Nonlinear Amygdala Response to Face Trustworthiness: Contributions of High and Low Spatial Frequency Information

Christopher P. Said, Sean G. Baron, and Alexander Todorov

Abstract

■ Previous neuroimaging research has shown amygdala sensitivity to the perceived trustworthiness of neutral faces, with greater responses to untrustworthy compared with trustworthy faces. This observation is consistent with the common view that the amygdala encodes fear and is preferentially responsive to negative stimuli. However, some studies have shown greater amygdala activation to positive compared with neutral stimuli. The first goal of this study was to more fully characterize the amygdala response to face trustworthiness by modeling its activation with both linear and nonlinear predictors. Using fMRI, we report a nonmonotonic response profile, such that the amygdala responds strongest to highly trustworthy and highly untrustworthy faces. This finding complicates future attempts to make inferences about mental states based on activation in the amygdala. The second goal of the study was to test for modulatory effects of image spatial frequency filtering on the amygdala response. We predicted greater amygdala sensitivity to face trustworthiness for low spatial frequency images compared with high spatial frequency images. Instead, we found that both frequency ranges provided sufficient information for the amygdala to differentiate faces on trustworthiness. This finding is consistent with behavioral results and suggests that trustworthiness information may reach the amygdala through pathways carrying both coarse and fine resolution visual signals. ■

INTRODUCTION

The amygdala is generally thought to be sensitive to emotional stimuli, with a particular emphasis on threatening or fear-related stimuli. This view has been supported by a wide range of studies using faces (Whalen et al., 2001; Morris et al., 1996; Adolphs, Tranel, Damasio, & Damasio, 1994) and emotionally valenced memories (Phelps & Anderson, 1997). Amygdala selectivity toward threatening stimuli has led some researchers to suggest its use as a biological marker of negative attitudes and emotions (Eberhardt, 2005). However, a few neuroimaging studies have reported greater amygdala activation for both positive and negative stimuli compared with neutral stimuli (Somerville, Wig, Whalen, & Kelley, 2006; Kim et al., 2004; Pessoa, McKenna, Gutierrez, & Ungerleider, 2002; Hamann, Ely, Grafton, & Kilts, 1999; Breiter et al., 1996). Pessoa et al. (2002) and Breiter et al. (1996) used fMRI to show greater amygdala activation for happy and fearful faces compared with neutral faces. Somerville et al. showed greater activation to faces paired with positive and negative biographical information compared with neutral biographical information. Hamann et al. (1999) used PET to show that correlations between amygdala activation during episodic memory encoding and subsequent retrieval were greater for emotionally negative and positive memories than emotionally neutral memories.

The first goal of this study was to characterize the amygdala response to one psychological dimension by modeling its activation with both linear and nonlinear predictors (Buchel, Holmes, Rees, & Friston, 1998). Nonlinear effects along a single dimension have been previously reported in the OFC response to monetary reward (Elliott, Newman, Longe, & Deakin, 2003) and in the amygdala response to attractiveness (Winston, O'Doherty, Kilner, Perrett, & Dolan, 2007).

We used a stimulus set of neutral faces that varied along the dimension of perceived trustworthiness. Trustworthiness judgments from faces are of interest because as argued by Engell, Haxby, and Todorov (2007), these judgments may reflect a general valence evaluation of faces. Valence evaluation permeates social judgments (Kim & Rosenberg, 1980; Rosenberg, Nelson, & Vivekananthan, 1968; cf. Osgood, Suci, & Tennenbaum, 1957) and is directly linked to approach/avoidance responses (Todorov, 2008; Chen & Bargh, 1999). Oosterhof and Todorov (submitted) recently provided evidence that trustworthiness judgments can serve as a good approximation of the valence evaluation of faces. In a principal components analysis of 14 different trait judgments of emotionally neutral faces, trustworthiness judgments had the highest loading on the first principal component, which accounted for more than 60% of the variance.

Princeton University

This component can be interpreted as a valence evaluation because all positive judgments loaded positively and all negative judgments loaded negatively. Trustworthiness judgments were highly correlated with this component even when the PCA excluded trustworthiness judgments.

The findings that trustworthiness judgments approximate the valence evaluation of faces are consistent with the findings from both patients with bilateral amygdala lesions (Adolphs, Tranel, & Damasio, 1998) and functional neuroimaging studies showing that the amygdala is involved in these judgments (Engell et al., 2007; Winston, Strange, O'Doherty, & Dolan, 2002). Specifically, both neuroimaging studies showed an increase in amygdala activation with decreasing facial trustworthiness. An increased response to untrustworthy faces is consistent with a prevailing view of the amygdala's function because highly untrustworthy faces may serve as threatening, negatively valenced stimuli in social interactions. However, when considering evidence that the amygdala also responds to positively valenced stimuli, we may expect an elevated response to highly trustworthy faces as well. We test this hypothesis by using faces from a wide range of perceived trustworthiness ratings. We created composite faces from the set used in Engell et al. (2007) to expand the range of perceived trustworthiness ratings.

The second goal of this study was to examine modulatory effects of spatial frequency information. There is evidence that perception of emotional expressions and facial identity is sensitive to different bands of spatial frequency information. Whereas perception of the former seems to depend on low spatial frequency (LSF) information (Schyns & Oliva, 1999), the latter seems to depend on both broad spatial frequency (BSF) and high spatial frequency (HSF) information (Nasanen, 1999; Costen, Parker, & Craw, 1996). This distinction is important because coarse LSF information may be carried through subcortical or magnocellular pathways to the amygdala. In fact, previous work on emotional expressions has shown that the amygdala response to fearful faces is driven mostly by LSF components (Vuilleumier, Armony, Driver, & Dolan, 2003; Winston, Vuilleumier, & Dolan, 2003), an effect that may confer an adaptive advantage for rapid or peripheral perception of fearful faces. LSF information about expressions may be carried through neurons in either subcortical (Hamm et al., 2003) or cortical (Bar, 2003; Pessoa et al., 2002) pathways.

To our knowledge, no neuroimaging studies have investigated spatial frequency modulation of amygdala activation to trustworthiness or other neutral face traits. In contrast to emotional expressions, which can be reduced to a few facial features (Whalen et al., 2004; Ekman, Friesen, & Ellsworth, 1972; Izard, 1971), trait judgments from neutral faces are more complex and not reducible to a few features (Todorov, Pakrashi, Loehr, & Oosterhof, submitted). Nevertheless, behavioral studies show rapid discrimination of neutral face trustworthiness (Willis & Todorov, 2006) and threat (Bar, Neta, & Linz, 2006). We, therefore, predicted that amygdala activation would better track face trustworthiness, which is rapidly detected, in the LSF and BSF components of the images compared with the HSF components.

METHODS

Stimuli

To create a set of facial images with an increased range of perceived trustworthiness, we composed novel faces from facial parts that had been rated on trustworthiness. Specifically, we started with a standardized set of neutral faces (Lundqvist, Flykt, & Ohman, 1998) that were used in the research of Engell et al. (2007). Each of the 66 faces (33 females) was divided into four facial parts (forehead, eyes, nose, and mouth), and each facial part was rated by 30 subjects on trustworthiness. Subjects were drawn from the Princeton University community and received course credit. Subjects were told to use their "gut instinct" to decide whether the person to whom the face segment belonged was trustworthy or not. The procedures are described in Todorov et al. (submitted). Based on the mean trustworthiness ratings, the facial parts were ranked according to their perceived trustworthiness for male and female faces. Then, facial parts with the same rank were combined to create a novel face with the constraint that these were parts of faces from the same gender. The final images were converted to grayscale and stripped of extraneous features, including hair. The edges of the faces were smoothed and blemishes were erased (Figure 1A and D).

To validate the novel composite faces, 39 subjects rated the faces on perceived trustworthiness using a 9-point Likert scale, ranging from 1 (not trustworthy at all) to 9 (extremely trustworthy). The mean trustworthiness judgments were highly correlated with the rank of the faces determined by the rank of the facial parts, $r = .79$, $p < .001$. The judgments were highly reliable (Cronbach's $\alpha = .95$), and the interrater agreement ($r = .37$) was higher for these faces than the original set of standardized faces ($r = .28$; Engell et al., 2007). We also asked the same 39 subjects to rate the faces on attractiveness on a 9-point scale, ranging from 1 (not attractive at all) to 9 (extremely attractive). In contrast to judgments of trustworthiness, the interrater agreement for judgments of attractiveness was lower for the composite faces ($r = .32$) than the agreement for the original faces ($r = .35$). These findings suggest that we were successful in optimizing face differences on the dimension of trustworthiness.

The 10 most trustworthy and the 10 most untrustworthy composite faces of both sexes were then selected for the fMRI study for a total of 40 individual faces. These faces are available from the authors upon request.

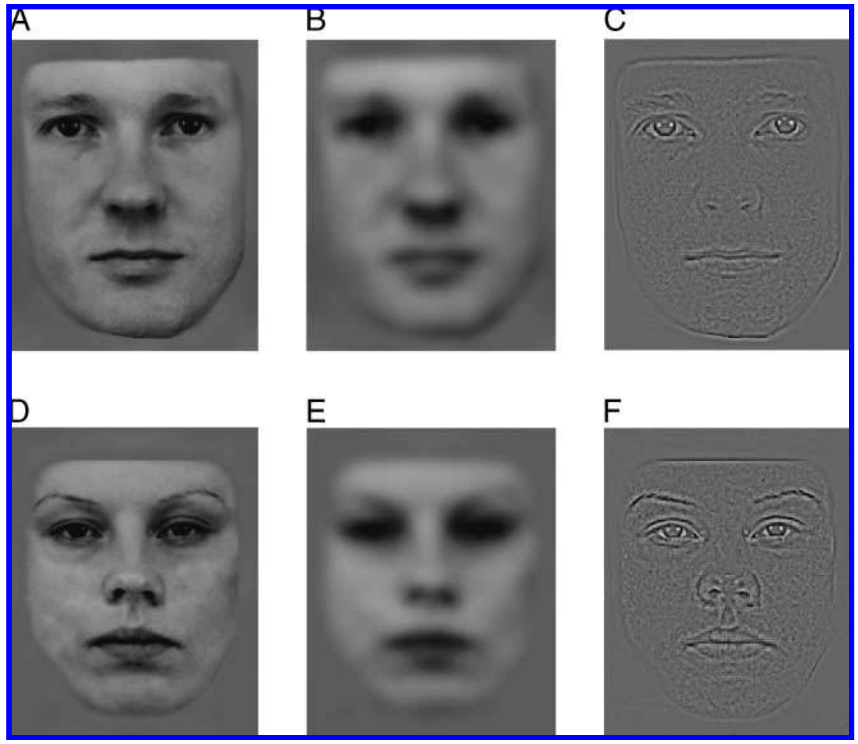

Figure 1. Example faces. (A) Original image. (B) Low-pass filtered image with a smooth cutoff of 8 cpi in the vertical direction. (C) High-pass filtered image with a smooth cutoff of 24 cpi in the vertical direction. (D–F) Similar examples from another face.

To create the LSF images, all faces were passed through a second-order Butterworth filter with a low-pass cutoff of eight cycles per image (cpi) in the vertical direction (Figure 1B). HSF images were similarly filtered with a high-pass cutoff of 24 cpi (Figure 1C). The original images were used as BSF stimuli. To eliminate luminance confounds, all faces were normalized to an average pixel intensity of 99 units on an 8-bit integer scale. Each face comprised 280 × 350 pixels and was projected through the bore of the magnet onto an angled mirror placed above the eyes. Each image subtended 8.0° of the visual field horizontally and 20.0° vertically.

Neuroimaging Subjects

Thirty-two subjects who had not participated in the face rating behavioral study (15 females, mean age = 22.8, *SD* = 6.8, range = 18–55) were recruited from the Princeton University community. One subject was excluded from this set for excessive head motion. All subjects were right-handed and had normal or corrected-to-normal vision. All subjects gave informed consent prior to the experiment and were fully debriefed at its completion in accordance with the policies of Princeton University's Institutional Review Panel.

Experimental Design and Imaging Paradigm

We used an event-related design consisting of 10 runs of 4 min 20 sec each, during which subjects rated a series of faces on perceived trustworthiness using button presses and a 4-point Likert scale. Subjects were asked to use their "gut instinct" and were told to expect images from the three different spatial frequency categories. For each of the three levels of frequency filtering, each of the 40 faces was presented twice, for a total of 240 events. Face gender, filtering level, and trustworthiness were counterbalanced across runs, such that an equal number of stimuli from each level of each factor was present in each run. We chose a relatively low face stimulus presentation interval of 200 msec to prohibit saccades. Such eye movements might confound activation in the superior colliculus, a region involved in eye movement control but which also belongs to the proposed subcortical visual stream (Morris, Ohman, & Dolan, 1999). Stimuli were presented for 200 msec followed by an 1800 msec fixation on green cross hairs during which subjects performed the button press rating. A white cross hair fixation of 4000 or 8000 msec then served as a variable ISI until the next stimulus.

The BOLD signal was used as a measure of neural activation (Kwong et al., 1992; Ogawa, Lee, Nayak, & Glynn, 1990). EPI were acquired with a Siemens 3.0-T Allegra scanner (Siemens, Erlangen, Germany) with a standard head coil (TR = 2000 msec, TE = 33 msec, flip angle = 90°, matrix size, 64 × 64). Whole brain coverage was achieved with 34 interleaved 3.6-mm axial slices with an interslice gap of 0.36 mm. At the beginning of each scan session, a high-resolution anatomical image was acquired (T1-MPRAGE, TR = 2500 msec, TE = 33 msec, flip angle = 8°, matrix size = 256 × 256).

Image Analysis

Data analysis was performed with the Analysis of Functional NeuroImages (Cox, 1996) unless otherwise indicated. After discarding the first five EPI images from each run to allow the MR signal to reach steady-state equilibrium, the remaining images were slice time corrected and then motion corrected to the first image of each run using an eight-parameter 3-D motion correction algorithm. Transient spikes in the signal were removed with the AFNI program 3dDespike. The images were then warped to the ICBM152 EPI template using the nonlinear spatial normalization algorithm provided by SPM5 (Wellcome Department of Imaging Neuroscience, London; http://www.fil.ion.ucl.ac.uk/spm). A 4-mm full width at half maximum (FWHM) smoothing kernel was then applied to the images before conversion to percent signal change from the mean.

To determine linear and quadratic effects of prior trustworthiness ratings, polynomial regression was performed according to the procedure described in Buchel et al. (1998). First, we created a zero-order time series indicating the presence of a face. We then created a first-order linear time series containing the original trustworthiness ratings and orthogonalized it to the zero-order time series. Both time series were convolved with an ideal hemodynamic response and entered into the general linear model (GLM). Additional time series representing subject head movement, time-dependent linear trends caused by scanner drift, and time-dependent quadratic trends caused by scanner drift were included as regressors of no interest. A t test was then performed on the parameter estimates supplied by the GLM from each subject to generate group level statistical parametric maps showing voxels that varied linearly with trustworthiness. Next, the same procedure was repeated with a quadratic time series containing the squared trustworthiness ratings orthogonalized to the zero-order and linear time series to show voxels that varied quadratically with face trustworthiness. Statistical parametric maps from both models were then thresholded at an uncorrected voxelwise p level of .01. Because we made a priori predictions about the amygdala, corrected significance of clusters was determined by a Monte Carlo simulation of null-hypothesis data in each amygdala. This simulation indicated that a minimum cluster size of 249 mm^3 was required to achieve corrected significance of $p <$ .05. For brain regions not part of our original prediction, a whole brain Monte Carlo simulation indicated a minimum cluster size of 2602 mm^3. To test for laterality effects, the parameter estimates for each subject were averaged across voxels within each amygdala, and a t test was performed to determine differences in mean linear response between left and right amygdala and differences in mean quadratic response between left and right amygdala.

To generate Figure 2C–D, faces were ranked by trustworthiness ratings and then binned into 10 trustworthiness groups. Ten regressors indicating the presence of a face from a particular bin were then entered into a GLM. The beta values for each bin were then averaged across subjects and plotted against the average

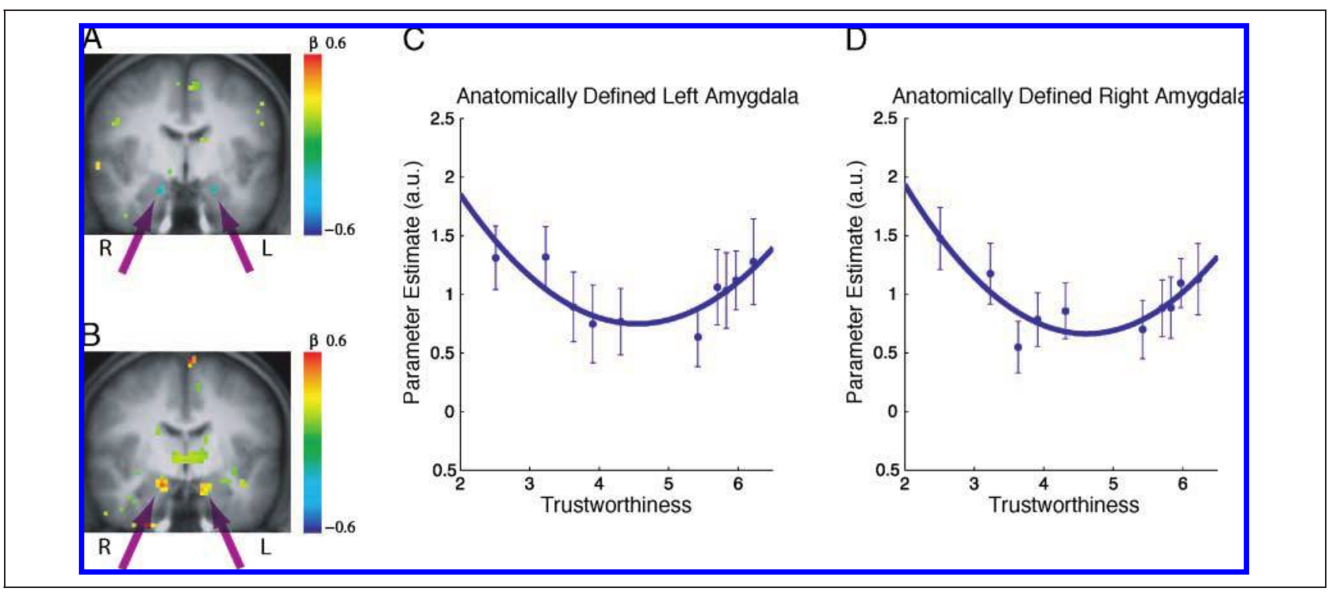

Figure 2. Amygdala response to trustworthiness with voxel-wise $p < .01$. (A) Coronal slice ($y = 6$ mm) showing small clusters in both amygdalae with a linear response to trustworthiness. The blue shading indicates more activation as trustworthiness decreases. These clusters were not large enough to pass significance after correction for multiple comparisons. (B) Coronal slice ($y = 8$ mm) showing significant clusters with quadratic effects of trustworthiness. (C and D) Results of a separate analysis on the anatomically defined left and right amygdala. Individual faces were binned into groups of four and collapsed across spatial frequency. The average trustworthiness rating of each bin is plotted against the average amygdala response, and a second-order polynomial with a linear term is fit to the plot. Error bars represent the *SEM*.

trustworthiness rating for each bin. This analysis does not make prior assumptions about the shape of amygdala response to face trustworthiness.

To test for modulatory effects of spatial frequency, we created a new model consisting of three time series indicating the presence of HSF, LSF, or BSF faces; two time series representing linear and quadratic transformation of trustworthiness; two time series representing the interaction of linear values with HSF and LSF faces; and two time series representing the interaction of quadratic values with HSF and LSF faces. Additionally, regressors of no interest were entered into the model, including eight time series representing subject head movement and two time series representing time-dependent linear and quadratic trends caused by scanner drift. These time series were then entered into a GLM, and the parameter estimates supplied by the regression were then extracted for further analysis.

RESULTS

Behavioral Results

For all three spatial bands, the trustworthiness judgments were highly reliable (Table 1). Subjects discriminated faces on trustworthiness for all spatial bands. The mean trustworthiness judgments for the BSF faces significantly correlated with the control trustworthiness judgments of unfiltered images collected outside the fMRI study as described above ($r = .90$ $p < .01$). Similarly, the mean trustworthiness judgments for the HSF faces correlated with control judgments ($r = .79$, $p < .01$), as did the mean trustworthiness judgments for the LSF faces ($r = .83$, $p < .01$).[1]

Consistent with the correlations to ratings made by control subjects outside the scanner, we also found significant correlations among the trustworthiness ratings made in the scanner. In these comparisons, there was greater agreement between the ratings of HSF faces and BSF faces than between the ratings of LSF faces and BSF faces. The mean judgments in the HSF condition correlated .91 with the mean judgments in the BSF condition, whereas the correlation between the mean judgments in the LSF and BSF conditions was .85. Using Williams' test for dependent correlations (Williams, 1959), this difference was significant, $t(37) = 2.1$, $p < .05$.

Table 1. Behavioral Data from the Three Spatial Frequency Conditions (Mean ± *SD*)

	Cronbach's α	*Correlation to BSF Ratings*	*Trustworthiness Rating (Scale of 1–4)*	*Reaction Time (msec)*
HSF	.94	.91	2.4 ± 0.42	1076 ± 180
LSF	.90	.85	1.9 ± 0.31	1088 ± 202
BSF	.97	(1)	2.8 ± 0.54	1085 ± 162

The same findings were obtained when the analysis was conducted on correlations for individual subjects. The average correlation between trustworthiness ratings of HSF images and BSF ratings was 0.45 (standard deviation [*SD*] = 0.19). The average correlation between trustworthiness ratings of LSF images and BSF images was 0.34 ($SD = 0.22$). Following a Fisher's z transformation on the set of correlations, a t test showed that the average correlation between HSF ratings and BSF ratings was significantly higher than the average correlation between LSF ratings and BSF ratings, $t(31) = 3.03$, $p < .01$.

Imaging Results

The right amygdala showed greater activation as trustworthiness decreased, although the cluster size was only marginally significant, corrected $p < .10$ (Figure 2A). A cluster in the left amygdala showing similar trends was not significant. There were no significant differences between the linear effect in the right amygdala and the left amygdala, $t < 1$. No other clusters with linear trends were large enough to pass corrections for multiple comparisons across the whole brain, although the largest nonamygdala clusters all showed positive linear trends, in contrast to clusters in both amygdalae (Table 2). The regions with other clusters included the medial prefrontal cortex (mPFC), the right caudate head, the left middle frontal gyrus, and the right posterior cingulate.

To test for nonlinear effects, we performed a second-order polynomial regression that treated zero-order and linear terms as regressors of no interest. As described above, group analysis on the second-order coefficients reveals regions showing a quadratic response to trustworthiness after the variance of linear effects has already been taken into account (Buchel et al., 1998). Significant clusters appeared in both the left and the right amygdala (Figure 2B). The positive sign of the parameter estimates within these clusters indicates that the response was highest at the extremes of the trustworthiness but lower in the middle of the trustworthiness scale. These clusters were larger and slightly more posterior than the clusters that varied linearly with trustworthiness. There were no significant effects of laterality across amygdalae, $t < 1$. Other large clusters displaying a quadratic response profile included the thalamus bilaterally, the right supramarginal gyrus, and the cerebellum (Table 3).

Figure 2C–D shows the results of a separate analysis on the anatomically defined left and right amygdala, in which individual faces were binned into groups of four and collapsed across spatial frequency. As noted in the Methods section, this analysis does not make prior assumptions about the shape of the amygdala response to face trustworthiness and, hence, does not bias the analysis toward finding a nonlinear relationship. Nevertheless, as shown in the plot of the average trustworthiness

Table 2. Brain Regions Containing the Four Largest Clusters That Show a Linear Response to Trustworthiness

	Volume (mm³)	*x*	*y*	*z*	*Mean t Value*
Positive linear relation with trustworthiness					
mPFC	1996	3.4	61.2	6.6	3.07
Right caudate head	1533	10.9	16.8	0.7	3.48
Left middle frontal gyrus	1461	−27.8	4.5	57.8	3.22
Right posterior cingulate	998	9.2	−41.7	27.4	3.16
Negative linear relation with trustworthiness					
Right amygdala	143*	18.7	−5.3	−18.2	2.94
Left amygdala	71	−21.0	−6	−17.2	2.93

The amygdala clusters are the only one to show greater activation as trustworthiness decreased. Cluster center of mass points are reported in Montreal Neurological Institute (MNI) coordinates.

*$p < .10$ small volume corrected (see Image Analysis section).

ratings of each bin against the average amygdala response (Figure 2C–D), the amygdala response changed as a quadratic function of face trustworthiness. This pattern was confirmed in an analysis of the amygdala's response at the level of individual subjects. The quadratic contrast was significant for both the right and the left amygdala, $F(1, 31) = 9.51$, $p < .004$, and $F(1, 31) = 7.76$, $p < .009$, respectively.

To test whether the quadratic term coefficients for the BSF, HSF, and LSF faces were significantly higher than zero in the amygdala, we created functional masks of amygdala clusters showing a quadratic response. These masks were unbiased with respect to spatial frequency because they were defined by the original regression model, which did not model spatial frequency category. The masks were intersected with anatomical masks of the amygdalae to ensure that the analysis was limited to voxels in the amygdala. Then, for each subject, we extracted the mean parameter estimates for the three categories of faces from a regression model that modeled the interactions of spatial frequency and trustworthiness. These parameters estimates were significantly higher than zero, $p < .05$. The only test that did not reach significance was for the HSF faces in the left amygdala, $t(31) = 1.83$, $p = .077$. However, the test was significant for the parameter estimates averaged across the left and right amygdala. To test whether the parameter estimates were significantly different for the three categories of faces, we submitted them to a 3 (Spatial frequency: BSF vs. HSF vs. LSF) × 2 (Laterality: left vs. right amygdala) ANOVA. None of the effects approached significance in this analysis.

To illustrate the predicted amygdala response in the voxels showing a significant quadratic trend, we averaged the parameter estimates for the presence of a face,

Table 3. Brain Regions Responding Quadratically to Trustworthiness

	Volume (mm³)	*x*	*y*	*z*	*Mean t Value*
Positive quadratic relation with trustworthiness					
Bilateral thalamus	4883**	0.2	−14.2	8.8	3.24
Right supramarginal gyrus	2922*	58.4	−43.3	35.9	3.16
Left uvula	2780*	−15.7	−69.5	−40.9	3.21
Left lateral amygdala/hippocampus	2780*	−31.1	−12.2	−11.7	3.17
Right amygdala	499****	18.2	−8.1	−12.9	3.22
Left amygdala	249***	−14.1	−7.3	−16.3	3.10

For all clusters, the parameter estimate on the quadratic term was positive, indicating greater activation for the extremes of the trustworthiness scale. Cluster center of mass points are reported in Montreal Neurological Institute (MNI) coordinates.

*$p < .05$.

**$p < .01$.

***$p < .05$ small volume corrected (see Image Analysis section).

****$p < .01$ small volume corrected.

the linear term, and the quadratic term across subjects for each of the three spatial frequency bands. Figure 3 plots the predicted response of the amygdala for the three spatial frequency bands as a function of the perceived trustworthiness of faces.

DISCUSSION

In this study, we report two novel findings. First, the amygdala response to face trustworthiness has linear and nonlinear components, with greatest responses to highly trustworthy and highly untrustworthy faces. Second, both LSF and HSF components of the image provide sufficient information for the amygdala to differentiate faces on trustworthiness.

The nonlinear nature of the response may be of particular relevance to researchers who use information about brain activation to make inferences about mental states. Although such inferences can be informative, particularly when behavioral data are unreliable, it is important to recognize that many brain structures have complex response profiles that may be overdetermined or nonmonotonic. Our finding of similar amygdala responses to stimuli on both ends of a psychological scale should complicate future inferences of this sort. The nonlinear response profile is inconsistent with functional descriptions of the amygdala as a detector of fear or negative stimuli. These results are consistent with the idea that the amygdala detects salience (Sander, Grafman, & Zalla, 2003) and directs attention toward emotionally relevant stimuli (Vuilleumier, 2005).

We note, however, that the nonlinear component found in our data appears to be stronger than that observed in previous studies on face trustworthiness. Although previous studies have not explicitly tested for nonlinear effects, published figures in Winston et al. (2002) appear to show mostly linear trends, and figures in Engell et al. (2007) show a trend that appears to be strongly linear. There may be two reasons for the discrepancy between our study and previous findings. First, we used a set of composite faces that were selected to include extremes along the trustworthiness scale, thus potentially covering a wider range than previous studies. Second, the figures described in previous articles were based on peak voxels or functional ROIs defined by a simple contrast or linear trend. Although these figures were meant to be illustrative and were not intended to rule out nonlinear effects, this procedure may bias the data extracted from these regions toward linear trends, particularly in light of our finding that cluster centers with a quadratic response are spatially displaced from cluster centers with a linear response. In our article, we use anatomically defined amygdalae to report nonmonotonic effects, thus eliminating bias toward linear trends (Figure 2C–D).

In amygdala clusters defined by significant quadratic trends (Figure 3), the response was more sensitive in the untrustworthy range than in the trustworthy range of faces, particularly in the right amygdala. That is, changes in trustworthiness on the negative end of the continuum evoked stronger changes in the amygdala response than changes on the positive end of the continuum. This suggests that the response to *untrustworthy* faces can be modeled as a negative linear trend, the trend observed in Engell et al. (2007). We conducted a conjunction analysis by isolating voxels that passed a threshold of $p <$ .05 in the linear trend analysis from both datasets. The three largest clusters showing a negative linear trend in both datasets were the left amygdala, the right amygdala, and the right inferior parietal lobule. The clusters in both amygdalae were significant at $p < .05$ after correction for multiple comparisons inside a small volume.

The finding that the amygdala is more sensitive to the detection of untrustworthy than trustworthy faces is

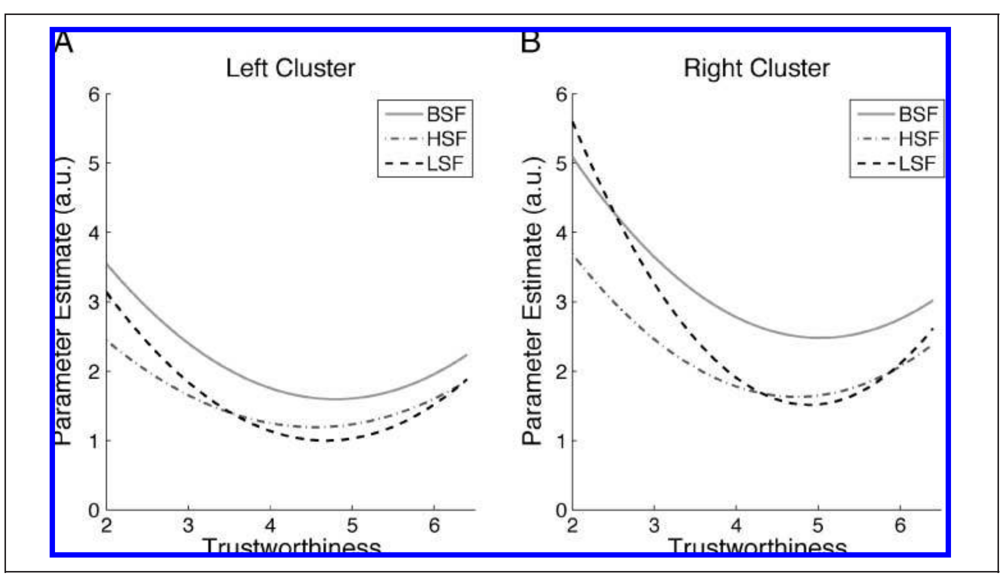

Figure 3. Modulatory effects of spatial frequency on amygdala clusters that showed a quadratic trend. (A and B) Predicted response in clusters in the left and right amygdala, respectively. Polynomial curves were generated by extracting the mean parameter estimates of zero-order, first-order, and second-order effects from each spatial frequency category (see Imaging Results section).

consistent with findings from bilateral amygdala damage patients (Adolphs et al., 1998). These patients show a specific bias to perceive untrustworthy faces as trustworthy, suggesting that damage to the amygdala is more severe for the perception of untrustworthy than trustworthy faces. The finding is also consistent with computer modeling and behavioral studies showing that people are more sensitive to differences in untrustworthiness than trustworthiness in faces (Oosterhof & Todorov, submitted).

The brain regions with the largest clusters showing a positive linear trend were the mPFC and the right caudate head. Regions with the largest clusters showing quadratic trends were the thalamus bilaterally and the right supramarginal gyrus.

The mPFC is a large region that is thought to be involved in reasoning about other minds (Amodio & Frith, 2006), although there is also evidence that it is responsive to reward value (Knutson, Taylor, Kaufman, Peterson, & Glover, 2005; Walton, Devlin, & Rushworth, 2004). The mPFC areas thought to be responsive to reward value, however, are generally more inferior than the cluster showing a linear trend in our data.

Quadratic trends in the thalamus could indicate that trustworthiness judgments can be computed in this structure. Indeed, the pulvinar—a posterior thalamic nucleus that shows considerable overlap with clusters showing quadratic trends—is thought to be part of the putative subcortical pathway to the amygdala (Morris et al., 1999). However, it is also possible that quadratic trends in the thalamus reflect attentional modulation from extensive cortico-thalamic back projections (O'Connor, Fukui, Pinsk, & Kastner, 2002).

Our second novel finding is evidence that both LSF and HSF components of the image provide sufficient information for the amygdala to differentiate face trustworthiness. We did not find evidence for the initial hypothesis that the amygdala would respond preferentially or more sensitively to BSF and LSF compared with HSF faces. Previous work on emotional expressions has shown that the amygdala response to fearful faces is driven mostly by LSF information (Vuilleumier et al., 2003). We suspect that the lack of a similar spatial frequency dissociation in our results could reflect differences between the visual properties of face fearfulness and face trustworthiness. Under this interpretation, face fearfulness may be carried in coarser visual information than face trustworthiness.

The amygdala neuroimaging results showing significant contributions from different frequency ranges are consistent with the behavioral results that ratings made on both HSF and LSF faces significantly correlated with ratings for BSF faces. We predicted that BSF ratings would correlate better with LSF ratings than with HSF. Although this prediction was confirmed when using the BSF ratings made by separate group of subject outside the scanner, the prediction was not confirmed when using the BSF ratings made inside the scanner. In both cases, however, the differences were small. In a previous study on face threat, correlations to BSF image ratings of threat were higher for LSF images compared with HSF images when masked faces were presented for 39 msec (Bar et al., 2006). However, this effect was not present for 1700 msec exposure times, perhaps due to differential recruitment of fast magnocellular pathways and slow parvocellular pathways in the two exposure time conditions. Because our study used unmasked stimuli presented for 200 msec, the small effects in our study are most likely due to the differences in exposure times and, perhaps, also to subtle differences in the dimensions of face threat and face trustworthiness.

Conclusion

The function of the amygdala is complex. Apparent discrepancies in amygdala research on valence may imply that the amygdala is sensitive to experimental parameters such as task or exposure time. Future research should systematically examine these and other effects. We have attempted to better characterize the function of the amygdala by modeling a nonlinear response function to face trustworthiness and by examining the modulatory effects of spatial frequency on this response function. Our finding that the amygdala responds most strongly to extremes on both ends of the trustworthiness scale has two methodological implications. First, weak fMRI contrasts or linear trends may be symptoms of an underlying quadratic relationship between the experimental parameter and the brain response. Second, the amygdala may not be as good an indicator of mental states as previously thought. Interpretation of its activation, as with any area of the brain, should proceed with the utmost caution.

Acknowledgments

We would like to thank Valerie Loehr and Richard Lopez for their help in preparing the composite faces and Andy Engell for his comments on an earlier draft of this article. We also thank Chris Moore, Patrik Vuilleumier, Carole Peyrin, and Valentin Wyart for their help and valuable discussions.

Reprint requests should be sent to Christopher P. Said or Alexander Todorov, Department of Psychology and Center for the Study of Brain, Mind, and Behavior, Princeton University, Princeton, NJ 08540, or via e-mail: csaid@princeton.edu. or atodorov@princeton.edu.

Note

1. Correlations between trustworthiness judgments at different spatial frequencies can be used to describe how much information about trustworthiness presented in BSF images is also carried in filtered images. Differences in the mean ratings among spatial frequencies do not address this issue but are nevertheless reported in Table 1.

REFERENCES

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature, 393,* 470–474.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature, 372,* 669–672.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268–277.

Bar, M. (2003). A cortical mechanism for triggering top–down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15,* 600–609.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6,* 269–278.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17,* 875–887.

Buchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage, 8,* 140–148.

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,* 215–224.

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics, 58,* 602–612.

Cox, R. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

Eberhardt, J. L. (2005). Imaging race. *American Psychologist, 60,* 181–190.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. New York: Pergamon Press.

Elliott, R., Newman, J. L., Longe, O. A., & Deakin, J. F. (2003). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: A parametric functional magnetic resonance imaging study. *Journal of Neuroscience, 23,* 303–307.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience, 19,* 1508–1519.

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

Hamm, A. O., Weike, A. I., Schupp, H. T., Treig, T., Dressel, A., & Kessler, C. (2003). Affective blindsight: Intact fear conditioning to a visual cue in a cortically blind patient. *Brain, 126,* 267–275.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., et al. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of Cognitive Neuroscience, 16,* 1730–1745.

Kim, M. P., & Rosenberg, S. (1980). Comparison of two structural models of implicit personality theory. *Journal of Personality and Social Psychology, 38,* 375–389.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *Journal of Neuroscience, 25,* 4806–4812.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences, U.S.A., 89,* 5675–5679.

Lundqvist, D., Flykt, A., & Ohman, A. (1998). *The Karolinska directed emotional faces*. Publication from Psychology Section, Department of Clinical Neuroscience, Karolinska Hospital, S-171, 76 Stockholm, Sweden.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383,* 812–815.

Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating ''unseen'' fear. *Proceedings of the National Academy of Sciences, U.S.A., 96,* 1680–1685.

Nasanen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research, 39,* 3824–3833.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience, 5,* 1203–1209.

Ogawa, S., Lee, T. M., Nayak, A. S., & Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine, 14,* 68–78.

Oosterhof, N. N., & Todorov, A. (submitted). *The functional basis of face evaluation.*

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

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences, U.S.A., 99,* 11458–11463.

Phelps, E. A., & Anderson, A. K. (1997). Emotional memory: What does the amygdala do? *Current Biology, 7,* R311–R314.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9,* 283–294.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences, 14,* 303–316.

Schyns, P., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition, 69,* 243–265.

Somerville, L. H., Wig, G. S., Whalen, P. J., & Kelley, W. M. (2006). Dissociable medial temporal lobe contributions to social memory. *Journal of Cognitive Neuroscience, 18,* 1253–1265.

Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. In M. Miller & A. Kingstone (Eds.), *The year in cognitive neuroscience 2008, Annals of the New York Academy of Sciences, 1124,* 208–224.

Todorov, A., Pakrashi, M., Loehr, V. R., & Oosterhof, N. N. (submitted). *Evaluating faces on trustworthiness: Automatic assessment of face valence.*

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9,* 585–594.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience, 6,* 624–631.

Walton, M. E., Devlin, J. T., & Rushworth, M. F. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience, 7,* 1259–1265.

Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S., et al. (2004). Human amygdala responsivity to masked fearful eye whites. *Science, 306,* 2061.

Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., & Rauch, S. L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion, 1,* 70–83.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society. Series B, Methodological, 21,* 396–399.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592–598.

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. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5,* 277–283.

Winston, J. S., Vuilleumier, P., & Dolan, R. J. (2003). Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. *Current Biology, 13,* 1824–1829.