Lateral OFC Activity Predicts Decision Bias due to First Impressions during Ultimatum Games

Hackjin Kim, Min-Jo Choi, and In-Ji Jang

Abstract

■ Despite the prevalence and potentially harmful consequences of first impression bias during social decision-making, its precise neural underpinnings remain unclear. Here, on the basis of the fMRI study using ultimatum games, the authors show that the responders' decisions to accept or reject offers were significantly affected by facial trustworthiness of proposers. Analysis using a model-based fMRI method revealed that activity in the right lateral OFC (lOFC) of responders increased as a function of negative decision bias, indicating a greater likelihood of rejecting otherwise fair offers, possibly because of the facial trustworthiness of proposers. In addition, lOFC showed changes in functional connectivity strength with amygdala and insula as a function of decision bias, and individual differences in the strengths of connectivities between lOFC and bilateral insula were also found to predict the likelihood of responders to reject offers from untrustworthy-looking proposers. The present findings emphasize that the lOFC plays a pivotal role in integrating signals related to facial impression and creating signal biasing decisions during social interactions. ■

INTRODUCTION

Numerous studies have indicated that our daily decisions are influenced by the facial impressions of others in various social settings. For example, attractive people are favored in mating situations (Green, Buchanan, & Heuer, 1984; Hatfield, Aronson, Abrahams, & Rottmann, 1966) and are considered to be happier, to have more prestigious jobs, and to have better personalities (Dion, Berscheid, & Walster, 1972). In addition, attractive people, compared with unattractive people, are more likely to be successful at job interviews (Dipboye, Arvey, & Terpstra, 1977), earn a higher salary (Hamermesh & Biddle, 1994; Frieze, Olson, & Russell, 1991), and receive less severe sentences in courts (Downs & Lyons, 1991; Sigall & Ostrove, 1975).

The effect of facial appearance on decision-making has also been investigated in diverse laboratory experiments using simple and systematic social interactions. According to a phenomenon called "beauty premium," attractive people tend to receive higher monetary offers from proposers in the ultimatum game (Solnick & Schweitzer, 1999), are more likely to be chosen as partners in the prisoner's dilemma game (Mulford, Orbell, Shatto, & Stockard, 1998), and are more likely to receive higher investments in the trust game (Wilson & Eckel, 2006) and in the public good game (Andreoni & Petrie, 2008). On the basis of the evidence mentioned above, one could speculate that, when facing a stranger, we tend to infer his or her personality and traits, such as trustworthiness, rather automatically based on appearance. Furthermore, such rapidly formed first impressions can bias perceivers' decisions, often without explicit awareness (van't Wout & Sanfey, 2008).

Recent studies have demonstrated that the human amygdala plays a key role in assessing facial trustworthiness by showing that patients with bilateral amygdala damage fail to rate untrustworthy strangers negatively (Adolphs, Tranel, & Damasio, 1998) and that amygdala activity increases when people see untrustworthy-looking faces even without explicitly evaluating facial impression (Engell, Haxby, & Todorov, 2007; Winston, Strange, O'Doherty, & Dolan, 2002). Furthermore, the manipulation of amygdala activity using intranasally administered oxytocin, a neuropeptide hormone implicated in various behaviors including female reproduction and social bonding, has been shown to effectively enhance trust and, thus, result in increased monetary investment to trustees during trust games (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005). Despite the unequivocal role played by the amygdala in "reading" of trustworthiness in faces and biasing decisions to trust others, few studies have directly attempted to investigate the exact neural mechanism whereby trustworthiness-related cues detected by the amygdala influence decisions in a social context.

The OFC integrates signals from diverse brain structures involved in processing emotional information, such as the nucleus accumbens, the amygdala, the insula, and the temporal cortex, and is considered to be strongly implicated in guiding a wide range of emotional and social decisions (O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001; Bechara, Damasio, & Damasio, 2000; Cavada & Schultz,

Korea University

2000; Schoenbaum, Chiba, & Gallagher, 1998). A number of animal and human studies have suggested that intact functioning of the OFC, with respect to decision guidance, is critically dependent on affective input signals from the amygdala (Hampton, Adolphs, Tyszka, & O'Doherty, 2007; Schoenbaum, Setlow, Saddoris, & Gallagher, 2003; Baxter, Parker, Lindner, Izquierdo, & Murray, 2000). According to one leading theory, the amygdala may form the associative values of cues linked to outcomes, whereas the OFC is important for holding these values in memory and updating them with newly acquired information to guide decision-making (Pickens et al., 2003). On the basis of this theory, we hypothesized that OFC may integrate input signals carrying facial values processed mainly by the amygdala, synthesize more complex and sophisticated values, and, thus, serve as a primary brain center for mediating facial impression bias during social interactions.

To systematically examine the role of facial trustworthiness in biased social decision-making in an experimental setting, we utilized an ultimatum game, a popular economic experiment, involving social interactions between two players, that has been widely used to test the effect of perceived unfairness on economic decisions (Guth, Schmittberger, & Schwarze, 1982). A recent neuroimaging study demonstrated that the insula, a key neural structure implicated in encoding negative emotion triggered by aversive stimuli, such as pain and stress (Evans et al., 2002; Peyron, Laurent, & Garcia-Larrea, 2000), encodes perceived unfairness for a given offer from a proposer (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). In addition, the PFC has also been shown to play a critical role in regulating emotional reaction to unfair offers during ultimatum games and to link this perceived unfairness to decision-making regarding the acceptance or rejection of offers (Koenigs & Tranel, 2007; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Sanfey et al., 2003).

In this study, we sought to identify the neural structures involved in the integration of the neural signals of the proposer's facial trustworthiness that bias the responders' decision-making, as determined by changes in acceptance rates. All 12 participants played 60 ultimatum games with 60 proposers. Sixty faces with high, medium, and low trustworthiness (20 faces in each category) were selected on the basis of normative trustworthiness ratings acquired from a preliminary experiment and were used as the faces of proposers for the ultimatum game in the fMRI experiment. Participants were instructed that 1 of the 60 trials would be randomly chosen, that the money made during this trial would be added to the monetary compensation paid to each participant at the end of the experiment, and that the proposer of the trial would be paid in accordance with decisions made during the experiment. We estimated amounts of facial impression bias using a functional fitting method to search for an area within the OFC that correlated with facial impression bias on a trial-by-trial basis. In addition, a network centering on an ROI within the OFC was investigated by functional connectivity analysis to identify specific neural correlates underlying individual variability in facial impression bias during the ultimatum games.

METHODS

Participants

Fifteen right-handed male participants (mean age = 23.4 years, *SD* = 2.13 years) were recruited by advertising. Considering recent findings regarding fairness in judgment based on gender differences (Singer et al., 2006), only male participants were recruited to eliminate potential gender effects when evaluating facial trustworthiness during ultimatum decisions. Three subjects were excluded for their failure to follow instructions. Therefore, the fMRI data of 12 subjects were subject to analysis. All participants were carefully screened for chronic mental illness and MR compatibility, and the present experiment was approved by the institutional review board of Korea University. All participants provided informed consent before study commencement.

Experimental Procedures

Preliminary Experiment

We ran a preliminary experiment to reliably select faces with high, medium, and low trustworthiness from among 355 photographs (black and white) of Korean men. In this experiment, we used Cogent 2000 stimulus presentation software (www.vislab.ucl.ac.uk/Cogent/) running on MATLAB 6.5, and all 355 faces were sequentially presented on a single computer screen for an unlimited time. Twenty-six participants evaluated each face on a 9-point scale (1 = *not trustworthy at all*, 9 = *very trustworthy*) shown below each face. For each photograph, we combined the rating results of all participants and obtained a normative (or mean) trustworthiness rating by transforming the results into *Z* scores and averaging across the 26 subjects. On the basis of the normative ratings of the 355 faces, we selected 60 faces (20 faces in each category) with high, medium, or low trustworthiness, and these 60 faces were used as proposers in the ultimatum game. The *Z* scores of face trustworthiness ratings obtained from the preliminary study were distributed as follows: low trustworthiness, −1.15 to −0.69 (mean = −0.859); medium trustworthiness, 0.04 to 0.07 (mean = 0.057); and high trustworthiness, 0.77 to 1.94 (mean = 1.024). One-way ANOVA confirmed a highly significant difference in mean *Z* scores between these three categories, $F(2, 57) = 499.6, p < .01$.

fMRI Experiment

Subjects received instructions about the ultimatum game before scanning, and it was further confirmed that all subjects fully understood the requirements of the game. All the participants played the role of responder during the ultimatum games. Participants were told that proposers were randomly selected from a list of university alumni and asked to respond to a written question that asked what fraction of KRW 10,000 (approximately USD 8) they would offer as proposers in the ultimatum game. Participants were also told that proposers would receive their shares via mail in accordance with the actual decisions they made during the ultimatum games.

Each participant played 60 single-shot ultimatum games in total. For each game, a proposer made an offer (a fraction of KRW 10,000) to the participant. Each game began by displaying the photograph of the proposer, and after 2–5 sec, an offer amount made by the proposer was presented below the photograph. The photograph and the offer amount then disappeared 2–5 sec after the offer amount was shown. This was followed by a display that asked participants to press one of two buttons to either accept or reject the offer (Figure 1A). According to the rules of ultimatum game, when a responder accepts an offer, the offer amount is added to the responder's account, whereas no money is added to either the responder's or proposer's account when an offer is rejected. Participants were told that 1 of the 60 trials would be randomly chosen, and the result of this trial would be added to the final monetary compensation paid at the end of the experiment.

Proposer's offer amounts ranged from KRW 1000 to KRW 5000 in increments of KRW 1000. For each category of facial trustworthiness, participants were given five trials with an offer of KRW 5000, five trials with KRW 4000, four trials with KRW 3000, three trials with KRW 2000, and three trials with KRW 1000. To minimize correlation between offer amounts and facial trustworthiness, the half of the faces in each face category were assigned to fair offers (KRW 4000 and KRW 5000) and the other half to unfair offers (KRW 1000, KRW 2000, and KRW 3000), and this assignment was counterbalanced across subjects.

Neuroimaging Procedures

fMRI Data Acquisition

Functional imaging was conducted on a 3-T Forte MRI scanner (ISOL Technology, Gwangju, South Korea; Oxford OR63) at Korea Advanced Institute of Science and Technology in Daejeon, Korea. Whole-brain T1-weighted anatomical scans (256 × 256 voxels, 1 × 1 × 1 mm resolution, 176 axial slices) and gradient-echo T2*-weighted EPIs with BOLD contrast (64 × 64 voxels, repetition time = 2000 msec, echo time = 30 msec, flip angle = 80°, 3 × 3 mm in-plane resolution, slice thickness = 4 mm, 24 oblique axial slices with no gap) were acquired of each participant. Each scan lasted approximately 10 min, depending on performance. The first three volumes of images were discarded to allow for equilibration effects.

Preprocessing

All image analyses were performed using SPM2 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, U.K.). Functional images were realigned to the first volume to correct for subject motion, spatially normalized to a standard T2* template with a resampled voxel size of 3 mm, and spatially smoothed using a Gaussian kernel with a FWHM of 6 mm.

Calculating Individual Model Estimation Errors

With the specific aim of identifying neural structures encoding decision biases caused by the proposer's facial trustworthiness, we reasoned that responders' decisions were the product of a complex computation that took into account facial trustworthiness and the offer amount. Therefore, to define trials during which facial trustworthiness

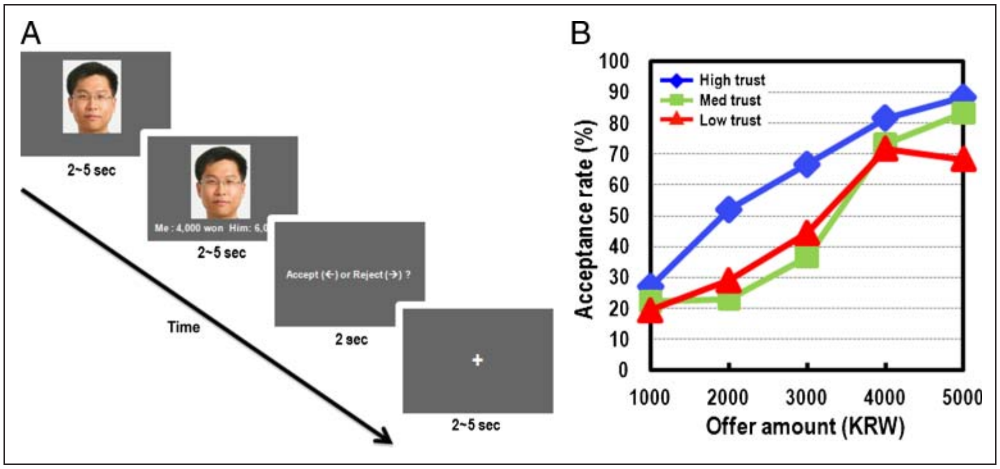

Figure 1. (A) A schematic diagram of the experimental design. Each trial began with a display of a proposer's face, and this was followed by an offer amount below the face. Participants were then asked to either accept or reject the offer by pressing appropriate keys. (B) Behavioral data showed a linearly increasing acceptance rate as a function of the increasing offer amount and that facial trustworthiness had a significant modulatory effect on acceptance rates.

biased decisions, we first estimated subject-specific decision probability models by estimating the optimal cubic functions that best fitted decisions as a function of the amount of payment offered by proposers (O'Doherty, Hampton, & Kim, 2007). Although we expected a sigmoid-like curve for decision probability as a function of offer amount, we tried to minimize any potential bias on the shape of the function and used a cubic function for fitting. However, as shown in Figure 2, it turned out that the model estimation resulted in individual decision probability curves with a sigmoid function shape for most of the participants, even without any preassumption about the shape of the function.

We assumed that the proposer's facial trustworthiness was a primary factor that contributed to any fitting errors between decisions and the estimated model, that is, the model estimation error (MEE). For example, one responder who showed no sign of facial impression bias by rejecting all offers of ≤KRW 2000 and accepting most of the others showed a near-perfect fitting between his decisions and the estimated decision probability curve (see Figure 2). The parameters of MEE were calculated by subtracting an estimated decision probability $f(x_i)$ for a given trial i with offer amount x from a decision D_i on a trial-by-trial basis (see equations below).

$$f(x_i) = ax_i^3 + bx_i^2 + cx_i + d$$

$$MEE(i) = D_i - f(x_i)\begin{cases} D_i = 1 : (accept) \\ D_i = 0 : (reject) \end{cases}$$

For example, for a trial with an offer of KRW 3000, if the estimated decision probability for the offer ($f(x_i)$) was 0.6 and the responder accepted the offer ($Di = 1$), then the MEE for this trial would be $1 - 0.6 = 0.4$. During individual fMRI data analysis, the parameters of MEE were convolved with a hemodynamic response function time-locked to the onset times of offer amounts and entered into a regression analysis against fMRI data. Although we focused on the onset of the offer display for the analyses, on the basis of the assumption that it is the time when the responders' decisions begin to be formed, there is a possibility that some regions showing activity at these events may simply encode information about facial trustworthiness and offer amounts. Therefore, to eliminate

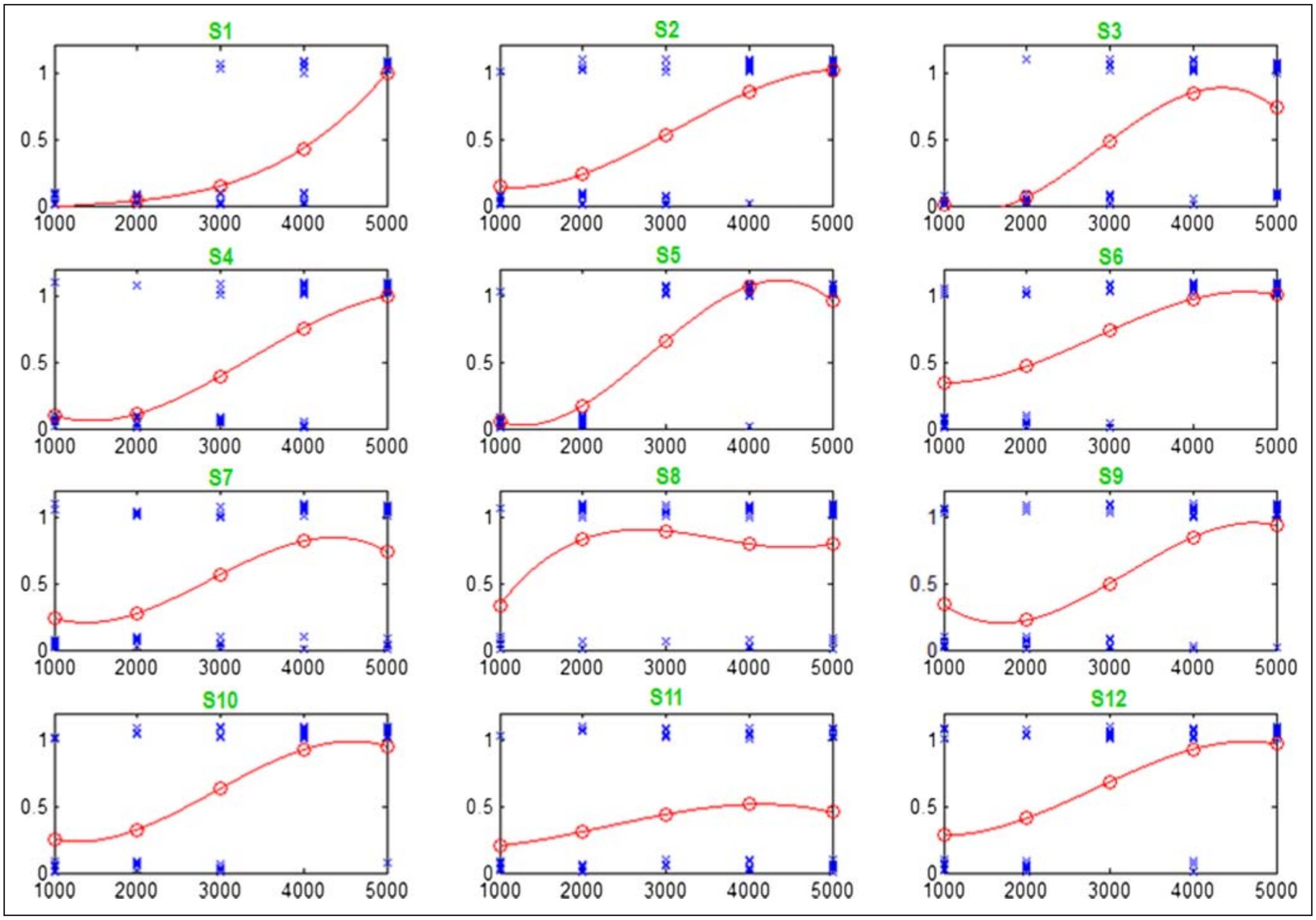

Figure 2. Subject-specific decision probabilities (red lines) estimated for each offer amount using optimal cubic functions best fitted to responders' decisions as a function of the amount offered by proposers. Subject numbers start from top left. The x axis indicates the amounts offered, and the y axis indicates decision probabilities (0 = *reject*, 1 = *accept*).

any brain regions simply correlated with normative facial trustworthiness ratings or decision probability for a given offer amount, we orthogonalized MEE parameters with respect to the parameters of facial trustworthiness and $f(x_i)$, the latter of which correlated with offer amount, to observe brain regions correlated uniquely with MEE. In addition, the six scan-to-scan motion parameters produced during realignment were entered into a regression analysis against fMRI data for individual participants to account for residual effects of movement. The results of each participant were taken to a random effects level by including the coefficient images of the MEE parameter obtained from each participant into a one-sample t test.

Determining Significance Level

Given our specific a priori hypothesis regarding the role of the OFC, we imposed a significance threshold of $p < .05$, corrected for multiple comparisons over the OFC volume (~17,900 mm^3; Kennedy et al., 1998), as determined by Monte Carlo simulations implemented in AlphaSim within AFNI software that generates volumetric cluster sizes corresponding to alpha levels (Cox, 1996). For any other activation within the frontal cortex, we used a search volume that combined superior, middle, and inferior frontal cortices (~88,100 mm^3; Kennedy et al., 1998). For all other ROIs, we restricted search volumes of AlphaSim to spheres with a radius of 10 mm (~4200 mm^3) centered on coordinates obtained in previous studies. More specifically, previous studies using ultimatum games were used to obtain the coordinates of the insula and ACC (Sanfey et al., 2003) and of the midbrain (Tabibnia, Satpute, & Lieberman, 2008). Furthermore, the coordinates of the amygdala were acquired from a study on the implicit assessment of facial trustworthiness (Engell et al., 2007).

Psychophysiological Interaction Analyses

To identify the neural structures communicating with the lateral OFC (lOFC), we performed psychophysiological interaction (PPI) functional connectivity analyses (Friston et al., 1997). The PPI analysis assesses temporal correlations between brain regions and their dependencies on experimental conditions and can be used to investigate functional coupling between two brain regions modulated by the experimental conditions. For PPI analyses, we extracted a deconvolved time series (Gitelman, Penny, Ashburner, & Friston, 2003) of the lOFC from each subject, and these showed significant correlations with MEE using a 5-mm-radius sphere around the peak voxel defined by the random-effects group analysis. Individual PPI general linear models contained three regressors: the deconvolved time series data from lOFC, the MEE parameter, and the product of the two regressors, that is, the PPI term. Each participant's coefficient image of the interaction term from the PPI analysis was then entered into the random-effects group analysis.

Correlation between Individual Variability of Facial Impression Bias and Functional Connectivity

Logistic regression analysis was carried out using two variables (i.e., normative facial trustworthiness ratings and proposers' offer amounts) to estimate the degree to which participants' decisions were influenced by the facial trustworthiness of proposers (β_{face}). Next, we regressed each participant's coefficient image of the interaction term from PPI analyses against the β coefficients of the facial trustworthiness predictor variable obtained from individual logistic regression analysis to investigate the neural structures showing differential functional connectivity with lOFC as a function of individual difference in facial impression bias (Etkin, Egner, Peraza, Kandel, & Hirsch, 2006).

RESULTS

Behavioral Results

Consistent with many previous studies based on ultimatum games, participants showed higher acceptance rates for fair offers and decreasing acceptance rates as offers became less fair. In addition, participants show slightly higher acceptance rates for offers from more trustworthy proposers. Behavioral data analyzed by two-way repeated measures ANOVA revealed that both facial trustworthiness ($F(2, 22) = 7.544, p < .01$) and offer amounts ($F(4, 44) = 12.043, p < .01$) had significant influences on acceptance rates (Figure 1B). No significant interaction effect was observed between facial trustworthiness and offer amount ($F(8, 88) = 0.993, p > .1$), but the effect of facial trustworthiness appeared to be more prominent for offer amounts of KRW 2000 and KRW 3000.

Subject-specific decision probabilities for each offer amount were estimated by computing optimal cubic functions best fitting individual responders' decisions as a function of the amounts offered. The parameters of MEE were then calculated by subtracting the estimated decision probabilities from actual decisions on a trial-by-trial basis, separately, for each responder (Figure 3A; see Methods for more details). The Jonckheere–Terpstra test for nonparametric trend analysis revealed a significant, linearly increasing trend in MEE as a function of an increase in normative facial trustworthiness ($JT = 3.77, p < .001$), thus verifying the contribution of facial trustworthiness to MEE (Figure 3B).

Neuroimaging Results

Brain Regions Correlated with MEE

For individual fMRI data analysis, the subject-specific parameters of MEE were convolved with a hemodynamic response function time-locked to the onset times of offer amounts

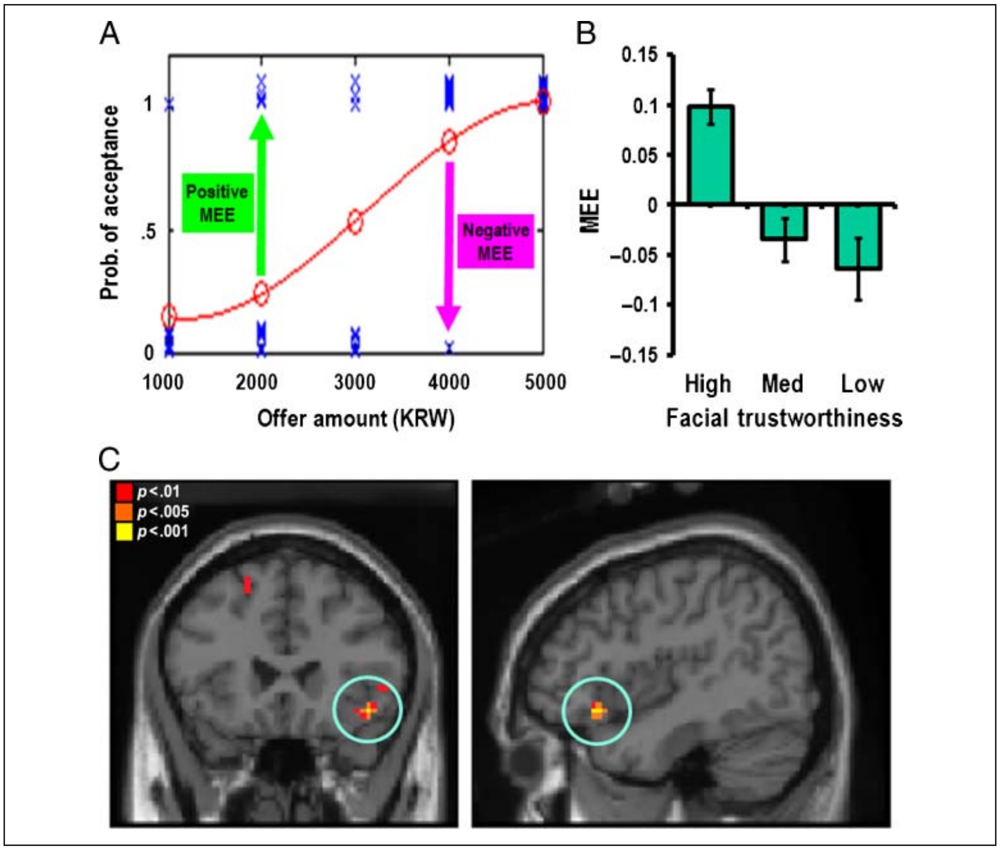

Figure 3. (A) The parameters of MEE were calculated by subtracting estimated decision probabilities from actual ultimatum decision values (blue Xs) on a trial-by-trial basis. (B) MEEs categorized by normative facial trustworthinesses of proposers. (C) Group analysis revealed that the activity of the lOFC was negatively correlated with MEE when offer amounts were revealed.

and entered into regression analysis against fMRI data. The application of the one-sample t test to individual subject regressor coefficient maps of the MEE parameter revealed the activity of the right lOFC (Figure 3C: $x = 45$, $y = 24$, $z = -12$, $Z = 3.57$) was correlated negatively with MEE, that is, activity increased as responder decisions deviated negatively from the estimated decision probability. In other words, lOFC showed elevated activity whenever responders rejected otherwise fair offers from proposers, possibly reflecting facial impression-related decision bias (see Figure 3B). In addition, positive correlations with MEE were found in the midbrain, possibly including the ventral tegmental area, which contains dopamine neurons (Figure 4A).

PPI Analysis Using lOFC as a Seed Point

We hypothesized that the lOFC integrates information about facial trustworthiness and offer amounts to compute complex signals that bias ultimatum decisions. To address how the lOFC creates signals biasing decisions, we examined the brain regions communicating with the lOFC using PPI functional connectivity analyses. Significant correlations with the PPI interaction term of deconvolved time series taken from the lOFC and negative MEE were found

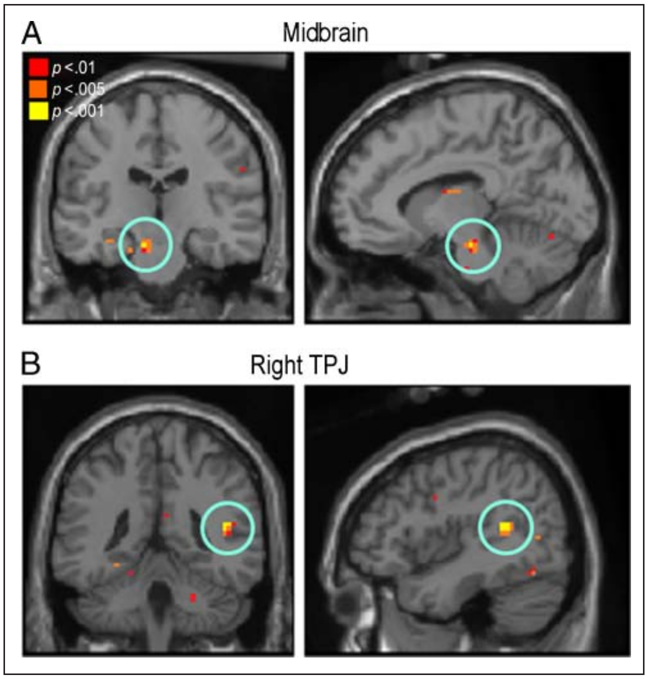

Figure 4. Brain regions that positively correlated with MEE were found in the midbrain (A: $x = -9$, $y = -21$, $z = -21$, $Z = 3.26$) and in the right TPJ (B: $x = 45$, $y = -45$, $z = 15$, $Z = 3.71$).

in the right posterior amygdala (Figure 5A: $x = 27$, $y = -6$, $z = -18$, $Z = 2.75$), the left anterior insula (Figure 5B: $x = -33$, $y = 15$, $z = -3$, $Z = 3.26$), the dorsal ACC (Figure 5C: $x = 9$, $y = 21$, $z = 39$, $Z = 3.24$), and the right dorsolateral PFC (Figure 5D: $x = 48$, $y = 9$, $z = 33$, $Z = 4.19$). These results suggest that the lOFC showed increased communications with these structures when subjects were more likely to reject otherwise fair offers because of the proposer's facial impressions (Table 1).

Neural Correlates of Individual Differences in Facial Impression Bias

We next considered that if the effect of facial trustworthiness on ultimatum decisions is critically dependent upon communication between the lOFC and the above-mentioned brain regions, then greater functional connectivity between them would be observed in responders who showed more decision bias because of facial trustworthiness. Thus, we performed logistic regression analyses with two predictor variables (i.e., normative facial trustworthiness ratings and offer amounts) and estimated the extent to which each responder's decisions were influenced by facial trustworthiness (β_{face}). Regressing each responder regressor coefficient map of the PPI interaction term against individual subjects' β_{face} revealed significant correlations in bilateral insula (Figure 6: left: $x = -42$, $y = 12$, $z = -9$, $Z = 3.71$; right: $x = 48$, $y = 9$, $z = -15$, $Z = 3.85$), indicating that people with greater facial-impression-related bias tend to show increased functional connectivity between bilateral insula and the OFC as their decisions deviated from model estimations.

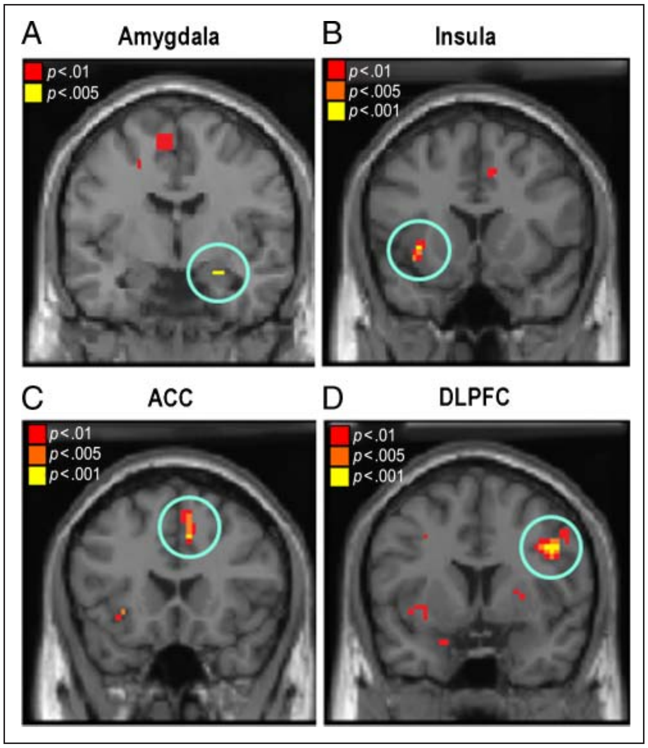

Figure 5. PPI analyses revealed significant MEE-modulated functional communications between the lOFC and the right amygdala (A: $x = 27$, $y = -6$, $z = -18$, $Z = 2.75$), the left insula (B: $x = -33$, $y = 15$, $z = -3$, $Z = 3.26$), the ACC (C: $x = 9$, $y = 21$, $z = 39$, $Z = 3.24$), and the right dorsolateral PFC (DLPFC; D: $x = 48$, $y = 9$, $z = 33$, $Z = 4.19$).

Table 1. List of All Brain Structures Exceeding the Threshold Determined by AlphaSim in This Study, except for $p < .001$ (Uncorrected) and $p < .005$ (Uncorrected)

Brain Region	*Z*	*Peak in MNI* *x*	*y*	*z*
Negative Correlation with Facial Trustworthiness at the Time of Face Display				
Anterior amygdala (R)	2.62*	21	0	−9
Correlation with Negative MEE at the Time of Offer Display				
LOFC (R)	3.57	45	24	−12
Correlation with Negative MEE at the Time of Offer Display				
Midbrain	3.26	−9	−21	−21
TPJ (R)	3.71**	45	−45	15
Regions Communicating with the lOFC				
Posterior amygdala (R)	2.75	27	−6	−18
Anterior insula (L)	3.26	−33	15	−3
ACC (R)	3.24	9	21	39
Dorsolateral prefrontal cortex (R)	4.19	48	9	33
Regions Communicating with the lOFC Modulated by Individual Difference in Decision Bias due to Facial Trustworthiness (β_{face})				
Anterior insula (R)	4.25	36	24	3
Anterior insula (L)	3.92	−45	3	21
Dorsal ACC (L)	3.49**	−12	9	45

L = left; R = right.

*$p < .005$ (uncorrected).

**$p < .001$ (uncorrected).

DISCUSSION

Consistent with previous literatures on first impression bias, this study demonstrates that an offer is more likely to be rejected when it is made by a less-trustworthy-looking proposer and that activity in the right lOFC predicts a trial-by-trial fluctuation of decision bias caused by the facial trustworthiness of the proposer. In addition, PPI analyses revealed that negative decision bias modulated the

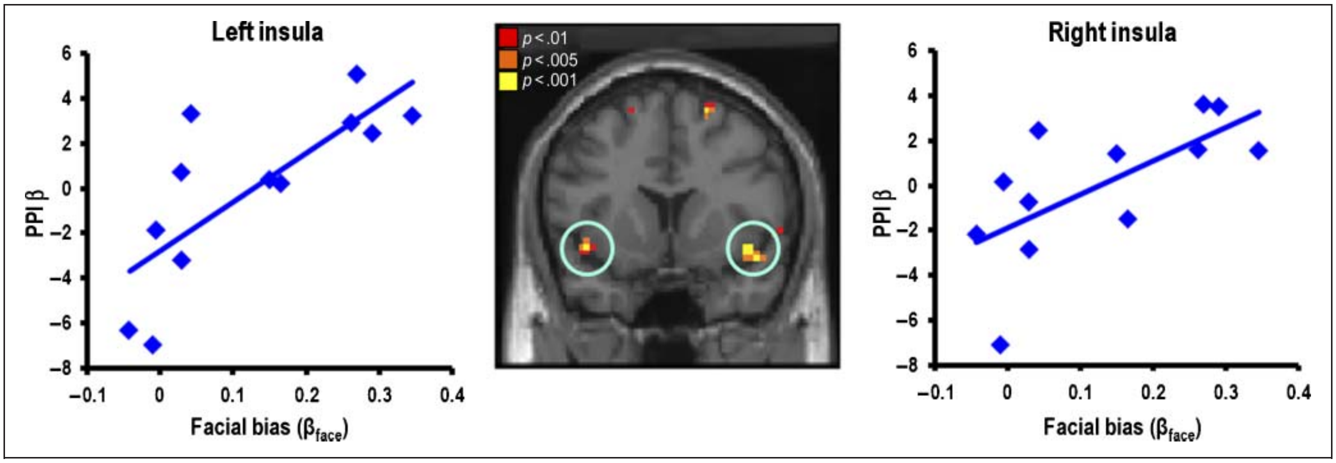

Figure 6. Individual variability in ultimatum decision bias on the basis of facial trustworthiness (β_{face}) predicted by MEE-modulated functional connectivity between the lOFC and the bilateral insula. Center: left, $x = -42, y = 12, z = -9, Z = 3.71$; right, $x = 48, y = 9, z = -15, Z = 3.85$).

functional communication between the lOFC and the amygdala and insula, which has been previously implicated in the encoding of facial trustworthiness and perceived unfairness, respectively. Finally, individual variability in the degrees to which decisions were biased because of facial trustworthiness was predicted by considering functional connectivity between the bilateral insula and the lOFC. In summary, the findings of this study provide evidence that the OFC plays a pivotal role in integrating various affectively charged subcortical input signals triggered by first impressions of strangers and biased economic decision-making in situations involving social interaction.

Facial Trustworthiness and Biasing of Ultimatum Decisions

As predicted, we confirmed the role of the proposer's facial trustworthiness in ultimatum decision-making by responders. This finding rather contrasts with that of a previous study that examined the role of facial attractiveness in ultimatum decisions and in which responders demanded higher minimum acceptable offers from attractive proposers (Solnick & Schweitzer, 1999). Of the many potential reasons that might have contributed to the observed discrepancy, perhaps the primary reason was that responders specified a minimum acceptance offer for each photograph in the previous study, whereas responders were asked to make actual ultimatum choices in this study. Perhaps, similar to a recent study, in which subtle environmental cues were found to influence moral decisions subconsciously (Schnall, Haidt, Clore, & Jordan, 2008), the criteria used to judge fairness may have been affected implicitly by facial trustworthiness in this study. Therefore, we believe that the decision biases caused by facial impressions in this study probably occurred automatically and that responders were not explicitly aware of the link between facial trustworthiness and their own choices in ultimatum games. It would be interesting for future studies to focus on the sources of the differences between studies that use self-reports on acceptable minimum offers and those that use ultimatum decisions.

Previous studies have indicated that facial trustworthiness and attractiveness tend to be correlated (van't Wout & Sanfey, 2008) and that attractive faces are more likely to be trusted (Andreoni & Petrie, 2008; Wilson & Eckel, 2006; Solnick & Schweitzer, 1999; Mulford et al., 1998), which is not surprising given that attractiveness has often been associated with various other positive traits (Langlois et al., 2000). We also examined the correlation between the facial attractivenesses and the facial trustworthinesses of the faces used in this study by asking an independent group of people to rate the same faces on an attractiveness scale. A highly significant correlation was found between the two ratings ($r = .94, p < .01$), which suggests that facial attractiveness potentially affected ultimatum decisions in this study. In fact, judgments of attractiveness and trustworthiness were found to be the two most rapidly made trait judgments (Willis & Todorov, 2006), and the extent to which the amygdala responds to variations of faces on specific trait dimensions was found to be a function of the general valence content of these dimensions (Engell et al., 2007). Therefore, the involvement of the amygdala in the modulation of the lOFC activity in response to the proposers' faces may not necessarily be due to facial trustworthiness per se but because of more general valence-related approach cues encompassing various types of traits.

The Role of the lOFC in the Biasing of Ultimatum Decisions due to Negative Facial Impressions

The OFC integrates signals from diverse brain structures involved in the processing of emotional information, such as the amygdala, the insula, and the temporal cortex, and has long been considered a key center for emotional and social decision-making (O'Doherty et al., 2001; Bechara et al., 2000; Cavada & Schultz, 2000; Schoenbaum et al., 1998). In particular, the lateral aspect of the OFC, observed in this study, has strong functional communication with the amygdala (Hsu, Bhatt, Adolphs, Tranel, & Camerer, 2005; Kim et al., 2004) and has been implicated in encoding negative valence in many previous studies (Elliott, Agnew, & Deakin, 2010; Kringelbach, 2005; Ursu & Carter, 2005; Kim et al., 2004; O'Doherty et al., 2001). Furthermore, the lOFC has been reported to encode aversive prediction errors (Seymour et al., 2005) to help organisms avoid certain aversive objects or help people avoid negative outcomes (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006) and to suppress previously learned desirable choices (Elliott et al., 2010; Wrase et al., 2007; Elliott, Dolan, & Frith, 2000). Of particular relevance to this study, the lOFC has been reported to be responsive to the negative aspect of facial stimuli, and its activity has been reported to be inversely correlated with facial attractiveness (Cloutier, Heatherton, Whalen, & Kelley, 2008; O'Doherty et al., 2003) to show an increased response to black versus white faces (Cunningham et al., 2004) and negative versus positive contextual information and to modulate amygdala responses to emotionally ambiguous faces (Kim et al., 2004). Consistent with these findings, we observed increased activity in the lOFC during the rejection of fair offers from untrustworthy looking proposers, which may have also required effective suppression of a natural impulse to maximize income by accepting any amount of money offered by proposers.

Anatomical data from animal studies suggest that strongest connections are between the posterior lOFC and the key brain regions critically involved in processing emotional information, such as the amygdala, the insula, and the temporal cortex (Augustine, 1996). Similarly, our PPI analysis findings suggest that the lOFC participates in the generation of signals that bias decisions via the integration of input signals from the amygdala and insula related to facial impressions. In addition, the lOFC appears to provide biasing signals to the dorsolateral PFC, which is supposedly one of the key neural structures linking perceived unfairness to final decisions during ultimatum games (Koenigs & Tranel, 2007; Knoch et al., 2006; Sanfey et al., 2003). The findings of this study extend our understanding of lOFC function by showing that the intimate functional connectivity between the neural components of the large-scale neural circuitry centered on the lOFC underlies first impression bias during social decision-making.

Contributions of the Amygdala and the Insula to lOFC Activity

When the proposers' faces were revealed at the onset of trials, we found that the right anterior amygdala activity ($x = 21$, $y = 0$, $z = -9$, $Z = 2.62$) was significantly correlated with normative facial trustworthinesses of the proposers' faces. This finding reinforces the notion that the amygdala participates in the encoding of facial trustworthiness (Engell et al., 2007; Winston et al., 2002; Adolphs et al., 1998), in biasing social evaluations/interactions based on facial impression (Schiller, Freeman, Mitchell, Uleman, & Phelps, 2009), and in biasing social decisions following pharmacological interventions (Baumgartner et al., 2008; Kosfeld et al., 2005). In light of a recent study on the role of the amygdala in encoding facial trustworthiness (Said, Baron, & Todorov, 2009), we examined the quadratic nature of amygdala response to facial trustworthiness but failed to replicate previous findings, probably because of the limited number of faces with high trustworthiness included in this study as compared with the intended inclusion of faces with high trustworthiness in the previous study.

Increasing numbers of theoretical and empirical studies on amygdala function indicate spatial and functional dissociation between the subregions of the human amygdala (Bach, Behrens, Garrido, Weiskopf, & Dolan, 2011; Gamer, Zurowski, & Buchel, 2010; Kim, Somerville, Johnstone, Alexander, & Whalen, 2003; Whalen et al., 2001). In this study, the anterior part of the amygdala was found to encode facial trustworthiness at the time when faces were revealed, whereas a PPI correlation with lOFC activity was detected in the posterior aspect of the amygdala when offers were shown. This raises the possibility that biologically significant social values detected in the proposers' faces may have been first detected and conveyed by the anterior amygdala to other brain regions involved in emotional and social decisions, including the posterior amygdala, the insula, and the OFC. These regions may then have communicated with each other to share the information related to facial impression and may also have generated more complex and sophisticated signals that biased decisions.

Given the strong reciprocal connections between the amygdala and the insula (Reep & Winans, 1982), the amygdala, which rapidly reads untrustworthiness in the proposers' faces, may have contributed to the activity in the insula, which is mainly responsible for encoding perceived unfairness on the basis of the amount of offer from a proposer in ultimatum games (Sanfey et al., 2003). It is noteworthy that, in this study, MEE-modulated functional connectivity between bilateral insula and the lOFC, but not between the amygdala and the lOFC, predicted individual variability in decision bias because of facial impressions, which is consistent with recent findings that insula activities reflecting perceived fairness vary across individuals (Singer et al., 2006; Sanfey et al., 2003). Intuitive facial-trustworthiness-related information, which was first detected by the amygdala and processed by the insula, would need to be converted into more concrete signals that are more directly available for guiding decisions, and our findings emphasize the role of the lOFC in such a process. It can be further inferred that individuals vary with respect to communication between the insula and the lOFC, which may be critical for constructing signals that bias decisions on the basis of facial impressions.

Areas Associated with Positive Facial Impression Bias

A correlation with positive MEE was found in the midbrain. This finding is consistent with that found in a recent study, in which similar midbrain activity was observed in response to high versus low fairness offers during ultimatum games (Tabibnia et al., 2008). It seems likely that the midbrain activity correlating with positive decision bias contributes to increase in perceived fairness when receiving offers from trustworthy-looking proposers. A correlation with a positive MEE was also observed at a lenient threshold ($p < .001$, uncorrected) in the right TPJ (Figure 4), which is a major brain region required for mentalizing and inferring the intentions of others (Ruby & Decety, 2003; Saxe & Kanwisher, 2003). In this study, no significant correlation with positive MEE was found in the PFC, which is probably because accepting as opposed to rejecting an offer, albeit unfair, may be a relatively easier default response in the ultimatum game. This possibility could perhaps be tested in a follow-up study using a more carefully designed behavioral task, where positive and negative decision biases are appropriately balanced in terms of required psychological effort.

Conclusions

Despite considerable research on the effect of first impressions on social decision-making, our understanding of the psychological and neural mechanisms underlying this behavior is surprisingly limited. This study shows that model-based fMRI (O'Doherty et al., 2007) can be useful to quantitatively define decision bias caused by facial impressions, because of its ability to accurately estimate targeted psychological states on a trial-by-trial basis based on the observed behavior in individual subjects. In summary, this study reveals, for the first time, the neural mechanism centered on the right lOFC, whereby the facial trustworthinesses of proposers influence the acceptance rates of responders in ultimatum games, perhaps by implicitly modulating responders' perceived unfairness. Supporting the novel idea that the main function of human reasoning is primarily argumentative (Mercier & Sperber, 2011) rather than the correction of misguided intuition (Kahneman, 2003), the present study may also provide neurobiological evidence that human reasoning, including judgments of fairness and morality, has its roots in intuitive inferential processing, which is typically subserved by evolutionarily primitive subcortical structures.

Acknowledgments

We thank Ming Hsu for his comments on the early version of the article and Shin Han and Young-Jae Hong for technical assistance. In addition, we also thank the two anonymous reviewers for their helpful comments. This study was supported by the Cognitive Neuroscience Program of the Korean Ministry of Science and Technology (M10644020003-06N4402-00310) and the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2011-327-H00038).

Reprint requests should be sent to Hackjin Kim, Department of Psychology, Korea University, Anam-Dong 5-Ga, Seongbuk-Gu, Seoul, 136-701, South Korea, or via e-mail: hackjinkim@korea.ac.kr.

REFERENCES

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

Andreoni, J., & Petrie, R. (2008). Beauty, gender, and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology, 29,* 73–93.

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research, Brain Research Reviews, 22,* 229–244.

Bach, D. R., Behrens, T. E., Garrido, L., Weiskopf, N., & Dolan, R. J. (2011). Deep and superficial amygdala nuclei projections revealed in vivo by probabilistic tractography. *Journal of Neuroscience, 31,* 618–623.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58,* 639–650.

Baxter, M. G., Parker, A., Lindner, C. C., Izquierdo, A. D., & Murray, E. A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *Journal of Neuroscience, 20,* 4311–4319.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex, 10,* 295–307.

Cavada, C., & Schultz, W. (2000). The mysterious orbitofrontal cortex. Foreword. *Cerebral Cortex, 10,* 205.

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. *Journal of Cognitive Neuroscience, 20,* 941–951.

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

Cunningham, W. A., Johnson, M. K., Raye, C. L., Chris Gatenby, J., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15,* 806–813.

Daw, N., O'Doherty, J., Dayan, P., Seymour, B., & Dolan, R. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441,* 876–879.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24,* 285–290.

Dipboye, R. L., Arvey, R. D., & Terpstra, D. E. (1977). Sex and physical attractiveness of raters and applicants as determinants of resume evaluations. *Journal of Applied Psychology, 62,* 288–294.

Downs, A. C., & Lyons, P. M. (1991). Natural observations of the links between attractiveness and initial legal judgments. *Personality & Social Psychology Bulletin, 17,* 541–547.

Elliott, R., Agnew, Z., & Deakin, J. F. (2010). Hedonic and informational functions of the human orbitofrontal cortex. *Cerebral Cortex, 20,* 198–204.

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex, 10,* 308–317.

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

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., & Hirsch, J. (2006). Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron, 51,* 871–882.

Evans, K. C., Banzett, R. B., Adams, L., McKay, L., Frackowiak, R. S., & Corfield, D. R. (2002). BOLD fMRI identifies limbic, paralimbic, and cerebellar activation during air hunger. *Journal of Neurophysiology, 88,* 1500–1511.

Frieze, I. H., Olson, J. E., & Russell, J. (1991). Attractiveness and income for men and women in management. *Journal of Applied Social Psychology, 21,* 1039–1057.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage, 6,* 218–229.

Gamer, M., Zurowski, B., & Buchel, C. (2010). Different amygdala subregions mediate valence-related and attentional effects of oxytocin in humans. *Proceedings of the National Academy of Sciences, U.S.A., 107,* 9400–9405.

Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *Neuroimage, 19,* 200–207.

Green, S. K., Buchanan, D. R., & Heuer, S. K. (1984). Winners, losers, and choosers. *Personality & Social Psychology Bulletin, 10,* 502–511.

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3,* 367–388.

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

Hampton, A. N., Adolphs, R., Tyszka, M. J., & O'Doherty, J. P. (2007). Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron, 55,* 545–555.

Hatfield, E., Aronson, V., Abrahams, D., & Rottmann, L. (1966). Importance of physical attractiveness in dating behavior. *Journal of Personality and Social Psychology, 4,* 508–516.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science, 310,* 1680–1683.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697–720.

Kennedy, D. N., Lange, N., Makris, N., Bates, J., Meyer, J., & Caviness, V. S., Jr. (1998). Gyri of the human neocortex: An MRI-based analysis of volume and variance. *Cerebral Cortex, 8,* 372–384.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *NeuroReport, 14,* 2317–2322.

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M., et al. (2004). Contextual modulation of amygdala responsivity to surprised faces. *Journal of Cognitive Neuroscience, 16,* 1730–1745.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314,* 829–832.

Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the ultimatum game. *Journal of Neuroscience, 27,* 951–956.

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature, 435,* 673–676.

Kringelbach, M. L. (2005). The human orbitofrontal cortex: Linking reward to hedonic experience. *Nature Reviews Neuroscience, 6,* 691–702.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126,* 390–423.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34,* 57–74; discussion 74-111.

Mulford, M., Orbell, J., Shatto, C., & Stockard, J. (1998). Physical attractiveness, opportunity, and success in everyday exchange. *American Journal of Social Psychology, 103,* 1565–1592.

O'Doherty, J. P., Hampton, A. N., & Kim, H. (2007). Model-based fMRI and its application to reward-learning and decision making. *Annals of the New York Academy of Sciences, 1104,* 35–53.

O'Doherty, J. P., Kringelbach, M., Rolls, E., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4,* 95–102.

O'Doherty, J. P., Winston, J., Critchley, H., Perrett, D., Burt, D., & Dolan, R. (2003). Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia, 41,* 147–155.

Peyron, R., Laurent, B., & Garcia-Larrea, L. (2000). Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Neurophysiologie Clinique, 30,* 263–288.

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C., & Schoenbaum, G. (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *Journal of Neuroscience, 23,* 11078–11084.

Reep, R. L., & Winans, S. S. (1982). Afferent connections of dorsal and ventral agranular insular cortex in the hamster *Mesocricetus auratus*. *Neuroscience, 7,* 1265–1288.

Ruby, P., & Decety, J. (2003). What you believe versus what you think they believe: A neuroimaging study of conceptual perspective-taking. *European Journal of Neuroscience, 17,* 2475–2480.

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

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., & Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300,* 1755–1758.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage, 19,* 1835–1842.

Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. A. (2009). A neural mechanism of first impressions. *Nature Neuroscience, 12,* 508–514.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality & Social Psychology Bulletin, 34,* 1096–1109.

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neuroscience, 1,* 155–159.

Schoenbaum, G., Setlow, B., Saddoris, M. P., & Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron, 39,* 855–867.

Seymour, B., O'Doherty, J., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., et al. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neuroscience, 8,* 1234–1240.

Sigall, H., & Ostrove, N. (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of crime on juridic judgment. *Journal of Personality and Social Psychology, 31,* 410–414.

Singer, T., Seymour, B., O'Doherty, J., Stephan, K., Dolan, R., & Frith, C. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439,* 466–469.

Solnick, S. J., & Schweitzer, M. E. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Process, 79,* 199–215.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science, 19,* 339–347.

Ursu, S., & Carter, C. (2005). Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: Implications for neuroimaging studies of decision-making. *Brain Research, Cognitive Brain Research, 23,* 51–60.

van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108,* 796–803.

Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., Wright, C. I., & Rauch, S. L. (2001). A functional MRI study of human amygdala responses to facial expressions of fear versus anger. *Emotion, 1,* 70–83.

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

Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly, 59,* 189–202.

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

Wrase, J., Kahnt, T., Schlagenhauf, F., Beck, A., Cohen, M., Knutson, B., et al. (2007). Different neural systems adjust motor behavior in response to reward and punishment. *Neuroimage, 36,* 1253–1262.