Evaluating Faces on Trustworthiness

An Extension of Systems for Recognition of Emotions Signaling Approach/Avoidance Behaviors

Alexander Todorov

Department of Psychology and Center for the Study of Brain, Mind and Behavior, Princeton University, Princeton, New Jersey, USA

People routinely make various trait judgments from facial appearance, and such judgments affect important social outcomes. These judgments are highly correlated with each other, reflecting the fact that valence evaluation permeates trait judgments from faces. Trustworthiness judgments best approximate this evaluation, consistent with evidence about the involvement of the amygdala in the implicit evaluation of face trustworthiness. Based on computer modeling and behavioral experiments, I argue that face evaluation is an extension of functionally adaptive systems for understanding the communicative meaning of emotional expressions. Specifically, in the absence of diagnostic emotional cues, trustworthiness judgments are an attempt to infer behavioral intentions signaling approach/avoidance behaviors. Correspondingly, these judgments are derived from facial features that resemble emotional expressions signaling such behaviors: happiness and anger for the positive and negative ends of the trustworthiness continuum, respectively. The emotion overgeneralization hypothesis can explain highly efficient but not necessarily accurate trait judgments from faces, a pattern that appears puzzling from an evolutionary point of view and also generates novel predictions about brain responses to faces. Specifically, this hypothesis predicts a nonlinear response in the amygdala to face trustworthiness, confirmed in functional magnetic resonance imaging (fMRI) studies, and dissociations between processing of facial identity and face evaluation, confirmed in studies with developmental prosopagnosics. I conclude with some methodological implications for the study of face evaluation, focusing on the advantages of formally modeling representation of faces on social dimensions.

Key words: **social cognition; face perception; trustworthiness**

Introduction

References to the belief that the nature of the mind and human personality can be inferred from facial appearance can be dated back to ancient Greece, Persia, Rome, and China (McNeill 1998). At the end of 18th century, Johann Kaspar Lavater, a Swiss pastor, wrote "Essays on physiognomy, designed to promote the knowledge and love of mankind," in which he described how to read a person's character from their face. The book caused a craze in Europe. It underwent more than 150 editions until 1940. As Darwin noted in his autobiography, he was almost denied the chance to take the historic *Beagle* voyage because of this book. The captain of the ship, a fan of Lavater, did not believe that a person with such a nose would possess "sufficient energy and determination" for the voyage (Darwin 1887/1950, p. 36). In the 19th century, the pseudo-science of physiognomy reached its apogee. Cesare Lombroso, who provided his "scientific" testimony at several trials, argued that "each type of crime is committed by men with particular physiognomic characteristics." For example, "thieves are notable for their expressive faces and manual dexterity, small wandering eyes that are often oblique in form, thick and close eyebrows, distorted or squashed noses, thin beards and hair, and sloping foreheads" (Lombroso 1876/2006, p. 51).

Although nowadays such notions strike most of us as ludicrous, there is abundant research in social psychology about the effects of facial appearance on social outcomes (e.g., Blair et al. 2004; Eberhardt et al. 2006; Hamermesh & Biddle 1994; Hassin & Trope 2000; Little et al. 2007; Langlois et al. 2000; Montepare & Zebrowitz 1998; Mueller & Mazur 1996; Zebrowitz 1999). Most of this work has been on the effects of attractiveness (e.g., Langlois et al. 2000), but specific

Address for correspondence: Alexander Todorov, Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544-1010. Voice: +609-258-7463; fax: +609-258-1113.
atodorov@princeton.edu

Ann. N.Y. Acad. Sci. 1124: 208–224 (2008).
doi: 10.1196/annals.1440.012

trait impressions also impact social outcomes. For example, inferences of competence, based solely on facial appearance, predict the outcomes of political elections (Ballew & Todorov 2007; Hall et al. in press; Todorov et al. 2005), and inferences of dominance predict military rank attainment (Mazur et al. 1984; Mueller & Mazur 1996).

In this article, I argue that evaluation of emotionally neutral faces is an extension of functionally adaptive systems for understanding the communicative meaning of emotional expressions and explore the implications of this hypothesis for brain responses to faces. I review evidence that a) trait judgments from faces, in particular judgments of trustworthiness, are highly efficient (Section I); b) trustworthiness judgments approximate the valence evaluation of faces that underlies multiple social judgments (Section II); c) the amygdala plays a key role in the automatic evaluation of faces on trustworthiness (Section III); and d) exaggerating the facial features that make a neutral face look trustworthy produces expressions of happiness, whereas exaggerating the facial features that make a face look untrustworthy produces expressions of anger (Section IV).

The emotion overgeneralization hypothesis can explain highly efficient but not necessarily accurate trait judgments from faces, a pattern that appears puzzling from an evolutionary point of view (Section V) and also generates novel predictions about brain responses to faces. Specifically, this hypothesis predicts a nonlinear response in the amygdala to face trustworthiness, a prediction confirmed in two functional magnetic resonance imaging (fMRI) studies (Section VI). Second, the hypothesis predicts that it should be possible to observe dissociations between processing of facial identity and face evaluation, a prediction confirmed in studies with developmental prosopagnosics who are unable to process facial identity but are able to make trustworthiness judgments (Section VII). I conclude with some methodological implications for the study of face evaluation, focusing on the advantages of formally modeling representation of faces on social dimensions (Section VIII).

I. The Efficiency of Trait Judgments from Faces

What is the minimal time exposure to a face sufficient for people to form a person impression? We studied five trait judgments from emotionally neutral faces: likeability, trustworthiness, competence, aggressiveness, and attractiveness (Willis & Todorov 2006). Prior studies have shown that judgments of physical attractiveness can be made after extremely brief presentations of faces (Locher et al. 1993; Olson & Marshuetz 2005). However, attractiveness is a property of facial appearance, and it is not clear whether these findings generalize to specific trait judgments such as trustworthiness and competence.

We included attractiveness judgments as a benchmark against which to compare the efficiency of other judgments. Faces were presented for 100, 500, or 1000 ms and participants were asked to make one of the five judgments. For all five judgments, judgments made after 100 ms exposure to faces closely agreed with control judgments made in the absence of time constraints. More importantly, this agreement did not improve with additional time exposure, suggesting that 100 ms exposure is sufficient for people to form a reliable person impression. Additional time exposure had the effect of increasing confidence in judgments.

The findings did not change when the analysis controlled for the shared variance of trait judgments with attractiveness. Importantly, the agreement for judgments of trustworthiness was as high as the agreement for judgments of attractiveness. The response times for these two judgments were also almost identical and faster than the response times for judgments of competence, likeability, and aggressiveness, suggesting that people are particularly efficient at making trustworthiness judgments.

In a series of subsequent studies focusing on judgments of trustworthiness (Todorov & Pakrashi, under review), we tested whether exposure times shorter than 100 ms are sufficient for these judgments. In fact, in three experiments we found that even after 33 ms exposure, judgments were better than chance in discriminating trustworthy-looking from untrustworthy-looking faces. In one of the experiments, we used 8 different presentation times ranging from subliminal presentation of faces (17 ms) to unlimited viewing time. As shown in FIGURE 1, judgments changed systematically as a function of time exposure. These judgments were almost perfectly described by a sigmoid function of time exposure, accounting for 95% of the variance. Participants were unable to discriminate between trustworthy-looking and untrustworthy-looking faces after subliminal exposure but were able to discriminate after 33 ms exposure. With the increase in exposure from 33 to 100 ms, the correlation between judgments made after limited exposure and control judgments made in the absence of time constraints increased dramatically. This correlation improved relatively little with the increase in exposure from 100 to 167 ms, and additional increases in time exposure did not improve the correlation at all.

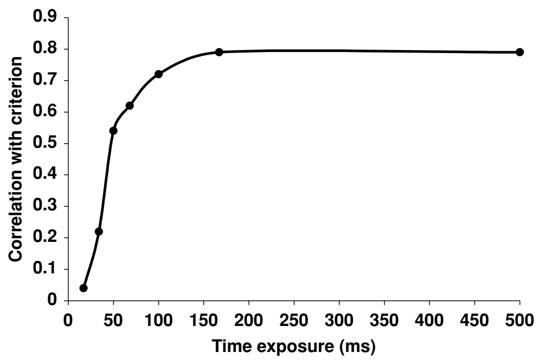

FIGURE 1. Correlation of judgments of trustworthiness made after limited time exposure to emotionally neutral faces and control judgments made in the absence of time constraints as a function of exposure time. Faces were presented for 17, 33, 50, 67, 100, 167, and 500 ms. Correlations at the level of mean trustworthiness judgment averaged across participants.

Bar and colleagues (2006) obtained similar findings for judgments of perceived threat in emotionally neutral faces. Judgments made after 39 ms (but not judgments made after 26 ms) correlated highly with judgments made after 1700 ms. As in our studies, participants were not able to make trait judgments after subliminal exposure to faces but were able to make these judgments after presentation times that are at the subjective threshold of visual awareness of faces (Pessoa et al. 2006a; Pessoa et al. 2006b).

Bar and colleagues also showed that after rapid exposures, threat judgments were predicted by judgments of low spatial frequency (LSF) face images but not by judgments of high spatial frequency (HSF) face images. This finding is interesting because it suggests that such judgments may be computed via fast coarse magnocellular pathways, guaranteeing a processing advantage for these judgments. Low frequency information may be carried through either superior colliculus/pulvinar subcortical pathways (Hamm et al. 2003; Vuilleumier 2005) or cortical pathways (Bar 2003; Pessoa 2005; Pessoa et al. 2002). In the case of emotional expressions, there is evidence that their perception depends on LSF information (Schyns & Oliva 1999), and functional neuroimaging studies suggest that the amygdala is particularly sensitive to this information in the discrimination of fearful from neutral faces (Vuilleumier et al. 2003; Winston et al. 2003). In one of our experiments, described in Section VI, we tested whether the amygdala response to face trustworthiness is modulated by spatial frequency information (Said et al., under review).

To summarize, person impressions are formed after face exposures shorter than 100 ms. Such exposure times are not sufficient for saccadic eye movements and, thus, do not allow for visual exploration of the face. In other words, trait impressions formed after rapid face presentations are "single glance" impressions.

II. What Do Judgments of Trustworthiness Measure?

As the first three panels of FIGURE 2 show, judgments of trustworthiness are highly correlated with other trait judgments. For example, for a set of standardized neutral faces (Lundqvist et al. 1998), trustworthiness judgments correlate 0.75 with judgments of attractiveness, −0.76 with judgments of aggressiveness, and 0.63 with judgments of intelligence. In fact, it is very difficult to find a trait judgment that is not correlated with trustworthiness. These high correlations suggest that trustworthiness judgments from faces may reflect the general evaluation of the face or, at least, approximate this evaluation rather well. Such valence evaluation permeates social judgments (Kim & Rosenberg 1980; Rosenberg et al. 1968; cf. Osgood et al. 1957) and is one of the organizing principles of person impressions (Wyer & Srull 1989).

A series of data-driven behavioral studies confirmed the hypothesis that trustworthiness judgments approximate the valence evaluation of faces (Oosterhof & Todorov, under review). First, we identified trait dimensions that are spontaneously used to characterize faces. Second, for each of these trait dimensions, a group of participants rated emotionally neutral faces. Third, the mean trait judgments were submitted to a principal component analysis. The first principal component accounted for 63% of the variance. All positive judgments (trustworthy, emotionally stable, responsible, sociable, caring, attractive, intelligent, and confident) had positive loadings. and all negative judgments (weird, mean, aggressive, unhappy, and dominant) had negative loadings on this component. Most important, judgments of trustworthiness had the highest loading (.94), suggesting that these judgments best approximate the valence dimension of face evaluation. This was the case even when we excluded trustworthiness judgments from the principal component analysis. The correlation between the first principal component (the evaluation factor) obtained from all trait judgments except trustworthiness and trustworthiness judgments was .94 (FIG. 2D), indicating that a single trustworthiness judgment was sufficient to summarize the evaluative information present in all other trait judgments.

Thus, it appears that in situations where no context is provided, trustworthiness judgments from faces

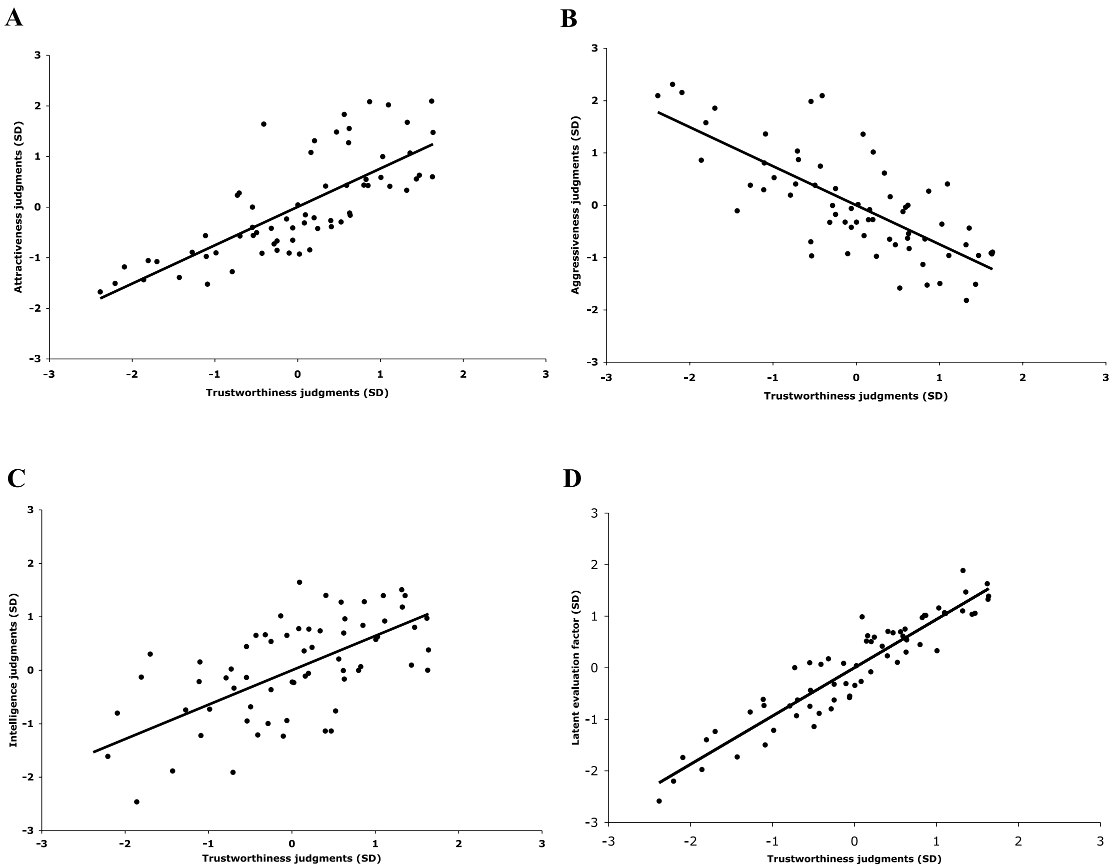

FIGURE 2. Scatter plots of judgments of trustworthiness from emotionally neutral faces and **(A)** judgments of attractiveness, **(B)** judgments of aggressiveness, **(C)** judgments of intelligence, and **(D)** a latent evaluation factor, which is a linear combination of judgments on 12 traits used to spontaneously categorize faces. Each point represents a face. The judgments are plotted in standardized scores. The line represents the best linear fit.

reflect inferences about the positivity/negativity of the face. If the valence evaluation of a stimulus is directly linked to automatic approach/avoidance responses (Chen & Bargh 1999), then judgments of trustworthiness may serve precisely this function in social interactions: determining whether to approach or avoid a stranger. In fact, Fenske and colleagues (2005) showed that momentary associations of motor approach/avoidance responses with faces affected subsequent trustworthiness judgments of these faces. Faces for which participants had to withhold a response were rated as less trustworthy than faces for which they did not have to withhold a response.

Adolphs and colleagues (1998) asked patients with bilateral amygdala–damage to make trustworthiness and approachability judgments ("how much would they want to walk up to that person and strike up a conversation") from faces. The results were identical for these two judgments, and the correlation between the mean control judgments was 0.89. Relative to three comparison groups (normal controls, brain-damaged patients with no damage in the amygdala, and patients with unilateral amygdala-damage), the bilateral amygdala damage patients judged untrustworthy-looking faces as trustworthy and "unapproachable" faces as approachable. This is consistent with experimental primate research showing that bilateral amygdala lesions lead to uninhibited approach behaviors in social interactions (Amaral 2002; Emery et al. 2001).

To summarize, the functional role of face evaluation is to prepare one's behavior in relation to the other person. When clear emotional cues broadcasting the intentions of the other person are absent, judgments of trustworthiness are tantamount to an approach/avoidance decision.

III. The Neural Underpinnings of Judgments of Trustworthiness

Three studies have implicated the amygdala in the computation of trustworthiness judgments (Adolphs et al. 1998; Engell et al. 2007; Winston et al. 2002). The amygdala is involved in multiple psychological functions (Phelps & LeDoux 2005) from learning of fear responses and consolidation of emotional memories (McGaugh 2004) to implicit evaluation of stimuli (Davis & Whalen 2001; Sander et al. 2003; Vuilleumier 2005; Whalen 1998). The findings suggesting that the amygdala plays a key role in the evaluation of face trustworthiness are consistent with the latter role.

As described above, patients with bilateral amygdala damage show a bias to perceive untrustworthy faces as trustworthy (Adolphs et al. 1998). Two subsequent functional neuroimaging studies (Engell et al. 2007; Winston et al. 2002) confirmed the involvement of the amygdala in face evaluation on trustworthiness and, further, showed that faces are spontaneously evaluated on this dimension. Winston and colleagues (2002) asked participants to make explicit trustworthiness judgments and implicit (with respect to trustworthiness) age judgments of unfamiliar faces. Independent of the task, the amygdala activity increased as the subjective *untrustworthiness* of the faces increased. The judgments of trustworthiness were collected after the imaging experiment and the brain responses were modeled as a function of these judgments.

We used a different implicit task and replicated the Winston et al. findings (Engell et al. 2007). In our experiment, participants were ostensibly engaged in a memory task. They were presented with blocks of faces and asked to indicate whether a test face was presented in the preceding block or not. Judgments of trustworthiness or person evaluation were never mentioned during the course of the experiment. Although the task did not demand face evaluation, the amygdala response to faces increased as the untrustworthiness of the faces increased (FIG. 3).

In the same study, we also tested whether the amygdala response is driven by face properties that signal untrustworthiness across individuals or by an individual's idiosyncratic perceptions of untrustworthiness. Although people agree when making trait judgments from faces, there is a large individual variation. For example, for the set of faces that we used, the average correlation between individual judgments and the mean judgments of the other individuals (consensus judgments) was 0.52. We collected trustworthiness judgments from a large sample of participants who did not participate in the fMRI study and used the mean judgments to predict the individual's amygdala response to the faces. These consensus judgments predicted the amygdala's response better than the individual's own judgments collected after the fMRI session. There was little residual variance explained by individual judgments in the amygdala's response after removing their shared variance with consensus judgments. Because consensus judgments reflect properties of the face rather than idiosyncratic perceptions of the judge (e.g., Hönekopp 2006), we argued that the amygdala response is driven in a bottom-up fashion by structural properties of the face that convey cues for untrustworthiness. This finding suggests that a more powerful approach to modeling the amygdala's response to face trustworthiness is to use formal models for representing face trustworthiness rather than rely on individual judgments, an approach that is outlined in the next section.

IV. The Origins of Face Evaluation

Secord (1958) suggested that one of the mechanisms for forming person impressions is "temporal extension"—misattributing a momentary state of the person to an enduring attribute. This hypothesis implies that person impressions may be grounded in momentary emotional expressions, and three empirical studies provide support for this idea. Knutson (1996) showed that emotional expressions were associated with trait judgments of dominance and affiliation. For example, faces expressing anger were rated as more dominant. Two subsequent studies replicated these findings (Hess et al. 2000; Montepare & Dobish 2003). Montepare and Dobish also showed that emotion ratings of neutral faces correlated with trait ratings of these faces.

Emotional expressions, among other things, broadcast behavioral intentions (e.g., Adams & Kleck 2005; Fridlund 1994). Expressions of anger communicate that the person should be avoided, and expressions of happiness communicate that the person can be approached. For example, recent studies show that angry faces trigger automatic avoidance responses (Adams et al. 2006; Marsh et al. 2005). If face evaluation reflects inferences of behavioral intentions related to one's approach/avoidance behavior, as argued in Section II, it may be derived from emotional expressions that signal approach/avoidance behaviors. Specifically, faces that are evaluated negatively may contain subtle features that resemble angry expressions and faces that are evaluated positively may contain subtle features that resemble happy expressions.

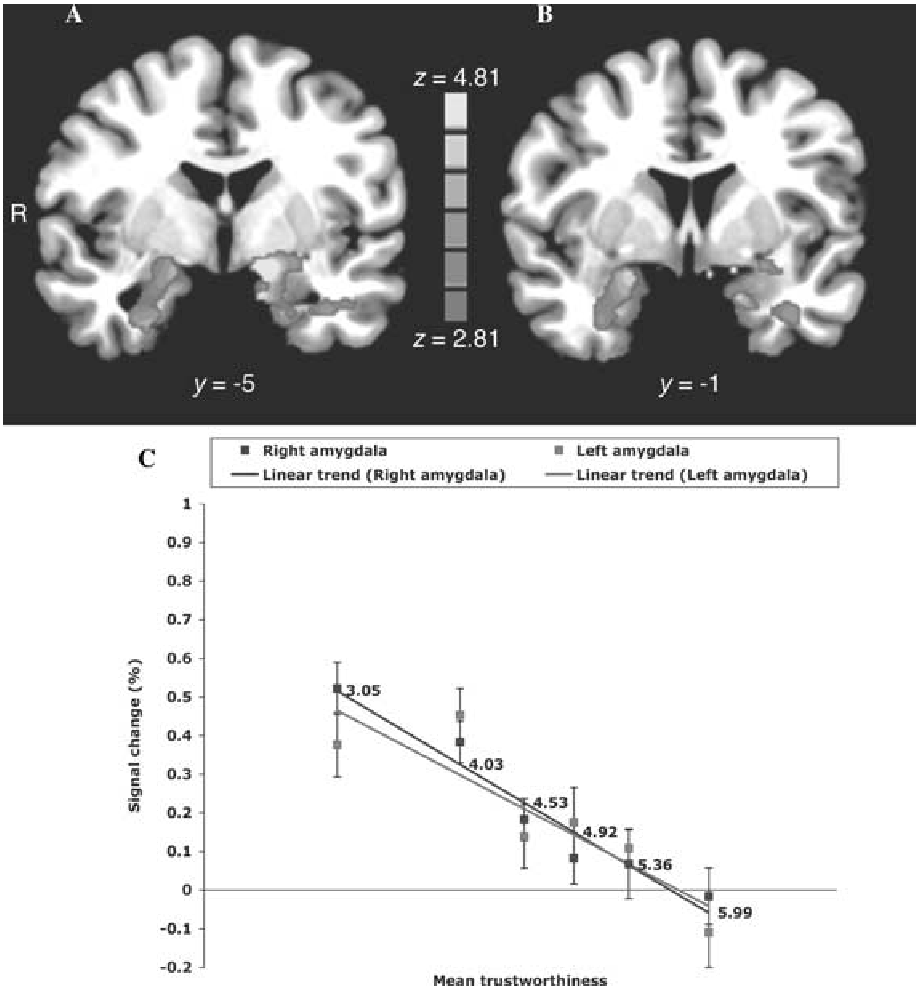

FIGURE 3. Amygdala response as a linear function of face trustworthiness. Activity is overlaid on slices of a standardized brain containing the peak response for **(A)** the left amygdala and for **(B)** the right amygdala. **(C)** Blood oxygenated level-dependent response (% signal change) in voxels in the amygdala showing a significant linear trend as a function of face trustworthiness. To extract percent signal change, faces were divided into 6 categories according to their perceived trustworthiness. For example, the mean for the least trustworthy faces was 3.05 ($SD = 0.45$) and the mean for the most trustworthy faces was 5.99 ($SD = 0.21$) on a scale ranging from 1 (untrustworthy) to 9 (trustworthy).

To test the hypothesis that face evaluation is an overextension of the mechanisms for understanding the communicative meaning of emotional expressions, we combined behavioral studies and computer modeling (Oosterhof & Todorov, under review). First, we built and validated a computer model for representing face trustworthiness. Second, using this model, we exaggerated the features that make a face look trustworthy or untrustworthy and tested how the perceptions of the face change as a function of these features.

We used a data-driven statistical model based on 3D laser scans of faces, in which faces are represented as points in a multidimensional space (Blanz & Vetter 1999; Singular Inversions 2006). We worked with 50 dimensions (50 independent principal components) representing 3D face shape. Using the model, we randomly generated 200 emotionally neutral faces with the constraint that the faces were Caucasian to avoid the influence of ethnic stereotypes on judgments. We asked participants to judge these faces on trustworthiness and used the mean judgments to find a vector (representing a weighted combination of the principal components) in the 50-dimensional face space whose direction is optimal in changing trustworthiness (FIG. 4A).

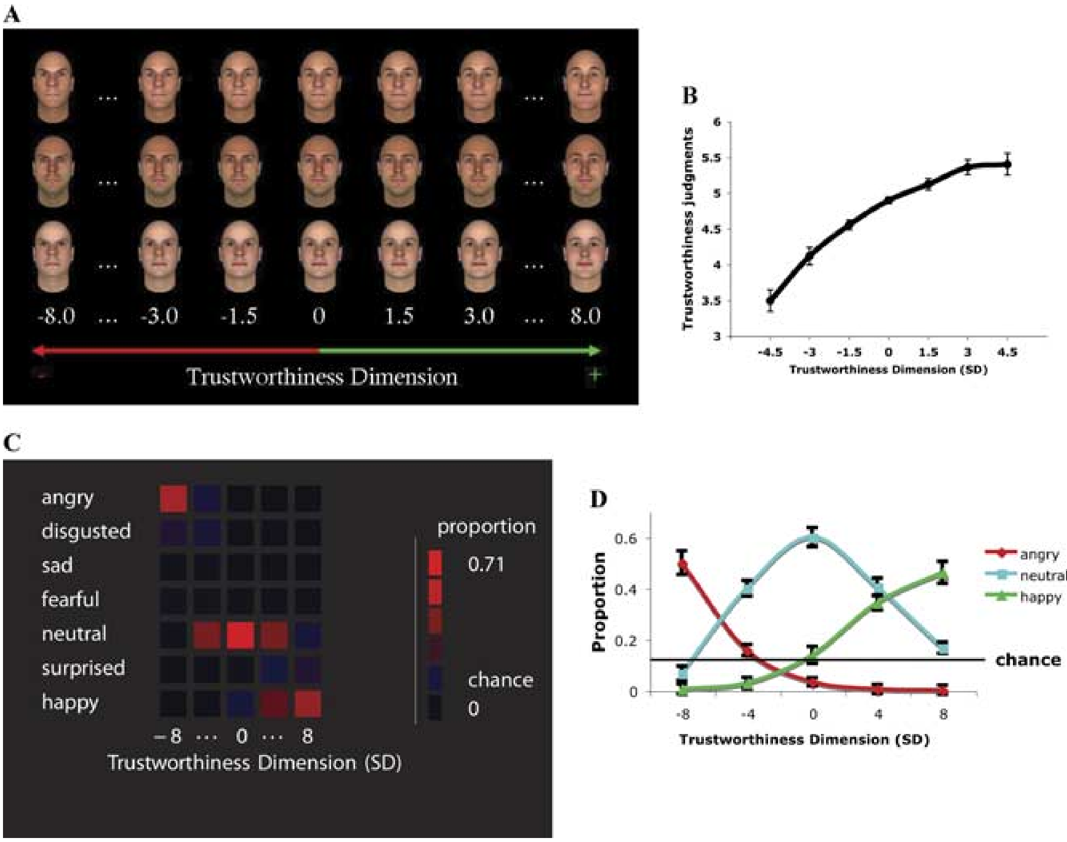

FIGURE 4. (A) Examples of faces with exaggerated trustworthiness features. The faces in the center column were randomly generated and then their features were exaggerated to decrease (left three columns) and increase (right three columns) their perceived trustworthiness. These changes were implemented in a computer model based on trustworthiness judgments of neutral faces. **(B)** Trustworthiness judgments of faces generated by the computer model. The judgments were made on a 9-point scale, ranging from 1 (not at all trustworthy) to 9 (extremely trustworthy). **(C)** Intensity color plot showing the emotion categorization of faces as a function of their trustworthiness. **(D)** Categorization of faces as angry, happy, and neutral as a function of their trustworthiness. The *x*-axis in the figures represents the extent of exaggeration of facial features in standard deviation units. Error bars show within-subjects standard error of the mean.

To validate that the model successfully manipulates face trustworthiness, we randomly generated neutral faces and produced untrustworthy and trustworthy versions for each face. Then, we asked participants to judge these faces on trustworthiness. As shown in FIGURE 4B, trustworthiness judgments of faces tracked the trustworthiness predicted by the model, although people were more sensitive to changes in trustworthiness at the low end of the spectrum than at the high end. A quadratic monotonic function accounted for 99% of the variance of the mean judgments. That is, although the physical distance between any two categories of faces was the same (1.5 SD), people were better at discriminating faces at the negative end of the trustworthiness dimension.

If face evaluation builds on the mechanisms underlying perception of emotional expressions, then exaggerating facial features that increase or decrease perceived trustworthiness should produce faces with emotional expressions (FIG. 4A). In other words, "neutral" expressions should contain subtle features that resemble emotional expressions and that people interpret to signify personality dispositions. To test this hypothesis, we presented participants with extreme trustworthy and untrustworthy versions of randomly generated neutral faces and asked them to classify the faces as neutral or as expressing one of the six basic emotions (FIG. 4C). The only responses above the chance level were for the categories of neutral, angry, and happy. The original faces (0 SD) were classified as neutral (FIG. 4D). As the facial features become more exaggerated, the neutral categorization approached chance, and this was particularly clear on the negative end of the continuum. As the facial features become more exaggerated in the negative direction (−8 SD), the faces were mostly classified as angry, whereas the trustworthy facial features become more exaggerated in the positive direction (+8 SD), the faces were mostly classified as happy.

To summarize, although the model for representing face trustworthiness was a) based on judgments of faces with neutral expressions and b) data driven without a priori assumptions about which facial parts are important for trustworthiness judgments, exaggerating the features specific for trustworthiness produced faces with emotional expressions. These findings suggest that face evaluation is an overextension of the ability to read emotional expressions. That is, trustworthiness judgments are an attempt to infer behavioral intentions and are derived from facial features that resemble emotional expressions signaling approach/avoidance behaviors. For the positive end of the trustworthiness continuum, these are expressions of happiness. For the negative end of the continuum, these are expressions of anger.

The emotion overgeneralization hypothesis can account for rapid, efficient trait judgments from faces that are not necessarily accurate, a pattern that appears puzzling from an evolutionary point of view (Section V). This hypothesis also generates novel predictions about brain responses to faces, namely the amygdala response to face trustworthiness should change as a nonlinear function (Section VI) and it should be possible to observe dissociations between processing of facial identity and face evaluation (Section VII).

V. Implications for the Accuracy of Trait Judgments from Faces

Granted the notoriety of physiognomy, is there evidence that judgments of trustworthiness from still images of faces are accurate? Berry (1990) found that judgments of honesty from faces correlated with self-reports and judgments of acquaintances. In the one study measuring the trustworthiness of actual behavior, judgments of honesty from faces accounted for 4% of the variance of behavior (Bond et al. 1994). Bond and colleagues (1994) found that participants who were rated as dishonest based on their photos were more likely to express willingness to participate in experiments that involved deceiving another participant. There are multiple possible pathways of developing behavioral patterns that confirm social expectations, and these can account for such correlations (Zebrowitz 1999). For example, according to a self-fulfilling prophecy perspective, a person with an untrustworthy appearance who is consistently treated as an untrustworthy individual may develop corresponding behavioral responses. Of course, the opposite prediction can also be made, namely that people may work hard to overcome stereotypes triggered by their appearance, and there is evidence for this self-defeating prophecy effect (Collins & Zebrowitz 1995; Zebrowitz et al. 1998a; Zebrowitz et al. 1998b).

In the one longitudinal study on the accuracy of trustworthiness judgments, Zebrowitz and colleagues (1996) failed to find correspondence between judgments of honesty from faces and clinical assessments of honesty. Additional analyses showed positive correlations for men with a stable appearance of honesty across the life span, but negative correlations for women with a stable appearance. It is interesting to note that some evolutionary theories predict negative correlations (Bond & Robinson 1988). Starting from the argument that deception confers an adaptive advantage, they argue that people with nondeceptive (trustworthy) faces can learn to be more successful liars and, correspondingly, can develop deceptive personality traits as a result of positive reinforcement of their deceptive behaviors.

From the point of view of the emotion overgeneralization hypothesis, it is not necessary to have a reliable relationship between trustworthiness judgments from faces and measures of personality. To the extent that these judgments are a measure of reading emotional cues in "neutral" faces that are misattributed to stable personality dispositions, one should not expect that they are accurate.

Finally, if one assumes that personality dispositions are combinations of emotion tendencies (e.g., Plutchik 1980), it should be possible to find positive correlations between facial appearance and personality dispositions. At the end of 19th century, Theodor Piderit, a German physician who rejected Lavater's physiognomy, argued that because emotional expressions involve exercising of facial muscles, frequent exercise of these muscles could leave its permanent trace on the face (cited in Fridlund 1994, p. 11). One hundred years later, Malatesta and colleagues (1987) found some support for this hypothesis. They showed that posed neutral expressions of a sample of elderly subjects were perceived as conveying specific emotions and that these perceptions correlated with self-reports of the frequency of experiencing the emotions. For example, people who looked angry when posing for neutral photos also reported high frequency of experiencing anger.

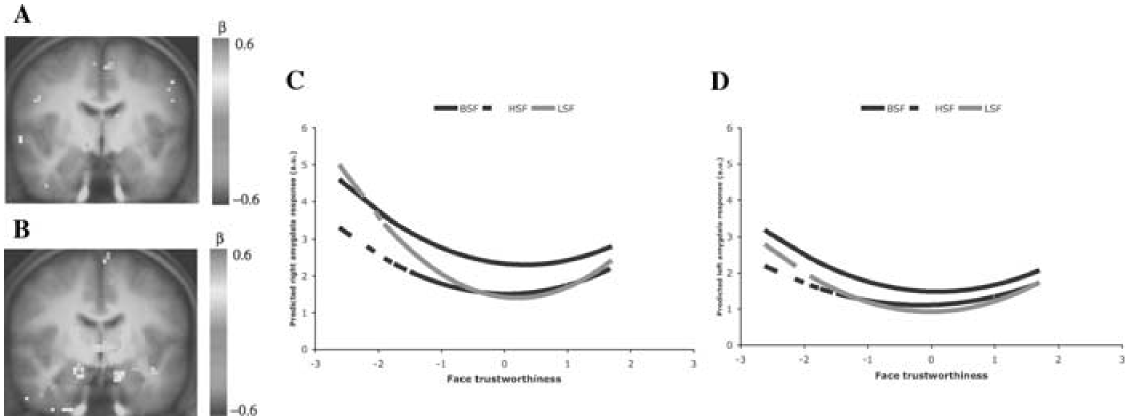

FIGURE 5. Amygdala response as a function of face trustworthiness. **(A)** Clusters in both amygdalae ($y = 6$) showing a linear response to trustworthiness. The blue shading indicates more activity as trustworthiness decreases. **(B)** Clusters ($y = 8$) showing significant quadratic effects of trustworthiness. Predicted quadratic response in clusters in the **(C)** right and **(D)** left amygdala as a function of face trustworthiness and spatial frequency. The curves were generated by extracting the mean parameter estimates of zero-order (presence of a face), linear, and quadratic effects from each spatial frequency category within the functionally defined region of interest.

VI. Implications for the Amygdala Response to Face Trustworthiness

In the first two functional neuroimaging studies on the evaluation of face trustworthiness (Engell et al. 2007; Winston et al. 2002), the amygdala response was modeled as a linear function of trustworthiness. As described above, in both studies, this response increased as the untrustworthiness of the faces increased. However, a number of functional neuroimaging studies have found a stronger amygdala response to happy than to neutral faces (e.g., Breiter et al. 1996; Pessoa, Japee, Sturman, et al. 2006; Winston et al. 2003; Yang et al. 2002). These findings, coupled with our findings that trustworthy faces have features resembling expressions of happiness, suggest that trustworthy faces can evoke a stronger amygdala response than faces in the middle of the trustworthiness dimension. That is, one should observe a nonlinear response to face trustworthiness with elevated responses to both extremely trustworthy and untrustworthy faces. Modeling the amygdala response as a linear function of face trustworthiness would miss this effect. Thus, the first implication of the emotion overgeneralization hypothesis for the amygdala's response is that this response should be a nonmonotonic function of face trustworthiness given sufficient range of trustworthiness.

In general, there is substantial evidence for the attention-grabbing power of negative information (e.g., Fiske 1980; Skowronski & Carlston 1989; Pratto & John 1991). In our own studies, as described above (FIG. 4B), we found that people are more sensitive to differences at the negative than at the positive end of the trustworthiness dimension (Oosterhof & Todorov, under review). From an evolutionary point of view, it is more important to detect untrustworthy than trustworthy individuals (cf., Cosmides & Tooby 1992), and our findings are consistent with this view. The second implication of the emotion overgeneralization hypothesis is that the amygdala's response should be more sensitive to differences at the negative than at the positive end of the trustworthiness dimension.

To test these predictions, we modeled the amygdala response as a quadratic function of face trustworthiness in two fMRI studies (Said et al. under review; Todorov et al. under review). In the first study, we found that the quadratic function provided a better fit of the amygdala response than the linear function in both left and right amygdala (FIG. 5). The amygdala response was stronger to both trustworthy and untrustworthy faces than to faces in the middle of the trustworthiness dimension. However, consistent with the previous findings of linear amygdala response to trustworthiness (Engell et al. 2007; Winston et al. 2002), the amygdala response was more sensitive to differences at the negative than at the positive end of the trustworthiness dimension (FIG. 5C & D).

In this study, we also tested whether the amygdala response was better predicted by LSF face images than by HSF images. We found that both LSF and HSF faces provided sufficient information to differentiate faces on trustworthiness and that this was the case for the behavioral and the functional neuroimaging data. Trustworthiness judgments from both the HSF and the LSF images correlated with judgments from unfiltered (broad spatial frequency faces) images. In addition, as shown in FIGURE 5C, D, the amygdala response for these images was similar, although at the negative end of the trustworthiness dimension, the slope was steeper for LSF images. One possible reason for the discrepancy with the behavioral findings of Bar et al. (2006), described in Section I, is that we used a longer presentation time in our experiment (200 ms as compared to 38 ms). In other words, it is likely that the advantage in processing of LSF images can only be seen after rapid exposure to faces. Our findings suggest that the amygdala response to face trustworthiness is robust with respect to the informational input for exposures as short as 200 ms.

In our second fMRI study (Todorov et al. under review), we used faces generated by a computer model for representing trustworthiness. The development of this model preceded the work of Oosterhof and Todorov (under review), which was described in Section IV. As it turned out, the two models converged on very similar representations of trustworthiness. We asked participants to judge computer-generated faces on trustworthiness and then regressed the mean judgments on facial features that can be controlled (e.g., moving up or down the inner part of the brow ridge) in the program for face manipulation (*FaceGen 3.1*, Singular Inversions 2006). Based on the four features that showed the highest correlation with the judgments and also had low correlations among them, we built a regression model for predicting face trustworthiness. Then we generated novel faces and manipulated the four features in both directions of increasing and decreasing face trustworthiness. The trustworthiness scores predicted by the regression model were used as regressors in the fMRI analysis.

As in Said and colleagues (under review), we found a cluster of voxels in the left amygdala that showed a non monotonic quadratic response to face trustworthiness (FIG. 6). However, this was limited to the left amygdala. A cluster of voxels in the right amygdala, which showed the same response in Engell and colleagues (2007), showed a significant linear trend. As the untrustworthiness of the faces increased, so did the right amygdala response. The findings suggest that the right amygdala may exhibit a more linear response than the left amygdala. In Engell et al., the linear trend was stronger for the right than the left amygdala, and in Said et al. (FIG. 5C & D), the slope of the response to untrustworthy faces was steeper in the right than in the left amygdala. Possible laterality differences should be addressed in future studies.

In the two studies described above, we found that the amygdala response to face trustworthiness may be better described as a quadratic rather than as a linear function of trustworthiness and that this response is more sensitive to differences at the negative than at the positive end of the trustworthiness dimension. It is interesting to note in the context of the latter finding that the relatively poor discrimination between trustworthy-looking and untrustworthy-looking faces in bilateral amygdala damage patients is due to a bias to perceive untrustworthy faces as trustworthy (Adolphs et al. 1998; see also Section VII). This also seems to be the case for people with Asperger syndrome (Adolphs et al. 2001; White et al. 2006).

The findings from patient studies and functional neuroimaging studies suggest that the amygdala is better tuned to detecting differences in the negative valence than in the positive valence of faces. This can also explain the linear trends observed in the prior studies. A linear trend can provide a reasonable fit of data that are better described by a quadratic function with a steeper slope at the negative than at the positive end of the trustworthiness dimension. This question is revisited in the section on methodological implications.

VII. Implications for Models of Face Perception

Models of face perception posit different neural pathways for processing of relatively invariant facial features essential for recognizing identity and gender and changeable facial features essential for recognizing emotions and gaze direction (Bruce & Young 1986; Haxby et al. 2000; Perrett et al. 1992; but see Calder & Young 2005 for an alternative view). Trait judgments from faces do not fit neatly in this categorization. The current findings suggest that face evaluation is subserved by the mechanisms underlying perception of emotional expressions. Thus, it should be possible to observe dissociations between processing of face evaluation and facial identity. For example, given that there are prosopagnosics who can recognize emotional expressions but not identity (Bentin et al. 2007; Damasio et al. 1990; Duchaine et al. 2003; Humphreys et al. in press; Tranel et al. 1988), there should be prosopagnosics who can make normal trustworthiness judgments.

In fact, data from four individuals with developmental prosopagnosia, a condition characterized by face recognition deficits due to a failure to develop face recognition mechanisms (Behrmann & Avidan 2005;

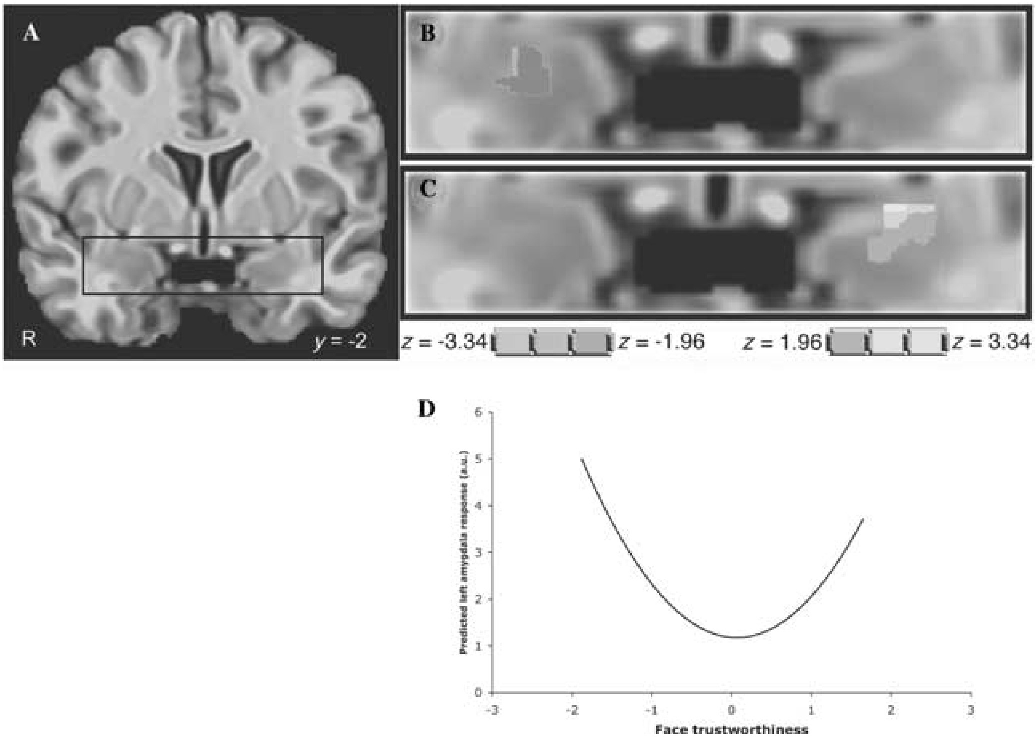

FIGURE 6. Amygdala response as a function of face trustworthiness. **(A)** Amygdala region of a standardized brain. **(B)** A cluster in the right amygdala showing a significant linear response. This cluster showed the same linear response in Engell et al. (2007). **(C)** A cluster in the left amygdala showing a significant quadratic response. **(D)** Predicted quadratic response in the left amygdala cluster as a function of face trustworthiness. The curve was generated by extracting the mean parameter estimates of zero-order (presence of a face), linear, and quadratic effects within the functionally defined region of interest.

Duchaine & Nakayama 2006a), confirmed this prediction (Todorov & Duchaine, under review). Although all four prosopagnosics had a normal performance on tests of low-level vision (Riddoch & Humpheys 1993), they had severe impairments in both long-term memory for faces and perception of facial identity. On five different measures—two measuring face memory problems and three measuring identity perception problems—each prosopagnosic individual was more than 2 standard deviations below the mean control performance (Duchaine & Nakayama 2006b; Yovel & Duchaine 2006). For example, as a group, they were 6.7 SD below the control mean on a recognition test of famous faces and 3.1 SD below the mean on a recognition test of newly learned faces (Bentin et al. 2007; Duchaine & Nakayama 2006b). On a task requiring the sorting of morphed faces in terms of similarity to a target face (Duchaine et al. in press), they were 3.3 SD below the control mean.

We asked the prosopagnosics to make trustworthiness judgments of three different sets of faces. The first set of faces was used to test patients with bilateral amygdala damage (Adolphs et al. 1998). As shown in FIGURE 7A, all four prosopagnosics showed normal performance on this set of faces. To compare their judgments with the judgments of the bilateral amygdala-damage patients studied by Adolphs and colleagues, we split the faces into the 50 most and the 50 least trustworthy faces. As described earlier, these patients show a bias to perceive untrustworthy-looking faces as trustworthy. Relative to our controls, their judgments of untrustworthy faces were 2.8 SD above the control mean. In contrast, the prosopagnosics' judgments of untrustworthy faces were within 1 SD of the control mean. For both groups, the judgments of trustworthy faces were within 1 SD of the control mean (FIG. 7B). Given that there are patients with bilateral amygdala damage who process the identity of faces normally (Adolphs et al. 1995), the results suggest a possible double dissociation between encoding identity and face evaluation. This is not surprising in light of the findings that these judgments reflect the detection of facial cues that resemble expressions of anger and happiness.

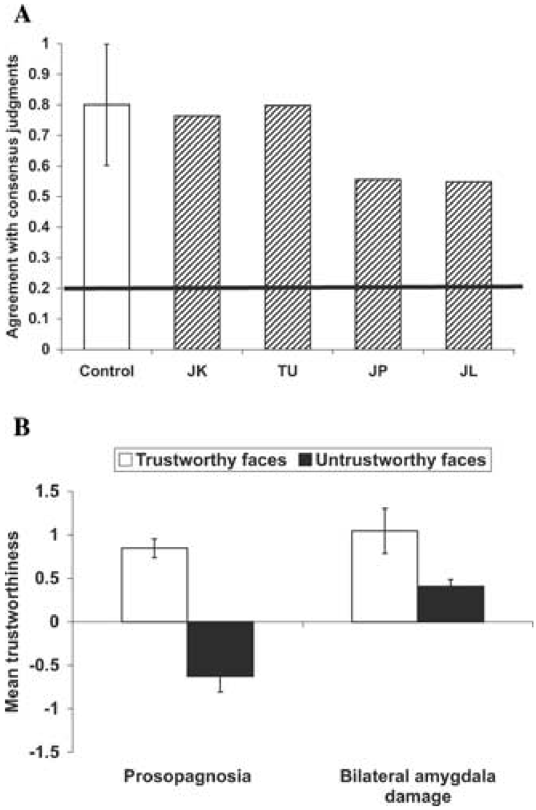

FIGURE 7. (A) Agreement between prosopagnosics' judgments and control judgments of trustworthiness of faces. The agreement is measured in Fisher z transformations of the raw correlations between individual judgments and the mean judgments of control participants. The error bar shows 1 standard deviation. Scores above the thick black line indicate significant correlations. **(B)** Mean perceived trustworthiness of faces categorized as trustworthy and untrustworthy for four patients with bilateral amygdala damage and four developmental prosopagnosics. The trustworthiness was measured on a scale from -3 (untrustworthy) to $+3$ (trustworthy). The error bars show standard error of the mean.

The faces in the first set varied on a number of dimensions including hair, expression, gaze, and age. It is possible that the prosopagnosics' normal performance on trustworthiness judgments can be accounted for by reliance on these cues. To rule out this explanation, we used two sets of standardized faces. In both sets, the faces had direct gaze, neutral expression, and a similar age. In the final set, the hair and facial blemishes of all faces were removed, and the faces were presented as grayscale images. Thus, the face sets consisted of increasingly homogeneous face images in order to force reliance on facial structure and complexion. For both sets of faces, two of the prosopagnosics showed typical trustworthiness judgments that agreed with control judgments. The judgments of the other two weakly agreed with control judgments, but even their judgments were within the normal range of control performance. The prosopagnosics with worse performance on the trustworthiness tests were not more severely impaired with the facial identity tests than the prosopagnosics with normal performance.

The findings suggest that the mechanisms underlying face evaluation are separable from the mechanisms underlying processing of facial identity. We tested developmental prosopagnosics, but these findings should be extended to individuals with acquired prosopagnosia and defined brain lesions in inferotemporal cortex.

VIII. Methodological Implications

The most important methodological implication of our findings is the value of using data-driven formal models for representing the variation of faces on social dimensions. First, by exaggerating the facial features that define the variation of a face on a specific dimension, one can discover the cues in the face that are critical for judgments on this dimension. We applied this technique to judgments of trustworthiness, but it can be applied to any other judgment. Second, from an experimental point of view, these models allow the researcher to have precise control over the facial stimuli and to generate an unlimited number of faces. This can lead to new discoveries and potential theory advancement, as well as facilitate comparisons across different studies. I illustrate the value of using models for face representation with three examples.

First, as described in Section IV (FIG. 4B), we found that participants were more sensitive to changes at the negative than at the positive end of the trustworthiness dimension, controlling for the distance between faces on the trustworthiness dimension. A computer model provides the means for objective scaling of differences on a specific dimension and, consequently, precise modeling of how judgments change as a function of this distance. Without a model for representing face trustworthiness, we would not have been able to discover that judgments are more sensitive to changes at the negative end of the trustworthiness dimension.

Second, the finding that the amygdala response was more sensitive to differences at the negative than at the positive end of the trustwothiness dimension (Said et al. under review; Todorov et al. under review) can partially explain the linear trends observed in the prior studies (Engell et al. 2007; Winston et al. 2002). However, one variable that is critical about the observed amygdala response and that can vary across studies is the range of face trustworthiness. Comparing behavioral ratings of faces across studies to test for differences in this range is not sufficient because people can shift their standard of judgment as a function of the specific set of faces (Biernat et al. 1991; Parducci 1965). In other words, two face sets may have the same behavioral ratings in two different studies but if judged within the same study may be very different. In fact, in a behavioral study conducted after the fMRI studies, we found that the faces used by Said and colleagues had a higher range of positivity than the faces used by Engell and colleagues (2007). Such variables external to the experiment may lead to different neural responses further complicating comparisons across studies. Having a formal model for representing face variations on the dimension of interest can solve this problem, because different face sets can be objectively scaled in terms of the model representation.

Third, as described in Section III, trait judgments from faces are highly correlated with each other. This makes it very difficult to disentangle the specific contributions of different judgments to neural responses. For example, Winston and colleagues (2007) recently found a nonlinear amygdala response to facial attractiveness. However, given that trustworthiness and attractiveness judgments are highly correlated (e.g., FIG. 2A), it is possible that this response was driven by the shared variance with perceptions of trustworthiness. The standard approach is to statistically control for the shared variance among different judgments, but this can reduce the statistical power of experiments and, in cases of specific trait judgments, it would be difficult to decide on an a priori basis what judgments should be controlled. The alternative is to experimentally rather than statistically unconfound contributions of different judgments. This can be easily achieved with formal models that can produce an unlimited number of faces varying on specific dimensions. Further, this would be a shift from exploratory correlational approaches (e.g., Engell et al. 2007) to theory validation approaches.

In our research, we focused on the spontaneous bottom-up valence evaluation of faces. This research should be extended to testing a) theory-driven multidimensional models of face evaluation using the same modeling tools and b) the role of top-down evaluation triggered by specific goals. For example, validated models of interpersonal perception posit two fundamental dimensions of dominance and affiliation (e.g., Wiggins et al. 1989). Similarly, models of perception of social groups posit two dimensions of competence and warmth that mapped into assessments of group status and group competition (Fiske et al. 2007).

Whether these dimensions obtained from behavioral studies have specific neural correlates is an open empirical question.

Finally, it would be important to integrate the present research with research on social categorization. Social category information such as age, gender, and race is rapidly extracted from facial information (e.g., Cloutier et al. 2005; Ito, in press; Ito et al. 2004; Ito & Urland 2003; Mason et al. 2006) and many models of person perception make a fundamental distinction between social category and individuating information (Bodenhausen & Macrae 1998; Brewer 1988; Fiske & Neuberg 1990; but see Kunda 1999). Social category information can be easily manipulated in models of face representation allowing for the identification of the relative contributions of social category and individuating (e.g., variations on trustworthiness) information to both behavioral and neural responses.

Conclusions

Evaluation permeates social judgments, and its functional role is to prepare the individual for an appropriate action. Evaluating emotionally neutral faces on trustworthiness approximates the basic valence evaluation of faces and involves the amygdala. I argue that this evaluation is grounded in the mechanisms for perception of emotional expressions. Trustworthiness judgments from faces reflect inferences of behavioral intentions that signal approach/avoidance behaviors. Specifically, these judgments are based on facial features that resemble emotional expressions—happiness and anger—that ordinarily signal approach and avoidance behaviors, respectively. This emotion overgeneralization hypothesis can account for the persistence and subjectively compelling character of trait impressions from faces, as well as for rapid, efficient judgments that are not necessarily accurate. Finally, this hypothesis makes novel predictions about brain responses to faces that were confirmed in fMRI studies and studies with prosopagnosics.

We are in the beginning of the study of the neural underpinnings of face evaluation. Although I focused on findings about the role of the amygdala in face evaluation, clearly there are multiple regions involved in this evaluation. This chapter outlines a set of tools that can be used to characterize the neural systems underlying face evaluation. We should be able to build formal representational models of any specific trait dimension (e.g., competence), decompose the face variation on this dimension to facial properties such as features resembling emotional expressions and masculine

/feminine face shape, and characterize the neural systems involved in the evaluation on this dimension.

Acknowledgments

This research was supported by National Science Foundation Grant BCS-0446846. I thank Ralph Adolphs for providing the stimuli used to test patients with bilateral amygdala damage and the raw data for my reanalysis and Sean Baron, Jason Barton, Andy Engell, Susan Fiske, Crystal Hall, Todd Heatherton, Valerie Loehr, Chris Olivola, Nick Oosterhof, Chris Said, Sara Verosky, and an anonymous reviewer for their comments on previous versions of this chapter.

Conflict of Interest

The author declares no conflicts of interest.

References

Adams, R. B., Ambady, N., Macrae, N., & Kleck, R. E. (2006). Emotional expressions forecast approach-avoidance behavior. *Motivation & Emotion*, *30*, 179–188.

Adams, R. B., & Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, *5*, 3–11.

Adolphs, R., Sears, L., & Piven, J. (2001). Abnormal processing of social information from faces in autism. *Journal of Cognitive Neuroscience*, *13*, 232–240.

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.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the USA, 104*, 17948–17953.

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.

Behrmann, M., & Avidan, G. (2005). Congenital prosopagnosia: Face blind from birth. *Trends in Cognitive Sciences*, *9*, 180–187.

Bentin, S., Degutis, J. M., D'Esposito, M., & Robertson, L. C. (2007). Too many trees to see the forest: Performance, event-related potential, and functional magnetic resonance imaging manifestations of integrative congenital prosopagnosia. *Journal of Cognitive Neuroscience*, *19*, 132–146.

Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. *Social Cognition*, *8*, 343–361.

Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*, 485–499.

Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, *15*, 674–679.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, 187–194.

Bodenhausen, G. V., & Macrae, C. N. (1998). Stereotype Activation and Inhibition. In R. S. Wyer, Jr. (Ed.). *Stereotype Activation and Inhibition* (pp. 1–52). Mahwah, NJ: Lawrence Erlbaum.

Bond, C. F., Berry, D. S., & Omar, A. (1994). The kernel of truth in judgments of deceptiveness. *Basic and Applied Social Psychology*, *15*, 523–534.

Bond, C. F., & Robinson, M. (1988). The evolution of deception. *Journal of Nonverbal Behavior*, *12*, 295–307.

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.

Brewer, M. C. (1988). A dual process model of impression formation. In R. Wyer, & T. Scrull (Eds.). *Advances in Social Cognition* (Vol. 1 pp. 1–36). Hillsdale, NJ: Erlbaum.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, *6*, 641–651.

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.

Cloutier, J., Mason, M. F., Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology*, *88*, 885–894.

Collins, M. A., & Zebrowitz, L. A. (1995). The contributions of appearance to occupational outcomes in civilian and military settings. *Journal of Applied Social Psychology*, *25*, 129–163.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 163–228). London: Oxford University Press.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience*, *13*, 89–109.

Darwin, F. (Ed.) (1950). *Charles Darwin's Autobiography*. New York: Henry Schuman.

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

Duchaine, B., & Nakayama, K. (2006a). Developmental prosopagnosia: A window to content-specific face processing. *Current Opinion in Neurobiology*, *16*, 166–173.

Duchaine, B., & Nakayama, K. (2006b). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*, 576–585.

Duchaine, B. C., Parker, H., & Nakayama, K. (2003). Normal emotion recognition in a developmental prosopagnosic. *Perception*, *32*, 827–838.

Duchaine, B., Yovel, G., & Nakayama, K. (in press). No global processing deficit in the Navon task in 14 developmental prosopagnosics. *Social*, *Cognitive*, *and Affective Neuroscience*.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*, 383–386.

Emery, N. J., Capitanio, J. P., Mason, W. A., Machado, C. J., Mendoza, S. P., & Amaral, D. G. (2001). The effects of bilateral lesions of the amygdala on dyadic social interactions in rhesus monkeys (Macaca mulatta). *Behavioral Neuroscience*, *115*, 515–544.

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

Fenske, M. J., Raymond, J. E., Kessler, K., Westoby, N., & Tipper, S. P. (2005). Attentional inhibition has social-emotional consequences for unfamiliar faces. *Psychological Science*, *16*, 753–758.

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

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

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category–based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–73.

Fridlund, A. J. (1994). *Human Facial Expression: An Evolutionary View*. San Diego, CA: Academic Press.

Hall, C. C., Goren, A., Chaiken, S., & Todorov, A. (In press). Shallow cues with deep effects: Trait judgments from faces and voting decisions. In E. Borgida, J. L. Sullivan, & C. M. Federico (Eds.), *The Political Psychology of Democratic Citizenship*. Oxford University Press.

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

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.

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

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.

Hess, U., Blairy, S., & Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *Journal of Nonverbal Behavior*, *24*, 265–283.

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

Humphreys, K., Avidan, G., & Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Experimental Brain Research, 176,* 356–373.

Ito, T. A. (In press). Perceiving social category information from faces: Using ERPs to study person perception. In A. Todorov, S. T. Fiske, & D. Prentice (Eds.), *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind*. Oxford University Press.

Ito, T. A., Thompson, E., & Cacioppo, J. T. (2004). Tracking the timecourse of social perception: The effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, *30*, 1267–1280.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*, 616–626.

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. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior*, *20*, 165–181.

Kunda, Z. (1999). Parallel processing of stereotypes and behaviors. In S. Chaiken & Y. Trope (Eds). *Dual-Process Theories in Social Psychology* (pp. 314–322). New York: Guilford Press.

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.

Little, A. C., Burriss, R. P., Jones, B. C., & Roberts, S. C. (2007). Facial appearance affects voting decisions. *Evolution and Human Behavior*, *28*, 18–27.

Locher, P., Unger, R., Sociedade, P., & Wahl, J. (1993). At first glance: Accessibility of the physical attractiveness stereotype. *Sex Roles*, *28*, 729–743.

Lombroso, C. (1876/2006). *Criminal Man*. Duke University Press.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces*. Psychology Section, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden.

Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging*, *2*, 64–69.

Marsh, A. A., Ambady, N., & Kleck, R. E. (2005). The effects of fear and anger facial expressions on approach- and avoidance-related behaviors. *Emotion*, *5*, 119–124.

Mason, M. F., Cloutier, J., & Macrae, C. N. (2006). On construing others: Category and stereotype activation from facial cues. *Social-Cognition*, *24*, 540–562.

Mazur, A., Mazur, J., & Keating, C. (1984). Military rank attainment of a West Point class: Effects of cadets' physical features. *American Journal of Sociology*, *90*, 125–150.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, *27*, 1–28.

McNeill, D. (1998). *The Face* (pp. 165–169). Boston: Little, Brown & Company.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior*, *27*, 237–254.

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.

Mueller, U., & Mazur, A. (1996). Facial dominance of West Point cadets as a predictor of later military rank. *Social Forces*, *74*, 823–850.

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion*, *5*, 498–502.

Oosterhof, N. N., & Todorov, A. (under review). *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.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*, 407–418.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *335*, 23–30.

Pessoa, L. (2005). To what extent are visual emotional stimuli processed without attention and awareness? *Current Opinion in Neurobiology*, *15*, 188–196.

Pessoa, L., Japee, S., Sturman, D., & Underleider, L. G. (2006a). Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cerebral Cortex*, *16*, 366–375.

Pessoa, L., Japee, S., & Underleider, L. G. (2006b). Visual awareness and the detection of fearful faces. *Emotion*, *5*, 243–247.

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences USA*, *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.

Plutchik, R. (1980). *Emotion: A Psychoevolutionary Synthesis*. New York: Harper & Row.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*, 380–391.

Riddoch, M. J., & Humpheys, G. W. (1993). *Birmingham Object Recognition Battery*. Hove, UK: Lawrence Erlbaum Associates.

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.

Said, C. P., Baron, S., & Todorov, A. (under review). *Nonlinear Amygdala Response to Face Trustworthiness: Contributions of High and Low Spatial Frequency Information.*

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

Secord, P. F. (1958). Facial features and inference processes in interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person Perception and Interpersonal Behavior* (pp. 300–315). Stanford, CA: Stanford University Press.

Singular Inversions (2006). *FaceGen 3.1 Ful l SDK Documentation.* http://facegen.com

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, *105*, 131–142.

Todorov, A., Pakrashi, M., Loehr, V. R., & Oosterhof, N. N. (under review). *Evaluating Faces on Trustworthiness: Automatic Assessment of Face Valence*.

Todorov, A., & Duchaine, B. (under review). *Recognizing Trustworthiness in Faces Without Recognizing Faces*.

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., & Pakrashi, M. (under review). *Trait Judgments From Faces: Rapid*, *Unreflective*, *and Robust*.

Tranel, D., Damasio, A. R., & Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology*, *38*, 690–696.

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.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, *7*, 177–188.

White, S., Hill, E., Winston, J., & Frith, U. (2006) An islet of social ability in Asperger Syndrome: Judging social attributes from faces. *Brain and Cognition*, *61*, 69–77.

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

Winston, J., O'Doherty, J., & Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage*, *20*, 84–97.

Winston, J., 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., Strange, B., O'Doherty, J., & Dolan, R. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of face. *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.

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

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

Yang, T. T., Menon, V., Eliez, S., Blasey, C., White, C. D. et al. (2002). Amygdalar activation associated with positive and negative facial expressions. *NeuroReport*, *13*, 1737–1741.

Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *Journal of Cognitive Neuroscience*, *18*, 580–593.

Zebrowitz, L. A. (1999). *Reading Faces: Window to the Soul?* Boulder, CO: Westview Press.

Zebrowitz, L. A., Andreoletti, C., Collins, M. A., Lee, S. Y., & Blumenthal, J. (1998a). Bright, bad, babyfaced boys: Appearance stereotypes do not always yield self-fulfilling prophecy effects. *Journal of Personality and Social Psychology*, *75*, 1300–1320.

Zebrowitz, L. A., Collins, M. A., & Dutta, R. (1998b). The relationship between appearance personality across the life span. *Personality and Social Psychology Bulletin*, *24*, 736–749.

Zebrowitz, L. A., Voinescu, L., & Collins, M. A. (1996). "Wide-eyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin*, *22*, 1258–1269.