Schizophrenia illness severity is associated with reduced loss aversion

Currie, James; Buruju, Dheeraj; Perrin, Jennifer S.; Reid, Ian C.; Steele, J. Douglas; Feltovich, Nick

Published in:
Brain Research

DOI:
10.1016/j.brainres.2017.03.006

Publication date:
2017

Document Version
Peer reviewed version

Schizophrenia Illness Severity is Associated with Reduced Loss Aversion

J Currie, D Buruju, J S Perrin, I C Reid, J D Steele, N Feltovich

ABSTRACT

Loss aversion, whereby losses weigh more heavily than equal-sized gains, has been demonstrated in many decision-making settings. Previous research has suggested reduced loss aversion in schizophrenia, but with little evidence of a link between loss aversion and schizophrenia illness severity. In this study, 20 individuals with schizophrenia and 16 control participants, matched by age and sex, played two versions of the Iterated Prisoners' Dilemma, one version with only positive payoffs and another version in which negative payoffs were possible, with the second version being derived from the first by subtracting a constant value from all payoffs. The control group demonstrated significantly lower cooperation rates under negative payoffs, compared with the version with only positive payoffs, indicative of loss aversion. The patient group on average showed no loss aversion response. Moreover, the extent of loss aversion in patients was found to be negatively correlated with schizophrenia illness severity, with less ill patients showing loss aversion more similar to controls. Results were found to be robust to the inclusion of potential confounding factors as covariates within rigorous probit regression analyses. Reduced loss aversion is a feature of schizophrenia and related to illness severity.

Key words:

Psychiatry

Decision-making

Game theory

Experimental economics

1. Introduction

Loss aversion is the tendency for individuals to weight losses more heavily than equal-sized gains in decision making. It has been amply documented in economic and psychological studies, found present in individual and strategic decision-making by healthy adults (Kahneman and Tversky 1984; Cachon and Camerer 1996; Rydval and Ortmann 2005; Feltovich 2011; Feltovich et al., 2012), as well as in children (Harbaugh et al., 2001) and non-human primates (Chen et al., 2006).

Here, we examine the association between loss aversion and schizophrenia using a novel version of the Iterated Prisoner's Dilemma paradigm to characterize loss aversion. There are good reasons to expect that loss aversion may be disrupted in schizophrenia. First, the dopamine hypothesis of schizophrenia attributes the psychosis of schizophrenia to abnormal dopaminergic activity in the brain, based on the common action of all antipsychotic medications and psychotomimetic effects of stimulant drugs (Kapur 2003; Gradin et al., 2011); while functional MRI work has found that, in healthy individuals, brain regions associated with loss aversion include dopamine projection areas, from the mesolimbic and mesocortical systems (Tom et al., 2007) including the amygdala and striatum (Canessa et al., 2013; Sokol-Hessner et al., 2013), areas implicated in the neural processing of gains and losses.

Second, schizophrenia has also been considered a disorder of *salience*. Aberrant salience theory proposes that psychosis, related to dopamine dysfunction, occurs due to abnormalities in the salience given to stimuli and internal representations, whether rewarding or aversive (Kapur 2003). For instance, it has been reported that schizophrenia patients show abnormal neural responses to stimuli in both threatening and safe contexts (Jensen et al., 2008; Holt et al., 2012). Aberrant salience may well contribute to an imbalance in the weighting given to losses and gains, necessary for loss aversion, and this provides further plausibility to the prediction that loss aversion is disrupted in schizophrenia.

There has also been work aimed at studying loss aversion in schizophrenia, which indicate that the weighting given to losses is reduced in schizophrenia (Trémeau et al., 2008; Brown et al., 2013). However, these studies were limited by a similarity between loss aversion and other well-known decision-making heuristics. Trémeau et al. (2008) report evidence consistent with reduced *exchange asymmetry* (also known as the *endowment effect*) in schizophrenia patients, compared to healthy controls, and attribute this difference to reduced loss aversion. However, recent research suggests that exchange asymmetry may actually be due to factors other than loss aversion (Plott and Zeiler 2005; Engelmann and Hollard 2010)[1]. Brown et al. (2013) report results from two tasks intended to measure loss aversion. Their first task, based on a paradigm from De Martino et al. (2006), is actually a measure of the *reflection effect* (Kahneman and Tversky 1979), the tendency to prefer more risk when making decisions over losses and less risk when making decisions over gains. Their second task, based on a paradigm from Tom et al. (2007), does measure loss aversion. However, whilst they do find reduced sensitivity to losses in schizophrenia patients versus controls, they find an even greater reduction in sensitivity to gains, so that estimated loss aversion does not significantly differ between the two groups. Also, neither Trémeau et al. (2008) nor Brown et al. (2013) report significant correlations between their measures of loss aversion and schizophrenia symptom severity. If altered loss aversion in schizophrenia does indeed relate to aberrant salience (as suggested by Trémeau et al., 2008), and with aberrant salience linked to dysfunction in dopamine firing (Kapur 2003; Gradin et al., 2011), one would expect that not only does loss aversion differ between schizophrenic patients and healthy controls, but also that it should vary within patients according to illness severity. Interestingly, Trémeau et al. (2008) did find that their loss aversion estimate in patients with schizophrenia was negatively correlated with age, duration of illness, number of months in hospital, and poorer performance in the Wisconsin Card Sorting Test, and they

1. Trémeau et al. (2008) never mention the terms *endowment effect* or *exchange asymmetry* in their paper, instead using *loss aversion* exclusively, while Brown et al. (2013) do characterize the earlier paper's results as pertaining to the *endowment effect* rather than *loss aversion*.

report a nearly significant result with the Positive And Negative Syndrome Scale (PANSS) Positive subscore ($r = -0.29$, $p = 0.08$).

Based on this previous research into loss aversion, and taking into consideration dopamine dysfunction and aberrant salience in schizophrenia, in our study we hypothesized that patients would exhibit reduced loss aversion compared to controls and that within the patient group loss aversion would correlate with schizophrenia illness severity, particularly measured through the PANSS.

2. Results

Our results are based on a novel implementation of the Iterated Prisoners' Dilemma paradigm, with participants playing against a computer player following pre-programmed strategies, and with the game paradigm involving two versions, each consisting of 50 rounds. One version (*'Gain frame'*) simulates a situation of plenitude, with all outcomes yielding gains. The other version (*'Loss frame'*), derived from the first by subtracting a constant from all payoffs, simulates a situation of hardship, with both losses and gains possible (Figure 1). This method of subtracting a constant has been used in previous research to test loss aversion and loss avoidance in other games of cooperation and competition, such as the Stag Hunt and Hawk-Dove games (Cachon and Camerer 1996; Feltovich 2011; Feltovich et al., 2012; Rydval and Ortmann 2005). Based on a game theoretic model incorporating psychological assumptions – see Supplementary Material (Brain Research online) – we show that as loss aversion becomes more pronounced, *Betray* choices become more likely, or equivalently, *Cooperate* choices become less likely. This provides a simple within-participant measure of loss aversion: the fraction of *Cooperate* choices in the *Gain* frame minus the corresponding fraction in the *Loss* frame, with larger values representing greater loss aversion. Clearly, since the *Gain* and *Loss* frames differ only through the subtraction of a constant in the payoffs, other factors between the two frames remain unchanged including the pre-programmed computer strategies. Additional benefits of the Iterated Prisoners' Dilemma paradigm include its relative simplicity and intuitive game mechanics, desirable given the cognitive impairments associated with schizophrenia (Aylward et al., 1984).

Figure 2 presents unconditional frequencies of *Cooperate* choices for controls and patients under both *Gain* and *Loss* frames[2]. As hypothesized, controls exhibit significant loss aversion, as

2. In about 1.5 percent of observations, the participant did not make a choice before the allotted time expired. When this happened, the participant received a payoff of 3 for that round in the *Gain* frame and -2 in the *Loss* frame, and the previous choice made by the participant was used for determining the computer's next-round choice. The results we report are based on all observations; however, the

Cooperate choices fall from 41.2% in the *Gain* frame to 32.9% in the *Loss* frame. This difference is significant ($p = 0.01$, Figure 2), corresponding to a mean loss aversion score of 8.3 percentage points (see also Figure 3). By contrast, patients' frequencies of *Cooperate* choices are not significantly different in the two frames (39.92% and 39.85%, *Gain* and *Loss* frames respectively, $p = 0.35$, Figure 2). The corresponding mean loss aversion score for patients of less than 0.1 percentage points is significantly lower than for controls ($p = 0.04$, Figure 3).

Figure 4 shows the relationship between patients' loss aversion scores and their PANSS Total scores. For the patient group, ordinary least squares regression shows a significant negative association ($R^2 = 0.32$; $p = 0.01$). As patient illness severity increases, loss aversion tends to reduce, with less ill patients showing loss aversion more similar to controls.

Table 2 shows results from two probit regression models, with *Cooperate* choice as the dependent variable. Model 1 has a minimal set of explanatory variables, while Model 2 also includes additional *a priori* variables of interest including PANSS Total (as the primary clinical variable of interest) and IQ (as a likely confounding factor, important to account for). Both models include variables for computer strategy and round number to control for learning. The results are shown as marginal effects (MEs). The ME of an independent variable indicates the change in probability of a *Cooperate* choice through a 1-unit increase in that variable (from 0 to 1 if the variable is binary), while keeping all other independent variables fixed.

The results for Model 1 confirm the results from Figures 2-3. Controls show significant loss aversion (ME 0.099, $p = 0.0001$), while patients do not exhibit such loss aversion (ME 0.006, $p = 0.39$). Moreover, the difference in loss aversion between groups is significant ($p = 0.008$). Overall cooperation levels did not differ significantly between the two groups in both Model 1 (Group effect ME -0.016, $p = 0.87$) and Model 2 (Group effect 0.078, $p = 0.71$), and so further analyses were not conducted looking into participants' cooperative strategies.

---

significance of our results and conclusions are robust to dropping these 'non-choices' from the dataset *post hoc*, with only negligible differences.

Due to consistent evidence of cognitive impairments in schizophrenia patients (Aylward et al., 1984), we expected matching the control and patient groups for IQ would not be feasible. Thus we incorporated IQ into Model 2 as an *a priori* variable of interest and likely confounder. However, rather than IQ, the primary clinical variable of interest in Model 2 was PANSS Total, which we included due to our hypothesis that loss aversion in patients would vary with illness severity. Here, the significant marginal effect for loss aversion (ME 0.195, $p = 0.0002$) at low PANSS Total scores (25^{th} percentile) reflects the response from controls. The non-significant marginal effect for loss aversion (ME -0.073, $p = 0.91$) at high PANSS Total scores (75th percentile) reflects the lack of such response in patients. The difference between these marginal effects is significant ($p = 0.006$) and substantial: moving from the 25^{th} percentile to the 75th percentile of PANSS Total results in an approximately 30-percentage-point (0.3) decrease in loss aversion.

Additionally, Model 2 was rerun with only schizophrenia patients, showing that patients responded appropriately to differing incentives between the two computer strategies. Cooperation was significantly reduced when patients faced the pre-programmed "seesaw" strategy, against which betrayal is more advantageous compared to the "neutral" strategy ($p = 0.02$, see Supplementary Material, Table S2) – evidence that patients properly understood the task and responded appropriately to changes in incentives. This is important, because it removes a potential alternative explanation for our results, namely, that reduced loss aversion in schizophrenia was due simply to reduced understanding of the task.

Using forward stepwise regression, further probit analyses were conducted for potential confounding factors and other clinical variables: age; sex; antipsychotic dose (chlorpromazine-equivalent); income; current IQ (WASI); pre-morbid IQ (NART); executive function (Hayling and Brixton); PANSS Total, and subscales PANSS Positive, Negative and General; depression (BDI and Hamilton); anxiety (STAI); pleasure (Snaith-Hamilton); psychopathy (Levenson); propensity to trust (PTS); time since diagnosis; and time as an inpatient. Each clinical variable or confounding factor was input into separate versions of regression Model 1 (Table 3 summarizes only model fit and loss

aversion marginal effects in relation to each clinical variable or confounding factor, with Model 1 in common with Table 2; full results for the marginal effects of all model variables are shown in Supplementary Material, Table S3). The following variables were found significantly positively correlated with the *Loss Aversion* measure in these individual regressions: age (p = 0.008); WASI (p = 0.0001); NART (p = 0.0006); and antipsychotic dose (p = 0.02). The following variables were found significantly negatively correlated with the *Loss Aversion* measure in these individual regressions: PANSS Total (p = 0.00004); PANSS Positive (p = 0.00001); PANSS Negative (p = 0.02); PANSS General (p = 0.0001); BDI (p = 0.002); Hamilton (p = 0.002); and STAI (p = 0.0002). The remaining variables tested were not significant, at $p > 0.10$.

Although PANSS Positive was most significantly correlated with the *Loss Aversion* measure in this analysis, PANSS Total provided best model fit, assessed by AIC (Akaike 1974). Adding non-PANSS variables individually and separately to the regression model with PANSS Total (Model 13) did not improve model fit according to AIC.

Substituting the three PANSS subscales (Positive, Negative and General) instead of PANSS Total did not improve model fit by AIC. However, it revealed that the significant (negative) correlation with the *Loss Aversion* measure was with PANSS Positive (p = 0.04) within this model, not with PANSS Negative (p = 0.55) or PANSS General (p = 0.61). More details about this analysis for PANSS subscales are found in Supplementary Material (Table S4). Therefore, from the clinical variables, PANSS Total provided the best fit to the data. On the other hand, PANSS Positive was most significantly correlated with loss aversion and appears to be driving the significant correlation for PANSS Total, rather than PANSS Negative and PANSS General.

Given the above positive correlation demonstrated between antipsychotic dose and the *Loss Aversion* measure (p = 0.02, Table 3), antipsychotic dose would thus be an important confounding factor necessary to consider; *post hoc* inclusion of antipsychotic dose as an additional covariate within probit regression Model 2 (Table 2), alongside the covariates PANSS Total and WASI-

estimated IQ as before, did not significantly affect the overall loss aversion results, particularly in the association with PANSS Total (results outlined in Supplementary Material, Table S5).

3. Discussion

Despite this study having relatively small numbers of participants in both groups, it does raise some interesting and fairly strong results. We acknowledge that we cannot rule out all possible alternative explanations for our results. For example, if individuals have diminishing returns to money payments and a constant benefit from following a social norm to cooperate – and if those diminishing returns are more pronounced for controls than patients – the same pattern of results we observed (reduced cooperation in the *Loss* frame by controls, but not by patients) would be expected. However, evidence does exist that perceived returns to money payments should be approximately constant over small gains and losses (Rabin and Thaler 2001), supporting our conclusions about loss aversion. This assumption of linearity was incorporated into our game theoretic model for loss aversion outlined in Supplementary Material (Brain Research online). This same limitation is shared even by some of the important studies on healthy loss aversion in the literature, such as Tom et al. (2007) and Canessa et al. (2012), which explicitly assume linear functions of the utility curve. At least one study, Sokol-Hessner et al. (2013), addresses this issue with lotteries involving gains and losses and with lotteries involving gains only, forgoing the need to assume linearity in the utility function, but a study of this type might prove very challenging to implement in experiments with schizophrenia patients, due to the large number of distinct decisions from each subject that is required to estimate a complete utility-of-money function. Regarding the previous studies on schizophrenia in this literature, these can alternatively be explained by any of the other causes of the endowment effect (Tremeau et al., 2008), the reflection effect (the first task from Brown et al., 2013), and diminishing returns to money (the second task from Brown et al., 2013). However, though each individual study on its own cannot be conclusive regarding reduced loss aversion in schizophrenia, the three studies together (including ours) – considered as a set of independent quasi-replications – provide compelling evidence, since there is little overlap across the three in what alternative explanations exist.

Compared to previous research focusing on loss aversion in schizophrenia (Brown et al., 2013; Trémeau et al., 2008), we used a different approach for studying loss aversion, involving two versions of the Iterated Prisoners' Dilemma differing only in subtracting a constant from all payoffs, similar to previous studies (Cachon and Camerer 1996; Feltovich 2011; Feltovich et al., 2012; Rydval and Ortmann 2005). In our study, the control group showed reduced cooperation in the version where losses are possible, compared to the version where only gains are possible, indicative of loss aversion, while this was not observed in schizophrenia patients. At first glance, this appears counterintuitive, as it implies that a psychiatric disorder produces more "rational" behavior. Subtracting a constant from all payoffs should not make any difference to decision-making, since the rank ordering of outcomes and Nash equilibria remain unaffected (Feltovich 2011; Feltovich et al., 2012). One explanation is that "irrational" loss aversion in the healthy controls is an adaptive response generally conducive to individual survival, since in times of plenty a betrayal or poor gamble will result in no worse than a "smaller dinner", while in times of scarcity a similar bad outcome could result in starvation.

This adaptive response was observed reduced in the schizophrenia group, with extent of loss aversion negatively correlated with illness severity, particularly measured by the PANSS Total and the underlying PANSS Positive. Our results implicate other symptom measures to a lesser extent, e.g. for depression and anxiety. This is unsurprising given comorbidities common in schizophrenia, which is why we included these measures. However, PANSS Total already captures depression and anxiety symptoms, as well as psychotic symptoms, so significant collinearity in these measurements would be expected. Indeed, recent research has found *increased* loss aversion in depression (Pammi et al., 2015), supporting the idea that the negative correlation seen here between loss aversion and depressive symptoms in schizophrenia is related to collinearity with PANSS Total, rather than depression per se. PANSS Total is the most representative of schizophrenia illness severity within our battery of clinical measures, and as per our hypothesis, we found PANSS most significantly

correlated with loss aversion in schizophrenia, although levels of anxiety and depression were also relatively high in correlation.

Indeed, the highly significant correlation seen between loss aversion and the underlying PANSS Positive would suggest that dopamine dysfunction may well be involved in our finding of reduced loss aversion in schizophrenia. This is supported by the dopamine hypothesis of schizophrenia, and salience theory of psychosis related to aberrant dopamine firing (Kapur 2003; Gradin et al., 2011), as well research findings that show dopamine brain regions and projections are involved in healthy loss aversion (Canessa et al., 2013; Sokol-Hessner et al., 2013; Tom et al., 2007) and involved in depressive loss aversion (Pammi et al., 2015), with the involvement of the dopamine system particularly discussed by Tom et al. (2007) and Pammi et al. (2015) in modulating rewarding and aversive stimuli. Genetic research also provides evidence for dopamine function in loss aversion (Voigt et al., 2015). However, the involvement of the norepinephrine system may also help explain our results, given the implications of this neurotransmitter system in loss aversion including in salience processing (Takahashi et al., 2013) and given our own findings that anxiety has a relatively high correlation with loss aversion (although not as highly correlated as PANSS Positive). Trémeau et al. (2008) already posited aberrant salience as a possible cause for diminished loss aversion in schizophrenia. Abnormal salience in response to gains and losses provides a plausible explanation for lack of loss aversion in schizophrenia, since aberrant valuations given to rewarding and aversive stimuli would lead to disruptions in the weighting given to gains and losses, necessary for loss aversion. Abnormal loss aversion has already been found in depression involving dopaminergic areas (Pammi et al., 2015); given that aberrant salience is implicated in depression (Soskin et al., 2013), disrupted loss aversion may very well be relevant to other mental disorders implicated with aberrant salience, such as substance addictions (Redish 2004) and also schizophrenia. Further research would elucidate this, particularly neuroimaging studies examining neural substrates. Specifically, functional MRI can be used to measure phasic (short timescale) dopamine responses, to model abnormalities in prediction error signals and salience (Berridge and Robinson 1998; Corlett et al., 2016; Kumar et al.,

2007; McClure et al., 2003; Montague et al., 1996; Murray et al., 2008; Pessiglione et al., 2006; Schultz 1998), and thus to investigate the role of these factors in the neural basis for loss aversion. Based on previously mentioned research, we would expect such functional abnormalities in schizophrenia particularly within the dopaminergic-rich brain areas and their projection regions.

4. Experimental Procedure

4.1 Recruitment

The study was approved by the local Research Ethics Committee and written informed consent obtained from all participants. Twenty patients with schizophrenia and sixteen healthy controls matched for age and sex were recruited from both inpatient and outpatients settings at Royal Cornhill Hospital, Aberdeen. The inclusion criteria were: aged 18−65; diagnosis of paranoid schizophrenia confirmed by two clinicians, according to DSM-IV (American Psychiatric Association 2000), for patients; no history of mental illness for healthy controls. The exclusion criteria were: serious medical condition; organic brain disorder; learning disability; current IQ < 70; history of alcohol or substance misuse in the last six months; past substance or alcohol dependence; other major mental illness; history of head injury; lack of capacity to consent; poor understanding of the game paradigm.

4.2 Clinical Interview

The following clinical rating scales were administered and used to quantify symptoms and other clinical factors, and to identify potential confounding factors: Positive And Negative Symptom Scale (PANSS) (Kay et al., 1987); Beck Depression Inventory (BDI) – II (Beck et al., 1996); Hamilton Rating Scale for Depression (Hamilton 1960); Snaith-Hamilton Pleasure Scale (Snaith et al., 1995); State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1970); Propensity to Trust Survey (PTS) (Evans and Revelle 2008); Wechsler Abbreviated Scale of Intelligence (WASI), two-subtest form, for current IQ (Wechsler 1999); National Adult Reading Test (NART) for 'pre-morbid' IQ (Nelson and Willison 1991); Hayling and Brixton tests for executive function (Burgess and Shallice 1997); and Levenson's Self-Report Psychopathy Scale (Levenson et al., 1995). The following information was also obtained: age; sex; income; employment status; antipsychotic dose converted to chlorpromazine-equivalent dose in accordance with published conversion values for

different antipsychotics (Gardner et al., 2010); time since diagnosis; and time spent as an inpatient. This demographic and clinical information is summarized in Table 1.

4.3 Game Paradigm

Participants played 100 rounds of the Iterated Prisoners' Dilemma, 50 in each of the *Gain* and *Loss* frames. The computer player followed pre-programmed strategies. In each block of 50 rounds, the computer opponent played 25 using a "neutral" strategy, against which neither *Cooperate* nor *Betray* would significantly outperform the other for the participant. In the other 25 rounds of each block, the computer used a "seesaw" strategy, against which *Betray* would significantly out-perform *Cooperate* for the participant – see Supplementary Material (Methods and Materials, Strategies and Figures S1, S2). Counterbalancing was used to control for order and learning effects. Cumulative scores and payoffs for the computer player were not displayed, to minimize the performance of the computer player as a point of reference. For further details, see Supplementary Material (Methods and Materials, Game Paradigm).

Consistent with standard experimental-economics methodology, participants were alerted before playing to the fact that the other player was a computer (although not informed of the specific computer strategies used); and we used real monetary incentives. Participants received a variable payment that was proportional to their final scores, so that total earnings ranged between 5-15 GBP. Performance-based payments have long been considered important (Siegel and Goldstein 1959): these encourage serious decision-making, and reduce the likelihood of increased variance in the data (Camerer and Hogarth 1999) and hypothetical bias (Cummings et al., 1997).

4.4 Analyses

All statistical testing was performed using Stata 12.1 (StataCorp). Results were considered statistically significant at ≤ 0.05. Continuous data were assessed for normality using Q-Q plots and the Skewness-Kurtosis normality test (Stata default settings) (D'Agostino et al., 1990; Royston 1991).

If data departed significantly from normal distribution, then non-parametric tests were used, otherwise parametric. Alternative to the Wilcoxon–Mann–Whitney test, the Fligner-Policello Robust Rank Order test was used due to evidence of preferability (Fligner and Policello II 1981; Feltovich 2003). To analyze cooperative choices over time, a random-effects probit regression was employed, for 'longitudinal/panel' data involving 'binary outcomes' (Stata default settings).

Loss aversion as a concept is one-sided by definition, and thus for statistical testing of loss aversion, one-tailed testing was performed. Otherwise, two-tailed testing was used. Between and within groups analyses were conducted to test the hypotheses. Further analyses were done for possible confounders, with best model fit assessed using the Akaike Information Criterion (AIC) (Akaike 1974).

Acknowledgments

This study was supported by funding from the Millar-McKenzie Trust and the BMA Margaret Temple grant. JDS has received research funding via an honorarium associated with a lecture from Wyeth. The authors have no conflicts of interest to declare.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans Auto Control. 19, 716-723.

American Psychiatric Association, 2000, Diagnostic and Statistical Manual of Mental Disorders, DSM-IV-TR. American Psychiatric Publishing, Arlington, Virginia.

Aylward, E., Walker, E., Bettes, B., 1984. Intelligence in schizophrenia: Meta-analysis of the research. Schizophrenia Bulletin. 10, 430-459.

Beck, A.T., Steer, R.A., Ball, R., Ranieri, W., 1996. Comparison of beck depression inventories -IA and -II in psychiatric outpatients. Journal of Personality Assessment. 67, 588-597.

Berridge, K.C., Robinson, T.E., 1998. What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? 28, 309-369.

Brown, J.K., Waltz, J.A., Strauss, G.P., McMahon, R.P., Frank, M.J., Gold, J.M., 2013. Hypothetical decision making in schizophrenia: The role of expected value computation and "irrational" biases. Psychiatry Research. 209, 142-149.

Burgess, P.W., Shallice, T., 1997, The Hayling and Brixton Tests. Thames Valley Test Company, Bury St. Edmunds.

Cachon, G.P., Camerer, C.F., 1996. Loss-avoidance and forward induction in experimental coordination games. The Quarterly Journal of Economics. 111, 165-194.

Camerer, C.F., Hogarth, R.M., 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. Journal of Risk and Uncertainty. 19, 7-42.

Canessa, N., Crespi, C., Motterlini, M., Baud-Bovy, G., Chierchia, G., Pantaleo, G., Tettamanti, M., Cappa, S.F., 2013. The functional and structural neural basis of individual differences in loss aversion. The Journal of Neuroscience. 33, 14307-14317.

Chen, M.K., Lakshminarayanan, V., Santos, L.R., 2006. How basic are behavioral biases? Evidence from capuchin monkey trading behavior. Journal of Political Economy. 114, 517-537.

Corlett, P.R., Honey, G.D., Fletcher, P.C., 2016. Prediction error, ketamine and psychosis: An updated model. Journal of Psychopharmacology. 30(11), 1145-1155.

Cummings, R.G., Elliott, S., Harrison, G.W., Murphy, J., 1997. Are hypothetical referenda incentive compatible? Journal of Political Economy. 105, 609-621.

D'Agostino, R.B., Belanger, A., D'Agostino Jr., R.B., 1990. A suggestion for using powerful and informative tests of normality. The American Statistician. 44, 316-321.

Engelmann, D., Hollard, G., 2010. Reconsidering the effect of market experience on the "Endowment effect". Econometrica. 78, 2005-2019.

Evans, A.M., Revelle, W., 2008. Survey and behavioral measurements of interpersonal trust. Journal of Research in Personality. 42, 1585-1593.

Feltovich, N., Iwasaki, A., Oda, S.H., 2012. Payoff levels, loss avoidance, and equilibrium selection in games with multiple equilibria: An experimental study. Economic Inquiry. 50, 932-952.

Feltovich, N., 2011. The effect of subtracting a constant from all payoffs in a hawk-dove game: Experimental evidence of loss aversion in strategic behavior. Southern Economic Journal. 77, 814-826.

Feltovich, N., 2003. Nonparametric tests of differences in medians: Comparison of the Wilcoxon–Mann–Whitney and robust rank-order tests. Experimental Economics. 6, 273-297.

Fligner, M.A., Policello II, G.E., 1981. Robust rank procedures for the behrens-fisher problem. Journal of the American Statistical Association. 76, 162-168.

Gardner, D.M., Murphy, A.L., O'Donnell, H., Centorrino, F., Baldessarini, R.J., 2010. International consensus study of antipsychotic dosing. American Journal of Psychiatry. 167, 686-693.

Gradin, V.B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., Hall, J., Steele, J.D., 2011. Expected value and prediction error abnormalities in depression and schizophrenia. Brain. 134, 1751-1764.

Hamilton, M., 1960. A rating scale for depression. Journal of Neurology, Neurosurgery & Psychiatry. 23, 56-62.

Harbaugh, W.T., Krause, K., Vesterlund, L., 2001. Are adults better behaved than children? Age, experience, and the endowment effect. Economics Letters. 70, 175-181.

Holt, D.J., Coombs III, G., Zeidan, M.A., Goff, D.C., Milad, M.R., 2012. Failure of neural responses to safety cues in schizophrenia. Arch Gen Psychiatry. 69, 893-903.

Jensen, J., Willeit, M., Zipursky, R.B., Savina, I., Smith, A.J., Menon, M., Crawley, A.P., Kapur, S., 2008. The formation of abnormal associations in schizophrenia: Neural and behavioral evidence. Neuropsychopharmacology. 33, 473-479.

Kahneman, D., Tversky, A., 1984. Choices, values, and frames. American Psychologist. 39, 341-350.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica. 47, 263-291.

Kapur, S., 2003. Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. American Journal of Psychiatry. 160, 13-23.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia Bulletin. 13, 261-276.

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., Steele, J.D., 2007. Abnormal temporal difference reward-learning signals in major depression. Brain. 131, 2084-2093.

Levenson, M.R., Kiehl, K.A., Fitzpatrick, C.M., 1995. Assessing psychopathic attributes in a noninstitutionalized population. Journal of Personality and Social Psychology. 68, 151-158.

McClure, S.M., Daw, N.D., Montague, P.R., 2003. A computational substrate for incentive salience. Trends in Neurosciences. 26, 423-428.

Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive hebbian learning. The Journal of Neuroscience. 76, 1936-1947.

Murray, G.K., Corlett, P.R., Clark, L., Pessiglione, M., Blackwell, A.D., Honey, G., Jones, P.B., Bullmore, E.T., Robbins, T.W., Fletcher, P.C., 2008. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Molecular Psychiatry. 13, 267-276.

Nelson, H.E., Willison, J.R., 1991, The Revised National Adult Reading Test - Test Manual. NFER-Nelson, Windsor.

Pammi, V.S.C., Rajesh, P.P.G., Kesavadas, C., Mary, P.R., Seema, S., Radhakrishnan, A., Sitaram, R., 2015. Neural loss aversion differences between depression patients and healthy individuals: A functional MRI investigation. The Neuroradiology Journal. 28, 97-105.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D., 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature. 442, 1042-1045.

Plott, C.R., Zeiler, K., 2005. The willingness to Pay–Willingness to accept gap, the "Endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. The American Economic Review. 95, 530-545.

Rabin, M., Thaler, R.H., 2001. Anomalies: Risk aversion. Journal of Economic Perspectives. 15(1), 219-232.

Redish, A.D., 2004. Addiction as a computational process gone awry. Science. 306, 1944-1947.

Royston, P., 1991. Sg3.5: Comment on sg3.4 and an improved D'agostino test. Stata Technical Bulletin. 3, 23-24.

Rydval, O., Ortmann, A., 2005. Loss avoidance as selection principle: Evidence from simple stag-hunt games. Economics Letters. 88, 101-107.

Schultz, W., 1998. Predictive reward signal of dopamine neurons. Journal of Neurophysiology. 80, 1-27.

Siegel, S., Goldstein, D.A., 1959. Decision-making behavior in a two-choice uncertain outcome situation. Journal of Experimental Psychology. 57, 37-42.

Snaith, R.P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., Trigwell, P., 1995. A scale for the assessment of hedonic tone the snaith-hamilton pleasure scale. British Journal of Psychiatry. 167, 99-103.

Sokol-Hessner, P., Camerer, C.F., Phelps, E.A., 2013. Emotion regulation reduces loss aversion and decreases amygdala responses to losses. Social Cognitive and Affective Neuroscience. 8, 341-350.

Soskin, D.P., Holta, D.J., Sacco, G.R., Fava, M., 2013. Incentive salience: Novel treatment strategies for major depression. CNS Spectrums. 18, 307-314.

Spielberger, C.D., Gorsuch, R.L., Lushene, R.E., 1970, Manual for the State-Trait Anxiety Inventory. Consulting Psychologists Press, Palo Alto, CA.

Takahashi, H., Fujie, S., Camerer, C., Arakawa, R., Takano, H., Kodaka, F., Matsui, H., Ideno, T., Okubo, S., Takemura, K., Yamada, M., Eguchi, Y., Murai, T., Okubo, Y., Kato, M., Ito, H., Suhara, T., 2013. Norepinephrine in the brain is associated with aversion to financial loss. Molecular Psychiatry. 18, 3-4.

Tom, S.M., Fox, C.R., Trepel, C., Poldrack, R.A., 2007. The neural basis of loss aversion in decision-making under risk. Science. 315, 515-518.

Trémeau, F., Brady, M., Saccente, E., Moreno, A., Epstein, H., Citrome, L., Malaspina, D., Javitt, D., 2008. Loss aversion in schizophrenia. Schizophrenia Research. 103, 121-128.

Voigt, G., Montag, C., Markett, S., Reuter, M., 2015. On the genetics of loss aversion: An interaction effect of BDNF Val66Met and DRD2/ANKK1 Taq1a. Behavioral Neuroscience. 129(6), 801-811.

Wechsler, D., 1999, Wechsler Abbreviated Scale of Intelligence. 1st Edition, The Psychological Corporation, San Antonio, TX.

Figure legends

Figure 1: Two versions of the Iterated Prisoners' Dilemma.

Figure 2: Aggregate frequency of *Cooperate* choices, all rounds, pooled opponent strategies. Error bars are mean + one standard deviation. The *p*-values are from Wilcoxon signed-ranks tests for paired samples (one-sided rejection regions).

Figure 3: Loss aversion (*Cooperate* frequency in *Gain* frame – *Loss* frame, all rounds, pooled opponent strategies). Error bars are mean + one standard deviation. The *p*-value is from a robust rank-order test (two-sided rejection region).

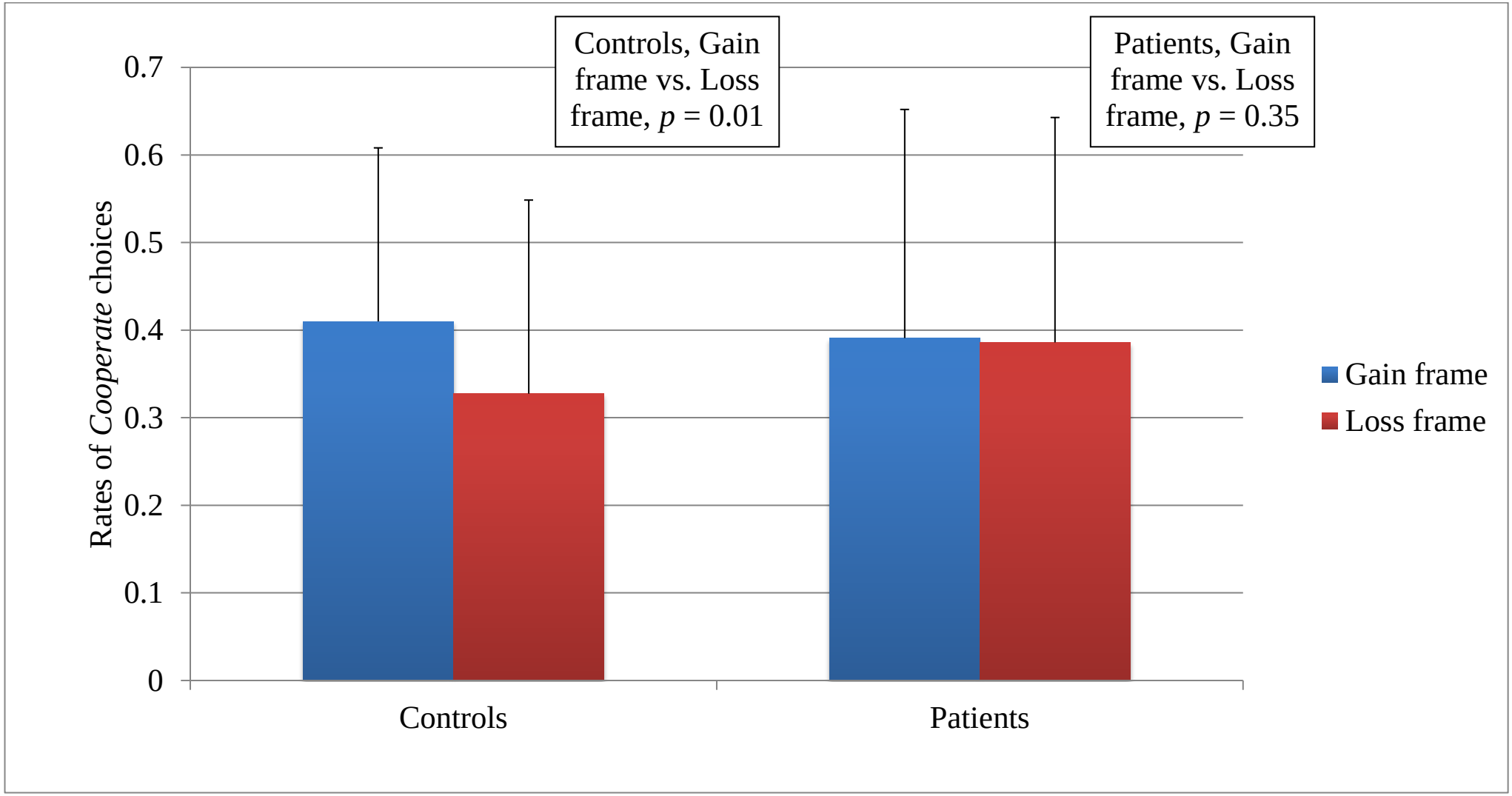
Figure 4: Scatter-plot of loss aversion and PANSS Total score in patient group, with ordinary least squares regression results.

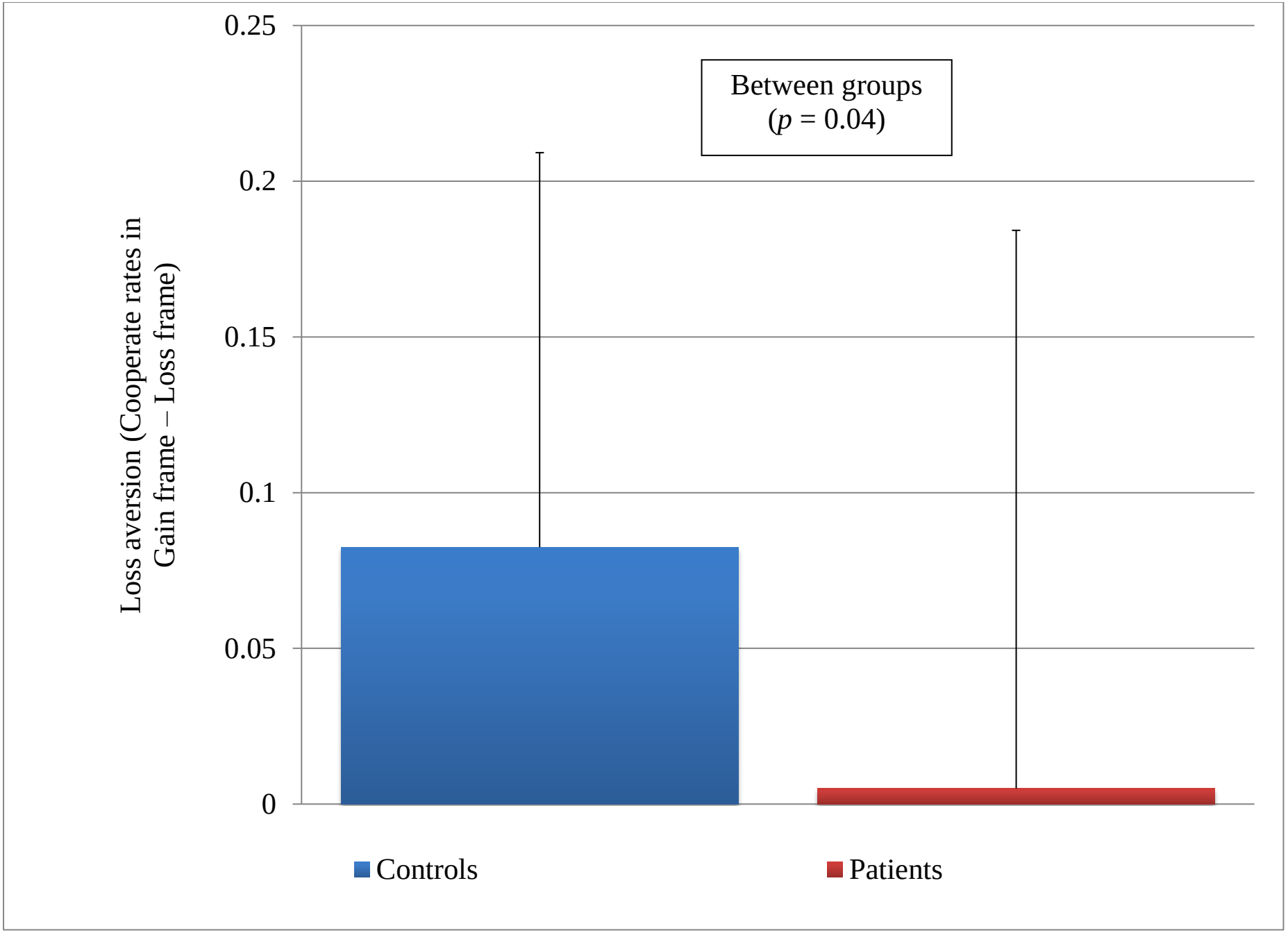
Gain frame

		Computer	
		Cooperate	Betray
Decision-maker	Cooperate	6, 6	2, 8
	Betray	8, 2	4, 4

Loss frame

		Computer	
		Cooperate	Betray
Decision-maker	Cooperate	1, 1	-3, 3
	Betray	3, -3	-1, -1


Controls, Gain frame vs. Loss frame, p = 0.01
Patients, Gain frame vs. Loss frame, p = 0.35
Rates of Cooperate choices
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Controls
Patients
Gain frame
Loss frame


0.25
0.2
0.15
0.1
0.05
0
Loss aversion (Cooperate rates in Gain frame – Loss frame)
Between groups
(p = 0.04)
Controls
Patients

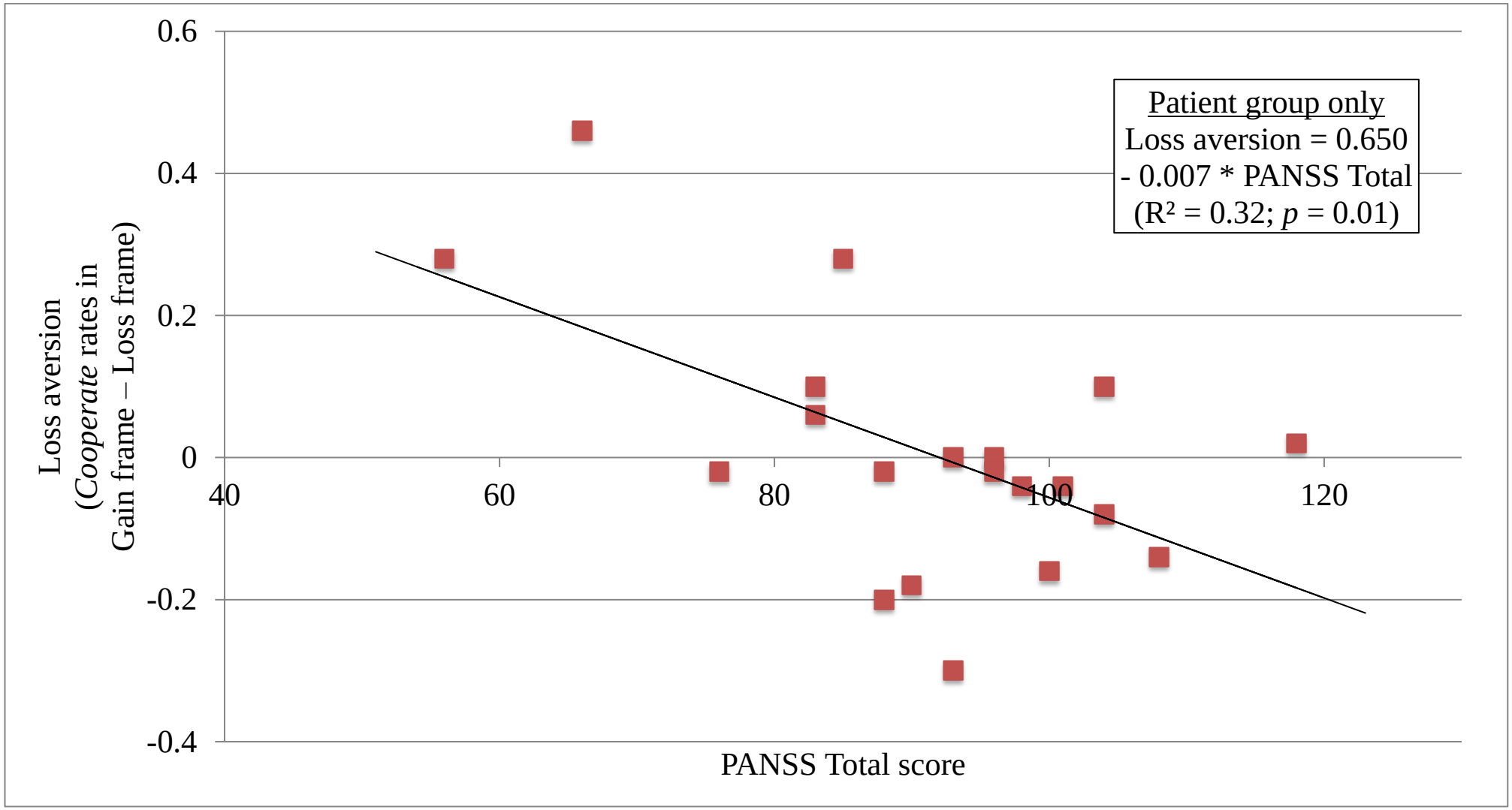

Patient group only
Loss aversion = 0.650
- 0.007 * PANSS Total
(R² = 0.32; p = 0.01)
Loss aversion
(Cooperate rates in
Gain frame – Loss frame)
0.6
0.4
0.2
0
-0.2
-0.4
40
60
80
100
120
PANSS Total score

Table 1: Demographics and clinical details

STT = Student's t-test RROT = Robust Rank-Order test TP = Two-sample test of proportions
SD = standard deviation IQR = interquartile range
*, **, *** indicate significance at ≤ 0.10, ≤ 0.05 and ≤ 0.01 levels.

	Patients (n = 20)	Controls (n = 16)	Significance
Mean age / years	44 .0 (SD 12.3)	42.9 (SD 10.5)	$p = 0.77$ (STT)
Sex	19 males (95%)	15 males (94%)	$p = 0.87$ (TP)
Employed	n = 0	n = 16	$p < 0.0001$ (TP) ***
Median monthly income after tax in GBP	525.00 (IQR 320.00)	1827.50 (IQR 500.00)	$p < 0.0001$ (RROT) ***
Mean antipsychotic dose (chlorpromazine-equivalent)	600.1 (SD 412.6)	0 (SD 0)	$p < 0.0001$ (RROT) ***
Median length of time since diagnosis (years)	19.3 (IQR 23.6)	0 (IQR 0)	$p < 0.0001$ (RROT) ***
Median total length of inpatient stays (months)	6.4 (IQR 10.0)	0 (IQR 0)	$p < 0.0001$ (RROT) ***
Mean NART IQ	110.2 (SD 9.0)	115.1 (SD 8.5)	$p = 0.10$ (STT) *
Mean WASI IQ	93.1 (SD 18.5)	113.3 (SD 9.1)	$p = 0.0002$ (STT) ***
Median Hayling score	4 (IQR 4)	5 (IQR 1)	$p = 0.01$ (RROT) ***
Median Brixton score	3.5 (IQR 5)	6 (IQR 1)	$p = 0.04$ (RROT) **
Median Positive PANSS score	22.5 (IQR 6.5)	7 (IQR 0.5)	$p < 0.0001$ (RROT) ***
Median Negative PANSS score	25.5 (IQR 6)	8 (IQR 3)	$p < 0.0001$ (RROT) ***
Median General PANSS score	44.5 (IQR 7.5)	18.5 (IQR 4.5)	$p < 0.0001$ (RROT) ***
Median Total PANSS score	93 (IQR 16.5)	34.5 (IQR 7)	$p < 0.0001$ (RROT) ***
Mean BDI (depression) score	13.5 (SD 9.1)	1.4 (SD 1.9)	$p < 0.0001$ (STT) ***
Mean Hamilton (depression) score	7.9 (SD 5.6)	0.3 (SD 0.5)	$p < 0.0001$ (STT) ***
Median STAI (anxiety) score	42 (IQR 19)	23.5 (IQR 6)	$p < 0.0001$ (RROT) ***
Median Snaith-Hamilton Pleasure Scale score	1.5 (IQR 2.5)	0 (IQR 1)	$p = 0.26$ (RROT)
Mean Levenson's Psychopathy score	20.1 (SD 10.7)	5.8 (SD 6.2)	$p < 0.0001$ (STT) ***
Median Propensity to Trust Survey score	34.5 (IQR 15.5)	55 (IQR 10)	$p < 0.0001$ (RROT) ***
Mean game score (cumulative payoffs)	272.0 (SD 51.6)	272.9 (SD 43.6)	$p = 0.95$

Note: means, standard deviations and Student's t-tests are presented for data with normal distributions; otherwise medians, interquartile ranges and non-parametric tests are presented.

Table 2: Probit marginal effects, standard errors (SEs) and *p* values.
*, **, *** indicate significance at ≤ 0.10, ≤ 0.05 and ≤ 0.01 levels; *p* values are from χ^2 tests.

Dependent variable: cooperate choice	**Model 1**	**Model 2**
Group effect [Binary variable: Patient Baseline: Control (Patient=0)]	-0.016 (SE 0.094) ***p = 0.87***	0.078 (SE 0.210) ***p = 0.71***
Loss (Aversion) effect [Binary variable: Gain frame Baseline: Loss frame (Gain = 0)]	0.047 (SE 0.017) ***p = 0.003*** ***	0.043 (SE 0.018) ***p = 0.007*** ***
Loss (Aversion) effect …		
… Controls (Patient = 0)	0.099 (SE 0.026) ***p = 0.0001*** ***	-0.086 (SE 0.063) ***p = 0.91***
… Patients (Patient = 1)	0.006 (SE 0.023) ***p = 0.39***	0 .154 (SE 0.044) ***p = 0.0002*** ***
Significance of difference	***p = 0.008*** ***	***p = 0.008*** ***
Loss (Aversion) effect …		
… PANSS Total = 35 (25th percentile)		0.195 (SE 0.054) ***p = 0.0002*** ***
… PANSS Total = 94.5 (75th percentile)		-0.073 (SE 0.053) ***p = 0.91***
Significance of difference		***p = 0.006*** ***
Strategy [Binary variable: Neutral strategy Baseline: Seesaw strategy (Neutral = 0)]	0.011 (SE 0.017) ***p = 0.52***	0.012 (SE 0.017) ***p = 0.49***
Round Number [Continuous variable: Round Number]	-0.005 (SE 0.001) $p = 6\times10^{-6}$ ***	-0.005 (SE 0.001) $p = 1\times10^{-5}$ ***
PANSS Total [Continuous variable: PANSS Total]		-0.004 (SE 0.004) ***p = 0.38***
WASI-estimated IQ [Continuous variable: WASI-estimated IQ]		-0.005 (SE 0.003) ***p = 0.12***
Observations	3600	3600
Log likelihood (AIC)	-2030.8 (4095.6)	-2008.4 (4082.8)

Note: The marginal effect for an independent variable indicates the change in probability (-1.0 to 1.0) of a cooperative choice through a 1-unit increase in that variable, while keeping all other independent variables fixed – and specifically for a binary variable this means switching from 0 (baseline) to 1 in the binary variable to indicate the change in probability.

The Loss Aversion effects – the marginal effects comparing the Gain frame to the Loss frame (baseline) – are displayed for Patients and Controls to test for Loss Aversion differences between these two groups (both Models 1 and 2), and also displayed for the 25th and 75th percentiles for PANSS Total to test for Loss Aversion differences between these two points on the PANSS scale (Model 2 only). With PANSS Total and IQ as covariates (Model 2), the Loss Aversion effects for Controls versus Patients become less important in interpretation, indeed reversed in direction, showing that patients don't have reduced loss aversion because they are called 'patients', but rather related to the severity of their illness captured in the model by the PANSS Total score. This reversal of Loss Aversion effects for Controls versus Patients does not arise from including IQ, since this reversal continues even after removing IQ completely from Model 2 (see Supplementary Material, Table S1).

Table 3: Probit models for clinical variables for forward stepwise regression
Loss aversion marginal effects (average marginal effects unless noted) and standard errors
Models all include the following variables: clinical variable (except Model 1), loss (aversion), group, strategy and rounds. Here, only marginal effects for loss (aversion) are presented in relation to clinical variables and potential confounders. Marginal effects for remaining model variables are comprehensively presented for these Models in the Supplementary Material (Table S3).
*, **, *** indicate significance at ≤ 0.10, ≤ 0.05 and ≤ 0.01 levels; *p*-values are from χ^2 tests.

Model No. [Clinical Variable]	**Log likelihood (AIC)**	**Marginal effects (with Standard Errors, SEs) and p-values**		
		Loss (Aversion) effect [Gain]		
		At 25th percentile [Clinical Variable]	**At 75th percentile [Clinical Variable]**	**Difference between 25th & 75th percentiles**
Model 1 [No clinical variable]	-2030.8 (4095.6)	No clinical variable		
Model 2	Not presented here, as not part of the stepwise regression, rather presented in Table 2~~1~~.			
Model 3 [Age]	-2021.8 (4093.7)	0.020 (SE 0.020) $p = 0.16$	0.084 (SE 0.022) $p = 0.0001$ ***	$p = 0.008$ ***
Model 4 [Income]	-2028.3 (4106.5)	0.072 (SE 0.026) $p = 0.003$ ***	0.020 (SE 0.026) $p = 0.22$	$p = 0.17$
Model 5 [Antipsychotic (AP) dose]	-2026.5 (4103.1)	-0.002 (SE 0.028) $p = 0.53$	0.080 (SE 0.024) $p = 0.0005$ ***	$p = 0.02$
Model 6 [WASI IQ]	-2016.8 (4083.6)	0.005 (SE 0.020) $p = 0.41$	0.113 (SE 0.024) $p < 0.0001$ ***	$p = 0.0001$ ***
Model 7 [NART IQ]	-2019.2 (4088.3)	0.007 (SE 0.021) $p = 0.36$	0.085 (SE 0.021) $p < 0.0001$ ***	$p = 0.0006$ ***
Model 8 [Hayling]	-2026.1 (4102.1)	0.046 (SE 0.021) $p = 0.01$ ***	0.050 (SE 0.027) $p = 0.03$ **	$p = 0.91$
Model 9 [Brixton]	-2026.4 (4102.8)	0.048 (SE 0.028) $p = 0.04$ **	0.047 (SE 0.021) $p = 0.01$ ***	$p = 0.97$
Model 10 [PANSS Positive]	-2017.3 (4084.5)	0.240 (SE 0.046) $p < 0.0001$ ***	-0.112 (SE 0.050) $p = 0.99$	$p = 0.00001$ ***
Model 11 [PANSS Negative]	-2022.4 (4094.8)	0.149 (SE 0.049) $p = 0.001$ ***	-0.031 (SE 0.038) $p = 0.79$	$p = 0.02$ **
Model 12 [PANSS General]	-2017.6 (4085.1)	0.195 (SE 0.049) $p = 0.0001$ ***	-0.106 (SE 0.046) $p = 0.99$	$p = 0.0001$ ***
Model 13 [PANSS Total]	-2015.9 (4081.8)	0.243 (SE 0.052) $p < 0.0001$ ***	-0.127 (SE 0.058) $p = 0.99$	$p = 0.00004$ ***
Model 14 [BDI]	-2024.9 (4099.8)	0.119 (SE 0.029) $p < 0.0001$ ***	0.008 (SE 0.021) $p = 0.34$	$p = 0.002$ ***
Model 15 [Hamilton]	-2022.9 (4095.9)	0.111 (SE 0.027) $p < 0.0001$ ***	0.016 (SE 0.020) $p = 0.21$	$p = 0.002$ ***
Model 16 [STAI] (anxiety)	-2023.0 (4096.0)	0.123 (SE 0.028) $p < 0.0001$ ***	-0.009 (SE 0.023) $p = 0.66$	$p = 0.0002$ ***
Model 17 [Snaith-Hamilton Pleasure Scale]	-2026.9 (4103.8)	0.069 (SE 0.023) $p = 0.001$ ***	0.036 (SE 0.018) $p = 0.02$ **	$p = 0.12$
Model 18 [Levenson's Psychopathy scale]	-2025.5 (4101.0)	0.040 (SE 0.023) $p = 0.04$ **	0.041 (SE 0.022) $p = 0.03$ **	$p = 0.97$
Model 19 [Propensity To Trust (PTS) scale]	-2028.4 (4106.7)	0.053 (SE 0.021) $p = 0.006$ ***	0.040 (SE 0.022) $p = 0.04$ **	$p = 0.62$
Model 20 [Length of time since diagnosis, years]	-2027.7 (4105.5)	0.049 (SE 0.022) $p = 0.01$ ***	0.045 (SE 0.028) $p = 0.05$ **	$p = 0.92$
Model 21 [Total length of inpatient stays, months]	-2026.9 (4103.8)	0.047 (SE 0.018) $p = 0.005$ ***	0.048 (SE 0.018) $p = 0.003$ ***	$p = 0.92$
		Female	Male	Difference (Female vs Male)
Model 22 [Sex]	-2030.1 (4110.2)	-0.021 (SE 0.067) $p = 0.62$	0.052 (SE 0.018) $p = 0.002$ ***	$p = 0.29$

Supplementary Material

Experimental Procedure

Game paradigm

Each participant played 100 rounds of the Repeated Prisoners' Dilemma involving simultaneous choices from the human participant and a computer player. Each round was 18 seconds long and divided into two stages: the 9-second *choice* stage and the 9-second *feedback* stage (see Screenshots below). At the start of the 9-second *choice* stage, there was a 5-second timer counting down, during which the participant was asked to make a choice to cooperate or betray, pressing the left mouse button to cooperate and the right mouse button to betray. Once the mouse was pressed, the choice was presented on the screen with appropriate graphics for the remainder of the 9-second *choice* stage. If the 5-second timer counted down to zero without a button press, the participant was advised that they had not made a choice due to being too slow. Once the 9-second *choice* stage finished, the 9-second *feedback* stage then started, during which information was presented with appropriate graphics about the choices made by both players and the resultant pay-offs. With both the *choice* stage and the *feedback* stage complete, the game continued into the next round. The current round and the human player's accumulated score were presented on the screen throughout the game. The computer player's pay-offs and score were not presented on the screen at any point during the game. It was made clear to participants before the game that they would be playing a computer player.

Participants played the Repeated Prisoners' Dilemma in two blocks of 50 rounds (2 x 15 minutes), one block consisting of the 'Gain frame' of the paradigm and the other block consisting of the 'Loss frame'. In each block, the computer played 25 rounds of one pre-programmed strategy, the "neutral" strategy, and 25 rounds of another pre-programmed strategy, the "seesaw" strategy.

Computer strategies

The "neutral" strategy and the "seesaw" strategy were both pre-programmed computer strategies. The "neutral" strategy was named such because of its neutral nature, whereby whether the human participant cooperated or betrayed their resultant score was largely unaffected – cooperation did not have a significant advantage over betraying and vice versa (Figure S1). The "seesaw" strategy was

named such because of its seesawing nature, in that the computer player would tend to play in the opposite manner to the human participant, tending to betray if the human participant cooperated and vice versa (Figure S2). Specifically, the "neutral" strategy mimicked the human player's previous-round response in 40 of the 50 rounds and made the opposite choice in the other 10 rounds ('Tit for Tat' with 20 percent switches). The "seesaw" strategy made the opposite choice to the human player's previous-round response in 40 of the 50 rounds and mimicked the human player in the other 10 ('Reverse Tit for Tat' with 20 percent switches). The switches occurred in pre-determined rounds (via a randomizing algorithm) for all participants: rounds 3, 5, 11, 15, 19, 31, 37, 44, 46 and 48 of the 'Gain frame'; and 4, 6, 9, 13, 16, 30, 31, 37, 44 and 47 of the 'Loss frame'. The switches were added to reduce the computer's predictability, in order to avoid possible IQ-linked effects from participants working out the computer strategy.

Counterbalancing

A counterbalancing procedure was followed, involving four permutations, due to there being two versions of the paradigm (the 'Gain frame' and 'Loss frame') and two pre-programmed computer strategies repeated in both versions of paradigm (the "neutral" strategy and the "seesaw" strategy). This counterbalancing was used to control for practice effects and learning that could occur as an individual progressed through the two blocks of 50 rounds. Individual participants were recruited into one of the four permutations (Figure S3), distributing equal numbers of participants into each permutation for each group. Therefore, for the 20 patients participating in the study, there were 5 patients in each permutation, and for the 16 healthy controls, there were 4 controls in each permutation.

Monetary incentives, instructions and other aspects

Participants were informed that if they completed the entire 100 rounds, they would be given money dependent on how well they did in the task, as is conventional in neuroeconomics experiments. Their total raw score was multiplied by 4 to give the number of pence (0.01 GBP) they were to receive. Participants received between £5 and £15 based on this calculation, with £5 being considered the minimum amount given for participation.

Participants were not provided information about the specifics of the computer strategies, but were informed about the mechanics of the game paradigm using a standard set of written instructions. To

ensure understanding, participants were also allowed to play a 5-round practice session of the paradigm, involving a completely random computer strategy.

The Repeated Prisoners' Dilemma game paradigm described was programmed on 'Presentation' software (© 2011 Neurobehavioral Systems, Inc., Albany, California, US).

Behavioural outcome measures

The following behavioural outcome measures were used:

(1) cooperative choice in a particular round (binary outcomes);

(2) cooperation frequency over a block of rounds – e.g. all rounds of the 'Gain frame' or 'Loss frame' (continuous outcomes);

(3) score (continuous outcomes).

Raw scores under the 'Gain frame' and 'Loss frame' were made comparable using a conversion calculation:

$$Converted\ Score = \frac{(Raw\ Score - Minimum\ Score)}{(Maximum\ Score - Minimum\ Score)} \times 100\% = Performance$$

The minimum score was taken to be the lowest possible score achievable under that frame (lowest pay-off × number of rounds), and the maximum score was taken to be the highest possible score achievable under that frame (highest pay-off × number of rounds). The above equation meant that the lowest possible converted score would be zero and the highest possible converted score would be 1.0 (or 100%). The converted score was therefore a comparative measure of performance in the task. The term *performance* is used to denote the converted score.

Note about hypotheses

Along with our hypotheses about loss aversion, we had originally hypothesised that patients with schizophrenia recruited via monetary incentives might show reduced overall levels of cooperative behaviour reflective of illness-related paranoia. However, due to restrictions placed by the Research Ethics Committee about only mentioning monetary incentives *after* consent is given, leading to likely exacerbations in selection bias, we later made no predictions about the direction of altered

cooperative behaviour, whether reduced or increased in patients compared to controls, when looking at the paradigm as a whole and without differentiating between the *Gain* and *Loss* frames.

Results

Figure S1: Performance versus Cooperation Rate for the "neutral" strategy

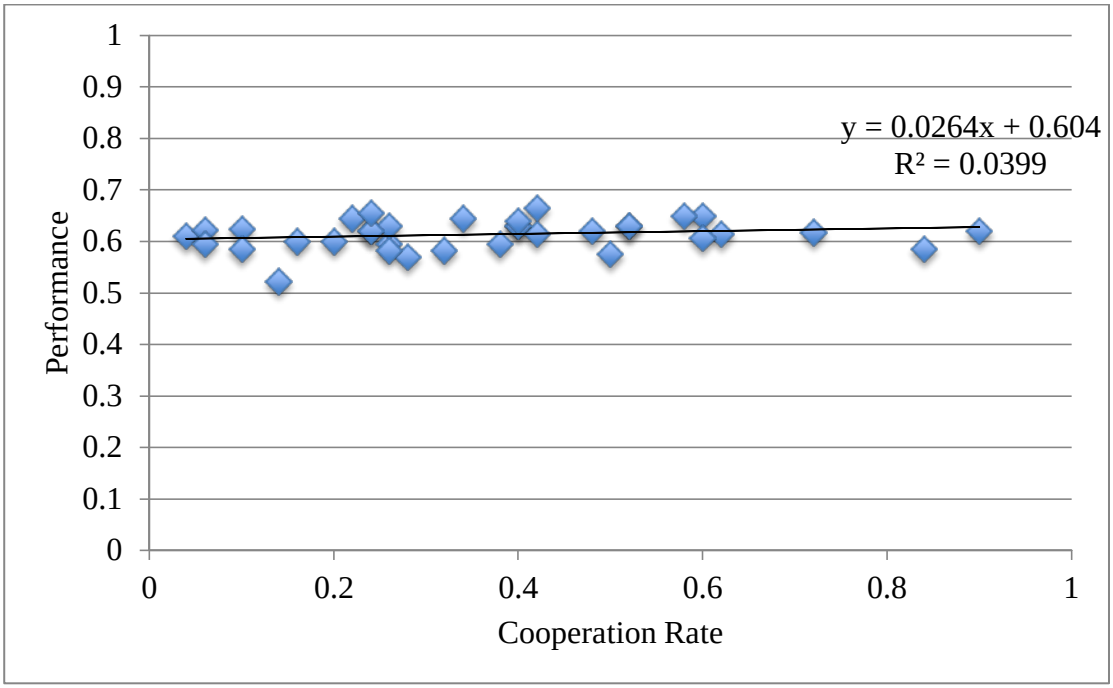


Figure S2: Performance versus Cooperation Rate for the "seesaw" strategy

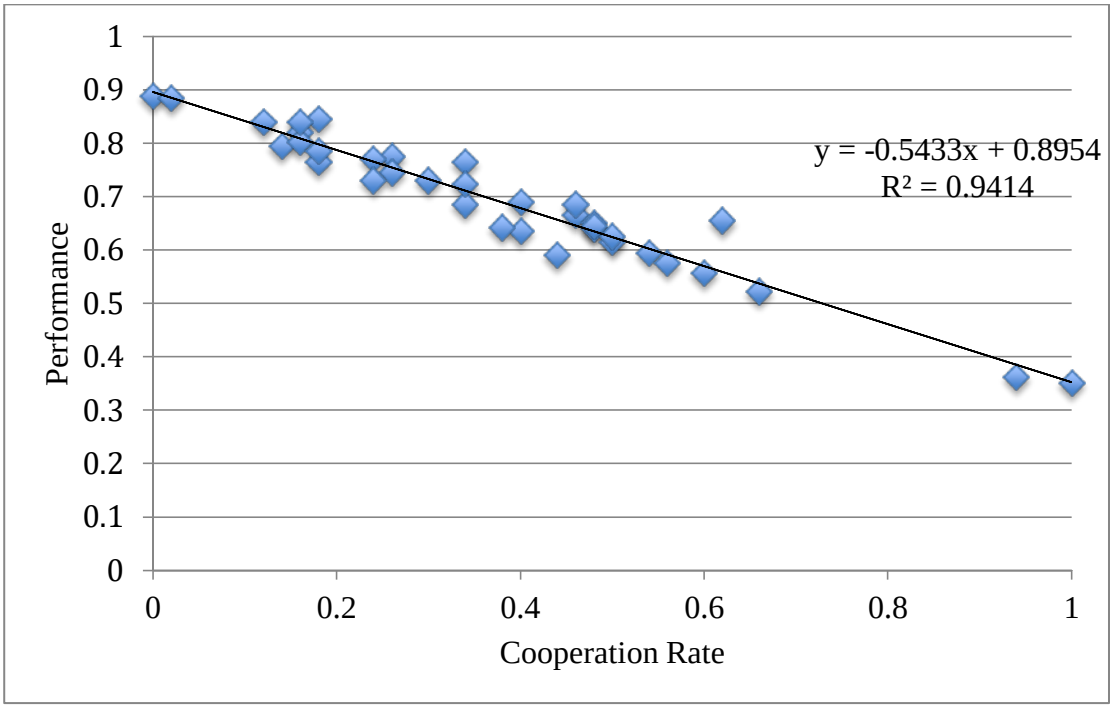


Table S1: Probit marginal effects, standard errors (SEs) and *p* values.
*, **, *** indicate significance at ≤ 0.10, ≤ 0.05 and ≤ 0.01 levels; *p* values are from χ^2 tests.

Dependent variable: cooperate choice	
Group effect [Binary variable: Patient Baseline: Control (Patient=0)]	0.052 (SE 0.221) ***p = 0.81***
Loss (Aversion) effect [Binary variable: Gain frame Baseline: Loss frame (Gain = 0)]	0.047 (SE 0.018) ***p = 0.004*** ***
Loss (Aversion) effect …	
… Controls (Patient = 0)	-0.091 (SE 0.066) ***p = 0.91***
… Patients (Patient = 1)	0.157 (SE 0.045) ***p = 0.0002*** ***
Significance of difference	***p = 0.006*** ***
Loss (Aversion) effect …	
… PANSS Total = 35 (25^{th} percentile)	0.243 (SE 0.052) $p = 1 \times 10^{-6}$ ***
… PANSS Total = 94.5 (75^{th} percentile)	-0.127 (SE 0.058) ***p = 0.99***
Significance of difference	***p < 0.0001*** ***
Strategy [Binary variable: Neutral strategy Baseline: Seesaw strategy (Neutral = 0)]	0.012 (SE 0.017) ***p = 0.50***
Round Number [Continuous variable: Round Number]	-0.005 (SE 0.001) $p = 5 \times 10^{-6}$ ***
PANSS Total [Continuous variable: PANSS Total]	-0.001 (SE 0.004) ***p = 0.72***
Observations	3600
Log likelihood (AIC)	-2015.9 (4081.8)

Table S2: Probit regression analysis for patients only
Probit marginal effects (average marginal effects unless noted) and standard errors
*, **, *** indicate significance at ≤ 0.10, ≤ 0.05 and ≤ 0.01 levels; *p*-values are from χ^2 tests.

Dependent variable: cooperate choice	
Loss (Aversion) effect [Binary variable: Gain frame Baseline: Loss frame (Gain = 0)]	0.0006 (SE 0.0244) *p* = 0.49
Strategy [Binary variable: Neutral strategy Baseline: Seesaw strategy (Neutral = 0)]	0.056 (SE 0.023) *p* = 0.02 **
Round Number [Continuous variable: Round Number]	-0.005 (SE 0.002) *p* = 0.001 ***
PANSS Total [Continuous variable: PANSS Total]	-0.002 (SE 0.005) *p* = 0.64
WASI-estimated IQ [Continuous variable: WASI-estimated IQ]	-0.005 (SE 0.004) *p* = 0.21
Observations	2000
Log likelihood (AIC)	-1084.8 (2219.6)

Table S3: Probit models for clinical variables for forward stepwise regression
Probit marginal effects (average marginal effects unless noted) and standard errors

Model 1, also presented in Table ~~1~~ 2 in the main text, shows the first step without any clinical variable, while including the loss (aversion) effect, group effect, and effects from strategies and over rounds. Model 2, from Table ~~1~~ 2 in the main text, is not presented here, as it does not form part of the stepwise regression. Models 3 to 22 show the second step adding individual clinical variables to model 1, including a comparison of loss aversion within the clinical variable. For all models, the log likelihood and Akaike Information Criterion (AIC) are shown to allow comparisons of model fit.

*, **, *** indicate significance at ≤ 0.10, ≤ 0.05 and ≤ 0.01 levels; *p*-values are from χ^2 tests.

Model No. [Clinical Variable]	Log likelihood (AIC)	**Marginal effects (with Standard Errors, SEs) and *p*-values**						
		Clinical Variable	Loss (Aversion) effect [Gain]	Loss (Aversion) effect [Gain]		Group [Patient]	Strategy [Neutral]	Round [Round]
				At 25th percentile [Clinical Variable]	At 75th percentile [Clinical Variable]			
Model 1 [No clinical variable]	-2030.8 (4095.6)		0.047 (SE 0.017) p = 0.01 ***			-0.016 (SE 0.094) p = 0.87	0.011 (SE 0.017) p = 0.52	-0.005 (SE 0.001) p < 0.0001 ***
Model 2	Not presented here, as not part of the stepwise regression, rather presented in Table ~~1~~ 2 in the main text.							
Model 3 [Age]	-2021.8 (4093.7)	-0.0006 (SE 0.0041) p = 0.89	0.047 (SE 0.018) p = 0.004 ***	0.020 (SE 0.020) p = 0.16	0.084 (SE 0.022) p = 0.0001 ***	-0.014 (SE 0.094) p = 0.88	0.012 (SE 0.017) p = 0.50	-0.005 (SE 0.001) p < 0.0001 ***
				Difference between LA for 25th & 75th percentiles [Age]				
				p = 0.008 ***				
Model 4 [Income]	-2028.3 (4106.5)	4×10^{-5} (SE 8×10^{-5}) p = 0.62	0.047 (SE 0.017) p = 0.003 ***	0.072 (SE 0.026) p = 0.003 ***	0.020 (SE 0.026) p = 0.22	0.027 (SE 0.125) p = 0.83	0.011 (SE 0.017) p = 0.52	-0.005 (SE 0.001) p < 0.0001 ***
				Difference between LA for 25th & 75th percentiles [Income]				
				p = 0.17				

Model No. [Clinical Variable]	Log likelihood (AIC)	**Marginal effects (with Standard Errors, SEs) and *p*-values**						
		Clinical Variable	Loss (Aversion) effect [Gain]	Loss (Aversion) effect [Gain]		Group [Patient]	Strategy [Neutral]	Round [Round]
				At 25th percentile [Clinical Variable]	At 75th percentile [Clinical Variable]			
Model 5 [Anti-psychotic (AP) dose]	-2026.5 (4103.1)	-6×10^{-5} (SE 16×10^{-5}) $p = 0.72$	0.048 (SE 0.018) $p = 0.003$ ***	-0.002 (SE 0.028) $p = 0.53$	0.080 (SE 0.024) $p = 0.0005$ ***	0.017 (SE 0.132) $p = 0.90$	0.011 (SE 0.017) $p = 0.51$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [AP dose]				
				$p = 0.02$ **				
Model 6 [WASI IQ]	-2016.8 (4083.6)	-0.004 (SE 0.003) $p = 0.21$	0.042 (SE 0.018) $p = 0.008$ ***	0.005 (SE 0.020) $p = 0.41$	0.113 (SE 0.024) $p < 0.0001$ ***	-0.092 (SE 0.110) $p = 0.40$	0.012 (SE 0.017) $p = 0.49$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [WASI]				
				$p = 0.0001$ ***				
Model 7 [NART IQ]	-2019.2 (4088.3)	0.0003 (SE 0.0054) $p = 0.96$	0.045 (SE 0.017) $p = 0.005$ ***	0.007 (SE 0.021) $p = 0.36$	0.085 (SE 0.021) $p < 0.0001$ ***	-0.013 (SE 0.099) $p = 0.89$	0.012 (SE 0.017) $p = 0.50$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [NART]				
				$p = 0.0006$ ***				
Model 8 [Hayling]	-2026.1 (4102.1)	0.019 (SE 0.027) $p = 0.47$	0.048 (SE 0.017) $p = 0.003$ ***	0.046 (SE 0.021) $p = 0.01$ ***	0.050 (SE 0.027) $p = 0.03$ **	0.015 (SE 0.102) $p = 0.88$	0.012 (SE 0.017) $p = 0.49$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [Hayling]				
				$p = 0.91$				

Model No. [Clinical Variable]	Log likelihood (AIC)	**Marginal effects (with Standard Errors, SEs) and *p*-values**						
		Clinical Variable	Loss (Aversion) effect [Gain]	Loss (Aversion) effect [Gain]		Group [Patient]	Strategy [Neutral]	Round [Round]
				At 25th percentile [Clinical Variable]	At 75th percentile [Clinical Variable]			
Model 9 [Brixton]	-2026.4 (4102.8)	-0.026 (SE 0.019) $p = 0.16$	0.047 (SE 0.017) $p = 0.003$ ***	0.048 (SE 0.028) $p = 0.04$ **	0.047 (SE 0.021) $p = 0.01$ ***	-0.073 (SE 0.099) $p = 0.46$	0.009 (SE 0.017) $p = 0.59$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [Brixton]				
				$p = 0.97$				
Model 10 [PANSS Positive]	-2017.3 (4084.5)	-0.006 (SE 0.013) $p = 0.63$	0.046 (SE 0.018) $p = 0.005$ ***	0.240 (SE 0.046) $p < 0.0001$ ***	-0.112 (SE 0.050) $p = 0.99$	0.065 (SE 0.193) $p = 0.74$	0.012 (SE 0.017) $p = 0.50$	-0.005 (SE 0.001) $p < 0.0001$ ***
				LA for 25th & 75th percentiles [Positive]				
				$p = 0.00001$ ***				
Model 11 [PANSS Negative]	-2022.4 (4094.8)	-0.008 (SE 0.012) $p = 0.51$	0.049 (SE 0.017) $p = 0.003$ ***	0.149 (SE 0.049) $p = 0.001$ ***	-0.031 (SE 0.038) $p = 0.79$	0.109 (SE 0.205) $p = 0.60$	0.011 (SE 0.017) $p = 0.53$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [PANSS Negative]				
				$p = 0.02$ **				
Model 12 [PANSS General]	-2017.6 (4085.1)	0.0001 (SE 0.0085) $p = 0.99$	0.046 (SE 0.018) $p = 0.004$ ***	0.195 (SE 0.049) $p = 0.0001$ ***	-0.106 (SE 0.046) $p = 0.99$	-0.021 (SE 0.212) $p = 0.92$	0.011 (SE 0.017) $p = 0.51$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [PANSS General]				
				$p = 0.0001$ ***				

Model No. [Clinical Variable]	Log likelihood (AIC)	**Marginal effects (with Standard Errors, SEs) and *p*-values**						
		Clinical Variable	Loss (Aversion) effect [Gain]	Loss (Aversion) effect [Gain]		Group [Patient]	Strategy [Neutral]	Round [Round]
				At 25th percentile [Clinical Variable]	At 75th percentile [Clinical Variable]			
Model 13 [PANSS Total]	-2015.9 (4081.8)	-0.001 (SE 0.004) $p = 0.72$	0.047 (SE 0.018) $p = 0.004$ ***	0.243 (SE 0.052) $p < 0.0001$ ***	-0.127 (SE 0.058) $p = 0.99$	0.052 (SE 0.221) $p = 0.81$	0.012 (SE 0.017) $p = 0.50$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [PANSS Total]				
				$p = 0.00004$ ***				
Model 14 [BDI]	-2024.9 (4099.8)	-0.002 (SE 0.007) $p = 0.78$	0.051 (SE 0.018) $p = 0.002$ ***	0.119 (SE 0.029) $p < 0.0001$ ***	0.008 (SE 0.021) $p = 0.34$	0.007 (SE 0.127) $p = 0.95$	0.011 (SE 0.017) $p = 0.54$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [BDI]				
				$p = 0.002$ ***				
Model 15 [Hamilton]	-2022.9 (4095.9)	-0.008 (SE 0.011) $p = 0.50$	0.050 (SE 0.018) $p = 0.002$ ***	0.111 (SE 0.027) $p < 0.0001$ ***	0.016 (SE 0.020) $p = 0.21$	0.039 (SE 0.126) $p = 0.76$	0.012 (SE 0.017) $p = 0.47$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [Hamilton]				
				$p = 0.002$ ***				
Model 16 [STAI] (anxiety)	-2023.0 (4096.0)	-0.003 (SE 0.005) $p = 0.58$	0.052 (SE 0.018) $p = 0.002$ ***	0.123 (SE 0.028) $p < 0.0001$ ***	-0.009 (SE 0.023) $p = 0.66$	0.026 (SE 0.120) $p = 0.83$	0.011 (SE 0.017) $p = 0.54$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [STAI]				
				$p = 0.0002$ ***				

Model No. [Clinical Variable]	Log likelihood (AIC)	**Marginal effects (with Standard Errors, SEs) and *p*-values**						
		Clinical Variable	Loss (Aversion) effect [Gain]	Loss (Aversion) effect [Gain]		Group [Patient]	Strategy [Neutral]	Round [Round]
				At 25th percentile [Clinical Variable]	At 75th percentile [Clinical Variable]			
Model 17 [Snaith-Hamilton Pleasure Scale]	-2026.9 (4103.8)	-0.029 (SE 0.026) $p = 0.26$	0.048 (SE 0.017) $p = 0.003$ ***	0.069 (SE 0.023) $p = 0.001$ ***	0.036 (SE 0.018) $p = 0.02$ **	0.009 (SE 0.095) $p = 0.92$	0.011 (SE 0.017) $p = 0.51$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [Snaith-Hamilton]				
				$p = 0.12$				
Model 18 [Levenson's Psycho-pathy scale]	-2025.5 (4101.0)	-0.013 (SE 0.004) $p = 0.003$ ***	0.046 (SE 0.017) $p = 0.003$ ***	0.040 (SE 0.023) $p = 0.04$ **	0.041 (SE 0.022) $p = 0.03$ **	0.163 (SE 0.095) $p = 0.09$ *	0.011 (SE 0.017) $p = 0.52$	-0.005 (SE 0.001) p < 0.0001 ***
				Difference between LA for 25th & 75th percentiles [Levenson's]				
				$p = 0.97$				
Model 19 [Propensity To Trust (PTS) scale]	-2028.4 (4106.7)	0.003 (SE 0.003) $p = 0.38$	0.047 (SE 0.017) $p = 0.004$ ***	0.053 (SE 0.021) $p = 0.006$ ***	0.040 (SE 0.022) $p = 0.04$ **	0.025 (SE 0.103) $p = 0.81$	0.012 (SE 0.017) $p = 0.50$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [PTS]				
				$p = 0.62$				
Model 20 [Length of time since diagnosis, years]	-2027.7 (4105.5)	-0.008 (SE 0.005) $p = 0.08$ *	0.047 (SE 0.017) $p = 0.003$ **	0.049 (SE 0.022) $p = 0.01$ ***	0.045 (SE 0.028) $p = 0.05$ **	-0.165 (SE 0.117) $p = 0.16$	0.010 (SE 0.017) $p = 0.56$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [time since diagnosis]				
				$p = 0.92$				

Model No. [Clinical Variable]	Log likelihood (AIC)	**Marginal effects (with Standard Errors, SEs) and *p*-values**						
		Clinical Variable	Loss (Aversion) effect [Gain]	Loss (Aversion) effect [Gain]		Group [Patient]	Strategy [Neutral]	Round [Round]
				At 25th percentile [Clinical Variable]	At 75th percentile [Clinical Variable]			
Model 21 [Total length of inpatient stays, months]	-2026.9 (4103.8)	-0.0007 (SE 0.0014) $p = 0.58$	0.047 (SE 0.017) $p = 0.003$ ***	0.047 (SE 0.018) $p = 0.005$ ***	0.048 (SE 0.018) $p = 0.003$ ***	-0.002 (SE 0.097) $p = 0.98$	0.012 (SE 0.017) $p = 0.49$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for 25th & 75th percentiles [length of inpatient stays]				
				$p = 0.92$				

Model No. [Clinical Variable]	Log likelihood (AIC)	**Marginal effects (with Standard Errors, SEs) and *p*-values**						
		Clinical Variable [Sex]	Loss (Aversion) effect [Gain]	Loss (Aversion) effect [Gain]		Group [Patient]	Strategy [Neutral]	Round [Round]
				Female [Sex = 0]	Male [Sex = 1]			
Model 22 [Sex]	-2030.1 (4110.2)	0.067 (SE 0.191) $p = 0.73$	0.048 (SE 0.017) $p = 0.003$ ***	-0.021 (SE 0.067) $p = 0.62$	0.052 (SE 0.018) $p = 0.002$ ***	-0.016 (SE 0.094) $p = 0.86$	0.011 (SE 0.017) $p = 0.52$	-0.005 (SE 0.001) $p < 0.0001$ ***
				Difference between LA for Female & Male				
				$p = 0.29$				

Table S4: Probit regression analysis for loss aversion with PANSS subscales
Probit marginal effects (average marginal effects unless noted) and standard errors
*, **, *** indicate significance at ≤ 0.10, ≤ 0.05 and ≤ 0.01 levels; *p*-values are from χ^2 tests.

Dependent variable: cooperate choice		
Group effect [Binary variable: Patient Baseline: Control (Patient=0)]		0.102 (SE 0.213) *p* = 0.63
Loss (Aversion) effect [Binary variable: Gain frame Baseline: Loss frame (Gain = 0)]		0.046 (SE 0.017) *p* = 0.004 ***
Loss (Aversion) effect …		
	… Controls (Patient = 0)	-0.079 (SE 0.069) *p* = 0.87
	… Patients (Patient = 1)	0.158 (SE 0.043) *p* = 0.0001 ***
Significance of difference		*p* = 0.01 ***
Loss (Aversion) effect …		
	… PANSS Positive = 7 (25th percentile)	0.202 (SE 0.076) *p* = 0.004 ***
	… PANSS Positive = 23.5 (75th percentile)	-0.064 (SE 0.060) *p* = 0.86
Significance of difference		*p* = 0.04 **
Loss (Aversion) effect …		
	… PANSS Negative = 8.5 (25th percentile)	0.068 (SE 0.060) *p* = 0.13
	… PANSS Negative = 26 (75th percentile)	0.013 (SE 0.039) *p* = 0.37
Significance of difference		*p* = 0.55
Loss (Aversion) effect …		
	… PANSS General = 19.5 (25th percentile)	0.063 (SE 0.039) *p* = 0.05 **
	… PANSS General = 45 (75th percentile)	0.017 (SE 0.057) *p* = 0.38
Significance of difference		*p* = 0.61
Strategy [Binary variable: Neutral strategy Baseline: Seesaw strategy (Neutral = 0)]		0.010 (SE 0.017) *p* = 0.56
Round Number [Continuous variable: Round Number]		-0.005 (SE 0.001) *p* < 0.0001 ***
PANSS Positive [Continuous variable: PANSS Positive]		-0.018 (SE 0.019) *p* = 0.35
PANSS Negative [Continuous variable: PANSS Negative]		0.015 (SE 0.015) *p* = 0.32
PANSS General [Continuous variable: PANSS General]		0.016 (SE 0.015) *p* = 0.29
Observations		3600
Log likelihood (AIC)		-2009.8 (4101.6)

Table S5: Probit analyses for loss aversion involving antipsychotic dose
Probit marginal effects (average marginal effects unless noted) and standard errors
*, **, *** indicate significance at ≤ 0.10, ≤ 0.05 and ≤ 0.01 levels; *p*-values are from χ^2 tests.

Dependent variable: cooperate choice	
Group effect [Binary variable: Patient Baseline: Control (Patient=0)]	0.068 (SE 0.225) ***p = 0.76***
Loss (Aversion) effect [Binary variable: Gain frame Baseline: Loss frame (Gain = 0)]	0.044 (SE 0.018) ***p = 0.007*** ***
Loss (Aversion) effect …	
… Controls (Patient = 0)	-0.054 (SE 0.059) ***p = 0.82***
… Patients (Patient = 1)	0.122 (SE 0.047) ***p = 0.004*** ***
Significance of difference	***p = 0.06*** ***
Loss (Aversion) effect …	
… PANSS Total = 35 (25th percentile)	0.199 (SE 0.054) ***p = 0.0001*** ***
… PANSS Total = 94.5 (75th percentile)	-0.073 (SE 0.054) ***p = 0.91***
Significance of difference	***p = 0.006*** ****
Strategy [Binary variable: Neutral strategy Baseline: Seesaw strategy (Neutral = 0)]	0.012 (SE 0.017) ***p = 0.47***
Round Number [Continuous variable: Round Number]	-0.005 (SE 0.001) $p = 8 \times 10^{-6}$ ***
PANSS Total [Continuous variable: PANSS Total]	-0.004 (SE 0.004) ***p = 0.36***
WASI-estimated IQ [Continuous variable: WASI-estimated IQ]	-0.005 (SE 0.003) ***p = 0.12***
Antipsychotic dose [Continuous variable: Antipsychotic dose]	0.00003 (SE 0.00016) ***p = 0.84***
Observations	3600
Log likelihood (AIC)	-2004.9 (4091.8)

Game theoretical mathematical model

Here we demonstrate mathematically that choice behavior in the two frames should diverge as the decision maker's degree of loss aversion increases, with more *Betray* choices under the *Loss* frame relative to the *Gain* frame. This provides us with a simple measure of loss aversion: the within-participant difference between the fraction of *Cooperate* choices in the *Gain* frame and the corresponding fraction in the *Loss* frame, with larger values associated with greater loss aversion.

An individual playing this game has utility given by $u_i(x_i, y_i) = v_i(x_i, y_i) + \psi_i(x_i, y_i)$, where x_i is her own action choice (either C=cooperate or B=betray) and $y_i \in \{C, B\}$ is her opponent's choice. The utility function's two components are $v_i(x_i, y_i)$, the value she gets from her monetary payoff ($\pi_i(x_i, y_i)$) from an outcome of (x_i, y_i), and $\psi_i(x_i, y_i)$, the psychological benefit she gets from that outcome. We assume that her value from the monetary payoff is linear in that payoff except possibly for loss aversion:

$$v_i(x_i, y_i) = \begin{cases} \pi_i(x_i, y_i) & \pi_i(x_i, y_i) \geq 0 \\ -\beta_i \cdot |\pi_i(x_i, y_i)| & \pi_i(x_i, y_i) < 0 \end{cases}$$

where $\beta_i \geq 1$ is her loss-aversion parameter (i.e., the factor by which losses are felt relative to equal-sized gains, so that $\beta_i = 1$ means no loss aversion). We make no specific assumptions about the form of $\psi_i(\cdot)$, other than that it depends only on the outcome:

$$\psi_i(x_i, y_i) = \begin{cases} \psi_{CCi} & (x_i, y_i) = (C, C) \\ \psi_{CBi} & (x_i, y_i) = (C, B) \\ \psi_{BCi} & (x_i, y_i) = (B, C) \\ \psi_{BBi} & (x_i, y_i) = (B, B) \end{cases}$$

The specific values of ψ_{CCi} ψ_{CBi} ψ_{BCi} and ψ_{BBi} could all be zero (for a player only motivated by monetary considerations), or some or all could be non-zero (conditional cooperation, altruism, reciprocity, efficiency seeking, competitiveness, inequity aversion).

Let $\Delta\psi_{Ci} = \psi_{CCi} - \psi_{BCi}$ and $\Delta\psi_{Bi} = \psi_{CBi} - \psi_{BBi}$; these are the gains in psychological benefit from choosing C instead of B, given that the opponent is choosing C and B respectively. Finally, suppose the individual believes that her opponent will choose C with probability $q \in (0,1)$ and B with probability 1-q.

Then, in the Loss game, her expected utility from choosing C is given by $u_i(C) = q[1+\psi_{CCi}] + (1-q)[-3\beta_i + \psi_{CBi}]$, while choosing B yields $u_i(B) = q[3+\psi_{BCi}] + (1-q)[-\beta_i + \psi_{BBi}]$. Then, if we define $\Delta u_i = u_i(C) - u_i(B)$, she prefers choosing C if $\Delta u_i > 0$; that is, if $q \cdot \Delta\psi_{Ci} + (1-q) \cdot \Delta\psi_{Bi} - 2[q + (1-q)\beta_i] > 0$, or equivalently $\beta_i < \frac{q \cdot \Delta\psi_{Ci} + (1-q) \cdot \Delta\psi_{Bi} - 2q}{2(1-q)} \equiv \beta_i^*$. Similarly, she prefers choosing B if $\beta_i > \beta_i^*$ and is indifferent between B and C if $\beta_i = \beta_i^*$. Thus, if $q \cdot \Delta\psi_C + (1-q) \cdot \Delta\psi_B > 2$, the individual would choose C if she were not loss averse (i.e., $\beta_i = 1$) and for values of β_i sufficiently close to 1, but beyond a threshold value of β_i, she would switch from C to B.

Since behaviour in the Gain game (for any β_i) is identical to that in the Loss game with $\beta_i = 1$, the above implies that individuals with sufficiently strong loss aversion will choose C less often in the Loss game than in the Gain game, while those with weak or no loss aversion will not behave differently in the two games.

Screenshots

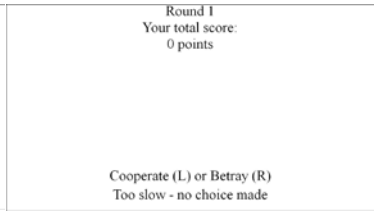
Choice Stage

5-second timer

Cooperative choice made

Betray choice made

No choice made



Feedback stage for 'Gain frame'

Computer player –

Cooperate

Betray

Human player –

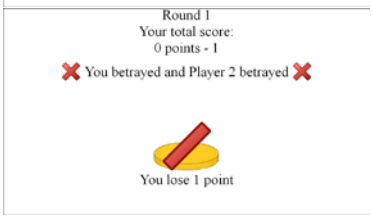
Cooperate

Betray

No choice

Feedback stage for ‘Loss frame’

Computer player –

Human player –	Cooperate	Betray
Cooperate		
Betray		
No choice