Brain Advance Access published April 23, 2013

BRAIN
A JOURNAL OF NEUROLOGY

Trust versus paranoia: abnormal response to social reward in psychotic illness

Paula M. Gromann,[1,2] Dirk J. Heslenfeld,[2] Anne-Kathrin Fett,[1,2] Dan W. Joyce,[1] Sukhi S. Shergill[1,*] and Lydia Krabbendam[2,*]

1 King's College London, Institute of Psychiatry, Department of Psychosis Studies, CSI Lab, UK
2 Department of Educational Neuroscience, Faculty of Psychology and Education and LEARN! Institute, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

*These authors contributed equally to this work.

Correspondence to: Paula M. Gromann,
King's College London, Institute of Psychiatry,
Department of Psychosis Studies,
CSI Lab, UK
E-mail: p.m.gromann@vu.nl

Psychosis is characterized by an elementary lack of trust in others. Trust is an inherently rewarding aspect of successful social interactions and can be examined using neuroeconomic paradigms. This study was aimed at investigating the underlying neural basis of diminished trust in psychosis. Functional magnetic resonance imaging data were acquired from 20 patients with psychosis and 20 healthy control subjects during two multiple-round trust games; one with a cooperative and the other with a deceptive counterpart. An *a priori* region of interest analysis of the right caudate nucleus, right temporo-parietal junction and medial prefrontal cortex was performed focusing on the repayment phase of the games. For regions with group differences, correlations were calculated between the haemodynamic signal change, behavioural outcomes and patients' symptoms. Patients demonstrated reduced levels of baseline trust, indicated by smaller initial investments. For the caudate nucleus, there was a significant game × group interaction, with controls showing stronger activation for the cooperative game than patients, and no differences for the deceptive game. The temporo-parietal junction was significantly more activated in control subjects than in patients during cooperative and deceptive repayments. There were no significant group differences for the medial prefrontal cortex. Patients' reduced activation within the caudate nucleus correlated negatively with paranoia scores. The temporo-parietal junction signal was positively correlated with positive symptom scores during deceptive repayments. Reduced sensitivity to social reward may explain the basic loss of trust in psychosis, mediated by aberrant activation of the caudate nucleus and the temporo-parietal junction.

Keywords: psychosis; social cognition; trust; neuroeconomics; fMRI

Introduction

Psychosis is a disorder that manifests itself in social interactions. This is most evident in the core symptoms of psychosis, especially paranoid delusions, which are characterized by a fundamental lack of trust. Trust is an essential and inherently rewarding aspect of successful social interactions. A fundamental lack of trust has long been regarded as a primary process underlying paranoid delusions (Erikson, 1953). However, trust has not been incorporated into cognitive models of psychosis, owing to the difficulty in probing the interactive nature of social processes experimentally (Adolphs, 2006).

Received October 22, 2012. Revised January 30, 2013. Accepted February 8, 2013.

Different approaches have been implemented to study socially relevant stimuli, ranging from passive watching (Michalopoulou *et al.*, 2008) and active associative learning (Evans *et al.*, 2010) towards actual social interactions (Fett *et al.*, 2012). The current development of neuroeconomics has shown that complex social interactions, such as trust, can be operationalized in economic exchange games (Harford and Solomon, 1969; Camerer, 2003; King-Casas *et al.*, 2005; Sanfey, 2007; King-Casas *et al.*, 2008). Recent reviews suggest that neuroeconomics offers objective and suitable paradigms to investigate the underlying mechanisms of social dysfunction in psychiatric disorders (King-Casas and Chiu, 2012; Sharp *et al.*, 2012).

The classic trust game involves the interaction of two anonymous players, based upon simple investment and repayment decisions (Berg *et al.*, 1995). The first player decides how much money to share with the second player. This shared amount is tripled, and the second player has to decide how much to repay to the first player. If both players cooperate, mutually beneficial outcomes become more likely; however, the second player could benefit at the expense of the other. Thus, it allows the examination of trust quantified by the amount of money being invested. Previous studies showed that healthy control subjects invest at least some of their money, and that this sign of trust is strongly reinforced by the reciprocity of the interacting partner (Croson and Buchan, 1999; Glaeser *et al.*, 2000; Scharleman *et al.*, 2001; Phan *et al.*, 2010).

Recent imaging studies showed that economic exchange games are associated with cortical regions associated with both social cognition (Frith and Frith, 2003; Gallagher and Frith, 2003; Gallese *et al.*, 2004) and reward networks (Rilling *et al.*, 2002; Singer *et al.*, 2004; King-Casas *et al.*, 2005). Mentalizing is essential for successful social interactions, and deficits in mentalizing have been linked to poor social functioning in psychosis (Fett *et al.*, 2011). Recent imaging data support the notion that reduced activation in the temporo-parietal junction and the medial prefrontal cortex may underlie the mentalizing impairments in psychosis (Lee *et al.*, 2011). Consequently, those brain regions may play an important role in the development of disturbed social interactions and diminished trust in psychosis.

Trust has been linked with activation in brain reward systems; the caudate nucleus was specifically linked to mutually positive interactions between healthy individuals (King-Casas *et al.*, 2005). This suggests a possible mechanism underlying disturbed social interactions in psychosis, bringing into play contemporary theories of dopamine function. Mesolimbic dopamine has a central role in reward, learning and motivation (Schultz, 2002), and is also thought to be crucial to the pathophysiology of psychotic symptoms (Davis *et al.*, 1991; Seeman and Kapur, 2000). Abnormalities of dopaminergic function may lead to aberrant salience signals, possibly underlying the development of psychotic symptoms (Kapur *et al.*, 2005). This leads to the hypothesis that aberrant sensitivity to social reward may underlie the basic lack of trust in psychosis. Using a multi-round trust game, we have shown that patients with psychosis engage in fewer mutually trusting interactions than healthy control subjects (Fett *et al.*, 2012).

The purpose of this study was to investigate the lack of trust manifest in psychosis at the neural level. Functional magnetic resonance imaging data were acquired from 20 patients with non-affective psychosis and 20 healthy control subjects, while participating in two multiple-round trust games. One game was played with a counterpart designed to respond with a cooperative playing style, the other game was based on a deceptive playing style. Compared with healthy control subjects, we expected to find in patients with psychosis (i) reduced baseline trust; (ii) reduced activation in the caudate nucleus in response to cooperative repayments; and (iii) reduced temporo-parietal junction and medial prefrontal cortex signals during cooperation and deception. As a secondary aim, we examined the link between haemodynamic signal change and symptoms as well as investment behaviour to identify if observed brain activation is related to specific symptoms. For the caudate, we focused on the link with baseline trust, measured by initial investments. Examining the mean investments seemed more relevant for the medial prefrontal cortex and the temporo-parietal junction, considering that mentalizing plays a role throughout the entire interactions, rather than the first rounds. The specific hypotheses were: (i) the magnitude of the brain response in the caudate nucleus is negatively correlated with the level of paranoia scores in patients; (ii) the initial investment is positively correlated with the caudate signal in control subjects, but not in patients; and (iii) the mean investments are positively correlated with the temporo-parietal junction and medial prefrontal cortex signals in control subjects, but not in patients.

Methods and materials

Subjects

Two groups of dextral male subjects aged between 18 and 50 years participated in the study: 20 patients with lifetime presence of non-affective psychosis according to Research Diagnostic Criteria, with illness duration of <15 years, and currently treated with atypical antipsychotics, and 20 control individuals without a personal history of psychosis or a family history of psychosis. The recruitment of participants took place through the South London and Maudsley (SLAM) NHS Trust. The SLAM PICuP research register was consulted to identify suitable patients, which is a research database for patients undergoing psychological treatment at the Maudsley Hospital, London. In order to select control subjects, a database of healthy volunteers was used, which has been created for this purpose at the Institute of Psychiatry, King's College London. Exclusion criteria included: current treatment with typical antipsychotics, current drug or alcohol abuse, a history of neurological disorder and serious intellectual impairment. Individuals were also screened with the imaging safety questionnaire and were excluded if they showed any contraindications to MRI, such as metal in the body or claustrophobia. For the control group, a lifetime or a family history of psychosis was used as an additional exclusion criterion. After complete description of the study to the subjects, written informed consent was obtained. The study received ethical approval by the Barking and Havering Local Research Ethics Committee.

Assessment

Psychotic symptoms

The positive, negative and general subscales of the Positive and Negative Syndromes Scale (Kay *et al.*, 1986) were used to assess the extent of psychotic symptoms. The persecution item of the Positive and Negative Syndromes Scale was used as an additional index for patients' paranoid symptoms.

Depressive symptoms

The Beck's Depression Inventory (Beck *et al.*, 1961) was used as a measure of co-morbid depression to ensure that patients were not suffering from severe depression.

General cognition

Two additional cognitive measures were used to control for the potential impact of general cognitive impairment on trust game behaviour. The Vocabulary subtest of the Wechsler Adult Intelligence Scale III (Wechsler, 1981) was used as an index for general cognitive ability. Working memory was estimated by the Letter Number Span of the Wechsler Adult Intelligence Scale III.

Experimental design

The trust game was a modified version of a previous multi-round trust game (King-Casas *et al.*, 2005). Subjects played the role of the first player. They played against the computer, but were led to believe that they would play with two different human partners. Subjects were asked to decide how much money to share with the other player. At the beginning of each round, subjects received the same starting budget consisting of £10. Any amount between £0 and £10 could be shared. The shared amount was tripled, and the second player had to decide how much to repay to the first player.

The computer algorithm consisted of two versions programmed in a probabilistic way, which reflected a cooperative and a deceptive style of playing. The decision on how much money should be returned depended on the previous investments of the investor. Specifically, in the cooperative strategy, the first repayment was either 100%, 150% or 200% of the amount invested. Each of these possible first repayments occurred with a probability of 33%. Subsequent repayment increased in a probabilistic way if the current investment reflected an increase in trust relative to the previous investment, but remained stable in all other situations. Hence, with each increase in trust from the side of the investor, the chance of a repayment of 200% increased with 10%. In the deceptive strategy, the first repayment was 50%, 75% or 100% of the amount invested. Each of these possible first repayments occurred with a probability of 33%. Subsequent repayments decreased in a probabilistic way if the current investment reflected an increase in trust relative to the previous investment, but remained stable in all other situations. Hence, with each increase in trust from the side of the investor, the chance of a repayment of 50% invested increased with 10%.

In total, all participants played two trust games, each consisting of 20 game trials and 20 null trials. The null trials were included as a baseline condition for the functional MRI analysis. The design and duration of each event within the null trials was identical to the game trials. Participants were told that the null trials were not related to the investment decisions. In one game, the computer playing style was cooperative, and in the second it was deceptive. The order of the games was counterbalanced across subjects.

A single round was set up as follows. Every trial started with an investment cue of £10 and the investment period of the subject (maximum 6 s). The invested amount was shown (2 s), followed by waiting period with a bar slowly filling itself with dots (2–4 s), and a fixation cross (500 ms). The partner's response was displayed (3 s), followed by the totals (3–5 s depending on the length of the partner's response). Each trial ended with a fixation cross (500 ms). In total, each trial lasted 18.5 s.

Scanning parameters

Imaging data were acquired using a 3 T GE Signa Neuro-optimized MR System at the Centre of Neuroimaging Science of the Institute of Psychiatry, King's College London. A quadrature birdcage head coil was used for radio frequency transmission and reception. Foam padding was placed around the subject's head in the coil to minimize head movement. Three hundred and seventy T_2*-weighted whole-brain echo-planar images sensitive to the blood oxygen level-dependent contrast were acquired with the following parameters: slice thickness = 2.4 mm; gap = 1 mm; repetition time = 2 s; echo time = 25 ms; flip angle = 75°; in-plane resolution = 3.4 mm; number of slices = 38; number of slices/DDAs = 4; matrix = 64 × 64. For anatomical reference, a coronal fast spoiled gradient echo image of the whole brain was obtained for each subject, which consisted of 196 slices acquired with the following parameters: slice thickness = 1.1 mm; gap = 0; repetition time = 7 s; echo time = 2.8 ms; flip angle = 20°; matrix = 256 × 256.

Statistical analyses

The SPSS software, version 17, was used to analyse the behavioural data. The average of the initial investments of the first round of both games was used as an index for baseline trust. The average of all investments was calculated for each game separately as an index for overall trusting behaviour.

The imaging data were analysed using BrainVoyager QX, version 2.3 (Brain Innovation). The functional scans were coregistered to each individual anatomical scan and converted to Talairach space. Preprocessing consisted of slice scan-time correction, 3D motion correction, temporal highpass filtering (0.01 Hz), and modest temporal Gaussian smoothing (3 s). Finally, spatial smoothing using a 3D Gaussian kernel (full-width at half-maximum = 6 mm) was performed. The preprocessed functional data were then resampled in standard space, resulting in normalized 4D volume time-course data. For each subject, a protocol was created defining the onsets and offsets of the events (real versus control investments with an onset at 2 s with duration of 4 s; real versus control repayments with an onset of 10.5 s after trial start and a duration of 5 s) for the different games. Using these protocols, design matrices were computed by convolving each event with a standard haemodynamic response function. *A priori* regions of interest were defined based on the Talairach coordinates from previous

research, identifying robust reward- and mentalizing-related activation in independent samples for the right caudate nucleus (Talairach coordinates 10, 9, 4; Knutson *et al.*, 2003), the right temporo-parietal junction (Talairach coordinates 51, −54, 27; Saxe and Kanwisher, 2003) and the medial prefrontal cortex (Talairach coordinates −3, 64, 20; Hampton *et al.*, 2008). Regions of interest were created with a 5 mm sphere centred around the published coordinates. Random-effects general linear model analyses were run, based on the individual design matrices and 4D volume time-course data, but restricted to the voxels contained by the regions of interest, after correction for serial correlations. For region of interests with a significant group difference, beta weights were extracted and subjected to further *post hoc* analyses in relation to symptoms (i.e. paranoid, positive, negative and general scores) and behavioural outcomes (i.e. initial investment for the caudate and mean investments for the temporo-parietal junction). These correlation analyses were conducted using adjusted alpha levels of 0.01 per test.

Furthermore, any effect of repayment magnitude on caudate activation was analysed using repeated measures ANOVA with repayment magnitude as the within-subjects variable, and group as the between-subjects factor.

An exploratory whole-brain, voxel-wise analysis focusing on the repayment phase of the cooperative and the deceptive game was conducted to investigate if there were group wise differences in regions outside the a priori defined region of interests.

Results

Demographics

Table 1 displays the means and standard deviations for the participant characteristics within each group. To ensure that age and indices of cognitive ability were distributed equally across the two groups, ANOVAs were run, comparing the demographic information obtained from patients and control subjects. There were no significant differences between patients and control subjects in terms of age [$F(1, 38) = 0.29$, not significant], Wechsler Adult Intelligence Scale vocabulary scores [$F(1, 38) = 0.6$, not significant], and Wechsler Adult Intelligence Scale letter-number span [$F(1, 38) = 2.7$, not significant].

Behavioural results

The variance of the individual investments was examined because the algorithms for the two games were programmed such that an investment of £10 sustained throughout the game would lead to similar repayments. There was no single subject who invested the maximum of £10 throughout all trust game rounds of the two games. Table 2 provides an overview of the means and standard deviations for the behavioural analyses. There was an effect of initial investments: patients invested significantly less during the first round than control subjects [$F(1, 38) = 8.071$, $P < 0.01$], indicating reduced levels of baseline trust in patients. Patients invested significantly less during the cooperative game [$F(1, 38) = 14.431$, $P < 0.01$]. No group differences were found for the deceptive game [$F(1, 38) = 0.033$, not significant].

Functional magnetic resonance imaging

For the right caudate nucleus (Fig. 1), there was a significant game × group interaction [$F(1, 38) = 4.834$, $P < 0.04$], with stronger activation in control subjects than patients during cooperative repayments [$t(38) = 2.144$, $P < 0.04$] and no significant differences for deceptive repayments [$t(38) = -0.541$, not significant]. The strength of the caudate signal during cooperative repayments correlated negatively with patients' paranoia scores (Pearson's $r = -0.555$, $P < 0.01$; Fig. 3), but not with negative (Pearson's $r = -117$, not significant), positive (Pearson's $r = 0.168$, not significant) or general symptom scores (Pearson's $r = 0.094$, not significant). When tested with a non-parametric measure, the correlation between caudate activation and paranoia scores revealed the same trend, but was not significant at the adjusted alpha level of 0.01 (Spearman's $rho = -0.409$, $P < 0.05$).

In control subjects the caudate signal correlated positively with the magnitude of the initial investment (Pearson's $r = 0.522$, $P < 0.01$), linking healthy baseline trust with the brain reward response in control subjects. The correlation between caudate signal strength and initial investment was not significant for patients (Pearson's $r = 0.011$, not significant).

There was a significant group effect for the right temporo-parietal junction [$F(1, 38) = 5.642$, $P < 0.03$; Fig. 2], with stronger activation in control subjects than patients during cooperative repayments [$t(38) = 2.064$, $P < 0.05$] as well as during deceptive repayments [$t(38) = 2.099$, $P < 0.05$]. The strength of the temporo-parietal junction signal during deceptive repayments correlated positively with patients' positive symptom scores (Pearson's $r = 0.516$, $P < 0.01$; Fig. 4), but not with negative (Pearson's $r = 0.391$, not significant), general (Pearson's $r = 0.449$, not significant), and paranoia symptom scores (Pearson's $r = 0.292$, not significant). There were no significant correlations between the temporo-parietal junction signal during cooperative repayments

Table 1 Participant characteristics

Measure	Possible range	Mean patients (SD)	Mean control subjects (SD)
Age	18–50	33.7 (7.8)	32.2 (9.1)
WAIS vocabulary	0–66	41.5 (8.9)	43.9 (10.7)
WAIS letter-number	0–21	11.2 (2.3)	12.3 (2.2)

WAIS = Wechsler Adult Intelligence Scale-Revised; SD = standard deviation.

Table 2 Behavioural measures

Measure	Mean patients (SD)	Mean control subjects (SD)
First investment	6.1 (2.2)	7.8 (1.4)
Mean investment during cooperative game	5.8 (2.3)	8 (1.7)
Mean investment during deceptive game	4.5 (1.7)	4.4 (1.2)

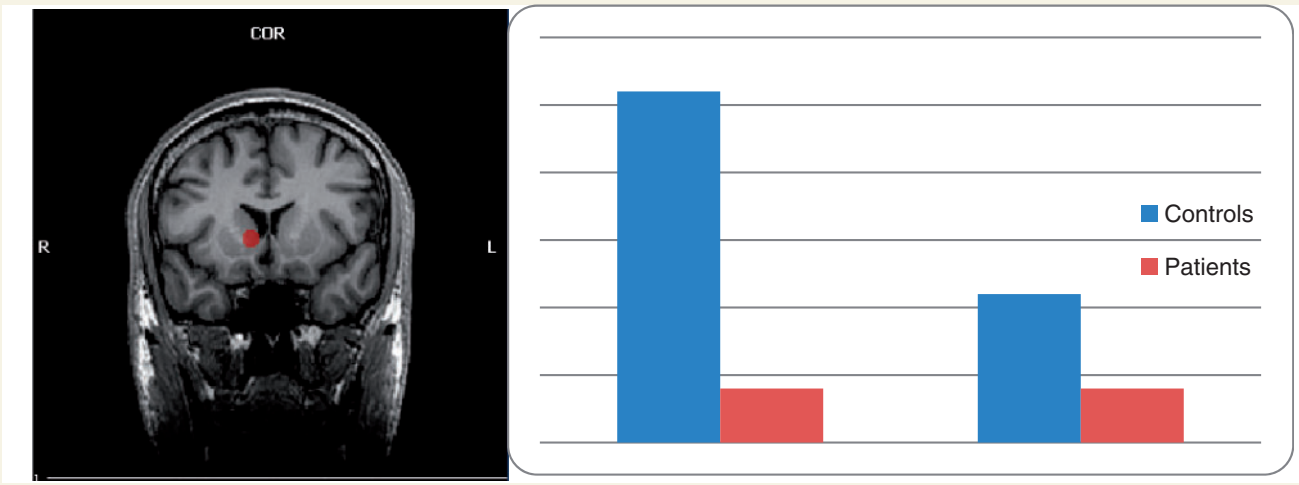

Figure 1 Location and percent signal change of the right caudate nucleus based on mean beta weights. COR = coronal; left = cooperative game; right = deceptive game.

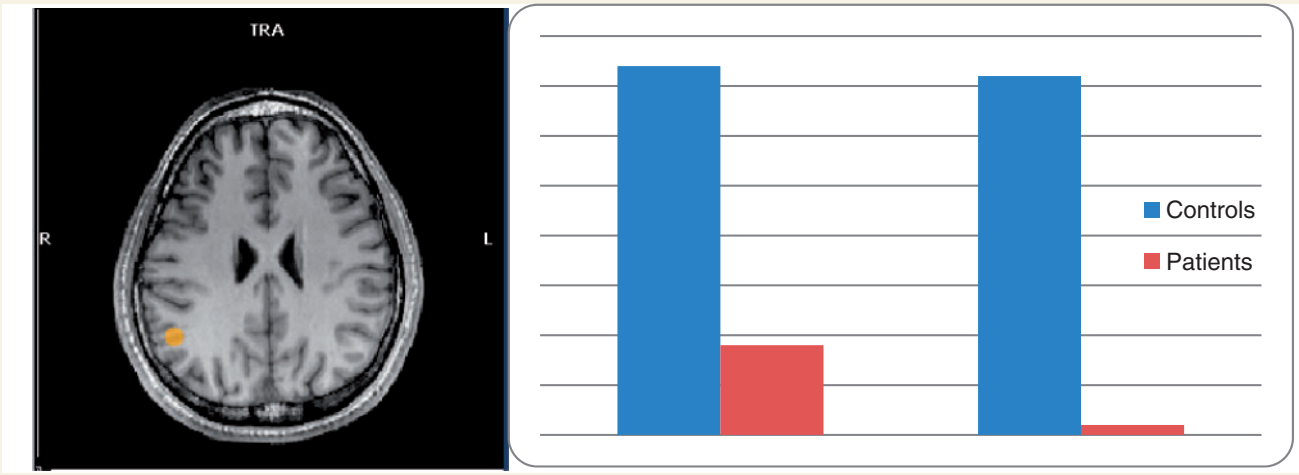

Figure 2 Location and per cent signal change of the right temporo-parietal junction based on mean beta weights. TRA = transverse; left = cooperative game; right = deceptive game.

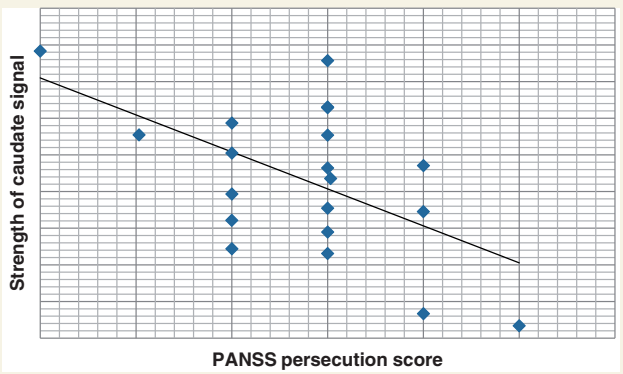

Figure 3 Scatterplot of the negative association between caudate signal strength and Positive and Negative Syndromes Scale (PANSS) persecution scores in patients.

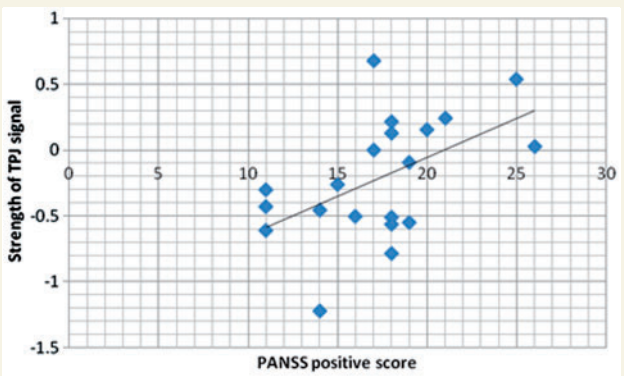

Figure 4 Scatterplot of the positive association between temporo-parietal junction (TPJ) signal strength and Positive and Negative Syndromes Scale (PANSS) positive scores in patients.

and any of the Positive and Negative Syndromes Scale symptom scores. To assess whether the observed association between temporo-parietal junction signal and positive symptoms was stronger for the deceptive than for the cooperative game, a repeated measures ANOVA was run for the patient group, yielding a trend-level significant temporo-parietal junction × positive symptoms interaction [$F(1, 37) = 3.583$, $P < 0.1$].

The temporo-parietal junction signal did not correlate significantly with the magnitude of the mean investment during the deceptive game (Pearson's $r = 0.378$, not significant).

For the medial prefrontal cortex, there was a significant main effect of game [$F(1, 38) = 7.297$, $P < 0.02$], with stronger activation for cooperative repayments than for deceptive repayments in both groups [$t(38) = 2.730$, $P < 0.01$]. There were no significant group differences for the medial prefrontal cortex [$F(1, 38) = 1.105$, not significant]. Figures 1 and 2 illustrate the size of the haemodynamic responses during cooperative versus deceptive repayments for the areas with significant group differences, i.e. the caudate nucleus and the temporo-parietal junction.

There was no significant effect of repayment magnitude on caudate activation [$F(3, 36) = 1.604$, not significant].

The exploratory whole-brain, voxel-wise analysis revealed significant task-related activation in three regions. Cooperative repayments were associated with stronger activation of the inferior parietal lobule (Talairach coordinates 44, −63, 48) and the middle temporal gyrus (61, −43, −1) in control subjects compared with patients; and deceptive repayments were associated with stronger activation of the inferior parietal lobule (Talairach coordinates 39, −53, 38) in control subjects compared with patients.

Discussion

This study examined the mechanisms underlying the lack of trust manifest in psychosis using a neuroeconomic game approach. In line with the strong link between paranoia and reduced trust, patients invested less during the first round of the games compared with control subjects. During this initial investment, subjects have no information on the behaviour of the other player, consequently, a reduced investment indicates reduced baseline trust in patients. This is in line with previous research and theories on the role of trust in psychosis (Erikson, 1953; Harford and Solomon, 1969).

Our imaging data show that receiving cooperative repayments is linked to stronger caudate activation in control subjects than in patients. The neural signal change correlated positively with the baseline trust index in control subjects, but not in patients. Combined with the finding of a negative association between paranoia scores and the strength of caudate activation, this provides a specific link between lack of trust and a reduced caudate signal in psychosis. No group differences were found for encounters with a deceptive partner. This is particularly interesting considering that the caudate forms part of the brain reward system and has been linked to greater activation in the generous condition of the trust game in healthy control subjects (King-Casas *et al.*, 2005). Consequently, this different activation pattern might suggest that patients have a reduced ability to perceive positive interactions as rewarding.

Patients also showed a reduced temporo-parietal junction signal in response to both cooperation and deception. This is in line with previous imaging data, showing impaired temporo-parietal junction activation during an on-line mentalizing task (Das *et al.*, 2012). Of note, the temporo-parietal junction has been specifically linked to mental state reasoning in a social context (Saxe and Kanwisher, 2003), in line with the notion that our subjects believed that they were interacting with real people. In the current study, the temporo-parietal junction signal change was associated with the severity of positive psychotic symptoms during deceptive repayments only, suggestive of a link between enhanced mentalizing activity during unfair social encounters and positive psychotic symptoms. However, this interpretation is based on a suggestive, but non-significant, interaction and hence requires replication in a larger sample.

Surprisingly, no group differences were established for the medial prefrontal cortex. Previous research suggests that medial prefrontal cortex impairments are directly linked to the mentalizing deficits observed in psychosis (Lee *et al.*, 2011). The lack of medial prefrontal cortex abnormalities in our study contradicts this notion. One explanation of this discrepancy might be that the medial prefrontal cortex is a better functioning region of the mentalizing network during social decision-making than the temporo-parietal junction. This would explain why patients exhibited similar medial prefrontal cortex activation as the healthy control subjects in our study, with a stronger signal for beneficial than non-beneficial social encounters. Alternatively, it is also possible that subtle medial prefrontal cortex impairments might be present in patients, which could not be detected in our study due to insufficient sample sizes.

The exploratory whole-brain analysis revealed reduced activation in patients in the inferior parietal lobule during cooperative and deceptive repayments, and additionally in the middle temporal gyrus during cooperative repayments. Abnormal activation in the inferior parietal lobule in schizophrenia has been linked to difficulties in self/other distinction and agency attribution (Shergill *et al.*, 2003, 2013; Brunet-Gouet and Decety, 2006), but given the exploratory nature of this analysis, the significance of this finding in the context of the trust game should be investigated in future studies.

The current study had a relatively moderate sample size ($n = 40$). Consequently, the results should be regarded as preliminary evidence and have to be interpreted with caution. Replication in a larger sample is required to obtain a more reliable account of the neural correlates of the lack of trust in patients with psychosis. Moreover, the generalizability of the current results is limited due to the strict inclusion criteria (i.e. only right-handed males, illness onset of <15 years, only atypical medication). However, these criteria were necessary in order to avoid potential confounding problems due to handedness, gender or medication.

One major drawback is that the design of our task does not allow for clear differentiation between social reward and more generic reward. Previous studies suggest that social reward during social interaction in the trust game can be distinguished

from utilitarian decision-making with evaluation of standard risk and reward. Recently, it has been shown that risk attitudes do not predict trust decisions during trust game interactions (Eckel and Wilson, 2004; Houser *et al.*, 2010). The neuropeptide oxytocin demonstrates specific effects on social learning, and not on learning in non-social risk games (Baumgartner *et al.*, 2008). Explicit social information has also been shown to modulate traditional reward learning systems in the striatum (Delgado *et al.*, 2005), indicating a clear distinction between social learning and reward learning. These studies support the notion that trust games tap into social rather than generic reward learning.

However, these social interactions can also be viewed as being underpinned by the mechanisms underlying reward-based learning. In accordance with this, a change in the timing of the caudate activation from the repayment phase towards the investment phase has been reported indexing the development of trust between interacting persons (King-Casas *et al.*, 2005). Other data have highlighted the correlation between social preferences and individual risk attitudes (Lauharatanahirun *et al.*, 2012), indicating that risk attitudes could influence decision-making in a social context. Combined with the finding of impaired reward prediction errors in psychosis (Murray *et al.*, 2008), this offers an alternative interpretation of the trust game paradigm, suggesting that trust game interactions may be influenced by reward processing and risk sensitivity. Future studies could usefully control for sensitivity to reward and risk in order to clarify these relationships.

By definition, the decision to trust the second player occurs at the very beginning of the trust game. Hence, higher initial investments reflect higher baseline trust. However, we chose to use the repayment phase as our point of interest because in a multi-round game, this is the time at which there is maximal mentalizing and planning for the next trial. In the current study, we found evidence for reduced baseline trust in patients, reflected by the lower initial investments compared with the healthy control subjects. Yet, it was not possible to investigate the deficit in baseline trust at a neural level due to an insufficient number of initial investment trials. Future imaging studies could overcome this using a single shot design with multiple trustees or a design with repeated investment trials without feedback as implemented in Fett *et al.* (2012).

Further research in this field should focus on risk groups such as individuals from the general population with subclinical psychotic symptoms or first-degree relatives of patients with psychosis. Previous research on first-degree relatives has revealed similar findings in the relatives as in the patients in terms of dopaminergic abnormalities (Hirvonen *et al.*, 2006; Huttunen *et al.*, 2008). Recently, evidence has been found for reduced trust in relatives at baseline, but trust levels similar to control subjects in the feedback condition, suggesting that cognitive flexibility may be a protective mechanism against transition from subclinical to clinical symptoms (Fett *et al.*, 2012). The neural basis of this transition still needs to be explored.

To conclude, we demonstrate for the first time that reduced sensitivity to social reward in psychosis is accompanied by attenuated caudate activation and this correlates with levels of paranoia. Moreover, there seems to be an impaired temporo-parietal junction signal in patients, which is linked to positive symptoms for situations of unfair social encounters. Overall, this points to aberrant reward and mentalizing mechanisms underlying disturbed social interactions in psychosis and contributing to paranoid delusions and overall symptomatology. Although speculative, this offers a new account of the origins of social cognition disturbances in psychosis. Further research on paranoia and its manifestations during social interactions is needed to gain more insight into one of the most devastating symptoms of psychosis.

Funding

Sukhwinder S. Shergill was funded by a MRC New Investigator Award, and supported by The Mental Health Biomedical Research Centre at SLAM NHS Foundation trust and King's College London. Lydia Krabbendam was funded by a VIDI grant from the Netherlands Organisation for Scientific Research (NWO). Tracey Bobin assisted with patient recruitment and testing.

References

Adolphs R. How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. Brain Res 2006; 1079: 25–35.

Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. Neuron 2008; 58: 639–50.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961; 4: 561–71.

Berg J, Dickhaut J, McCabe K. Trust, reciprocity and social history. Games Econ Behav 1995; 10: 122–42.

Brunet-Gouet E, Decety J. Social brain dysfunctions in schizophrenia: a review of neuroimaging studies. Psychiatry Res 2006; 148: 75–92.

Camerer CF. Psychology and economics. Strategizing in the brain. Science 2003; 300: 1673–5.

Croson R, Buchan N. Gender and culture: international experimental evidence from trust games. Am Econ Rev 1999; 89: 386–91.

Das P, Lagopoulos J, Coulston CM, Henderson AF, Malhi GS. Mentalizing impairment in schizophrenia: a functional MRI study. Schizophr Res 2012; 134: 158–64.

Davis KL, Kahn RS, Ko G, Davidson M. Dopamine in schizophrenia: a review and reconceptualization. Am J Psychiatry 1991; 148: 1474–86.

Delgado MR, Frank RH, Phelps EA. Perceptions of moral character modulate the neural systems of reward during the trust game. Nat Neurosci 2005; 8: 1611–8.

Eckel CC, Wilson RK. Is trust a risky decision? J Econ Behav Organ 2004; 55: 447–65.

Erikson EH. Growth and crises of the healthy personality. New York: Knopf; 1953.

Evans S, Shergill SS, Averbeck BB. Oxytocin decreases aversion to angry faces in an associative learning task. Neuropsychopharmacology 2010; 35: 2502–9.

Fett AK, Viechtbauer W, Dominguez MD, Penn DL, Van Os J, Krabbendam L. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. Neurosci Biobehav Rev 2011; 35: 573–88.

Fett AK, Shergill SS, Joyce DW, Riedl A, Strobeld M, Gromann PM, et al. To trust or not to trust: the dynamics of social interaction in psychosis. Brain 2012; 135: 976–84.

Frith U, Frith CD. Development and neurophysiology of mentalizing. Philos Trans R Soc Lond B Biol Sci 2003; 358: 459–73.

Gallagher HL, Frith CD. Functional imaging of 'theory of mind'. Trends Cogn Sci 2003; 7: 77–83.

Gallese V, Keysers C, Rizzolatti G. A unifying view of the basis of social cognition. Trends Cogn Sci 2004; 8: 396–403.
Glaeser E, Laibson D, Scheinkman J, Soutter C. Measuring trust. Q J Econ 2000; 115: 811–46.
Hampton AN, Bossaerts P, O'Doherty JP. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc Natl Acad Sci USA 2008; 105: 6741–6.
Harford T, Solomon L. Effects of a "reformed sinner" and a "lapsed saint" strategy upon trust formation in paranoid and non-paranoid schizophrenic patients. J Abnorm Psychol 1969; 74: 498–504.
Hirvonen J, van Erp TG, Huttunen J, Aalto S, Någren K, Huttunen M, et al. Brain dopamine D1 receptors in twins discordant for schizophrenia. Am J Psychiatry 2006; 163: 1747–53.
Houser D, Schunk D, Winter J. Distinguishing trust from risk: an anatomy of the investment game. J Econ Behav Organ 2010; 74: 72–81.
Huttunen J, Heinimaa M, Svirskis T, Nyman M, Kajander J, Forsback S, et al. Striatal dopamine synthesis in first-degree relatives of patients with schizophrenia. Biol Psychiatry 2008; 63: 114–7.
Kapur S, Mizrahi R, Li M. From dopamine to salience to psychosis: linking biology, pharmacology and phenomenology of psychosis. Schizophr Res 2005; 79: 59–68.
Kay SR, Opler LA, Fiszbein A. Positive and Negative Syndrome Scale (PANSS) rating manual. New York: Department of Psychiatry, Albert Einstein College of Medicine; 1986.
King-Casas B, Chiu PH. Understanding interpersonal function in psychiatric illness through multiplayer economic games. Biol Psychiatry 2012; 72: 119–25.
King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR. The rupture and repair of cooperation in borderline personality disorder. Science 2008; 321: 806–10.
King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR. Getting to know you: reputation and trust in a two-person economic exchange. Science 2005; 308: 78–83.
Knutson B, Fong GW, Bennett SM, Adams CM, Hommer D. A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. Neuroimage 2003; 18: 263–72.
Lauharatanahirun N, Christopoulos GI, King-Casas B. Neural computations underlying social risk sensitivity. Front Hum Neurosci 2012; 6: 1–7.
Lee J, Quintana J, Nori P, Green MF. Theory of mind in schizophrenia: exploring neural mechanisms of belief attribution. Soc Neurosci 2011; 6: 569–81.
Michalopoulou PG, Surguladze S, Morley LA, Giampietro VP, Murray RM, Shergill SS. Facial fear processing and psychotic symptoms in schizophrenia: functional magnetic resonance imaging study. Br J Psychiatry 2008; 192: 191–6.
Murray GK, Corlett PR, Clark L, Pessiglione M, Blackwell AD, Honey G, et al. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Mol Psychiatry 2008; 13: 239–76.
Phan KL, Sripada CS, Angstadt M, McCabe K. Reputation for reciprocity engages the brain reward center. Proc Natl Acad Sci USA 2010; 107: 13099–104.
Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C. A neural basis for social cooperation. Neuron 2002; 35: 395–405.
Sanfey AG. Social decision-making: insights from game theory and neuroscience. Science 2007; 318: 598–602.
Saxe R, Kanwisher N. People thinking about people: the role of the temporoparietal junction in "theory of mind". Neuroimage 2003; 19: 1835–42.
Scharleman J, Eckel C, Kacelnik A, Wilson R. The value of a smile: game theory with a human face. J Econ Psychol 2001; 22: 617–40.
Schultz W. Getting formal with dopamine and reward. Neuron 2002; 36: 241–63.
Seeman P, Kapur S. Schizophrenia: more dopamine, more D2 receptors. Proc Natl Acad Sci USA 2000; 97: 7673–5.
Sharp C, Monterosso J, Montague PR. Neuroeconomics: a bridge for translational research. Biol Psychiatry 2012; 72: 87–92.
Shergill SS, Brammer MJ, Fukuda R, Williams SC, Murray RM, McGuire PK. Engagement of brain areas implicated in processing inner speech in people with auditory hallucinations. Br J Psychiatry 2003; 182: 525–31.
Shergill SS, White T, Joyce DW, Bays PM, Wolpert DM, Frith CD. Modulation of somatosensory processing by action. Neuroimage 2013; 70: 356–62.
Singer T, Kiebel SJ, Winston JS, Dolan RJ, Frith CD. Brain responses to the acquired moral status of faces. Neuron 2004; 41: 653–62.
Wechsler D. Wechsler adult intelligence scale-revised. New York: Psychological Corporation; 1981.