Title: Learning to trust: Social feedback normalizes trust behavior in first episode psychosis and clinical high-risk

Authors: Imke L.J. Lemmers-Jansen [a] , Anne-Kathrin J. Fett [a,b,c,d], Esther Hanssen [a], Dick J. Veltman [e,f], & Lydia Krabbendam [c,d]

Affiliations:

[a] Department of Educational and Family studies, Faculty of Behavioral and Movement Sciences, and Institute for Brain and Behavior Amsterdam, Vrije Universiteit Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

[b] Department of Psychology, City, University of London, Northampton Square, London EC1V 0HB, United Kingdom

[c] Department of Clinical, Neuro and Developmental Psychology, Faculty of Behavioral and Movement Sciences, and Institute for Brain and Behavior Amsterdam, Vrije Universiteit Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

[d] Department of Psychosis Studies, King's College London, Institute of Psychiatry, Psychology and Neuroscience, 16 De Crespigny Park, London SE5 8AF, Great Britain

[e] Department of Psychiatry, VU Medical Center, Van der Boechorststraat 7 , 1081 BT Amsterdam, The Netherlands

[f] Neuroscience Campus Amsterdam, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

Email addresses: imke.jansen@vu.nl, Anne-Kathrin.Fett@city.ac.uk, esther.hanssen@vu.nl, dj.veltman@vumc.nl, lydia.krabbendam@vu.nl

Manuscript consist of:

Words: 6010

Tables: 2

Graphs: 3

Abbreviations

FEP = First episode psychosis

CHR = Clinical high-risk

CAARMS = Comprehensive Assessment of At-Risk Mental States

PANSS = Positive and Negative Syndromes Scale

WAIS = Wechsler Adult Intelligence Scale

ROI = Region of interest

TPJ = Temporo-parietal junction

mPFC = medial prefrontal cortex

Abstract

Background: Psychosis is characterized 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 mentalizing areas. We investigated whether first episode psychosis patients (FEP) and CHR show similar abnormalities in the neural and behavioral 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 mentalizing 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

1 Introduction

Psychosis is characterized 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 et al, 2016a; Velthorst et al, 2016b; 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 et al, 2016a; Velthorst et al, 2016b). 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 et al, 2011; Horat et al, 2017), including impaired mentalizing (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 off-line 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 behavior in an on-line 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 pay-off 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 mentalizing (Declerck et al, 2013; Frith & Frith, 2006; Gallagher & Frith, 2003) and reward processing (Fehr & Camerer, 2007; King-Casas et al, 2005; Rilling & Sanfey, 2011). Mentalizing appears to be important during both the investment and repayment phase, where estimations of the other's behavior 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 et al, 2015; Fett et al, 2016; Fett et al, 2012; 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 behavioral 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 behavior in these patient groups. Based on the existing trust game literature, we hypothesized that similar to relatives (Fett et al, 2012) and (chronic) patients (Fett 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 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 hypothesized to find 3) attenuated activation in brain areas associated with mentalizing 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 hypothesized to find 4) positive

symptoms related positively to brain activation in mentalizing and negatively to reward areas in both patient groups. In addition, associations of brain activations with negative symptoms were investigated.

2 Methods

2.1 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 (Association & 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 subthreshold 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 behavioral 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.

2.2 Measures

2.2.1 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 behavior of their counterpart was administered, to check if participants believed that they were playing a real person.

2.2.2 Positive and Negative Syndrome Scale (PANSS; Kay et al, 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 interrater reliability was calculated (r = .85). All PANSS data were rated by the same two researchers.

2.2.3 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.

2.2.4 Trust Manipulation Check Questionnaire

The trust game was followed by a questionnaire to investigate participants' opinions on the behavior 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 behavior 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.

2.3 Procedure

This research was approved 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 2.2.4). Participants received an image of their brain, 25€ for participation and reimbursement of their travel costs.

2.4 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).

2.5 Data Analysis

2.5.1 Behavioral data

Demographic and behavioral data were analyzed using Stata 13 (StataCorp, 2013) with regression analyses and chi-square tests. We analyzed 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.

2.5.2 Imaging data

Imaging data were analyzed using Statistical Parametric Mapping 8 (SPM, 2009). Functional images for each participant were preprocessed as follows: realign and unwarp, coregistration 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 (version0.43; 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 (http://www.quantitativeskills.com/sisa/calculations/bonfer.htm), resulting in adjusted *p*-values (see Table 2) (Li et al, 2014; Woudstra et al, 2013). Additional whole-brain analyses were performed, to investigate activation outside the predefined ROIs.

2.5.3 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 behavior and brain activation, additional post-hoc analyses within patient groups were performed.

3 Results

3.1 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).

3.2 Behavioral 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).

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.

3.3 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.

3.4 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 (β = .37, *p* = .03). There was a trend towards significance indicating that FEP had higher negative symptoms than CHR ($p < .07$).

3.4.1 Associations between behavioral 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).

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 visualize 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.

3.4.2 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.

4 Discussion

This study investigated the behavioral 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. Behaviorally, 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.

4.1 Behavioral mechanisms of trust

Importantly, and in line with previous research both FEP and CHR showed reduced baseline trust toward unknown others (Fett et al, 2015; Fett et al, 2016; Fett et al, 2012; 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, 2016; Fett et al, 2012) 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 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

388 the trials. The slightly reduced response to negative feedback in FEP, resembling results of Fett et al
389 (2016), might be explained by the differences in first investment (FEP starting significantly lower).

390 4.1.1 Symptoms

391 Symptom severity on average was similar in the two patient groups. FEP showed substantial
392 variability in symptom severity, which reflects the fact that we included both hospitalized patients
393 and ambulant patients who were in a rehabilitation trajectory. CHR experienced more paranoia
394 than FEP, and FEP tended to have more negative symptoms than CHR. These differences might be
395 explained by medication effects: 64% of the FEP were on atypical antipsychotic medication,
396 probably dampening positive symptoms, whereas 47% of the CHR was on other psychotropic
397 medication (see Table 1). First episode patients with highest negative symptoms, as opposed to
398 milder symptoms, showed almost no adjustment of trust in response to positive feedback. Intact
399 feedback learning mechanisms in FEP were associated with milder negative symptom severity. The
400 association between negative symptoms and problems in social functioning and responding to
401 feedback has been well established in psychosis (Addington & Addington, 2005; Campellone et al,
402 2016; Milev et al, 2005; Strauss et al, 2013; Voges & Addington, 2005; Waltz et al, 2011), possibly
403 reflecting a lack of (social) motivation or depression. In the unfair condition positive symptoms
404 were associated with less decrease in trust in both FEP and CHR. This suggests that positive
405 symptoms interfere with learning from negative social feedback, contradicting earlier findings that
406 found no associations of positive symptoms with learning to trust (Fett et al, 2016).

407 4.2 Neural mechanisms of trust

408 On the neural level CHR activated the TPJ significantly more than the other two groups. The TPJ
409 forms part of the mentalizing system (Fletcher et al, 1995; Frith & Frith, 2006; Van Overwalle,
410 2009), and was previously found to be activated in the trust game (King-Casas et al, 2005; 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 mentalizing system, this could suggest that unfair interactions elicit increased mentalizing 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 mentalizing and emotion processing areas despite similar behavioral 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 behavioral data, showing that CHR adapted adequately to negative social feedback, the increased TPJ activity could indicate a cognitive mechanism by which increased mentalizing 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 mentalizing might be associated with reduced baseline trust. However, due to the small sample size, this result must be interpreted with caution.

4.3 Clinical high-risk

Following the procedure of previous CHR investigations (Fusar-Poli et al, 2010a; 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 et al, 2010; Fusar-Poli et al, 2010b; Nelson et al, 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 behavioral 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).

4.4 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 mentalizing 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 behavior. 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 mentalizing task in the scanner could establish direct links with the current outcomes, pointing in the direction of differential mentalizing processes.

5 Conclusion

Summarizing, 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 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 mentalizing 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.

Funding: This work was supported by funding of the Hersenstichting Nederland [KS2011(1)-75], a VIDI and VICI grant from the Netherlands Organization for Scientific Research (NWO) [452-07-007,453-11-005]; and a ERC Consolidator grant (648082 SCANS) to Prof. Lydia Krabbendam. Anne-Kathrin Fett was supported by a Netherlands Organization for Scientific Research (NWO) VENI grant (451-13-035) and a NARSAD Young Investigator Grant from the Brain & Behaviour research Foundation (24138).

Addington, J. & Addington, D. (2005) Patterns of premorbid functioning in first episode psychosis: relationship to 2-year outcome. *Acta Psychiatrica Scandinavica*, 112(1), 40-46.
Association, A. P. & Association, A. P. (2000) DSM-IV-TR: Diagnostic and statistical manual of mental disorders, text revision. *Washington, DC: American Psychiatric Association*, 75, 78-85.
Ballon, J. S., Kaur, T., Marks, I. I. & Cadenhead, K. S. (2007) Social functioning in young people at risk for schizophrenia. *Psychiatry research*, 151(1), 29-35.
Berg, J., Dickhaut, J. & McCabe, K. (1995) Trust, reciprocity, and social history. *Games and economic behavior*, 10(1), 122-142.
Bora, E. & Pantelis, C. (2013) Theory of mind impairments in first-episode psychosis, individuals at ultra-high risk for psychosis and in first-degree relatives of schizophrenia: systematic review and meta-analysis. *Schizophrenia research*, 144(1), 31-36.
Broome, M. R., Day, F., Valli, I., Valmaggia, L., Johns, L., Howes, O., Garety, P. & McGuire, P. (2012) Delusional ideation, manic symptomatology and working memory in a cohort at clinical high-risk for psychosis: a longitudinal study. *European Psychiatry*, 27(4), 258-263.
Broome, M. R., Fusar-Poli, P., Matthiasson, P., Woolley, J. B., Valmaggia, L., Johns, L. C., Tabraham, P., Bramon, E., Williams, S. C. & Brammer, M. J. (2010) Neural correlates of visuospatial working memory in the 'at-risk mental state'. *Psychological medicine*, 40(12), 1987-1999.
Broome, M. R., Woolley, J. B., Johns, L. C., Valmaggia, L. R., Tabraham, P., Gafoor, R., Bramon, E. & McGuire, P. K. (2005) Outreach and support in south London (OASIS): implementation of a clinical service for prodromal psychosis and the at risk mental state. *European Psychiatry*, 20(5), 372-378.
Brüne, M., Özgürdal, S., Ansorge, N., von Reventlow, H. G., Peters, S., Nicolas, V., Tegenthoff, M., Juckel, G. & Lissek, S. (2011) An fMRI study of "theory of mind" in at-risk states of psychosis: comparison with manifest schizophrenia and healthy controls. *Neuroimage*, 55(1), 329-337.
Campellone, T. R., Fisher, A. J. & Kring, A. M. (2016) Using social outcomes to inform decision-making in schizophrenia: Relationships with symptoms and functioning. *Journal of abnormal psychology*, 125(2), 310.
Cannon, T. D., Cadenhead, K., Cornblatt, B., Woods, S. W., Addington, J., Walker, E., Seidman, L. J., Perkins, D., Tsuang, M. & McGlashan, T. (2008) Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Archives of general psychiatry*, 65(1), 28-37.
Corcoran, C., Kimhy, D., Parrilla-Escobar, M., Cressman, V., Stanford, A., Thompson, J., David, S. B., Crumbley, A., Schobel, S. & Moore, H. (2011) The relationship of social function to depressive and negative symptoms in individuals at clinical high risk for psychosis. *Psychological medicine*, 41(02), 251-261.
Cornblatt, B. A., Auther, A. M., Niendam, T., Smith, C. W., Zinberg, J., Bearden, C. E. & Cannon, T. D. (2007) Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophrenia bulletin*, 33(3), 688-702.
Couture, S. M., Penn, D. L. & Roberts, D. L. (2006) The functional significance of social cognition in schizophrenia: a review. *Schizophrenia bulletin*, 32(suppl 1), S44-S63.
Csukly, G., Polgár, P., Tombor, L., Réthelyi, J. & Kéri, S. (2011) Are patients with schizophrenia rational maximizers? Evidence from an ultimatum game study. *Psychiatry research*, 187(1), 11-17.
Declerck, C. H., Boone, C. & Emonds, G. (2013) When do people cooperate? The neuroeconomics of prosocial decision making. *Brain and Cognition*, 81(1), 95-117.
Demjaha, A., Valmaggia, L., Stahl, D., Byrne, M. & McGuire, P. (2010) Disorganization/cognitive and negative symptom dimensions in the at-risk mental state predict subsequent transition to psychosis. *Schizophrenia bulletin*, 38(2), 351-359.
Derks, J. (2015) Adolescent Social Cognition (doctoral dissertation).

Derntl, B., Michel, T. M., Prempeh, P., Backes, V., Finkelmeyer, A., Schneider, F. & Habel, U. (2015) Empathy in individuals clinically at risk for psychosis: brain and behaviour. *The British Journal of Psychiatry*, bjp. bp. 114.159004.

Fehr, E. & Camerer, C. F. (2007) Social neuroeconomics: the neural circuitry of social preferences. *Trends in cognitive sciences*, 11(10), 419-427.

Fett, A.-K., Gromann, P., Shergill, S. & Krabbendam, L. (2015) Trust vs. Paranoia: The Dynamics of Social Interaction in Early and Chronic Psychosis, *SCHIZOPHRENIA BULLETIN*. OXFORD UNIV PRESS GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND.

Fett, A.-K., Shergill, S., Korver-Nieberg, N., Yakub, F., Gromann, P. M. & Krabbendam, L. (2016) Learning to trust: trust and attachment in early psychosis. *Psychological medicine*, 46(7), 1437.

Fett, A.-K., Shergill, S. S., Joyce, D. W., Riedl, A., Strobel, M., Gromann, P. M. & Krabbendam, L. (2012) To trust or not to trust: the dynamics of social interaction in psychosis. *Brain*, 135(3), 976-984.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. & Frith, C. D. (1995) Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition*, 57(2), 109-128.

Frith, C. D. & Frith, U. (2006) The neural basis of mentalizing. *Neuron*, 50(4), 531-534.

Fusar-Poli, P., Borgwardt, S., Bechdolf, A., Addington, J., Riecher-Rössler, A., Schultze-Lutter, F., Keshavan, M., Wood, S., Ruhrmann, S. & Seidman, L. J. (2013) The Psychosis High-Risk State: A Comprehensive State-of-the-Art Review. *JAMA psychiatry*, 70(1), 107.

Fusar-Poli, P., Broome, M. R., Matthiasson, P., Woolley, J. B., Johns, L., Tabraham, P., Bramon, E., Valmaggia, L., Williams, S. & McGuire, P. (2010a) Spatial working memory in individuals at high risk for psychosis: longitudinal fMRI study. *Schizophrenia Research*, 123(1), 45-52.

Fusar-Poli, P., Broome, M. R., Woolley, J. B., Johns, L. C., Tabraham, P., Bramon, E., Valmaggia, L., Williams, S. & McGuire, P. (2011) Altered brain function directly related to structural abnormalities in people at ultra high risk of psychosis: longitudinal VBM-fMRI study. *Journal of psychiatric research*, 45(2), 190-198.

Fusar-Poli, P., Byrne, M., Valmaggia, L., Day, F., Tabraham, P., Johns, L., McGuire, P. & Team, O. (2010b) Social dysfunction predicts two years clinical outcome in people at ultra high risk for psychosis. *Journal of psychiatric research*, 44(5), 294-301.

Fusar-Poli, P., Nelson, B., Valmaggia, L., Yung, A. R. & McGuire, P. K. (2014) Comorbid depressive and anxiety disorders in 509 individuals with an at-risk mental state: impact on psychopathology and transition to psychosis. *Schizophrenia bulletin*, 40(1), 120-131.

Gallagher, H. L. & Frith, C. D. (2003) Functional imaging of 'theory of mind'. *Trends in cognitive sciences*, 7(2), 77-83.

Goldman, H. H., Skodol, A. E. & Lave, T. R. (1992) Revising axis V for DSM-IV: a review of measures of social functioning. *Am J Psychiatry*, 149(9), 1148-1156.

Green, M. F., Horan, W. P. & Lee, J. (2015) Social cognition in schizophrenia. *Nature Reviews Neuroscience*, 16(10), 620.

Green, M. F. & Leitman, D. I. (2008) Social cognition in schizophrenia. *Schizophrenia Bulletin*, 34(4), 670-672.

Gromann, P. M., Heslenfeld, D. J., Fett, A.-K., Joyce, D. W., Shergill, S. S. & Krabbendam, L. (2013) Trust versus paranoia: abnormal response to social reward in psychotic illness. *Brain*, awt076.

Gromann, P. M., Shergill, S., de Haan, L., Meewis, D., Fett, A.-K., Korver-Nieberg, N. & Krabbendam, L. (2014) Reduced brain reward response during cooperation in first-degree relatives of patients with psychosis: an fMRI study. *Psychological medicine*, 44(16), 3445-3454.

Hertwig, R. & Ortmann, A. (2001) Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24(03), 383-403.

615 Horat, S. K., Favre, G., Prévot, A., Ventura, J., Herrmann, F. R., Gothuey, I., Merlo, M. C. & Missonnier, P.
616 (2017) Impaired social cognition in schizophrenia during the Ultimatum Game: An EEG study.
617 *Schizophrenia Research*.
618 Ising, H. K., Kraan, T. C., Rietdijk, J., Dragt, S., Klaassen, R. M., Boonstra, N., Nieman, D. H., Willebrands-
619 Mendrik, M., van den Berg, D. P. & Linszen, D. H. (2016) Four-year follow-up of cognitive behavioral
620 therapy in persons at ultra-high risk for developing psychosis: the Dutch early detection intervention
621 evaluation (EDIE-NL) trial. *Schizophrenia bulletin*, 42(5), 1243-1252.
622 Jang, J. H., Shin, N. Y., Shim, G., Park, H. Y., Kim, E., Jang, G.-E., Kwon, S. J., Hur, J.-W., An, S. K. & Kwon, J.
623 S. (2011) Longitudinal patterns of social functioning and conversion to psychosis in subjects at ultra-high
624 risk. *Australian and New Zealand Journal of Psychiatry*, 45(9), 763-770.
625 Johnson, N. D. & Mislin, A. A. (2011) Trust games: A meta-analysis. *Journal of Economic Psychology*,
626 32(5), 865-889.
627 Juckel, G., Friedel, E., Koslowski, M., Witthaus, H., Özgürdal, S., Gudlowski, Y., Knutson, B., Wrase, J.,
628 Brüne, M. & Heinz, A. (2012) Ventral striatal activation during reward processing in subjects with ultra-
629 high risk for schizophrenia. *Neuropsychobiology*, 66(1), 50-56.
630 Kay, S., Fiszbein, A. & Opler, L. (1987) The positive and negative syndrome scale (PANSS) for
631 schizophrenia. *Schizophrenia bulletin*, 13(2), 261.
632 Kelleher, I., Keeley, H., Corcoran, P., Lynch, F., Fitzpatrick, C., Devlin, N., Molloy, C., Roddy, S., Clarke, M.
633 C. & Harley, M. (2012) Clinicopathological significance of psychotic experiences in non-psychotic young
634 people: evidence from four population-based studies. *The British Journal of Psychiatry*, 201(1), 26-32.
635 King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R. & Montague, P. R. (2005) Getting to
636 know you: reputation and trust in a two-person economic exchange. *Science*, 308(5718), 78-83.
637 Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A. & Grafman, J.
638 (2007) Neural correlates of trust. *Proceedings of the National Academy of Sciences*, 104(50), 20084-
639 20089.
640 Lavoie, M.-A., Lacroix, J. B., Godmaire-Duhaime, F., Jackson, P. L. & Achim, A. M. (2013) Social cognition
641 in first-degree relatives of people with schizophrenia: a meta-analysis. *Psychiatry Research*, 209(2), 129-
642 135.
643 Lemmers-Jansen, I. L. J., Krabbendam, L., Veltman, D. J. & Fett, A.-K. J. (2017) Boys vs. girls: Gender
644 differences in the neural development of trust and reciprocity depend on social context. *Developmental*
645 *Cognitive Neuroscience*.
646 Li, W., Tol, M. J., Li, M., Miao, W., Jiao, Y., Heinze, H. J., Bogerts, B., He, H. & Walter, M. (2014) Regional
647 specificity of sex effects on subcortical volumes across the lifespan in healthy aging. *Human brain*
648 *mapping*, 35(1), 238-247.
649 Locey, M. L., Jones, B. A. & Rachlin, H. (2011) Real and hypothetical rewards. *Judgment and decision*
650 *making*, 6(6), 552.
651 Madden, G. J., Begotka, A. M., Raiff, B. R. & Kastern, L. L. (2003) Delay discounting of real and
652 hypothetical rewards. *Experimental and clinical psychopharmacology*, 11(2), 139.
653 Marjoram, D., Job, D. E., Whalley, H. C., Gountouna, V.-E., McIntosh, A. M., Simonotto, E., Cunningham-
654 Owens, D., Johnstone, E. C. & Lawrie, S. (2006) A visual joke fMRI investigation into Theory of Mind and
655 enhanced risk of schizophrenia. *Neuroimage*, 31(4), 1850-1858.
656 McCleery, A., Horan, W. P. & Green, M. F. (2014) Social Cognition During the Early Phase of
657 Schizophrenia. *Social Cognition and Metacognition in Schizophrenia*, 49.
658 McGorry, P. & van Os, J. (2013) Redeeming diagnosis in psychiatry: timing versus specificity. *The Lancet*,
659 381(9863), 343-345.
660 Milev, P., Ho, B.-C., Arndt, S. & Andreasen, N. C. (2005) Predictive values of neurocognition and negative
661 symptoms on functional outcome in schizophrenia: a longitudinal first-episode study with 7-year follow-
662 up. *American Journal of Psychiatry*, 162(3), 495-506.

663 Modinos, G., Allen, P., Frascarelli, M., Tognin, S., Valmaggia, L., Xenaki, L., Keedwell, P., Broome, M.,
664 Valli, I. & Woolley, J. (2014) Are we really mapping psychosis risk? Neuroanatomical signature of
665 affective disorders in subjects at ultra high risk. *Psychological medicine*, 44(16), 3491-3501.
666 Möller, H.-J., Llorca, P.-M., Sacchetti, E., Martin, S. D., Medori, R., Parellada, E. & Group, S. S. (2005)
667 Efficacy and safety of direct transition to risperidone long-acting injectable in patients treated with
668 various antipsychotic therapies. *International clinical psychopharmacology*, 20(3), 121-130.
669 Morosini, P., Magliano, L., Brambilla, L., Ugolini, S. & Pioli, R. (2000) Development, reliability and
670 acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale
671 (SOFAS) to assess routine social funtioning. *Acta Psychiatrica Scandinavica*, 101(4), 323-329.
672 Morrison, A. P., French, P., Stewart, S. L., Birchwood, M., Fowler, D., Gumley, A. I., Jones, P. B., Bentall, R.
673 P., Lewis, S. W. & Murray, G. K. (2012) Early detection and intervention evaluation for people at risk of
674 psychosis: multisite randomised controlled trial. *Bmj*, 344, e2233.
675 Nelson, B., Yuen, K. & Yung, A. (2011) Ultra high risk (UHR) for psychosis criteria: are there different
676 levels of risk for transition to psychosis? *Schizophrenia research*, 125(1), 62-68.
677 Niendam, T. A., Bearden, C. E., Zinberg, J., Johnson, J. K., O'Brien, M. & Cannon, T. D. (2007) The course
678 of neurocognition and social functioning in individuals at ultra high risk for psychosis. *Schizophrenia*
679 *bulletin*, 33(3), 772-781.
680 Phillips, L. J., Nelson, B., Yuen, H. P., Francey, S. M., Simmons, M., Stanford, C., Ross, M., Kelly, D., Baker,
681 K. & Conus, P. (2009) Randomized controlled trial of interventions for young people at ultra-high risk of
682 psychosis: study design and baseline characteristics. *Australian & New Zealand Journal of Psychiatry*,
683 43(9), 818-829.
684 Pukrop, R., Schultze-Lutter, F., Ruhrmann, S., Brockhaus-Dumke, A., Tendolkar, I., Bechdolf, A.,
685 Matuschek, E. & Klosterkötter, J. (2006) Neurocognitive functioning in subjects at risk for a first episode
686 of psychosis compared with first-and multiple-episode schizophrenia. *Journal of Clinical and*
687 *Experimental Neuropsychology*, 28(8), 1388-1407.
688 Rietdijk, J., Klaassen, R., Ising, H., Dragt, S., Nieman, D., Van de Kamp, J., Cuijpers, P., Linszen, D. & Van
689 der Gaag, M. (2012) Detection of people at risk of developing a first psychosis: comparison of two
690 recruitment strategies. *Acta psychiatrica scandinavica*, 126(1), 21-30.
691 Rilling, J. K. & Sanfey, A. G. (2011) The neuroscience of social decision-making. *Annual review of*
692 *psychology*, 62, 23-48.
693 Ruhrmann, S., Schultze-Lutter, F. & Klosterkötter, J. (2010) Probably at-risk, but certainly ill—advocating
694 the introduction of a psychosis spectrum disorder in DSM-V. *Schizophrenia research*, 120(1), 23-37.
695 Saxe, R. & Kanwisher, N. (2003) People thinking about thinking people: the role of the temporo-parietal
696 junction in "theory of mind". *Neuroimage*, 19(4), 1835-1842.
697 Shim, G., Kang, D.-H., Sun Chung, Y., Young Yoo, S., Young Shin, N. & Soo Kwon, J. (2008) Social
698 functioning deficits in young people at risk for schizophrenia. *Australian and New Zealand Journal of*
699 *Psychiatry*, 42(8), 678-685.
700 Smieskova, R., Marmy, J., Schmidt, A., Bendfeldt, K., Riecher-Rossler, A., Walter, M., E Lang, U. &
701 Borgwardt, S. (2013) Do subjects at clinical high risk for psychosis differ from those with a genetic high
702 risk?-A systematic review of structural and functional brain abnormalities. *Current medicinal chemistry*,
703 20(3), 467-481.
704 SPM (2009) *Statistical Parametric Mapping*. London, UK: Wellcome Trust Centre for Neuroimaging.
705 Available online: www.fil.ion.ucl.ac.uk/spm [Accessed.
706 StataCorp (2013) *Stata Statistical Software* Release 13. TX: StataCorp LP.
707 Strauss, G. P., Waltz, J. A. & Gold, J. M. (2013) A review of reward processing and motivational
708 impairment in schizophrenia. *Schizophrenia bulletin*, 40(Suppl_2), S107-S116.

Thompson, A., Papas, A., Bartholomeusz, C., Allott, K., Amminger, G. P., Nelson, B., Wood, S. & Yung, A. (2012) Social cognition in clinical "at risk" for psychosis and first episode psychosis populations. *Schizophrenia research*, 141(2), 204-209.
Valmaggia, L., Stahl, D., Yung, A., Nelson, B., Fusar-Poli, P., McGorry, P. & McGuire, P. (2013) Negative psychotic symptoms and impaired role functioning predict transition outcomes in the at-risk mental state: a latent class cluster analysis study. *Psychological medicine*, 43(11), 2311-2325.
van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. & Crone, E. A. (2011) Changing brains, changing perspectives the neurocognitive development of reciprocity. *Psychological Science*, 22(1), 60-70.
van der Gaag, M., Nieman, D. H., Rietdijk, J., Dragt, S., Ising, H. K., Klaassen, R. M., Koeter, M., Cuijpers, P., Wunderink, L. & Linszen, D. H. (2012) Cognitive behavioral therapy for subjects at ultrahigh risk for developing psychosis: a randomized controlled clinical trial. *Schizophrenia Bulletin*, 38(6), 1180-1188.
van Os, J. & Linscott, R. J. (2012) Introduction: the extended psychosis phenotype—relationship with schizophrenia and with ultrahigh risk status for psychosis. *Schizophrenia Bulletin*, 38(2), 227-230.
van Os, J. & Murray, R. M. (2013) Can we identify and treat "schizophrenia light" to prevent true psychotic illness. *BMJ*, 346(jan18 1), f304-f304.
van Os, J. & Reininghaus, U. (2016) Psychosis as a transdiagnostic and extended phenotype in the general population. *World Psychiatry*, 15(2), 118-124.
Van Overwalle, F. (2009) Social cognition and the brain: a meta-analysis. *Human brain mapping*, 30(3), 829-858.
Velthorst, E., Fett, A.-K. J., Reichenberg, A., Perlman, G., van Os, J., Bromet, E. J. & Kotov, R. (2016a) The 20-Year Longitudinal Trajectories of Social Functioning in Individuals With Psychotic Disorders. *American Journal of Psychiatry*, appi. ajp. 2016.15111419.
Velthorst, E., Nieman, D. H., Becker, H. E., van de Fliert, R., Dingemans, P. M., Klaassen, R., de Haan, L., van Amelsvoort, T. & Linszen, D. H. (2009) Baseline differences in clinical symptomatology between ultra high risk subjects with and without a transition to psychosis. *Schizophrenia Research*, 109(1), 60-65.
Velthorst, E., Reichenberg, A., Kapara, O., Goldberg, S., Fromer, M., Fruchter, E., Ginat, K., de Haan, L., Davidson, M. & Weiser, M. (2016b) Developmental Trajectories of Impaired Community Functioning in Schizophrenia. *JAMA psychiatry*, 73(1), 48-55.
Vlaev, I. (2012) How different are real and hypothetical decisions? Overestimation, contrast and assimilation in social interaction. *Journal of Economic Psychology*, 33(5), 963-972.
Voges, M. & Addington, J. (2005) The association between social anxiety and social functioning in first episode psychosis. *Schizophrenia research*, 76(2), 287-292.
Waltz, J. A., Frank, M. J., Wiecki, T. V. & Gold, J. M. (2011) Altered probabilistic learning and response biases in schizophrenia: behavioral evidence and neurocomputational modeling. *Neuropsychology*, 25(1), 86.
Wechsler, D. (1997) WAIS-III Dutch Translation. Lisse: Swets & Zeitlinger.
Wigman, J. T., van Nierop, M., Vollebergh, W. A., Lieb, R., Beesdo-Baum, K., Wittchen, H.-U. & van Os, J. (2012) Evidence that psychotic symptoms are prevalent in disorders of anxiety and depression, impacting on illness onset, risk, and severity—implications for diagnosis and ultra–high risk research. *Schizophrenia Bulletin*, sbr196.
Wood, S. J., Yung, A. R., McGorry, P. D. & Pantelis, C. (2011) Neuroimaging and treatment evidence for clinical staging in psychotic disorders: from the at-risk mental state to chronic schizophrenia. *Biological psychiatry*, 70(7), 619-625.
Woods, S. W., Addington, J., Cadenhead, K. S., Cannon, T. D., Cornblatt, B. A., Heinssen, R., Perkins, D. O., Seidman, L. J., Tsuang, M. T. & Walker, E. F. (2009) Validity of the prodromal risk syndrome for first psychosis: findings from the North American Prodrome Longitudinal Study. *Schizophrenia bulletin*, 35(5), 894-908.

Woudstra, S., van Tol, M.-J., Bochdanovits, Z., van der Wee, N. J., Zitman, F. G., van Buchem, M. A., Opmeer, E. M., Aleman, A., Penninx, B. W. & Veltman, D. J. (2013) Modulatory effects of the piccolo genotype on emotional memory in health and depression. *Plos one*, 8(4), e61494.

Yung, A. R., Phillips, L. J., Nelson, B., Francey, S. M., Yuen, H. P., Simmons, M. B., Ross, M. L., Kelly, D., Baker, K. & Paul Amminger, G. (2011) Randomized controlled trial of interventions for young people at ultra high risk for psychosis: 6-month analysis. *Journal of Clinical Psychiatry*, 72(4), 430.

Yung, A. R., Phillips, L. J., Yuen, H. P., Francey, S. M., McFarlane, C. A., Hallgren, M. & McGorry, P. D. (2003) Psychosis prediction: 12-month follow up of a high-risk ("prodromal") group. *Schizophrenia research*, 60(1), 21-32.

Yung, A. R., Woods, S. W., Ruhrmann, S., Addington, J., Schultze-Lutter, F., Cornblatt, B. A., Amminger, G. P., Bechdolf, A., Birchwood, M. & Borgwardt, S. (2012) Whither the attenuated psychosis syndrome? *Schizophrenia bulletin*, 38(6), 1130-1134.

Yung, A. R., Yung, A. R., Pan Yuen, H., Mcgorry, P. D., Phillips, L. J., Kelly, D., Dell'olio, M., Francey, S. M., Cosgrave, E. M. & Killackey, E. (2005) Mapping the onset of psychosis: the comprehensive assessment of at-risk mental states. *Australian and New Zealand Journal of Psychiatry*, 39(11-12), 964-971.

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 $p < .05$

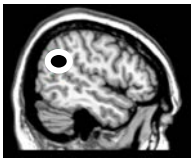
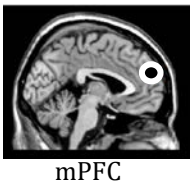
*^ = FEP > CHR at $p < .07$

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

800 depression item forms part of the general subscale (G6). For analyses, these item were investigated
801 separately.
802

Table 2

Region of Interest (ROI) Analyses Outcome

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

*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