Neuropsychologia 50 (2012) 2205–2211

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

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

Response of face-selective brain regions to trustworthiness and gender of faces

Giulia Mattavelli [a,c,*], Timothy J. Andrews [b,c], Aziz U.R. Asghar [c,d], John R. Towler [b], Andrew W. Young [b,c]

[a] Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy
[b] Department of Psychology, University of York, Heslington, YO10 5DD York, UK
[c] York Neuroimaging Centre, The Biocentre, University of York, Heslington, YO10 5DD York, UK
[d] Hull York Medical School, University of Hull, HU6 7RX Hull, UK

ARTICLE INFO

Article history:
Received 20 September 2011
Received in revised form
4 May 2012
Accepted 23 May 2012
Available online 30 May 2012

Keywords:
Amygdala
Face-selective regions
Face perception
Trustworthiness

ABSTRACT

Neuropsychological and neuroimaging studies have demonstrated a role for the amygdala in processing the perceived trustworthiness of faces, but it remains uncertain whether its responses are linear (with the greatest response to the least trustworthy-looking faces), or quadratic (with increased fMRI signal for the dimension extremes). It is also unclear whether the trustworthiness of the stimuli is crucial or if the same response pattern can be found for faces varying along other dimensions. In addition, the responses to perceived trustworthiness of face-selective regions other than the amygdala are seldom reported. The present study addressed these issues using a novel set of stimuli created through computer image-manipulation both to maximise the presence of naturally occurring cues that underpin trustworthiness judgments and to allow systematic manipulation of these cues. With a block-design fMRI paradigm, we investigated neural responses to computer-manipulated trustworthiness in the amygdala and core face-selective regions in the occipital and temporal lobes. We asked whether the activation pattern is specific for differences in trustworthiness or whether it would also track variation along an orthogonal male–female gender dimension. The main findings were quadratic responses to changes in both trustworthiness and gender in all regions. These results are consistent with the idea that face-responsive brain regions are sensitive to face distinctiveness as well as the social meaning of the face features.

1. Introduction

Faces are multi-dimensional stimuli conveying crucial information for social interaction and people are highly skilled at making social judgements to faces (Bruce & Young, 2012). For example, judgements of trustworthiness from facial appearance are remarkably consistent across different observers, and can even be made with very brief presentations (Bar, Neta, & Linz, 2006; Willis & Todorov, 2006). Theories of social perception link this rapid evaluation of trustworthiness to a more general conception of primate behaviour in which individuals in a social group are evaluated for potential threat (warmth, or approachability) and their capacity to enact any such threat (competence, or dominance) (Fiske, Cuddy, & Glick, 2007; Todorov, 2008). In such models the evaluation of trustworthiness is closely linked to approachability, and studies show that ratings of trustworthiness and approachability are highly correlated (Oosterhof & Todorov, 2008; Santos & Young, 2008a, 2008b).

Neuropsychological studies have demonstrated a role for the amygdala in processing trustworthiness and approachability (Adolphs, 1999; Adolphs, Baron-Cohen, & Tranel, 2002; Cristinzio, Sander, & Vuilleumier, 2007). Patients with amygdala damage rate untrustworthy-looking faces as more approachable and trustworthy than do neurologically normal participants, consistent with a more general problem in the evaluation of potential threat and danger in the environment (Adolphs, Tranel, & Damasio, 1998; Feinstein, Adolphs, Damasio, & Tranel, 2010).

The role of the amygdala in evaluating trustworthiness has also been supported by functional neuroimaging studies, but with mixed results. Early studies showed greater response in the amygdala for untrustworthy as compared to trustworthy faces (Winston, Strange, O'Doherty, & Dolan, 2002) with a linear trend in amygdala activation for increasing untrustworthiness (Engell, Haxby, & Todorov, 2007). Other studies have found U-shaped, quadratic responses in the amygdala (Said, Baron, & Todorov, 2008; Todorov, Baron, & Oosterhof, 2008), with increased responses at the extremes of the trustworthiness dimension; however, these studies reported both

* Corresponding author at: Department of Psychology, University of Milano-Bicocca, Piazza Ateneo Nuovo, 1, 20126 Milano, Italy.
Tel.: +39 02 6448 3866; fax: +39 02 6448 3706.
E-mail address: g.mattavelli2@campus.unimib.it (G. Mattavelli).

0028-3932/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neuropsychologia.2012.05.024

linear and non-linear components in amygdala activation, preventing unequivocal conclusions concerning how the amygdala processes this social dimension. Interestingly, U-shaped functions are also apparent to faces that vary along other social dimensions such as dominance (Said, Dotsch, & Todorov, 2010).

These contrasting activation patterns across differences in perceived trustworthiness lead to different interpretations of the role of the amygdala in social evaluation. A linear response is in line with the hypothesis that the amygdala is activated by arousing and threatening signals (Gläsher & Adolphs, 2003; Lane et al., 1997) and involved in evaluating the valence of negative stimuli (Todorov & Engell, 2008). On the other hand, a U-shaped quadratic pattern is more consistent with the hypothesis that the amygdala is activated by salient social cues independent from whether they have a positive or negative valence (Said et al., 2008). The quadratic pattern is also consistent with the idea that faces are represented in a multidimensional space in which the origin represents the average face and more distinctive faces are represented away from the origin (Said et al., 2010; Valentine, 1991). From this perspective, the linear and nonlinear responses to trustworthiness in previous studies could be due to uncontrolled variation in the distinctiveness of faces (Said, Haxby, & Todorov, 2011).

A key aim of the present study was therefore to address these different perspectives on the way that the amygdala represents information about faces by comparing the neural responses to trustworthiness and a control face dimension (male–female). To do this we developed a novel set of naturalistic face stimuli varying in perceived trustworthiness and along an orthogonal male–female dimension. Previous studies have used face photographs, which cannot vary relevant stimulus dimensions systematically, or computer-synthesised faces that, whilst useful, form highly constrained sets that may not utilise all of the cues that are naturally available to human observers. Our stimuli were derived from prototype images created with a photograph averaging technique, in order to maximise the presence of naturally occurring cues that underpin trustworthiness and gender judgments. These prototypes were then systematically manipulated through image-morphing to create independent dimensions of trustworthiness and gender.

Neural responses to these novel sets of stimuli were tested using a block design fMRI paradigm, to take advantage of its greater statistical power compared to event-related designs (Sergerie, Chochol, & Armony, 2008). If the social meaning of facial trustworthiness cues is crucial to determining the neural responses, we would expect the patterns of activation to vary with the trustworthiness of the faces, but not with changes in gender. On the other hand, if the distinctiveness of the face is important, then a similar pattern of activation should be evident for variation in both the social and control dimensions (Said et al., 2010, 2011).

A second aim of our study was to determine whether the pattern of response was specific to the amygdala or was evident in other face-responsive regions of the brain. Most previous studies have drawn conclusions based only on responses from the amygdala region itself, but it is crucial to correctly interpreting these amygdala responses to know whether they are similar or different in form from the responses of other regions involved in face perception. We therefore analysed responses from core face-selective regions of the occipital and temporal lobes (Haxby, Hoffman, & Gobbini, 2000) as well as the amygdala itself.

2. Material and methods

2.1. Participants

Twenty healthy volunteers (ten male, ten female, mean age = 22.9 years, range 18–35) took part in the experiment. All participants were right-handed, with a western cultural background, and had normal or corrected to normal vision with no history of neurological illness. The study was approved and conducted following the guidelines of the Ethics Committee of the York Neuroimaging Centre, University of York. All participants gave written consent prior to their participation.

2.2. Experiment stimuli

Fig. 1 shows the complete matrix of images from which the stimuli used in the experiment were selected. The matrix was created as follows. Photographs of 500 adult male and 500 adult female faces were collected from the internet. The photographs varied in pose, age and expression, to allow as wide a range of cues as possible to be present in the images. However, photographs of famous people were excluded, to eliminate potential influences of prior knowledge about the person. Moreover, only Caucasian adult faces were chosen, to reduce potential cultural influences. The 1000-face photographs were rated for trustworthiness (using 1–7 scales) by six independent raters. From these ratings the 15 highest and 15 least trustworthy male faces and the 15 highest and 15 least trustworthy female faces were selected, subject to constraints that the photographs included no spectacles, were as close to frontal view as possible, showed no beards or moustaches, and that there were no more than two faces with hats in each set. There was no matching on any other characteristics, with free variation of all other aspects. The faces in each set of 15 photographs were then averaged using PsychoMorph software (Tiddeman, Burt, & Perrett, 2001) to create four prototypes (high and low trustworthy male, high and low trustworthy female). Image continua were then created for trustworthiness of male faces (from very high to very low trustworthiness) and for trustworthiness of female faces by caricaturing each prototype at two levels to increase its distance from the opposite prototype and by anti-caricaturing each prototype at two levels to decrease distance from the opposite prototype. For example, the highly trustworthy male prototype was caricatured to enhance its trustworthiness by increasing differences from the low trustworthy male prototype and it was anti-caricatured to diminish its trustworthiness by decreasing differences from the low trustworthy male prototype. In this way, a quasi-linear continuum of 10 male face-like images of varying trustworthiness was created, and a corresponding continuum of 10 female face-like images of varying trustworthiness.

These continua of 10 images were then presented in random order and rated for trustworthiness (on a 1–7 low–high trustworthy scale) by 10 raters (5 male, 5 female, mean age = 20.4 years, S.D. = 0.55) who did not otherwise participate in the study. The correlation between rated trustworthiness and position on the appropriate continuum was 0.94 for the male images and 0.95 for the female images, showing that the caricaturing and anti-caricaturing manipulations were successful in creating continua varying systematically in perceived trustworthiness. However, it was also necessary to match continua needed for the present experiment so that the male and female prototype images were of equivalent high or equivalent low trustworthiness. We therefore selected a male and a female image that were rated equally low in trustworthiness, and a male and a female image that were rated equally high. These matched pairs of male and female images formed the four new prototypes used to generate Fig. 1. They are shown at highlighted positions in Fig. 1 corresponding to the intersections of the second and ninth rows with the second and ninth columns. The rest of the 10×10 matrix was generated by morphing the faces between the prototypes along the trustworthiness and the gender dimensions and adding a caricatured image in each of the four directions. If we consider the prototypes to represent 0% and 100% on each dimension, the manipulation used generated images with the following percentages along the gender (horizontal) and trustworthiness (vertical) axes of Fig. 1: −15%, 0%, 15%, 30%, 45%, 55%, 70%, 85%, 100%, and 115%. On this scale, values falling outside the 0–100% range represent caricatures with respect to the opposite prototype.

2.3. Imaging parameters

Scanning was performed at the York Neuroimaging Centre at the University of York with a 3 T HD MRI system with an eight channels phased array head coil (GE Signa Excite 3.0 T, High resolution brain array, MRI Devices Corp., Gainesville, FL). Axial images were acquired for functional and structural MRI scans. For fMRI scanning, echo-planar images were acquired using a T2* weighted gradient echo sequence with blood oxygen level-dependent (BOLD) contrast (TR = 3 s, TE = 32.7 ms, flip-angle = 90°, acquisition matrix 128×128, field of view = 288×288 mm). Whole head volumes were acquired with 38 contiguous axial slices, each with an in-plane resolution of 2.25×2.25 mm and a slice thickness of 3 mm. The slices were positioned for each participant to ensure optimal imaging of the temporal lobe regions, where the amygdala is situated. T1-weighted images were acquired for each participant to provide high-resolution structural images using an Inversion Recovery (IR = 450 ms) prepared 3D-FSPGR (Fast Spoiled Gradient Echo) pulse sequence (TR = 7.8 s, TE = 3 ms, flip-angle = 20°, acquisition matrix = 256×256, field of view = 290×290 mm, in-plane resolution = 1.1×1.1 mm, slice thickness = 1 mm). To improve co-registration between fMRI and the 3D-FSPGR structural a high resolution T1 FLAIR was acquired using the same physical dimensions as the fMRI protocol (TR = 2850 ms, TE = 10 ms,

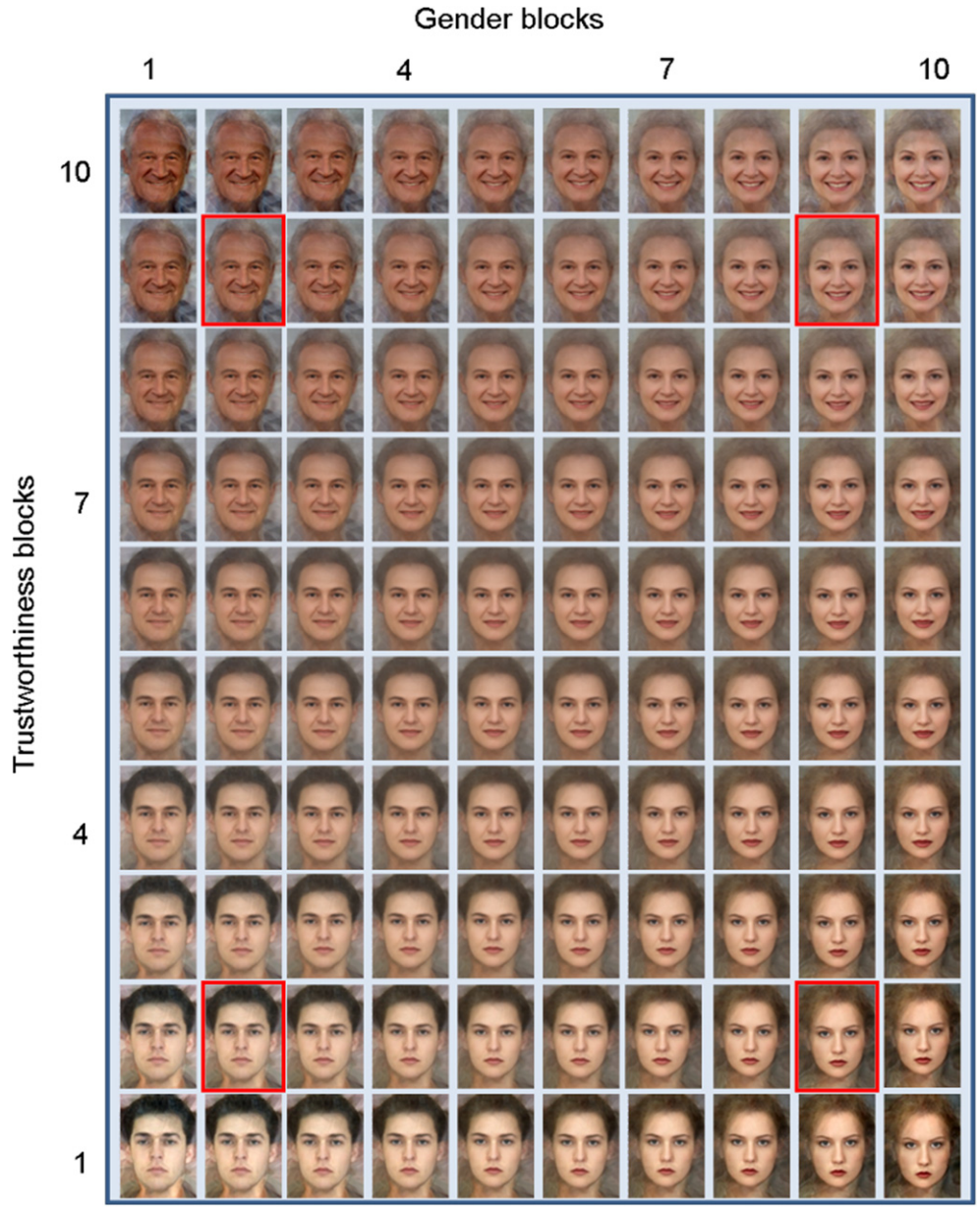

Fig. 1. Matrix of faces created by computer image manipulation. Images in red squares represent the prototypes used to produce the matrix of 10 levels of face gender (rows) and 10 levels of face trustworthiness (columns). Four trustworthiness conditions and four gender conditions were selected for the fMRI experiment, in order to cover the full range of each of the dimensions. Stimuli for the trustworthiness blocks were the rows labelled with numbers 1, 4, 7 and 10 of the matrix, thus including 10 different face images with the same trustworthiness level but varying in terms of gender. Gender blocks consisted of columns 1, 4, 7 and 10, each with 10 faces varying in trustworthiness but constant in terms of gender.

acquisition matrix 256 × 224 interpolated to 512 giving effective in plain resolution of 0.56 mm).

2.4. Localiser scan

In order to identify brain regions responding selectively to faces, participants performed a separate localiser scan (see Andrews, Davies-Thompson, Kingstone, & Young, 2010). Twenty blocks with ten images were run, using Neurobehavioural System Presentation 13.0 software. Each block contained images from one of five different categories: faces, bodies, objects, places or Fourier-scrambled images derived from the previous categories. Face images were taken from the Psychological Image Collection at Stirling (PICS; http://pics.psych.stir.ac.uk/) and bodies were selected from a body images collection at Bangor (http://pages.bangor.ac.uk/~pss811/page7/page7.html). Images of other categories were taken from website sources. Each image was presented for 700 ms followed by a 200 ms fixation cross, giving a block duration of 9 s for the 10 images. Stimulus blocks were interleaved with resting periods of 9 s with a fixation cross superimposed on a grey screen. The five conditions were repeated four times in a counterbalanced order.

2.5. Trustworthy/gender scan

The experiment aimed to test whether the response patterns in the amygdala and face-selective regions are specific to the trustworthiness dimension or if similar patterns appear for faces varying along an independent and orthogonal male–female dimension. A block design was used with eight conditions divided into four trustworthiness conditions and four gender conditions. Each of the four trustworthiness blocks comprised the images from a row of the stimulus matrix shown in Fig. 1 (rows labelled as 1, 4, 7 and 10 were selected) and therefore involved faces varying in terms of gender but with the same trustworthiness level. Each of the four gender blocks consisted of a column from the stimulus matrix (columns 1, 4, 7 and 10 were selected) and therefore involved faces varying in level of trustworthiness but not in terms of gender. Consequently, the eight presented conditions sampled the full range of each of the two orthogonal dimensions. The blocks for each condition were repeated five times in a counterbalanced order. Within each block the 10 images were presented in a pseudorandom order for 1 s each followed by a 200 ms fixation cross, giving a total block duration of 12 s; blocks were interleaved with a 12 s fixation cross on a grey screen. To monitor attention during the scan session a red spot detection task was

used. In one or two images per block a small red spot appeared; subjects were instructed to look at the stimuli and press with the right index finger a response button whenever they saw the red spot. Subjects responded correctly to the majority of the red spot trials (mean accuracy = 98.6%, S.D. = 2.87).

After the fMRI scan a behavioural task was run to check how each participant perceived the stimuli. Participants were asked to rate on a 7-point scale the trustworthiness (1 = very untrustworthy, 7 = very trustworthy) and the masculinity–femininity (1 = high masculine, 7 = high feminine) of the images used in the experiment. These two sets of ratings were completed separately in a counterbalanced order.

2.6. fMRI data analysis

Image analyses were performed by means of FEAT (FMRI Expert Analysis Tool), part of FSL (http://www.fmrib.ox.ac.uk/fsl). For each participant the following pre-statistic processing was applied: motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), slice-timing correction using Fourier-space time-series phase-shifting, non-brain removal using BET (Smith, 2002), spatial smoothing using a Gaussian kernel (FWHM 5 mm in the localiser scan and 6 mm in the main experiment), grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative factor; high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 60.0 s in the localiser scan and sigma = 120.0 s in the main experiment).

Face-selective regions comprising the core components identified by Haxby et al. (2000) were individually defined in each participant's brain using the localiser scan by averaging the four contrasts faces > bodies, faces > objects, faces > places and faces > scrambled images. The average of these four contrasts in each participant was thresholded at $Z > 2.6$ ($p < .005$, uncorrected). In this way, the fusiform face area (FFA), occipital face area (OFA) and right posterior superior temporal sulcus (pSTS) could be indentified at the level of each single participant. These regions of interest (ROIs) were defined from the thresholded statistical images (see Andrews et al., 2010). The FFA, OFA and pSTS each appeared as a contiguous cluster of voxels in each participant located respectively in the inferior fusiform gyrus, in the posterior occipital cortex and in the superior temporal lobe. A different approach had to be taken to define the amygdala, which is not reliably identified through a functional localiser at the individual level. A face-responsive ROI in the amygdala was therefore defined by considering the statistical map of amygdala activation at the group level, resulting from the four contrasts averaged and thresholded at $Z = 3$ ($p \leq .001$, uncorrected), which was back-transformed into the individual MRI space for each participant.

Within these functionally identified face-selective regions (amygdala, OFA, FFA, pSTS) derived from the functional localiser scan, data from the main experiment were analysed by extracting the time-course of the filtered MR data as per cent signal change in each voxel and then averaging the voxels within each ROI for each participant. The average time-course for the different conditions was calculated and data were normalised relative to the zero time point for that stimulus block. The peak of activation, considered as the average of the response between 9 s and 15 s after block onset, was used for the analyses.

For each ROI the following analyses were performed to test the linear and quadratic responses. First, a linear regression and a second-order polynomial were fitted to the responses at group level in order to investigate the activation pattern in each region. Second, a linear regression and a second-order polynomial were fitted to each individual participant's responses and paired *t*-tests were used to test differences between the *R*-squared of the two fitted equations in each ROI. Finally, paired sample *t*-tests were performed to compare the linear and quadratic regressions for the gender and trustworthiness dimensions.

3. Results

3.1. Behavioural data

The post-scan behavioural ratings were analysed to check that the participants in the fMRI experiment rated the stimuli in line with what was intended. The trustworthiness and gender ratings of each participant were correlated with the four trustworthiness and the four gender levels included in the fMRI scan. One participant was excluded from the following analyses because of a very low correlation score for the trustworthiness rating ($r = 0.01$), whereas for the remaining participants the correlations were always > 0.8 for both dimensions (mean $r = 0.96$, for trustworthiness rating; mean $r = 0.98$, for gender rating). In this post-scan behavioural task, participants rated the stimuli on a 7-point scale separately for the trustworthiness and gender dimensions. We were interested in the stimuli included in the fMRI experiment, hence levels 1, 4, 7 and 10 of each dimension (see Fig. 1). The mean rating for trustworthiness level 1 was 2.36 (S.D. = 1.27), 3.94 (S.D. = 0.77) for level 4, 4.87 (S.D. = 0.84) for level 7, and 6.14 (S.D. = 1.14) for level 10. The mean rating for gender level 1 was 1.53 (S.D. = 0.38), 2.85 (S.D. = 0.67) for level 4, 5.26 (S.D. = 0.72) for level 7, and 6.37 (S.D. = 0.91) for level 10. The mean rating for both the dimensions significantly correlated with the trustworthiness ($r = .995$, $p = .005$) and gender levels ($r = .99$, $p = .01$).

3.2. Localiser scan

Fig. 2 shows the location of regions within the amygdala, the occipital and temporal lobes (FFA, OFA, pSTS) that showed face-selective activity from a whole-brain group analysis of the localiser scan data. Mean MNI coordinates and size of each region across participants are reported in Table 1. The FFA and OFA were identified in all of the 19 participants and right pSTS in 18 participants.

3.3. Trustworthy/gender scan

Fig. 3 shows the peak response in each ROI for faces varying along the trustworthiness and gender dimensions. Since both hemispheres showed similar response patterns in FFA, OFA and amygdala, the responses in the right and left hemispheres were combined for these regions. In contrast, the pSTS region could only be reliably identified in the right hemisphere. For the trustworthiness dimension, results at group level showed bigger *R*-squared values for the quadratic polynomial than for the linear regression in all the face-selective regions. The same pattern of greater overall quadratic than linear responses for all regions was also seen for the gender dimension (Table 2).

Quadratic and linear regressions were then fitted to the individual responses in each ROI and paired sample *t*-tests confirmed that the *R*-squared values for the quadratic polynomial were significantly higher than the *R*-squared for the linear regression for both the dimensions in all the regions (amygdala: trustworthiness [$t(18) = 4.97$, $p < .001$], gender [$t(18) = 6.33$, $p < .001$]; FFA: trustworthiness [$t(18) = 5.07$, $p < .001$], gender [$t(18) = 5.1$, $p < .001$]; OFA: trustworthiness [$t(18) = 6.12$, $p < .001$], gender [$t(18) = 4.12$, $p = .001$]; right pSTS: trustworthiness [$t(17) = 4.15$, $p = .001$], gender [$t(17) = 5.27$, $p < .001$]).

Having established the general pattern of quadratic rather than linear response in all ROIs, it is of interest to ask whether the quadratic component was more pronounced for one dimension than the other. However, there were no significant differences comparing quadratic *R*-squared between the two dimensions of gender and trustworthiness in any region (paired sample *t*-tests, $p > .05$).

4. Discussion

In the present study we investigated the response pattern in the amygdala and the core face-selective brain regions to faces varying in a social (trustworthiness) and a control (male–female) gender dimension. A novel set of stimuli were created that consisted of naturalistic face images, which varied systematically in perceived trustworthiness and gender. The behavioural results showed that this method captures the multiplicity of cues that are used to evaluate variations in trustworthiness. There were high correlations between rated trustworthiness and vertical position of the images shown in Fig. 1. Since the essence of the method used to derive the prototype images was simply averaging face photographs rated as high or low in trustworthiness, the continued presence of high and low trustworthiness in the averaged

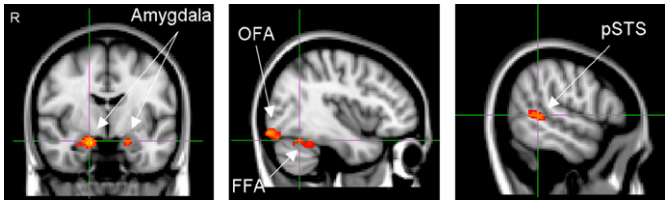

Fig. 2. Location of the face-selective regions (amygdala, FFA, OFA, pSTS) in a whole-brain group analysis of the localiser scan. Statistical parametrical maps thresholded at $Z=3$ ($p \leq .001$, uncorrected) resulting from the average of four contrasts (faces > bodies, faces > objects, faces > places and faces > scrambled images) are represented. Images follow the radiological convention, with the right hemisphere represented on the left side.

Table 1
MNI coordinates and size of face-selective regions. The left and right amygdala were defined at the group level. FFA, OFA and pSTS were defined in each participant; values represent the mean (S.D.) across all 19 participants.

Region		*n*	MNI coordinates (*x*, *y*, z)			Size (cm^3)
Amygdala	R	19	18	−6	−18	4.44
	L	19	−18	−10	−18	1.24
FFA	R	19	42 (4)	−56 (8)	−23 (5)	2.23 (1.48)
	L	18	−42 (4)	−58 (7)	−23 (4)	1.35 (1.06)
OFA	R	19	39 (6)	−81 (9)	−14 (5)	2.19 (1.93)
	L	18	−37 (5)	−83 (5)	−18 (5)	1.42 (1.25)
pSTS	R	18	50 (8)	−53 (8)	5 (6)	0.82 (0.79)

prototype images shows that the cues that convey these impressions must have been reasonably consistently present in the original photographs. Inspection of Fig. 1 suggests that the trustworthiness dimension involves cues that include a combination of age, skin colour and hostile expression. Using computer models, Oosterhof and Todorov (2008) have already shown that trustworthiness evaluation is sensitive to emotional expression. In the same study maturity cues did not correlate with trustworthiness evaluation when internal features, linked with the trustworthiness features, were masked. However, Oosterhof and Todorov's (2008) stimuli were synthesised computer images with a limited range of ages and smoothing of texture cues such as wrinkles that can signify a loss of elasticity in the skin. In contrast, our stimuli were created with a data-driven approach without any a priori constraint, thus taking into account the multiplicity of naturally occurring cues which influence trustworthiness judgments, and age seems to be part of this evaluation.

To determine how the brain responded to the stimulus set, a localiser scan was used to functionally define the amygdala and the core face-selective regions (OFA, FFA, pSTS) of the occipital and temporal lobes (Haxby et al., 2000). The main findings from this analysis were: (i) the amygdala responded to varied trustworthiness with a U-shaped quadratic function; (ii) the amygdala also showed a U-shaped pattern of response to changes in gender; (iii) FFA, OFA and right pSTS showed a similar U-shaped pattern for both the trustworthiness and gender dimensions.

Differences in the pattern of response in the amygdala to variations in trustworthiness have been reported in different

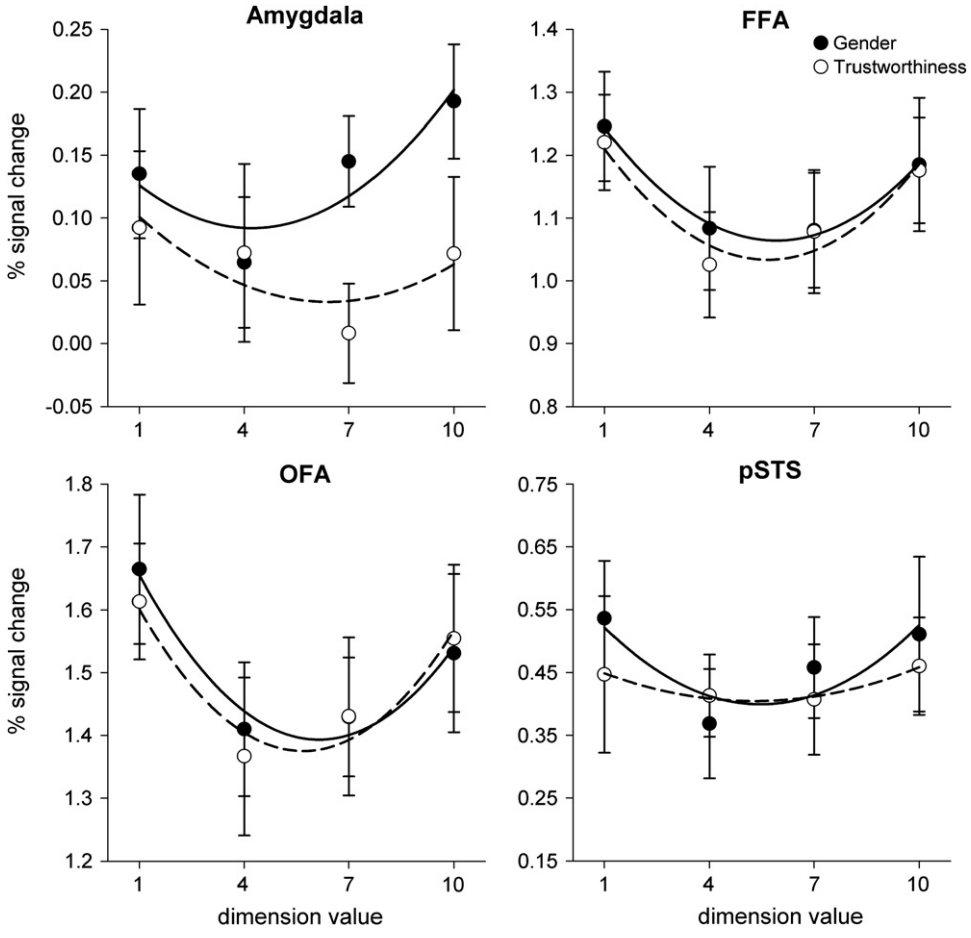

Fig. 3. Response to trustworthiness and gender dimensions in the four ROIs defined by the localiser scan. U-shaped lines represent the quadratic polynomial that best fitted the data in the trustworthiness and gender dimensions; bars represent standard errors of the mean.

Table 2
R-squared values for the quadratic polynomial and linear regressions for the two dimensions.

	Trustworthiness		Gender	
	Quadratic	Linear	Quadratic	Linear
Amygdala	0.63	0.20	0.80	0.38
FFA	0.91	0.01	0.99	0.09
OFA	0.92	0.02	0.95	0.18
pSTS	0.97	0.03	0.74	0

neuroimaging studies. Some studies have reported a greater response in the amygdala for untrustworthy as compared to trustworthy faces (Winston et al., 2002) with a linear trend in amygdala activation for increasing untrustworthiness (Engell et al., 2007), whereas other reports have found U-shaped quadratic responses in the amygdala (Said et al., 2008; Todorov et al., 2008). Our results provide support for a U-shaped quadratic response to trustworthiness in the amygdala. Critically, we replicated this finding for a functionally defined face-selective region within the amygdala using a novel set of naturalistic face images which varied systematically in perceived trustworthiness and using the higher statistical power gained from a fMRI block design.

The specificity of amygdala responses to social cues in faces has remained an issue of debate. U-shaped amygdala responses to a face dimension different from trustworthiness have been previously reported by Winston, O'Doherty, Kilner, Perrett, and Dolan (2007), who found greater activation in the right amygdala when highly attractive or unattractive faces were presented compared to moderately attractive faces; although a correlation between attractiveness and face valence could potentially have influenced this result (Todorov & Engell, 2008). Another fMRI study with computer-generated images showed a quadratic response for faces that varied along a dimension orthogonal to trustworthiness with lower social relevance (Said et al., 2010). Our results confirm these previous findings since we found non-linear activation in the amygdala to realistic faces morphed along a continuum of perceived trustworthiness and a comparison gender dimension. Notably, the entirely data-driven approach used to create the present stimuli offers independent confirmation that findings with more artificial stimulus sets can be considered reliable, and of course enhances the ecological validity of our results by allowing us to systematically manipulate the dimensions of interest without constraints on the range of cues naturally available in face perception.

As well as the amygdala, we were also interested in clarifying the response pattern in face-selective regions in the occipital and temporal lobes. To do this we used the functional localiser scan to define bilateral FFA, bilateral OFA and right pSTS in each participant, and then extracted the per cent signal change during the main experiment within each ROI. These face-selective regions form Haxby et al.'s (2000) core system for face perception, and all of them showed U-shaped activations; again with no significant difference between the two dimensions. Previous studies mostly focussed on the activation within the amygdala (Engell et al., 2007; Todorov et al., 2008) and hypothesised that the activity in the posterior face-selective regions was modulated by the amygdala (Todorov & Engell, 2008). Only Said et al. (2010) used a separate localiser scan to functionally define the face-selective regions on an individual subject level and, similarly to our results, they reported quadratic activations in FFA for their social and non-social dimensions. However, responses in OFA and pSTS were less clear in Said et al.'s (2010) study, showing a non significant quadratic trend in OFA and a quadratic effect in pSTS for the social dimension but not for the non-social dimension. In contrast, our results show a common quadratic pattern in all the face-selective regions, which might be taken to suggest that these areas are equally important for the perceptual analysis of the stimuli. The different experimental designs used in our and in Said et al.'s (2010) study might potentially account for the different effects found in OFA and pSTS. Besides this, though, the features of the control dimension could have a key role in understanding activations in these regions. Our stimuli were varied along two dimensions that are both well recognisable as face categories, trustworthiness and gender, whereas Said et al. (2010) used computer modelling to generate a control dimension orthogonal to the social dimension but not definable as a specific face category. Therefore, it may prove to be the case that ability to identify face variations as ecologically relevant dimensions is important to eliciting quadratic responses in OFA and pSTS.

Overall, our findings can be interpreted in line with the concept that faces are represented by a multidimensional space in which each face represents a particular location (Valentine, 1991). The origin of the face space reflects the average face and faces are more distinctive as the distance from the origin increases. Neuroimaging support for this perspective was reported by Loffler, Yourganov, Wilkinson, and Wilson (2005), who showed that the response of face-selective regions increases with the geometric distance from the average face. The U-shaped function shown here and in other studies provide support for the idea that responses from the amygdala and other face-selective regions are at least in part driven by coding the difference between the presented faces and an average face, regardless of the specific social meaning of the stimuli (Said et al., 2010, 2011).

An alternative explanation might be that trustworthiness and gender are both important dimensions which require specific coding. However, the hypothesis of a multidimensional representation for face stimuli at present seems more likely in light of previous findings of increased fMRI signal for increasing distinctiveness in face geometry (Loffler et al., 2005) and reports of quadratic activations for different face dimensions manipulated both with computer models and with photographs (Said et al., 2010; Winston et al., 2007). Although we did not explicitly control the distinctiveness of our stimuli, it is likely that the way they were generated would lead to images that lie closer to the centre of Fig. 1 being closer to an average face (more 'typical' in appearance) and those falling toward the periphery of Fig. 1 being more distinctive. To check whether this was the case, we asked a separate group of 10 participants to rate the images included in the matrix along the distinctiveness-typicality dimension. These ratings confirmed that perceived face distinctiveness increased moving from the centre to the edges of the matrix along both the dimensions. Indeed, rated distinctiveness was highly correlated with the U-shaped regressor for both the trustworthiness ($r=0.88$, $p=.001$) and the gender dimensions ($r=0.92$, $p<.001$).

Previous studies have interpreted non-linear responses to trustworthiness in the amygdala in terms of detecting and evaluating socially salient stimuli that are relevant for guiding approach and avoidance behaviour (Sander, Jordan, & Zalla 2003; Todorov, 2008; Vuilleumier, 2005). The concept of face distinctiveness is not in conflict with the idea that the amygdala is involved in evaluating and directing attention toward relevant stimuli. Instead, it suggests that the approach/avoidance system is not in itself sufficient to explain how multiple facial cues are processed by the brain, whereas the distance from an average face in terms of distinctiveness could be a simple and efficient property for highlighting stimuli that require additional evaluation (Said et al., 2010). Our results add support to this view. In particular, we found a common response in the amygdala and posterior face-selective regions to orthogonal

dimensions with different social content, suggesting that all these areas are involved in coding face stimuli in terms of their distinctiveness as well as the social cues conveyed by facial features. Nonetheless, the theoretical explanation of why these regions are sensitive to this feature and the mechanisms underlying face evaluation remain difficult issues. Face distinctiveness could be considered an important cue per se; indeed it is spontaneously encoded from faces and less typical faces are better recognised (Santos and Young, 2005; Valentine, 1991). Therefore, our results could be interpreted by considering that faces at the extremes of the stimuli matrix were processed as perceptually salient because of their distinctiveness, independently of their being varied along the trustworthiness or gender dimensions. This could have driven the quadratic response in the amygdala, because of its sensibility to the personal impact of the stimuli (Ewbank, Barnard, Croucher, Ramponi, & Calder, 2009). This hypothesis is in line with the idea of the amygdala as detector of relevant events (Sander et al., 2003) and can account for different effects reported in previous fMRI studies, such as increased amygdala response when participants received increasing reward or punishment in a competitive game (Zalla et al., 2000), or quadratic amygdala activation when socio-biological facial features like self-resemblance and race were varied (Platek & Krill., 2009). On the other hand, amygdala activation is reported to increase linearly when modulated by the intensity of gustative or olfactory stimuli (Anderson et al., 2003; Small et al., 2003), or by the rated intensity of emotional faces (Sato, Yoshikawa, Kochiyama & Matsumara, 2004) and socially relevant concepts (Cunningham, Raye, & Johnson, 2004). Further studies could investigate whether the effects in the posterior face-selective regions are due to a modulatory influence from the amygdala (Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004) or directly depend on the distance of faces from the average face (Loffler et al., 2005).

In summary, our results help clarify how different face-selective brain regions respond to face stimuli in order to code cues that can be socially relevant. In line with the idea of the amygdala as a salient stimuli detector (Sander et al., 2003), we replicated previous findings of quadratic responses to face trustworthiness (Said et al., 2008; Todorov et al., 2008). However, the activation pattern turned out not to be specific for this social dimension. Similar responses were observed in the amygdala and posterior face-selective regions (OFA, FFA, right pSTS) for faces varying along a gender dimension, suggesting that the images may be processed in terms of their distinctiveness from an average face. Future studies could explore this possibility by asking whether the average face against which the images are coded as more or less distinctive is represented by a general average of the faces seen in a population or by the average of the faces presented in a specific context. This should be possible by creating an average face for the experiment that differs from the general population average of faces encountered in daily life.

References

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Neurosciences*, *3*(12), 469–479.

Adolphs, R., Baron-Cohen, S., & Tranel, D. (2002). Impaired recognition of Social Emotion following amygdala damage. *Journal of Cognitive Neuroscience*, *14*, 1264–1274.

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

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*, 196–202.

Andrews, T. J., Davies-Thompson, J., Kingstone, A., & Young, A. W. (2010). Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *The Journal of Neuroscience*, *30*(9), 3544–3552.

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

Bruce, V., & Young, A. (2012). *Face Perception*. Hove, East Sussex: Psychology Press.

Cristinzio, C., Sander, D., & Vuilleumier, P. (2007). Recognition of emotional face expressions and amygdala pathology. *Epileptologie*, *24*, 130–138.

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in processing of attitudes. *Journal of Cognitive Neuroscience*, *16*, 1–13.

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

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

Feinstein, J. S., Adolphs, R., Damasio, A., & Tranel, D. (2010). The human amygdala and the induction and experience of fear. *Current Biology*, *21*, 34–38.

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

Gläsher, J., & Adolphs, R. (2003). Processing of the arousal subliminal and supraliminal emotional stimuli by the human amygdala. *The Journal of Neuroscience*, *23*(32), 10274–10282.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–232.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841.

Lane, R. D., Reiman, E. M., Bradley, M. M., Lang, P. J., Ahern, G. L., Davidson, R. J., et al. (1997). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia*, *35*(11), 1437–1444.

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

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*, 11087–11092.

Platek, S. M., & Krill, A. L. (2009). Self-face resemblance attenuates other-race face effect in the amygdala. *Brain Research*, *1284*, 156–160.

Said, C. P., Baron, S. G., & Todorov, A. (2008). Nonlinear amygdala response to face trustworthiness: contribution of high and low spatial frequency information. *Journal of Cognitive Neuroscience*, *21*(3), 519–528.

Said, C., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and non-social face dimension. *Neuropsychologia*, *48*, 3596–3605.

Said, C. P., Haxby, J. V., & Todorov, A. (2011). Brain system for assessing the affective value of faces. *Philosophical Transaction of the Royal Society*, *366*, 1660–1670.

Sander, D., Jordan, G., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Reviews in the Neuroscience*, *14*, 303–316.

Santos, I. M., & Young, A. W. (2005). Exploring the perception of social characteristics in faces using the isolation effect. *Visual Cognition*, *12*, 213–247.

Santos, I. M., & Young, A. W. (2008a). Effect of inversion and negation on social inferences from faces. *Perception*, *37*, 1061–1078.

Santos, I. M., & Young, A. W. (2008b). Exploring the perception of social characteristics in faces using the isolation effect. *Visual Cognition*, *12*, 213–247.

Sato, W., Yoshikawa, S., Kochiyama, T., & Matsumara, M. (2004). The amygdala processes the emoziona significante of facial expressions: an fMRI investigation using the interaction between expression and face direction. *NeuroImage*, *22*, 1006–1013.

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neuroscience and Behavioural Reviews*, *32*, 811–830.

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., & Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gestation. *Neuron*, *39*, 701–711.

Smith, S. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155.

Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Computer graphics in facial perception research. *IEEE Computer Graphics and Applications*, *21*, 42–50.

Todorov, A. (2008). Evaluating faces on trustworthiness. An extension of system for recognition of emotions signalling approach/avoidance behaviors. *Annals of the New York Academy of Science*, *1124*, 208–224.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. *Scan*, *3*, 119–127.

Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. *Scan*, *3*, 303–312.

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

Vuilleumier, P. (2005). How brain beware: neural mechanisms of emotional attention. *Trends in Cognitive Science*, *9*, 585–594.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience*, *7*, 1271–1278.

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

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

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

Zalla, T., Koechlin, E., Pietrini, P., Basso, G., Aquino, P., Sirigu, A., et al. (2000). Differential amygdala responses to inning and loosing: a functional magnetic resonance imaging study in humans. *European Journal of Neuroscience*, *12*, 1764–1770.