

# Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala

Andrew D. Engell, James V. Haxby, and Alexander Todorov

## Abstract

Deciding whether an unfamiliar person is trustworthy is one of the most important decisions in social environments. We used functional magnetic resonance imaging to show that the amygdala is involved in implicit evaluations of trustworthiness of faces, consistent with prior findings. The amygdala response increased as perceived trustworthiness decreased in a task that did not demand person evaluation. More importantly, we tested whether this response is due to an individual's idiosyncratic perception or

to face properties that are perceived as untrustworthy across individuals. The amygdala response was better predicted by consensus ratings of trustworthiness than by an individual's own judgments. Individual judgments accounted for little residual variance in the amygdala after controlling for the shared variance with consensus ratings. These findings suggest that the amygdala automatically categorizes faces according to face properties commonly perceived to signal untrustworthiness. ■

## INTRODUCTION

People form person impressions from minimal information (e.g., Uleman, Blader, & Todorov, 2005; Todorov & Uleman, 2002, 2003, 2004; Hassin & Trope, 2000; Ambady, Hallahan, & Rosenthal, 1995; Carlston, Skowronski, & Sparks, 1995; Carlston & Skowronski, 1994; Ambady & Rosenthal, 1992), and faces are a particularly rich source of social information. One hundred milliseconds of exposure to a neutral face is sufficient for people to make a variety of trait judgments such as trustworthiness, competence, and aggressiveness (Willis & Todorov, 2006), and the time exposure can be even shorter for some of these judgments (Bar, Neta, & Linz, 2006). Trait inferences from faces are important because they often predetermine the course of social interactions, and behavioral research has documented that the effects of facial appearance on social outcomes are pervasive (Todorov, Mandisodza, Goren, & Hall, 2005; Langlois et al., 2000; Zebrowitz, 1999; Montepare & Zebrowitz, 1998; Hamermesh & Biddle, 1994).

Despite the wealth of behavioral data about the significance of trait impressions from faces, there is little research about the neural mechanisms underlying these impressions. The large body of cognitive neuroscience research on face perception focuses either on face categorization (e.g., Haxby et al., 2001; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997) and recognition of facial identity (Hoffman & Haxby, 2000) or on recognition of expressions of emotions (e.g., Adolphs, 2002; Calder, Lawrence, & Young, 2001; Blair, Morris,

Frith, Perrett, & Dolan, 1999; Sprengelmeyer, Rausch, Eysel, & Przuntek, 1998; Lane, Reiman, Ahern, Schwartz, & Davidson, 1997; Phillips et al., 1997; Morris et al., 1996). An exception is research on perceptions of trustworthiness. Adolphs, Tranel, and Damasio (1998) showed that patients with bilateral amygdala damage cannot discriminate between trustworthy- and untrustworthy-looking faces, suggesting that the amygdala plays a key role in decisions of trustworthiness. Consistent with this finding, in a functional magnetic resonance imaging (fMRI) study with normal individuals, Winston, Strange, O'Doherty, and Dolan (2002) confirmed the involvement of the amygdala in judgments of trustworthiness. Specifically, the study showed an increased amygdala response to faces that the participants subsequently rated as untrustworthy, implying that the amygdala automatically tracks the trustworthiness of faces.

The amygdala is involved in multiple psychological functions (Phelps & LeDoux, 2005) from learning of fear responses (e.g., LeDoux, 2000) and consolidation of emotional memories (McGaugh, 2004) to implicit evaluation of stimuli (Vuilleumier, 2005; Sander, Grafman, & Zalla, 2003) and providing general vigilance functions (Amaral, 2002; Davis & Whalen, 2001; Whalen, 1998). The latter functional role is entirely consistent with findings suggesting that the amygdala plays a key role in perceptions of trustworthiness in faces. Deciding whether an unfamiliar person is trustworthy is one of the most important decisions routinely faced in social environments. Perceived trustworthiness determines whether to approach or avoid the person and serves as a gating mechanism for social interactions. In this study, we sought to replicate the findings of Winston et al. (2002) using more stringent fMRI procedures and to explore the determinants of the

---

Princeton University

amygdala response to untrustworthy faces. Specifically, we tested whether this response is due to an individual's idiosyncratic perception or to face properties that are perceived as untrustworthy across individuals.

In Adolphs et al. (1998), faces were categorized as trustworthy and untrustworthy based on ratings of trustworthiness averaged across normal individuals (consensus ratings). In the study by Winston et al. (2002), the neural response to faces was modeled as a function of idiosyncratic judgments of trustworthiness collected after the fMRI session. This distinction is important because consensus trait ratings of faces reflect properties of the face stimuli that signal untrustworthiness across individuals. Individual judgments, on the other hand, reflect both face properties commonly perceived to signal untrustworthiness (consensus contributions to judgments) and face properties that signal untrustworthiness to the specific individual (idiosyncratic contributions to judgments). Psychometric studies of trait judgments from faces show that these judgments contain both shared variance with other judges and variance unique to the individual judge (Honekopp, 2006). The amygdala can respond to face properties that signal untrustworthiness to the specific individual, as reflected in their judgments of trustworthiness (Winston et al., 2002), or to face properties that signal untrustworthiness across individuals, as reflected in aggregated ratings of trustworthiness.

Although individual judgments of trustworthiness are correlated, there is large individual variation. We conducted a study, described in detail below, in which 129 participants rated a standardized set of neutral faces (Karolinska Directed Emotional Faces set; Lundqvist, Flykt, & Ohman, 1998) on trustworthiness. The average correlation between individual judgments and the mean for the remaining individuals was .52. Thus, consensus ratings of trustworthiness aggregated across individuals accounted for 27% of the variance of individual judgments. If the amygdala responds to face properties that signal untrustworthiness across individuals, consensus ratings of trustworthiness should predict the amygdala response to novel faces in individuals better than their own judgments. In other words, individual judgments should predict amygdala activity only to the extent that they share variance with consensus ratings.

The first goal of the current fMRI experiment was to investigate what better predicts an individual's amygdala response to novel faces: the individual's own judgments of trustworthiness or consensus ratings of trustworthiness (i.e., the average rating of many individuals). In addition to the main fMRI study, we conducted two behavioral experiments. In the first, we established the consensus ratings of trustworthiness of the faces used in the fMRI study and demonstrated that consensus estimates are robust with respect to the number of raters. In the second, we showed that there is reliable variance in individual judgments of trustworthiness that can be attributed to idiosyncratic face perceptions.

The second goal of the fMRI experiment was to provide a conceptual replication of the findings of Winston et al. (2002), demonstrating more conclusively that the amygdala automatically tracks the trustworthiness of novel faces. In their study, participants viewed faces and were instructed to either attend to the age or the trustworthiness of the displayed face. These two trial types were presented in different blocks and allowed the authors to compare the amygdala response to implicit (age trials) and explicit (trustworthiness trials) trustworthiness judgments. They found that the amygdala tracks the trustworthiness of faces even when the perceiver is not engaged in explicit trustworthiness judgments. However, it is possible that the original finding was due to the participants attending to trustworthiness even when instructed to attend to age. That is, once participants were instructed to make explicit trustworthiness judgments, they might have been unable to ignore this goal during the implicit (age) trials, consistent with behavioral experiments showing that goals can be easily primed (Gollwitzer & Bargh, 2005). Given the counterbalancing of the blocks across participants, implicit trials followed explicit trials for half of the participants. Collapsing across the latter implicit trials and the "pure" implicit trials (age trials presented before trustworthiness trials) might have overestimated the role of the amygdala in implicit judgments of trustworthiness. The current design did not include any explicit trustworthiness judgments until after the fMRI session had concluded.

In the current fMRI study, participants (different from those in the behavioral experiments) viewed the faces rated in the first behavioral study. Because we were interested in implicit decisions of trustworthiness, participants were told that the study was about face memory. Specifically, they were presented with blocks of novel faces and asked whether a subsequently presented test face was presented in the block (Figure 1). Thus, the fMRI task did not demand any person evaluation. Following Winston et al. (2002), individual judgments of trustworthiness were collected by having participants rate the faces after the completion of the fMRI scanning session.

## METHODS

### Behavioral Studies

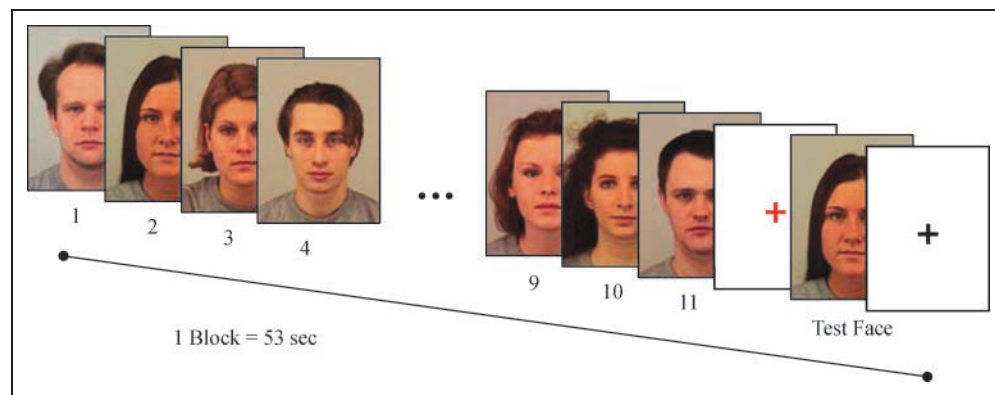
#### *Subjects*

One hundred twenty-nine undergraduate students from Princeton University participated in the first study for partial course credit. Another 15 undergraduate students participated in the second study.

#### *Stimuli*

Pictures were taken from the Karolinska Directed Emotional Faces set (Lundqvist et al., 1998). This database

**Figure 1.** An example of a stimulus presentation block. Each block comprised a series of 11 faces presented for 1 sec each. The interstimulus interval was randomly chosen (see Methods for parameters). A red fixation cross appeared following the presentation of the 11th face. At 52 sec from the start of the block, the red fixation cross was replaced with a 1-sec presentation of the “test face.” Each block concluded with a 12-sec presentation of a fixation cross.



consists of 70 photographs of amateur actors, 35 women and 35 men, between 20 and 30 years of age. In the pictures, all actors wore gray T-shirts and had no beards, mustaches, earrings or eyeglasses, or visible make-up. We used frontal head-shot photographs of individuals with neutral expressions and a direct gaze (see Figure 1 for examples). Of the 70 photographs, two photographs of men were excluded due to poor lighting quality. For the experiments, we also excluded two photographs of women in order to have an equal number of male and female photographs. The remaining 66 photographs were used in the behavioral studies and in the fMRI study.

We further conducted a validation study to establish that the faces were perceived as neutral and not as expressing emotions. Each face was categorized by 15 participants into one of seven mutually exclusive categories (angry, fearful, disgusted, neutral, surprised, happy, and sad). In addition to the neutral expressions used in the behavioral studies and the fMRI study, the validation study included the same faces expressing the six basic emotions (angry, fearful, disgusted, surprised, happy, and sad). Thus, there were 462 trials (66 faces  $\times$  7 expressions: angry, fearful, disgusted, neutral, surprised, happy, and sad). For each face, participants were asked to decide among the seven categories. The study began with a demonstration of different facial expressions and seven practice trials. The order of trials was randomized for each participant. The interstimulus interval was 1 sec.

All of the faces used in the behavioral and fMRI studies were perceived as neutral. On average, each face was categorized as neutral 95.05% ( $SD = 8$ ). Thus, categorizations of the faces as expressing emotions accounted for less than 5% of the responses. The correct categorization was significantly higher than the chance expectation of 14% (1 out of 7 responses),  $t(65) = 97.56$ ,  $p < .001$ , and all of the misclassification errors (e.g., a neutral face categorized as sad) were significantly lower than chance,  $t_s > 17.22$ ,  $p_s < .001$ . It should also be noted that none of the six possible misclassification

errors correlated with the consensus ratings of trustworthiness. For example, it was not the case that neutral faces that were categorized as angry were judged as less trustworthy.

#### *Procedures for Trait Judgments Studies*

In both studies, participants were told that the study was about first impressions and that there is no right or wrong answer. They were encouraged to rely on their “gut feeling” and to work as quickly as possible. The order of the faces was randomized for each participant. Each face was presented at the center of the screen with a question above the photograph “How trustworthy is this person?” and a response scale below the photograph. The response scale ranged from 1 (*not at all*) to 9 (*extremely*). The face was presented on the screen until the participant’s response. The next trial was presented after 1 sec. The average of the trustworthiness judgments (“consensus ratings”) from the first study was used to weight the regressors in the fMRI analysis.

In the second study, participants rated the faces twice in two different blocks in order to estimate test–retest reliability for individual participants and the variance that can be attributed to idiosyncratic face perceptions. The order of the faces was randomized within each block and participants were asked to judge an unrelated set of faces on a different trait dimension between the two trustworthiness blocks to reduce the likelihood that they would be responding by memory in the second trustworthiness block.

#### **fMRI Study**

##### *Subjects*

Sixteen participants (5 women, mean age = 22.4 years), different from the participants in the behavioral studies, were recruited from the community in and around

Princeton University. The participants were all right-handed and had normal or corrected-to-normal vision. All participants 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.

### *fMRI Task*

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 presented in random order. These were the 66 faces used in the first behavior study. The order of the faces was randomized for each participant. The faces were projected onto a screen at the rear of the bore of the magnet. Subjects viewed these images via an angled mirror attached to the radio-frequency coil and placed above their eyes. All runs began with a 12-sec presentation of a fixation cross. Within a block (see Figure 1), each of the 11 faces was presented for 1 sec in a jittered event-related fashion. Each interstimulus interval (ISI) was randomly chosen from an exponential distribution with target mean ISI = 3.5 sec and a minimum ISI = 1.5 sec. At the conclusion of each block, a red fixation cross appeared on a white screen until a predetermined time point (52 sec from the beginning of the block), at which time another face (the test face) was presented for 1 sec. 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-sec rest period in order to allow hemodynamic activity to return to baseline levels.

### *Behavioral Task*

After subjects were removed from the scanner, they were shown each of the 66 faces and asked to rate each face on trustworthiness using a Likert scale. These "individual judgments" were used to weight the regressors in the fMRI analysis.

### *Image Acquisition*

The blood oxygenation level-dependent signal was used as a measure of neural activation. Echo-planar images were acquired with a Siemens 3.0-T Allegra head-dedicated scanner (Siemens, Erlangen, Germany) with a standard "bird-cage" head coil (TR = 2000 msec, TE = 30 msec, 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 msec, TE = 4.3 msec, flip angle = 8°, matrix size = 256 × 256) was acquired for use in registering activity

to each subject's anatomy and for spatially normalizing data across subjects.

### *Image Analysis*

One subject was excluded from the analysis due to excessive head motion. Data were analyzed with analysis of functional neuroimages (AFNI) (Cox, 1996) using standard preprocessing procedures. Subject motion was corrected using a six-parameter 3-D motion-correction algorithm following slice scan-time correction. The data were low-pass filtered with a frequency cutoff of 0.1 Hz subsequent to spatial smoothing with a 6-mm full width at half minimum 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. This regressor controlled for general response to faces relative to the baseline. In order to test for event-dependent responses, a second regressor was created that scaled the predicted response by the individual or consensus rating for each face. That is, the first regressor estimated the mean response to all faces, whereas the second regressor estimated how the response to each face was modulated by trustworthiness. In addition, regressors of noninterest were included in the multiple regression model to factor out variance associated with mean, linear, and quadratic trends in each run as well as subject head motion. The nine-parameter landmark method of Talairach and Tournoux (1988) was used to spatially normalize the activation maps across subjects. An independent-samples *t* test was performed on the coefficients supplied by the multiple regression analysis for each subject to test the significance of coefficients across all subjects. We used the AlphaSim program included in AFNI in order to correct for multiple comparisons. A minimum cluster size of 162 mm<sup>3</sup> to achieve corrected significance of  $p < .05$  was determined by a Monte Carlo simulation within our region of interest (bilateral amygdala), with a voxelwise threshold of  $p < .01$ .

The initial analysis comprised two independent regression models. In the first model, the neural response to faces was modeled as a function of the individual trustworthiness judgments of these faces. In the second model, the neural response was modeled as a function of the consensus ratings of trustworthiness. In the subsequent analysis, we first removed the effect of consensus ratings on individual judgments. Specifically, for each subject, we regressed their judgments on consensus ratings and used the regression residuals as a measure of "idiosyncratic judgments" of trustworthiness. Then, we modeled the neural response as a function of both consensus ratings and idiosyncratic judgments. It should be noted that these two regressors were completely uncorrelated. This regression model yielded coefficients for the linear trends of both consen-

sus ratings and idiosyncratic judgments. These coefficients represented the modulation of the signal relative to the mean response to faces within each voxel. Our criteria for regions showing significant linear modulation were cluster size  $>270 \text{ mm}^3$  or 10 functional voxels and  $p < .005$  (uncorrected).

Finally, to compare the strength of the coefficients, we conducted a region-of-interest analysis. For each subject, we computed the average coefficients for the consensus ratings and for the idiosyncratic judgments across all voxels within an anatomically defined mask of the amygdala. The amygdala mask was defined separately for each subject by tracing the perimeter of the amygdala as seen on coronal slices. The anterior border was defined as the slice in which the gray matter of the amygdala was no longer clearly distinguishable from the uncus. From the anterior border, the mask extended posteriorly below the optic tract and above the hippocampus until the amygdala was no longer distinguishable from the hippocampus. After defining the masks for each subject, they were transformed into Talairach space in order to report the spatial coverage. Each amygdala extended on average in the anterior–posterior plane from  $y = 0.4$  ( $SD = 1.8$ ) to  $y = 10.4$  ( $SD = 1.1$ ), in the superior–inferior plane from  $z = -5.8$  ( $SD = 1.3$ ) to  $z = -23.5$  ( $SD = 1.8$ ), and laterally from  $x = \pm 9.5$  ( $SD = 1.8$ ) to  $x = \pm 31.8$  ( $SD = 2.1$ ).

These coefficients were submitted to a 2 (predictor: consensus ratings vs. idiosyncratic judgments)  $\times$  2 (laterality: left vs. right amygdala) analysis of variance.

## RESULTS

### Behavioral Data

#### Study 1

The average (consensus) ratings of trustworthiness were highly reliable (Cronbach's  $\alpha = .98$ ). The high reliability is not surprising given the large number of raters ( $n = 129$ ). However, the consensus ratings were robust with respect to the number of raters. To estimate this robustness, we did a bootstrapping simulation with sample sizes of 10, 20, and 30 raters. From the 129 raters, we randomly drew 10 samples for each of the respective sample sizes. Then, for each sample, we correlated the mean judgments (sample consensus) with the mean for all raters (consensus). The average correlations were .91 ( $SD = .02$ ) for a sample of 10 raters, .95 ( $SD = .01$ ) for a sample of 20 raters, and .97 ( $SD = .006$ ) for a sample of 30 raters. Thus, very few raters are needed to obtain a reliable estimate of consensus ratings of trustworthiness.

Although the consensus ratings were highly reliable, there was a large individual variation. To estimate the shared variance between consensus ratings and individual judgments, for each participant we computed the correlation between their judgments and the mean of the judgments of the remaining participants. Then, we

computed the mean and standard deviation ( $SD$ ) for the 129 correlations. The mean correlation was .52 and the  $SD$  was .18. Thus, consensus ratings of trustworthiness aggregated across individuals accounted for 27% of the variance of individual judgments. As in the case of the stability of consensus ratings, data simulations showed that the correlations between individual judgments and consensus ratings were robust with respect to the number of raters needed to estimate the consensus. First, we randomly divided the sample of 129 raters into two groups of 15 and 114 raters, respectively. We used the judgments of the 15 raters (this was the sample size for the fMRI study) to estimate the average correlation between individual judgments and consensus ratings. From the remaining raters, we randomly drew 10 samples for each of the sample sizes of 10, 20, and 30 raters to estimate the consensus. The average correlations between individual judgments and consensus ratings were .47 ( $SD = .02$ ) for a sample of 10 raters, .50 ( $SD = .01$ ) for a sample of 20 raters, and .51 ( $SD = .02$ ) for a sample of 30 raters.

#### Study 2

Although there is prior evidence that individual trait judgments from faces reflect both meaningful idiosyncratic and consensus contributions (Honekopp, 2006), we conducted a behavioral study, using the same analytic procedures as in the fMRI study, to estimate whether the variance attributed to idiosyncratic components is meaningful. The average correlation between the individual judgments at Time 1 and Time 2 was .49 ( $SD = .14$ ). The average correlation between the trustworthiness judgments at Time 1 and the consensus ratings was .47 ( $SD = .11$ ). More importantly, the correlation between the consensus ratings and the individual judgments at Time 2 was .51 ( $SD = .14$ ). Thus, consensus ratings were as strong predictor of individual judgments at Time 2 as the individual judgments at Time 1. Because individual judgments share variance with consensus ratings, we removed the effect of consensus on individual judgments (at both Time 1 and Time 2) in regression analyses conducted for each rater. We used the residuals of these analyses as a measure of the idiosyncratic contributions to individual judgments. The residuals at Time 1 predicted the residuals at Time 2, suggesting that idiosyncratic contributions to individual judgments reflect reliable variance attributable to the individual rather than to measurement error only. The correlations between residuals at Time 1 and residuals at Time 2 were positive for all 15 raters, and the average correlation ( $r = .32$ ,  $SD = .13$ ) was significantly higher than zero,  $t(14) = 8.76$ ,  $p < .001$ , although smaller than the correlations between consensus and judgments at Time 2,  $t(14) = 4.61$ ,  $p < .001$ , and judgments at Time 1 and judgments at time 2 including the effect of consensus,  $t(14) = 9.32$ ,  $p < .001$ . (For all significance tests

involving correlations, we used the Fisher  $z$  transformations of the raw correlations.)

### fMRI Participants

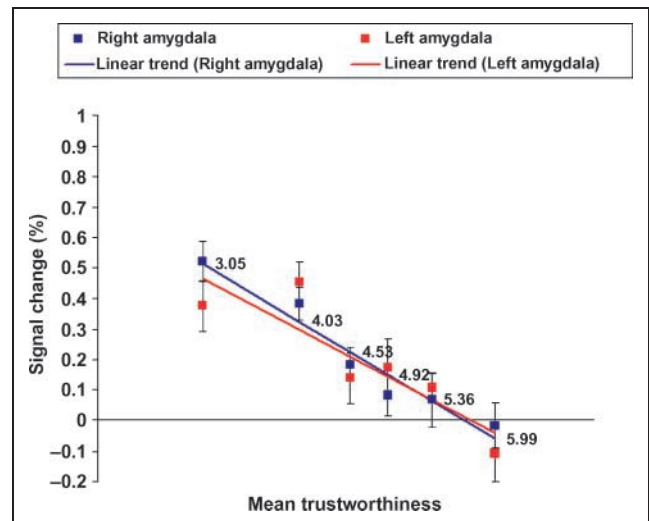
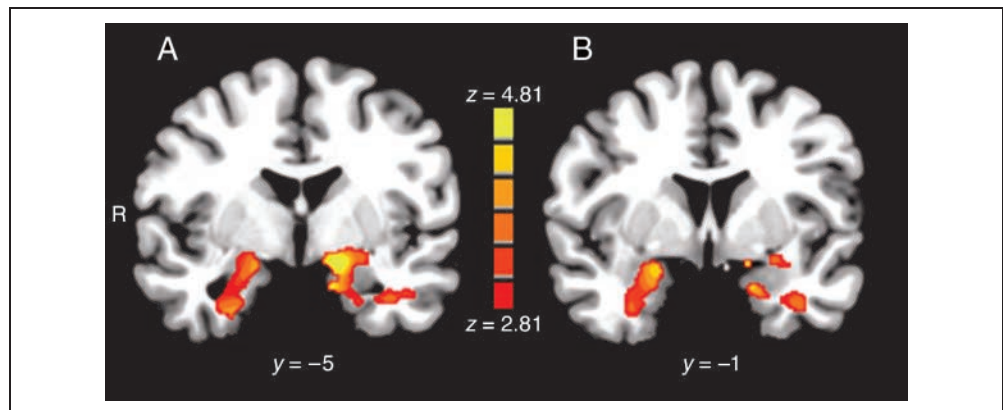
The reliability of the judgments of trustworthiness of the 15 fMRI participants was high (Cronbach's  $\alpha = .81$ ). The correlations between the judgments of individual participants and the consensus ratings from the first behavioral study ranged from  $.26, p < .036$  to  $.66, p < .001$ , with an average correlation of  $.46$  ( $SD = .12$ ). These correlations were comparable to the correlations from the behavioral study.

### fMRI Data

In two separate regression analyses, the amygdala response was modeled as a linear function of individual judgments and of consensus ratings. As shown in Figure 2, the consensus ratings predicted a linear response in the left and right amygdala ( $p < .005$ , uncorrected;  $p < .05$  corrected within region of interest). The amygdala response increased as the consensus ratings of trustworthiness decreased in both regions (Figure 3).

Replicating Winston et al. (2002), as shown in Figure 4B, the individual judgments predicted a linear response in the left amygdala ( $p < .01$ , uncorrected for multiple comparisons). The amygdala response to novel faces increased as the subjectively judged trustworthiness decreased. Although both individual judgments and consensus ratings predicted the linear trend, the area predicted by the consensus judgments was larger and statistically more robust, as can be seen in the comparison of Figure 4B and C. The slice used for Figure 4 is posterior to the maximum activation for the consensus ratings (Figure 2 and Table 1), but this was the slice on which we can depict both the activations for consensus ratings and individual judgments.

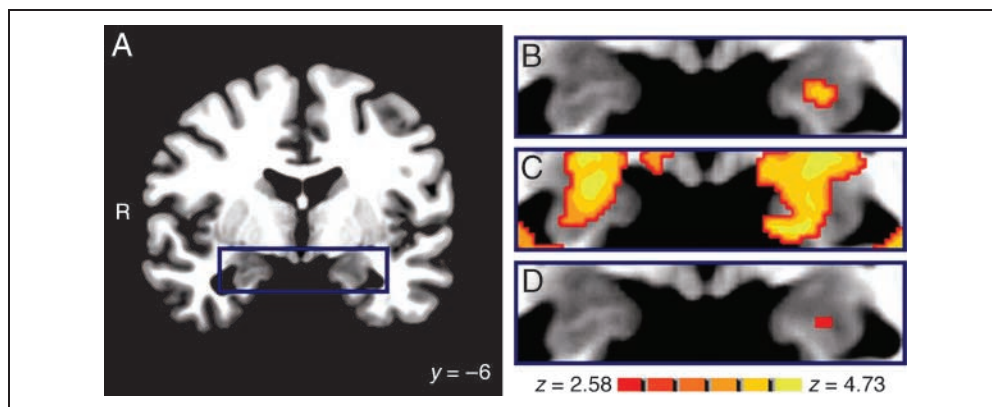
**Figure 2.** Implicit amygdala response as a function of consensus ratings of trustworthiness. The results of a  $t$  test performed on the coefficients of the linear trend regressors yielded by the regression analysis on the individual data from each of 15 subjects for the consensus ratings of trustworthiness. Activity is overlaid on slices of a standardized brain containing the peak response for the (A) left amygdala and for the (B) right amygdala. The activity extends into the medial temporal lobe including the parahippocampal gyrus and the uncus.



**Figure 3.** Blood oxygenated level-dependent response (% signal change) in voxels in the amygdala showing a significant linear trend as a function of consensus ratings of trustworthiness of novel faces. To extract percent signal change, faces were divided into six categories of 11 faces according to their perceived trustworthiness. For example, the mean for the 11 least trustworthy faces was 3.05 ( $SD = .45$ ) and the mean for the 11 most trustworthy faces was 5.99 ( $SD = .21$ ) on a scale ranging from 1 (*untrustworthy*) to 9 (*trustworthy*).

As demonstrated above, individual judgments correlated with consensus ratings. To test whether the effect of individual judgments can be accounted for by their shared variance with consensus ratings, we removed the effect of consensus ratings from the individual judgments in a regression analysis. The residuals of this analysis are a measure of idiosyncratic perceptions of trustworthiness uncorrelated with consensus ratings. Modeling the amygdala response as a linear function of these idiosyncratic judgments revealed a negligible cluster of voxels in the left amygdala ( $p < .01$ , uncorrected for multiple comparisons; Figure 4D).

**Figure 4.** Implicit amygdala response as a function of perceived trustworthiness of faces. The results of a  $t$  test performed on the coefficients of the linear trend regressors yielded by the regression analysis on the individual data from each of 15 subjects. (A) Amygdala region of a standardized brain. Areas in the amygdala showing a significant linear change as a function of (B) individual judgments of trustworthiness of novel faces, (C) consensus ratings of trustworthiness, and (D) individual judgments of trustworthiness after removing shared variance with consensus ratings.



To test whether the regression coefficients for the linear trend are significantly different, we conducted a region-of-interest analysis on all voxels in the amygdala that showed a linear trend at  $p < .05$  for either the consensus ratings or the idiosyncratic judgments. We compared across subjects the average coefficients in a 2 (predictor: consensus vs. idiosyncratic)  $\times$  2 (laterality:

left vs. right amygdala) analysis of variance. This analysis showed that the consensus ratings coefficients ( $M = -0.10$ ,  $SD = 0.04$ ) were significantly larger than the idiosyncratic judgments coefficients ( $M = -0.06$ ,  $SD = 0.06$ ),  $F(1, 28) = 9.04$ ,  $p = .006$ ; other  $F$ s  $< 1$ . Because there were more significant voxels for the linear trend of consensus ratings, the above test biases the analysis in

**Table 1.** Regions Showing Significant Linear Modulation Correlated with Consensus Ratings and Idiosyncratic Judgments

	<i>t</i> Value	Cluster Size ( $\text{mm}^3$ )	<i>x</i>	<i>y</i>	<i>z</i>
<i>Regions showing significant linear modulation correlated with consensus ratings</i>					
Right amygdala <sup>a</sup>	6.83*	794	24	-1	-18
Left amygdala <sup>b</sup>	4.93*	595	-16	-5	-19
Right parahippocampal gyrus	6.13*	831	27	-28	-15
Left parahippocampal gyrus	6.22*	875	-36	-25	-15
Left uncus/parahippocampal gyrus	5.69*	1182	-30	1	-25
Right middle temporal gyrus	4.94*	807	54	-67	2
Left inferior temporal gyrus	4.76*	355	-52	-65	-9
Right middle occipital gyrus	5.08*	554	35	-87	5
Left middle occipital gyrus	6.00*	2514	-29	-84	2
Left middle occipital gyrus	5.25*	308	44	-70	-6
Right cuneus	4.72*	277	11	-94	22
<i>Regions showing significant linear modulation correlated with idiosyncratic judgments</i>					
Left middle frontal gyrus	4.74*	307	-43	54	10

Results are from the multiple regression model with both consensus and idiosyncratic predictors in the model ( $p = .005$ , cluster size  $> 270 \text{ mm}^3$  or 10 functional voxels). The coordinates of the voxel with maximum  $t$  value are reported in Talairach space.

<sup>a</sup>This activation extended into the right uncus (full volume =  $1677 \text{ mm}^3$ ).

<sup>b</sup>This activation extended into the left uncus (full volume =  $858 \text{ mm}^3$ ).

\* $p < .001$ .

favor of these ratings. In order to remove this bias, we conducted the same analysis on all voxels in the amygdala. Although the average regression coefficients were substantially reduced, the difference between the coefficients for the consensus and idiosyncratic judgments ( $M = -0.05$ ,  $SD = 0.03$  vs.  $M = -0.03$ ,  $SD = 0.03$ , respectively) remained practically the same,  $F(1, 28) = 9.62$ ,  $p = .004$ ; other  $F$ s  $< 1$ .

Table 1 lists all brain regions showing a significant linear trend from the regression analysis using both consensus ratings and idiosyncratic perceptions as predictors of neural activity. In addition to the observed significant linear trend in the amygdala, consensus ratings predicted a linear trend in a number of other regions in the occipital visual cortex, the temporal cortex, and parahippocampal areas. In all cases, the neural response to faces increased as their trustworthiness decreased. In contrast, the idiosyncratic judgments predicted activity only in the frontal cortex. There were no clusters that met our criteria for significance in the amygdala for the idiosyncratic regressor. However, at  $p = .01$ , this regressor yielded a  $7 \text{ mm}^3$  cluster in the left amygdala. At this threshold, the idiosyncratic judgments also predicted activity in the left temporal pole ( $219 \text{ mm}^3$  cluster).

## DISCUSSION

Important social decisions can be made without deliberation, without intention to make the decision, and in contexts not requiring the decision (Hassin, Uleman, & Bargh, 2005; Bargh & Chartrand, 1999). Judgments of trustworthiness are formed after as little as 100 msec exposure to a novel face (Willis & Todorov, 2006), characterizing these judgments as fast, unreflective, effortless “system 1” processes in contrast to slow, deliberate, effortful “system 2” processes (Willis & Todorov, 2006; Todorov et al., 2005; Kahneman, 2003). Consistent with this research and replicating previous work (Winston et al., 2002), our findings suggest that the amygdala automatically tracks the trustworthiness of novel faces. The amygdala response to faces increased as their trustworthiness decreased despite the fact that all faces displayed neutral expressions and participants engaged in a memory task that did not require person evaluation. These findings are consistent with the notion that the amygdala is involved in rapid assessments of trustworthiness that serve to bias attention (e.g., Vuilleumier, 2005; Anderson & Phelps, 2001) or approach/avoidance behavior.

The amygdala is particularly sensitive to visual facial information (Adolphs & Tranel, 2003). Bilateral amygdala damage patients show impairments in discrimination of facial expressions of emotions but not in discrimination of verbal descriptions of emotions (Anderson & Phelps, 2000; Adolphs, Tranel, Damasio, & Damasio, 1995) and, in fact, facial information can interfere with their judg-

ments of the meaning of social interactions in complex scenes (Adolphs & Tranel, 2003). Similarly, although the amygdala is activated by the trustworthiness of faces as shown here, it is not activated by descriptions of persons as “immoral” or untrustworthy (Delgado, Frank, & Phelps, 2005). However, in the case of trait judgments from faces, it is not clear to what face properties the amygdala is responding. In contrast to facial expressions of emotions that can be characterized by specific configurations of facial features (Ekman, 1982), trait judgments are not well characterized and there is large individual variation. Such judgments comprise both consensus and idiosyncratic components (Honekopp, 2006), and the amygdala response can, therefore, be due to either an individual’s idiosyncratic perception, face properties that are perceived as untrustworthy across individuals, or a combination of the two. We showed that the amygdala response is better predicted by consensus ratings of trustworthiness averaged across many individuals than by the participants’ own judgments of trustworthiness. The analysis, in which consensus ratings and idiosyncratic judgments were used to predict amygdala activity, showed that individual judgments of trustworthiness account for little residual variance in the amygdala after controlling for the shared variance with consensus ratings.

Clearly, consensus ratings are a better predictor of amygdala activity than individual judgments (Figure 4). One possible explanation for this result is that individual judgments consist of meaningful variance shared with other raters and measurement error. However, this explanation is inconsistent with prior findings (Honekopp, 2006) and the findings of our second behavioral study. In this study, we showed that idiosyncratic face perceptions contribute meaningful variance to individual judgments in addition to variance shared with other raters. At the same time, these idiosyncratic contributions were smaller than the consensus contributions. The latter finding suggests that increasing statistical power would facilitate the identification of neural correlates of idiosyncratic perceptions of trustworthiness. Future studies would need larger samples and multiple trustworthiness judgments for each participant. Collecting multiple judgments of the same faces from fMRI participants would lead to more reliable estimates of their idiosyncratic perceptions. For example, in the current study, the measures of idiosyncratic perceptions also contained error variance. Future investigations can estimate this error by asking participants to provide repeated judgments of the faces.

Consensus in judgments reflects properties of the stimulus, rather than idiosyncratic perception of the judge, and these findings suggest that the amygdala automatically codes face properties that signal untrustworthiness across individuals in a process akin to the amygdala response to emotional faces (Adolphs, 2002; Morris et al., 1996). Ultimately, it should be possible to specify the physical characteristics of the face that



trigger perceptions of untrustworthiness. However, in the absence of formal representational models of social attributes of faces specifying how faces should be measured, this would be a difficult task. In fact, research on facial attractiveness suggests that behavioral ratings of face averageness and face symmetry, attributes that are much more easily specified than trait judgments, systematically covary with attractiveness and show higher validity in predicting attractiveness judgments than physical measures of face averageness and symmetry (Rhodes, 2006). Given that consensus estimates are robust with respect to the number of judges, as shown in our first behavioral study, behavioral ratings of different trait attributes can be used to search for neural correlates of these attributes even though we currently lack sufficient understanding of the physical properties of the face that give rise to particular trait inferences.

Our findings are consistent with the notion that the amygdala is involved in the initial automatic assessments of trustworthiness based on facial features, but that other neural systems, which are more closely linked to idiosyncratic perception, modulate the amygdala's influence on behavior. The feasibility of such a role for the amygdala is supported by evidence that it is engaged in processing of fearful faces when these faces are unattended (Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004; Williams, Morris, McGlone, Abbott, & Mattingley, 2004; Vuilleumier, Armony, Driver, & Dolan, 2001; but see Pessoa, McKenna, Gutierrez, & Underleider, 2002), when the faces are presented at rates which prevent explicit individual judgments (Whalen et al., 1998, 2004; but see Pessoa, Japee, Sturman, & Underleider, 2006), and when the faces are suppressed from conscious perception in a binocular rivalry procedure (Pasley, Mayes, & Schultz, 2004; Williams et al., 2004).

This leads to the interesting possibility that the amygdala response to untrustworthy faces can be dissociated from individual judgments of trustworthiness and, perhaps, overt behavior. This possibility is consistent with research in social psychology showing that implicit and explicit social judgments can be dissociated (Greenwald & Banaji, 1995). For example, implicit prejudice responses (not consciously endorsed by the person) correlate with differences between the amygdala responses to Black and White faces (Phelps et al., 2000), and attempts to control implicit responses correlate with activity in frontal regions (Cunningham et al., 2004). At the same time, there are differences between trait judgments from Caucasian faces as in the present research and responses to faces of members of stereotyped outgroups. Whereas in the latter case there can be a clear cognitive conflict between implicit and explicit responses, in the former case there are no reasons for such a conflict to be present.

An important goal for future research is to identify the network of regions implicated in idiosyncratic judg-

ments of trustworthiness. For example, in our study, these judgments predicted activity in the lateral frontal cortex (Table 1). At a lower significance threshold of  $p = .01$ , these judgments also predicted activity in the left temporal pole, a region that is involved in face recognition and retrieval of biographical information (Gallagher & Frith, 2003; Nakamura & Kubota, 1996). To the extent that idiosyncratic perceptions of trustworthiness originate in memories of familiar faces and that knowledge of familiar people is automatically retrieved in face perception (Gobbini & Haxby, 2007; Gobbini, Leibenluft, Santiago, & Haxby, 2004), it is possible that the trustworthiness of novel faces is computed as a function of their similarity to familiar faces. One possibility is that the initial response to faces is determined by face properties commonly perceived to signal untrustworthiness and then modified by idiosyncratic perceptions determined by retrieval of specific face memories. Interestingly, in a recent article, Sergerie, Lepage, and Armony (2006) found that the right amygdala was involved in encoding of faces, whereas the left amygdala was involved in the retrieval of faces. We found that the effect of consensus ratings was stronger in the right than in the left amygdala, and that individual judgments predicted activity only in the left amygdala.

### Methodological Implications

One of the defining features of social judgments is their subjective nature (Ross & Nisbett, 1991). Yet, people often agree in their judgments and it may be easier to model the commonalities in judgments than their idiosyncratic character. To draw an analogy with research on emotions, most progress has been achieved by focusing on specific processes rather than on the conscious experience of emotions (LeDoux, 2000). This article suggests a general methodological approach for the study of neural correlates of automatic social judgments. We showed that behavioral judgments aggregated across participants can predict individual neural responses to novel faces. This finding suggests that neural responses to implicitly inferred social attributes in one group of participants can be modeled as a function of the judgments of another group of participants. Faces can be characterized on multiple social dimensions (Willis & Todorov, 2006; e.g., trustworthiness, attractiveness, aggressiveness) and large behavioral studies can be used to map the social space of the face. For example, most trait judgments are correlated and behavioral studies, using multiple judges, multiple traits, and factor analytic techniques, can identify orthogonal dimensions of social perception of faces. Then, the empirically derived dimensions can be used to search for corresponding neural correlates in the brain.

This approach can also address the question about the specificity of the amygdala response to untrustworthy-looking faces. It may be the case that the amygdala is

responding to the general valence of the face. For example, Kim, Somerville, Johnstone, Alexander, and Whalen (2003) found that surprised faces that were perceived as negative (assessed by subjective ratings after the fMRI session) evoked stronger activity in the right amygdala than surprised faces that were perceived as positive. Addressing the question about the specificity of the amygdala response would require a comprehensive approach in which faces are rated on multiple dimensions, including traits and emotions. It may turn out that one of the fundamental factors of face perception is evaluation (cf. Osgood, Suci, & Tennenbaum, 1957), and that this evaluation is tracked by the amygdala. The present findings suggest that such an approach is possible.

Although existing neural models of face perception (Haxby, Hoffman, & Gobbini, 2000) have been extended to incorporate the role of prior person knowledge in face perception (Gobbini & Haxby, 2007; Todorov, Gobbini, Evans, & Haxby, 2007), they have not incorporated the role of trait inferences. These inferences fall in between the two major functional paths of face processing: identity recognition (tracking persons over time) and emotion detection (social communication). Trait inferences share with identity recognition the fact that they are about the person, and share with emotion detection the fact that they are affectively based. Characterizing these inferences is essential for building comprehensive models of person perception and social cognition and, ultimately, understanding the social brain.

## Acknowledgments

We thank Manish Pakrashi, Valerie Loehr, and Sean Baron for their help with this project. This research was supported by NSF Grant BCS 0446846.

Reprint requests should be sent to Alexander Todorov, Department of Psychology, Princeton University, Princeton, NJ 08540, or via e-mail: atodorov@princeton.edu.

## REFERENCES

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*, 169–177.

Adolphs, R., & Tranel, D. (2003). Amygdala damage impairs emotion recognition from scenes only when they contain facial expressions. *Neuropsychologia*, *41*, 1281–1289.

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. R. (1995). Fear and the human amygdala. *Journal of Neuroscience*, *15*, 5879–5891.

Amaral, D. G. (2002). The primate amygdala and the neurobiology of social behavior: Implications for understanding social anxiety. *Biological Psychiatry*, *51*, 11–17.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, *69*, 518–529.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256–274.

Anderson, A. K., & Phelps, E. A. (2000). Expression without recognition: Contributions of the human amygdala to emotional communication. *Psychological Science*, *11*, 106–111.

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, *411*, 305–309.

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

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *41*, 521–528.

Blair, R. J. R., Morris, J. S., Frith, C. D., Perrett, D. I., & Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain*, *122*, 883–893.

Calder, A. J., Lawrence, A. D., & Young, A. W. (2001). Neuropsychology of fear and loathing. *Nature Reviews: Neuroscience*, *2*, 352–363.

Carlston, D. E., & Skowronski, J. J. (1994). Saving in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*, 840–856.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: II. On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*, 420–436.

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

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of Black and White faces. *Psychological Science*, *15*, 806–813.

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

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*, 1611–1618.

Ekman, P. (1982). *Emotion in the human face* (2nd ed.). Cambridge: Cambridge University Press.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of “theory of mind.” *Trends in Cognitive Sciences*, *7*, 77–83.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*, 32–41.

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage*, *22*, 1628–1635.

Gollwitzer, P. M., & Bargh, J. A. (2005). Automaticity in goal pursuit. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 624–646). New York: Guilford Press.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Hamermesh, D., & Biddle, J. (1994). Beauty and the labor market. *American Economic Review*, *84*, 1174–1194.

Hassin, R., & Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, *78*, 837–852.

- Hassin, R., Uleman, J. S., & Bargh, J. A. (Eds.). (2005). *The new unconscious*. New York: Oxford University Press.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233.
- Hoffman, E., & Haxby, J. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*, 80–84.
- Honekopp, 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.
- Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist*, *58*, 697–720.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.
- Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *NeuroReport*, *14*, 2317–2322.
- Lane, R. D., Reiman, E. M., Ahern, G. L., Schwartz, G. E., & Davidson, R. J. (1997). Neuroanatomical correlates of happiness, sadness, and disgust. *American Journal of Psychiatry*, *154*, 926–933.
- Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*, 390–423.
- LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.
- Lundqvist, D., Flykt, A., & Ohman, A. (1998). *The Karolinska directed emotional faces* [Database of standardized facial images]. Psychology Section, Department of Clinical Neuroscience, Karolinska Hospital, S-171 76 Stockholm, Sweden.
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*, 605–610.
- McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, *27*, 1–28.
- Montepare, J. M., & Zebrowitz, L. A. (1998). Person perception comes of age: The salience and significance of age in social judgments. *Advances in Experimental Social Psychology*, *30*, 93–161.
- 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.
- Nakamura, K., & Kubota, K. (1996). The primate temporal pole: Its putative role in object recognition and memory. *Behavioural Brain Research*, *77*, 53–77.
- Osgood, C. E., Suci, G. I., & Tennenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Pasley, B. N., Mayes, L. C., & Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron*, *42*, 163–172.
- Pessoa, L., Japee, S., Sturman, D., & Underleider, L. G. (2006). Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cerebral Cortex*, *16*, 366–375.
- Pessoa, L., McKenna, M., Gutierrez, E., & Underleider, 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., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*, *48*, 175–187.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*, 729–738.
- Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrews, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*, 495–498.
- Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199–226.
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation*. New York: McGraw-Hill.
- Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*, 303–316.
- Sergerie, K., Lepage, M., & Armony, J. L. (2006). A process-specific functional dissociation of the amygdala in emotional memory. *Journal of Cognitive Neuroscience*, *18*, 1359–1367.
- Sprengelmeyer, R., Rausch, M., Eysel, U. T., & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society of London, Series B*, *265*, 1927–1931.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain* (M. Rayport, Trans.). New York: Thieme Medical Publishers.
- Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*, 163–173.
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*, 1623–1626.
- Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actor's faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1065.
- Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actor's faces. *Journal of Experimental Social Psychology*, *39*, 549–562.
- Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*, 482–493.
- Uleman, J. S., Blader, S., & Todorov, A. (2005). Implicit impressions. In R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 362–392). New York: Oxford University Press.
- 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. (2001). Effects of attention and emotion on face processing in the human brain: An event-related study. *Neuron*, *30*, 829–841.
- Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*, 1271–1278.
- Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, *7*, 177–188.

- 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., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expression modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*, 411–418.
- Williams, M. A., Morris, A. P., McGlone, F., Abbott, D. F., & Mattingley, J. B. (2004). Amygdala responses to fearful and happy facial expressions under conditions of binocular rivalry. *Journal of Neuroscience*, *24*, 2898–2904.
- 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., 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.
- Zebrowitz, L. A. (1999). *Reading faces: Window to the soul?* Boulder, CO: Westview Press.