

Brain systems for assessing the affective value of faces

Christopher P. Said, James V. Haxby and Alexander Todorov

Phil. Trans. R. Soc. B 2011 **366**, 1660-1670

doi: 10.1098/rstb.2010.0351

References

[This article cites 70 articles, 18 of which can be accessed free](#)

<http://rstb.royalsocietypublishing.org/content/366/1571/1660.full.html#ref-list-1>

[Article cited in:](#)

<http://rstb.royalsocietypublishing.org/content/366/1571/1660.full.html#related-urls>

Subject collections

Articles on similar topics can be found in the following collections

[cognition](#) (237 articles)

[neuroscience](#) (296 articles)

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

Review

Brain systems for assessing the affective value of faces

Christopher P. Said^{1,*}, James V. Haxby² and Alexander Todorov³

¹*Computational Neuroimaging Laboratory, Center for Neural Science and Department of Psychology, New York University, New York, NY, USA*

²*Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755, USA*

³*Psychology Department, Princeton University, Princeton, NJ 08540, USA*

Cognitive neuroscience research on facial expression recognition and face evaluation has proliferated over the past 15 years. Nevertheless, large questions remain unanswered. In this overview, we discuss the current understanding in the field, and describe what is known and what remains unknown. In §2, we describe three types of behavioural evidence that the perception of traits in neutral faces is related to the perception of facial expressions, and may rely on the same mechanisms. In §3, we discuss cortical systems for the perception of facial expressions, and argue for a partial segregation of function in the superior temporal sulcus and the fusiform gyrus. In §4, we describe the current understanding of how the brain responds to emotionally neutral faces. To resolve some of the inconsistencies in the literature, we perform a large group analysis across three different studies, and argue that one parsimonious explanation of prior findings is that faces are coded in terms of their typicality. In §5, we discuss how these two lines of research—perception of emotional expressions and face evaluation—could be integrated into a common, cognitive neuroscience framework.

Keywords: face; social cognition; cognitive neuroscience

1. INTRODUCTION

To successfully navigate the social world, people need to be able to accurately infer the mental and emotional states of other people. One of the richest sources of such inferences is the face. In addition to conveying information about person identity, age and gender, faces convey information about momentary mental states and, consequently, potential intentions of others. These states provide relevant information to the perceiver about appropriate behaviours. In this paper, we focus both on inferences about emotions (e.g. anger) and inferences about social personality traits (e.g. aggressiveness). Traditionally, the perceptions of emotional expressions and personality traits in faces have been studied separately. However, as we document in §2, there is good empirical evidence that these two types of perceptions may rely on similar mechanisms. Both types of perceptions are inherently imbued with affect and, most likely, are in the service of inferring intentions [1].

In this review, we describe the brain systems for recognizing facial expressions and for the evaluation of neutral faces, focusing on both the existing state of knowledge and the many questions that remain

open, particularly about mechanism. In §2, we review behavioural evidence about the relationships between the perception of facial expressions and the perception of traits in neutral faces. In §3, we discuss the neural systems underlying the perception of emotional expressions. Specifically, we discuss new evidence for a distributed network of regions processing facial expressions, and how it may be partially segregated from networks processing facial identity. We then describe how an existing computational model of bodily motion perception can be applied to dynamic facial expressions. In §4, we describe the existing knowledge of the cognitive neuroscience of face evaluation. In addition to reviewing prior studies, we report the results of an analysis across several functional magnetic resonance imaging (fMRI) studies. We then provide an account of how face typicality can account for previous inconsistencies in the literature [2]. In §5, we make suggestions about where future research on the topics could be directed and how brain research on perception of emotional expressions and perception of traits in faces could be integrated.

2. THE COMMON ORIGIN OF PERCEPTION OF EMOTIONAL EXPRESSIONS AND TRAITS IN FACES

Faces with neutral expressions are socially relevant, as they are believed by many human observers to carry

* Author for correspondence (chris.said@nyu.edu).

One contribution of 10 to a Theme Issue ‘Face perception: social, neuropsychological and comparative perspectives’.

information about personality traits. Although the accuracy of this information is low [3], people make rapid and reliable judgements about the trustworthiness, competence, aggressiveness and many other traits in neutral faces [4,5]. Judgements of these personality traits affect important real-world outcomes, such as election decisions and job promotions [6–8]. Together with the judgements of facial attractiveness, we refer to the process of making these judgements as ‘face evaluation’.

Although the processes underlying perceptions of facial expressions and face evaluation are obviously not identical, there is a growing body of evidence that the perception of traits in neutral faces is related to the perception of facial expressions, and may rely on the same mechanisms. We review three types of evidence: (i) emotion and trait judgements from neutral faces are highly correlated and have similar dimensional structures; (ii) computer vision techniques confirm that most face traits contain structural similarities to emotional expressions; and (iii) behavioural adaptation studies show that adapting to emotional expressions affects trait judgements from faces. These findings are accommodated by the emotion overgeneralization hypothesis, according to which certain properties of neutral faces may bear slight resemblances to emotional expressions [9]. When observers are confronted with these neutral faces, their emotional expression recognition system overgeneralizes and misattributes traits to the faces [10,11].

In an early test of the idea that trait judgements are related to perceptions of emotional expressions, Knutson [12] showed that trait judgements of faces expressing basic emotions were affected by the specific emotions. For example, angry faces were perceived as more dominant. More interestingly, emotion and trait judgements are highly correlated even in neutral faces [10], as predicted by the emotion overgeneralization hypothesis.

Here, we demonstrate this effect with correlations between judgements of the basic emotions and judgements on an empirically derived set of traits shown to be important for spontaneous characterization of neutral faces. We start with data from previous studies, in which participants were asked to rate 66 neutral faces on 14 trait dimensions [13]. A separate group of 30 participants who had not previously seen the faces was asked to rate them on the six basic emotions: anger, happiness, disgust, surprise, fear and sadness. Because the faces were neutral, participants were asked to detect ‘subtle’ emotions, using a 9-point scale ranging from 1 (not at all) to 9 (moderately).

As shown in figure 1*a*, emotion judgements were highly correlated with trait judgements. Out of the 84, 55 trait-emotion correlations were significant. For example, judgements of happiness were positively correlated with all positive trait judgements and negatively correlated with all negative judgements ($ps < 0.05$). That is, faces that were evaluated positively were perceived as happier. Although the correlations were weaker, the pattern of correlations was similar for judgements of surprise. Judgements of anger were positively correlated with negative judgements and negatively correlated with positive judgements. The

pattern of correlations was the same for judgements of disgust. Judgements of fear were negatively correlated with judgements of emotional stability and confidence, and positively correlated with judgements of unhappiness ($ps < 0.05$). Judgements of sadness were negatively correlated with judgements of emotional stability, sociability, intelligence, confidence and dominance, and positively correlated with judgements of unhappiness ($ps < 0.05$).

Additionally, a principal components analysis (PCA) showed that the two types of judgements have similar dimensional structure (figure 2). The first principal component (PC) for trait judgements accounted for 64 per cent of the variance and contrasted positive (e.g. trustworthiness) with negative judgements (e.g. weirdness). The first PC for emotion judgements accounted for 51 per cent of the variance and contrasted judgements of positive emotions with judgements of negative emotions. The correlation between these two components derived from different sets of judgements was 0.60. The second PC for trait judgements—with highest loadings for dominance, confidence and aggressiveness—accounted for 18.2 per cent of the variance. The second PC for emotion judgements accounted for 29.3 per cent of the variance and contrasted judgements of anger with judgements of fear, surprise and sadness. The correlation between these two components was 0.58.

The strong relations between emotion and trait judgements are consistent with the hypothesis that evaluation of faces on trait dimensions is an overgeneralization of perception of cues to emotional states. That is, similarity of facial features to emotional expressions is attributed to personality traits. However, it is also possible that these correlations can be accounted for by common semantic properties of the emotion and trait judgements rather than by perceptual similarity. To show that the relations between trait and emotion judgements cannot be accounted for entirely by semantic similarities, Said *et al.* [14] used a Bayesian network classifier to detect the subtle presence of features resembling emotions in the set of rated faces. The network accepts as input a feature vector containing the displacements between automatically chosen landmarks and the same landmarks of a prototypical neutral face. The output of the network is a set of probabilities corresponding to each basic emotion: happiness, surprise, anger, disgust, fear and sadness. Because the classifier was applied to neutral faces, the probabilities associated with specific emotions were very low. Nevertheless, these probabilities could be used to predict trait judgements of faces. As shown in figure 1*b*, the pattern of correlations was similar to the pattern of correlations for emotion and trait judgements, although the relationships were weaker. Out of the 84, 27 probabilities–trait judgement correlations were significant. These findings suggest that perceptions of most face traits (e.g. perceived trustworthiness, competence, friendliness, etc.) are based on structural similarities to emotional expressions. Further support for this idea has been garnered by Zebrowitz *et al.* [15], who have also used face image analysis techniques to demonstrate a relationship between face traits and emotional expressions.

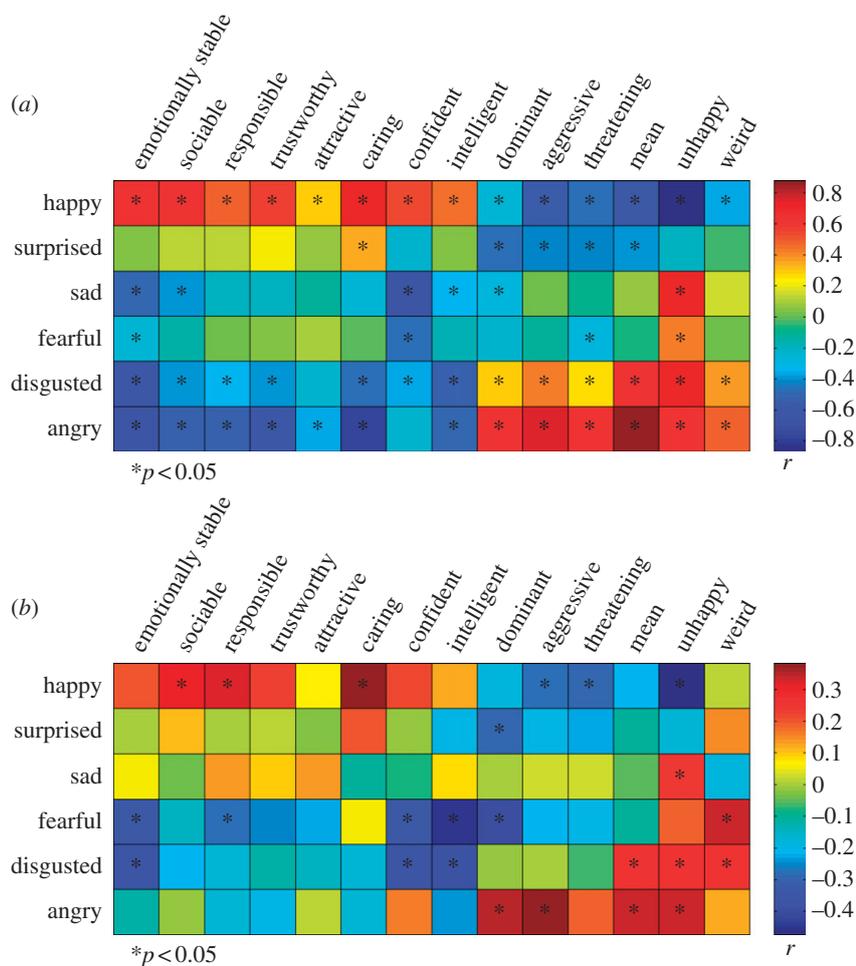


Figure 1. (a) Correlations between trait and emotion judgements of emotionally neutral faces. (b) Correlations between trait judgements and a classifier's probabilities of emotional expressions in these faces. Traits are ordered by their loadings on the first PC derived from a PCA of the traits.

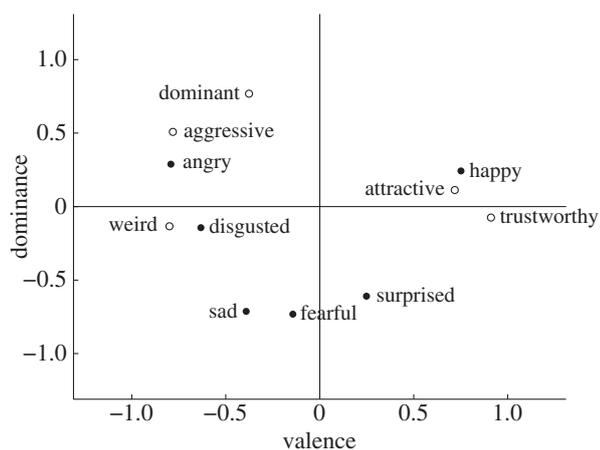


Figure 2. A plot of emotion judgements and a selected set of trait judgements in a two-dimensional space derived from a PCA on all trait and emotion judgements collectively. A representative subset of the traits is shown to avoid overcrowding.

Oosterhof & Todorov [13] used a data-driven approach to show that perceived face trustworthiness in neutral faces—a dimension that closely approximates the valence evaluation of faces—is largely

driven by the degree of resemblance to anger and happiness. Building on this work, Engell *et al.* [16] used a behavioural adaptation paradigm to test the hypothesis that the same neural systems subserve perception of these emotions and evaluation of emotionally neutral faces. The main assumption of this paradigm is that extended exposure to a stimulus category (adaptation) creates a visual after-effect so that the perception of subsequently viewed novel, ambiguous stimuli is shifted away from the adapting category to the extent that the same neuronal populations subserve the perception of the adapting and novel stimuli. In fact, Engell *et al.* found that although participants judged emotionally neutral faces as more trustworthy after adaptation to angry faces, they judged the emotionally neutral faces as less trustworthy after adaptation to happy faces. These judgements were unaffected by adaptation to fearful faces.

In this section, we discussed three types of evidence for the hypothesis that the perception of traits in neutral faces may originate in the perception of facial expressions. However, despite this behavioural evidence, cognitive neuroscience research has approached the study of emotional expressions and the study of face evaluation as independent problems. Next, we review research on the neural systems underlying emotion recognition.

3. CORTICAL SYSTEMS FOR THE PERCEPTION OF FACIAL EXPRESSIONS

Visual information about faces is first processed in the early visual system and is then believed to move to the occipital face area. From there, information is then sent to the inferior and lateral temporal lobes and much of the frontal cortex [17]. It is important to ask how—and to what degree—this network is functionally segregated. We argue that a representational space for facial expressions is stored predominantly in the superior temporal sulcus (STS), whereas the representational space for facial identity is stored predominantly in the fusiform gyrus.

It is known that facial expressions are represented at least in part in the STS. In humans, this region responds more to facial expressions than to neutral faces [18]. The STS adapts to repetitions of the same facial expression [19] and contains distributed representations of facial expressions that can be decoded at the resolution of fMRI [20].

Several studies have attempted to determine the segregation of function, if any, in the STS and in the fusiform cortex. Single-cell recordings in monkeys have shown that neurons tuned to facial expressions are located mostly in the STS, whereas neurons tuned to facial identity are located mostly in the inferior temporal (IT) cortex [21,22]. In humans, repetition of facial expression leads to fMRI adaptation in the anterior STS, whereas repetition of facial identity leads to adaptation in the fusiform and STS [19]. Furthermore, developmental prosopagnosics with severely impaired facial identity recognition have been shown to perform normally on tests of facial expression recognition [23,24]. Together, these studies provide good evidence for a partial segregation of function in the STS and fusiform gyrus.

It is important to note some caveats in the above dissociation, and to stress that the distinction between the STS and fusiform functionality is only partial. First, developmental prosopagnosics do not offer the best model for functional localization, as they do not possess lesions to a normally developed system. As has been noted by others, there are no clear cases of acquired lesions affecting facial expression recognition but not identity recognition, or vice versa [25]. Second, there is evidence that facial expressions are also represented in the fusiform gyrus. Like the STS, this region responds more to facial expressions than to neutral faces [18] and shows fMRI adaptation to the repetition of facial expression [19]. Indeed, one intracranial electroencephalography study showed better decoding of happy and fearful faces on the surface of human ventral temporal cortex than on the lateral temporal sites [26]. However, this difference might be explained by the fact that cortical surface electrodes are less well suited to recording activity from within a sulcus (e.g. the STS) than from the surface of a gyrus (e.g. the fusiform gyrus) because of the proximity of the neural source to the recording site and to the orientation of the evoked dipole.

While it is clear that the STS and fusiform area have some sort of role in facial expression processing, the mechanisms are far from clear. Giese & Poggio [27] have proposed a computational model of biological

motion recognition. The model—which can successfully detect walking, running and limping—specifies both a motion pathway and a form pathway. Near the beginning of the motion pathway, neurons in the middle temporal (MT) area, the middle superior temporal (MST) area and the kinetic occipital (KO) area are tuned to detect local regions of optic flow. These detectors then send signals to optic flow pattern detectors in the STS, whose activations can be modelled as Gaussian radial basis functions of the input. This type of function simply describes how particular STS neurons could have graded preferences for particular patterns of optic flow. In the case of facial expressions, a smile detector might respond optimally to local upward optic flow at the corners of the lips and compression around the eyes.

The STS and fusiform face area (FFA) also respond to static images of facial expressions. Extrapolating from the model, these neurons could be form detectors sensitive to particular configurations of the face surface. These form detectors could be asymmetrically connected to each other, such that each form detector pre-excites the form detector for a subsequent face configuration. Together, these form detectors encode a sequence of face configurations corresponding to a particular facial expression. The activations of these detectors could then be summed by a leaky integrator in the STS, which would therefore act as a motion pattern detector.

The Giese and Poggio model quantitatively specifies how form and motion detectors in the STS can be used to recognize body motion. We suspect that a similar model could be used for facial expression recognition. However, a model of visual processing of facial expressions in the STS does not imply that the region's response to facial expressions is strictly visual. The STS is a functionally heterogeneous region that receives input from multiple modalities [28]. Thus, some of its selectivity for facial expressions may be best explained by their emotional content rather than their physical appearance [25]. Future experiments could attempt to test how much of the STS encoding of facial expressions is visual and how much of it is emotional. As the visual relationships and emotional relationships are highly correlated [20], this question will be difficult to test.

In addition to the STS, an extended network for facial expression perception reaches the frontal operculum (FO), premotor cortex and somatosensory cortex [17]. The FO responds more to facial expressions than to neutral faces [18,29–32], and contains distributed representations of facial expressions that can be decoded at the spatial resolution of fMRI [20]. Additionally, transcranial magnetic stimulation (TMS) of the nearby right somatosensory cortex disrupts the perception of facial expressions, but not identity [33].

One possible role for the FO is to serve as a control region that enforces category distinctions in the STS. There is evidence that the ventrolateral prefrontal cortex, which overlaps with FO, exerts top-down control on the temporal lobes by selecting conceptual information during semantic memory tasks [34].

Another possibility is that the response in FO reflects activation of the mirror neuron system.

Mirror neurons fire both when an action is observed and when the same action is produced [35–39]. While it is important not to overstate the importance or prevalence of mirror neurons [40], there is reason to believe that they could be involved in the results. Mirror neurons are especially concentrated in monkey area F5, which is believed to be homologous to the FO in humans [41]. Facial expressions are particularly well suited for the mirror neuron hypothesis, as it is known that humans mimic the facial expression of people they are interacting with [42] and produce micro-expressions when simply looking at expressive face images [43,44]. It is thus possible that the activity in FO related to specific facial expressions may be owing to mirror neurons that fire upon the perception of expressions, and which might also drive micro-expression production in response. In fact, many of the studies that found activity in the FO during facial expression perception also found it during the production of facial expressions [29,31]. However, adaptation studies may be needed to establish that the same neurons—rather than different neurons within the same voxel—are involved in both action perception and production [45,46].

In sum, we argue that after initial processing in the occipital face area (OFA), face expression processing occurs in the STS and to a lesser extent in the FFA. In these two regions, asymmetrically connected form detectors might collectively encode motion patterns characteristic of facial expressions. In the STS, optic flow pattern detectors might detect preferred patterns of activation in earlier, local optic flow detectors. This region may have a more general role in processing emotional input from several modalities [25]. Additionally, the FO and nearby motor and somatosensory areas show activity during the perception of facial expression and may have a causal role in their recognition [33]. Since the monkey FO (area F5) is known to contain mirror neurons in monkeys, we argued that the human mirror neuron system in these regions might facilitate facial expression recognition.

4. BRAIN SYSTEMS FOR THE EVALUATION OF EMOTIONALLY NEUTRAL FACES

Having discussed the functional segregation of neural systems involved in facial expression recognition, we now turn our attention to the evaluation of emotionally neutral faces. We start our review of the face trait literature with attractiveness. While this property of faces has not been strongly tied to facial expressions, it is one of the most well studied social properties in neutral faces.

(a) *Brain systems for evaluating facial attractiveness*

Several fMRI studies have attempted to identify brain regions that show variable responses to different levels of facial attractiveness. A hypothesis in most of these studies is that perceptions of attractiveness should be related to activation in reward-related brain regions. Consistent with this hypothesis, the medial orbitofrontal cortex (mOFC) activated reliably across these studies, with greater activation as attractiveness

increased [47–50]. Conversely, the lateral orbitofrontal cortex (lOFC) showed greater activation with decreasing levels of attractiveness [47,49]. This dissociation has been interpreted in light of evidence that mOFC activates in response to abstract monetary reward while the lOFC activates to abstract monetary punishment [51]. According to this interpretation, mOFC activates to attractive faces because they are rewarding, and lOFC activates to unattractive faces because they are not rewarding. However, the distinction between lOFC and mOFC is not a strict dissociation. There is evidence that rewarding gustatory stimuli can activate the lOFC [52] and a fairly lateral OFC response to attractive faces has been found in one study [53].

The nucleus accumbens (NAcc) has also been reported to respond more strongly to attractive faces in several studies [47,53] and, like the mOFC, it has been interpreted with reference to its known role in reward processing [54,55]. However, many attractiveness studies have not reported NAcc activation [48–50,56]. The reason for this discrepancy is not entirely clear, but one possibility is that the studies that found it only showed faces of the opposite gender of the subject, whereas the studies that did not find it showed both genders to each subject. As proposed by Cloutier *et al.* [47], it may be possible that opposite gender-only paradigms put subjects in more of a mate-seeking context in which attractive faces of opposite gender are particularly rewarding.

The anterior cingulate cortex (ACC) shows greater activation to attractive faces than to unattractive faces, according to two studies [47,50]. Because the ACC is known to generate and monitor autonomic states [57,58], it is possible that its activity during the presentation of attractive faces reflects autonomic arousal. Support for this hypothesis comes from a study [50], in which only males showed increased pupil dilation—an indicator of autonomic arousal—and increased ACC activation to attractive faces.

(b) *Brain systems for evaluating face valence and trustworthiness*

Most studies on perceptions of face trustworthiness have focused on the amygdala, following a study by Adolphs *et al.* [59]. Adolphs *et al.* tested three bilateral amygdala damage patients, other brain damage controls and normal controls on perceptions of approachability and trustworthiness. Relative to the controls, bilateral amygdala damage patients showed a specific bias to give high ratings of trustworthiness and approachability to faces that were judged by normal controls as untrustworthy and unapproachable. In addition, participants who are given an intranasal dose of oxytocin, which is believed to work in part by dampening amygdala activity [60], make higher judgements of trustworthiness than controls [61]. Interestingly, in contrast to bilateral amygdala damage patients, some developmental prosopagnosics are able to make normal trustworthiness judgements [62]. As in the case of facial expressions, these findings suggest that the neural systems that underlie face evaluation and processing of facial identity are at least partially dissociable.

Several fMRI studies with normal participants have confirmed the amygdala's involvement in the

perception of trustworthiness [63,64]. Winston *et al.* [64] showed that the amygdala's response decreased with the trustworthiness of faces, as assessed by the subjects' judgements of the faces after the brain imaging session. This was the case independent of the subjects' task in the scanner: judging trustworthiness or age of faces. Engell *et al.* [63] replicated the findings that the activation in the amygdala decreased with face trustworthiness, using only an implicit task to rule out the possibility that performance on implicit trials was influenced by prior performance on explicit trials. They also showed that the amygdala's response to face trustworthiness was driven by structural properties of the face that signal trustworthiness across perceivers rather than by idiosyncratic components of trustworthiness judgements. Todorov *et al.* [65] and Todorov & Engell [66] replicated these findings, using faces generated by a computer model of face trustworthiness.

Although all initial fMRI studies reported a linear amygdala response to face trustworthiness, two subsequent studies found a nonlinear response [65,67]. Using an explicit evaluation task, Said *et al.* [67] showed that the response to extremely trustworthy and untrustworthy faces was larger than the response to faces in the middle of the continuum. Using an implicit evaluation task, Todorov *et al.* [65] found a similar quadratic response in the left but not the right amygdala.

(c) *An analysis across studies using computer-generated faces*

To try to better summarize the existing research on studies of face trustworthiness that used computer-generated faces, we performed a large group analysis on the trustworthiness results from three fMRI studies, focusing initially on studies that used artificial faces [2,68] (figure 3). In the first study, 22 subjects viewed computer-generated faces designed to parametrically vary on trustworthiness and dominance while performing a one-back task. In the second study, a separate group of 22 subjects viewed the same faces while indicating whether they would like to approach or avoid each face. In the third study, 23 subjects viewed a set of faces designed to vary on trustworthiness as well as a set of faces designed to vary on an orthogonal face dimension. All of the faces that varied on trustworthiness were produced by a validated statistical model [13]. For each subject and for each voxel, the parameter estimates for the linear and quadratic responses to trustworthiness were entered into a *t*-test. In total, 67 subjects were entered into the analysis. We applied a strict uncorrected voxel-wise cut-off of $p < 10^{-4}$ to the brain maps. To correct for multiple comparisons, we used the AFNI program AlphaSim to determine the minimum cluster size needed for corrected significance of $p < 0.01$. This simulation revealed a minimum cluster size of 0.4 ml.

We observed both linear and quadratic responses to face trustworthiness in extensive areas throughout the brain (figure 4). The lateral occipital cortex, fusiform gyrus, middle temporal gyrus, inferior parietal lobule and putamen all showed linear responses to

face trustworthiness, with greater responses to untrustworthy faces than trustworthy faces (table 1). All of these regions showed the response bilaterally. The bilateral left occipital cortex, fusiform gyrus, amygdala, hippocampus and middle temporal gyrus all showed quadratic responses to face trustworthiness, with greater responses to highly trustworthy and highly untrustworthy faces than to faces in the middle of the continuum (table 2).

(d) *Face trustworthiness, or general distance from the average face?*

The analysis presented above was based on studies that used faces generated by a computer model of trustworthiness [13], which was derived from a multi-dimensional face space—a vector space where each dimension can be thought of as a physical property of faces, and every face can be represented as a point in the space [69]. The origin of the space contains the average face. According to the model, the trustworthiness of a face near the average face can be increased maximally by moving it in one direction in face space, and decreased maximally by moving it in the opposite direction. Hence, in the context of a multi-dimensional face space, face trustworthiness is confounded with the distance from the average face. In fact, there is evidence that the fusiform response increases with distance from the average face [70,71].

To test whether the amygdala and posterior face-selective regions are responding to facial properties that convey specific social signals or to general distance from the average face, Said *et al.* [2] directly compared trustworthiness with a control dimension that was perceived to be less socially relevant but was matched on face distance to trustworthiness. To represent trustworthiness, we used a valence dimension that was obtained from a PCA of nine different social judgements of faces and was highly correlated with judgements of trustworthiness ($r = 0.91$). As shown in figure 5, the amygdala and the FFA did not discriminate between these two dimensions, suggesting that their responses were at least partially driven by the distance to the average face.

These findings suggest that the BOLD activation in these brain regions is more closely tied to the distance of a face from the mean, a variable that can be roughly described as face typicality, rather than any particular face trait. While face typicality is different from the traits that are usually tested in social experiments, it does carry socially relevant information. In fact, descriptive judgements of typicality ('how likely would you be to see a person who looks like this walking down the street?') were highly correlated with 13 out of 14 evaluative, social judgements (e.g. trustworthiness, attractiveness, weirdness) [2].

(e) *Face typicality can explain inconsistencies in the literature*

A typicality explanation can account for a major inconsistency in the literature. Some studies on the amygdala and high-level visual areas have found linear response functions to social traits, whereas others have found highly nonlinear U-shaped response



Figure 3. Parametric manipulation of face trustworthiness. From left to right: -3 , -1 , 1 and 3 standard deviations away from the mean face.

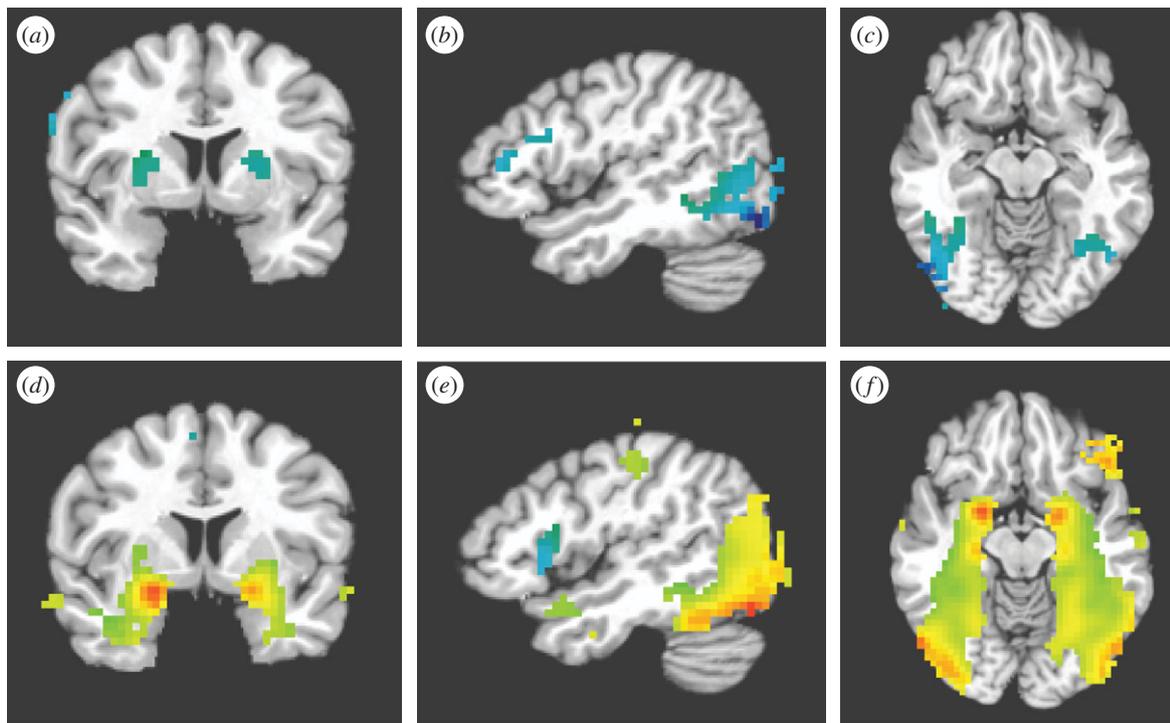


Figure 4. Brain responses to faces parametrically manipulated to vary on perceived trustworthiness. (a–c) Statistical overlay of regions showing a linear response to face trustworthiness in a group analysis of 67 subjects across three studies. Cool colours indicate a negative linear trend with greater responses to untrustworthy faces than trustworthy faces. Uncorrected $p < 10^{-4}$. (d–f) Statistical overlay of regions showing a quadratic response to face trustworthiness. Hot colours indicate a positive quadratic term. Cool colours indicate a negative quadratic term. Uncorrected $p < 10^{-4}$.

functions. Most studies that observed nonlinear, U-shaped responses used faces generated by a computer model or some other procedure for creating synthetic faces. For these faces, the typicality function is nonlinear: faces at the middle of the continuum are perceived as more typical. Most studies that observed linear responses used real faces. For these faces, the typicality function is linear: faces at the positive end of the continuum are perceived as more typical. Interestingly, the one study that used real faces and observed a U-shaped response in the amygdala [50] used normal and extremely attractive faces. We suspect that the extremely attractive faces would be perceived as less typical than average faces.

(f) *Future directions for research on face evaluation*

Because of the diversity of results, looking for an ‘attractiveness network’ or ‘trustworthiness network’ may not be the best approach for future research. Instead, one

challenge is to specify how social variables in faces interact with multiple other variables—including typicality, context and task variables—to generate neural responses.

A more difficult issue stems from the fact that most of the studies above can only describe the relationship between face trustworthiness and the mean BOLD response in large brain areas. While this is an important first step, traditional fMRI data can only provide limited constraints on theoretical neuroscience models of face evaluation. One fairly extreme model is that dedicated neurons in the fusiform gyrus represent face valence/trustworthiness with a firing rate code, after receiving input from form-sensitive neurons. The non-monotonic quadratic responses described above could be explained by this model if we assume (i) that some of these neurons prefer positive/trustworthy faces; (ii) that some of these neurons prefer negative/untrustworthy faces; and (iii) that their responses are convex nonlinear functions of valence/trustworthiness. This way, a single voxel containing both types of neurons

Table 1. Regions showing a significant linear response to face trustworthiness in a group analysis of 67 subjects across three studies. Corrected $p < 0.01$. Coordinates of the peak voxel are in the left–posterior–inferior (LPI) convention.

volume (ml)	region	X	Y	Z
20.0	right occipital cortex, fusiform gyrus, middle temporal gyrus, inferior parietal lobule	46.5	-70.5	-12.5
3.7	left inferior parietal lobule	-28.5	-73.5	47.5
3.2	right middle frontal gyrus	58.5	25.5	17.5
2.2	left fusiform gyrus	-34.5	-55.5	-15.5
1.8	right putamen	19.5	7.5	8.5
1.3	left occipital cortex	-34.5	-91.5	11.5
0.9	left putamen	-19.5	4.5	8.5
0.5	right inferior frontal gyrus	58.5	10.5	26.5

Table 2. Regions showing a significant quadratic response to face trustworthiness in a group analysis of 67 subjects across three studies. Corrected $p < 0.01$. Coordinates of the peak voxel are in the LPI convention.

volume (ml)	region	X	Y	Z
70.3	left occipital cortex, fusiform gyrus, amygdala, hippocampus, middle temporal gyrus, and superior occipital gyrus	-43.5	-73.5	-15.5
55.7	right occipital cortex, fusiform gyrus, amygdala, hippocampus, and middle temporal gyrus	37.5	-76.5	-15.5
12.7	medial prefrontal cortex	1.5	55.5	5.5
11.0	right postcentral gyrus	28.5	-40.5	65.5
5.7	left superior parietal lobule	-7.5	-58.5	62.5
5.4	left inferior frontal gyrus	-43.5	25.5	-21.5
5.1	right insula*	34.5	19.5	5.5
3.5	left cingulate gyrus	-4.5	-49.5	29.5
3.3	right medial frontal gyrus*	4.5	10.5	47.5
1.5	left insula*	-34.5	16.5	8.5
1.0	right anterior middle temporal gyrus	61.5	-4.5	-9.5
0.9	right anterior superior temporal gyrus	40.5	10.5	-15.5
0.5	right posterior caudate	16.5	-19.5	23.5
0.5	left anterior middle temporal gyrus	-61.5	-10.5	-12.5

*negative quadratic term.

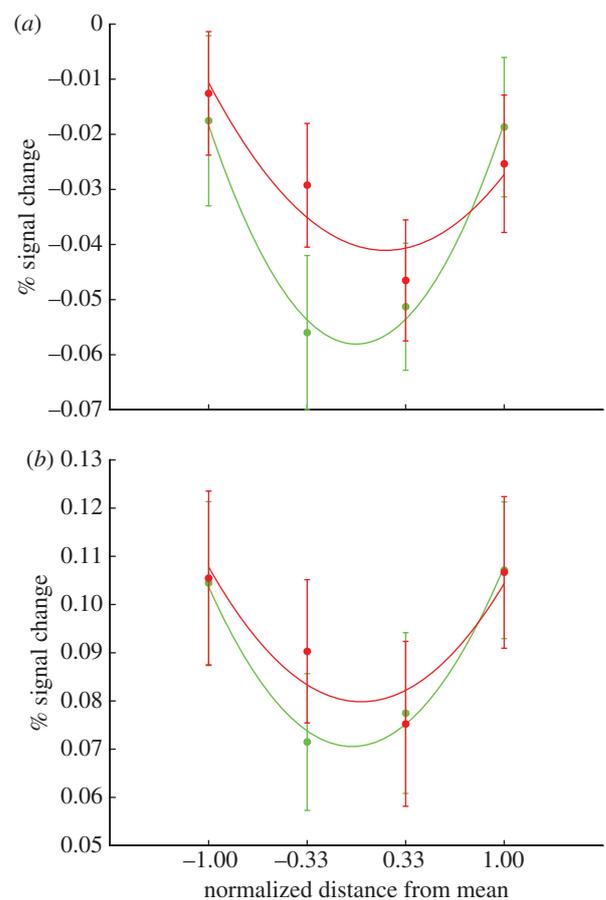


Figure 5. The response to both the high-social dimension (green; valence) and the low-social dimension (red; control) in the amygdala and the FFA. (a) The bilateral amygdala, defined anatomically by the AFNI Talairach atlas; (b) the bilateral FFA, defined by a face localizer.

would show a quadratic, non-monotonic response function. However, different models could also explain the data. For example, voxel-level functions could reflect attention-based amplification of responses by form-sensitive neurons. Under this possibility, face valence/trustworthiness is not represented in visual areas. Instead, attentional feedback from other areas amplifies visual system responses to socially relevant faces. Neither of these possibilities—which are just two of many—can be disentangled by the existing research, although fMRI adaptation studies would help.

5. FUTURE RESEARCH DIRECTIONS

As we documented in §2, there is good behavioural evidence that perceptions of emotional expressions and perceptions of social traits in faces are related. Yet, these two research topics have not been pursued in a common, cognitive neuroscience framework. While behavioural studies can provide support for the emotional overgeneralization hypothesis, further support can be obtained by examining the brain areas that are involved in expression perception and trait perception. There are several suggestive preliminary findings. In an fMRI study, Winston *et al.* [64] showed that the STS activates more during explicit ratings of face trustworthiness compared with ratings of face age. As the STS is also involved in facial

expression perception, this result is consistent with the overgeneralization hypothesis. Additional causal evidence for an STS role in face trait perception has been found in a TMS study. Dzhelyova *et al.* [72] applied repetitive TMS to the OFA and the posterior STS in two face tasks: judging gender and judging trustworthiness. Whereas TMS applied to the right posterior STS interfered with trustworthiness judgements—making participants slower, TMS applied to the right OFA interfered with gender judgements. Said *et al.* [2] found the STS differentiates between the social and control dimensions discussed above, but failed to replicate this finding in other studies. Finally, the amygdala response to face trustworthiness is consistent with overgeneralization. Just as the amygdala responds to both angry and happy faces more than neutral faces [73], so it responds more to highly trustworthy and highly untrustworthy faces. However, while findings in the amygdala are consistent with overgeneralization, we stress that they are also consistent with non-perceptual arousal or valence effects that are shared by both facial expressions and face traits.

Clearly, the perceptions of facial expressions and face traits are not identical processes. The stimuli are different, and some dissociations emerge in the neural responses. For instance, the FO, which is important for the perception of facial expressions [18,20], is not necessary for trustworthiness judgements [72]. However, the growing body of behavioural and computer vision studies, together with reports of substantially overlapping neural responses, suggest that these two processes are closely linked.

In all cases, the most fruitful approach would be to pursue these two phenomena within a common, conceptual framework. For example, Calder & Young [25] discussed how multi-dimensional face space models can account for a wide variety of face perceptions (e.g. identity, gender, age, emotions, etc.). Such models can provide a common framework for cognitive neuroscience studies on the perception of emotional expressions and face evaluation. Another approach would be to use pattern classification techniques [74] to directly test for similarities in the representations of emotional expressions and face traits. For example, such techniques can be applied to brain states during the perception of expressive faces and emotionally neutral faces. As we described in §3, it is possible to classify different emotional expressions [20]. Classifiers that are trained to detect the pattern of activation associated with the perception of particular expressions can then be tested on the perception of neutral faces that vary on their evaluation. If overgeneralization is occurring in a particular brain area, then the classifier may detect similar activation patterns during the viewing of neutral faces varying on trait dimensions and during the viewing of corresponding emotional expressions.

6. CONCLUSION

In this overview, we discussed the cognitive neuroscience of the recognition of facial expressions and evaluation of emotionally neutral faces. While a large number of studies have attempted to research these topics, many questions remain unanswered,

particularly those about mechanism. Nevertheless, some conclusions can be reached. First, we used several lines of behavioural evidence to argue that the perceptions of facial expressions and face traits are related, and may rely on the same mechanisms. Second, we discussed the broadly distributed network of cortical systems underlying the perception of facial expressions. We argued that there is partial functional segregation of the STS and fusiform gyrus, with the STS more responsible for facial expressions and the fusiform gyrus more responsible for facial identity. In the STS, facial expressions may be recognized by a network of local optic flow detectors, optic flow pattern detectors and form detectors [27]. The STS may also have a more general role in processing emotional input from several modalities [25]. The FO and nearby motor and somatosensory cortex also show activity during the perception of facial expression and may have a causal role in their recognition, perhaps through engagement of the mirror neuron system. Third, we discussed the cognitive neuroscience of the evaluation of neutral faces. In attractiveness research, some of the most consistent findings are that the mOFC, the NAcc and the ACC all respond more to more attractive faces. In research on face trustworthiness, some of the most consistent findings are that the amygdala and large parts of the visual system are sensitive to face trustworthiness, although the response function is not always the same across studies. We argued that these inconsistencies can be explained in terms of the typicality of the faces used in the particular studies. How these responses should be interpreted is still largely an open question. Future research—particularly studies relying on fMRI adaptation, TMS or patient populations—could be very illuminating.

REFERENCES

- 1 Todorov, A. 2008 Evaluating faces on trustworthiness: an extension of systems for recognition of emotions signaling approach/avoidance behaviors. *Year Cogn. Neurosci.* **1124**, 208–224.
- 2 Said, C. P., Dotsch, R. & Todorov, A. 2010 The amygdala and FFA track both social and non-social face dimensions. *Neuropsychologia* 3596–3605. (doi:10.1016/j.neuropsychologia.2010.08.009)
- 3 Olivola, C. Y. & Todorov, A. 2010 Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *J. Exp. Social Psychol.* **46**, 315–324. (doi:10.1016/j.jesp.2009.12.002)
- 4 Willis, J. & Todorov, A. 2006 First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* **17**, 592–598. (doi:10.1111/j.1467-9280.2006.01750.x)
- 5 Zebrowitz, L. A. & Montepare, J. M. 2008 Social psychological face perception: why appearance matters. *Soc. Pers. Psychol. Compass* **2**, 1497–1517. (doi:10.1111/j.1751-9004.2008.00109.x)
- 6 Ballew II, C. C. & Todorov, A. 2007 Predicting political elections from rapid and unreflective face judgments. *Proc. Natl Acad. Sci. USA* **104**, 17 948–17 953. (doi:10.1073/pnas.0705435104)
- 7 Mazur, A., Mazur, J. & Keating, C. 1984 Military rank attainment of a West Point class: effects of cadets physical features. *Am. J. Sociol.* **90**, 125–150. (doi:10.1086/228050)
- 8 Todorov, A., Mandisodza, A. N., Goren, A. & Hall, C. C. 2005 Inferences of competence from faces predict

- election outcomes. *Science* **308**, 1623–1626. (doi:10.1126/science.1110589)
- 9 Zebrowitz, L. A. 1997 *Reading faces: window to the soul?* Boulder, CO: Westview Press.
 - 10 Montepare, J. M. & Dobish, H. 2003 The contribution of emotion perceptions and their overgeneralizations to trait impressions. *J. Nonverbal Behav.* **27**, 237–254. (doi:10.1023/A:1027332800296)
 - 11 Zebrowitz, L. A. 2004 The origins of first impressions. *J. Cult. Evol. Psychol.*, **2**, 93–108. (doi:10.1556/JCEP.2.2004.1-2.6)
 - 12 Knutson, B. 1996 Facial expressions of emotion influence interpersonal trait inferences. *J. Nonverbal Behav.* **20**, 165–182. (doi:10.1007/BF02281954)
 - 13 Oosterhof, N. N. & Todorov, A. 2008 The functional basis of face evaluation. *Proc. Natl Acad. Sci. USA* **105**, 11 087–11 092. (doi:10.1073/pnas.0805664105)
 - 14 Said, C. P., Sebe, N. & Todorov, A. 2009 Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion* **9**, 260–264. (doi:10.1037/a0014681)
 - 15 Zebrowitz, L. A., Kikuchi, M. & Fellous, J. M. 2010 Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *J. Pers. Soc. Psychol.* **98**, 175–189. (doi:10.1037/a0017990)
 - 16 Engell, A. D., Todorov, A. & Haxby, J. V. 2010 Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception* **39**, 931–941. (doi:10.1068/p6633)
 - 17 Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. 2000 The distributed human neural system for face perception. *Trends Cogn. Sci.* **4**, 223–233. (doi:10.1016/S1364-6613(00)01482-0)
 - 18 Engell, A. D. & Haxby, J. V. 2007 Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia* **45**, 3234–3241. (doi:10.1016/j.neuropsychologia.2007.06.022)
 - 19 Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R. & Dolan, R. J. 2004 fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *J. Neurophysiol.* **92**, 1830–1839. (doi:10.1152/jn.00155.2004)
 - 20 Said, C. P., Moore, C. D., Engell, A. D., Todorov, A. & Haxby, J. V. 2010 Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J. Vis.*, **10**. (doi:10.1167/10.5.11)
 - 21 Hasselmo, M. E., Rolls, E. T. & Baylis, G. C. 1989 The role of expression and identity in the face-selective responses of neurons in the temporal visual-cortex of the monkey. *Behav. Brain Res.* **32**, 203–218. (doi:10.1016/S0166-4328(89)80054-3)
 - 22 Perrett, D. I., Hietanen, J. K., Oram, M. W. & Benson, P. J. 1992 Organization and functions of cells responsive to faces in the temporal cortex. *Phil. Trans. R. Soc. Lond. B* **335**, 23–30. (doi:10.1098/rstb.1992.0003)
 - 23 Bentin, S., Degutis, J. M., D'Esposito, M. & Robertson, L. C. 2007 Too many trees to see the forest: performance, event-related potential, and functional magnetic resonance imaging manifestations of integrative congenital prosopagnosia. *J. Cogn. Neurosci.* **19**, 132–146. (doi:10.1162/jocn.2007.19.1.132)
 - 24 Duchaine, B. C., Parker, H. & Nakayama, K. 2003 Normal recognition of emotion in a prosopagnosic. *Perception* **32**, 827–838. (doi:10.1068/p5067)
 - 25 Calder, A. J. & Young, A. W. 2005 Understanding the recognition of facial identity and facial expression. *Nat. Rev. Neurosci.* **6**, 641–651. (doi:10.1038/nrn1724)
 - 26 Tsuchiya, N., Kawasaki, H., Oya, H., Howard III, M. A. & Adolphs, R. 2008 Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. *PLoS ONE* **3**, e3892. (doi:10.1371/journal.pone.0003892)
 - 27 Giese, M. A. & Poggio, T. 2003 Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* **4**, 179–192. (doi:10.1038/nrn1057)
 - 28 Hein, G. & Knight, R. T. 2008 Superior temporal sulcus—it's my area: or is it? *J. Cogn. Neurosci.* **20**, 2125–2136. (doi:10.1162/jocn.2008.20148)
 - 29 Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C. & Lenzi, G. L. 2003 Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl Acad. Sci. USA* **100**, 5497–5502. (doi:10.1073/pnas.0935845100)
 - 30 Kitada, R., Johnsrude, I. S., Kochiyama, T. & Lederman, S. J. 2010 Brain networks involved in haptic and visual identification of facial expressions of emotion: an fMRI study. *Neuroimage* **49**, 1677–1689. (doi:10.1016/j.neuroimage.2009.09.014)
 - 31 Montgomery, K. J. & Haxby, J. V. 2008 Mirror neuron system differentially activated by facial expressions and social hand gestures: a functional magnetic resonance imaging study. *J. Cogn. Neurosci.* **20**, 1866–1877. (doi:10.1162/jocn.2008.20127)
 - 32 Sprengelmeyer, R., Rausch, M., Eysel, U. T. & Przuntek, H. 1998 Neural structures associated with recognition of facial expressions of basic emotions. *Proc. R. Soc. Lond. B* **265**, 1927–1931. (doi:10.1098/rspb.1998.0522)
 - 33 Pitcher, D., Garrido, L., Walsh, V. & Duchaine, B. C. 2008 Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *J. Neurosci.* **28**, 8929–8933. (doi:10.1523/JNEUROSCI.1450-08.2008)
 - 34 Martin, A. 2007 The representation of object concepts in the brain. *Ann. Rev. Psychol.* **58**, 25–45. (doi:10.1146/annurev.psych.57.102904.190143)
 - 35 Ferrari, P. F., Gallese, V., Rizzolatti, G. & Fogassi, L. 2003 Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur. J. Neurosci.* **17**, 1703–1714. (doi:10.1046/j.1460-9568.2003.02601.x)
 - 36 Montgomery, K. J., Isenberg, N. & Haxby, J. V. 2007 Communicative hand gestures and object-directed hand movements activated the mirror neuron system. *Soc. Cogn. Affect Neurosci.* **2**, 114–122. (doi:10.1093/scan/nsm004)
 - 37 Montgomery, K. J., Seherman, K. R. & Haxby, J. V. 2009 The well-tempered social brain. *Psychol. Sci.* **20**, 1211–1213. (doi:10.1111/j.1467-9280.2009.02428.x)
 - 38 Nakamura, K. et al. 1999 Activation of the right inferior frontal cortex during assessment of facial emotion. *J. Neurophysiol.* **82**, 1610–1614.
 - 39 Rizzolatti, G. & Craighero, L. 2004 The mirror-neuron system. *Ann. Rev. Neurosci.* **27**, 169–192. (doi:10.1146/annurev.neuro.27.070203.144230)
 - 40 Dinstein, I., Thomas, C., Behrmann, M. & Heeger, D. J. 2008 A mirror up to nature. *Curr. Biol.* **18**, R13–R18. (doi:10.1016/j.cub.2007.11.004)
 - 41 Johnson-Frey, S. H., Maloof, F. R., Newman-Norlund, R., Farrer, C., Inati, S. & Grafton, S. T. 2003 Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron* **39**, 1053–1058. (doi:10.1016/S0896-6273(03)00524-5)
 - 42 Buck, R. 1984 *The communication of emotion*. New York, NY: Guilford.
 - 43 Dimberg, U. 1982 Facial reactions to facial expressions. *Psychophysiology* **19**, 643–647. (doi:10.1111/j.1469-8986.1982.tb02516.x)
 - 44 Dimberg, U., Thunberg, M. & Elmehed, K. 2000 Unconscious facial reactions to emotional facial

- expressions. *Psychol. Sci.* **11**, 86–89. (doi:10.1111/1467-9280.00221)
- 45 Grill-Spector, K. & Malach, R. 2001 fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol. (Amst.)* **107**, 293–321. (doi:10.1016/S0001-6918(01)00019-1)
- 46 Krekelberg, B., Boynton, G. M. & van Wezel, R. J. A. 2006 Adaptation: from single cells to BOLD signals. *Trends Neurosci.* **29**, 250–256. (doi:10.1016/j.tins.2006.02.008)
- 47 Cloutier, J., Heatherton, T. F., Whalen, P. J. & Kelley, W. M. 2008 Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. *J. Cogn. Neurosci.* **20**, 941–951. (doi:10.1162/jocn.2008.20062)
- 48 Kranz, F. & Ishai, A. 2006 Face perception is modulated by sexual preference. *Curr. Biol.* **16**, 63–68. (doi:10.1016/j.cub.2005.10.070)
- 49 O’Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M. & Dolan, R. J. 2003 Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* **41**, 147–155. (doi:10.1016/S0028-3932(02)00145-8)
- 50 Winston, J. S., O’Doherty, J., Kilner, J. M., Perrett, D. I. & Dolan, R. J. 2007 Brain systems for assessing facial attractiveness. *Neuropsychologia* **45**, 195–206. (doi:10.1016/j.neuropsychologia.2006.05.009)
- 51 O’Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J. & Andrews, C. 2001 Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* **4**, 95–102. (doi:10.1038/82959)
- 52 O’Doherty, J. P., Deichmann, R., Critchley, H. D. & Dolan, R. J. 2002 Neural responses during anticipation of a primary taste reward. *Neuron* **33**, 815–826. (doi:10.1016/S0896-6273(02)00603-7)
- 53 Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O’Connor, E. & Breiter, H. C. 2001 Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* **32**, 537–551. (doi:10.1016/S0896-6273(01)00491-3)
- 54 Breiter, H. C. *et al.* 1997 Acute effects of cocaine on human brain activity and emotion. *Neuron* **19**, 591–611. (doi:10.1016/S0896-6273(00)80374-8)
- 55 Schultz, W. 2000 Multiple reward signals in the brain. *Nat. Rev. Neurosci.* **1**, 199–207. (doi:10.1038/35044563)
- 56 Kampe, K. K. W., Frith, C. D., Dolan, R. J. & Frith, U. 2001 Reward value of attractiveness and gaze: making eye contact enhances the appeal of a pleasing face, irrespective of gender. *Nature* **413**, 589–589. (doi:10.1038/35098149)
- 57 Critchley, H. D. 2004 The human cortex responds to an interoceptive challenge. *Proc. Natl Acad. Sci. USA* **101**, 6333–6334. (doi:10.1073/pnas.0401510101)
- 58 Teves, D., Videen, T. O., Cryer, P. E. & Powers, W. J. 2004 Activation of human medial prefrontal cortex during autonomic responses to hypoglycemia. *Proc. Natl Acad. Sci. USA* **101**, 6217–6221. (doi:10.1073/pnas.0307048101)
- 59 Adolphs, R., Tranel, D. & Damasio, A. R. 1998 The human amygdala in social judgment. *Nature* **393**, 470–474. (doi:10.1038/30982)
- 60 Kirsch, P. *et al.* 2005 Oxytocin modulates neural circuitry for social cognition and fear in humans. *J. Neurosci.* **25**, 11 489–11 493. (doi:10.1523/JNEUROSCI.3984-05.2005)
- 61 Theodoridou, A., Rowe, A. C., Penton-Voak, I. S. & Rogers, P. J. 2009 Oxytocin and social perception: oxytocin increases perceived facial trustworthiness and attractiveness. *Horm. Behav.* **56**, 128–132. (doi:10.1016/j.yhbeh.2009.03.019)
- 62 Todorov, A. & Duchaine, B. 2008 Reading trustworthiness in faces without recognizing faces. *Cogn. Neuropsychol.* **25**, 395–410. (doi:10.1080/02643290802044996)
- 63 Engell, A. D., Haxby, J. V. & Todorov, A. 2007 Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. *J. Cogn. Neurosci.* **19**, 1508–1519. (doi:10.1162/jocn.2007.19.9.1508)
- 64 Winston, J. S., Strange, B. A., O’Doherty, J. & Dolan, R. J. 2002 Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat. Neurosci.* **5**, 277–283. (doi:10.1038/nn816)
- 65 Todorov, A., Baron, S. G. & Oosterhof, N. N. 2008 Evaluating face trustworthiness: a model based approach. *Social Cogn. Affect. Neurosci.* **3**, 119–127. (doi:10.1093/scan/nsn009)
- 66 Todorov, A. & Engell, A. D. 2008 The role of the amygdala in implicit evaluation of emotionally neutral faces. *Social Cogn. Affect. Neurosci.* **3**, 303–312. (doi:10.1093/scan/nsn033)
- 67 Said, C. P., Baron, S. G. & Todorov, A. 2009 Nonlinear amygdala response to face trustworthiness: contributions of high and low spatial frequency information. *J. Cogn. Neurosci.* **21**, 519–528. (doi:10.1162/jocn.2009.21041)
- 68 Todorov, A., Said, C. P., Oosterhof, N. N. & Engell, A. D. In press. Neural systems for face evaluation. *J. Cogn. Neurosci.*
- 69 Valentine, T. 1991 A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol. A* **43**, 161–204.
- 70 Leopold, D. A., Bondar, I. V. & Giese, M. A. 2006 Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* **442**, 572–575. (doi:10.1038/nature04951)
- 71 Loffler, G., Yourganov, G., Wilkinson, F. & Wilson, H. R. 2005 fMRI evidence for the neural representation of faces. *Nat. Neurosci.* **8**, 1386–1390. (doi:10.1038/nn1538)
- 72 Dzhelyova, M. P., Ellison, A. & Atkinson, A. P. In press. Event-related repetitive TMS reveals distinct, critical roles for right OFA and bilateral posterior STS in judging the sex and trustworthiness of faces. *J. Cogn. Neurosci.*
- 73 Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M., Hyman, S. & Rosen, B. 1996 Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* **17**, 875–887. (doi:10.1016/S0896-6273(00)80219-6)
- 74 Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L. & Pietrini, P. 2001 Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430. (doi:10.1126/science.1063736)