Chapter 5

Learning to trust:

Social feedback normalises trust behaviour in first episode psychosis and clinical high-risk

Imke L.J. Lemmers-Jansen , Anne-Kathrin J. Fett, Esther Hanssen, Dick J. Veltman, and Lydia Krabbendam

Published in
Psychological Medicine (2018), p 1-11

Abstract

Background: Psychosis is characterised by problems in social functioning that exist well before illness onset, and in individuals at clinical high-risk (CHR) for psychosis. Trust is an essential element for social interactions that is impaired in psychosis. In the trust game, chronic patients showed reduced baseline trust, impaired response to positive social feedback, and attenuated brain activation in reward and mentalising areas. We investigated whether first episode psychosis patients (FEP) and CHR show similar abnormalities in the neural and behavioural mechanisms underlying trust.

Methods: Twenty-two FEP, 17 CHR, and 43 healthy controls performed two trust games, with a cooperative and an unfair partner in the fMRI scanner. Region of interest analyses were performed on mentalising and reward processing areas, during the investment and outcome phases of the games.

Results**:** Compared to healthy controls, FEP and CHR showed reduced baseline trust, but like controls, learned to trust in response to cooperative and unfair feedback. Symptom severity was not associated with baseline trust, however in FEP associated with reduced response to feedback. The only group differences in brain activation were that CHR recruited the temporo-parietal junction (TPJ) more than FEP and controls during investment in the unfair condition. This hyper-activation in CHR was associated with greater symptom severity.

Conclusions: Reduced baseline trust may be associated with risk for psychotic illness, or generally with poor mental health. Feedback learning is still intact in CHR and FEP, as opposed to chronic patients. CHR however show distinct neural activation patterns of hyper-activation of the TPJ.

Keywords:

First episode psychosis, clinical high-risk, trust, social feedback, fMRI

Introduction

Psychosis is characterised by problems in social functioning (Couture et al., 2006; Fett et al., 2012). Lower social functioning is already present in childhood in individuals who continue to develop psychosis and has also been reported in individuals at high-risk for psychosis (Ballon et al., 2007; Corcoran et al., 2011; Cornblatt et al., 2007; Velthorst, Fett, et al., 2016; Velthorst, Reichenberg, et al., 2016; Yung et al., 2003). Clinical high-risk patients (CHR) are already in care for other psychopathology, reporting psychotic-like symptoms, but have not yet experienced – and maybe never will –full-blown psychosis. In CHR, the developmental course of social functioning is predictive of the conversion to psychosis (Cannon et al., 2008; Jang et al., 2011; Niendam et al., 2007). Understanding the mechanisms underlying deficits in social functioning in at-risk states and first-episode psychosis (FEP) is crucial for understanding transition and outcome prognosis. Intervening at these early stages targeting social functioning can improve outcome and possibly delay (or prevent) transition.

Social functioning relates to establishing relationships, both vocational and private (Velthorst, Fett, et al., 2016; Velthorst, Reichenberg, et al., 2016). Patients show a steep decline in these domains starting about five years before illness onset. The basis of social functioning is the ability to interact in an appropriate way with other people. Previous research studying online social interactions in psychosis has suggested two possible explanatory mechanisms for impairments in social interactions; these are a reduced sensitivity to rewarding effects of social contact (Campellone et al., 2016; Fett et al., 2012; Gromann et al., 2013), and an impaired social cognitive ability (Csukly, Polgár, Tombor, Réthelyi, & Kéri, 2011; Horat et al., 2017), including impaired mentalising (Green et al., 2015). Social cognitive skills (Couture et al., 2006; Green & Leitman, 2008) are necessary for the formation and maintenance of relationships and for building trust in other people. Like patients with psychosis, CHR show deficits in a variety of these skills (Bora & Pantelis, 2013; Lavoie et al., 2013; McCleery et al., 2014), albeit to a lesser degree. Research has mainly focused on offline cognitive skills, without investigating them in real interactions. In the last decade, interactive designs have been widely used, that have the strength to capture social cognitive skills, as well as the rewarding effects of social behaviour in an online setting. We therefore investigated cooperative and unfair social interactions and the neural correlates of trust, directly comparing FEP and CHR to controls, using an interactive trust game to test whether these groups display similar underlying mechanisms of reduced social interactions.

The trust game investigates real-time social interactions (Berg et al., 1995). In the game, the first player (investor), receives a certain endowment, e.g., €10. He or she can give any amount between €0 and €10 to the second player, the trustee. The given

amount is tripled and the trustee then can return any part of this amount to the investor. The best payoff for the trustee is reached by keeping the money. Thus, investing requires trust that a fair repayment will be made. The iterative game allows for the investigation of baseline trust (i.e. first investment), and the development of trust based on cooperative and unfair social feedback. Key-processes involved in the trust game are thought to be mentalising (Declerck et al., 2013; Frith & Frith, 2006; Gallagher & Frith, 2003) and reward processing (Fehr & Camerer, 2007; King-Casas et al., 2005; Rilling & Sanfey, 2011). Mentalising appears to be important during both the investment and repayment phase, where estimations of the other's behaviour are made. Reward learning signals have been shown to shift from the repayment phase to the investment phase in an iterative trust game (King-Casas et al., 2005). Hence, we investigated both the investment and repayment phase (Figure 1).

Research in healthy subjects has shown that participants initially invest more than half of their endowment (Berg et al., 1995; Johnson & Mislin, 2011). Studies from our lab have shown that baseline trust tends to be lower in patients than controls (Fett et al., 2016; Gromann et al., 2013). Both positive (Fett et al., 2012) and negative (Fett et al., 2016) symptoms have been associated with lower baseline trust, suggesting that reduced trust may reflect either paranoia or a lack of social motivation. The ability to learn from social feedback seems to depend on context (cooperative or unfair partner's responses) and illness duration: Early psychosis patients were able to adjust their trust to similar levels as controls, whereas chronic patients showed an insensitivity to positive feedback. In unfair interactions, early and chronic psychosis patients responded adequately to negative feedback (Campellone et al., 2016; Fett, Gromann, et al., 2015; Fett et al., 2012; Fett et al., 2016; Gromann et al., 2013). Understanding the mechanisms of trust in early psychosis stages may provide insights in focal points to target in social functioning interventions.

At the neural level, reduced caudate activation in chronic patients has been reported in cooperative interactions. Relatives of patients with psychosis, despite behavioural outcomes similar to controls, also showed reduced recruitment of the caudate and insula. These results possibly reflect reduced sensitivity to social reward processing mechanisms in both patients and relatives (Gromann et al., 2013; Gromann et al., 2014), which could account for social impairments. Associations of neural activity with positive symptoms have been reported (Gromann et al., 2013).

This study set out to investigate whether CHR and FEP patients, similar to chronic patients, show reduced baseline trust and to explore the neural mechanisms underlying trust behaviour in these patient groups. Based on the existing trust game literature, we hypothesised that similar to relatives (Fett et al., 2012) and (chronic) patients (Fett,

Gromann, et al., 2015) 1) FEP and CHR will show lower levels of baseline trust, and 2) CHR and FEP are able to learn from positive and negative feedback given by the counterpart and adjust their levels of trust accordingly (Fett, Gromann, et al., 2015; Gromann et al., 2013). In both cases, we expected CHR to perform in between FEP and controls. In addition, the associations of symptoms with baseline trust and changes in trust in FEP and CHR were examined. On the neural level we hypothesised to find 3) attenuated activation in brain areas associated with mentalising and reward (learning) in FEP compared to controls. Based on the trust literature in relatives and imaging research in CHR (Smieskova et al., 2013), we expected intermediate activation in CHR. Based on the findings by Gromann et al. (2013), we hypothesised to find 4) positive symptoms related positively to brain activation in mentalising and negatively to reward areas in both patient groups. In addition, associations of brain activations with negative symptoms were investigated.

Methods

Subjects

Twenty-six FEP patients with non-affective psychosis, aged 16-22, and 17 CHR, aged 16-31, were recruited in the Amsterdam and The Hague area. Forty-nine healthy control participants (aged 16-31) were recruited to match both patient populations on age and gender. Patients were contacted through their caregivers at the academic medical center Amsterdam (AMC), the Amsterdam early intervention team psychosis and PsyQ, The Hague. FEP were diagnosed at the AMC, according to the DSM-IV criteria (American Psychiatric Association, 2000). FEP were included within 18 months of the diagnosis (mean 5.6 months). CHR were help seeking individuals that were referred to PsyQ by their general practitioners or other mental health institutions. After an initial diagnosis based on their complaints, all new admissions (between age 14-35) were screened for an "at-risk mental state" (ARMS) with the Comprehensive Assessment of At-Risk Mental States [CAARMS; (Yung et al., 2005)], a semi-structured interview that assesses psychotic experiences in the last year before assessment. Only the positive symptoms sub-scale was used, and both intensity and frequency of the symptoms were assessed. Patients met ARMS criteria for attenuated psychotic symptoms (a) with subthreshold intensity, when scoring 3-5 on severity and 3-6 on frequency of symptoms (all CHR participants), (b) or based on sub-threshold frequency, when scoring 5-6 on severity and 3 or above on frequency (N = 2, in combination with intensity), or based on vulnerability, i.e., schizotypal personality disorder, familial history of psychosis, or a drop in social functioning [N = 3, in combination with intensity; for details, see (Yung et al., 2005), p. 966]. Patients were diagnosed as having a full-blown psychosis when

scoring higher than 6. Additionally, patients had to display marked problems in socially useful activities (work and study), relationships, and self-care, indicated by a score below 55 on the Social and Occupational Functioning Assessment Scale [SOFAS; mean score 46.9; (Goldman et al., 1992; Morosini et al., 2000); see also (Rietdijk et al., 2012)]. CHR were included within one year after CAARMS assessment (mean 4.8 months). Fourteen CHR patients also took part in a larger study (EU-GEI), with post-measurements at 6, 12 and 24 months using the CAARMS. Of two CHR participants, follow-up data are missing. Symptoms of depression and anxiety are often the primary presenting complaints of CHR patients, rather than the attenuated psychotic symptoms (Modinos et al., 2014). Similar to other CHR samples (Fusar-Poli et al., 2014; Kelleher et al., 2012; Morrison et al., 2012; Wigman et al., 2012; Woods et al., 2009), the current CHR sample had comorbid diagnoses of anxiety (5), personality (3), eating (2) and mood (2) disorders, trauma (2), and ADHD (3). Exclusion criteria for all participants were an IQ < 80 and contraindications for scanning. For FEP additional exclusion criteria were a primary diagnosis of a mood disorder, and comorbidity with autism spectrum disorder. Healthy control participants were excluded if they had a (family) history of psychopathology, which was assessed with self-report, and by a systematic interview with questions regarding past and present mental help seeking, depressed and psychotic symptoms, and intake of medication.

We excluded four FEP and six controls due to invalid behavioural data (1 FEP and 2 controls), unusable or missing imaging data (3 FEP and 4 controls). The analysis sample consisted of 22 FEP, 17 CHR and 43 controls.

Measures

Trust game

Participants played the role of investor in two multi-round trust games. They were told that they were connected to their anonymous counterpart via the Internet. In reality, they played against a computer, programmed to respond either in a cooperative or in an unfair way. In the cooperative condition, the return was 100%, 150% or 200% of the invested amount, with increasing likelihood of a 200% repayment after each increase of investment. In the unfair condition the return was 75% or 50%, with increasing likelihood of a 50% repayment after increase of investment. The two games were presented in counterbalanced order. Each game consisted of 20 experimental and 20 control trials (Figure 1). For a detailed description of the paradigm see Lemmers-Jansen et al. (2017). After the trust game, a questionnaire to investigate participants' opinions on the behaviour of their counterpart was administered, to check if participants believed that they were playing a real person.

Figure 1 Graphical Overview of the Trust Game

Note: Top row represents the visual stimuli in the game trials; middle row are the separate phases of the trust game, including durations; bottom row represents the visual stimuli in the control trials. Printed with permission of Lemmers-Jansen et al. (2017).

Positive and Negative Syndrome Scale [PANSS (Kay, Fiszbein, & Opler, 1987)]

The 30-item PANSS semi-structured interview was used for rating symptoms in the two weeks prior to testing. The PANSS distinguishes between positive, negative, and general symptoms (Kay et al., 1987). Items are scaled on a 7-point Likert scale, ratings 3 and higher indicating clinical values. All FEP and 13 CHR completed the interview. Interviews were taken by four researchers and audio tapes (and if consented video tapes) were made. Responses were rated on the basis of the recordings and notes taken during the interview, by two researchers. Based on the first participants an inter-rater reliability was calculated (r = .85). All PANSS data were rated by the same two researchers.

Wechsler Adult Intelligence Scale [WAIS; (Wechsler, 1997)]

To control for confounding effects of intelligence, the vocabulary subscale of the WAIS was included, a measure of verbal comprehension, consisting of 33 words that had to be defined by the participants. Answers were coded as fully correct (2 points), partially correct (1) or wrong (0), resulting in a maximum score of 66.

Trust Manipulation Check Questionnaire

The trust game was followed by a questionnaire to investigate participants' opinions on the behaviour of their counterpart, and to check if they believed that they were playing a real person ("What do you think this task was about?", "What were the main causes that influenced your behaviour during the task?", "Did you use a strategy during the tasks? If so, which strategy?", and "Did you think that the counterparts made fair

choices?"). The last questions required answers specified per counterpart. If a participant referred to the two counterparts as persons in his/her responses, we regarded the manipulation as successful. If participants reported on any of the questions that they had doubts, or did not believe that the counterpart was real, the manipulation was coded as failed. Four controls, three FEP and one CHR did not believe the manipulation.

Procedure

This research was approved of by the Medical Ethical Committee of the VU Medical Center Amsterdam. All participants signed an informed consent, and completed several questionnaires that are not included in the current paper. We then administered the PANSS to FEP and CHR patients. Prior to scanning, participants received oral and written instructions for the trust game, illustrated with screenshots of the game to illustrate the task. Participants played several practice rounds on a computer, accompanied by additional feedback, to ensure understanding of the task. Subsequently participants were scanned for about an hour. First, participants performed the trust game, followed by the structural scan. After this period of relative rest, they performed a second task [the Social Mindfulness Paradigm, to be reported in a separate paper (in preparation)], followed by a resting state scan. Immediately after scanning, a manipulation check for the trust game was administered (see Methods, Trust Manipulation Check Questionnaire). Participants received an image of their brain, 25€ for participation and reimbursement of their travel costs.

fMRI data acquisition

Imaging data were obtained at the Spinoza Center Amsterdam, using a 3.0 T Philips Achieva whole body scanner (Philips Healthcare, Best, The Netherlands) equipped with a 32 channel head coil. A T2* EPI sequence (TR = 2.31, TE = 27.63, FA = 76.1º, FOV 240mm, voxel size 2.5 x 2.5 x 2.5, 40 slices, 0.3 mm gap) was used, which resulted in 325 images per condition. A T1-weighted scan was obtained for anatomical reference (TR = 8.2, TE = 3.8, FA = 8º, FOV 240*188mm, voxel size 1 x 1 x 1, 220 slices).

Data Analysis

Behavioural data

Demographic and behavioural data were analysed using Stata 13 (StataCorp, 2013) with regression analyses and chi-square tests. We analysed group differences in first investment (baseline trust), and the development of investments (changes in trust) across repeated interactions (indicated by trial number) in each game (cooperative and unfair). We used multilevel random regression analyses to account for multiple observations [investments (level 1); within participants (level 2)]. All analyses were controlled for age and WAIS score.

Imaging data

Imaging data were analysed using Statistical Parametric Mapping 8 (SPM, 2009). Functional images for each participant were pre-processed as follows: realign and unwarp, co-registration with individual structural images, segmented for normalization to an MNI template and smoothing with a 6 mm Gaussian kernel (FWHM). At fist-level, a general linear model was used to construct individual time courses for the investment and repayment phase per condition, using an event-related design. For each trial we defined the investment as the period of stimulus onset to the moment of investment, and the repayment phase as the period during which the partner's return was displayed (Figure 1). Trials from both the cooperative and unfair conditions were contrasted with control trials. Additionally, cooperative trials were contrasted with unfair trials, to directly compare the differences in response to cooperative and unfair feedback.

A priori ROI analyses were performed. ROIs were derived from Gromann et al. (2014). Talairach coordinates were converted to MNI space (tal2mni under MatLab), resulting in the following ROIs: right caudate (MNI coordinates 16, 17, 7), superior temporal sulcus (STS; 62, -58, 5) and TPJ (51, -57, 26), left insula (-33, 14, 0), and medial prefrontal cortex (mPFC; -3, 65, 25). We tested group differences using MarsBaR (version 0.43; http://MarsBaR.sourceforge.net). An adjusted *p*-value was calculated, taking the correlation between the β-values into account by using the Simple Interactive Statistical Analysis Bonferroni tool resulting in adjusted *p*-values [see Table 2; see (http://www.quantitativeskills.com/sisa/calculations/bonfer.htm); (Li et al., 2014; Woudstra et al., 2013)]. Additional whole brain analyses were performed, to investigate activation outside the predefined ROIs.

Associations with symptoms

Group differences in the association of first investment and development of investments with symptoms (paranoia item, positive and negative PANSS subscales) were investigated. The persecution item (P6) and the depression item (G6) were used

as an additional index for paranoid ideation, as previously reported (Gromann et al., 2013), and depression, based on CHR comorbidity. Second, beta weights of the ROIs (average over all voxels) were associated with symptoms. To further explore the association of symptoms with behaviour and brain activation, additional post-hoc analyses within patient groups were performed.

Table 1

Participant Characteristics and Baseline Trust in the Trust Game

		Controls	CHR	FEP
		N = 43	*N* = 17	*N* = 22
Gender, *N male* (%)		22 (51.16 %)	7 (41.18 %)	14 (63.64 %)
Age, mean (*SD)*		21.06 (2.74)	23.78 (2.42) **	19.88 (1.54)
WAIS vocabulary, mean (*SD)*		41.77 (11.39)	41.71 (12.16)	32.5 (10.53) **
PANSS	total *(SD)*		58.92 (11.84)	61.32 (14.51)
	mean (*SD)*		1.95 (.37)	2.09 (.51)
- Positive	total (*SD*)		13.38 (2.59)	13.23 (5.72)
	mean (*SD)*		1.91 (.39)	1.89 (.82)
- Negative	total (*SD*)		13.69 (3.73)	17.18 (5.85)
	mean (*SD)*		1.96 (.53)	2.45 (.84)*^
- General	total (*SD*)		31.85 (6.07)	30.91 (7.66)
	mean (*SD)*		2.0 (.38)	1.93 (.48)
- Paranoia	mean (*SD)*		3.38 (1.50)	2.23 (1.41)*
- Depression	mean (SD)		4 (1.73)	2.89 (1.49)
Medicated *N* (%)			8 (47%)*	16 (73%)
- Atypical anti-psychotics (%)				13 (81.5%)
- Typical and atypical antipsychotics (%)				1 (6.25%)
- Anti-depressant (%)			3 (37.5%)	-
- SSRI (%)			3 (37.5%)	-
- Benzodiazepine (%)			2 (25%)	1 (6.25%)
- Sertraline (%)			-	1 (6.25%)
Baseline trust, mean (*SD)*		7.02 (1.81)**	5.82 (2.32)	5.52 (2.02)

* = significant difference between FEP and CHR at $p < .05$

** = significantly different from both other groups at $p < .05$

*^ = FEP > CHR at $p < .07$

Note: CHR = clinical high-risk; FEP = first episode psychosis; SD = standard deviation; WAIS = Wechsler Adult Intelligence Scale; PANSS = Positive and Negative Syndrome Scale; SSRI = selective serotonin reuptake inhibitors. The paranoia item forms part of the positive subscale (P6), the depression item forms part of the general subscale (G6). For analyses, these item were investigated separately.

Results

Participant characteristics

Participant characteristics and baseline trust are displayed in Table 1. CHR were significantly older than controls (β = .40, p < .001) and FEP (β = .57, p < .001), and FEP scored significantly lower on the WAIS vocabulary scale than controls (β = -.39, p = .003) and CHR (β = -.31, p = .01). The time between diagnosis and inclusion in the study did not differ significantly between FEP and CHR (β = -.35, p = .21).

Behavioural results

Based on the group differences, all analyses were controlled for age and WAIS vocabulary score. Four controls, three FEP and one CHR did not believe they played against a human counterpart. Analyses without these subjects yielded similar results. Below the results of the complete sample are reported.

Group differences in baseline trust - the first investment of the first game - were found (β = -.27, p = .02; Table 1), with FEP and CHR showing lower baseline trust than controls (FEP: β = -0.24, p = .04; CHR: β = -0.25, p < .05). CHR and FEP did not differ significantly from each other (p = 0.8).

To investigate the development of trust over trials we performed a three-way interaction "trial number-by-group-by-condition" on investment. This interaction was not significant and therefore removed from the model. Significant trial number-by-group, condition-by-group and trial number-by-condition interactions on investment were found (b = .03, 95% CI [.031, .05], p = .001; b = .49, 95% CI [.28, .71], p < .001; b = -.20, 95% CI [-.24, -.17], p < .001, respectively), indicating that the development of trust differed between groups (Figure 2).

In the *cooperative condition* there was a significant group-by trial number interaction on investment (b = .03, 95% CI [.006, .05], p = .01), with FEP showing significantly stronger increase than controls (b = .08, 95% CI [-.07, .08], p = .03). Controls and CHR did not differ significantly from each other (p = .9). Analysis by group showed that all groups increased investments significantly (all p's <.01; see Figure 2a). In the *unfair condition*, analysis revealed a significant group-by trial number interaction on investment (b = .03, 95% CI [.01, .06], p = .005), with FEP showing significantly less decrease than the other groups (b = -.97, 95% CI [-1.89, -.06], p = .04). All groups decreased investments significantly (all p's <.001; see Figure 2b).

a b

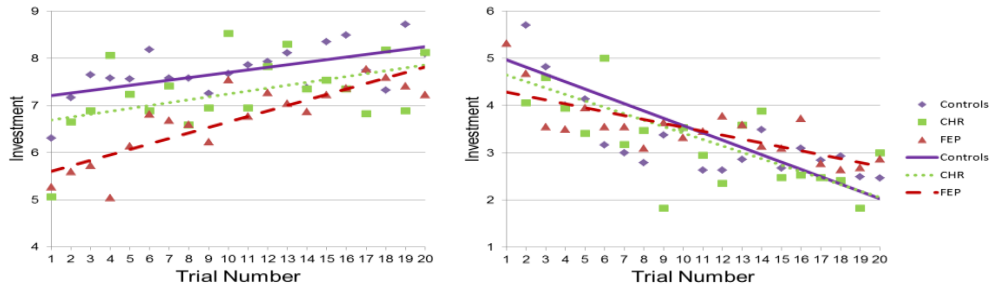


Figure 2 Changes in Investment over Trials in the (a) Cooperative and (b) Unfair Condition

All analyses were also conducted with medication type (no medication; atypical anti-psychotics; combination of typical and atypical; other psychotropic medication) as a grouping variable. No differences in baseline trust, nor in adjustment of trust were found between the medication groups, and no interactions with symptoms on trust were found.

Imaging results

ROI analyses revealed significant group differences in the right TPJ only, showing more activation in CHR compared to controls and FEP during the investment phase in the unfair condition (Table 2).

Furthermore, CHR activated the TPJ and mPFC more than controls, when investing in an unfair partner compared to a cooperative partner. Other ROIs showed no group differences. During cooperative investment, cooperative and unfair repayment no significant group differences in ROI activation were found. All ROI analyses were also conducted between medication groups. No significant differences in activation were found.

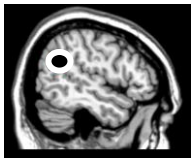
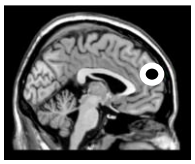
Additional exploratory whole brain analyses, based on a significance level of $p < .05$ FWE cluster corrected, did not reveal significant group differences. Results with a more lenient threshold are presented in Supplementary Table S1.

Symptoms

CHR and FEP did not differ significantly from each other in terms of overall and positive symptoms, and on the depression item. Only on the paranoia item, CHR scored significantly higher than FEP ($\beta = .37$, $p = .03$). There was a trend towards significance indicating that FEP had higher negative symptoms than CHR ($p < .07$).

Table 2

Region of Interest (ROI) Analyses Outcome

Condition	ROI	*p*	Contrast	Location
Unfair Investment *				
	TPJ	.019	CHR > FEP	
	TPJ	<.001	CHR > Controls	TPJ
Unfair investment > cooperative investment **				
	TPJ	.005	CHR > Controls	
	mPFC	.007	CHR > Controls	mPFC

Note: * = adjusted significance level for multiple comparisons of $p = .027$.

** = adjusted significance level of $p = .012$.

Montreal Neurological Institute (MNI) coordinates: TPJ = temporo-parietal junction, 51, -57, 26; mPFC = medial prefrontal cortex, -3, 65, 25. CHR = clinical high-risk; FEP = first episode psychosis

Associations between behavioural outcomes and symptoms

No group-by-symptoms interactions on first investment were found (all *p*s > .24). After removing the interaction from the model, no main effects of symptoms on baseline trust were found (all *p*s > .16).

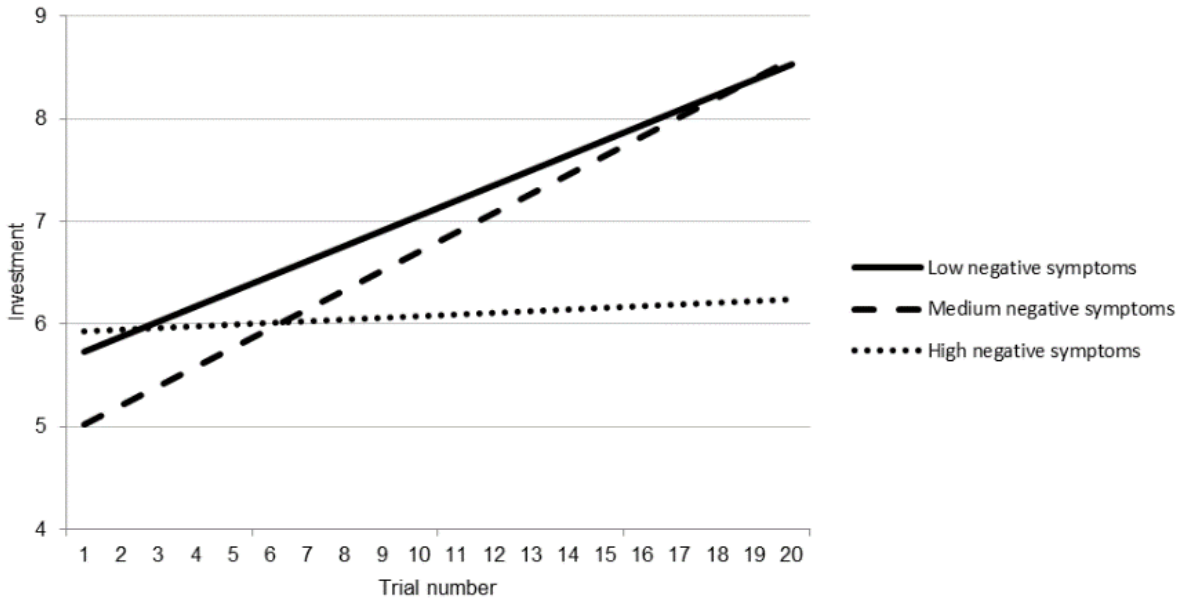


Figure 3 Interaction Between Trial Number and Negative Symptoms on Investment in the Cooperative Condition in First Episode Psychosis Patients (FEP)

In the *cooperative condition*, the group-by-trial number-by-symptoms models showed a significant interaction for negative symptoms only (b = -.02, 95% CI [-.03, -.001], p = .03), indicating that negative symptoms impacted upon the development of trust differentially in the three groups. Post-hoc analyses showed a significant association between symptoms and changes in investments over trials in FEP (b = -.01, 95% CI [-.02, -.003], p = .004), but not in CHR. To visualise this association, we divided the negative symptoms in three levels (Figure 3). Analysis indicated that the only highest level of negative symptoms interfered with increasing investments. No significant interactions with positive symptoms, paranoia, or depression were found.

In the *unfair condition* the group-by-trial number-by-symptoms models did not show significant interactions. After removing the 3-way interaction from the model, the interaction of positive symptoms with trial number became significant (b = .01, 95% CI [.0002, .012], p = .04). Higher positive symptoms were associated with less decrease in investments in FEP and CHR. Associations of decreasing investment with negative symptoms, paranoia, and depression showed no significant group differences. Analyses within each patient group revealed a significant association between depression and investment over trials in FEP (b = .03, 95% CI [.007, .056], p = .01), showing that FEP with a more severe depression score adjusted their investment less to the negative feedback than FEP with milder depression scores.

Associations of ROI beta weights and symptoms

Beta weights of the ROIs showing group differences (Table 2) were correlated with the positive and negative subscales, and PANSS paranoia and depression score. No group differences were found in the association between symptoms and ROI activation. After removing the group-by-symptoms interaction from the model there was a positive association at trend level for TPJ activation and paranoia in the unfair investment phase (β = .32, p = .07), indicating that in both patient groups the TPJ was increasingly activated in patients with higher paranoia.

Beta weights of the TPJ were unevenly distributed. Therefore, Spearman rank correlation was used for the exploratory analyses per group. This analysis revealed significant associations between symptoms and beta weights of the TPJ for CHR, but not for FEP. Specifically, CHR showed a significant positive association between the paranoia item, positive, and negative symptoms, and TPJ activation during unfair investments (ρ = .57, p = .04; ρ = .70, p = .008; ρ = .64, p = .02, respectively). No associations between symptoms and mPFC activation were found.

Discussion

This study investigated the behavioural and neural mechanisms associated with trust and the association with symptoms in a high-risk (CHR) and first episode psychosis (FEP) sample using an interactive trust game. Participants played two trust games, with a cooperative and an unfair pre-programmed partner. Behaviourally, FEP and CHR only differed from controls, and not from each other, showing reduced basic trust, that is initial trust before partner feedback is revealed. No impairments in the development of trust in response to feedback over trials were found in either patient group, compared to controls. Only in FEP associations between trust development and symptoms were found. On the neural level, CHR recruited the TPJ more than the other groups during investment in the unfair condition, suggesting differential processing as compared to healthy controls and FEP.

Behavioural mechanisms of trust

Importantly, and in line with previous research both FEP and CHR showed reduced baseline trust toward unknown others (Fett, Gromann, et al., 2015; Fett et al., 2012; Fett et al., 2016; Gromann et al., 2013). Reduced baseline trust has been found in individuals at genetic risk for psychosis, (Fett et al., 2012), but contradicted by another (Gromann et al., 2014). Contrary to previous research (Fett et al., 2012; Fett et al., 2016) baseline trust in FEP and CHR was not associated with symptom severity. The association with positive symptoms found by Fett et al. (2012) was at trend level (.09), providing only tentative support. The fact that reduced baseline trust has also been found in individuals at genetic risk for psychosis and now in CHR in combination with the lack of an association with symptoms, tentatively suggests that reduced baseline trust is linked to the risk for psychosis trait, rather than a consequence of the illness (a state marker) that would be associated with (temporal fluctuations of) symptoms.

Feedback learning in cooperative and unfair interactions is still intact in CHR and FEP, as opposed to chronic psychosis (Campellone et al., 2016; Fett, Gromann, et al., 2015). The development of trust in response to positive feedback by the game partner showed, as predicted, that FEP and CHR increased their levels of trust significantly. The same pattern was found in response to negative feedback: over game rounds FEP and CHR decreased their trust to the same level as controls. FEP showed steeper increase in positive interactions than controls, possibly because they were more sensitive to the effects of the positive feedback given they initially had lower expectations, as reflected in lower baseline trust. Furthermore, a ceiling effect might result in a less steep increase for controls, with 23% of the control participants investing the maximum of 10 in 75% or more of the trials. The slightly reduced response to negative feedback in FEP,

resembling results of Fett et al. (2016), might be explained by the differences in first investment (FEP starting significantly lower).

Symptoms

Symptom severity on average was similar in the two patient groups. FEP showed substantial variability in symptom severity, which reflects the fact that we included both hospitalised patients and ambulant patients who were in a rehabilitation trajectory. CHR experienced more paranoia than FEP, and FEP tended to have more negative symptoms than CHR. These differences might be explained by medication effects: 64% of the FEP were on atypical antipsychotic medication, probably dampe-ning positive symptoms, whereas 47% of the CHR was on other psychotropic medication (see Table 1). First episode patients with highest negative symptoms, as opposed to milder symptoms, showed almost no adjustment of trust in response to positive feedback. Intact feedback learning mechanisms in FEP were associated with milder negative symptom severity. The association between negative symptoms and problems in social functioning and responding to feedback has been well established in psychosis (Addington & Addington, 2005; Campellone et al., 2016; Milev et al., 2005; Strauss et al., 2013; Voges & Addington, 2005; Waltz, Frank, Wiecki, & Gold, 2011), possibly reflecting a lack of (social) motivation or depression. In the unfair condition positive symptoms were associated with less decrease in trust in both FEP and CHR. This suggests that positive symptoms interfere with learning from negative social feedback, contradicting earlier findings that found no associations of positive symptoms with learning to trust (Fett et al., 2016).

Neural mechanisms of trust

On the neural level CHR activated the TPJ significantly more than the other two groups. The TPJ forms part of the mentalising system (Fletcher et al., 1995; Frith & Frith, 2006; Van Overwalle, 2009), and was previously found to be activated in the trust game (King-Casas et al., 2005; Frank Krueger et al., 2007; Saxe & Kanwisher, 2003; Van den Bos et al., 2011). In chronic patients reduced TPJ activation was associated with more positive symptoms (Gromann et al., 2013). In our sample, TPJ activation did not differ between FEP and controls, suggesting a decline in TPJ response with longer illness duration. CHR however, showed increased activation in this area compared to FEP and controls during unfair investment. CHR also showed more TPJ and mPFC activation than controls during investments towards the unfair counterpart, as compared to cooperative counterpart. Since both areas form part of the mentalising system, this could suggest that unfair interactions elicit increased mentalising in CHR. Gromann et al. (2013), in contrast,

found the mPFC to be activated more in cooperative interactions in both patients and controls. Increased neural activation in patients at-risk for psychosis during mentalising and emotion processing areas despite similar behavioural performance was previously found (Brüne et al., 2011; Derntl et al., 2015; Marjoram et al., 2006). The elevated TPJ activation in CHR was associated with higher symptoms in all domains, thus associating higher illness severity with greater neural activity. This association was not found in FEP, possibly suggesting different underlying mechanisms between groups. In combination with the behavioural data, showing that CHR adapted adequately to negative social feedback, the increased TPJ activity could indicate a cognitive mechanism by which increased mentalising helps to respond adequately to negative feedback, indicating more effort, or an inefficient use of the TPJ. The data do not point to compensating mechanisms, since they would suggest deficiencies or reduced processing in other parts of the brain. These were not found in the ROI analyses, nor in the additional whole brain analyses (see supplementary Table S1). The results show no evidence for reduced sensitivity to social reward in FEP and CHR, and suggest that altered mentalising might be associated with reduced baseline trust. However, due to the small sample size, this result must be interpreted with caution.

Clinical high-risk

Following the procedure of previous CHR investigations (Fusar-Poli et al., 2010; McGorry & Van Os, 2013; Phillips et al., 2009; Rietdijk et al., 2012; Shim et al., 2008; Thompson et al., 2012; Valmaggia et al., 2013; Van der Gaag et al., 2012; Wood et al., 2011), we included participants assessed with the CAARMS, and with a score below 55 on the SOFAS. Our sample was comparable to other samples in terms of comorbidities (Corcoran et al., 2011; Fusar-Poli et al., 2014; Ising et al., 2016; Modinos et al., 2014; Morrison et al., 2012; Woods et al., 2009). One year after testing, e.g. around two years after initial assessment, CHR participants were re-assessed with the CAARMS, to investigate their current status. One of the CHR had made the transition to psychosis. Of two CHR transition data were missing. In this aspect, our high-risk group differed from other high-risk groups. Variant transition rates have been reported in comparable samples with regard to assessment and age range (Broome et al., 2012; Broome et al., 2005; Demjaha, Valmaggia, Stahl, Byrne, & McGuire, 2010; Fusar-Poli et al., 2010; Nelson, Yuen, & Yung, 2011). Transition rates in similar referred samples are under 10% (Rietdijk et al., 2012; Yung et al., 2011). Patients already received treatment for their primary problems, including cognitive behavioural therapy (CBT) for their CHR status (psychotic symptoms). This has shown to be an adequate strategy to reduce symptoms, increasing their social functioning skills, to reduce the transition rates (by 46%), and to

increase chances for remission (Cannon et al., 2008; Ising et al., 2016; Jang et al., 2011; Niendam et al., 2007; Van Os & Murray, 2013).

In a recent discussion on CHR it has been argued that the presence of psychotic symptoms is possibly more important than transition in the assessment of CHR (Van Os & Reininghaus, 2016). Many patients in care for anxiety and depression report psychotic symptoms (Van Os & Linscott, 2012; Van Os & Reininghaus, 2016; Velthorst et al., 2009; Wigman et al., 2012; Woods et al., 2009), but do not transition to psychosis. The current sample fits previous descriptions, making it a representative sample of patients with psychotic symptoms and generally poor mental health. The addition of psychotic symptoms renders these patients at-risk for developing psychopathology, without the direct consequence of developing a psychotic disorder (Fusar-Poli et al., 2013; Yung et al., 2012). In many cases subclinical psychotic experiences are transitory (Van Os & Reininghaus, 2016). However, the presence of psychotic symptoms is associated with a poorer prognosis, showing that these patients are certainly in need of special care (McGorry & Van Os, 2013; Ruhrmann et al., 2010; Valmaggia et al., 2013; Van Os & Linscott, 2012; Van Os & Reininghaus, 2016).

Limitations and future directions

Several limitations should be considered. First, current results should be interpreted with caution, due to the small sample size, especially of the CHR sample. Our CHR results should therefore be considered as a first step in research on real-time social interaction in high-risk individuals, that warrants replication and extension in future research [see also (Broome et al., 2010; Fusar-Poli et al., 2011; Juckel et al., 2012)]. With a larger CHR sample, the number of converters will increase, allowing for an investigation of converters vs. non-converters. Additionally, a larger sample would provide the possibility to subdivide the CHR sample on the basis of different symptomatology (Fusar-Poli et al., 2014; Valmaggia et al., 2013), yielding more insight in the factors causing social problems, and explaining transition trajectories. Transition in these studies was not explained by comorbid anxiety and depression, but by the severity of CAARMS score, social dysfunction and increased negative symptoms. Direct comparison between at-risk subjects with and without progression into a diagnosis of schizophrenia could also elucidate whether greater activation of the mentalising network in CHR is serving as a compensatory mechanism, or could also be linked to transition to psychosis. Larger samples could have revealed group differences that were not apparent in this sample. Further, FEP symptom severity was rather mild, possibly due to responsiveness to antipsychotic treatment. Similar symptom severity has been found in stable and medicated patients (Möller et al., 2005). Including a broader range

of symptom severity would increase the validity of the sample. Methodologically, a limitation is that participants were not paid on performance. There is some evidence that real payment has a different effect on decisions and related brain activity than hypothetical payment (Hertwig & Ortmann, 2001; Johnson & Mislin, 2011; Vlaev, 2012), but other studies have found no differences (Locey et al., 2011; Madden et al., 2003). Plausibly, hypothetical payment may influence the strength, but not the direction of the effect (Derks, 2015). Furthermore, eight participants did not believe they were playing against a human counterpart. This might have influenced their behaviour. However, analyses without these participants did not change the results. Adequate increase and decrease of investments in response to cooperative and unfair feedback showed that overall the experimental manipulation of the counterpart was effective. Additionally, including an online mentalising task in the scanner could establish direct links with the current outcomes, pointing in the direction of differential mentalising processes.

Conclusion

Summarising, baseline trust is impaired in FEP and CHR, indicating that reduced baseline trust is associated with the risk for psychotic illness trait, or with poor mental health in general, rather than a consequence of psychotic disorder. CHR performed in between controls and FEP (Pukrop et al., 2006; Thompson et al., 2012), resembling most the patient group. In contrast to chronic patients (Fett, Gromann, et al., 2015), reward learning was not impaired in FEP and CHR [see also (Juckel et al., 2012)]. This suggests that with positive social feedback, the lack of initial trust in FEP and CHR can be restored, at least in the context of the trust game. The neural results pointed to globally intact neural mechanisms associated with trust, except for changes in brain areas associated with mentalising in CHR. However, these findings should be considered as preliminary, and more research in the field is needed to replicate and extend our findings.

Acknowledgements

We would like to thank Prof. Lieuwe de Haan and his team at the Amsterdam Medical Center (AMC), Prof. Mark Van der Gaag and Dr. Helga Ising at PsyQ, The Hague, Onno Ackema and the department of early psychosis of the AMC, and the VIP team for helping with recruitment of both patient groups, and Tinka Beemsterboer and colleagues at the Spinoza Centre for Neuroimaging, Roeterseiland Amsterdam for their help during scanning, and all participants for completing the testing session and providing us with valuable material.

Supplementary Material

Table S1

Whole Brain Group Differences per Condition of the Trust Game

Condition	Region	Cluster size	MNI coordinates X	Y	Z	Hemi-sphere	*Z*	*p*
Cooperative repayment, CHR > controls								
	Frontal lateral gyrus	29	27	13	33	R	4.16	.25
			32	16	25		3.62	
	TPJ	15	42	-42	33	R	4.20	.71
	Parahippocampus	19	-33	-22	-20	L	4.61	.54
	Hippocampus		-38	-15	-18	L	3.41	
	Posterior cingulate	11	10	-45	28	R	3.62	.87
Cooperative repayment, controls > FEP								
	Precuneus	11	-13	-42	5	L	4.02	.87
Unfair investment, CHR > controls								
	Frontal inferior gyrus	28	47	21	28	R	3.56	.40
	Frontal inferior gyrus		50	16	35	R	3.42	
	Temporal middle gyrus	19	52	-62	23	R	4.13	.15
	TPJ	13	50	-65	33	R	3.21	
	TPJ	25	57	-40	35	R	3.62	.48
			50	-45	38	R	3.15	
	Middle cingulate	23	2	-40	40	R	3.69	.54
	Posterior cingulate	29	-1	-47	23	L	3.85	.38
	Precuneus		7	-45	20	R	3.53	
Unfair investment, CHR > FEP								
	Postcentral	28	35	-22	10	R	4.06	.40
Unfair repayment, FEP > controls								
	Supra marginal gyrus	28	60	-20	40	R	4.45	.29
	Postcentral gyrus		50	-20	35	R	3.31	
	Superior temporal pole	12	50	16	-15	R	4.24	.84
	Occipital middle gyrus	13	30	-80	35	R	3.43	.80
	Occipital superior gyrus		25	-82	25	R	3.33	
Unfair repayment, FEP > CHR								
	Occipital middle gyrus	16	-26	-85	3	L	3.62	.68
	Occipital inferior gyrus	21	40	-85	0	R	3.60	.49
	Occipital middle gyrus		35	-87	8	R	3.36	
	Occipital middle gyrus	10	30	-80	35	R	3.52	.91
Unfair repayment, CHR > controls								
	Lingual gyrus	10	-1	-70	-5	L	3.80	.91

(continues on next page)

Supplementary Material

Table S1 (continued)

Whole Brain Group Differences per Condition of the Trust Game

Condition	Region	Cluster size	MNI coordinates X	Y	Z	Hemi-sphere	*Z*	*p*
Unfair repayment, CHR > FEP								
	Middle cingulate	14	20	-22	45	R	3.97	.76
Cooperative repayment > unfair repayment, CHR > controls								
	Frontal middle gyrus	23	35	26	23	R	3.61	.39
	Frontal inferior operculum		32	13	28	R	3.42	
Cooperative repayment > unfair repayment, CHR > FEP								
	Precuneus	39	-3	-42	58	L	3.98	.01
Cooperative repayment > unfair repayment, controls > FEP								
	Occipital middle gyrus	15	30	-77	10	R	3.58	.7
	Supplemental motor aria	14	2	-5	50	R	3.62	.74
	Supplemental motor aria		-3	-12	50	L	3.34	
Unfair investment > cooperative investment, CHR > controls								
	TPJ	11	50	-62	20	R	3.62	.9
	ACC	13	7	31	18	R	3.83	.85
	Middle cingulate	46	-3	-27	40	L	3.56	.14
	Middle cingulate		5	-22	30	R	3.51	
	Posterior cingulate	31	2	-47	20	R	4.10	.34
	Precuneus	16	2	-52	38	R	3.61	.75
	Precuneus		-3	-50	50	L	3.24	
Unfair investment > cooperative investment, CHR > FEP								
	Occipital middle gyrus	16	27	-87	23	R	3.68	.75

Note: FEP = first episode psychosis; CHR = clinical high-risk; MNI = Montreal Neurological Institute; TPJ = temporo-parietal junction; ACC = anterior cingulate cortex. All contrasts reported were performed at an uncorrected $p = .001$, with a cluster-size threshold of $k = 10$.