these regions was negatively correlated with all positive judgments of faces and positively correlated with all negative judgments. Moreover, after controlling for the valence content of these judgments, these correlations were no longer significant. These findings suggest that in addition to medial temporal lobe regions, regions in occipital and inferotemporal cortex were recruited during valence evaluation of faces.

Although the patterns of responses in the amygdala and the regions in the occipital and inferotemporal cortex were similar (Figures 1 and 3), the magnitude of correlations

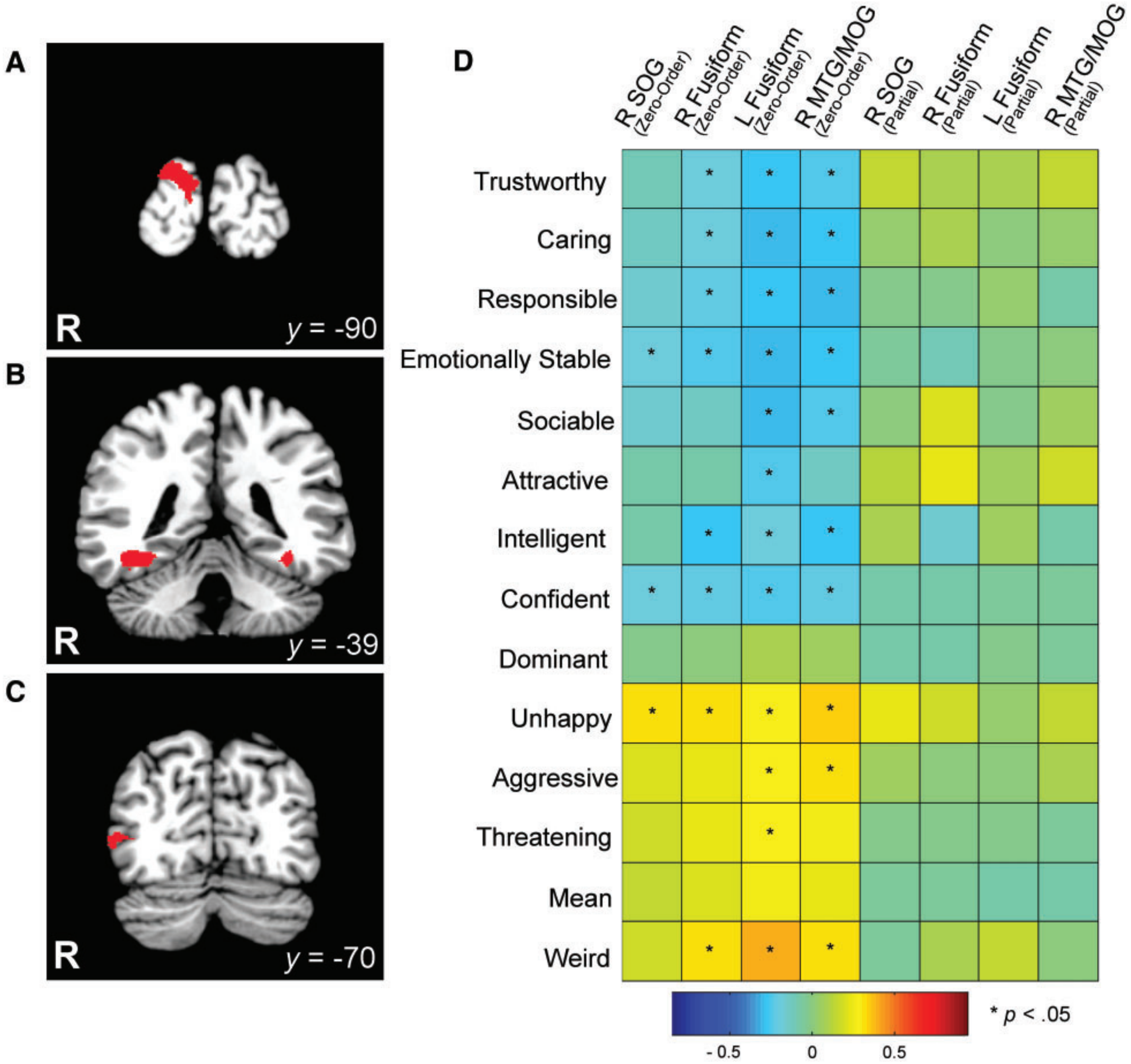

Fig. 3 The relation between the response in regions in temporal and occipital cortices to emotionally neutral faces and variations of these faces on trait dimensions. Coronal brain slices showing face-responsive voxels in right superior occipital gyrus (**A**), bilateral fusiform gyri (**B**) and right middle temporal/occipital gyrus (**C**). An intensity color plot showing correlations between the response in these regions to emotionally neutral faces and trait judgments of these faces (**D**). The first four columns show zero-order correlations and the second four columns show partial correlations controlling for the valence content of the judgments. The traits are ordered according to their correlations with the valence component (Table 1).

Table 3 Zero-order and partial correlations between mean face activation in face-responsive regions and a valence component derived from a principal components analysis of trait judgments of faces

Region	Zero-order	Partial correlation
Right superior occipital gyrus	−0.22*	0.03
Right fusiform gyrus	−0.31**	−0.08
Left fusiform gyrus	−0.40**	−0.13
Right middle temporal/occipital gyrus	−0.37**	−0.10

The partial correlations control for the shared variance with the amygdala activation to the faces. $^{*}P < 0.10$; $^{**}P < 0.05$

between the latter and face evaluation was weaker than the magnitude of correlations between the former and face evaluation. Further, the correlation between valence evaluation of faces and the activation in these regions was no longer significant after the analysis controlled for the amygdala's response to faces. In contrast, the correlation with the amygdala remained significant after controlling for the activation in these regions. These findings are consistent with the hypothesis that the activity in occipital and inferotemporal regions was modulated by the amygdala (Morris *et al.*, 1998; Vuilleumier *et al.*, 2004; Sabatinelli *et al.*, 2005; Vuilleumier, 2005). Anatomical evidence from tracing studies of the

Table 4 Average variance accounted for by trait judgments in face-responsive regions, and correlations between trait variance accounted for by the valence component and variance in face-responsive regions accounted for by trait judgments

Region	Average variance (%)	Correlation
Right superior occipital gyrus	3.8	0.23
Right fusiform gyrus	6.6	0.29
Left fusiform gyrus	10.2	0.86*
Right middle temporal/occipital gyrus	9.1	0.61*
Right amygdala	15.9	0.90*
Left amygdala	14.7	0.79*

*$P < 0.05$

macaque brain shows that the projections from the amygdala to visual cortex are more extensive than those from visual cortex to the amygdala (Amaral *et al.*, 2003). Whereas the amygdala receives visual input only from temporal visual areas (from area TE to the lateral nucleus of the amygdala), it projects to multiple areas in both temporal and occipital visual areas, including early visual areas (from the basal nucleus to these areas).

It is not possible to prove the causal influence of the amygdala on regions in occipital and inferotemporal cortex with correlational evidence such as the one presented here and from functional connectivity studies (Morris *et al.*, 1998; Sabatinelli *et al.*, 2005). However, Vuilleumier *et al.* (2004) showed that whereas patients with hippocampal lesions show enhanced response in regions in occipital and inferotemporal cortex to emotionally salient but unattended stimuli, patients with amygdala lesions did not show this response. These regions included SOG, lateral occipital cortex and FG. Given these findings, one would expect that patients with amygdala lesions would not show enhanced responses to negatively evaluated faces in face-responsive regions in occipital and inferotemporal cortex.

As shown in Figure 1C and D, the response of the amygdala to face valence was linear. Yet, there have been three recent studies reporting a U-shaped, nonlinear amygdala response to face trustworthiness (Todorov *et al.*, 2008; Said *et al.*, in press) and face attractiveness (Winston *et al.*, 2007). Specifically, the activation was stronger to faces at the extremes of the dimensions than to faces at the middle of the dimensions. There are, at least, two hypotheses about the conditions under which the amygdala's response to face valence may be quadratic. According to the first hypothesis, the nature of the evaluation—implicit *vs* explicit—may be critical. In contrast to the current study, participants in Said *et al.*'s study explicitly evaluated the faces on trustworthiness and this may have biased attention to extreme faces. In a recent study, Cunningham *et al.* (2008) observed similar quadratic responses in the amygdala in a valence evaluation task of famous people. When participants focused on the positivity of the evaluation, the response was enhanced to positive stimuli; when they focused on the negativity, the response was enhanced to negative stimuli.

However, this hypothesis cannot account for the all of the data. In Todorov *et al.* (2008), the task was the same as the task used in the current study. According to the second hypothesis, the range of face valence used in a particular study may determine the nature of the amygdala's response. For wider ranges of face valence, the response may be quadratic. For example, we compared the trustworthiness of the faces used in Todorov *et al.* (2008) and the faces used in Engell *et al.* (2007) in a computer model that can assign trustworthiness values to faces based on shape information (Oosterhof and Todorov, 2008). The range of trustworthiness in the former study was from −3.26 to 2.64 in standard deviation units, whereas the range in the latter study was from –1.79 to 1.53. Studies on attractiveness also typically use extreme faces (O'Doherty *et al.*, 2003; Winston *et al.*, 2007) and given the high correlation between attractiveness and face valence (Table 1), this can lead to nonlinear responses in the amygdala as observed by Winston *et al.* (2007).

Although the current findings suggest that the amygdala is involved in valence evaluation of faces, they do not rule out the possibility that the amygdala is involved in arousal evaluation. For example, studies using olfactory (Anderson *et al.*, 2003) and gustatory (Small *et al.*, 2003) stimuli show that the amygdala responds to the intensity of stimuli rather than to their valence (pleasantness/unpleasantness). These findings are consistent with U-shaped responses in the amygdala to faces. There are, at least, two possibilities that can reconcile our findings with hypotheses that the amygdala is primarily involved in processing of arousing stimuli. First, for visual stimuli, arousal and valence are highly negatively correlated. Negative stimuli (e.g. an angry face) are more arousing than neutral and often positive stimuli (e.g. a smiling face), consistent with well-documented negativity biases in processing of social stimuli (Fiske, 1980; Skowronski and Carlston, 1989; Pratto and John, 1991; Rozin and Royzman, 2001). The stimuli in the current study were not sampled to represent different levels of arousal and, as noted above, were not highly positive. Thus, their valence was most likely correlated with their arousal value. Second, it is possible that the amygdala involvement in processing of stimuli depends on their sensory modality. Prior studies testing the arousal hypothesis used olfactory and gustatory stimuli because it is relatively easy to manipulate their pleasantness and intensity independently. Even for olfactory stimuli, the amygdala's response to arousing stimuli may depend on their valence. For example, Winston *et al.* (2005) showed that whereas the amygdala responded more strongly to high- than to low-arousal stimuli for both pleasant and unpleasant odors, it did not discriminate between high- and low-arousal stimuli for neutral odors.

Both linear and U-shaped responses in the amygdala are consistent with a common attentional mechanism according to which the amygdala biases attention toward stimuli that are of current motivational significance to the person

(LaBar *et al.*, 2001; Vuilleumier, 2005; Cunningham *et al.*, 2008). Interestingly, early studies in social cognition showed that allocation of attention to social stimuli exhibits non-linear quadratic response to people as a function of their extremeness rather than their valence (Fiske, 1980) and more recent studies show that evaluative processes are context dependent (Ferguson and Bargh, 2004).

Yet, when no specific context is provided as in the current study, faces are automatically evaluated in terms of their valence. As argued by both Whalen (1998) and Amaral (2002), one of the primary functions of the amygdala may be to provide continuous vigilance by evaluating objects and agents prior to interacting with them. The valence evaluation of stimuli guides approach/avoidance behavior (Chen and Bargh, 1999), and the valence evaluation of emotionally neutral faces is derived from similarity to emotional expressions signaling approach/avoidance behaviors (Oosterhof and Todorov, 2008; Todorov, 2008). Evaluation processes in the amygdala may not only enhance attention and processing of stimuli in perceptual areas (Anderson and Phelps, 2001; Vuilleumier *et al.*, 2004), but may also influence approach/avoidance decisions via interactions with orbital frontal cortex (Baxter *et al.*, 2000). In fact, macaque monkeys with bilateral amygdala lesions exhibit uninhibited approach behaviors during social interactions (Emery *et al.*, 2001). Together, these findings suggest that an important role of the primate amygdala is to influence the likelihood of social engagement with unfamiliar conspecifics.

REFERENCES

Adolphs, R., Tranel, D., Damasio, A.R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470–74.

Amaral, D.G. (2002). The primate amygdala and the neurobiology of social behavior: implications for understanding social anxiety. *Biological Psychiatry*, *51*, 11–7.

Amaral, D.G., Behniea, H, Kelly, J.L. (2003). Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. *Neuroscience*, *118*, 1099–120.

Anderson, A.K., Christoff, K., Stappen, I., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience*, *6*, 196–202.

Anderson, A.K., Phelps, E.A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*, 305–9.

Bar, M., Neta, M., Linz, H. (2006). Very first impressions. *Emotion*, *6*, 269–78.

Baxter, M.G., Parker, A., Lindner, C.C.C., Izquierdo, A.D., Murray, E.A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *Journal of Neuroscience*, *20*, 4311–9.

Chen, M., Bargh, J.A. (1999). Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*, 215–24.

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

Cunningham, W.A., Van Bavel, J.J., Johnsen, I.R. (2008). Affective flexibility: evaluative processing goals shape amygdala activity. *Psychological Science*, *19*, 152–60.

Davis, M., Whalen, P.J. (2001). The amygdala: vigilance and emotion. *Molecular Psychiatry*, *6*, 13–34.

Emery, N.J., Capitanio, J.P., Mason, W.A., Machado, C.J., Mendoza, S.P., Amaral, D.G. (2001). The effects of bilateral lesions of the amygdala on dyadic social interactions in rhesus monkeys (Macaca mulatta). *Behavioral Neuroscience*, *115*, 515–44.

Engell, A.D., Haxby, J.V., Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in human amygdala. *Journal of Cognitive Neuroscience*, *19*, 1508–19.

Ferguson, M.J., Bargh, J. A. (2004). Liking is for doing: the effects of goal pursuit on automatic evaluation. *Journal of Personality and Social Psychology*, *87*, 557–72.

Fiske, S.T. (1980). Attention and weight in person perception: the impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*, 889–906.

Fiske, S.T., Cuddy, A.J.C., Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*, 77–83.

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 199–209.

Kim, M.P., Rosenberg, S. (1980). Comparison of two structural models of implicit personality theory. *Journal of Personality and Social Psychology*, *38*, 375–89.

LaBar, K.S., Gitelman, D.R., Parrish, T.B., Kim, Y.-H., Nobre, A.C., Mesulam, M.-M. (2001). Hunger selectively modulates corticolimbic activation to food stimuli in humans. *Behavioral Neuroscience*, *115*, 493–500.

Lundqvist, D., Flykt, A., Ohman, A. (1998). *Karolinska Directed Emotional Faces* [Database of standardized facial images]. Psychology Section, Department of Clinical Neuroscience, Karolinska Hospital, S-171 76 Stockholm, Sweden.

Morris, J.S., Friston, K.J., Buchel, C., et al. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain*, *121*, 47–57.

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D.M., Dolan, R.J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, *41*, 147–55.

Oosterhof, N.N., Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the USA*, *105*, 11087–92.

Osgood, C.E., Suci, G.I., Tennenbaum, P.H. (1957). *The Measurement of Meaning*. Urbana: University of Illinois Press.

Phelps, E.A., LeDoux, J.E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, *48*, 175–87.

Pratto, F., John, O.P. (1991). Automatic vigilance: the attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*, 380–91.

Rosenberg, S., Nelson, C., Vivekananthan, P.S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*, 283–94.

Rozin, P., Royzman, E.B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320.

Sabatinelli, D., Bradley, M.M., Fitzsimmons, J.R., Lang, P.J. (2005). Parallel amygdala and inferotemporal activation reflect emotional intensity and fear relevance. *NeuroImage*, *24*, 1265–70.

Said, C.P., Baron, S., Todorov, A. (in press). Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience*.

Skowronski, J.J., Carlston, D.E. (1989). Negativity and extremity biases in impression formation: a review of explanations. *Psychological Bulletin*, *105*, 131–42.

Small, D.M., Gregory, M.D., Mak, Y.E., Gitelman, D., Mesulam, M.M., Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron*, *39*, 701–11.

Talairach, J., Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York, NY: Thieme Medical Publishers.

Todorov, A. (2008). Evaluating faces on trustworthiness: an extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Annals of the New York Academy of Sciences*, *1124*, 208–24.

Todorov, A., Baron, S., Oosterhof, N.N. (2008). Evaluating face trustworthiness: a model based approach. *Social, Cognitive, and Affective Neuroscience*, *3*, 119–27.

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585–94.

Vuilleumier, P., Richardson, M., Armony, J.L., Driver, J., Dolan, R.J. (2005). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*, 1271–8.

Whalen, P.J. (1998). Fear, vigilance, and ambiguity: initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, *7*, 177–88.

Winston, J.S., Gottfried, J.A., Kilner, J.M., Dolan, R.J. (2005). Integrated neural representations of odor intensity and affective valence in human amygdala. *Journal of Neuroscience*, *25*, 8903–7.

Winston, J.S., O'Doherty, J., Kilner, J.M., Perrett, D.I., Dolan, R.J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, *45*, 195–206.

Winston, J.S., Strange, B., O'Doherty, J., Dolan, R. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of face. *Nature Neuroscience*, *5*, 277–83.

Wiggins, J.S., Philips, N., Trapnell, P. (1989). Circular reasoning about interpersonal behavior: evidence concerning some untested assumptions underlying diagnostic classification. *Journal of Personality and Social Psychology*, *56*, 296–305.

Wyer, R.S. Jr, Srull, T.K. (1989). *Memory and Cognition in Social Context*. Hillsdale, NJ: Erlbaum.