NeuroImage 54 (2011) 1685–1691

Contents lists available at ScienceDirect

NeuroImage

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

Emotional conflict in interpersonal interactions

María Ruz [*], Pío Tudela

University of Granada, Spain

ARTICLE INFO

Article history:
Received 19 April 2010
Revised 15 July 2010
Accepted 17 August 2010
Available online 22 August 2010

Keywords:
Emotional conflict
Social expectations
Ultimatum game
Theory of mind
fMRI

ABSTRACT

Facial displays of emotions can help to infer the mental states of other individuals. However, the expectations we generate on the basis of people's emotions can mismatch their actual behaviour in certain circumstances, which generates conflict. In the present study, we explored the neural mechanisms of emotional conflict during interpersonal interactions. Participants had to accept or reject economic offers made by several partners who displayed emotional expressions. On every trial, a cue informed participants of whether they could trust the emotion of their partner or not. Trustworthy (low-conflict) partners with happy facial expressions were cooperative and those with angry expressions did not cooperate. Untrustworthy (high-conflict) partners, on the other hand, cooperated when their expression was angry and did not cooperate when they displayed a happy emotion. Behavioural responses were faster for trustworthy than for untrustworty partners. High-conflict partners activated the anterior cingulate and the anterior insula. In turn, trustworthy partners were associated with activations in the left precuneus. Our results suggest that the emotion displayed by another person affects our decision-making in social contexts. When emotional expressions are linked to their natural consequences, they engage ToM processes. In contrast, untrustworthy emotional expressions engage conflict-related brain regions.

Introduction

These past years have witnessed an extensive growth of the research focused on the neural mechanisms that allow us to detect conflict and to deal with it. Most of the paradigms that have been used to study this effect include variations of the Stroop (1935) or flanker (Eriksen and Eriksen, 1974) tasks. In these tasks, different stimuli or different dimensions of the same stimulus are associated to non-overlapping responses. This generates competition between prepotent processes and situationally task-relevant goals. Hence, conflict arises when these representations lead to incompatible courses of action. For example, if we are asked to name the ink colour in which a word is printed and its meaning refers to a different hue (e.g. the word red printed in yellow), our responses are slower than when both (ink and word; e.g. the word red printed in red) match. The difference in reaction time (RT) between incongruent and congruent conditions is termed *conflict effect*. The materials that elicit conflict can be of different nature, and research to date has mostly focused on *cognitive* and *emotional* kinds of conflict. Whereas cognitive conflict involves stimuli that have no affective connotations (e.g. arrows pointing left or right), emotional conflict typically requires the suppression of distracters with an emotional valence (e.g. faces displaying happy or fearful expressions).

In neural terms, several reports have shown the pervasive involvement of two brain areas in conflictive situations: the Anterior Cingulate (ACC) and Dorsolateral Prefrontal (DLPFC) cortices. An influential theory on the field, the conflict-monitoring hypothesis (Botvinick et al., 2001) proposes that the role of the ACC would be to signal the occurrence of conflicts between competing active representations and to trigger adjustments in cognitive control, which in turn would be implemented by the DLPFC (see Egner, 2008 for a complementary approach). Results also seem to indicate that there is a regional specialization in the brain for different types of conflict (e.g. Whalen et al., 1998). Cognitive conflict engages dorsal parts of the ACC (dACC) and the lateral prefrontal cortex (LPFC), whereas emotional conflict engages the rostral ACC (rACC; Egner et al., 2008; Ochsner et al., 2009). In addition, both types of conflict activate common regions of the dACC.

Emotions play an essential role in social interactions (Olsson and Ochsner, 2008). It is well known that people use several social cues to try to predict the mental states of others (Frith and Frith, 2006), and their emotions are among these (Van Kleef et al., 2010). Along evolution, we have learned that positive emotions such as happiness predict positive consequences whereas negative emotions such as anger predict that bad things may happen (Darwin, 1872). However, there are situations in which these relations do not hold. For example, we may learn that some people are untrustworthy and thus their emotional displays cannot be taken as indicative of their future actions. In certain contexts, these people can conceal their true emotional state and express a different one. These situations may generate conflict between the natural reactions that their emotions generate on us and their actual meaning in the

[*] Corresponding author. Dept. of Experimental Psychology, Campus de Cartuja s/n, 18071 Granada, Spain.
E-mail address: mruz@ugr.es (M. Ruz).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.08.039

current context. Yet, to date it has not been explored which brain areas are involved in this sort of interpersonal emotional conflict, nor their underlying neural dynamics.

Interpersonal interactions have long been studied experimentally using behavioural bargaining games developed within the field of Behavioural Game Theory (Camerer, 2003). In all these games, participants have to use whatever information is available to try to predict what their partners are going to do, and act accordingly to obtain the maximum benefit. One of them is the Ultimatum Game (Güth et al., 1982). In the original game, one player (the *proposer*) splits a certain amount of money into two sums, one for him and the other for another player, the *responder*. He, in turn, can either accept the offer (and thus they both win their respective amounts) or reject it (and both get nothing). Results using this game have shown that people reject more unfair offers than would be expected from a rational perspective (Camerer, 2003). In addition, the activation in the anterior insula can be used to predict the decision made by participants (Sanfey et al., 2003).

We adapted the UG in several aspects to evoke a situation in which the non-conflictive (trustworthy) and conflictive (untrustworthy) emotions of the partners in a bargaining game had to be used to predict the valence (good or bad) of their economic offers. Rather than learning about the trustworthiness of the emotions of their partners by trial and error, participants were explicitly told whether they could trust their partners or not by means of a symbolic cue. For humans, instruction is a reliable means of exchanging relevant information. For example, previous research in the field of fear learning has shown that instructing participants about the association between a conditioned stimulus and an electric shock produces similar levels of learning than personally acquiring or observing such relation (Olsson and Phelps, 2004, 2007). Thus, in the current study we investigated the effect that the instruction of trustworthiness had on decisions made on the basis of the emotions displayed by other people.

In this modified game, participants always played the role of responders to divisions of money provided by alleged partners, which allowed measuring their choices and their speed. They were asked to use the emotions (happiness and anger) conveyed by trustworthy and untrustworthy partners to anticipate their most likely behaviour. In the low-conflict condition emotions predicted their ‘natural’ consequences, whereas in the high-conflict condition emotions predicted the opposite. Note that, in contrast to previous investigations, the present study focuses on the emotions displayed by the partners in the game, rather than in the emotions that participants themselves may feel during the decision-making process (see Van Kleef et al., 2010).

The present study had several goals. First, we aimed at obtaining behavioural markers of emotional conflict during interpersonal interactions. Second, we wanted to investigate whether this type of conflict engages brain areas similar to those reported in the previous literature exploring non-social conflict. Finally, we were interested in studying how the trustworthiness of the partners modulated the pattern of functional interactions between relevant brain areas.

Methods

Participants

Eighteen right-handed participants (20–31 year old, 8 men), with normal or corrected-to-normal vision, were recruited from the University of Oxford community. They all signed a consent form approved by the Central Office for Research Ethics Committee (COREC 06/Q1607/33).

Stimuli and procedure

Participants played the role of the *responder* in a modified ‘Ultimatum Game’ (UG) with many different partners (*proposers*). They were instructed that the offers that they were going to receive were taken from the responses of participants who completed several standardized questionnaires related to social situations and trustworthiness. Half of these offers would be beneficial to them (‘good offers’), and the other half would be beneficial to the partners (‘bad offers’). Their goal was to sum more money than all of their partners together, and if they won they would receive an extra £5 as a reward. For each of the partners, they would receive information regarding how trustworthy they were by means of a cue (a square or a triangle) presented at the beginning of every trial. For trustworthy partners, a smile would mean that most likely the offer would be good, and an angry expression would predict a probable bad offer. Untrustworthy partners, on the other hand, would smile in anticipation of bad offers and would have an angry expression before a probable good offer. Participants, who were explicitly informed of the relation between the partners trustworthiness and the emotion they displayed, had to use the information provided by the cue together with the emotion expressed by their partner to accept or reject the offers *before* they were presented. The offer was presented afterwards (see Fig. 1 for a display of the sequence of the events in a trial). If they had accepted the offer, they would keep their share of the division and their partner would take the other amount. If they had rejected it instead, no amount would be added to any account for that trial. In addition, participants were asked to respond to their partner's face as fast as possible; they were told that if they took too long, the highest amount in the offer would be added to their partner's account.

Triangular and squared black shapes were used as trustworthiness cues, which indicated whether the partner for that trial was either trustworthy or untrustworthy (counterbalanced across participants). One hundred and sixty faces from the Karolinska Directed Emotional Faces database (Lundqvist et al., 1998; 50% female), displaying happy or angry (50%) facial expressions, were used as partners who offered the participants a split of a sum of money. There were 32 different offers, displayed as a green and a blue number (from 1 to 9) separated by a slash symbol. The difference between the two numbers was always 1.[1] The left–right location of the highest number and the colours were matched across trials. The participants were assigned the amount coded in one colour and the partners the other number/amount (the colour assignment was counterbalanced across participants). In half of the trials, the highest number was displayed in the colour that corresponded to the participant and in the other half in the partner's colour. This manipulation was orthogonal to the emotion displayed by the partners and to the cue. The predictions of the cue and the emotional expression of the target were valid in 80% of the trials. That is, offers were good to the participants in 80% of the trials in which a trustworthy partner smiled or an untrustworthy partner had an angry expression, and offers were bad in 80% of the trials in which a trustworthy partner had an angry expression or an untrustworthy partner smiled.

A PC running Presentation 0.70 displayed the stimuli, which were viewed by the participants in the screen mounted at the back of the scanner by means of a mirror placed on top of the image acquisition coil. The delays between cue, target and offers, and between trials, were jittered to allow the deconvolution of cue-related, target-related and offer-related signals. Trials were presented in random order. Each trial comprised the following events (see Fig. 1). A cue was flashed in the centre of the screen for 500 ms, followed by an interval displaying the fixation point with a 4–10 s duration that varied pseudorandomly in a quasi-logarithmic fashion in steps of 500 ms (50%: 4–5.5 s, 33.3%: 6–7.5 s, 16.7%: 8–9.5 s; mean = 6.15 s). The picture of the partner for that trial then replaced the fixation point for 500 ms, after which another variable interval was presented, with the same structure as

[1] Offers used in experimental studies of the Ultimatum Game are usually divided into fair (the difference between the two amounts is small) and unfair (the difference between the two amounts is large) types. In our study we restricted our stimuli to fair offers to avoid the variability associated to this manipulation.

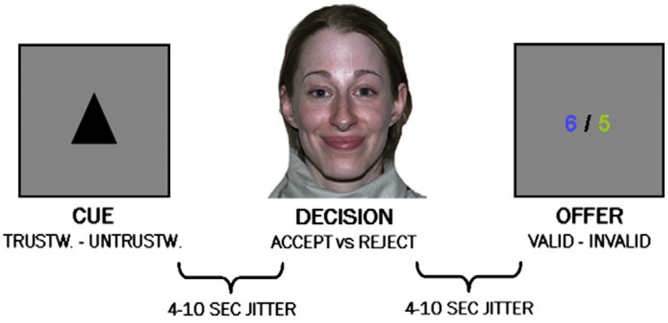

Fig. 1. Display of sequence of events in a trial.

the previous one. Finally, the offer was presented for 500 ms, and was followed by a third variable interval with the same jitter as the previous ones. This type of distribution of delays has the advantages of controlling for temporal expectations and keeping the task to an endurable length whilst allowing a good separation of hemodynamic responses to cues and targets within trials (Friston et al., 1998). On average, a trial lasted 19.9 s. In total, there were 160 trials (53 min).

Participants used the index and middle fingers of their right hand to make speeded decision responses (accept or reject the offers) to the facial displays by pressing one of two buttons on a custom-made MRI-compatible button box. Before performing the task in the scanner, participants completed a short training session (10 min), which had the same parameters as the main task but used a different set of faces.

Image acquisition and preprocessing

Magnetic-resonance images were acquired using a 3 T Trio scanner at the Oxford Centre for Clinical Magnetic Resonance Research (OCMR). Functional images were obtained with a one-shot T2*-weighted echoplanar imaging (EPI) sequence (time until echo [TE] = 22 ms, flip angle = 90°, repetition time [TR] = 2.1 s). Thirty-five interleaved sagittal slices with a thickness of 4 mm, tilted 30 degree to the anterior–posterior commissural line to optimize sensitivity to orbitofrontal cortex and medial temporal lobes (Deichmann et al., 2003), covered the entire brain (64 × 64 matrix with a field of view of 192 × 192 mm, voxel size of 3 × 3 × 4 mm). The experiment was performed in a single run consisting of 1528 image volumes. The first 5 images were discarded to allow for saturation of the signal. In addition, we acquired a standard structural image of each participant using a high-resolution T1-weighted sequence (TR = 15 ms; TE = 6.9 ms; 1 × 1 mm in-plane resolution and 1.5-mm slice thickness).

Analyses

Image analysis was performed with SPM5 (Welcome Department of Imaging Neuroscience, University of London, UK). Functional images were slice-time corrected, and realigned and unwarped using a least-squares approach and a six-parameter (rigid body) spatial transformation to correct for motion artifacts. High-resolution anatomical T1 images were then coregistered with the realigned functional images (Friston et al., 1995). Functional images were spatially normalized into the Montreal Neurological Institute [MNI]-space using the default EPI template in SPM5, and the resultant parameters were applied to the participants' structural images. Functional images were spatially smoothed using an 8 mm^3 full-width-at-half-maximum Gaussian kernel, to account for anatomical variability between participants and to conform the data to a Gaussian model (Hopfinger et al., 2000).

Statistical analysis was performed with a General Linear Model for each participant with corrections for serial autocorrelations using the AR(1) model. The model included regressors for the cues (TC: trustworthy cue, UC: untrustworthy cue), faces (TH: trustworthy happy, UH: untrustworthy happy, TA: trustworthy angry, UA: untrustworthy angry) and offers (VO: valid offer, IO: invalid offer) which were convolved with the standard hemodynamic response function. The two different cues (for trustworthy and untrustworthy partners) were modeled as events with a duration that encompassed the whole cue-face interval. Facial displays and offers were modeled as events with zero duration. Trials with errors and missing responses were grouped together as separate events of no-interest with an extended duration for the whole trial (encompassing cue, faces and offers). A high-pass filtering (128 s) was applied to remove low-frequency confounds.

We performed the contrasts of interest separately at the individual level, which were then entered in a second-level group analyses (random-effects). For all the analyses performed, voxel-wise statistical thresholds were set at $p < 0.001$ (uncorrected). To guard against Type I errors, only clusters with 10 or more voxels were considered (Forman et al., 1995). This cluster size was always larger than the number of voxels expected by chance in each cluster as calculated by the SPM package (Friston et al., 1996). These contrasts aimed to identify the brain areas involved in the preparation for high or low conflict generated by the cues (UC > TC; TC > UC), and the role of conflict in decision-making when trustworthy and untrustworthy partners were presented (UP > TP; TP > UP).

In addition, to explore the interactions between conflict and emotions, we contrasted the conditions in which the same predictions (good or bad offers) were conveyed by emotions that are naturally associated (low-conflict) or not (high-conflict) with such consequences (i.e. TH > UA; UA > TH; TA > UH; UH > TA).[2]

As a complement to the contrasts between conditions, we explored how the level of conflict influenced the pattern of functional interactions of one key conflict-related brain region (ACC), by means of a psychophysiological interaction (PPI) analysis (Friston et al., 1997) as implemented in SPM5. This analysis evaluates how the correlation of the time course of brain activity in a 'seed' region with other brain areas changes depending on the experimental conditions. For each subject, we extracted the time course of activity from a 6 mm radius volume of interest around the peak voxel in the ACC identified in the main target group contrasts (UP > TP; this region was also activated by all the faces collapsed across trustworthiness conditions, z = 4.22). Therefore, our PPI analyses revealed which brain areas showed patterns of activations that covaried with ACC activity depending on the level of conflict that the emotion of the partners generated.

Results

Behavioural

Responses that did not help participants to win the game (i.e., those in which they rejected offers that were good for them or those in which they accepted offers that were beneficial to their partners) were considered errors. Overall accuracy in the task was 87%. Performance was more accurate in the trustworthy (90%; SD = 9) than in the untrustworthy condition (84%; SD = 12), $F_{1,17} = 15.57$, $p < 0.001$. The effect of emotion and the interaction between cue type and emotion were not significant, $F_{1,17} = 3.30$, p = 0.08 and $F < 1$, respectively.

[2] Activations generated by the offers are not presented in this paper. Tables describing these results are available upon request to the authors.

The average RT was 1365 ms. Responses were faster in the trustworthy (1302 ms; SD = 535) than in the untrustworthy (1429 ms; SD = 544) condition, $F_{1,17}=24.09$, $p<0.001$, and were also faster for happy (1275 ms; SD = 487) than for angry (1456; SD = 580) partners, $F_{1,17}=28.83$, $p<0.001$. This difference, however, was larger in the trustworthy (299.7 ms, $F_{1,17}=48.74$, $p<0.001$) than in the untrustworthy condition (108.4 ms, $F_{1,17}=5.52$, $p<0.05$), as evidenced by the significant interaction between type of cue and emotion, $F_{1,17}=20.93$, $p<0.001$, see Fig. 2.

Neuroimaging

No significant differences were found between cues that predicted trustworthy or untrustworthy partners at the specified threshold. Deciding whether to accept or reject offers coming from untrustworthy compared to trustworthy partners engaged several brain regions, including the dorsal anterior cingulate cortex, the medial and lateral prefrontal cortex and the anterior insula bilaterally. Some of these areas have been previously involved in conflict processing, specially the ACC. Trustworthy partners, on the other hand, activated the precuneus (see Table 1 and Fig.3).

To examine the interactions between conflict and emotional processing, we contrasted the conditions in which good or bad offers were conveyed by emotions that are naturally associated (low-conflict) or not (high-conflict) with such consequences. When good offers were predicted by happy partners, compared when angry partners predicted the same offer (TH>UA), activation was found in the precuneus. The opposite contrast (UA>TH) revealed activations in the bilateral anterior insula, orbitofrontal and inferior parietal cortex, as well as in the bilateral thalamus. The other two contrasts (TA>UH; UH>TA) did not yield significant supra-threshold results (see Table 2).

The connectivity (PPI) analyses showed that the pattern of interactions of the ACC changed depending on the level of conflict. When participants interacted with untrustworthy partners, the activation of the ACC was most strongly coupled with the mid-cingulate/SMA and the posterior rolandic operculum. In contrast, ACC activity during interactions with trustworthy or low-conflict partners was associated with an increase in activity in some brain regions previously linked with theory of mind operations (ToM), such as the anterior ventromedial prefrontal cortex and the precuneus bilaterally (see Table 3 and Fig. 4).

Discussion

The present study sought to develop a novel paradigm to study the neural basis of emotional conflict during interpersonal interactions. Our results show that participants took longer and made more errors when the emotions displayed by the partners in the game did not lead to their natural consequences. At a neural level, untrustworthy partners activated the dorsal ACC and bilateral frontal areas, as well as the anterior insula. These results matched our predictions regarding the conflictive nature of emotional displays that are not followed by their natural consequences. Trustworthy partners, on the other hand, recruited the precuneus. In addition, the pattern of interactions of the ACC with other regions in high and low-conflict conditions lent further support to the notion that partners whose emotions could be trusted engaged a ToM-related brain circuit.

Table 1
Effect of emotional conflict.

Contrast/region	BA	X	Y	Z	Z-score	Voxels
Untrustworthy > trustworthy partners						
L sup med frontal	32	−6	33	39	3.88	16
R sup med frontal	32/7	15	51	33	3.96	16
R ant insula	47	36	24	−6	4.23	90
L ant cingulate	24	−6	18	33	4.00	21
L inf frontal operculum/insula	48	−39	18	18	3.67	17
R thalamus		12	−15	12	3.67	13
L cerebellum		−51	−57	51	3.62	21
Trustworthy > untrustworthy partners						
L precuneus	30	−12	−54	18	3.71	13

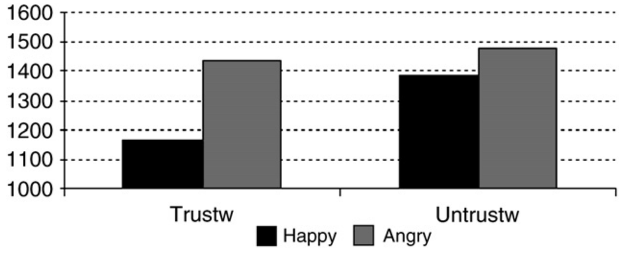

Fig. 2. Reaction Times (RT) for trustworthy (Trustw) and untrustworthy (Untrustw) partners displaying happy and angry facial expressions.

The involvement of the ACC in conflictive situations has been reported in many studies (see Botvinick et al., 2004). Several inter-related functions have been proposed to explain the role of this area, which include the detection of conflict (Botvinick et al., 2001), the monitoring of performance and evaluation of action outcomes (Rushworth et al., 2004) or the configuration of priorities for a new task (Hyafil et al., 2009), among others. As mentioned in the introduction, emotional and cognitive conflict seem to engage dissociable and common regions of the ACC. Whereas tasks that require the suppression of conflictive emotional information activate the rostral ACC, paradigms in which the information to suppress is non-emotional engage more dorsal parts of this brain region (Mohanty et al., 2007; Egner et al., 2008; Ochsner et al., 2009). The results

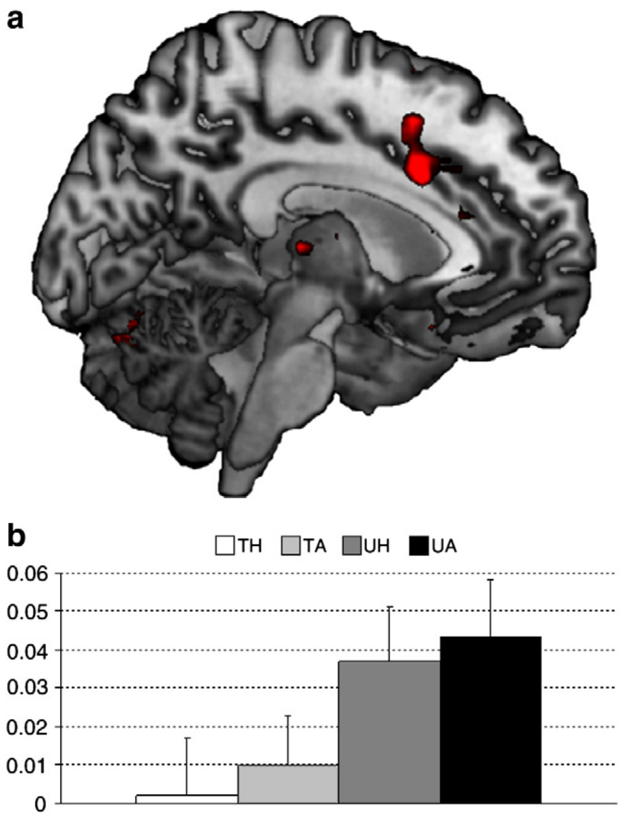

Fig. 3. a. Sagittal display of activations in the anterior cingulate (ACC) and right thalamus in the contrast of untrustworthy vs. trustworthy partners. b. Percent signal change (plus standard error) in the ACC cluster for the conditions Trustworthy happy (TH), Trustworthy angry (TA), Untrustworthy happy (UH) and Untrustworthy angry (UA).

Table 2
Interactions between conflict and emotion.

Contrast/region	BA	X	Y	Z	Z-score	Voxels
Trustworthy happy > untrustworthy angry partners						
Precuneus	30	−6	−48	12	4.56	102
Untrustworthy angry > trustworthy happy partners						
L inf orbitofrontal	47	−42	42	−3	4.08	32
L sup med frontal	32	−6	27	42	3.61	11
L inf frontal operculum	44	−54	18	30	3.48	29
L ant insula	48	−36	18	15	4.77	138
R ant insula	48	36	27	3	4.51	80
L inf parietal	40	−42	−57	48	3.67	91
L thalamus		−6	−18	6	3.83	37
R thalamus		12	−18	12	4.35	51

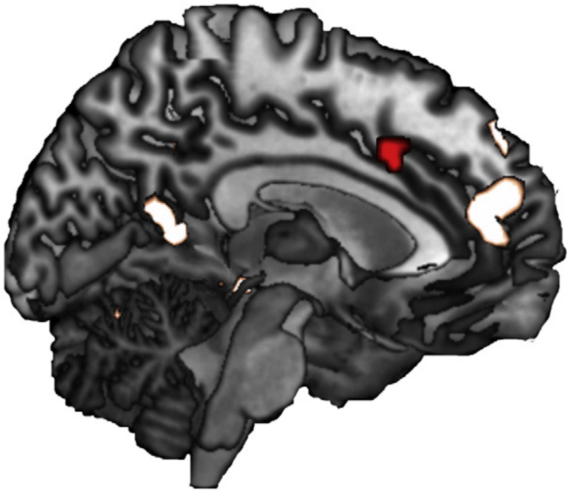

Fig. 4. Results of the analysis of psychophysiological interactions (PPI) seeded in the anterior cingulate cortex (ACC; displayed in red). This brain area interacts more strongly with the anterior ventromedial prefrontal cortex and the precuneus (displayed in white) in the trustworthy compared to the untrustworthy condition.

of our study may seem at odds with such dissociative pattern. Even when the emotions displayed by the partners were the source of conflict in our paradigm, we did not find reliable activations in the rACC at the specified threshold[3] (other studies using emotional information have also failed to find activation in the rACC, e.g. Haas et al., 2006). There are, however, differences between the paradigms that may explain this discrepancy. Previous studies of emotional conflict have used variants of the Stroop or flanker tasks, which required the suppression of the irrelevant emotional content of *distracters* (e.g. Whalen et al., 1998; Egner et al., 2008; Ochsner et al., 2009). In one of them, for example, participants were required to evaluate the valence of the expression of faces displaying happy and angry emotions while ignoring the meaning of the words "happy" or "fear", presented in capital letters on the centre of the screen (Egner et al., 2008). In our task, participants had to pay attention to the emotions displayed by the partners in all conditions, as it was the clue to infer whether their offer were going to be good or bad. What needed to be suppressed in the high-conflict condition was the natural tendency to associate happiness and anger with positive and negative outcomes, respectively. This conflict activated the dACC with a peak in MNI coordinates $x=-6$, $y=18$, $z=33$, which is close to the reported peak for a region of the ACC that is recruited by both emotional and cognitive kinds of conflict (−6, 12, 40; Egner et al., 2008; see also Luo et al., 2007). Thus, the involvement of this brain region in the untrustworthy condition supports the idea that emotions that do not predict the outcomes that are naturally associated to them engender conflict during interpersonal interactions.

We also found differences in the anterior insula. Although this brain area has been implicated in a wide array of processes (see Craig, 2009), its involvement in cognitive control (e.g. Dosenbach et al., 2007, 2008; Roberts and Hall, 2008) and interpersonal relations (e.g. Olsson and Ochsner, 2008) is well established. In the social neuroscience literature the anterior insula is usually linked to visceral feedback in response to negative social interactions (e.g. Rilling et al., 2008b) such as unreciprocated cooperation (Rilling et al., 2008a) or unfair offers in the Ultimatum Game (Sanfey et al., 2003). In our paradigm, however, trustworthy and untrustworthy partners predicted outcomes of equal valence, as both offered the same proportion of good and bad offers to participants. Thus, the activation of the insula in high-conflict situations is not easily explained as a simple reaction to interactions that predict a negative outcome in economic terms, as conditions were matched on this respect. In our experiment, the trigger for the activation of this brain area seems to be the mismatch between the valence naturally associated to emotions and their actual consequences in the game, rather than the mere prospects of receiving unfair or bad offers. In support of this idea, offers that did not match the initial expectations (invalid offers) activated the insula bilaterally, with a peak in the right hemisphere (MNI $x=33$, $y=24$, $z=-3$) quite close to the activation (MNI $x=36$, $y=24$, $z=-6$) found for untrustworthy partners. This result opens the possibility that previous reported results of insula activations may be explained in part by a mismatch between a natural tendency to expect fair offers from partners in a game and their actual unfair offers. Further research however would be needed to add support to this claim.

Another potential reason explaining the involvement of the insula may be its relation to trustworthiness processing. Winston et al. (2002) showed that the activation of the right anterior insula (and bilateral amygdala) displayed a negative correlation with trustworthiness ratings of facial displays, even when participants were performing an unrelated gender judgment task (see also Todorov et al., 2008). Differences in facial characteristics that may relate to trustworthiness judgments cannot explain our results, however. In contrast to previous studies, the assignment of faces to the trustworthiness conditions was arbitrary and fully counterbalanced across participants in our experiment. Faces were assigned to trustworthy and untrustworthy conditions depending on whether their facial displays of emotions could be trusted or not, and hence this difference in the reliability of the emotions of the partners may have contributed to the activation of the anterior insula in the untrustworthy condition. In any case, the engagement of this area in the high-conflict condition suggests that it is involved in top-down control needed to override prepotent responses linked to emotions during interpersonal interactions (see also Lee et al., 2008; Levens and Phelps, 2010).

In addition, our results brought out an interesting relation between ToM and the trustworthiness of emotional facial displays. A network of brain areas is reliably activated in tasks that require participants to attribute mental states to other people. These include the medial prefrontal cortex (MPFC; Amodio and Frith, 2006), precuneus, and the superior temporal sulci, among others (Gallagher and Frith, 2003; Saxe,

Table 3
ACC connectivity in high and low emotional conflict.

Contrast/Region	BA	X	Y	Z	Z-score	Voxels
Untrustworthy > trustworthy						
L mid-cingulate/SMA	6/24	−3	6	42	4.14	60
L post rolandic operculum	48	−48	−21	21	3.70	10
L postcentral gyrus	2/3	−39	−24	48	4.48	104
Trustworthy > untrustworthy						
Anterior ventromedial PFC	32/10	−3	57	24	3.79	78
R sup temporal sulcus	21/22	63	−21	−3	3.61	15
R precuneus	23	18	−48	24	3.99	11

[3] Note, however, than when the threshold is lowered to $p<0.005$ (uncorrected) we observe the involvement of a region in the rACC (UP>TP, $t=3.71$, 8 contiguous voxels).

in press; see also Spreng et al., 2009). These areas are also engaged by tasks that require the evaluation of emotions (Ochsner et al., 2004). During interactive games, these regions are more activated when participants play with alleged human partners compared to computers (Rilling et al., 2004). They are also involved in the attribution of false beliefs (e.g. Gobbini et al., 2007), moral judgments (Young and Saxe, 2007) or the generation of intentional representations (Abraham et al., 2008). The analysis of connectivity (PPI, see Methods) that we performed in our data showed that the ACC, a key region in cognitive control, interacted more strongly with the MPFC, the right STS and bilateral precuneus for trustworthy partners compared to untrustworthy ones. These results suggest that, when participants played with people whose emotions could be trusted, they relied on ToM mechanisms to make inferences about the emotional mental states of their partners (see Saxe, 2006).

The opposite contrast showed that the ACC interacted with the SMA for untrustworthy, or high-conflict, partners. This area could be involved in implementing task-relevant response mapping rules (e.g. Ullsperger and von Cramon, 2004), which would be needed to overcome the tendency to respond in line with the natural consequences associated with emotions. Thus, the coupling between ToM regions and the ACC is higher during trustworthy conditions because the information conveyed by these areas may prove more helpful in these situations, whereas other areas related to cognitive control may coordinate with the ACC during conflict between emotions.

Conclusions

Emotions are a key ingredient of social interactions among people. This study presents a novel approach to study the neural basis of high and low emotional conflict during interpersonal exchanges. Our results show that conflict-related brain areas such as the ACC are activated when the emotions displayed by other people during social interactions do not predict their natural consequences. In addition, the anterior insula is also involved in conflictive social interactions. However, when we can trust the emotions of others, we observe activations in ToM-related brain areas, such as the ventromedial PFC. Thus, our study extends previous results by introducing social interactions in the field of emotional conflict research, and by showing that the trustworthiness that we explicitly ascribe to other people modulates the engagement of conflict and ToM-related brain regions.

Acknowledgments

We are grateful to Anna C. Nobre's lab support during data collection. Financial support for this research was provided by the Spanish Ministry of Science and Innovation (Ministerio de Ciencia e Innovación, MICINN) through a Ramón y Cajal research fellowship to MR and a research grant to both authors (ref. SEJ2007-63247). MR and PT are also supported by the Consolider-Ingenio 2010 program (ref. CSD2008-00048) and the Junta de Andalucía excellence research program (ref. P09.HUM.5422).

References

Abraham, A., Werning, M., Rakoczy, H., von Cramon, D.Y., Schubotz, R.I., 2008. Minds, persons, and space: an fMRI investigation into the relational complexity of higher-order intentionality. Conscious. Cogn. 17, 438–450.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7, 268–277.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D., 2001. Conflict monitoring and cognitive control. Psychol. Rev. 108, 624–652.

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

Camerer, C.F., 2003. Behavioural Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton.

Craig, A.D., 2009. How do you feel — now? The anterior insula and human awareness. Nat. Rev. Neurosci. 10, 59–70.

Darwin, C., 1872. The Expression of Emotions in Man and Animals. John Murray, London.

Deichmann, R., Gottfried, J.A., Hutton, C., Turner, R., 2003. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage 19, 430–441.

Dosenbach, N.U., Fair, D.A., Miezin, F.M., Cohen, A.L., Wenger, K.K., Dosenbach, R.A., Fox, M.D., Snyder, A.Z., Vincent, J.L., Raichle, M.E., Schlaggar, B.L., Petersen, S.E., 2007. Distinct brain networks for adaptive and stable task control in humans. Proc. Natl Acad. Sci. USA 104, 11073–11078.

Dosenbach, N.U., Fair, D.A., Cohen, A.L., Schlaggar, B.L., Petersen, S.E., 2008. A dual-networks architecture of top-down control. Trends Cogn. Sci. 12, 99–105.

Egner, T., 2008. Multiple conflict-driven control mechanisms in the human brain. Trends Cogn. Sci. 12, 374–380.

Egner, T., Etkin, A., Gale, S., Hirsch, J., 2008. Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. Cereb. Cortex 18, 1475–1484.

Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a non-search task. Percept. Psycho. 16, 143–149.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647.

Friston, K.J., Frith, C.D., Frackowiak, R.S., Turner, R., 1995. Characterizing dynamic brain responses with fMRI: a multivariate approach. Neuroimage 2, 166–172.

Friston, K.J., Holmes, A., Poline, J.B., Price, C.J., Frith, C.D., 1996. Detecting activations in PET and fMRI: levels of inference and power. Neuroimage 4, 223–235.

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.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. Neuroimage 7, 30–40.

Frith, C.D., Frith, U., 2006. How we predict what other people are going to do. Brain Res. 10791, 36–46.

Gallagher, H.L., Frith, C.D., 2003. Functional imaging of 'theory of mind'. Trends Cogn. Sci. 7, 77–83.

Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V., 2007. Two takes on the social brain: a comparison of theory of mind tasks. J. Cogn. Neurosci. 19, 1803–1814.

Güth, Schmittberger R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. J Econ Behav Org 3, 367–388.

Haas, B.W., Omura, K., Constable, R.T., Canli, T., 2006. Interference produced by emotional conflict associated with anterior cingulate activation. Cogn. Affect. Behav. Neurosci. 6, 152–156.

Hopfinger, J.B., Büchel, C., Holmes, A.P., Friston, K.J., 2000. A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. Neuroimage 11, 326–333.

Hyafil, A., Summerfield, C., Koechlin, E., 2009. Two mechanisms for task switching in the prefrontal cortex. J. Neurosci. 29, 5135–5142.

Lee, T.W., Dolan, R.J., Critchley, H.D., 2008. Controlling emotional expression: behavioural and neural correlates of nonimitative emotional responses. Cereb. Cortex 18, 104–113.

Levens, S.M., Phelps, E.A., 2010. Insula and Orbital Frontal Cortex Activity Underlying Emotion Interference Resolution in Working Memory. J. Cogn. Neurosci. Jan 4. [Epub ahead of print].

Lundqvist, D., Flykt, A., Öhman, A., 1998. The Karolinska Directed Emotional Faces — KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet. ISBN: 91-630-7164-9.

Luo, Q., Mitchell, D., Jones, M., Mondillo, K., Vythilingam, M., Blair, R.J., 2007. Common regions of dorsal anterior cingulate and prefrontal-parietal cortices provide attentional control of distracters varying in emotionality and visibility. Neuroimage 38, 631–639.

Mohanty, A., Engels, A.S., Herrington, J.D., Heller, W., Ho, M.H., Banich, M.T., Webb, A.G., Warren, S.L., Miller, G.A., 2007. Differential engagement of anterior cingulate cortex subdivisions for cognitive and emotional function. Psychophysiology 44, 343–351.

Ochsner, K.N., Knierim, K., Ludlow, D.H., Hanelin, J., Ramachandran, T., Glover, G., Mackey, S.C., 2004. Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. J. Cogn. Neurosci. 16, 1746–1772.

Ochsner, K.N., Hughes, B., Robertson, E.R., Cooper, J.C., Gabrieli, J.D., 2009. Neural systems supporting the control of affective and cognitive conflicts. J. Cogn. Neurosci. 21, 1842–1855.

Olsson, A., Ochsner, K.N., 2008. The role of social cognition in emotion. Trends Cogn. Sci. 12, 65–71.

Olsson, A., Phelps, E.A., 2004. Learned fear of "unseen" faces after Pavlovian, observational, and instructed fear. Psychol. Sci. 15, 822–828.

Olsson, A., Phelps, E.A., 2007. Social learning of fear. Nat. Neurosci. 10, 1095–1102 Review.

Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2004. The neural correlates of theory of mind within interpersonal interactions. Neuroimage 22, 1694–1703.

Rilling, J.K., Goldsmith, D.R., Glenn, A.L., Jairam, M.R., Elfenbein, H.A., Dagenais, J.E., Murdock, C.D., Pagnoni, G., 2008a. The neural correlates of the affective response to unreciprocated cooperation. Neuropsychologia 46, 1256–1266.

Rilling, J.K., King-Casas, B., Sanfey, A.G., 2008b. The neurobiology of social decision-making. Curr. Opin. Neurobiol. 18, 159–165.

Roberts, K.L., Hall, D.A., 2008. Examining a supramodal network for conflict processing: a systematic review and novel functional magnetic resonance imaging data for related visual and auditory stroop tasks. J. Cogn. Neurosci. 20, 1063–1078.

Rushworth, M.F., Walton, M.E., Kennerley, S.W., Bannerman, D.M., 2004. Action sets and decisions in the medial frontal cortex. Trends Cogn. Sci. 8, 410–417.

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

Saxe, R., 2006. Uniquely human social cognition. Curr. Opin. Neurobiol. 16, 235–239.

Saxe, R. in press. Theory of Mind (Neural Basis). Encyclopedia of consciousness.

Spreng, R.N., Mar, R.A., Kim, A.S., 2009. The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. J. Cogn. Neurosci. 21, 489–510.

Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18, 643–662.

Todorov, A., Baron, S.G., Oosterhof, N.N., 2008. Evaluating face trustworthiness: a model based approach. Soc. Cogn. Affect. Neurosci. 3, 119–127.

Ullsperger, M., von Cramon, D.Y., 2004. Neuroimaging of performance monitoring: error detection and beyond. Cortex 40, 593–604.

Van Kleef, G.A., De Dreu, C.W., Manstead, A.S.R., 2010. An interpersonal approach to emotion in social decision making: the emotions as social information model. Adv Exp Soc Psych 42, 45–96.

Whalen, P.J., Bush, G., McNally, R.J., Wilhelm, S., McInerney, S.C., Jenike, M.A., Rauch, S.L., 1998. The emotional counting Stroop paradigm: a functional magnetic resonance imaging probe of the anterior cingulate affective division. Biol. Psychiatry 44, 1219–1228.

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

Young, L., Saxe, R., 2007. An FMRI investigation of spontaneous mental state inference for moral judgment. J. Cogn. Neurosci. 21, 1396–1405.