

To trust or not to trust: the dynamics of social interaction in psychosis

Anne-Kathrin J. Fett,^{1,2,3} Sukhi S. Shergill,² Dan W. Joyce,² Arno Riedl,^{4,5,6,7} Martin Strobel,⁴ Paula M. Gromann^{1,2} and Lydia Krabbendam^{1,2}

- 1 Department of Educational Neuroscience, Faculty of Psychology and Education, VU University Amsterdam, Van der Boechorststraat 1, Amsterdam 1081 BT, The Netherlands
- 2 Department of Psychosis Studies, King's College London, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK
- 3 Department of Psychiatry and Neuropsychology, School of Mental Health and Neuroscience, Maastricht University, P.O. Box 616 (Location VIJ), 6200 MD Maastricht, The Netherlands
- 4 Department of Economics, School of Business and Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands
- 5 Munich Society for the Promotion of Economic Research - CESifo GmbH, Poschingerstr. 5, 81679 Munich, Germany
- 6 Institute for the Study of Labor (IZA), Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany
- 7 Network for Studies on Pensions, Aging and Retirement (NetSPAR), P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Correspondence to: Dr L. Krabbendam,
Faculty of Psychology and Education; Department of Educational Neuroscience,
VU University Amsterdam; Van der Boechorststraat 1,
1081 BT Amsterdam,
The Netherlands
E-mail: lydia.krabbendam@vu.nl

Psychotic illness is a disorder of social interaction unique to humans. However, up to now research has failed to pin down the exact determinants of the complex and interactive processes associated with the development of trust and reciprocity in psychosis. Utilizing a novel multi-round version of an interactive trust game experiment, we show that patients with psychosis and healthy relatives with a heightened risk for the illness exhibit lower baseline levels of trust compared with healthy controls. This effect partly overlapped with a reduced general intelligence. Furthermore, patients were unable to modify their trusting behaviour neither in response to information about the general trustworthiness of their interaction partner, nor in response to their partners' specific direct behavioural feedback. Relatives, in contrast, modified their trusting behaviour towards similar levels as healthy subjects in response to both. The results show that behavioural flexibility in response to socially relevant information is a critical determinant of success in the instantiation and maintenance of social relationships. A lack thereof may drive social dysfunction and the progression from subclinical symptoms to a full-blown psychosis. This offers a testable mechanistic hypothesis for progression from prodrome to psychotic illness, and may provide a therapeutic avenue to grapple the psychotic symptoms of social dysfunction.

Keywords: schizophrenia; social cognition; trust; neuroeconomics; family study

Abbreviations: PANSS = Positive and Negative Syndrome Scale

Introduction

Psychosis is marked by severe impairments in social functioning. This is evident in its core symptoms, such as social withdrawal or

paranoid delusions, which are characterized by hostility and fundamental impairment of trust in others. Trust and reciprocity are essential components of mutually beneficial human interactions. They critically influence competitive and cooperative behaviour

and are vital for an individual's successful functioning in a complex social environment. Despite its importance, research has failed to incorporate the interactive character of the social cognitive impairment in psychosis into its experimental paradigms. Most studies employed 'off-line' measures, such as questions about stories or cartoons with mental characters (Adolphs, 2006). While these approaches have yielded valuable insights into deficits of emotion recognition and mentalizing, they cannot capture the dynamic development of trust and trustworthiness in actual social interactions. Trust previously appeared too complex to probe experimentally. Recently, the development of neuroeconomics and the adaptation of experimental economics and game theoretical approaches (Camerer, 2003) have shown that complex social interactions can be operationalized in economic exchange games (Rilling *et al.*, 2002; King-Casas *et al.*, 2005).

Such games usually depict situations where mutually beneficial trust and trustworthiness between two people can be undermined by a conflict of interest. The classical trust game introduced by Berg *et al.* (1995) has been shown to be a good direct behavioural measure of trust and trustworthiness (Glaeser *et al.*, 2000; Karlan, 2005; King-Casas *et al.*, 2005, 2008; Kosfeld *et al.*, 2005; Fehr, 2009; Houser *et al.*, 2010). It involves the interaction of two anonymous persons, an investor and a trustee and comprises the investment and reciprocation of money between the two. Various outcomes are possible, depending on whether the players are trusting and trustworthy, respectively. Mutually beneficial outcomes occur when the investor trusts and the trustee behaves trustworthily. However, trusting is risky because the trustee can achieve the highest earnings when reciprocating untrustworthy (i.e. keeping most or all of the received money). For healthy individuals, it has been shown that investor trust is strongly reinforced by trustworthiness of the partner (King-Casas *et al.*, 2005, 2008). In addition, healthy individuals base their behaviour on pre-established perceptions, such as the belief that the partner has a good moral character (Delgado *et al.*, 2005). These results make it especially interesting to study psychosis with the trust game, because there is evidence that the integration of new experiences and information into prior belief systems and decision-making goes awry in affected patients. First, deficits in the ability to integrate new evidence and contextual information into pre-established belief systems have been observed (Cohen *et al.*, 1996; Mortiz and Woodward, 2006; Woodward *et al.*, 2008; Veckenstedt *et al.*, 2011). Secondly, patients with psychosis exhibit difficulties using reward and feedback learning to guide decision-making in various settings (Waltz and Gold, 2007; Fletcher and Frith, 2009). An aberrant sensitivity to the rewarding value of social interactions and a reduced sensitivity to significant contextual information may underlie social dysfunction and the inability to correct negative beliefs about others on the basis of contradictory information.

There is now ample evidence that the mechanisms that are associated with symptoms of psychosis also operate in individuals at lower levels of a hypothesized psychosis continuum, as for example, healthy first-degree relatives of affected patients. A lack of trust may clinically manifest itself as paranoid delusions and social withdrawal, but at the subclinical level as suspiciousness, feelings of ill will or resentment (Fenigstein and Vanable, 1992; Langdon

and Coltheart, 1999; Peters *et al.*, 1999; Johns and van Os, 2001; Krabbendam *et al.*, 2004; Dael *et al.*, 2005). Epidemiological studies have shown that subclinical psychotic experiences are much more frequent among the general population (i.e. up to 15%) than the rates of clinical psychotic disorder (i.e. ~1%; Janssen *et al.*, 2003; Krabbendam *et al.*, 2004; Versmissen *et al.*, 2008). Investigating the subclinical phenotype is conceptually important because it will help to elucidate mechanisms of transition over the psychosis continuum, but also methodologically important, as it permits the investigation of cognitive processes without the confounding effects of antipsychotic medication. If transitory psychotic experiences are widely present in the general population, the question of which factors protect against the transition to clinical status becomes crucially important (Krabbendam *et al.*, 2005). Based on cognitive models of psychosis and previous research on cognitive biases and belief flexibility (Hemsley, 1993; Cohen *et al.*, 1996; Mortiz and Woodward, 2006; Woodward *et al.*, 2008; Veckenstedt *et al.*, 2011), we hypothesize that the evolution from subclinical psychotic symptoms to clinical psychosis might be associated with a failure in the utilization of bottom-up (i.e. feedback learning) or top-down reasoning (i.e. integration of *a priori* information).

Materials and methods

Participants

The experimental group consisted of 29 patients with diagnoses within the Diagnostic and Statistical Manual cluster schizophrenia or other psychotic disorders (16 schizophrenia, 4 schizoaffective disorder, 1 schizophreniform disorder, 5 undifferentiated type and 3 psychotic disorder not otherwise specified) according to the research diagnostic criteria (Spitzer, 1978), 24 healthy first-degree relatives of patients with non-affective psychosis (13 siblings, 2 offspring and 9 parents) and 35 healthy controls from the general population without a personal or family history of psychosis. Additionally, 176 students participated as trustees in the trust game. Patients were recruited from the assertive community treatment teams, outpatient clinics of the catchment area or patient associations. Relatives were sampled through participating patients or associations for relatives of patients. Control subjects were recruited through adverts in the local area. Students were recruited via adverts within the university. The study was approved by the Medical Ethics Committee of the University Medical Centre Maastricht.

The inclusion criteria for all participants were: (i) ability to give informed consent; (ii) age between 18 and 60 years; (iii) sufficient command of the Dutch language; (iv) having an IQ > 80; (v) no lifetime history of psychosis for controls and first-degree relatives; and (vi) no family history of psychosis for healthy controls. Participants were excluded if they had: (i) a diagnosis of current drug or alcohol dependence; (ii) brain disease or damage; or if they (iii) used psychotropic medication (except patients).

Measures

Patients' diagnoses were confirmed with the Comprehensive Assessment of Symptoms and History (Andreasen *et al.*, 1992). A more detailed assessment of their symptoms included the Positive

and Negative Syndrome Scale (PANSS; Kay *et al.*, 1987). The PANSS is based on findings that schizophrenia comprises at least two distinct syndromes. The positive syndrome consists of productive symptoms, such as paranoid delusions and hallucinations, while the negative syndrome consists of deficit features, such as emotional withdrawal or blunted affect. Subclinical psychotic symptoms were assessed with the Community Assessment of Psychic Experiences (Stefanis *et al.*, 2002). The Green Paranoid Thought Scale (Green *et al.*, 2008) was used for a further specification of psychotic symptoms. The exclusion criterion substance dependence was confirmed with the Composite International Diagnostic Interview (Robins *et al.*, 1989). The vocabulary subtest of the Wechsler Adult Intelligence Scale III (Wechsler, 1997) was used as indicator of general cognitive ability. Patients were also interviewed about demographic information.

The trust game

We experimentally investigated trusting behaviour in patients, first-degree relatives and a healthy control group by means of a modified version of a multi-round trust game (Berg *et al.*, 1995; King-Casas *et al.*, 2005, 2008). Participants in the experimental groups were investors throughout the game. The role of trustees was taken on by students (see 'Material and Methods' section). At the beginning of each round of the trust game, both the investor and the trustee received €10 to ensure that investors did not make their initial investment choice to avoid inequality. Investors were then asked to transfer any (integer) amount between €0 and €10 to the anonymous trustee. The transferred amount was tripled by the experimenter. After having received that amount, the trustee decided whether to honour the investor's trust and send all or some of the received money back or behave untrustworthy and keep all or most of the received money.

In order to examine whether social dysfunction is driven by aberrant bottom-up feedback learning or top-down processing of socially relevant information, we implemented three conditions. First, the 'baseline condition' (five rounds) was used to get an unbiased estimate of basic trust. Therefore, the investor was not informed about how much the trustee reciprocated in a round and feedback about the reciprocated amounts was given only at the end of the whole experiment. The investors were aware of this when making their decisions. Second, the 'context condition' (five rounds) was used to investigate the effects of *a priori* information about the partners' trustworthiness. The investor was matched with a new trustee and, before making the investments, shown truthful information about this trustees' trustworthiness during the baseline condition. Specifically, the investor was informed whether the trustee, on an average, returned less, as much as, or more than the amount invested by the investor. To ensure no confounds in the investigation of the contextual effect, the investor received no feedback about the trustees' actual trustworthiness in the current condition. Third, the 'feedback condition' (10 rounds) was used to obtain a measure of the effect of bottom-up feedback learning on prior beliefs and the development of trust and reciprocity in consecutive interactions. In this condition, the investor received direct feedback about the amount of money that was reciprocated by the trustee in response to each investment. Investors interacted with the same trustee within each condition, but with different trustees across the three conditions to minimize carry-over effects from one condition to another. The same set of trustees played with two different investors in the baseline and context condition, because the lack of feedback in the former precluded carry-over effects between the conditions. As information about the average trustworthiness in the baseline condition was given in the context condition, all investors were paired up with a new trustee in the feedback condition. All participants were

informed about this before making their decisions in the respective condition.

Specifically, we hypothesized that psychosis risk would be associated with a loss of basic trust, a diminished sensitivity to contextual information and an inability to engage in trusting interactions. This would be reflected in smaller invested amounts in the baseline condition, a reduced effect of information about the counterparts' trustworthiness on investment behaviour and reduced reciprocal trust in response to observed trustworthy behaviour by the partner in the feedback condition.

Procedure

Participants read the information material and gave written informed consent during a first testing session that comprised a brief interview about demographic information, the Comprehensive Assessment of Symptoms and History, the PANSS, the Composite International Diagnostic Interview, the Green Paranoid Thought Scale and the Wechsler Adult Intelligence Scale subtest. The trust game was computerized in z-Tree (Fischbacher, 2007) and conducted in the Behavioural and Experimental Economics laboratory of the Department of Economics at Maastricht University in a second session. The lab was equipped with cabinets that prevented participants from seeing or communicating with each other in order to circumvent unwanted effects of confounding factors, such as age, gender, race or patient status on the trust game behaviour of both players. Computers in each cabinet were connected to a network, and participants were paired up anonymously. Experiment instructions for investors and trustees were presented in the cabinets (see Supplementary Material). The instructions were identical, except that investors were told that they would be taking on role A throughout the game, while trustees were informed that they would take on role B throughout the game. To make sure that all participants understood the rules of the experiment correctly, they had to complete a comprehension questionnaire with calculations and questions before the start of the experiment (see Supplementary Material). The earnings depended on the decisions made. In the baseline and context condition, no feedback about earnings was given. In the feedback condition, participants' personal earnings from each investment round were displayed on the computer screen once a game round was completed. At the end of the experiment, the money earned in each of the 20 investment rounds (including baseline and context condition) was shown to the participants. The earnings from one randomly chosen round were actually paid out to each participant. All participants were informed of the payout procedure before they made their decisions. The average earnings were €13.19 for investors and €9.09 for trustees. Investors additionally received vouchers worth €37.50 for attending the first testing session and the experiment. Trustees had fewer expenses in terms of time and were compensated with a €10 show-up fee.

Data analysis

The analysis was conducted in STATA 11.0 (Statacorp, 2009). Investments in the baseline and context condition were skewed to the left (i.e. the mass of the investments was at the higher end of the possible investments) and transformed into a three-level variable (tertile based) for normalization. This transformation was based on the points that divided the ordered distribution into three parts, each containing one-third of the population. Despite clear hypotheses about the directionality of the effects, we chose to conduct conservative two-tailed statistical testing to allow for the possibility of a relationship in both directions.

Analyses of the baseline and context condition

Multilevel random regression analyses were used to investigate differences in trusting behaviour across groups and conditions, while controlling for multiple observations [investments (Level 1); within subjects (Level 2)]. In the context condition, only cooperative context was analysed because in 81% of the cases the trustees reciprocated average amounts that were equal to or higher than the average investments of the baseline condition. The effect of context information on the amount of mean investments (tertile based) in comparison to the baseline condition was analysed by means of multilevel random regression analyses within each participant group. Age and gender were included in all analyses as *a priori* confounders. Associations with symptoms were analysed using the PANSS-positive and negative symptom score within patients, and the respective dimensions on the Community Assessment of Psychic Experiences within relatives and controls.

Analyses of the feedback condition

For the analysis of the interactions in the feedback condition, we classified each decision made by the investor in each round as being in one of the following four categories: (i) trust honouring: the trustee did not decrease the repaid amount from two rounds ago to the preceding round and, in response, the investor did not decrease the investment from the preceding to the current round; (ii) trust repairing: the trustee did decrease the reciprocated amount from two rounds ago to the preceding round and the investor, nevertheless, did not decrease the investment from the preceding to the current round; (iii) trust disrupting: the trustee did not decrease the reciprocated amount from two rounds ago to the preceding round but the investor, nevertheless, did decrease the investment from the preceding to the current round; and (iv) distrust reciprocating: the trustee did decrease the reciprocated amount from two rounds ago to the preceding round and, in response, the investor did decrease the investment from the preceding to the current round. Table 1 summarizes these behavioural categories. Multiple regression analyses were used to investigate group differences in the probabilities of engaging in the four different trusting behaviours. Age and gender were included in all analyses as *a priori* confounders. Group differences in the frequency distributions of the four different trusting behaviours (trust honouring, trust repairing, trust disrupting and distrust reciprocating) were calculated with chi-square tests.

Table 1 Contingency table of the behavioural categories in the feedback condition

Investor	Trustee	
	$\Delta T(t) \geq 0$	$\Delta T(t) < 0$
$\Delta I(t) \geq 0$	Trust honouring	Trust repairing
$\Delta I(t) < 0$	Trust disrupting	Distrust reciprocating

An investment in round t was defined as $I(t)$. The amount the trustee returned in round t was defined as $T(t)$. For investors, the change in investments from t to $t - 1$ was defined as $\Delta I(t) = I(t) - I(t - 1)$. Similarly, the change in the reciprocated amount relative to the received amount by the trustee in the preceding rounds, from $t - 1$ to $t - 2$ was defined as $\Delta T(t) = T(t - 1)/(I(t - 1) \times 3) - T(t - 2)/(I(t - 2) \times 3)$. The matrix is assumed to be stationary (i.e. probabilities of contingent behaviours do not change over time). The underlying assumption is a Markov decision process, where the probability with which the investor executes a specific behaviour is contingent on the behaviour of the trustee in the preceding rounds (Sutton and Barto, 1998).

Depending on the statistical test that has been used, effect sizes are expressed as the regression coefficient b and the 95% confidence interval.

Results

The three groups did not differ in age and estimated IQ. There were significant gender differences. The group of relatives included a lower percentage of males than the patient group [$\chi^2(1) = 6.47, P < 0.01$], but healthy controls did not significantly differ from patients or relatives [$\chi^2(1) = 3.08, P = 0.08$ and $\chi^2(1) = 0.96, P = 0.32$, respectively, Table 2]. Relatives and controls did not differ in the level of reported psychopathology, as indicated by the PANSS, the Community Assessment of Psychic Experiences and the Green Paranoid Thought Scale. Sample characteristics and demographic data are displayed in Table 2.

Table 3 shows means and standard deviations (SD) of the investments in Euro for the original and transformed (three-level, tertile based) data by experimental group and condition (see 'Material and methods' section).

Baseline and context condition

The descriptive statistics in Table 3 suggest higher baseline trust of controls (mean = 6.69) in comparison to relatives and patients (mean = 5.61 and 5.63, respectively). This impression was corroborated by multilevel random regression analyses comparing baseline trust pairwise between the three groups (Table 4). Patients invested significantly less than controls ($\beta = -0.32, P < 0.05$) and the investments of relatives were marginally significantly smaller than those of controls ($\beta = -0.28, P = 0.10$). Investments of relatives were similar to those of patients and did not differ significantly ($\beta = -0.04, P = 0.86$). The difference between patients and controls was reduced to marginal significance when IQ was added to the model as a covariate ($\beta = -0.26, P = 0.10$). None of the other effects changed. Investments in the context condition show an interesting change in pattern. Prior positive information on the trustees' trustworthiness significantly increased investments of both controls (mean = 7.73, $\beta = 0.13, P < 0.05$) and relatives (mean = 6.97, $\beta = 0.19, P < 0.05$), but had no effect on patients' investments (mean = 5.64, $\beta = 0.07, P = 0.37$; Tables 3 and 4). In line with our hypotheses, the findings support that psychosis risk is associated with reduced baseline trust. Importantly, however, only clinical psychosis was associated with a diminished sensitivity to contextual information.

Across the baseline and context condition, positive psychotic symptoms (see 'Material and methods' section) were associated with lower investments in patients with marginal significance ($\beta = -0.35, P = 0.09$), but negative symptoms were not ($\beta = 0.03, P = 0.89$). Similarly, in relatives subclinical positive symptoms were associated with lower investments with marginal significance ($\beta = -2.2, P = 0.06$), while subclinical negative symptoms were not ($\beta = 0.12, P = 0.74$). Within controls, no association was present between subclinical positive symptoms and investments ($\beta = -0.40, P = 0.46$). However, a significant association

Table 2 Participant characteristics

Variable	Group			Test statistic
	Controls, <i>n</i> = 35 Mean (SD)	Relatives, <i>n</i> = 24 Mean (SD)	Patients, <i>n</i> = 29 Mean (SD)	
Age	41.3 (12.4)	42.3 (11.6)	39.7 (8.41)	$b = -0.74$ (–3.51 to 2.02)
Male gender (%)	43 ^{a,t,b}	21 ^{b,**}	55 ^a	$\chi^2(2) = 6.49$; $P = 0.04$
Wechsler Adult Intelligence Scale vocabulary	13.51 ^a (2.85)	13.21 ^{a,b} (2.60)	12.31 ^{a,t,b} (3.04)	$b = -0.59$ (–1.30 to 0.11)
Green Paranoid Thought Scale	34.74 ^a (12.06)	30.66 ^a (7.31)	45.83 ^b (21.64)	$b = 5.29$ (1.35–9.23)
Cape positive	0.22 ^a (0.2)	0.12 ^a (0.10)	0.76 ^{b,**} (0.48)	$b = 0.26$ (0.17–0.35)
Cape negative	0.64 ^a (0.35)	0.53 ^a (0.34)	1.15 ^{b,**} (0.44)	$b = 0.24$ (0.14–0.35)
PANSS sum positive	7.08 ^a (0.37)	7.13 ^a (0.34)	11.48 ^{b,**} (3.63)	$b = 2.14$ (1.57–2.71)
PANSS sum negative	7.46 ^a (1.56)	7.42 ^a (0.97)	9.79 ^{b,**} (3.69)	$b = 1.13$ (0.53–1.74)

Different superscripts (i.e. a, b versus a, a) indicate group differences that are significant at [†] $P < 0.10$; ^{**} $P < 0.01$; 95% CI = 95% confidence interval for regression coefficient *b*.

Values are given as *b* unless otherwise specified.

Table 3 Means and standard deviations of the untransformed and transformed investments by experimental group and condition

Condition	Control <i>n</i> = 35 Mean (SD)	Relative <i>n</i> = 24 Mean (SD)	Patient <i>n</i> = 29 Mean (SD)
Baseline			
Transformed	1.23 (0.81)	0.93 (0.87)	0.92 (0.80)
Euro	6.69 (3.22)	5.61 (3.29)	5.63 (2.87)
Cooperative context			
Transformed	1.38 (0.72)	1.18 (0.81)	0.87 (0.76)
Euro	7.73 (2.70)	6.97 (3.08)	5.64 (3.10)
Feedback			
Transformed	1.17 (0.83)	1.11 (0.78)	0.89 (0.76)
Euro	6.87 (3.50)	6.81 (3.16)	5.95 (3.35)

Original investments are in Euro. Transformed investments are based on a three-level variable that has been derived from the tertile split of the investments in Euro.

was present with negative subclinical symptoms ($\beta = 0.66$, $P = 0.02$; Table 4).

Feedback condition

During the feedback condition, investors were given immediate information about the extent of the trustees' trustworthiness at the end of each round. For each investor, we calculated an estimate of the conditional probability of engaging in the four different trusting behaviours (Table 1 and Fig. 1). The distributions of the frequencies of these four behavioural contingencies differed between the experimental groups [$\chi^2(6) = 17.37$, $P < 0.01$]. Across the four categories, patients showed a differential investment behaviour than controls [$\chi^2(6) = 16.17$, $P < 0.01$]. Relatives' behaviour was intermediate to controls and patients, although not significantly different from controls [$\chi^2(6) = 4.80$, $P = 0.19$] or patients [$\chi^2(6) = 3.76$, $P = 0.29$]. Figure 1 suggests that in contrast to healthy controls and relatives, patients engage less often in trust honouring behaviour.

Multiple regression analyses by category showed that patients honoured trust less frequently compared with controls ($\beta = -0.18$, $P < 0.05$). The differences between patients and relatives and relatives and controls were not significant ($\beta = -0.09$, $P = 0.28$ and $\beta = -0.09$, $P = 0.35$, respectively). The effects did not change when IQ was added to the regression model as a covariate. No significant differences were found for any of the other behavioural categories (all $P > 0.11$). These findings support our hypothesis that clinical psychosis is associated with a reduced ability in the utilization of bottom-up reasoning through feedback learning of positive social interactions.

Discussion

In this study, we investigated trust and reciprocity in an incentivized multi-round trust game of 20 consecutive real time social interactions. This method delivers a novel, objective behavioural measure of the dynamics of social interactions and the subjective experience of (dis)trust. The findings in the 'baseline' and 'context condition' provided support for our hypotheses and are in line with the real world social functioning deficits that are frequently seen in patients with schizophrenia (Mathews and Barch, 2010). The results revealed lower basic trust in patients and first-degree relatives compared with healthy controls, suggestive of a familial substrate.

Positive information on the trustees' trustworthiness had no effect on the patients' degree of trust. In contrast, controls and relatives made significantly higher investments when being informed that they interacted with a trustworthy counterpart. When receiving direct behavioural feedback on the partners' trustworthiness patients were insufficiently able to modify their trusting behaviour in response. They honoured trustworthiness of their interaction partner less often than controls. As shown by the increase of mean investments from baseline, relatives modified their trusting behaviour towards the level of controls when they were provided with direct behavioural feedback. In addition, they showed an intermediate probability of trust honouring behaviour.

Table 4 Group differences of the investments in the baseline and context condition and the associations between mean investments across the conditions with symptoms and context effects by group

	Controls versus patients <i>b</i> (95% CI)	Controls versus relatives <i>b</i> (95% CI)	Relatives versus patients <i>b</i> (95% CI)
Baseline investment	−0.32* (−0.73 to −0.1)	−0.28 [†] (−0.71 to 0.16)	−0.04 (−0.51 to 0.42)
Context investment	−0.52** (−0.88 to −0.17)	−0.17 (−0.53 to 0.19)	−0.35 [†] (−0.75 to 0.04)

	Control	Relative	Patient
Mean investments—positive symptoms	−0.40 (−1.45 to 0.65)	−2.2 [†] (−4.42 to 0.09)	−0.35 [†] (−0.76 to 0.05)
Mean Investments—negative symptoms	0.66* (0.11–1.21)	0.12 (−0.61 to 0.86)	0.03 (−0.40 to 0.45)
Context effect	0.13* (0.01–0.25)	0.20* (0.01–0.39)	0.07 (−0.08 to 0.22)

[†] $P < 0.10$; * $P < 0.05$; ** $P < 0.01$; 95% CI = 95% confidence interval for the regression coefficient *b*.

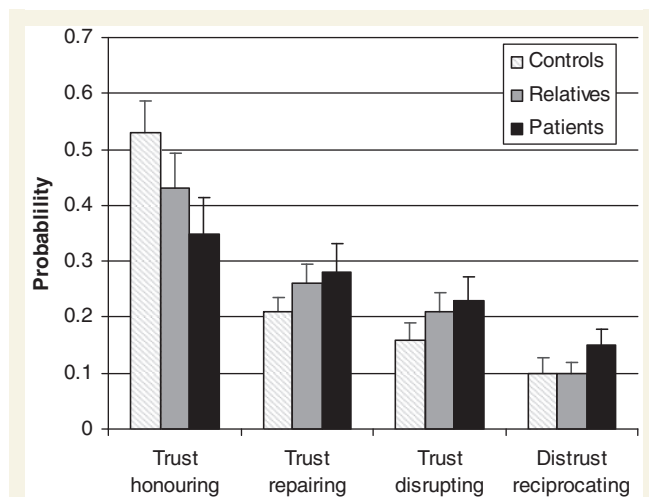


Figure 1 Behavioural contingencies within the feedback condition. Distributions of the conditional probabilities of engaging in the four different trusting behaviours by group. Error bars represent SEs.

Interestingly, patients' responses were marginally significantly associated with their level of subjective positive psychotic symptoms, rather than negative symptoms. A similar pattern was present in relatives; their subclinical positive symptoms were associated with lower investments but their subclinical negative symptoms were not. For the control group, no association was present between subclinical positive symptoms and investments. However, their subclinical negative symptoms were associated with the investments, indicating a fundamentally different relationship between these characteristics and investments in the general population, as compared to those with a liability for psychosis.

Our findings suggest that deficits in social functioning may partly be explained within a framework of social information processing (Hemsley, 1993). To initiate social interactions that maximize mutual individual gains, investor and trustee must induce trust and trustworthiness by investing high amounts and by reciprocating trustworthy in a repeated manner. The maintenance and repair of such a cooperative climate during the course of the

interactions requires subjects to understand social gestures and their consequences (i.e. mentalizing) and to act upon the signals by adapting their responses accordingly. Patients did not only make the lowest investments over all three conditions, they also were the least successful in adapting their decisions to their partners' trustworthy behaviour. Research in healthy subjects suggested that the amount of trial-by-trial learning can be altered by prior information about the moral character of the interaction partner (Delgado *et al.*, 2005). Deficits in the usage of contextual information (Servan-Schreiber *et al.*, 1996; Stratta *et al.*, 1998), impaired top-down processing (Silverstein *et al.*, 2006) and an impaired ability in the integration of new evidence have previously been reported in patients with schizophrenia (Mortiz and Woodward, 2006; Woodward *et al.*, 2008; Veckenstedt *et al.*, 2011) and recent research suggested that patients had an inflexible decision-making style (Veckenstedt *et al.*, 2011). In line with this research, patients' failure to adapt their investment decisions to the positive contextual information about their interaction partner in our study can be explained by impaired top-down processing and inflexible *a priori* beliefs about the malevolent nature of other humans. Our results indicate that deficits in trust may thus be maintained by a failure to extinguish negative beliefs about others. In relatives, unimpaired top-down contextualizing ability may act as a protective mechanism that counteracts the transition from subclinical psychotic experiences to clinical symptoms. Within their interactions, patients tended to behave less positive towards the partner. Specifically, they honoured trustworthiness less than relatives and controls. Patients' failure to utilize direct behavioural feedback of their partner to guide predictive coding as sensitive as relatives and controls suggests that their social interactions are also disturbed by aberrant outcome based bottom-up processing of social cues.

The first question here is whether these differences in trust game behaviour are due to general cognitive decline in patients. In our study, general intelligence slightly reduced the group difference between patients and controls in the baseline condition, so that it became marginally significant. This does not necessarily imply that cognitive abnormalities cannot be causally implicated in the development of psychosis, but rather suggests that social cognitive impairment, such as apparent in the trust game, overlaps

to some extent with general cognitive impairment in the association with psychosis. This is in line with previous research showing that although overlapping, social and neurocognitive functioning may still be regarded as, at least partly, independent factors (Van Hooren *et al.*, 2008).

The other question is whether the deficit in the lack of 'normal' responsiveness to new evidence is distinct from deficits in learning and reasoning evident in non-social cognition or whether it is the same core deficit but now demonstrated in an active social interaction. First, the underlying mechanisms of the trust game are different from those seen in pure associative learning/probabilistic reasoning tasks, such as the beads task, because in the trust game feedback by the trustee is dependent on investor behaviour. For example, a lower baseline trust sets off less cooperative interactions. Therefore, the current results do not only reflect insensitivity to feedback but also an impaired ability to initiate mutually trusting relationships. Accordingly, social interactions in the trust game are associated with the ability to represent the intentions and goals of others (Sripada *et al.*, 2009). Research has also shown that this social learning can be differentiated from non-social learning through modulation using the 'social' neuropeptide oxytocin. Oxytocin enhanced the effects of social learning, while not affecting learning in a non-social risk game (Baumgartner *et al.*, 2008).

To the extent that behaviour in the trust game is influenced by deficits in feedback learning, it is important to note that these are dependent on the type of stimuli used. For example, data from a study of probabilistic learning and inference in schizophrenia using a stochastic sequence learning and the beads task (Averbeck *et al.*, 2011) suggests that patients demonstrate decreased performance during the learning task; however, they were able to use feedback to change behaviour—but more slowly than healthy subjects. In a similar vein, previous work looking at a stochastic reward learning task (Evans *et al.*, 2011) found that patients with schizophrenia were able to use feedback to improve their performance, evident in significantly above chance performance. In summary, the trust game is associated with specific social mechanisms that are not evident during non-social tasks. Furthermore, the lack of 'normal' responsiveness to feedback is not a consistent finding throughout non-social conditions, and when it is evident, it varies depending on the nature of the stimuli. This suggests that the deficits evident in our study presented here are not driven by a non-specific non-social core deficit, but are specific to the social interaction in this study.

A question that can be raised is whether the patients' failure to adjust their trusting behaviour to their interaction partner may ground in alterations in brain functioning, which are associated with psychotic illness. Neuroimaging studies in healthy populations have showed associations between trust game performance and the activation in brain areas that have been related to reward learning and the ability to adapt behaviour to changing circumstances (Rolls, 2000; Rahman *et al.*, 2001; Rilling *et al.*, 2002; King-Casas *et al.*, 2005, 2008; Sanfey, 2007; Phan *et al.*, 2010). A deregulation of dopamine, the main neurotransmitter in these processes, is thought to play a key role in the pathophysiology of psychosis by altering reward learning mechanisms (Carlsson, 1988; Berridge and Robinson, 1998; Kapur *et al.*, 2005; Smith *et al.*,

2006). Possibly, the observed association between positive symptoms and reduced contextual sensitivity is due to aberrant social reward learning, i.e. insensitivity to the motivational aspects of social cues. As a consequence, the ability to make predictions of the partners' behaviour and the translation of these predictions into optimal decisions could be impaired and may give rise to the formation and instantiation of psychotic symptoms. Future neuroimaging studies may help to identify the neural basis of these processes.

Our paradigm allowed for an investigation of trust in consecutive real time interactions. It enabled us to investigate the effects of contextual information and direct behavioural feedback of the interaction partner, while isolating other contextual factors that could have influenced decision-making (e.g. gender, nationality or verbal ability). It has proven useful as a means to investigate social interaction in patients and healthy individuals and can aid the understanding of the association between social cognitive impairment and social dysfunction and, possibly, psychotic symptoms. As a direct behavioural measure, the trust game has clear advantages over self-report and it also allows for manipulations of 'online' social behaviour. It offers a more objective way to assess social interactions and the subjective experience of distrust and is sensitive enough to pick up plain behavioural signs independent of frequently impaired verbalizing abilities. Being more objective than self-report, the trust game results also do not suffer the potential confounds of under-reporting of patients and family members that may be caused by pressure to act normal.

This first of a kind investigation linked psychosis liability to impairments in trust by establishing different behavioural patterns among patients, relatives and healthy controls. We see this as a first necessary step to explore the underlying mechanisms of (dis)trusting behaviour and social malfunctioning in these groups. In line with earlier research, our findings indicate an underlying familial substrate of disturbed trusting behaviour (Cesarini *et al.*, 2008). Yet, psychotic illness is also associated with negative social experiences, which may further reinforce distrusting behaviour. To clarify the role of trust and reciprocity in the onset and maintenance of specific symptoms, it is important to examine whether these behavioural alterations are present at illness onset or whether they are associated with a prolonged illness duration and symptom fluctuation.

In conclusion, healthy social interactions and partnership building are characterized by behavioural flexibility in response to socially relevant cues. Crucially, a lack of flexibility differentiated patients with psychosis from those at risk of the illness. While lower levels of trust were initially present in those at risk, they showed a comparable sensitivity to contextual information and direct behavioural feedback as controls. Hence, behavioural flexibility in response to social cues does not only determine the success of social interactions in healthy people but also in those with subclinical psychotic experiences. These results strongly suggest that behavioural flexibility may act protectively against the transition from a subclinical into a clinical state. A lack thereof could underlie social dysfunction in patients and may also contribute to the instantiation and maintenance of positive symptoms. This offers a testable mechanistic hypothesis for progression from prodrome to psychotic illness, and may provide a

therapeutic avenue to grapple the psychotic symptoms of social dysfunction.

Acknowledgements

We would like to thank all the participants who contributed to this study.

Funding

Netherlands Organization for Scientific Research (VIDI, grant number MAGW 452-07-007, to L.K.); HEFCE Clinical Senior Lecturer Award (to S.S.S.) and Oesterreichische Nationalbank (grant number 11780 to A.R.).

Supplementary material

Supplementary material is available at *Brain* online.

References

- Adolphs R. How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Res Rev* 2006; 1079: 25–35.
- Andreasen NC, Flaum M, Arndt S. The Comprehensive Assessment of Symptoms and History (CASH). An instrument for assessing diagnosis and psychopathology. *Arch Gen Psychiatry* 1992; 49: 615–23.
- Averbeck BB, Evans S, Chouhan SV, Bristow E, Shergill SS. Probabilistic learning and inference in schizophrenia. *Schiz Res* 2011; 127: 115–22.
- Baumgartner T, Heinrichs M, Vonlanthen A, Fischbacher U, Fehr E. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* 2008; 58: 639–50.
- Berg J, Dickhaut J, McCabe K. Trust, reciprocity and social history. *Games Econ Behav* 1995; 10: 122–42.
- Berridge KC, Robinson TE. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Rev* 1998; 28: 309–69.
- Camerer CF. Psychology and economics. Strategizing in the brain. *Science* 2003; 300: 1673–75.
- Carlsson A. The current status of the dopamine hypothesis of schizophrenia. *Neuropsychopharmacology* 1988; 1: 179–86.
- Cesarini D, Dawes CT, Fowler JH, Johannesson M, Lichtenstein P, Wallace Br. Heritability of cooperative behavior in the trust game. *Proc Natl Acad Sci USA* 2008; 105: 3721–26.
- Cohen JD, Braver TS, O'Reilly RC. A computational approach to prefrontal cortex, cognitive control and schizophrenia: recent developments and current challenges. *Philos Trans R Soc Lond B Biol* 1996; 351: 1515–27.
- Delgado MR, Frank RH, Phelps EA. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci* 2005; 8: 1611–8.
- Evans S, Shergill SS, Chouhan V, Bristow E, Collier T, Averbeck BB. Patients with schizophrenia show increased aversion to angry faces in an associative learning task. *Psychol Med* 2011; 41: 1471–9.
- Fenigstein A, Vanable PA. Paranoia and self-consciousness. *J Pers Soc Psychol* 1992; 62: 129–38.
- Fehr E. On the economics and biology of trust. *J Eur Econ Assoc* 2009; 7: 235–66.
- Fischbacher U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp Econ* 2007; 10: 171–78.
- Fletcher PC, Frith CD. Perceiving is believing: a bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci* 2009; 10: 48–58.
- Glaeser E, Laibson D, Scheinkman J, Soutter C. Measuring trust. *Q J Econ* 2000; 115: 811–46.
- Green C, Freeman D, Kuipers E, Bebbington P, Fowler D, Dunn, Garety PA. Measuring ideas of persecution and reference: the Green et al paranoid thought scales (GPTS). *Psychol Med* 2008; 38: 101–11.
- Hemsley DR. A simple (or simplistic?) cognitive model for schizophrenia. *Behav Res Ther* 1993; 31: 633–45.
- Houser D, Schunk D, Winter J. Distinguishing trust from risk: an anatomy of the investment game. *J Econ Behav Organ* 2010; 74: 72–81.
- Janssen I, Krabbendam L, Jolles J, van Os J. Alterations in theory of mind in patients with schizophrenia and non-psychotic relatives. *Acta Psychiatr Scand* 2003; 108: 110–7.
- Johns LC, van Os J. The continuity of psychotic experiences in the general population. *Clin Psychol Rev* 2001; 21: 1125–41.
- Kapur S, Mizrahi R, Li M. From dopamine to salience to psychosis-linking biology, pharmacology and phenomenology of psychosis. *Schizophr Res* 2005; 79: 59–68.
- Karlan DS. Using experimental economics to measure social capital and predict financial decisions. *Amer Econ Rev* 2005; 95: 1688–99.
- Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophr Bull* 1987; 13: 261–76.
- King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR. The rupture and repair of cooperation in borderline personality disorder. *Science* 2008; 321: 806–10.
- King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR. Getting to know you: reputation and trust in a two-person economic exchange. *Science* 2005; 308: 78–83.
- Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. Oxytocin increases trust in humans. *Nature* 2005; 435: 673–76.
- Krabbendam L, Hanssen M, Bak M, van Os J. Psychotic features in the general population. Risk factors for what? In: Gattaz WF, Hafner H, editors. Search for the causes of schizophrenia. Darmstadt, Germany: Dietrich Steinkopff Verlag; 2004. p. 54–78.
- Krabbendam L, Myin-Germeys I, Bak M, van Os J. Explaining transitions over the hypothesized psychosis continuum. *Aust N Z J Psychiatry* 2005; 39: 180–86.
- Krabbendam L, Myin-Germeys I, van Os J. The expanding psychosis phenotype. *Int J Psychol Psychol Ther* 2004; 4: 411–20.
- Langdon R, Coltheart M. Mentalising, schizotypy, and schizophrenia. *Cognition* 1999; 71: 43–71.
- Mathews JR, Barch DM. Emotion responsivity, social cognition, and functional outcome in schizophrenia. *J Abnorm Psychol* 2010; 119: 50–59.
- Mortiz M, Woodward TS. A generalized bias against disconfirmatory evidence in schizophrenia. *Psych Res* 2006; 142: 157–65.
- Peters ER, Joseph SA, Garety PA. Measurement of delusional ideation in the normal population: introducing the PDI (Peters et al. Delusions Inventory). *Schizophr Bull* 1999; 25: 553–76.
- Phan KL, Sripada CS, Angstadt M, McCabe K. Reputation for reciprocity engages the brain reward center. *Proc Natl Acad Sci USA* 2010; 107: 13099–104.
- Rahman S, Sahakian BJ, Cardinal RN, Rogers RD, Robbins TW. Decision making and neuropsychiatry. *Trends Cogn Sci* 2001; 5: 271–77.
- Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C. A neural basis for social cooperation. *Neuron* 2002; 35: 395–405.
- Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, et al. The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Arch Gen Psychiatry* 1989; 45: 1069–77.
- Rolls ET. The orbitofrontal cortex and reward. *Cereb Cortex* 2000; 10: 284–94.
- Sanfey A. Social decision-making: insights from game theory and neuroscience. *Science* 2007; 318: 598–601.
- Servan-Schreiber D, Cohen JD, Steingard S. Schizophrenic deficits in the processing of context. *Arch Gen Psychiatry* 1996; 53: 1105–12.

- Silverstein S, HatashitaWong M, Schenkel L, Wilkniss S, Kovács I, Fehér A, et al. Reduced top-down influences in contour detection in schizophrenia. *Cognit Neuropsychiatry* 2006; 11: 112–32.
- Smith A, Li M, Becker S, Kapur S. Dopamine, prediction error and associative learning: a model-based account. *Network* 2006; 17: 61–84.
- Spitzer RL. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry* 1978; 35: 773–82.
- Sripada CS, Angstadt M, Banks S, Nathan PJ, Liberzon I, Phan KL. Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *NeuroReport* 2009; 20: 984–9.
- Statacorp. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP; 2009.
- Stefanis NC, Hanssen M, Smirnis NK, Avramopoulos DA, Evdokimidis IK, Stefanis CN, et al. Evidence that three dimensions of psychosis have a distribution in the general population. *Psychol Med* 2002; 32: 347–58.
- Stratta P, Daneluzzo E, Bustini M, Casacchia M, Rossi A, Barch DM, et al. Schizophrenic deficits in the processing of context. *Arch Gen Psychiatry* 1998; 55: 186–88.
- Sutton RS, Barto AG. *Reinforcement learning: an introduction*. Cambridge, MA/London, England: The MIT Press; 1998.
- van Dael F, Versmissen D, Janssen I, Myin-Germeys I, van Os J, Krabbendam L. Data gathering: biased in psychosis? *Schizophr Bull* 2005; 32: 341–51.
- Van Hooren S, Versmissen D, Janssen I, Myin-Germeys I, a Campo J, Mengelers R, et al. Social cognition and neurocognition as independent domains in psychosis. *Schizophr Res* 2008; 103: 257–65.
- Veckenstedt R, Randjbar S, Vitzthum F, Hottenrott B, Woodward T, Moritz S. In corrigibility, jumping to conclusion, and decision threshold in schizophrenia. *Cogn Neuropsychiatry* 2011; 16: 174–92.
- Versmissen D, Janssen I, Myin-Germeys I, Mengelers R, a Campo J, van Os J, et al. Evidence for a relationship between mentalising deficits and paranoia over the psychosis continuum. *Schizophr Res* 2008; 99: 103–10.
- Waltz JA, Gold JM. Probabilistic reversal learning impairments in schizophrenia: further evidence of orbitofrontal dysfunction. *Schizophr Res* 2007; 93: 296–303.
- Wechsler D. *WAIS-III: Wechsler Adult Intelligence Scale*. 3rd edn., Administration and Scoring Manual. San Antonio, TX: Psychological Corporation; 1997.
- Woodward TS, Moritz S, Menon M, Klinge R. Belief inflexibility in schizophrenia. *Cogn Neuropsychiatry* 2008; 13: 267–77.