# Brain Responses to the Acquired Moral Status of Faces

Tania Singer,* Stefan J. Kiebel, Joel S. Winston,
Raymond J. Dolan, and Chris D. Frith
Wellcome Department of Imaging Neuroscience
Institute of Neurology
University College of London
London WC1N 3BG
United Kingdom

## Summary

We examined whether neural responses associated with judgments of socially relevant aspects of the human face extend to stimuli that acquire their significance through learning in a meaningful interactive context, specifically reciprocal cooperation. During fMRI, subjects made gender judgments on faces of people who had been introduced as fair (cooperators) or unfair (defector) players through repeated play of a sequential Prisoner's Dilemma game. To manipulate moral responsibility, players were introduced as either intentional or nonintentional agents. Our behavioral (likability ratings and memory performance) as well as our imaging data confirm the saliency of social fairness for human interactions. Relative to neutral faces, faces of intentional cooperators engendered increased activity in left amygdala, bilateral insula, fusiform gyrus, STS, and reward-related areas. Our data indicate that rapid learning regarding the moral status of others is expressed in altered neural activity within a system associated with social cognition.

## Introduction

Human evolution can be viewed in terms of the species' increasing ability to function effectively within a social context. By consequence, our brain has evolved a specialized ability for social cognition ("the processing of information which culminates in the accurate perception of the dispositions and intentions of other individuals" [Brothers, 1990; page 28]). Brothers proposed that the amygdala, the orbitofrontal cortex (OFC), inferotemporal face-responsive regions, and superior temporal sulcus (STS) represent areas primarily involved in the processing of socially relevant information. Adolphs (2003) extended this proposal differentiating higher-order sensory cortices such as fusiform gyrus and superior temporal sulcus involved in detailed perceptual processing with the amygdala, ventral striatum, and orbitofrontal cortex linking sensory representations of stimuli to their motivational value. Anterior cingulate cortex as well as insular-somatosensory cortices have also been associated with feeling states that reflect representation of changes in bodily states arising from processing emotion-eliciting stimuli (e.g., Adolphs, 1999; Critchley et al., 2000, 2001; Damasio, 1994; Dolan, 2002).

Functional imaging studies using emotional and so-cially relevant stimuli provide evidence for the importance of the above proposed neural circuitry for social cognition. For example, activation in posterior STS is seen during perception of biological motion as well as during more abstract tasks involving inferences about intentions, beliefs, and feelings of other persons (for reviews, see Allison et al., 2000; Frith and Frith, 2003; Puce and Perrett, 2003). Further evidence for this circuitry derives from imaging studies focusing on the processing of socially salient features in the human face such as emotional expressions (Morris et al., 1996; Phillips et al., 1997; Winston et al., 2003), facial attractiveness (Aharon et al., 2001; O'Doherty et al., 2003), trustworthiness (Winston et al., 2002), or racial identity (Hart et al., 2000; Phelps et al., 2000; Phelps, 2001). These studies highlight engagement of the amygdala, dorsal and ventral striatum, orbitofrontal cortex (OFC), insula, and higher-order visual areas such as fusiform face areas and superior temporal sulcus (STS) in deliberative and implicit social judgments.

In everyday situations, individuals base their social judgment not only on features inherent in the human face, but also on inferences about the other people arising from socially meaningful contexts. One salient example is the degree of perceived cooperativeness of another person. Reciprocal exchange or cooperation is ubiquitous in behavior among primates. In consequence, the ability to recognize and discriminate against those who do not reciprocate cooperation is crucial for effective social interactions, allowing us to know who to approach or avoid (see also Cosmides and Tooby, 2000).

Analytical models and computer simulations developed in the fields of theoretical biology and economics provide evidence of evolutionary advantage in the emergence of mutual cooperation or "direct reciprocity" (Axelrod, 1984; Axelrod and Hamilton, 1981; or "reciprocal altruism," e.g., Trivers, 1971) as well as "indirect reciprocity" (e.g., Alexander, 1987; Nowak and Sigmund, 1998a). The possibility of reputation formation or "image scoring" in individuals has also been suggested to contribute to the emergence of indirect reciprocity in human societies (Milinski et al., 2002; Nowak and Sigmund, 1998b; Nowak et al., 2000; Wedekind and Milinski, 2000).

Empirical data from social dilemma experiments suggest that people have strong aversions against being deceived when playing with others for monetary payoffs. For example, individuals involved in Public Good Games are willing to punish noncooperators or "free-riders" even if punishment is costly and does not provide any benefits for the punisher (e.g., Fehr and Gächter, 2000, 2002). In accordance with theories of inequity aversion (Fehr and Schmidt, 1999), there is even evidence that monkeys also reject unfair offers and react negatively when observing a conspecific obtaining a more attractive reward for equal effort (Brosnan and De Waal, 2003). Neural evidence for inequity aversion in humans has been provided in a fMRI study showing enhanced activity in brain areas related to emotions (bilateral anterior insula and ACC) in response to unfair offers of human partners in the Ultimatum Game (Sanfey et al., 2003). In

*Correspondence: t.singer@fil.ion.ucl.ac.uk

contrast, a study by Rilling et al. (2002), in which people played a version of a simultaneous Prisoner's Dilemma Game, showed enhanced activity in reward-related areas such as striatum, rostral ACC, and OFC during mutual cooperation. Striatum and rostral ACC were activated solely when playing with human but not computer partners (though the interaction was not explicitly tested). Reward-related areas such as ventral striatum and OFC as well as amygdala were identified in other fMRI studies using financial rewards (Breiter et al., 2001; Delgado et al., 2000; Elliott et al., 2000, 2003; Knutson et al., 2000; O'Doherty et al., 2001), in line with animal literature on reward processing (e.g., Schultz, 2000).

In the present study, we used event-related functional magnetic resonance imaging (fMRI) to investigate implicit social judgments of people who differ in their acquired moral status. Moral status was here defined by differential histories of fair or unfair behavior during social interactions. Thus, and in contrast to previous imaging studies that have mainly focused on social judgments of unfamiliar faces with regard to emotional expression, attractiveness, race, or trustworthiness (e.g., Hart et al., 2000; O'Doherty et al., 2003; Morris et al., 1996; Winston et al., 2002), we aimed to investigate processing of socially relevant cues that acquired their significance through learning in an interactive context. To achieve this goal, subjects were scanned while they made gender judgments of faces of people previously familiarized through repeated playing of a version of a sequential Prisoner's Dilemma game. In this game, one person ("first mover," who was always the scanned participant) has to decide whether to cooperate or to defect, and in a second step the other person (the "second mover" identified by a facial photograph) decides whether to reciprocate the first movers' trust or to defect. In terms of monetary payoffs, mutual cooperation is better than mutual defection. However, after a cooperative first mover's choice, the best payoff for the second mover is given through defection, so that the second mover has a strong incentive to betray the trust of the first mover. The social learning phase consisted of repeated interactions with people who either reciprocated the subject's trust (cooperator faces), never reciprocated trust (defector faces), or participated in Null Games (neutral faces). To disentangle the effects due to the moral status of a person from effects of monetary reward, half of the players were introduced as intentional agents (intentional task condition), whereas the other half were said to follow a given response sequence (nonintentional task condition). On theoretical and empirical grounds, we predicted that the face of a cooperative or defecting person would elicit "automatic" emotional responses that reflect evaluative processes. Furthermore, we predicted that discrete brain circuitry related to social cognition including the amygdala, orbitofrontal cortex, striatum, insula, fusiform gyrus, and superior temporal sulcus would be implicated in the perception of defectors (cheaters) as well as cooperators (friends), and especially in the perception of faces introduced as agents with moral choice (intentional condition) as compared to nonintentional agents.

## Results

### Behavioral Data
Following the scanning session, we assessed several behavioral measures. We report behavioral analyses based on the pooled sample of the behavioral pilot and the scanned sample (n = 21). Statistical tests revealed no significant differences between the behavioral patterns observed in these two samples. Results of a memory task are reported for the scanned sample as the format of the task was changed after the pilot by adding a "don't know" option (see Experimental Procedures).

### Involvement in Intentional versus Nonintentional Task Condition
Analyses of the debriefing questionnaire (Table 1) revealed that subjects (n = 21) rated themselves as being more emotionally involved with [mean ratings = 1.5 versus −.23; paired t test, $t(20) = 8.21$, $p < 0.001$] and more angry about [mean ratings = 1.3 versus −1.0; paired t test, $t(20) = 8.63$, $p < 0.001$] defectors in the intentional compared to the nonintentional task condition. In addition, subjects rated the playing partners to be more real in the intentional compared to the nonintentional task condition [mean ratings = 1.33 versus .04; paired t test, $t(20) = 4.37$, $p < 0.001$]. These differences are summarized in a composite measure (mean over the three questions) reflecting overall involvement in both task conditions [mean ratings = 1.38 versus −.39; paired t test, $t(20) = 9.02$, $p < 0.001$] and displayed for each of the three questions in Figure 1A.

### Likeability Ratings
A repeated measures analysis of variance with the two within subject factors "social behavioral status" (cooperators, neutral, defectors) and "task condition" (intentional, nonintentional) revealed a significant main effect for social behavioral status [$F(2,40) = 48.68$; $p < 0.001$]. No significant main effect of task condition was evident, but there was a significant interaction between social behavioral status and task condition [$F(2,40) = 7.46$; $p = 0.002$]. Post hoc contrasts confirmed that faces of cooperators were rated as more likeable than neutral faces [$F(1,20) = 18.18$; $p < 0.001$], and that faces of defectors were rated as less likeable than neutral faces [$F(1,20) = 36.97$; $p < 0.001$]. As illustrated in Figure 1B, the significant interaction was based on a significant difference between defectors and neutral faces in intentional compared to the nonintentional task condition [$F(1,20) = 15.48$; $p = 0.001$].

### Memory Performance
As illustrated in Figure 1C, memory for cooperator and defector faces was better than for neutral faces. A repeated measures analysis of variance with the two within subject factors "social behavioral status" (cooperators, neutral, defectors) and task condition (intentional, nonintentional) revealed a main effect for social behavioral status [$F(2,20) = 14.06$; $p < 0.001$], no main effect of task condition [$F(1,10) = 3.747$; $p = 0.08$], and a significant interaction between social behavioral status and task condition [$F(2,20) = 3.87$; $p = 0.03$]. Post hoc contrasts revealed that the significant main effect of social behavioral status was based on better memory for cooperator compared to neutral faces [$F(1,10) = 33.557$; $p < 0.001$]. Collapsed over both task conditions, there were no sig-

Table 1. Debriefing Questionnaire

Following rating scale was given for questions 1, 2, 3, and 5.

| | −2 | −1 | 0 | +1 | +2 |
|---|---|---|---|---|---|
| | not at all | | | | very much |

1. How interesting did you find the whole study?
2. How much have you been emotionally involved during the run in which
        (a) persons were free to decide?
        (b) persons were forced to decide?
3. How angry have you been about noncooperating players in the run in which
        (a) persons were free to decide?
        (b) persons were forced to decide?
4. In which aspects did both runs differ for you (e.g., application of different playing strategies, different attentional focus, different emotional involvement, different feelings, different thoughts)?
5. How real (animated) were the persons you were playing within the run in which
        (a) persons were free to decide?
        (b) persons were forced to decide?
6. Did you believe that your playing partners were real living persons?
        Yes        No

nificant differences of memory performance for defectors and neutral faces [$F(1,10) = 2.22$; $p = .167$]. This lack of difference is explained by a significant interaction reflecting differences for defectors and neutral faces in intentional compared to the nonintentional task condition [$F(1,10) = 5.21$; $p = 0.04$]. Although memory for cooperator faces was better in the intentional compared to nonintentional task condition [$t(10) = 2.63$, $p = .02$], the interaction for cooperator compared to neutral faces did not attain significance.

## Neuroimaging Results
### Effects of Social Behavioral Status
The comparison between affective (cooperators and defectors) and neutral (null games) faces revealed significant effects in ventromedial prefrontal cortex, left insula, left amygdala, and left fusiform/lateral inferior temporal gyrus (for details, see Table 2).

To identify areas with greater responses to cooperator relative to neutral faces, we compared cooperator and neutral faces, collapsed over task conditions. As shown in Figure 2, significant predicted effects were evident in left insula extending to OFC, left amygdala, and left putamen (for details, see Table 2).

The comparison between defector and neutral faces collapsed over task conditions revealed significant effects in ventro-medial prefrontal cortex (for details, see Table 2).

### Social Effects: Intentional Compared to Nonintentional Task Condition
To determine effects specific to the intentional compared to nonintentional task condition (e.g., due to com-
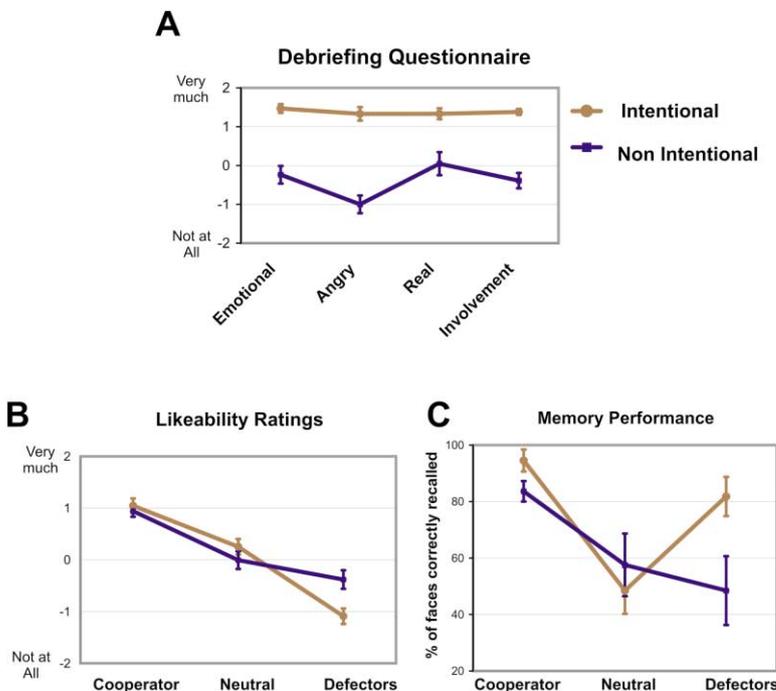


Figure 1. Postscan Behavioral Assessment of Emotional Involvement as well as Likeability Ratings and Memory Performance for Cooperators, Defectors, and Neutral Faces in Intentional and Nonintentional Task Condition

(A) The subjective responses provided by the subjects when asked (a) how much they felt emotionally involved; (b) how angry they were about noncooperating players; and (c) how real and animated the players appeared in the intentional and nonintentional task condition. Subjects used a rating scale ranging from $+2$ = very much, 0 = neutral, through to $-2$ = not at all.

(B) The subjective likeability ratings for faces of cooperators, defectors, and neutral in intentional and nonintentional task condition, using a scale ranging from $+2$ = very likeable, 0 = neutral, through to $-2$ = not likeable at all.

(C) Percentage of cooperator, defector, and neutral faces correctly recalled in the intentional and nonintentional task condition. Subjects were asked whether the presented person had "mostly cooperated" and could respond with "yes" (in case of cooperators), "no" (in case of defectors), "null" (in case of people involved in Null games), or "don't remember."

Table 2. Cerebral Foci of Activation for Contrasts of Interest

| Brain Regions | Coordinates of Peak Activation (mm) | | | |
| --- | --- | --- | --- | --- |
| | x | y | z | Z Scores |
| Affective-Neutral Faces | | | | |
| Ventro-medial OFC | 0 | 51 | −21 | 4.84 |
| Left anterior insula/OFC | −42 | 18 | −12 | 4.36 |
| Left amygdala | −21 | 0 | −21 | 3.34* |
| Left fusiform gyrus/lateral inferior temporal gyrus | −45 | −66 | −21 | 3.83 |
| Inferior frontal sulcus | −36 | 3 | 39 | 4.02 |
| Inferior frontal gyrus | −45 | 36 | −6 | 3.37 |
| Cooperator-Neutral Faces | | | | |
| Left insula | −39 | 18 | −15 | 4.38 |
| Amygdala/ventral putamen | −18 | 3 | −15 | 3.92 |
| Left amygdala | −21 | 0 | −18 | 3.89* |
| Left putamen | −24 | 6 | −9 | 3.81 |
| Left inferior PFC | −45 | 39 | −6 | 3.59 |
| Defector-Neutral Faces | | | | |
| Ventro-medial PFC | 0 | 54 | −21 | 3.85 |
| Inferior frontal sulcus | −33 | 6 | 39 | 4.20 |

All values, p < 0.001 uncorrected. *p < 0.05 corrected for multiple comparisons across a small volume of interest.

pliance or violation of social rules), we examined effects as a function of task condition (intentional versus nonintentional task condition). No significant differences between intentional and nonintentional task condition were evident when comparing the combined affective faces or when comparing defector faces alone. For cooperator faces alone, the same comparison revealed significant

activations in posterior STS, bilateral fusiform gyrus, bilateral insula cortex, right and left lateral OFC, and ventral striatum including right and left nucleus accumbens (see Figure 3; for details see Table 3). Differential activity in left amygdala was observed above threshold only in two voxels (−21, −9, −33; z = 3.40; p < 0.001, uncorrected) and below threshold for coordinates identified in previous comparisons (−21, 0, −18; z = 2.71; p = 0.003, uncorrected).

To further determine whether responses were modulated by social context, we examined the interaction between social behavioral status (affective faces-neutral faces) and task condition (intentional versus nonintentional). No significant activations were evident in regions of interest for the interaction involving affective compared to neutral faces as well as for the interaction involving defector or cooperator compared to neutral faces alone. At a threshold of p < 0.05 (uncorrected), however, significant effects in amygdala, putamen, and insula—brain regions all identified in the main effect of cooperator-neutral faces—reflected greater activity for faces of cooperators versus neutral faces in intentional compared to nonintentional task condition.

## Discussion

We investigated whether brain circuitry identified as relevant for social judgments of facial expression, attractiveness, trustworthiness, or racial identity (Aharon et al., 2001; Hart et al., 2000; O'Doherty et al., 2003; Morris et al., 1996; Phillips et al., 1997; Winston et al., 2002, 2003) also provide a basis for a moral judgments acquired through social learning in an interactive context.

## A Left Amygdala/putamen

### (i)
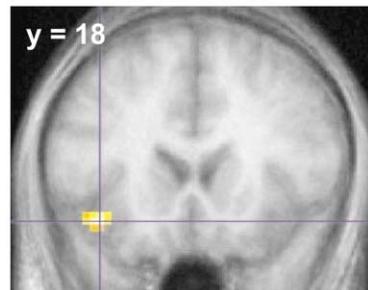


### (ii)
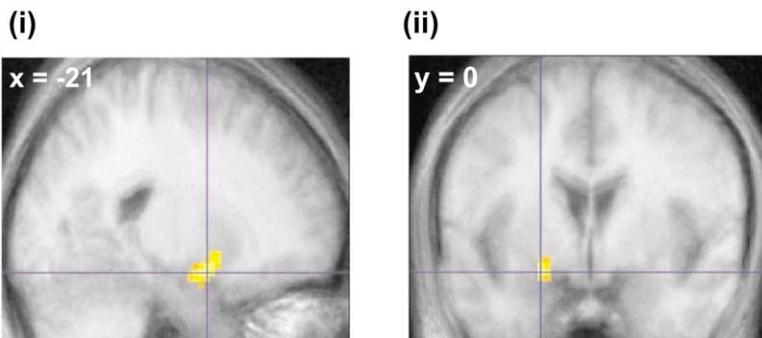


## B Left Putamen
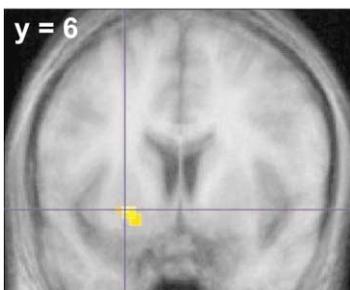


## C Left insula



Figure 2. Brain Responses to Faces of Cooperators Compared to Faces of People Involved in Neutral Null Games

The figure illustrates (A) activation in left amygdala extending anteriorly into ventral striatum with (B) a separate peak in this region showing significantly greater responses to faces of cooperators compared to neutral faces of people involved in null games. Also shown in (C) is activity in left insula. The threshold is set at p < 0.001 uncorrected.

## A Right fusiform Posterior STS



## B Bold signal change in posterior STS



## C Right fusiform Right insula



## D Left insula Ventral striatum
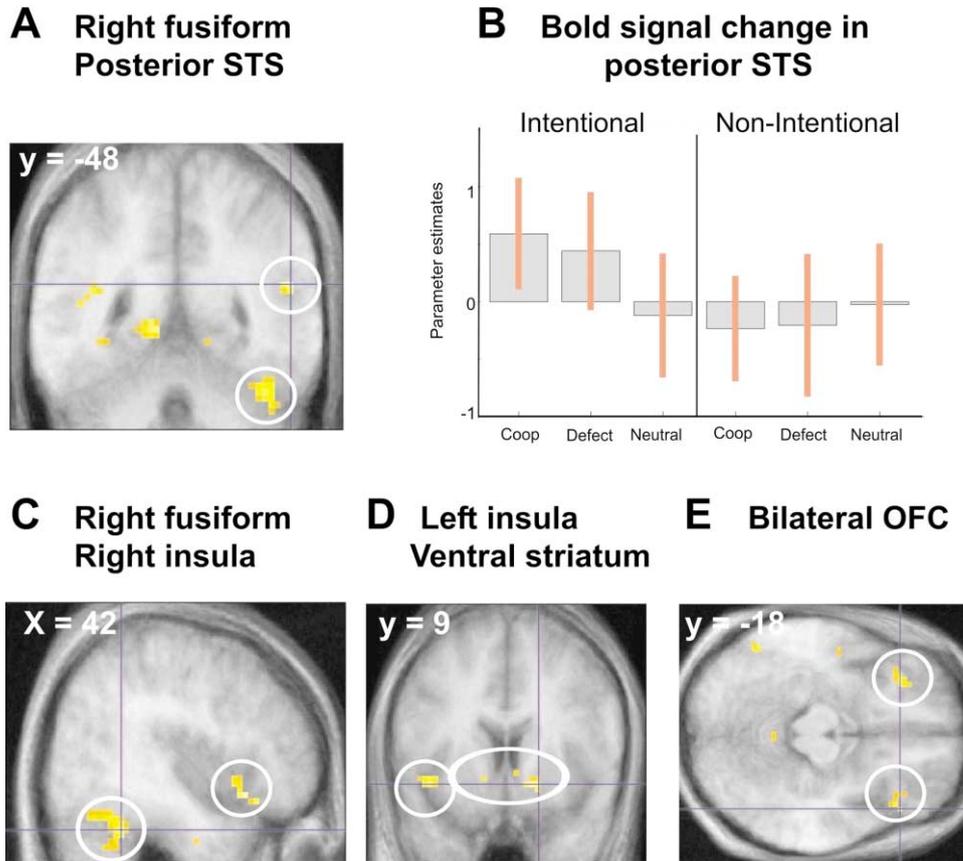


## E Bilateral OFC



Figure 3. Brain Responses to Faces of Cooperators Who Were Introduced as Intentional Agents Compared to Cooperators Who Were Said to Follow a Prescribed Response Sequence

Highlighted regions reflect brain areas that show significantly greater effects when subjects made gender judgments about faces of cooperators that have been endowed with intentionality. Specific social effects were present on a $p < 0.001$ level in right fusiform gyrus (see A and C) and right posterior STS (see A). Panel (B) illustrates BOLD signal change in posterior STS (see also A) for cooperator, defector, and neutral faces as a function of intentional and nonintentional task condition. Significant activity was also shown in bilateral insula (see C and D), ventral striatum with distinct peaks in right putamen, as well as in left and right nucleus accumbens (see D) and left and right lateral OFC (see E). STS, superior temporal sulcus; OFC, orbitofrontal cortex.

We chose mutual cooperation and defection (cheating) during interactive game play because analytical models, computer simulations, as well as empirical data from game theory, economics, and evolutionary biology (Axelrod and Hamilton, 1981; Fehr and Gächter, 2000, 2002; Nowak et al., 2000; Trivers, 1971; Wedekind and Milinski, 2000) point to the evolutionary importance of mutual cooperation and fairness in human societies coupled with an ability to readily distinguish between friend and foe among conspecifics (Cosmides and Tooby, 2000).

Behavioral assessments after scanning confirmed the saliency of perceived fairness of conspecifics as indicated by cooperative or defective behavior during social interactions. Compared to faces involved in null games (neutral behavioral status), faces of cooperators and defectors were rated as much more and much less likeable, respectively. Note that differential affective judgments cannot be attributed to differences in facial features of the pictures given that all faces had neutral facial expressions and were randomly distributed over subjects. Acquired saliency of cooperators and defectors was further corroborated in free recall performances showing that memory for cooperators and defectors

was better than for neutral faces. This finding is consistent with previous findings showing better memory for people introduced as cheaters with the help of short vignettes (Mealey et al., 1996). Interestingly, in the present study, memory for cooperators exceeded that for defectors, pointing perhaps to an enhanced saliency of mutual cooperation in human interaction.

The most robust neural responses were evident to presentation of cooperator rather than defector faces, and particularly to cooperator faces introduced as intentional agents. Our findings of activation in amygdala, striatum including putamen and nucleus accumbens, lateral orbitofrontal cortex, insula, fusiform gyrus, and STS in response to cooperator faces are highly consistent with general models of social cognition (e.g., Adolphs 2003; Brothers, 1990). The findings extend these models to processing of stimuli that acquire their social relevance through meaningful interactive social contexts. The identified regions were engaged automatically by the mere presence of socially salient faces since subjects were not explicitly asked to make social judgments about the fairness of the persons depicted. These findings are in agreement with previous fMRI studies dem-

Table 3. Cerebral Foci of Activation for Intentional versus Nonintentional Cooperator Faces

| Brain Regions | Coordinates of Peak Activation (mm) | | | Z Scores |
| --- | --- | --- | --- | --- |
| | x | y | z | |
| Predicted Brain Regions | | | | |
| Right fusiform gyrus | 42 | −48 | −33 | 3.70 |
| Left middle temporal gyrus | −51 | −66 | 0 | 4.11 |
| Left lateral fusiform gyrus | −51 | −63 | −21 | 4.13 |
| Right posterior STS | 54 | −48 | 18 | 3.96 |
| Left anterior STS | −51 | −9 | −15 | 3.47 |
| Left anterior insula | −45 | 6 | −6 | 4.08 |
| Left posterior insula | −39 | −24 | −3 | 3.96 |
| Right anterior insula/OFC | 42 | 21 | −15 | 3.97 |
| Right putamen/ | 15 | 9 | −9 | 3.92 |
| Right nucleus accumbens | 6 | 9 | −3 | 3.24 |
| Left nucleus accumbens | −12 | 9 | −6 | 3.33 |
| Left lateral OFC | −33 | 30 | −21 | 3.73 |
| Right lateral OFC | 36 | 33 | −15 | 3.43 |
| Other Brain Regions | | | | |
| Parieto-occipital fissure | 21 | −51 | 12 | 4.05 |
| Parieto-occipital fissure | −9 | −78 | 42 | 3.47 |
| Left lingual gyrus | −33 | −45 | −9 | 3.93 |
| Left retrosplenial cortex | −9 | −54 | −6 | 3.78 |
| Right retrosplenial cortex | 21 | −45 | −3 | 3.49 |
| Inferior temporal gyrus | 39 | −6 | −39 | 3.79 |
| Anterior middle temporal gyrus | 63 | −9 | −27 | 3.55 |
| Precentral sulcus | 30 | −3 | 54 | 3.58 |
| Left cuneus | −6 | 75 | 12 | 3.44 |

All values, p < 0.001 uncorrected.

onstrating automatic, task-independent affective responses during perception of socially salient features such as facial expressions of emotions, trustworthiness, attractiveness, or racial identity (O'Doherty et al., 2003; Phelps et al., 2000; Vuilleumier et al., 2001; Whalen et al., 1998; Winston et al., 2002, 2003).

It should be noted that the enhanced activity we observed was not only based on reward-related activation due to a learned association between the presented faces and losing or winning money. Activation in reward-related areas such as ventral striatum (including nucleus accumbens and putamen) and bilateral orbitofrontal cortex was also observed when testing for the effect of moral status in comparing cooperator faces of players in the intentional with those in the nonintentional task condition. Both task conditions were identical during learning and by consequence involved the same amount of monetary losses and gains. The difference depended on an attribution regarding the intentionality of the players. In contrast to the intentional task condition, subjects were not able to morally assess the actions of the players in the nonintentional task condition. Our behavioral analyses revealed that these tasks were perceived as different by our subjects. They reported that they were more emotionally involved and more angry about defectors in the intentional compared to the nonintentional task condition (see Figure 1A). These findings are in line with assumptions of inequity aversions in humans (Fehr and Schmidt, 1999).

The finding of enhanced reward-related activity in striatum including putamen and nucleus accumbens as well as bilateral OFC based on the perception of intentional players is in accordance with findings from extensive animal literature on reward processing (for a review, see Schultz, 2000) and several fMRI studies in humans using different kinds of rewarding stimuli such as abstract monetary rewards (Delgaldo et al., 2000; Elliott et al., 2000, 2003; O'Doherty et al., 2001), attractive faces (Aharon et al., 2001; O'Doherty et al., 2003), or pleasant sensory stimuli (Rolls et al., 1997; O'Doherty et al., 2002). In conjunction with results of a previous fMRI study in which striatal activation was observed during mutual cooperation in an iterated version of a simultaneous Prisoner's Dilemma game when playing against a human partner but not against a computer partner (Rilling et al., 2002), the observation of task-specific reward-related activation suggests that social fairness is experienced as rewarding per se. Thus, neural responses in these regions extend beyond players' processing of outcomes, such as monetary gain, usually associated with mutual cooperation in human societies, and reflect emotional responses based on social and moral judgments.

Another region commonly associated with reward processing as well as processing of emotionally and socially salient stimuli (Adolphs et al., 1998; Bachevalier, 2000; Baron-Cohen et al., 1999; Hart et al., 2000; Phelps et al., 2000; Winston et al., 2002) is the amygdala. In the present study, we observed left amygdala activation in response to cooperator compared to neutral faces. Only weak evidence (below our 6 voxel extent threshold at p < 0.001) for an interaction between social behavioral status and task condition in amygdala was found. Further replication will be necessary before strong conclusions about specific involvement of amygdala in social judgment can be drawn. Nevertheless, amygdala activation in response to positive stimuli, as represented by faces of fair conspecifics, is of considerable importance because it contrasts with a predominant involvement in processing threat-related and emotionally negative stimuli (e.g., Adolphs, 1999; LeDoux, 1998, 2000; Morris et al., 1996; Phan et al., 2002). For example, Winston et al. (2002) observed greater amygdala activation in response to untrustworthy compared to trustworthy faces. Rilling et al. (2002) did not find amygdala activation during mutual cooperation in the repeated Prisoner's Dilemma game. Our finding points to social saliency of fair behavior in social judgment as a potent activator of the amygdala and suggests that a negative bias observed in emotion studies may reflect on the difficulty in equating the salience of positive and negative stimuli.

One suggested role for the amygdala is feedback modulation of sensory cortices during processing of salient stimuli (e.g., Adolphs, 2003; Anderson and Phelps, 2001; Dolan, 2002; Morris et al., 1998). The observation of bilateral activation in face-responsive regions of the fusiform gyrus to socially relevant faces (intentional versus nonintentional cooperators) is in line with such an interpretation. Note that bilateral fusiform gyrus activation was also observable at a p < 0.05 uncorrected level when testing specific social effects by computing the interaction between social status and task condition. Similarly, other fMRI studies report enhanced fusiform activation at similar coordinates in response to trustworthiness (Winston et al., 2002) or emotional expression

in face stimuli (e.g., Breiter et al., 1996; Morris et al., 1998; Vuilleumier et al., 2001; Winston et al., 2003).

In addition to fusiform activation, we observed activation in the right posterior STS specific to implicit social judgments of cooperator faces introduced as free intentional agents. It has been suggested that STS has a role in processing dynamic aspects of faces (e.g., facial expressions, eye gaze) and the whole body as well as in the detection of intentions (for reviews, see Allison et al., 2000; Frith and Frith, 2003; Puce and Perrett, 2003). For example, posterior STS activation close to the coordinates observed in the present study has been reported during so-called mentalizing or "theory of mind tasks" requiring the understanding of intentions, feelings, and goals of others, be it persons or moving geometrical shapes engaged in complex interactions (Castelli et al., 2000: right STS: 60, −56, 12; Gallagher et al., 2000: right STS: 60, −48, 22; Schultz et al., 2003: right STS: 51, −57, 15; for a review, see also Frith and Frith, 2003). Interestingly, and in contrast to the present study, all mentalizing studies required subjects to make some sort of explicit social judgments about the animacy, goals, intentions, or desires of other beings—be it a person or animated geometrical shapes. Recent evidence suggests, indeed, that activity in STS may be mediated by explicit attention to socially salient features (see also Schultz et al., 2003) and that, in contrast to the amygdala, right posterior STS activation is greater when subjects make explicit judgments about trustworthiness compared to age judgments on facial stimuli (e.g., Winston et al., 2002). In the present study, however, we observed greater right posterior STS activity during gender judgments associated with intentional compared to nonintentional agents. We suggest that our subjects did indeed attribute more intentionality to the players as indicated by the results from the debriefing questionnaire (Table 1). Subjects rated such persons to be more real and animated in the intentional compared to players introduced in the nonintentional task condition (for similar behavioral findings, see Opfer, 2002). Our data support a general role for STS for processing socially salient signals ranging from simple human behavior to mental states and intentions underlying the behavior of conspecifics. Moreover, our data extends the proposed role of STS by suggesting that it may not only be involved during explicit social inferences, but also triggered automatically by socially salient signals.

Finally, we observed activation in insula cortex: left anterior insula when comparing affective to neutral faces and bilateral anterior insula when comparing cooperator faces in intentional versus nonintentional task conditions. Activation in anterior insula has been reported in a wide range of imaging studies using social or emotional stimuli (Büchel et al., 1999; Critchley et al., 2001; Hart et al., 2000; O'Doherty et al., 2003; Phelps et al., 2000; Phillips et al., 1997; Sanfey et al., 2003). Models of social cognition and emotion (e.g., Adolphs, 1999; Critchley et al., 2004; Churchland, 2002; Damasio, 1994; Dolan, 2002) have proposed that insular cortex provides a matrix for the conscious representation of "feeling states." In accordance with such a proposal, our debriefing questionnaire shows that subjects reported feelings of emotional involvement during interaction with the intentional players.

Surprisingly, we did not observe reliable activation in regions of interest in response to faces of defectors. The only exception was activation in medial OFC when comparing defector and neutral faces. However, we cannot conclude that detecting cheaters is less salient than recognizing friends. Theoretical accounts, and indeed our behavioral findings, would contradict such a conclusion. Indeed, subjects rated defectors as significantly less likeable than neutral faces. Social effects as indicated by the interaction between "social behavioral status" and "task condition" showed stronger negative likeability ratings and better memory for defector compared to neutral faces in the intentional compared to nonintentional task condition. Furthermore, subjects in the debriefing questionnaire indicated feelings of anger about defectors in the intentional task. These results speak to a strong saliency for defectors and suggest a lack of statistical power for defectors given the smaller number of defector compared to cooperator faces. To enhance credibility and cooperation in the scanned subjects, each run comprised five cooperative players and only three defecting players. It is noteworthy that Rilling and colleagues using a simultaneous version of the Prisoner's Dilemma game (see above) likewise reported more robust effects for mutual cooperation compared to defection. It would be of interest to demonstrate similar activations in response to defectors in follow-up experiments optimized for reliable assessment of defection in human interactions.

In conclusion, behavioral and functional brain data indicate saliency of perceived fairness for human interactions and point to an inherently rewarding value for mutual cooperation in human societies. Furthermore, our functional imaging data give support for neural circuitry proposed in recent models of social cognition (Adolphs, 1999, 2003; see also Brothers, 1990). In accord with the three-level model of social cognition proposed by Adolphs (2003), we demonstrate enhanced activity in higher-order visual cortices such as fusiform gyrus and posterior STS in response to socially salient stimuli (faces of cooperators). In addition, we show activation of amygdala, orbitofrontal cortex, and striatum, three areas postulated to link incoming sensory stimuli to the motivational evaluation necessary to guide behavior. On the basis of these data, we conclude that previously identified neural systems involved in social cognition extend to stimuli that acquire their salience through rapid social learning that reflects the moral evaluation of others in interactive contexts.

## Experimental Procedures

### Subjects
A total of 18 healthy subjects were scanned. The reported analyses were based on 11 healthy right-handed normal subjects (6 male, 5 female). The mean age of subjects was 25.6 (SD = 4.81; age range = 21–36). All subjects gave informed consent and the study was approved by the local research ethics committee. Given that the social cognition paradigm used required that the subjects believed that they were interacting with real people, a debriefing questionnaire was developed to assess the success of this critical manipulation. Three subjects had to be excluded on the basis of their answers in the debriefing questionnaire. Two subjects indicated that they did not believe that they were actually playing with real people during the game (see question six of the debriefing questionnaire outlined

in detail in the next section), and one indicated that he had doubts. A further three subjects were excluded because they showed more than 12 no-trust decisions in either or both task conditions and had by consequence fewer opportunities to learn the emotional properties of the relevant faces during the social learning phase (see below). Finally, one subject was excluded because of technical problems during data acquisition.

In a behavioral pilot study, an additional 10 subjects (2 men, 8 women, mean age = 26.8 years; SD = 9.37; age range = 20–52 years) were tested on the behavioral tasks only. From the 15 original participants, two subjects had to be excluded because of more than 12 nontrusting decisions, one subject due to technical problems, and two subjects because of a lack of belief in the experimental manipulation. If not otherwise stated, the behavioral analyses are based on the pooled sample of 21 subjects, and the imaging results on the scanned cohort of 11 subjects.

### Stimulus Set
Sixty-seven face stimuli were selected from a larger pool of Caucasian faces (the Karolinska Directed Emotional Faces). Equal numbers of female and male faces were selected, all of them depicting greyscale frontal images of their faces with frontal gaze direction and neutral facial expression. All stimuli had equal size and were equated for luminance and centered on a black background.

### Experimental Paradigm
The experiment was divided into four parts: an instruction phase, a social learning phase, a retrieval phase, and a behavioral assessment phase.

In the instruction phase, subjects were told that they would play interactive computer games via the internet with different people from different research institutions around Europe. Subjects were instructed that they would be randomly paired with one of the players in the pool for one game and they would be unable to predict with whom they would play before each game. To increase credibility, subjects were then photographed and told that their pictures would be edited to be adapted to the frame of the game and subsequently fed into the computer to allow the other player to see them as well. The rules of the game were then explained in detail, and a questionnaire indicated that subjects understood the rules and principles of the game.

The social learning phase included two runs (intentional and nonintentional task condition) consisting of 56 games each, resulting in a total of 112 games. The game was a sequential iterated Prisoner's Dilemma game, in which a first player can trust a second player by sending his/her starting points (transferred to money at the end of the game) to the other player, knowing that each point sent will be tripled. The second player then has the opportunity to reciprocate by sending money back, which is also tripled, resulting in equal payoffs for both players. The second player, however, can decide to keep the money and maximize his own earnings at the expense of the first player. Subjects in the scanner were always the first player and made their investment decision before they knew with whom they would be paired. During a run, subjects were repeatedly connected with five cooperator faces (players who always reciprocated trust), three defector faces (players who always kept their money), and three neutral faces involved in null games. Null games were introduced to allow for a neutral baseline condition controlling for contextual factors associated with the game. Subjects were told that due to technical reasons in some games they would take an investment decision, see their game partner as always, but would not receive any response from their partner and in consequence would neither lose nor win any money. Throughout a run, each of the 11 relevant faces was repeated four times in pseudo-random order. Twelve additional games were added to the sequence to compensate for possible noninvestment decisions by the scanned subject. Points were converted into real money at the end of the experiment. Each point corresponded to one pence and each subject received 20 pounds in total. To ensure equal payments for every subject, we rounded up payment for the subjects who opted for conservative strategies (i.e., defected frequently).

Subjects were instructed that they would play two runs of games. In one they would be connected to a platform where all players were free to decide after they have seen the face and the investment decision of the scanned subjects (intentional task condition). In the other they would be connected to a platform where players were forced to follow a given response sequence determined by a computer based on observed population distributions (nonintentional task condition). The order of task condition was counterbalanced between subjects and faces were distributed randomly over subjects. The latter controlled for possible differences in facial features of the face images such as attractiveness, facial expression, or trustworthiness.

The retrieval phase involved subjects being scanned while presenting seven repetitions of each of the 22 faces (5 cooperators, 3 defectors, 3 neutral in intentional and nonintentional task condition) introduced during the social learning phase. The faces were randomly presented and interspersed with 39 null event trials (25%) in which a fixation cross was presented for 3000 ms (total number of trials was 193 events). Faces were presented for 500 ms followed by a fixation cross for another 2500 ms. Subjects were asked to judge as quickly as possible by pressing one of two buttons with their right hand whether the perceived face was a female or male gender.

After scanning, subjects were asked to take part in a computerized behavioral assessment involving two separate tasks and to fill out a questionnaire at the end. The first task was self-paced and subjects rated likeability of all 22 faces (not likeable at all, not likeable, neutral, likeable, very likeable). The second task was a forced choice memory task in which subjects were asked to remember what the players have actually done during playing. The question was "Did this person mostly cooperate?" and subjects had to choose between "yes," "no," "don't know" or "null" when they remembered the face being involved in a null game. The debriefing questionnaire included six questions and is displayed in Table 1. The results of questions 2, 3, and 5 are displayed in Figure 1. Questions 1 and 4 served exploratory purposes and question 6 was an exclusion criterium of people not believing in the manipulation.

### Image Acquisition and Analysis
Images were acquired using a 1.5 Tesla Siemens Sonata MRI scanner to acquire gradient-echo, echoplanar T2*-weighted echo-planar images (EPI) with blood oxygenation level-dependent (BOLD) contrast. Each volume comprised 35 axial slices of 2 mm thickness with 1 mm slice gap and $3 \times 3$ mm in-plane resolution. Volumes were acquired continuously every 3.15 s. Each run began with 6 "dummy" volumes discarded for analyses. At the end of each scanning session, a T1-weighted structural image was acquired.

The images were analyzed using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK) using an event-related model (Josephs et al., 1997). To correct for motion, functional volumes were realigned to the first volume (Friston et al., 1995a), spatially normalized to a standard template with a resampled voxel size of $3 \times 3 \times 3$ mm, and smoothed using a Gaussian kernel with a full width at half maximum (FWHM) of 10 mm. In addition, high pass temporal filtering with a cut-off of 128 s was applied. After preprocessing, statistical analysis was carried out using the general linear model (Friston et al., 1995b). The presentation of each face was modeled by convolving a delta function at each event onset with a canonical hemodynamic response function (HRF) and its temporal and dispersion derivatives to create regressors of interest. Residual effects of head motion were corrected for by including the six estimated motion parameters for each subject as regressors of no interest. Contrast images were then calculated by applying appropriate linear contrasts to the parameter estimates for the parametric regressor of each event. These contrast images were then entered into a one-sample t test across the 11 subjects (random effects analysis).

The experiment constituted a $3 \times 2$ factorial design with the first factor representing social behavioral status (cooperator, defector, and neutral faces) and the second factor being task condition (intentional and nonintentional task condition). Statistical parametric maps of the main effects of social behavioral status and task condition as well as the interactions between both factors were inspected. In addition, linear contrasts between cooperator and defector faces in the intentional versus nonintentional task condition were computed.

We report results in a priori regions of interest (amygdala, OFC, fusiform, STS, striatum, and insula) at p < 0.001 uncorrected for multiple comparisons with an extent threshold of more than 6 contiguous voxels. In the case of the amygdala, we also indicate if this region survives correction for multiple comparisons within a small volume defined around the region of interest using the theory of Gaussian random fields (Worseley et al., 1996). The volume of interest for amygdala was defined by drawing a mask around the region bilaterally of approximately 10 cm³. For descriptive purposes, we report activations outside regions of interest at p < 0.001 with an extent threshold of more than 6 contiguous voxels.

**References**

Adolphs, R. (1999). Social cognition and the human brain. Trends Cogn. Sci. *3*, 469–479.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. Nat. Rev. Neurosci. *4*, 165–178.

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

Aharon, I., Etcoff, N., Ariely, D., Chabris, C.F., O'Connor, E., and Breiter, H.C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence 1. Neuron *32*, 537–551.

Alexander, R.D. (1987). The biology of moral systems (New York: de Gruyter).

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. Trends Cogn. Sci. *4*, 267–278.

Anderson, A.K., and Phelps, E.A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. Nature *411*, 305–309.

Axelrod, R. (1984). The evolution of cooperation (New York: Basic Books).

Axelrod, R., and Hamilton, W.D. (1981). The evolution of cooperation. Science *211*, 1390–1396.

Bachevalier, J. (2000). The amygdala, social behavior, and autism. In The Amygdala: A Functional Analysis, J. Aggleton, ed. (New York: Guilford Press), pp. 509–544.

Baron-Cohen, S., Ring, H.A., Wheelwright, S., Bullmore, E.T., Brammer, M.J., Simmons, A., and Williams, S.C. (1999). Social intelligence in the normal and autistic brain: an fMRI study. Eur. J. Neurosci. *11*, 1891–1898.

Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hymen, S.E., and Rosen, B.R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. Neuron *17*, 875–887.

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., and Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron *30*, 619–639.

Brosnan, S.F., and De Waal, F.B. (2003). Monkeys reject unequal pay. Nature *425*, 297–299.

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. Concepts Neurosci. *1*, 27–51.

Büchel, C., Dolan, R.J., Armony, J.L., and Friston, K.J. (1999). Amygdala-hippocampal involvement in human aversive trace conditioning revealed through event-related functional magnetic resonance imaging. J. Neurosci. *19*, 10869–10876.

Castelli, F., Happe, F., Frith, U., and Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage *12*, 314–325.

Churchland, P.S. (2002). Self-representation in nervous systems. Science *296*, 308–310.

Cosmides, L., and Tooby, J. (2000). The cognitive neuroscience of social reasoning. In The New Cognitive Neuroscience, M.S. Gazzaniga, ed. (Cambridge, Massachusetts: MIT Press), pp. 1259–1276.

Critchley, H.D., Corfield, D.R., Chandler, M.P., Mathias, C.J., and Dolan, R.J. (2000). Cerebral correlates of autonomic cardiovascular arousal: a functional neuroimaging investigation in humans. J. Physiol. *523*, 259–270.

Critchley, H.D., Mathias, C.J., and Dolan, R.J. (2001). Neuroanatomical basis for first- and second-order representations of bodily states. Nat. Neurosci. *4*, 207–212.

Critchley, H.D., Wiens, S., Rotshtein, P., Ohman, A., and Dolan, R.J. (2004). Neural systems supporting interoceptive awareness. Nat. Neurosci., in press Published online January 18, 2004. 10.1038/nn1176.

Damasio, A.R. (1994). Descartes' Error. (New York: Putman).

Delgado, M.R., Nystrom, L.E., Fissell, C., Noll, D.C., and Fiez, J.A. (2000). Tracking the haemodynamic response to reward and punishment in the striatum. J. Neurophysiol. *84*, 3072–3077.

Dolan, R.J. (2002). Emotion, cognition, and behavior. Science *298*, 1191–1194.

Elliott, R., Friston, K.J., and Dolan, R.J. (2000). Dissociable neural responses in human reward systems. J. Neurosci. *20*, 6159–6165.

Elliott, R., Newman, J.L., Longe, O.A., and Deakin, J.F.W. (2003). Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. J. Neurosci. *23*, 303–307.

Fehr, E., and Gächter, S. (2000). Cooperation and punishment in public goods experiments. Am. Econ. Rev. *90*, 980–994.

Fehr, E., and Gächter, S. (2002). Altruistic punishment in humans 2. Nature *415*, 137–140.

Fehr, E., and Schmidt, K.M. (1999). A theory of fairness, competition, and cooperation. Quar. J. Econ. *114*, 817–868.

Friston, K.J., Ashburner, J., Poline, J.B., Frith, C.D., Heather, J.D., and Frackowiak, R.S. (1995a). Spatial registration and normalization of images. Hum. Brain Mapp. *2*, 165–189.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., and Frackowiak, R.S. (1995b). Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. *2*, 189–210.

Frith, U., and Frith, C.D. (2003). Development and neurophysiology of mentalizing. Philos. Trans. R. Soc. Lond. B Biol. Sci. *358*, 459–473.

Gallagher, H.L., Happe, F., Brunswick, N., Fletcher, P.C., Frith, U., and Frith, C.D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. Neuropsychologia *38*, 11–21.

Hart, A.J., Whalen, P.J., Shin, L.M., McInerney, S.C., Fischer, H., and Rauch, S.L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. Neuroreport *11*, 2351–2355.

Josephs, O., Turner, R., and Friston, K. (1997). Event-related fMRI. Hum. Brain Mapp. *5*, 243–248.

Knutson, B., Westdorp, A., Kaiser, E., and Hommer, D. (2000). fMRI visualization of brain activity during a monetary incentive delay task. Neuroimage *12*, 20–27.

LeDoux, J.E. (1998). The Emotional Brain (London: Weidenfeld & Nicholson).

LeDoux, J.E. (2000). Emotion circuits in the brain. Annu. Rev. Neurosci. *23*, 155–184.

Mealey, L., Daood, C., and Krage, M. (1996). Enhanced memory for faces of cheaters. Ethol. Sociobiol. *17*, 119–128.

Milinski, M., Semmann, D., and Krambeck, H.J. (2002). Reputation helps solve the 'tragedy of the commons'. Nature *415*, 424–426.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J., and Dolan, R.J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. Nature *383*, 812–815.

Morris, J.S., Ohman, A., and Dolan, R.J. (1998). Conscious and un-conscious emotional learning in the human amygdala. Nature *393*, 467–470.

Nowak, M.A., and Sigmund, K. (1998a). The dynamics of indirect reciprocity. J. Theor. Biol. *194*, 561–574.

Nowak, M.A., and Sigmund, K. (1998b). Evolution of indirect reci-procity by image scoring. Nature *393*, 573–577.

Nowak, M.A., Page, K.M., and Sigmund, K. (2000). Fairness versus reason in the ultimatum game. Science *289*, 1773–1775.

O'Doherty, J., Kringelbach, M.L., Rolls, E.T., Hornak, J., and An-drews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. Nat. Neurosci. *4*, 95–102.

O'Doherty, J., Deichmann, R., Critchley, H.D., and Dolan, R.J. (2002). Neural responses during anticipation of a primary taste reward. Neuron *33*, 815–826.

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

Opfer, J.E. (2002). Identifying living and sentient kinds from dynamic information: the case of goal-directed versus aimless autonomous movement in conceptual change. Cognition *86*, 97–122.

Phan, K.L., Wager, T., Taylor, S.F., and Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. Neuroimage *16*, 331–348.

Phelps, E.A. (2001). Faces and races in the brain. Nat. Neurosci. *4*, 775–776.

Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Funayama, E.S., Gatenby, J.C., Gore, J.C., and Banaji, M.R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. J. Cogn. Neurosci. *12*, 729–738.

Phillips, M.L., Young, A.W., Senior, C., Brammer, M., Andrew, C., Calder, A.J., Bullmore, E.T., Perrett, D.I., Rowland, D., Williams, S.C., et al. (1997). A specific neural substrate for perceiving facial expres-sions of disgust. Nature *389*, 495–498.

Puce, A., and Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. Philos. Trans. R. Soc. Lond. B Biol. Sci. *358*, 435–445.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. Neuron *35*, 395–405.

Rolls, E.T., Francis, S., Bowtell, R., Browning, D., Clare, S., Smith, T., and McGlone, F. (1997). Taste and olfactory activation of the orbitofrontal cortex. Neuroimage *5*, 199.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., and Cohen, J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. Science *300*, 1755–1758.

Schultz, W. (2000). Multiple reward signals in the brain. Nat. Rev. Neurosci. *1*, 199–207.

Schultz, R.T., Grelotti, D.J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., and Skudlarski, P. (2003). The role of the fusiform face area in social cognition: implications for the pathobiology of autism. Philos. Trans. R. Soc. Lond. B Biol. Sci. *358*, 415–427.

Trivers, R. (1971). The evolution of reciprocal altruism. Q. Rev. Biol. *46*, 39–57.

Vuilleumier, P., Armony, J.L., Driver, J., and Dolan, R.J. (2001). Ef-fects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron *30*, 829–841.

Wedekind, C., and Milinski, M. (2000). Cooperation through image scoring in humans. Science *288*, 850–852.

Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., and Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J. Neurosci. *18*, 411–418.

Winston, J.S., Strange, B.A., O'Doherty, J., and Dolan, R.J. (2002). Automatic and intentional brain responses during evaluation of trust-worthiness of faces. Nat. Neurosci. *5*, 277–283.

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

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., and Evans, A.C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. *4*, 58–73.