Neuropsychologia 48 (2010) 3596–3605

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

The amygdala and FFA track both social and non-social face dimensions

Christopher P. Said [a], Ron Dotsch [b], Alexander Todorov [c,*]

[a] Department of Psychology and Center for Neural Science, New York University, New York, NY, United States
[b] Department of Psychology, Behavioural Science Institute, Radboud University Nijmegen, The Netherlands
[c] Department of Psychology and Princeton Neuroscience Institute, Princeton University, Green Hall, Princeton, NJ 08540, United States

ARTICLE INFO

Article history:
Received 18 May 2010
Received in revised form 5 August 2010
Accepted 9 August 2010
Available online 18 August 2010

Keywords:
Amygdala
Face perception
FFA
OFA
pSTS
Social cognition

ABSTRACT

The amygdala is thought to perform a number of social functions, and has received much attention for its role in processing social properties of faces. In particular, it has been shown to respond more to facial expressions than to neutral faces, and more to positively valenced and negatively valenced faces than faces in the middle of the continuum. However, when these findings are viewed in the context of a multidimensional face space, an important question emerges. Face space is a vector space where every face can be represented as a point in the space. The origin of the space represents the average face. In this context, positively valenced and negatively valenced faces are further away from the average face than faces in the middle of the continuum. It is therefore unclear if the amygdala response to positively valenced and negatively valenced faces is due to their social properties or to their general distance from the average face. Here, we compared the amygdala response to a set of faces that varied along two dimensions centered around the average face but differing in social content. In both the amygdala and much of the posterior face network, we observed a similar response to both dimensions, with stronger responses to the extremes of the dimensions than to faces near the average face. These findings suggest that the responses in these regions to socially relevant faces may be partially due to general distance from the average face.

1. Introduction

The amygdala has been implicated in a number of social functions (Adolphs & Spezio, 2006; Adolphs, Tranel, & Damasio, 1998; LeDoux, 2007; Morris et al., 1996). For example, patients with lesions in the amygdala have problems identifying expressions of fear (Adolphs, Gosselin, Buchanan, Tranel, Schyns, & Damasio, 2005; Adolphs, Tranel, Damasio, & Damasio, 1994), feel comfortable invading the personal space of other people in dyadic social interactions (Kennedy, Glascher, Tyszka, & Adolphs, 2009), and judge faces that appear to most people untrustworthy as trustworthy (Adolphs et al., 1998). Consistent with the human findings, monkeys with experimentally induced amygdala lesions demonstrate uninhibited social interaction (Amaral, 2003). Human functional neuroimaging studies have provided a wealth of data supporting the importance of the amygdala in social perception (Breiter et al., 1996; Canli, Sivers, Whitfield, Gotlib, & Gabrieli, 2002; Costafreda, Brammer, David, & Fu, 2008; Cunningham, Van Bavel, & Johnsen, 2008; Pessoa, Japee, Sturman, & Ungerleider, 2006; Whalen et al., 2004; Whalen, Rauch, Etcoff, Mclnerney, Lee, & Jenike, 1998; Winston, O'Doherty, & Dolan, 2003; Winston, O'Doherty, Kilner, Perrett, & Dolan, 2007).

Following human lesion studies (Adolphs et al., 1998; Todorov & Duchaine, 2008), there have been a number of functional neuroimaging studies implicating the amygdala in social judgments from faces (Engell, Haxby, & Todorov, 2007; Said, Baron, & Todorov, 2009; Todorov, Baron, & Oosterhof, 2008; Todorov & Engell, 2008; Winston, Strange, O'Doherty, & Dolan, 2002). Most of these studies used judgments of trustworthiness. For example, Engell et al. (2007) used such judgments to predict brain responses to faces in a task that did not require an explicit evaluation of the faces. Nevertheless, the amygdala response increased with decreases in the perceived trustworthiness of faces. While the initial studies primarily reported a negative linear response in the amygdala, most recent studies have found a quadratic non-monotonic response (Said, Baron, et al., 2009; Said, Haxby, & Todorov, submitted for publication; Todorov, Said, Oosterhof, & Engell, submitted for publication; Todorov, Baron, & Oosterhof, 2008). Faces that are highly untrustworthy and highly trustworthy elicit the strongest responses, while faces near the middle of the continuum elicit the weakest responses. The same response function was also observed in the inferior temporal cortex. As we show in the present study, this apparent contradiction in the literature may be due to differences in the stimulus properties of the faces used in the respective studies.

* Corresponding author.
E-mail address: atodorov@princeton.edu (A. Todorov).

0028-3932/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2010.08.009

It is experimentally useful to measure face trustworthiness, because this trait is an excellent approximation of the valence evaluation of faces (Oosterhof & Todorov, 2008; Todorov, Pakrashi, & Oosterhof, 2009). Therefore, one interpretation of the neuroimaging findings implicating the amygdala in trustworthiness evaluation is that the amygdala is specifically tracking facial properties that define face valence.

However, when these findings are viewed in the context of a multidimensional face space, some important questions emerge. Face space is a vector space where each dimension can be thought of as a physical property of faces, and every face can be represented as a point in the space (Valentine, 1991). The origin of the space represents the average face. According to the model of face trustworthiness proposed by Oosterhof and Todorov (2008), the trustworthiness of a face near the average face can be increased maximally by moving it in one direction in face space, and decreased maximally by moving it in the opposite direction. In this respect, the finding that the amygdala responds more strongly to highly trustworthy and untrustworthy faces than to faces in the middle of the continuum is confounded by the fact that highly trustworthy and highly untrustworthy faces are further away from the average face than faces near the middle of the continuum. Indeed, electrophysiology and fMRI studies have shown that the fusiform response increases with distance from the average face (Leopold, Bondar, & Giese, 2006; Loffler, Yourganov, Wilkinson, & Wilson, 2005). Therefore, it is unknown if observations about the trustworthiness dimension can be attributed specifically to facial properties that convey specific social signals, or if they are instead due to general distance from the average face, regardless of the dimension.

In this fMRI experiment, we compare the valence response profile to the response profile for a control dimension that is perceived to be less socially relevant but is matched on face distance to the valence dimension. As described in the methods section, the valence dimension was obtained from a principal components analysis (PCA) of nine different social judgments of faces (see Table S6 in Oosterhof & Todorov, 2008). The control dimension was selected from a large number of randomly generated dimensions that were orthogonal to all social dimensions. To compare the responses for the valence and control dimensions, we use both a whole brain approach and a region of interest (ROI) approach. Specifically, we targeted the amygdala, the fusiform face area (FFA), the occipital face area (OFA), and the face-selective regions of the posterior superior temporal sulcus (pSTS), as these have all been implicated in trustworthiness judgments or general face processing (Haxby, Hoffman, & Gobbini, 2000; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997). We expected to find a larger quadratic response to valence than to the control dimension in the amygdala and the FFA (Fig. 1).

There are many physical and psychophysical metrics that can be used to measure the distance between faces. It is therefore impossible to find a control dimension in which the range is matched to the valence range on all possible metrics. For metrics in which the ranges are unmatched, it is most conservative to have a smaller range for valence. This provides a stringent test of our hypothesis, as any effect driven by general distance along that metric will be stronger for the other dimension.

A series of preliminary experiments were used to measure the properties of the valence dimension and the control dimension. First, we show that there is less change in perceived trustworthiness, threat, and dominance for the control dimension than the valence dimension. Second, we show that for all the metrics we tested, the range of the control dimension used in the fMRI experiment was either matched to the range of the valence dimension, or unmatched in a conservative direction. Two of the metrics were tested experimentally. A third metric, which could be measured directly on the 3D meshes that defined the shape of our stimuli, was the average vertex displacement of the face mesh.

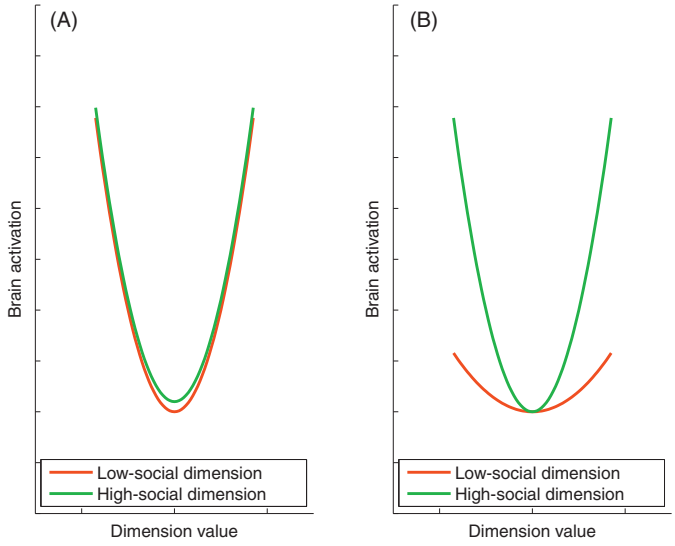

Fig. 1. Possible outcomes. (A) A similar quadratic response to both the high-social dimension (valence) and the low-social dimension (control). (B) A larger quadratic response for the high-social dimension (valence) than the low-social dimension (control).

2. Preliminary experiments

2.1. Stimuli

Faces were generated with FaceGen software and custom code. FaceGen defines face shape using a 50-dimensional face space, where each dimension is a component from a PCA performed on the 3D face vertex positions–defined on a face mesh–of a large sample of laser-scanned human faces. Because the top 50 components account for most of the shape variance, any face can be reasonably approximated as a point in this space. Only face shape was manipulated; the reflectance properties were held fixed. Oosterhof and Todorov (2008) obtained trait ratings for a large number of faces sampled from this space. A separate PCA performed on these trait ratings revealed that more than 54% of the variance could be explained by the first principal component, which can be referred to as valence (see Table S6 in Oosterhof & Todorov, 2008). Trustworthiness judgments were highly correlated with this component (>.90) even when the component was estimated without trustworthiness judgments in the PCA. Next, to build dimensions corresponding to social judgments, Oosterhof and Todorov performed a multiple regression with judgments as the dependent variable and the 50 face shape dimensions as the predictor variables. The same approach was used for building a model of face valence. Specifically, the first principal component derived from the PCA of social judgments was regressed on the shape dimensions. The valence dimension was then defined as the vector of coefficients from this regression. This vector can be added to any face in order to change its predicted value on valence. Under the assumptions of the linear model, a unit change along this dimension is expected to result in a maximal change in the valence of the face.

The control dimension was chosen as a dimension in a face space orthogonal to the valence dimension. First, we randomly generated 100 face dimensions that were orthogonal to valence and 9 other social face dimensions (e.g., threat, competence, extraversion, etc.)

Fig. 2. The faces used in the fMRI experiment. Top row: faces at values of −3, −1, 1, and 3 SDs away from the average face along the valence dimension. Bottom row: faces at values of −5, −1.67, 1.67, and 5 SDs away from the average face along the control dimension.

that were created using the methods described in Oosterhof and Todorov (2008). Second, we visually inspected the variations along these dimensions and identified the dimension that seemed to produce the least amount of change in the subjective valence of the faces. Third, we conducted a series of tests with this dimension and the valence dimension to validate that these dimensions are differentially sensitive to social judgments and yet comparable on distance to the average face.

Based on a series of pilot experiments, the stimuli in the fMRI experiment were positioned at −3, −1, 1, and 3 SDs away from the average face along the valence dimension and −5, −1.67, 1.67, and 5 SDs away from the average face along the control dimension (Fig. 2). We found that the range used for the control dimension involved an average 3D vertex displacement that was higher than the average vertex displacement along the valence dimension range (.089 compared to .086). This small difference would, if anything, work against our hypothesis.

2.2. Subjective ratings of social personality traits in both dimensions

In the first preliminary study, we examined how subjective ratings of social personality traits changed along both dimensions, using the same faces presented in the fMRI experiment. The three traits we tested were trustworthiness, dominance, and threat, as these are thought to explain much of the variance in face evaluation (Oosterhof & Todorov, 2008). Fifteen subjects (11 female; average age = 19.3; SD = 1.1) were each shown the 4 faces that varied on the valence dimension and the 4 faces that varied on the control dimension. The experiment was divided into three blocks in which subjects rated either the trustworthiness, dominance, or threat of the faces. The order of the faces was randomized across subjects. On each block, subjects saw a random sequence of 24 face presentations (3 presentations for each of the 8 faces). Each face was separated by a 1 s fixation cross. Subjects rated each face on a 9-point scale, where 1 was marked as "extremely low" on the trait, and 9 was marked as "extremely high" on the trait. The results, which are plotted in Fig. 3(A)–(C), show that subjective trustworthiness, dominance, and threat changed more with the valence dimension than with the control dimension ($p < .05$ for all traits).

2.3. Subjective ratings of stimulus features in both dimensions

In the second preliminary study, we examined how subjective ratings of relatively non-social stimulus features changed along both dimensions. Ten subjects (5 female; average age = 23.8; SD = 7.2) participated in this experiment. The procedure was the same as in the first experiment, except that subjects rated the visual impact (e.g., "whether you feel the content of the image created an instant sense of impact on you personally"), interestingness, and unusualness of the faces (Ewbank, Barnard, Croucher, Ramponi, & Calder, 2009). We found that there were no significant differences between the valence dimension and the control dimension in the slopes of these ratings (Fig. 3(D)–(F)). When we fit a second order polynomial to the ratings, we found that there was a significantly more positive quadratic component in the impact ratings for the valence dimension compared to the control dimension. There were no differences in the quadratic components for the other ratings.

As can be seen in Fig. 3, there was more change in the valence dimension than in the control dimension for social personality judgments that varied linearly with the dimensions (Fig. 3(A)–(C)). The dimensions differed to a much smaller extent for stimulus feature judgments that typically increased with distance from the average face (Fig. 3(D)–(F)). This dissociation is consistent with our selection of a control dimension that contains less social information than the valence dimension, while still being matched on distance from the average face.

2.4. Perceived identity change

In the third preliminary study, we measured how much perceived identity varies along the two dimensions. If perceived identity changes more along one dimension than another, then this dimension will have a larger psychological range along this metric. Twenty-three subjects (14 female; average age = 19.3; SD = 1.0) participated in this experiment. On each trial, subjects were shown the average face next to a face sampled from one of 40 positions along either of the dimensions in face space. The positions were drawn from equally spaced intervals along the ranges for each dimension used in the fMRI experiment. Each face was separated

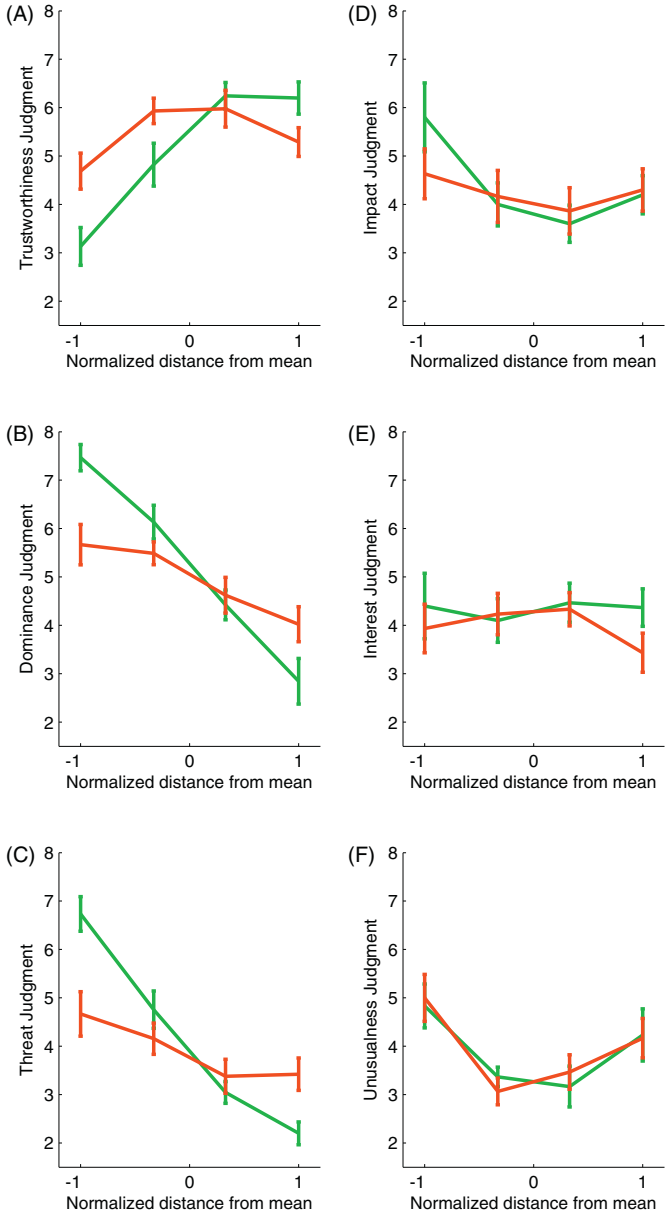

Fig. 3. Results from two behavioral experiments comparing the high-social dimension (valence) to the low-social dimension (control) along the ranges used in the fMRI experiment. The normalized *x*-axis represents an unnormalized range of [−3 3] SDs along the high-social dimension and an unnormalized range of [−5 5] SDs along the low-social dimension. The *y*-axis represents average subjective judgments ratings on a 9-point scale. Error bars represent standard error of the mean. A through C Subjective ratings on three social personality traits. D through F Subjective ratings on three non-social, stimulus features.

by a 1 s fixation cross, and the order of the faces was randomized for each subject. Subjects were asked to indicate whether the two faces could be from the same identity or different identities. Fig. 4 plots the fraction of time each face was rated as the same identity. As faces move toward the negative values of the two dimensions, the amount of perceived change in identity is matched for both dimensions. As faces move toward the positive values of the two dimensions, there is less change in perceived identity for the valence dimension than for the control dimension. Since any fMRI effect driven by identity-based distance from the average face will be stronger for the control dimension, this difference works against our hypothesis.

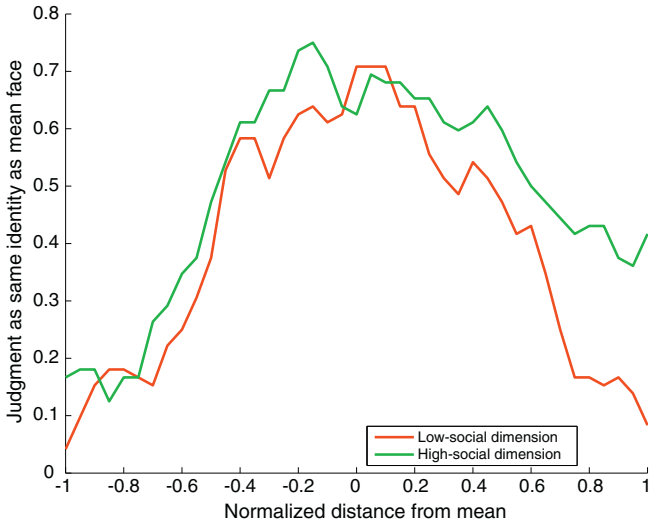

Fig. 4. Results from a behavioral experiment ($n = 23$) in which subjects were asked to choose whether each face was the same identity or a different identity from the average face (presented adjacently). The normalized *x*-axis represents an unnormalized range of [−3 3] SDs along the high-social dimension (valence) and an unnormalized range of [−5 5] SDs along the low-social dimension (control). Along the *y*-axis, 'same' is coded as 1 and 'different' is coded as 0. Results are smoothed with a 3-point moving average.

2.5. Difference threshold count

It is possible to count the number of DLs (from the German *differenze limen*, or difference threshold, sometimes referred to as just-noticeable-differences) along a range. This approach differs from the identity change approach in that the range calculation relies on the discrimination ability of the participants rather than their subjective decisions. The DL is defined as the distance in face space at which subjects are 75% accurate at discriminating between two faces. The DL is different at different positions in face space. Early theorists believed that the DL increased logarithmically with the magnitude of the stimulus. However, it is unclear if this law applies to the distance from the center of face space. Therefore, we estimated the DL at each of several positions along each dimension. A function was then fit to the DL estimates and the number of DLs was then counted along the range for each stimulus dimension (Gescheider, 1997).

Thirty subjects (21 female; average age = 19.6, SD = 1.3) participated in the experiment. The experiment consisted of two blocks, one for the valence dimension and one for the control dimension. The order of the blocks was randomized across subjects. Each block was associated with a "comparison face". The comparison face was at the extreme end (+8 units) of the dimension for that block. On each trial, participants saw two faces next to each other for 2000 ms and were then asked to indicate which face looked more similar to the comparison face. Faces were centered around seven different base positions (−6, −4, −2, 0, 2, 4, 6 SDs), with the goal of estimating the DL at each base position. For the purposes of exploration, this range exceeded the range that was ultimately used in the fMRI experiment (−3 to 3 SDs for the valence dimension, and −5 to 5 SDs for the control dimension). For each base position, eight difference values Δ were used (0.2, 0.6, 1.0, 1.4, 1.8, 2.2, 2.6, 3.0 SDs). The difference value Δ refers to the distance around the base value. Thus, for base = 6 and Δ = 0.2, the subject was asked whether the face at 5.9 or the face at 6.1 looked closer to the comparison face. In this case, the correct response was the face at 6.1, as it is more similar to the comparison face (which was at 8 SDs). Each base and Δ combination was tested twice, once with the more positive face on the left and once with the more positive face on the right.

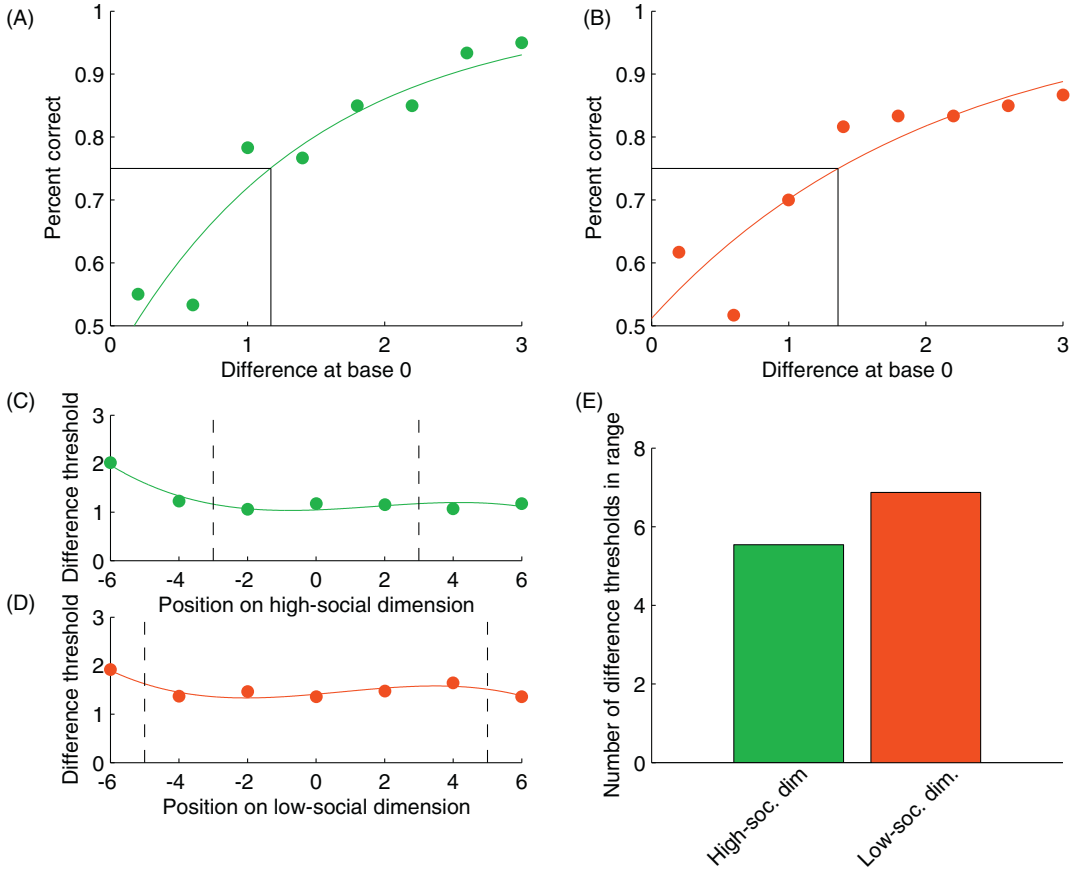

Fig. 5. (A). Example of percent correct scores for face pairs separated by distance Δ around base = 0 for the high-social dimension (valence). An exponential function is fit using least squares, and the difference threshold is the Δ value at which the percent correct is estimated to be 75%. (B) Example for the low-social dimension (control). (C) Plot of DLs for each base along the high-social dimension. The dotted lines indicate the range of faces used in the fMRI experiment. (D) Same as C, but for the low-social dimension. (E) Total number of DLs along the ranges used in the fMRI experiment.

For each base position, the relation between percent correct and Δ can be fit with an exponential function $P = 1 - a\,e^{b\Delta + c}$. The DL is the Δ at which the percent correct is estimated to be 75%. Examples of these functions for base 0 of the valence and the control dimensions can be seen in Fig. 5(A) and (B). Next, a cubic function was fit to describe the relationship between DL and base position (Fig. 5(C) and (D)). Finally, using a step of .01 DLs, the number of DLs was counted along the range for each dimension. The range used for the valence dimension contained a total of 5.54 DLs, whereas the range for the control dimension contained a total of 6.87 DLs (Fig. 5(E)). This difference will work against our hypothesis about the fMRI results.

In summary, the preliminary experiments allowed us to obtain appropriate ranges for two dimensions to be used in the fMRI experiment. There was more change in subjective social judgments for the social, valence dimension than the control dimension, particularly for perceived trustworthiness, threat, and dominance. On a variety of metrics of face distance, the ranges were either matched or unmatched in a conservative direction. There was slightly more change in 3D vertex displacement for the control range than the valence range. There was a matched amount of perceived identity change for the two dimensions at the negative ends of the ranges, but a larger perceived identity change for the control dimension than the valence dimension at the positive ends of the ranges. Finally, there were more DLs along the control dimension range than the valence dimension range. All of these differences work against our hypothesis.

3. fMRI methods

3.1. Experiment design

Twenty-four subjects (13 female; average age = 23.3; SD = 2.5) participated in the fMRI experiment. Each subject was presented with a rapid sequence of faces in the scanner. There were a total of 8 types of faces centered around the average face: −3, −1, 1, and 3 SDs along the valence dimension and −5, −1.67, 1.67, and 5 SDs along the control dimension. The faces were presented with a type 1–index 1 sequence (Aguirre, 2007; Finney & Outhwaite, 1956). This type of sequence ensures that every face was preceded by every other face an equal number of times. Each face was presented for 1450 ms and was followed by a 50 ms fixation intertrial interval. Sixty-eight rest trials, each 3000 ms long, were distributed throughout the sequence and included in the order counterbalancing. Subjects performed a one-back task in which they pressed a button whenever a face was shown twice in a row.

To increase subject comfort, the experiment was divided into eight "runs" of 4 min and 20 s each. The number of trials corresponding to each type of face, as well as the number of rests, was equally balanced among runs 1–2, runs 3–4, runs 5–6, and runs 7–8. To allow time for the MR signal to reach steady-state equilibrium and for the subjects to orient to the task, by design, each run began with 11 dummy trials, which were excluded from the analysis. This procedure also reinstates the adaptation effect for

the first valid trial of every run and minimizes effects caused by the discontinuity between runs (Aguirre, 2007).

After the main 8 runs, subjects also participated in two runs of a face localizer task in order to identify the FFA, OFA, and face-selective pSTS areas. Each localizer run consisted of 32 alternating blocks of faces and chairs. Each block was 14 s long, and each block was preceded by a 14 s block of rest. For half the subjects, the first block was faces. For the other half, the first block was chairs. Subjects performed a one-back task.

3.2. fMRI data acquisition

The blood oxygenation level-dependent (BOLD) signal was used as a measure of neural activation (Kwong et al., 1992; Ogawa, Lee, Nayak, & Glynn, 1990). Echo planar images (EPI) were acquired with a Siemens 3.0 T Allegra scanner (Siemens, Erlangen, Germany). Functional data was obtained at a resolution of 3 mm × 3 mm × 4 mm, with TR = 2000 ms, TE = 30 ms, and flip angle = 80°. For the purposes of cross-subject spatial registration, a whole brain high-resolution T1-weighted structural scan was acquired at the end of each experiment (TR = 2500 ms, TE = 33 ms, flip angle = 8°).

3.3. fMRI data pre-processing

Image analysis was performed with AFNI (Cox, 1996). After discarding the first five functional images from each run to allow the MR signal to reach steady-state equilibrium, the remaining images were slice time-corrected and then motion corrected to the last image of the last run using a six parameter 3D motion correction algorithm. Next, the functional data was despiked using the AFNI program 3dDespike, smoothed with an 8 mm full width at half maximum Gaussian kernel, and then normalized to percent signal change from the mean.

The anatomical data was aligned to the unsmoothed functional data using the AFNI program align_epi_anat.py and then transformed to Talairach space using the function @auto_tlrc. Functional datasets were then subjected to the same spatial transformation.

3.4. fMRI whole brain analysis

To determine the effects of the valence and control dimensions across the whole brain, the general linear model was used on each voxel. The set of regressors were (1) linear effects of the valence dimension; (2) quadratic effects of the valence dimension; (3) linear effects of the control dimension; (4) quadratic effects of the control dimension; (5) linear carryover effects of the valence dimension, to account for variance caused by adaptation to the previous trial; (6) linear carryover effects of the control dimension; (7) a "stimulus on" regressor, indicating whenever a face was present; (8) a dummy regressor for the first 11 trials; (9–14) six regressors for subject head motion. Regressors 1–8 were then convolved with a canonical gamma-variate hemodynamic response function before being entered into the model.

Next, we performed group analysis on the regression coefficients extracted from the polynomial regression. The coefficients were submitted to single-sample *t*-tests to determine whether the linear and quadratic effects of the two dimensions were significant.

To determine cluster size significance, we performed Monte Carlo simulations of null hypothesis data using the AFNI program AlphaSim. In order to account for spatial correlation among voxels, each randomly generated image was smoothed by a Gaussian kernel corresponding to the smoothness of the residual image left over from the group analysis. We then used the program to determine the distribution of cluster sizes defined by a voxelwise-threshold of $p < .001$. Whole brain simulations revealed that less than 5% of null clusters exceeded 540 mm^3, which we therefore used as our cutoff for significance.

3.5. Analysis of fMRI localizer runs

The data from the two localizer runs were preprocessed in the same way as the main runs. The runs were then concatenated in time and submitted to a GLM consisting of the following regressors: (1) a boxcar regressor for faces; (2) a boxcar regressor for chairs and (3–8) six regressors for motion parameters. Regressors 1 and 2 were convolved with a canonical gamma-variate hemodynamic response function. Face-selective regions were defined by the contrast of the face parameter estimate against the chair parameter estimate. Using a voxelwise-threshold of $p < .005$, the FFA could be identified bilaterally in 23 of the 24 subjects. The remaining subject did not show an FFA even at lower thresholds. This subject was excluded from further analysis of the FFA ROI but retained for group analysis of whole brain data. The right OFA was identified in 22 subjects and the left OFA was identified in 17 subjects. Like the FFA, we used an OFA that collapsed across laterality, where possible. This procedure allowed us to identify an OFA in 23 of the subjects. For the pSTS, we used a more liberal threshold of $p < .05$, as this region typically shows a weaker and less reliable face-selective response than the FFA and OFA (Fox, Iaria, & Barton, 2009). We were able to identify the face-selective right pSTS in 23 subjects and the face-selective left pSTS in 15 subjects. After collapsing across laterality, we obtained pSTS ROIs in 23 subjects.

3.6. ROI analysis

We used a total of 4 *a priori* ROIs: The bilateral amygdala was defined anatomically, and the FFA, OFA, and pSTS were defined with a separate functional localizer. For all 4 ROIs, we extracted the mean response across voxels to each of the 8 individual faces. The responses for each voxel were defined as the face regression coefficients from a whole brain GLM containing 17 regressors: (1–8) eight regressors for each of the individual faces; (9–14) six regressors for participant head motion; (15) linear carryover effects of the valence dimension, to account for variance caused by adaptation to the previous trial; (16) linear carryover effects of the control dimension; (17) a dummy regressor for the first 11 trials. For each of the ROIs, we tested for linear and quadratic effects of each of the dimensions as follows: For each subject and each dimension, we fit a second order polynomial to the responses. We then performed a single-sample *t*-test on the linear coefficients and quadratic coefficients to test for their significance at the group level. We then used paired *t*-tests to determine if there were significant differences between the dimensions, both for the linear and the quadratic coefficients.

4. fMRI results

4.1. Whole brain analysis

Consistent with previous experiments (Said, Baron, et al., 2009; Todorov et al., submitted for publication), the bilateral amygdala and the bilateral fusiform gyri showed a quadratic response to face valence/trustworthiness with stronger responses to positively valenced and negatively valenced faces than to faces at the middle of the continuum. Many frontal areas also showed a quadratic response, although in some cases these were characterized by a negative quadratic term, indicating stronger responses to faces at the middle of the continuum. Other areas that showed a positive quadratic response included the left angular gyrus, the right medial frontal gyrus, the left inferior frontal gyrus, and the left cingulate. A full list of regions showing a quadratic response to face valence is provided in Table 1.

Table 1
Locations of clusters showing a quadratic effect of face valence. Areas with a positive quadratic term are indicated by a '+' sign in the second column. Areas with a negative quadratic term are indicated by a '−' sign. All cluster sizes are significant at $p < .05$, after correction for multiple comparisons.

Region	Orientation of quadratic function	Volume (mm^3)	x	y	z
Right insula	−	7884	34.5	19.5	5.5
Right medial frontal gyrus	−	6291	1.5	13.5	47.5
Left angular gyrus	+	5238	−43.5	−70.5	32.5
Right amygdala	+	4293	22.5	1.5	−12.5
Right medial frontal gyrus	+	3618	1.5	55.5	2.5
Left inferior frontal gyrus	+	2403	−52.5	34.5	−0.5
Left inferior frontal gyrus	+	2349	−25.5	13.5	−15.5
Left insula	−	2052	−31.5	22.5	5.5
Left cingulate	+	1647	−13.5	−43.5	35.5
Left cingulate	+	1431	−1.5	−61.5	29.5
Right posterior cingulate	+	1242	1.5	−43.5	14.5
Left fusiform gyrus	+	1080	−43.5	−43.5	−15.5
Left precuneus	+	945	−4.5	−55.5	59.5
Right caudate	−	783	7.5	4.5	11.5
Right fusiform gyrus	+	594	43.5	−46.5	−21.5

Table 2
Locations of clusters showing a quadratic effect of the control dimension. Areas with a positive quadratic term are indicated by a '+' sign in the second column. Areas with a negative quadratic term are indicated by a '−' sign. All cluster sizes are significant at $p < .05$, after correction for multiple comparisons.

Region	Orientation of quadratic function	Volume (mm^3)	x	y	z
Right inferior frontal gyrus	−	24300	55.5	13.5	2.5
Right superior medial gyrus	−	12069	4.5	13.5	44.5
Left insula	−	4671	−37.5	16.5	5.5
Left middle occipital gyrus	+	2349	−46.5	−70.5	32.5
Right inferior parietal lobule	−	1404	40.5	−49.5	41.5
Left precentral gyrus	−	1296	−37.5	−25.5	62.5
Right fusiform gyrus	+	1188	49.5	−64.5	−3.5
Right hippocampus	−	675	25.5	−19.5	−12.5
Right caudate nucleus	+	675	13.5	4.5	11.5
Left thalamus	−	594	−10.5	−19.5	11.5

Several regions showed quadratic responses to the control dimension, including the right fusiform gyrus and left occipital gyrus (Table 2). Although this analysis did not find a significant quadratic response in the amygdala, a whole brain test revealed no significant differences between the quadratic response to valence and the quadratic response to the control dimension in either the fusiform region or the amygdala. Instead, the test revealed only one significant cluster in the cingulate gyrus. This result is probably spurious and, in any case, not relevant to our hypothesis, because the difference was driven mostly by high responses to the faces at the middle of the control dimension continuum.

There were no significant linear effects of valence. There was a significant linear effect of the control dimension near the left precuneus ($x = -28.5$, $y = 46.5$, $z = 20.5$) and in the left hippocampus ($x = -34.5$, $y = -28.5$, $z = -6.5$). Both of these areas showed increasing activation as the value of the control dimension increased.

4.2. ROI analysis

It is possible that the lack of significant differences between dimensions in the fusiform and other areas may be due to intersubject differences in the location of face-selective regions. To address this issue, we tested for differences between the dimensions in three face-selective ROIs defined on an individual subject level (FFA, OFA, pSTS). Additionally, we tested for responses in the amygdala, which was defined anatomically.

We found no linear effects of the valence dimension or the control dimension in any of the ROIs. There was a quadratic effect of the valence dimension in the FFA ($t(22) = 3.9$, $p < .05$), pSTS ($t(22) = 3.2$, $p < .05$), and amygdala ($t(23) = 4.1$, $p < .05$), and a marginally significant effect in the OFA ($t(22) = 1.8$, $p = .08$) (see Fig. 6). There was a significant quadratic effect of the control dimension in the FFA ($t(22) = 2.8$, $p < .05$) and amygdala ($t(23) = 3.1$, $p < .05$), but not in the pSTS or OFA. We did not find any significant differences between the quadratic responses to the two dimensions in the FFA, OFA, or amygdala. However, there was a significant difference in the pSTS ($t(22) = 2.6$, $p < .05$).

5. Discussion

In this experiment, we attempted to test whether the fusiform and amygdala quadratic response to face valence was specific to facial properties conveying social signals, or if it instead reflected a general increase in activation to faces away from the center of face space. To test this, we showed participants a set of faces that varied on the valence dimension, and a set of faces that varied on an orthogonal dimension with low social relevance. Since there are many different metrics for face distance, it was not possible to match the range of faces on every metric. On several of the face distance metrics we tested, the ranges of the two dimensions were matched. On the other metrics, the range was greater for the control dimension, thus working against the hypothesis that the quadratic response is socially specific.

Contrary to our expectations, neither the amygdala nor the FFA showed a significant difference in the strength of their quadratic responses to these two dimensions. Instead, both dimensions revealed a quadratic response in the amygdala and the FFA, as well as a quadratic response in their perceived unusualness. In the current experimental context, where no specific social goals were activated and no explicit face evaluation was demanded, the response magnitude in these regions reflected the distance from the average face. This result is more in line with proposals that the amygdala tracks stimulus intensity (Anderson et al., 2003) than with proposals that it tracks stimulus valence. It is also broadly consistent with proposals that the amygdala is a component of a vigilance system that is preferentially invoked in ambiguous learning situations (Whalen, 1998), and that its response depends on the motivational relevance of the stimuli (Cunningham et al., 2008).

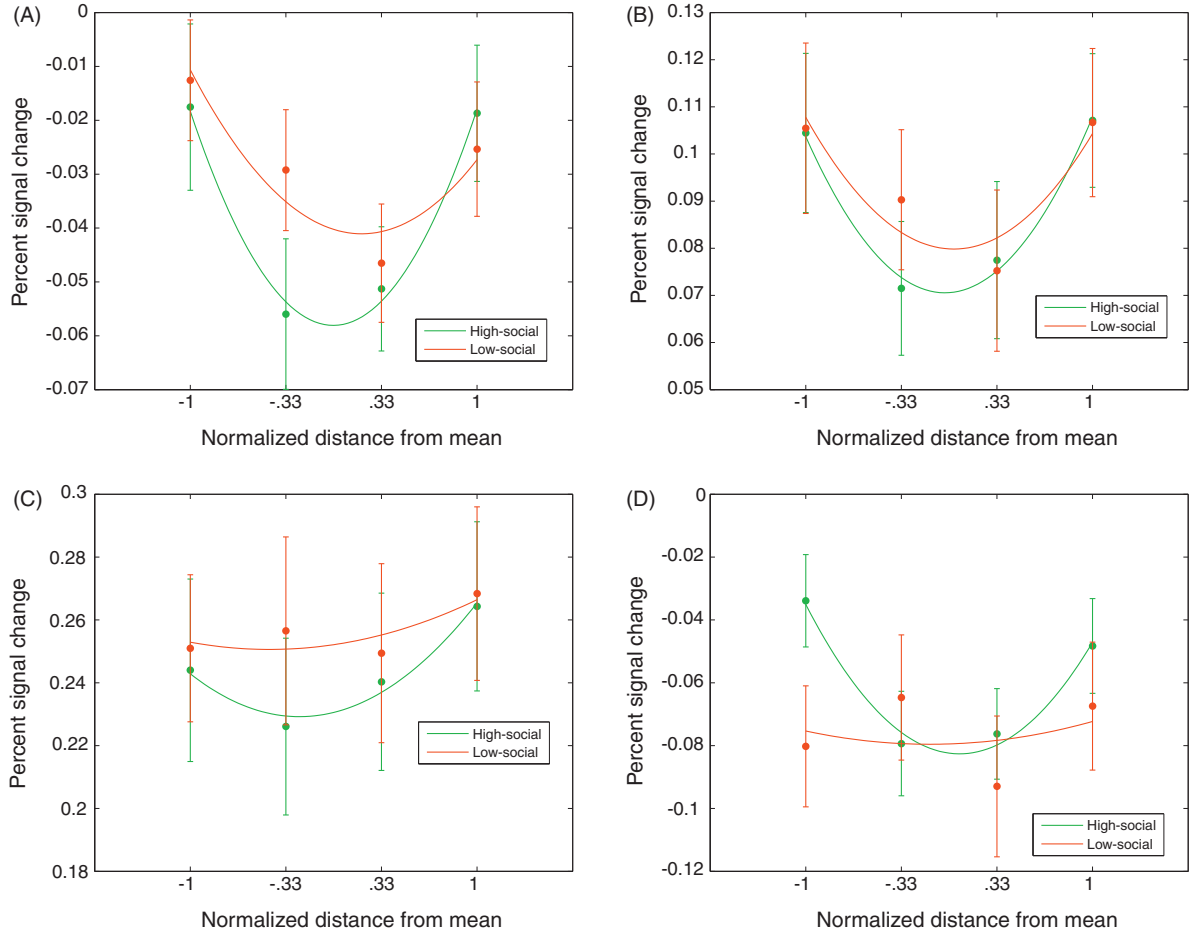

Fig. 6. The response to both the high-social dimension (valence) and the low-social dimension (control) in four ROIs. (A) The bilateral amygdala, defined anatomically by the AFNI Talairach atlas, (B) the bilateral FFA, (C) the bilateral OFA, and (D) the bilateral pSTS face-selective region.

Finally, it supports the idea that the amygdala tracks face typicality. As we show below, a typicality explanation can account for previous inconsistencies in the literature.

5.1. Brain areas responsive to the dimensions

Superficially, the amygdala showed a nonsignificant trend of a greater quadratic response to the valence dimension than the control dimension (Fig. 6(A)). However, this trend was driven by differences between faces near the center of face space, and not by faces nearer to the outside of face space, as our hypothesis would predict (Fig. 1(B)).

We also observed a significant difference between the two dimensions in the pSTS region, but not the OFA. In an fMRI study, Winston et al. (2002) showed that the pSTS activates more during explicit ratings of face trustworthiness compared to ratings of face age. Moreover, a recent study showed that transcranial magnetic stimulation (TMS) applied to the right pSTS interfered with trustworthiness judgments—making participants slower, whereas TMS applied to the right OFA interfered with gender judgments (Dzhelyova, Ellison, & Atkinson, submitted for publication), suggesting that pSTS plays a causal role in these judgments. Our finding is also consistent with face perception models that assume that the pSTS is critical for perception of emotional expressions (Haxby et al., 2000; Said et al., submitted for publication) and behavioral findings that perceptions of trustworthiness are based on facial similarity to emotional expressions (Oosterhof & Todorov, 2008, 2009; Said, Sebe, & Todorov, 2009).

However, while the studies cited above show that the pSTS is important for making trustworthiness decisions, they do not necessarily support a U-shape response function, as we observe here. Indeed, we urge some caution in interpreting our pSTS result, since no U-shape response function was observed in the pSTS in two previous experiments with stimuli and designs that were similar to the present experiment (Todorov et al., submitted for publication). One explanation of this inconsistency is that previous studies, which did not functionally localize the pSTS, failed to find a pSTS response to trustworthiness because of intersubject differences in the location of the face-selective pSTS region. These differences could blur out and reduce the group responses across anatomically aligned brain volumes. However, this explanation is not fully supported by the existing evidence. First, intersubject differences in location would most likely weaken the response in the pSTS rather than eliminate it. However, even at very low thresholds, we did not observe a clear pSTS response in the previous studies. Second, we were unable to replicate the current finding using a face localizer and a different task (detecting changes in color of a fixation cross). Clearly, more research is needed on the role of pSTS in face evaluation.

5.2. High social and low social dimensions

It is important to interpret the results in the amygdala, the FFA, and the OFA, where we observed no significant differences between the two dimensions. A strong interpretation is that the quadratic effect of valence is driven entirely by general distance from the average face. However, according to several metrics the range of

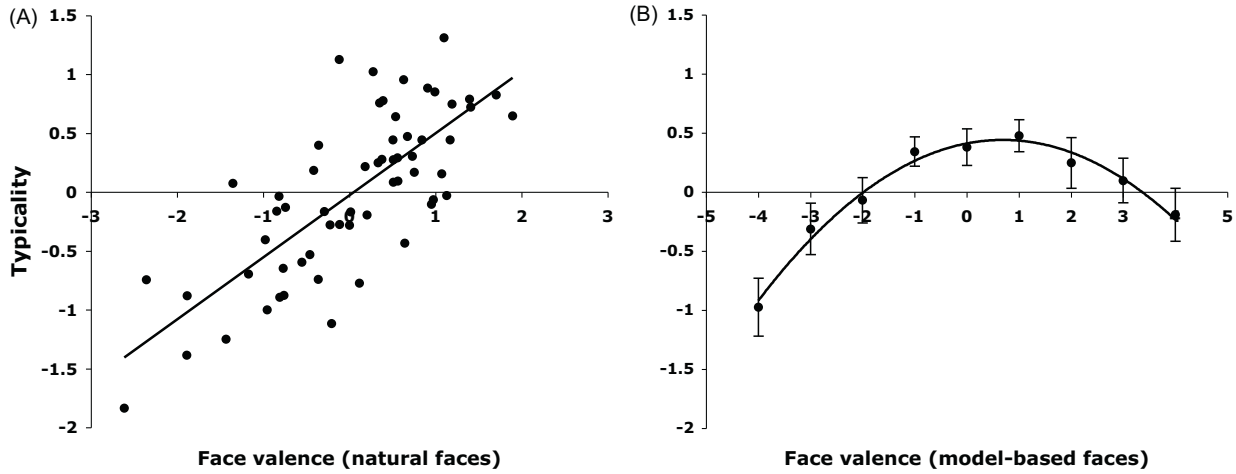

Fig. 7. The relationship between perceptions of face typicality and face valence for real faces (A) and computer generated faces (B). Both face valence and typicality are shown in standardized scores. The face valence for the real faces was obtained by a principle component analysis of 14 social judgments of the faces. The first principled component accounted for 63% of the variance of these judgments and was interpreted as face valence. The computer faces were generated by a model of face valence (see methods for details). Each point in the plots represents a face. The error bars in plot B represent the standard error of the mean.

faces we used along the control dimension was greater than the range used for valence. Thus, it is possible to obtain our results if the fMRI response encodes both social content and one of these metrics. Specifically, it is possible that a greater response to positively valenced and negatively valenced faces was counterbalanced by a greater response to control faces positioned even more distantly from the center of face space. Under this interpretation, quadratic responses to valence/trustworthiness would be partially, but not entirely, confounded by general distance from the average face. It is also worth noting that the control dimension contained some social information (Fig. 3(A)–(C)), thus reducing the magnitude of the social manipulation. Finally, since we only used one control dimension, it is possible that there are other control dimensions that would not show a quadratic effect.

5.3. Linear and non-linear response profiles

The findings here do not invalidate the linear responses to trustworthiness found in several previous studies (Engell et al., 2007; Said, Baron et al., 2009; Todorov et al., submitted for publication; Winston et al., 2002). While the results in this study showed no linear response, there are important differences between the studies. As we describe below, the stimuli used in the different studies differ substantially in their perceived typicality profile. We explain the differences under the assumption that the amygdala response magnitude depends on face typicality, although we suspect the explanation would also work well with correlates of typicality such as impact, intensity, or unusualness (Anderson et al., 2003; Ewbank et al., 2009).

In the present study and other studies that used artificial faces, the amygdala responded to both high valence and low valence faces. According to a typicality explanation of amygdala responses, this result implies that face typicality should be low for both high valence and low valence artificial faces (see Fig. 3(F) for judgments of unusualness). However, it is not clear whether a similar, non-linear relationship exists for real faces. If the relationship between perceptions of typicality and valence of real faces is linear, then the predicted brain response should be linear, as observed in most studies that used real faces. To test this hypothesis, we collected typicality judgments of (a) the set of real faces used in our prior fMRI studies that have found negative linear responses to face valence (Engell et al., 2007; Todorov & Engell, 2008) and (b) a set of 9 faces that varied on the valence dimension. Eighteen subjects (9 female; average age = 20.7; SD = 4.7) rated the real faces and 20 subjects (12 female; average age = 22.9; SD = 7.5) rated the artificial faces. Whereas typicality judgments were linearly related to face valence of real faces (Fig. 7(A)), they were quadratically related to face valence of artificial faces (Fig. 7(B)). Thus, these results can resolve apparent contradictions in the literature: studies that observed mostly linear responses to face valence used faces that had a linear relationship between valence and typicality. Studies that observed mostly quadratic responses to face valence used faces that had a quadratic relationship between valence and typicality. In all cases, typicality explains the amygdala response better than valence.

5.4. Social meaning in typicality

Although we failed to find positive evidence for coding of social properties of faces, it is important to note how difficult it is to find a face dimension that is free of social meaning. We produced 100 face dimensions that were orthogonal to dimensions with validated social meanings and selected one of these 100 dimensions that appeared to be the least sensitive to changes in face evaluation. Yet, variations along this dimension led to differences in social judgments (Fig. 3(A)–(C)). Faces are imbued with affective meaning (Todorov et al., 2008) and most changes in appearance could lead to changes in face evaluation. In this context, one interpretation of our findings is that during the long fMRI experiment, faces that deviated from the average face along the control dimension acquired just as much social relevance as faces that deviated along the valence dimension. Under this interpretation, the similar quadratic responses in the brain could be still explained by social factors.

More generally, although we can experimentally de-confound face typicality and perceived social value of faces, face typicality will be highly correlated with social attributions from faces in real life. In fact, for the set of real faces (Fig. 7(A)), perceived face typicality was linearly and significantly related to 13 out of 14 social judgments of these faces (see Todorov & Engell, 2008 for the list of these judgments). These correlations ranged in magnitude from .32 (negative relation with judgments of unhappiness) to .92 (negative relation with judgments of weirdness). That is, evaluative social judgments were highly correlated with non-evaluative, descriptive judgments of face typicality ("how likely would you be to see a person who looks like this walking down the street?"). These findings show that face typicality is an important principle of organizing social information.

Moreover, assuming that faces are neurally represented in something akin to face space, coding faces along their typicality would be a simple way for organizing face information. Although to extract face typicality, the brain only needs to learn the statistical properties of faces, face typicality nevertheless would correlate with a number of important social variables. For example, average faces are generally perceived to be healthier and more attractive, and face averageness prominently figures in evolutionary models of face perception (Rhodes, 2006; Thornhill & Gangestad, 1999). As a result, coding based on the typicality of the faces could provide information relevant to social perception.

Acknowledgements

We thank Jenny Porter and Olivia Kang for their help with the experiments. This research was supported by NSF grant 0823749.

References

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature*, *433*(7021), 68–72.

Adolphs, R., & Spezio, M. (2006). Role of the amygdala in processing visual social stimuli. *Progress in Brain Research*, *156*, 363–378.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*(6684), 470–474.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669–672.

Aguirre, G. K. (2007). Continuous carry-over designs for fMRI. *NeuroImage*, *35*(4), 1480–1494.

Amaral, D. G. (2003). The amygdala, social behavior, and danger detection. *Annals of the New York Academy of Sciences*, *1000*, 337–347.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience*, *6*(2), 196–202.

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*(5), 875–887.

Canli, T., Sivers, H., Whitfield, S. L., Gotlib, I. H., & Gabrieli, J. D. E. (2002). Amygdala response to happy faces as a function of extraversion. *Science*, *296*(5576), 2191–12191.

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

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, An International Journal*, *29*(3), 162–173.

Cunningham, W. A., Van Bavel, J. J., & Johnsen, I. R. (2008). Affective flexibility: Evaluative processing goals shape amygdala activity. *Psychological Science*, *19*(2), 152–160.

Dzhelyova, M. P., Ellison, A., & Atkinson, A. P. (submitted for publication). Event-related repetitive TMS reveals distinct, critical rols for right OFA and bilateral pSTS in judging the trustworthiness and sex of faces.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*(9), 1508–1519.

Ewbank, M. P., Barnard, P. J., Croucher, C. J., Ramponi, C., & Calder, A. J. (2009). The amygdala response to images with impact. *Social Cognitive and Affective Neuroscience*, *4*(2), 127–133.

Finney, D. J., & Outhwaite, A. D. (1956). Serially balanced sequences in bioassay. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *145*(921), 493–507.

Fox, C. J., Iaria, G., & Barton, J. J. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. *Human Brain Mapping*, *30*(5), 1637–1651.

Gescheider, G. A. (1997). Psychophysics: The Fundamentals. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Kennedy, D. P., Glascher, J., Tyszka, J. M., & Adolphs, R. (2009). Personal space regulation by the human amygdala. *Nature Neuroscience*, *12*(10), 1226–1227.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., et al. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(12), 5675–5679.

LeDoux, J. (2007). The amygdala. *Current Biology*, *17*(20), R868–R874.

Leopold, D. A., Bondar, I. V., & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*(7102), 572–575.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, *8*(10), 1386–1390.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–610.

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*(6603), 812–815.

Ogawa, S., Lee, T. M., Nayak, A. S., & Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, *14*(1), 68–78.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092.

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*(1), 128–133.

Pessoa, L., Japee, S., Sturman, D., & Ungerleider, L. G. (2006). Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cerebral Cortex*, *16*(3), 366–375.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199–226.

Said, C. P., Baron, S. G., & Todorov, A. (2009). Non-linear amygdala response to face trustworthiness: Contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience*, *21*(3), 519–528.

Said, C. P., Haxby, J., & Todorov, A. (submitted for publication). Brain systems for the recognition of facial expressions and evaluation of emotionally neutral faces. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*.

Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion*, *9*(2), 260–264.

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends Cognitive Sciences*, *3*(12), 452–460.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience*, *3*(2), 119–127.

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, *25*(3), 395–410.

Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Social Cognitive and Affective Neuroscience*, *3*(4), 303–312.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833.

Todorov, A., Said, C. P., Oosterhof, N. N., & Engell, A. D. (submitted for publication). Untitled.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *43*(2), 161–204.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science*, *7*(6), 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*(5704), 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 expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*(1), 411–418.

Winston, J. S., O'Doherty, J., & Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage*, *20*(1), 84–97.

Winston, J. S., O'Doherty, J., Kilner, J. M., Perrett, D. I., & Dolan, R. J. (2007). Brain systems for assessing facial attractiveness. *Neuropsychologia*, *45*(1), 195–206.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*(3), 277–283.