NeuroImage 49 (2010) 1690–1698

Contents lists available at ScienceDirect

NeuroImage

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

Differential neural responses to faces physically similar to the self as a function of their valence

Sara C. Verosky *, Alexander Todorov *

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

ARTICLE INFO

Article history:
Received 18 March 2009
Revised 4 September 2009
Accepted 7 October 2009
Available online 15 October 2009

ABSTRACT

Behavioral studies show that people self-enhance across a number of domains, including self-face recognition. We used functional magnetic resonance imaging (fMRI) to investigate whether response to physical similarity to the self would differ depending on whether the self-face was morphed with a positive (trustworthy) or negative (untrustworthy) novel face. Participants were presented with morphs of their faces (20%, 40%, 50%, 60%, and 80% self) and asked to decide whether the morph looked like them or the other face. Participants were more likely to identify the trustworthy than the untrustworthy morphs as looking like the self. Moreover, there were large differences in brain activation to trustworthy and untrustworthy morphs. As similarity of the untrustworthy morphs to the self decreased, the response in a number of regions, including bilateral posterior superior temporal sulcus/inferior parietal lobule, right inferior frontal gyrus, and bilateral middle/inferior temporal gyrus, increased. In contrast, there was little evidence for changes in activation as a function of the similarity to trustworthy faces. That is, these regions seemed to differentiate between the self and untrustworthy faces to a much greater extent than between the self and trustworthy faces, despite the fact that the task did not demand evaluation of the faces. The findings suggest that comparing the self to others who are viewed as positive versus negative triggers different psychological processes.

The self can be similar to another person in physical appearance or in terms of less tangible internal states—like thoughts, attitudes, or beliefs (e.g. Mitchell et al., 2005, 2006). Recognizing one's own face entails matching an image of the self to an internal representation of the self. Children develop the ability to recognize the self in a mirror around 18 to 24 months of age (Amsterdam, 1972). Great apes, dolphins, elephants, and magpies have also shown evidence of mirror self-recognition (Gallup, 1970; Reiss and Marino, 2001; Plotnik et al., 2006; Prior et al., 2008), prompting debate over exactly what type of self-representation is needed for this behavior to occur (Mitchell, 1993). The goal of the current study was to investigate the role of perceived psychological similarity in self-recognition in humans. Perception of social stimuli is inherently evaluative (Kim and Rosenberg, 1980; Todorov et al., 2006, 2008; Zajonc, 1980), and it is possible that faces that differ in valence but are physically similar to the self are processed differently. Specifically, we were interested in whether regions in the brain that are active when viewing the self-face simply respond to physical similarity to the self or whether the response depends on the evaluation of the person the self is similar to.

People's automatic associations with the self tend to be positive (Greenwald and Banaji, 1995; Hetts et al., 1999). For example, stimuli associated with the self, such as the letters in one's own name, tend to be evaluated more positively than stimuli that are not associated with the self (Koole et al., 2001). Recently, Epley and Whitchurch (2008) demonstrated that these self-enhancement effects extend to recognition of the self-face. These authors morphed photographs of participants' faces with photographs of attractive or unattractive faces. They found that participants thought it was more likely that the attractive morphs were the self than their actual faces or the unattractive morphs. Participants were also faster to select the attractive versions of their faces out of a line-up of distracter faces than their actual faces or the unattractive versions of their faces.

People also evaluate novel faces that resemble the self more positively. For example, DeBruine (2005) found that participants perceived morphs of their faces as more trustworthy than morphs of other faces. Additional studies have also shown that facial resemblance increases trusting behavior in economic games (DeBruine, 2002; Krupp et al., 2008). That is, biases in the perception of trustworthiness resulted in more prosocial behaviors towards people who resembled the self. Studies also show that participants are more likely to vote for political candidates whose faces have been experimentally manipulated to resemble their own faces (Bailenson et al., 2009).

These findings clearly show that faces that resemble the self are evaluated more positively than faces that do not resemble the self. However, the processes involved may be different when the self is similar to positive as compared to negative others. To the extent that the self is evaluated positively, facial morphs with positive others may

* Corresponding authors. Department of Psychology, Princeton University, Princeton, NJ 08540, USA.
E-mail addresses: sverosky@princeton.edu (S.C. Verosky), atodorov@princeton.edu (A. Todorov).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.10.017

be perceived as psychologically more similar to the self than morphs with negative others. This hypothesis is consistent with research on social judgments showing that stimuli can be evaluated differently depending on whether they are seen in the context of similar or dissimilar stimuli (Herr et al., 1983; Stapel et al., 1997).

While a number of studies have investigated neural responses to the self-face (Platek et al., 2004, 2006; Sugiura et al., 2005, 2008; Uddin et al., 2007), only three studies to date have examined responses to faces varying in physical similarity to the self and in all cases these faces were likely to be evaluated as positive or neutral. In the first such study, Kircher et al. (2001) had participants view morphs of their own faces and novel faces that contained 0% to 30% or 70% to 100% of the self. Face perception is categorical, such that morphs that contain more than two-thirds of a given face are identified as looking like that person (Beale and Keil, 1995; Levin and Beale, 2000). During a behavioral study, Kircher et al. (2001) determined that these categorical boundaries held for the self-face and then during the fMRI study they used this morphing technique to prevent habituation to the self-face. These authors found greater response to the faces containing large as compared to small percentages of the self in a number of regions including the left inferior frontal gyrus (IFG)/dorsolateral prefrontal cortex, the left inferior parietal lobule (IPL), and the left fusiform gyrus.

More recently, Uddin et al. (2005) morphed participants' faces with the faces of gender-matched friends in increments of 20% and asked participants to indicate whether each morph looked more like the self or more like the familiar other. When responses to the three morphs containing the most self were compared to responses to the three morphs containing the least self, there was greater activation in the right IFG, the right inferior and superior parietal lobule, and the inferior occipital gyrus for the self versus the familiar other. Further analysis revealed that these areas responded more strongly as the amount of self present in a face increased. Finally, in a study with a similar design, Uddin et al. (2008) investigated neural responses to morphs of the self and a stranger in children with autism and typically developing children. All of the children showed activation in the right IFG for the morphs containing greater percentages of the self as compared to rest.[1]

The IFG and the IPL showed increased activation for the self as compared to other in both Kircher et al. (2001) and Uddin et al. (2005), but the activation in Kircher et al. (2001) was left lateralized, while the activation in Uddin et al. (2005) was right lateralized. Further, there was little overlap in the other areas that were active in these two studies. It is possible that the different patterns of activation were due to differences in the level of morphing of the stimuli or to differences in the tasks.[2]

The right IFG and the right IPL have also been found to be active in a number of self-face studies that did not use morphing (Platek et al., 2004, 2006; Sugiura et al., 2005, 2008). In addition, a recent study by Morita et al. (2008) indicates that the right IFG is involved in the evaluation of positive and negative photographs of the self. Finally, a study by Uddin et al. (2006) demonstrated that rTMS to the right parietal lobule leads to a slight decrease in sensitivity to the self-face in a self-other discrimination task. Thus, these areas seem to be important for self-face recognition.

These previous studies compared neural responses to the self-face to responses to the faces of familiar others, who are likely to be seen in a positive manner, or to the faces of strangers, who are likely to be seen as neutral. In contrast, the current study investigated neural responses to physical similarity to faces that were likely to be evaluated positively as well as faces that were likely to be evaluated negatively. Work by Todorov and Engell (2008) and Oosterhof and Todorov (2008) demonstrates that trustworthiness judgments provide a good approximation of the valence evaluation of faces, and therefore we morphed participants' faces with faces that were either trustworthy or untrustworthy in appearance. The trustworthy and untrustworthy faces were created using a data-driven model of trustworthiness (Oosterhof and Todorov, 2008), which matched the two categories of faces in terms of physical differences from an average face, thereby controlling against the possibility that one category of faces was more physically similar to participants' own faces.

In the experiment, participants were shown the morphs of their faces and the positively or negatively valenced faces and asked to decide whether each morph looked more like them or more like the other person. The similarity to the self was manipulated by incrementally increasing the percentage of the self-face in the morphs (20% to 80%). As prior behavioral studies have shown (Beale and Keil, 1995; Kircher et al., 2001), facial morphs that contain more than 70% of one face are within the category boundary of this face. Thus, we expected that the extreme morphs would be perceived as the self (80%) and the other person (20%), respectively. Based on the behavioral work showing that the self tends to be seen as positive (Epley and Whitchurch, 2008), we expected that participants would be more likely to perceive the trustworthy than the untrustworthy morphs as looking like the self.

For the neuroimaging data, we were interested in whether activation in areas responsive to physical similarity would be modulated by the trustworthiness of the face that the self was morphed with. Based on previous studies (e.g. Uddin et al., 2005, 2006), we expected that the responses in the right IFG and the right IPL would increase with the increase in similarity of trustworthy faces to the self, but we were not sure whether this would hold true for the untrustworthy faces. If these areas track simple physical similarity to the self, they should respond in a similar way to all morphs, regardless of the valence of the face. However, if comparing the self to positive versus negative faces triggers different processes, these areas might show a different pattern of response for the untrustworthy as compared to trustworthy morphs.

Methods

Participants

Thirty participants (8 men, mean age = 22, SD = 2.91) were recruited from the Princeton University community. All participants were right-handed, had normal or corrected-to-normal vision, and reported no history of neurological or psychiatric disorders. All participants gave informed consent in accordance with the procedure approved by Princeton University's institutional review board. Although participants were recruited without regard to ethnicity, the model of face trustworthiness used to create the novel trustworthy and untrustworthy faces was based on ratings of Caucasian faces and therefore the endpoint faces used in the current study were Caucasian. Eight of the recruited participants were not Caucasian (5 were African-American and 3 were Asian) and preliminary analyses revealed that ethnicity affected both the behavioral and neuroimaging data.[3] Because the non-Caucasian group was non-homogenous and small

[1] This study did not report the contrast between faces containing large percentages of the self and other faces. However, based on the graphs of signal change for the self versus rest and the other face versus rest, it appears that the autistic children showed greater signal change for the self as compared to other in the right IFG, but the typically developing children did not.

[2] It should also be noted that Kircher et al. (2001) had a relatively small sample size of six fMRI participants.

[3] For the behavioral data, the Caucasian participants differentiated between the trustworthy and untrustworthy morphs, but the non-Caucasian participants did not. For the imaging data, the two groups showed substantial overlap in regions that were active for the linear effect of trustworthiness, but less overlap in regions that were active for the main effect of trustworthiness and the quadratic effect of similarity. There are multiple interpretations of these differences. Future work needs to examine these effects using morphing with both ingroup and outgroup faces, as well as larger and homogeneous samples of non-Caucasian participants.

in size, this paper focuses on the data from Caucasian participants only. Finally, the data from one participant were excluded because of excessive head motion and the data from another were excluded because of lack of variance in their behavioral data. Thus, the final sample consisted of 20 participants (7 men, mean age = 22, SD = 3.14).

Stimuli

The trustworthy and untrustworthy faces were created using a computer model of face trustworthiness developed by Oosterhof and Todorov (2008) and implemented in the Facegen Modeller program (http://facegen.com) version 3.1. The face model of Facegen (Blanz and Vetter, 1999; Singular Inversions, 2006) is based on a database of male and female faces that were laser scanned in 3D. Using a principal component analysis, a model was constructed so that each face can be represented by a limited number of independent components. Specifically, in this model each face can be represented as a point in 50D space, and novel faces can be generated as a linear combination of the components. Oosterhof and Todorov (2008) used trustworthiness ratings of novel, emotionally neutral faces to build a model of face trustworthiness. Specifically, using these ratings, they built a new dimension in the 50D space, which was optimal in changing face trustworthiness (see also Todorov et al., 2008). Subsequent behavioral studies validated the model. For example, the mean trustworthiness judgments for −3 and +3 SD faces were 3.99 (SD = 0.74) and 5.95 (SD = 0.62), respectively.

For the current study, we generated 20 random faces for each gender. The race of the faces was set to European because the model of face trustworthiness was based on ratings of Caucasian faces and the attractiveness was set to slightly higher than average to make the morphs more similar to photofitted real faces (cf. Todorov et al., 2008). We then set the trustworthiness for half of the faces to three standard deviations below zero and the trustworthiness for the other half to three standard deviations above zero.

Photographs of each participant were taken using a Canon PowerShot S5 1S digital camera. Participants were asked to remove glasses and to maintain neutral facial expressions while their pictures were being taken. Models of participants' faces were generated in Facegen using the Photofit option on one forward facing and two profile pictures for each person. The model of each participant's face was morphed with all of the novel faces of the same gender using the Tween option in Facegen (Fig. 1A). The model of each participant's face consisted of two separable parts: one representing the shape and the other representing the texture of the face. The Tween option only morphs the shape of two faces, and therefore each participant's texture was overlaid over all of the morphs of their face. Finally, to prevent changes in eye or skin color from providing a marker for similarity between faces, the eye color for all the faces was set to black and luminance of the images was adjusted so that all of the morphs for a given participant had the same average luminance. More specifically, both eye color and skin color provide strong cues to facial identity and in order to prevent participants from simply relying on these single features in their similarity judgments, we kept this information constant across morphs.

There were two levels of face trustworthiness (trustworthy or untrustworthy), 10 different identities for each level, and five levels of morphing (20% self, 40% self, 50% self, 60% self, and 80% self), resulting in 100 morphed faces for each participant.

Experimental design and image acquisition

The experiment consisted of five event-related time series, each lasting 5 min 20 s. An equal number of morphs from each condition were presented in each time series. Each morph was presented twice over the course of the experiment, resulting in a total of 20 trials per condition.

The stimuli were projected onto a screen at the rear of the bore of the magnet and participants viewed the images via an angled mirror placed above the eyes. At the beginning of each trial, participants were shown a small image of their own face, next to a small image of the face that their face was morphed with. The endpoint faces were included in order to constrain participants' choices and thereby make the task less difficult. This probe remained in the center of the screen for 2 s and was immediately followed by a larger morph of the two faces (Fig. 1B). Participants were asked to press a button under their right index finger if the morph looked more like them or to press a button under their right middle finger if it looked more like the person who their face was morphed with. The morph remained on the screen for 2 s and was followed by a jittered inter-trial interval (ITI) of 2, 4, or 6 s. The morphs were presented in a different random order for each participant, while the ITIs were presented in the same fixed random order across all participants.

Echo planar images (EPI) were acquired using a Siemens 3.0 Tesla Allegra head-dedicated scanner (Siemens, Erlangen, Germany) with a Nova Medical NM-011 Head Transit Coil with receive-only array system (Nova Medical, Wilmington, MA). Thirty-three interleaved 3-mm axial slices with an interslice gap of 1 mm were used to achieve full-brain coverage (TR = 2000 ms, TE = 30 ms, flip angle = 90°, matrix size 64 × 64). At the end of each scan session, a high-resolution anatomical image (T1-MPRAGE, TR = 2500 ms, TE = 4.38 ms, flip angle = 8°, matrix size = 256 × 256) was acquired for registering functional activation to the participant's anatomy and for spatial normalization.

fMRI analysis

Preprocessing and statistical analysis of the fMRI data were conducted with Analysis of Functional NeuroImages (AFNI; Cox, 1996). The first four EPI images from each run were discarded to allow the MR signal to reach steady-state equilibrium. Subject motion was corrected using a six-parameter 3-D motion correction algorithm following slice scan-time correction. Transient spikes were removed from the signal using the AFNI program 3dDespike and subsequently the data were low-pass filtered with a frequency cutoff of 0.1 Hz. Finally, the data were spatially smoothed with an 8-mm full-width at half maximum (FWHM) Gaussian kernel and the signal was normalized to percent signal change from the mean.

Polynomial regression was conducted to test for linear and quadratic effects of similarity to the self (Buchel et al., 1998). We used a zero-order regressor, which indicated the presence of a face, to model the mean response to all faces relative to baseline. We used a first-order, linear regressor, which scaled responses according to their distance from the self, to model physical similarity to the self, and a second-order, quadratic regressor, which scaled responses according to their distance from the middle of the morphing continuum, to model task difficulty. The linear and quadratic regressors were centered around zero and orthogonalized to each other.

Additionally, a regressor for the trustworthiness of the faces with which the self-face was morphed was included. This regressor indicated the presence of a trustworthy face to model the main effect of trustworthiness. We also modeled the interaction of face trustworthiness with similarity by including regressors for the interaction of face trustworthiness with the linear and quadratic trends of similarity. Finally, we included a regressor modeling missed responses, which made up less than 5% of trials overall.

These seven regressors were convolved with an ideal hemodynamic response and entered into the general linear model (GLM). In addition, time series representing subject head movement and mean, linear, and quadratic trends in each run caused by scanner drift were included in the model as regressors of non-interest.

Correction for multiple comparisons was performed using the program AlphaSim, which is part of the AFNI package. The Monte

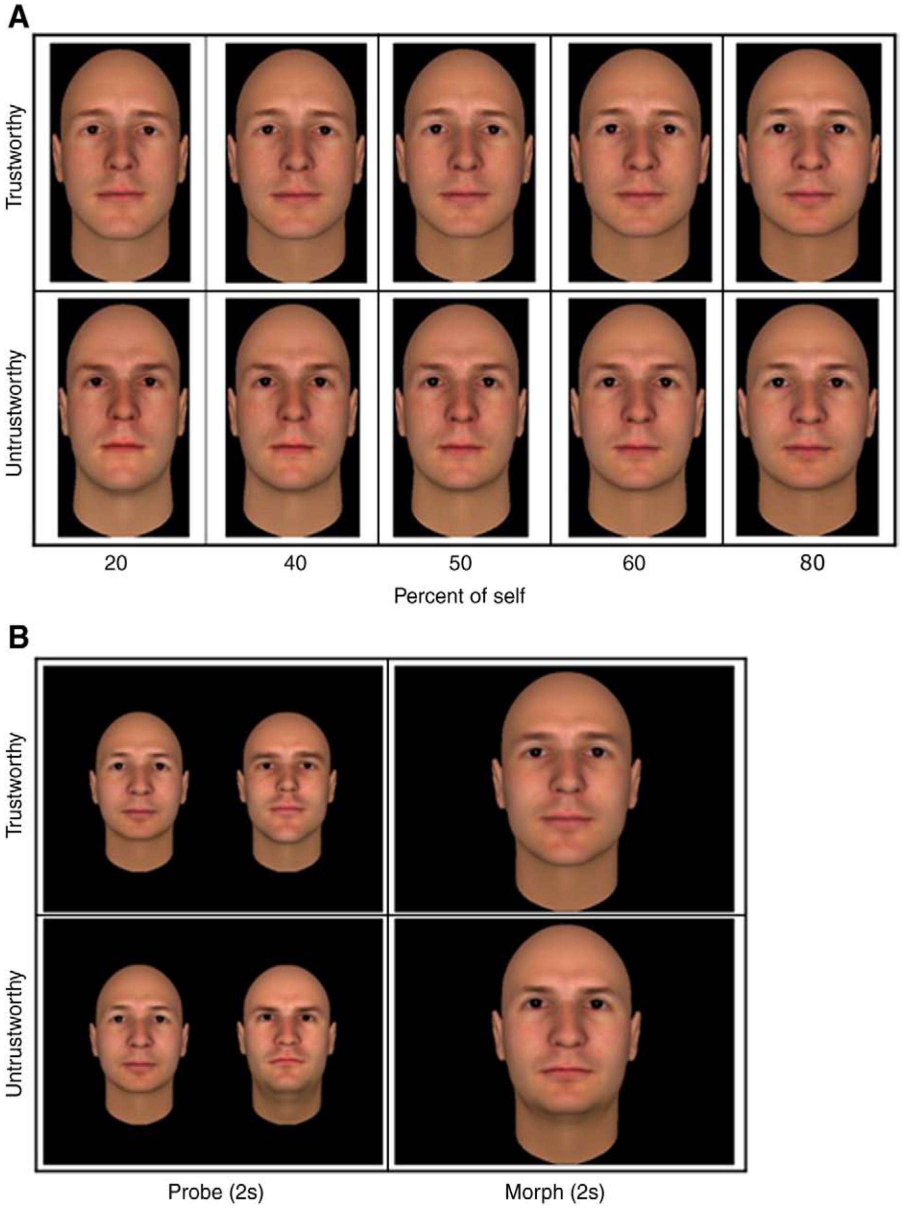

Fig. 1. Example of stimuli used in the experiment. (A) The top row shows a participant's face morphed with a trustworthy face and the bottom row shows a participant's face morphed with an untrustworthy face. (B) The top row shows examples of the probe and morph for trials where a participant's face was morphed with a trustworthy face, while the bottom row shows the probe and morph for trials with an untrustworthy face.

Carlo simulation indicated that a minimum cluster size of 1476 mm^3 was needed to attain a corrected significance of $p<.05$, with a voxelwise threshold of $p<.005$.

A second analysis, where each of the 10 conditions – 2 (face trustworthiness) × 5 (self similarity) – was modeled by a separate regressor, was also conducted. In this analysis, the regressors representing the 10 conditions and a regressor representing missed responses were convolved with the ideal hemodynamic response and entered into the general linear model, along with the regressors of non-interest described above. The second analysis was used to extract percent signal change for each condition in functionally defined regions in the first analysis.

Subsequently, participants' activation maps for both of the analyses were warped to Talairach–Tournoux space. For the first analysis, independent samples *t*-tests were performed on the coefficients supplied by the general linear model for each participant to generate group-level statistical parametric maps showing the significance of the coefficients across participants. A functional mask was created for each area of activation. Percent signal change for each of the 10 conditions was extracted from these ROIs using the coefficients from the spatially normalized individual level activation maps from the second, separate analysis.

Results

Behavioral data

As shown in Fig. 2A, the percentage of identification of the morphs as the self increased monotonically as a function of the percentage of the self-face in the morphs, from near zero for morphs containing 20% of the self to near 100% for morphs containing 80% of the self. Correspondingly, the 2 (face trustworthiness) × 5 (self similarity) repeated measures ANOVA on the participants' responses revealed a large effect of similarity (Greenhouse–Geisser corrected $F(2.61, 49.67) = 292.07$, $p<.0001$). Across all levels of morphing, there was a bias to perceive trustworthy morphs as more similar to the self than

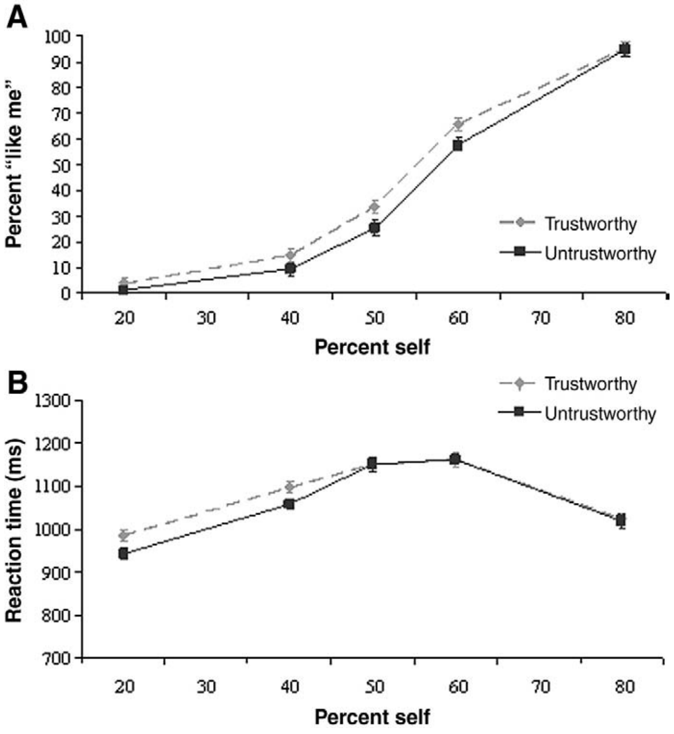

Fig. 2. Behavioral results for similarity judgments and response times ($N = 20$). (A) Percent of the time participants identified the morph as the self as a function of the percent of the self present in the morphs and trustworthiness of the other face. (B) Mean response times for similarity judgments as a function of the percent of the self present in the morphs and trustworthiness of the other face. Error bars represent standard error of the mean.

untrustworthy morphs ($F(1, 19) = 7.19$, $p = .015$). Although the bias was most pronounced for the intermediate levels of morphing, the interaction between trustworthiness and similarity was not significant (Greenhouse–Geisser corrected $F(3, 56.91) = 1.58$, $p = .20$).

An analysis of the response times showed that participants were slower to respond to morphs in the middle of the continuum than to morphs at the extremes of the continuum (Fig. 2B), (Greenhouse–Geisser corrected $F(2.57, 48.91) = 51.02$, $p < .0001$; $F\ (1, 19) = 98.60$, $p < .001$, for the quadratic trend). Participants were also slightly slower to respond to the trustworthy ($M = 1088.39$ ms, SD = 278.91) as compared to untrustworthy morphs overall ($M = 1070.40$, SD = 267.26; $F(1, 19) = 5.46$, $p = .03$). Although the difference in response times to trustworthy versus untrustworthy morphs was greater for morphs containing smaller as compared to larger percentages of the self, the interaction between trustworthiness and similarity was not significant (Greenhouse–Geisser corrected $F(2.91, 55.27) = 1.58$, $p = .21$).

fMRI data

Linear effects of similarity

Both right and left pSTS extending into right and left IPL, the right IFG, the left middle temporal gyrus extending into the left inferior temporal gyrus/left fusiform gyrus, and right inferior temporal gyrus showed a significant negative linear effect of similarity, responding more strongly to increasing percentages of the other person (Table 1).[4] The cluster in the right IFG was not large enough to survive correction for multiple comparisons, but it is included here as an a priori region of interest. No regions showed the opposite effect.

Table 1
Brain regions showing a linear effect of similarity to the self.

Region	Cluster size (mm^3)	*x*	*y*	*z*	*t*-value
Right STS/supramarginal gyrus (IPL)	8856	55.5	−37.5	5.5	−5.66
Left supramarginal gyrus (IPL)/STS	2538	−58.5	−46.5	20.5	−5.51
Left posterior middle temporal/ superior temporal gyrus	2052	−37.5	−58.5	23.5	−4.93
Left middle temporal gyrus, extending into inferior temporal gyrus and fusiform gyrus	1728	−52.5	−16.5	−15.5	−4.86
Right inferior/middle temporal gyrus	1620	67.5	−4.5	−18.5	−4.20
Right inferior frontal gyrus (IFG)	783	52.5	31.5	−6.5	−4.03

Note. Regions are reported at $p = .05$ corrected for multiple comparisons, with voxelwise $p = .005$. The *t*-values correspond to the voxels with maximum activation and the coordinates for these voxels are reported in Talairach space. The activation in the IFG did not survive correction for multiple comparisons, but it is listed because it was an a priori region of interest.

As described below, the linear response in these regions was primarily driven by the responses to the untrustworthy morphs. Although no areas that were active for the interaction of similarity and trustworthiness survived the correction for multiple comparisons, the GLM analysis showed that many of the voxels that were active for the linear effect of similarity were also active for the interaction of the linear effect and trustworthiness (Fig. 3A).

To better understand the pattern of response, percent signal change was extracted from each area that showed a significant linear response (Table 1) and submitted to a 2 (face trustworthiness) × 5 (self similarity) repeated measures ANOVA. All of the regions showed an interaction between the linear effect of similarity and trustworthiness ($p < .05$ in all cases). For the untrustworthy faces, the regions responded more strongly as the morphs contained increasing percentages of the other person ($p < .005$ in all cases). For the trustworthy faces, most of the areas did not show a change in responses as a function of similarity (Figs. 3B–D). However, the right IFG showed a slight increase in signal change as the morphs looked more like the self, although this positive linear trend was not significant ($F(1, 19) = 2.14$, $p = .16$; Fig. 3E).

Quadratic effects of similarity

Three areas showed a quadratic response to similarity to the self (Table 2; Fig. 4A). The dorsal anterior cingulate cortex, the left anterior insula extending well into the left prefrontal cortex, and the right middle frontal gyrus responded more strongly to morphs in the middle of the similarity continuum as compared to those at either end. When signal change data from these regions were analyzed, none of them showed a significant interaction between trustworthiness and similarity.

Finally, a region in the left thalamus that was active for the interaction of the quadratic effect of similarity and trustworthiness survived the correction for multiple comparisons (Table 2). This area showed a quadratic effect of similarity for the trustworthy faces ($F(1, 20) = 23.86$, $p < .001$), responding more strongly to the faces in the middle of the continuum, but not for the untrustworthy faces ($F < 1$).

Discussion

In this study, participants were presented with morphs of their faces with trustworthy and untrustworthy faces and asked to decide whether each morph looked more like them or more like the other face. Not surprisingly, as similarity to the self increased, participants were more likely to identify the morphs as the self than as the other person (Fig. 2A). Consistent with prior studies on categorical perception of faces (Beale and Keil, 1995; Kircher et al., 2001; Levin and Beale, 2000), morphs containing 20% of the self were almost always identified as the other face and morphs containing 80% of the

[4] None of the regions that showed a main effect of trustworthiness survived the correction for multiple comparison.

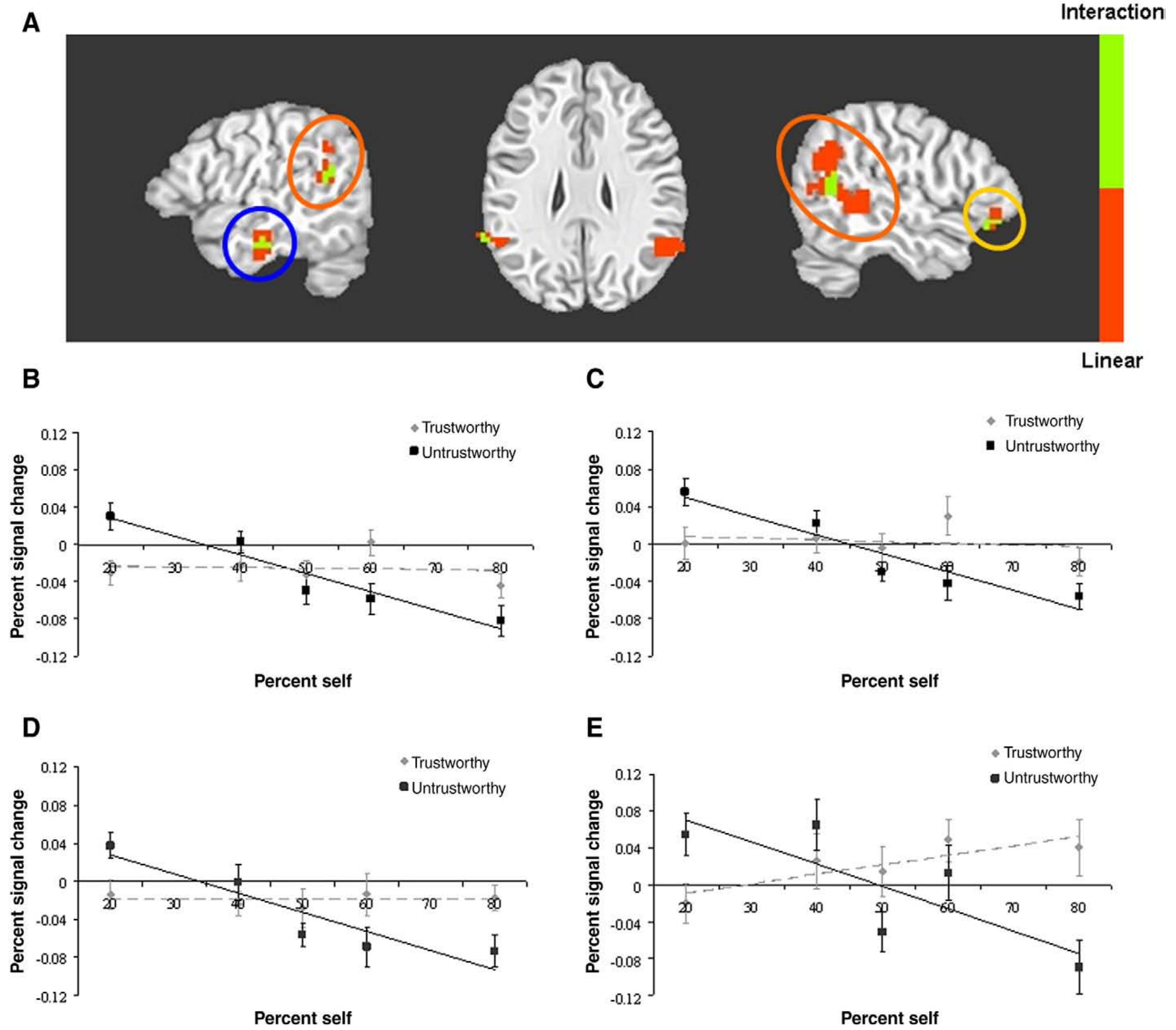

Fig. 3. Brain regions showing a significant linear effect of similarity to the self ($p = .05$, corrected for multiple comparisons). (A) Sagittal ($x = -51$ for the slice on the left, $x = 51$ for the slice on the right) and axial ($z = 28$) slices showing areas that were active for the linear effect of similarity to the self. Voxels that were active for both the linear effect and the interaction with face trustworthiness are shown in green, while those that were only active for the linear effect are shown in orange. (B) Percent signal change by condition from all active voxels in the left pSTS/IPL region (circled in orange on the left sagittal slice in panel A). Error bars in this graph and the following graphs represent standard error of the mean. (C) Percent signal change by condition from all active voxels in the right pSTS/IPL region (circled in orange on the right sagittal slice in panel A). (D) Percent signal change from all active voxels in the left middle/inferior temporal gyrus extending into the left fusiform gyrus (circled in blue on the right sagittal slice in panel A). Error bars represent standard error of the mean. (E) Percent signal change from all active voxels in the right IFG (circled in yellow on the right sagittal in panel A). Error bars represent standard error of the mean.

self were almost always identified as the self. More importantly, participants were more likely to identify the trustworthy than the untrustworthy morphs as looking like the self, even though the trustworthy and untrustworthy faces were matched in terms of physical characteristics. This finding is consistent with previous behavioral work demonstrating that people recognize their own faces as more attractive than they actually are (Epley and Whitchurch, 2008). Together these findings indicate that the tendency to see the self as positive extends to judgments of physical appearance. Judgments of face trustworthiness and face attractiveness are highly correlated and therefore it is not surprising that this positivity bias affects ratings of the self on both dimensions. Indeed it is likely that the positivity bias will generalize to other valence-related dimensions as well (cf. Todorov and Engell, 2008).

In addition to affecting the behavioral responses, the valence of the faces affected the neural responses in regions that tracked physical similarity to the self. Regions in the right and left pSTS extending into IPL, the right IFG, the right inferior/middle temporal gyrus, and the left middle temporal gyrus extending into inferior temporal and fusiform gyri showed a linear effect of similarity to the self that was modulated by the trustworthiness of the other face. As similarity of the untrustworthy morphs to the self decreased, the response in these regions increased (Fig. 3). In contrast, there was little evidence for changes in activation as a function of the similarity to trustworthy faces. In other words, these regions seemed to differentiate between the self and untrustworthy faces to a much greater extent than between the self and trustworthy faces. Thus, even though the task did not demand explicit evaluation of the faces, comparing the self to

Table 2
Brain regions showing quadratic effects of similarity to the self.

Region	Cluster size (mm^3)	*x*	*y*	*z*	*t*-value
Quadratic effect					
Left cingulate/superior frontal gyrus	9855	−10.5	16.5	32.5	6.33
Left anterior insula, extending into inferior/middle frontal gyrus	9315	−34.5	7.5	5.5	5.10
Right middle frontal gyrus	1755	43.5	16.5	26.5	4.64
Interaction between the quadratic effect of similarity and face trustworthiness					
Left thalamus, extending from the left mammillary body to the left pulvinar	1701	−10.5	−16.5	−0.5	5.65

Note. Regions are reported at $p = .05$ corrected for multiple comparisons, with voxelwise $p = .005$. The *t*-values correspond to the voxels with maximum activation and the coordinates for these voxels are reported in Talairach space.

untrustworthy versus trustworthy others led to qualitatively different patterns of response in these regions.

Consistent with previous work that has demonstrated that the IFG and IPL are involved in recognition of the self-face (Kircher et al., 2001; Uddin et al. 2005), we found that these regions were sensitive to physical similarity to the self. However, while previous studies found increased response to physical similarity in these regions, our strongest finding was the decreased response to physical similarity for untrustworthy morphs. There was a trend in the opposite direction for trustworthy morphs in the right IFG, but this trend did not reach significance. Our findings suggest that these regions do not simply respond to physical similarity to the self. Rather, their response most likely depends on both the type of task (e.g. comparing oneself to another person) and the evaluation of the self-resembling face.

Recent work by Morita et al. (2008) indicates that the right IFG is involved in self-evaluation and potentially sheds light on the current pattern of results. These authors showed participants flattering and unflattering photographs of the self and of other people. The right IFG was more active when viewing the self as compared to others and activation in this area was negatively correlated with ratings of embarrassment. In addition, Uddin et al. (2005) found that viewing morphs of the self and a familiar other that contain large percentages of the self led to increased activation in this region. In other words, whereas distancing oneself from a negative image was related to decreased activation in the right IFG, endorsing a positive or neutral morph as looking like the self was related to increased activation. In line with these previous two studies, we found that viewing untrustworthy morphs that looked like the self led to decreased activation in this region, while viewing trustworthy morphs that looked like the self led to increased activation, although the latter did not reach significance. Thus, taken together, these three studies suggest that this region may play not only a role in representations of the self with respect to appearance but also in evaluation of these representations.

In sum, the brain imaging data show that positive and negative morphs of the self are processed in a qualitatively different fashion. This is consistent with psychological models that suggest that comparison of targets to standards that are perceived as similar leads to an informational focus on similarities, while comparison to standards that are perceived as different leads to a focus on differences (Mussweiler, 2003). Given that the self is usually seen as positive (Greenwald and Banaji, 1995; Hetts et al., 1999; Koole et al., 2001), comparison of the self-face to trustworthy faces may have led to different expectations and a focus on similarity, whereas comparison of the self-face to untrustworthy faces may have led to a focus on dissimilarity. Such processes might be able to account for the greater neural differentiation between the self and untrustworthy faces than between the self and trustworthy faces. Future studies need to

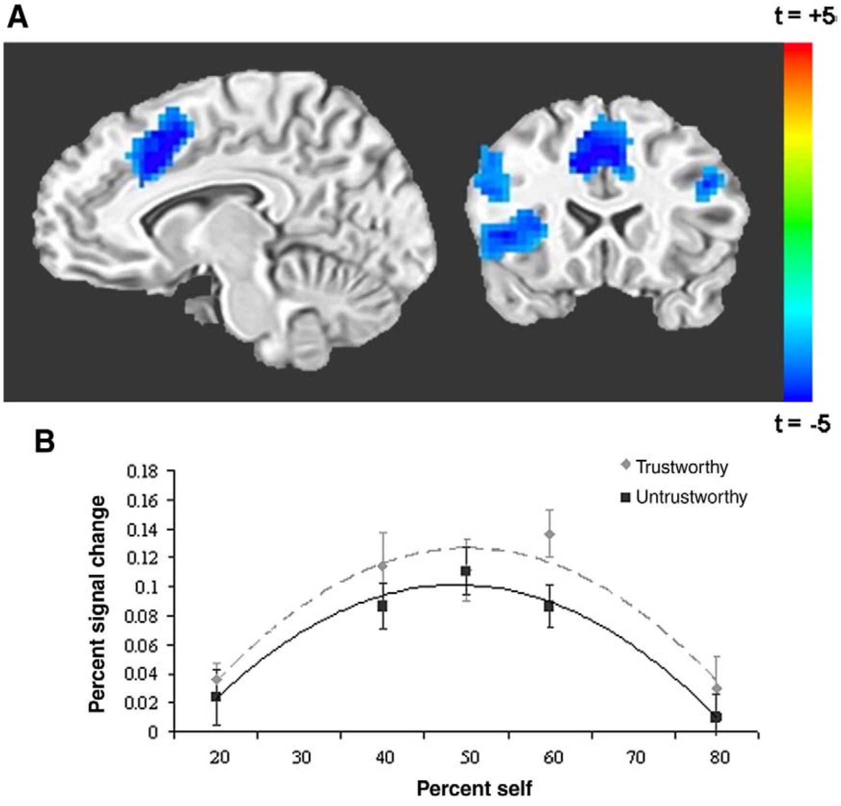

Fig. 4. Brain regions showing a significant quadratic effect of similarity to the self ($p = .05$, corrected for multiple comparisons). (A) Sagittal ($x = -8$) and coronal ($y = 18$) slices showing regions that were active for the quadratic effect of similarity to the self. (B) Percent signal change from the voxels in the dorsal ACC that were active for the quadratic effect of similarity to the self (shown in the sagittal slice in panel A). Error bars represent standard error of the mean.

address the exact nature of the psychological mechanisms producing these effects.

In contrast to the regions that were active for the linear effect of similarity, the regions that were active for the quadratic effect – except for the thalamus – showed the same pattern of response for trustworthy and untrustworthy morphs. Specifically, regions in the prefrontal cortex and the dorsal anterior cingulate cortex (ACC) responded more strongly to morphs in the middle of the similarity continuum than to those at the ends. When the morphs contained large percentages of either the self or the other person, participants' decision as to whether the morph looked more like the self or more like the other person was straightforward. However, when the morphs contained nearly equivalent percentages of the self and other, the decision was more difficult. This was reflected in the pattern of reaction times, with participants being slower to respond to morphs in the middle as compared to the ends of the continuum (Fig. 2B). The regions that were active for the quadratic response showed this same pattern and it is likely that they tracked task difficulty (Botvinick et al., 2004; Miller and Cohen, 2001).

For the reaction times, participants were slightly slower to respond to the trustworthy than to untrustworthy morphs overall, suggesting that the decision may have been slightly more difficult for the trustworthy faces. Interestingly, the areas that showed a quadratic response typically responded slightly more strongly to the trustworthy than to untrustworthy morphs, although these effects were not statistically significant (Fig. 4B). In addition, a region in the left thalamus, extending from the left mammillary body to the left pulvinar nucleus, showed an exaggerated version of this pattern. This area showed a quadratic response to trustworthy faces – once again responding more strongly to morphs in the middle of the continuum – but not for untrustworthy faces. Although the peak of the activation was in the mammillary body, it is worth noting that the pulvinar is involved in visual attention (Saalmann and Kastner, 2009) and that it may have modulated responses in the other active regions.

Self-resemblance as a proxy for kinship

From an evolutionary perspective, self-resemblance is important because it provides a cue to genetic relatedness (DeBruine et al., 2008). In fact, as mentioned in the introduction, research has shown that self-resemblance increases trustworthiness ratings of faces and cooperative behavior (DeBruine, 2002, 2005; Krupp et al., 2008). In addition, neuroimaging work has started exploring the neural basis of kin detection based on facial similarity, with studies finding overlapping regions of activation for the self-face and self-resembling faces (Platek et al., 2005, 2008, 2009). The implication of these studies is that similarity automatically induces positive evaluation. However, the current study suggests that the context can have a large effect on how self-resembling faces are evaluated (see also DeBruine, 2005; Platek, 2007). This is not inconsistent with the idea that similar neural mechanisms underlie positive evaluation of self-resembling others and positive evaluation of the self (cf. Platek et al, 2009). However, different types of tasks may lead to different expectations, which in turn can lead to different neural responses to physical similarity to the self.

Different types of similarity to the self

We focused on one particular type of similarity with the self, namely facial resemblance, but there are many other types of similarity that can affect both perception of the self and other people. For example, Mitchell et al. (2006) constructed descriptions of other people that either matched the values and preferences of participants or did not, thereby increasing or decreasing the psychological similarity with the self. Then, participants were asked to judge how likely they and the similar and dissimilar others would be to agree with different opinions (e.g. “enjoy having a roommate from a different country”). Mitchell et al. found that judgments about the self and about the similar other recruited a ventral region of the MPFC, while judgments about the dissimilar other recruited a dorsal region of the MPFC. Based on prior findings suggesting that self relevant information is processed in ventral MPFC (Kelley et al., 2002; Mitchell et al., 2005), they have argued that judgments of similar others recruit the same regions that process self-referential information.

The current findings, coupled with Mitchell et al. (2006) findings, suggest a possible dissociation between regions that track visual similarity with the self (right IFG) and regions that track personality similarity (ventral MPFC). Additional findings from Mitchell et al. (2005) suggest that the key factor that can produce this dissociation in these areas is the focus of the task (e.g. physical similarity vs. mental states) rather than how similarity between the self and others was determined. Specifically, Mitchell et al. (2005) showed that ventral MPFC was more strongly recruited when making mental state judgments about similar as compared to dissimilar faces, as determined by the participants' ratings of similarity. Indeed, in a review of the relevant literature, Lieberman (2007) noted that lateral prefrontal and lateral parietal areas are generally recruited during tasks that require processing of externally focused information about the self, for instance in detecting mismatches between vision and proprioceptive feedback, while the MPFC is recruited when processing internally focused information about the self, for instance reflecting on one's current experience. Similarly, Uddin et al. (2007) point out that lateral prefrontal and lateral parietal areas overlap with areas that are part of the mirror neuron system, which is thought to be involved in understanding the physical actions of others, while the areas recruited in self-referential thought overlap with areas of the MPFC, which are involved in understanding psychological aspects of others.

Conclusions

Perception of social stimuli is inherently evaluative (Todorov et al., 2006, 2008) and this is particularly true when these stimuli are related to the self (Epley and Whitchurch, 2008; Greenwald and Banaji, 1995). We observed that when the self was compared to an untrustworthy face, most brain regions showed increased response with the decrease in similarity of the morphs to the self. In contrast, when the self was compared to a trustworthy face, the same brain regions did not show a change in their response. That is, these regions differentiated between the self and untrustworthy faces to a much greater extent than between the self and trustworthy faces, consistent with people's inherent biases to perceive the self as positive.

Acknowledgments

We would like to thank Nick Oosterhof for his help with the creating of the stimuli for the study, and Chris Said, Sean Baron, Nick Oosterhof, and Andy Engell for their help with the data collection.

This research was supported by NSF grant BCS 0823749.

References

Amsterdam, B., 1972. Mirror self-image reactions before age two. Dev. Psychobiol. 5, 297–305.

Bailenson, J.N., Iyengar, S., Yee, N., Collins, N.A., 2009. Facial similarity between voters and candidates causes influence. Public Opin. Quart. 72, 935–961.

Beale, J.M., Keil, F.C., 1995. Categorical effects in the perception of faces. Cognition 57, 217–239.

Blanz, V., Vetter, T., 1999. A morphable model for the synthesis of 3D faces. Proceedings of the 26th annual conference on computer graphics and interaction techniques. Los Angeles, Addison Wesley Longman, pp. 187–194.

Botvinick, M.M., Cohen, J.D., Carter, C.S., 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn. Sci. 8, 539–546.

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

Cox, R., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comp. Biomed. Res. 29, 162–173.

DeBruine, L.M., 2002. Facial resemblance enhances trust. Proc. Royal Soc. London B 269, 1307–1312.

DeBruine, L.M., 2005. Trustworthy but not lust-worthy: context-specific effects of facial resemblance. Proc. Royal Soc. London B 272, 919–922.

DeBruine, L.M., Jones, B.C., Little, A.C., Perrett, D.I., 2008. Social perception of facial resemblance in humans. Arch. Sex. Behav. 37, 64–77.

Epley, N., Whitchurch, E., 2008. Mirror, mirror, on the wall: self-enhancement in self-recognition. Pers. Soc. Psychol. Bull. 34, 1159–1170.

Gallup, G.G., 1970. Chimpanzees: self-recognition. Science 167, 86–87.

Greenwald, A.G., Banaji, M.R., 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. Psychol. Rev. 102, 4–27.

Herr, P.M., Sherman, S.J., Fazio, R.H., 1983. On the consequences of priming: assimilation and contrasts effects. J. Exp. Social Psychol. 19, 323–340.

Hetts, J.J., Sakuma, M., Pelham, B.W., 1999. Two roads to positive regard: implicit and explicit self-evaluation and culture. J. Exp. Social Psychol. 35, 512–559.

Kelley, W.M., Macrae, C.N., Wyland, C.L., Caglar, S., Inati, S., Heatherton, T.F., 2002. Finding the self? An event-related fMRI study. J. Cogn. Neurosci. 14, 785–794.

Kim, M.P., Rosenberg, S., 1980. Comparison of two structural models of implicit personality theory. J. Pers. Soc. Psychol. 38, 375–389.

Kircher, T.T.J., Senior, C., Phillips, M.L., Rabe-Hesketh, S., Benson, P.J., Bullmore, E.T., Brammer, M., Simmons, A., Bartels, M., David, A.S., 2001. Recognizing one's own face. Cognition 78, B1–B15.

Koole, S.L., Dijksterhuis, A., van Knippenberg, A., 2001. What's in a name: implicit self-esteem. J. Pers. Soc. Psychol. 80, 614–627.

Krupp, D.B., DeBruine, L.M., Barclay, P., 2008. A cue of kinship promotes cooperation for the public good. Evol. Hum. Behav. 29, 49–55.

Levin, D.L., Beale, J.M., 2000. Categorical perception occurs in newly learned faces, other-race faces and inverted faces. Perception and Psychophysics 62, 386–401.

Lieberman, M.D., 2007. Social cognitive neuroscience: a review of core processes. Annu. Rev. Psychol. 58, 259–289.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.

Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005. The link between social cognition and self-referential thought in the medial prefrontal cortex. J. Cogn. Neurosci. 17, 1306–1315.

Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50, 655–663.

Mitchell, R.W., 1993. Mental models of mirror-self-recognition: two theories. New Ideas Psychol. 11, 295–325.

Morita, T., Itakura, S., Saito, D.N., Nakashita, S., Harada, T., Kochiyama, T., Sadato, N., 2008. The role of the right prefrontal cortex in self-evaluation of the face: a functional magnetic resonance imaging study. J. Cogn. Neurosci. 20, 342–355.

Mussweiler, T., 2003. Comparison processes in social judgment: mechanisms and consequences. Psychol. Rev. 110, 472–489.

Oosterhof, N.N., Todorov, A., 2008. The functional basis of face evaluation. Proc. Nat. Acad. Sci. U.S.A. 105, 11087–11092.

Platek, S.M., 2007. Facial resemblance exaggerates sex-specific jealousy-based decisions. Evol. Psychol. 5, 223–231.

Platek, S.M., Keenan, J.P., Gallup, G.G., Mohamed, F.B., 2004. Where am I? The neurological correlates of self and other. Cogn. Brain Res. 19, 114–122.

Platek, S.M., Keenan, J.P., Mohamed, F.B., 2005. Sex differences in the neural correlates of child facial resemblance: an event-related fMRI study. NeuroImage 25, 1336–1344.

Platek, S.M., Loughead, J.W., Gur, R.C., Busch, S., Ruparel, K., Phend, N., Panyavin, I.S., Langleben, D.D., 2006. Neural substrates for functionally discriminating self-face from personally familiar faces. Hum. Brain Mapp. 27, 91–98.

Platek, S.M., Krill, A.L., Kemp, S.M., 2008. The neural basis of facial resemblance. Neurosci. Lett. 437, 76–81.

Platek, S.M., Krill, A.L., Wilson, B., 2009. Implicit trustworthiness ratings of self-resembling faces activate brain centers involved in reward. Neuropsychologia 47, 289–293.

Plotnik, J.M., de Waal, F.B.M., Reiss, D., 2006. Self-recognition in an Asian elephant. Proc. Nat. Acad. Sci. U.S.A. 103, 17053–17057.

Prior, H., Schwarz, A., Gunturkun, O., 2008. Mirror-induced behavior in the Magpie (*Pica pica*): evidence of self-recognition. PLoS Biol. 6, e202. doi:10.1371/journal.pbio.0060202.

Reiss, D., Marino, L., 2001. Mirror self-recognition in the bottlenose dolphin: a case of cognitive convergence. Proc. Nat. Acad. Sci. U.S.A. 98, 5937–5942.

Saalmann, Y.B., Kastner, S., 2009. Gain control in the visual thalamus during perception and cognition. Curr. Opin. Neurobiol. 19, 408–414.

Singular Inversions, 2006. FaceGen 3.1 Full Software Development Kit Documentation. Retrieved June 5, 2007, from http://www.facegen.com.

Stapel, D.A., Koomen, W., van der Plight, J., 1997. Categories of category accessibility: the impact of trait concept versus exemplar priming on person judgments. J. Exp. Social Psychol. 33, 47–76.

Sugiura, M., Watanabe, J., Maeda, Y., Matsue, Y., Fukuda, H., Kawashima, R., 2005. Cortical mechanisms of visual self-recognition. NeuroImage 24, 143–149.

Sugiura, M., Sassa, Y., Jeong, H., Horie, K., Sato, S., Kawashima, R., 2008. Face-specific and domain general characteristics of cortical responses during self-recognition. NeuroImage 42, 414–422.

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

Todorov, A., Harris, L.T., Fiske, S.T., 2006. Toward socially inspired social neuroscience. Brain Res. 1079, 76–85.

Todorov, A., Baron, S., Oosterhof, N.N., 2008. Evaluating face trustworthiness: a model based approach. Social, Cognitive, & Affective Neuroscience 3, 119–127.

Todorov, A., Said, C.P., Engell, A.D., Oosterhof, N.N., 2008. Understanding evaluation of faces on social dimensions. Trends Cogn. Sci. 12, 455–460.

Uddin, L.Q., Kaplan, J.T., Molnar-Szakacs, I., Zaidel, E., Iacoboni, M., 2005. Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: an event-related fMRI study. NeuroImage 25, 926–935.

Uddin, L.Q., Molnar-Szakacs, T., Zaidel, E., Iacoboni, M., 2006. rTMS to the right inferior parietal lobule disrupts self-other discrimination. Social Cognitive & Affective Neuroscience 1, 65–71.

Uddin, L.Q., Iacoboni, M., Lange, C., Keenan, J.P., 2007. The self and social cognition : the role of cortical midline structures and mirror neurons. Trends Cogn. Sci. 11, 153–157.

Uddin, L.Q., Davies, M.S., Scott, A.A., Zaidel, E., Bookheimer, S.Y., Iacoboni, M., Dapretto, M., 2008. Neural basis of self and other representation in autism: an fMRI study of self-face recognition. PLoS ONE 3, e3526. doi:10.1371/journal.pone.0003526.

Zajonc, R.B., 1980. Feeling and thinking: preferences need no inferences. Am. Psychol. 35, 151–175.