**Comparison of social prediction error signals obtained from fMRI data during a trust game in patients with schizophrenia and healthy controls**

**1.Introduction**

Schizophrenia is a complex and multifaceted mental disorder characterized by a range of cognitive deficits. Cognitive impairments, including deficits in memory, attention, and executive function, precede the clinical diagnosis of schizophrenia, highlighting their role as core features of the disorder1,2. Social cognitive dysfunction, a subset of cognitive impairment, refers to difficulties in understanding and processing social information, which can manifest as challenges in recognizing emotions, understanding social cues, managing complex tasks, and engaging in effective communication3,4. Social cognition is of great importance in adapting to an environment with numerous stimuli and changes. Difficulties in social functioning lead to social isolation, and interpersonal problems exacerbate feelings of loneliness and depression. These further compound the overall impact of the disorder on quality of life, causing relapses, significantly impairing quality of life, and negatively affecting rehabilitation processes5–7.

Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS), conducted by NIMH, suggested that social cognition be examined in five main areas. These areas are: emotion processing, social perception, social knowledge, attribution bias, and theory of mind8,9. Due to the complexity of social cognition, it is difficult to assess it with clinical interviews or scales. To solve this problem, researchers use neuroeconomics games that are based on game theory. Interpersonal interactions can be simulated in a laboratory environment with these games. Then, social decision making can be experimentally investigated and the underlying neuroscientific mechanisms can be elucidated10.

The trust game is one of the most studied neuroeconomics games for assessing social cognition11. This game typically involves two players: an investor and a trustee. The investor decides to invest, give an endowment to the trustee, or keep it himself. If the investor invests the money, the amount sent is usually increased. The trustee then decides to share the money with an investor or keep it all to himself. In a study of patients with schizophrenia, patients with schizophrenia invested less than controls during play with a cooperative partner but did not differ when playing with a deceptive partner. These findings suggest that they are impaired in adapting to changing environmental conditions, have reduced behavioral flexibility after negative feedback, and develop poor strategies. Further studies using functional magnetic resonance imaging (fMRI) have provided valuable insights in dysfunctional social cognition (REF).

In a trust game, a safe environment refers to a context in which the participant expects cooperative behavior from their partner, such as when interacting with a close friend or a consistently fair trustee. During the anticipation phase of the trust game in a safe environment, healthy individuals showed increased coupling between the right parietal cortex, the fusiform gyrus (FG), and the inferior/middle temporal gyrus. During cooperative responses in trust games, lower BOLD signals were observed in the right TPJ in patients with schizophrenia. The temporoparietal junction (TPJ) plays a crucial role in processes such as empathy, theory of mind, and social decision-making tasks12–15. This diminished activation may reflect the underlying cognitive deficits that characterize the disorder, such as impaired theory of mind and difficulties in processing social cues16.

Social interaction and social functioning involve many cognitive processes, including social perception, understanding the actions of others, observational social learning, and social decision making. To adapt to a rapidly changing world, the brain makes numerous predictions. Predictive mechanisms play a role in social communication and interpersonal interactions, affecting daily functioning. Motor control, perceptual inference, and reward-based learning occur through constant updating of predictions based on error17–19 called prediction error (PE). Prediction error refers to a mismatch between the prior prediction and the actual results. If the outcome is better than expected, it's called positive PE, and if the outcome is worse than expected, it is called negative PE. Effective learning occurs by monitoring prediction errors20. Being able to predict the consequences of our actions is essential for effective decision-making. The TPJ has been shown to play a role in social prediction errors21. Studies have shown that patients with schizophrenia demonstrate deficits in learning from social feedback, which may be linked to abnormal activation in the TPJ and other related brain regions22. These neural disruptions can lead to maladaptive social behaviors, such as distrust and social withdrawal, which further exacerbate the challenges faced by individuals with schizophrenia in their daily lives.

There are many studies indicating that PE is impaired in schizophrenia. Deactivation was detected in the right superior frontal lobe and inferior frontal lobes during PE compared to healthy controls. Increased activation was seen in the PCC, which is part of the DMN, indicating impairment in the estimation of self-worth in the schizophrenia group23. In negative prediction error, differences in mPFC and right middle frontal cortex activations have been observed23,24. In positive prediction error, impairments are manifested by deactivation in PFC, which may explain the avolition among negative findings25. It has also been suggested that impaired PE calculations in schizophrenia may cause delusions26,27. According to the incentive salience hypothesis, increased firing of chaotic or stress-related dopaminergic afferents in the striatum of schizophrenia patients has been proposed to attribute salience to irrelevant stimuli28–30. Over-attribution of meaning to irrelevant cues may affect thought content and mood, leading to perceptual distortions and the formation of delusions23. This condition has also been thought to affect cognitive processes in schizophrenia patients by causing paradoxical learning (over-learning of irrelevant and neutral information and under-learning of rewarding events)25. Moreover, the role of dopamine in schizophrenia cannot be overlooked. The dopamine hypothesis suggests that dysregulation of dopaminergic pathways contributes to the cognitive and emotional symptoms observed in schizophrenia25 Specifically, the TPJ's interaction with dopaminergic systems may influence social reward processing, which is critical for establishing trust and cooperation in social contexts31.

The interplay between neural mechanisms, social cognition, and dopamine dysregulation highlights schizophrenia's multifaceted nature. The TPJ emerges as a critical region for understanding the social deficits associated with the disorder, particularly in the context of trust and cooperation. The neural underpinnings of social cognition in schizophrenia with PE contribute to a deeper understanding of the disorder.

In this study, we aimed to compare TPJ activation between schizophrenia patients and healthy controls during the trust game. To achieve this, we designed a trust game combined with a reinforcement model that was structured first to induce positive PE, then negative PE. Prediction errors were extracted from the response screen of the game and linked with BOLD signals obtained via fMRI.

Our goal was to examine whether patients with schizophrenia invest differently during the trust game and to identify brain regions that exhibit altered activation patterns related to PE. We hypothesized that TPJ activation would be reduced in the schizophrenia group and that this reduction would be associated with abnormal prediction error processing.

**2.Methods**

**2.1.Participants**

Clinically stable, right-handed schizophrenia patients were recruited from the Ege University Department of Psychiatry Outpatient Unit. All patients had remained on a consistent treatment regimen for at least six months before participation, without further alteration in medical treatment The study was conducted under the Declaration of Helsinki, and written informed consent was obtained from all participants.

Patients with current depressive episodes, substance use disorders, or significant medical conditions affecting cerebral blood flow were excluded. Healthy controls were recruited through community advertisements and had no history of psychiatric disorders and no family history of schizophrenia. Demographic characteristics of all participants are presented in Table 1.

**2.2 Psychometric Assessment**

Schizophrenia diagnoses were verified using the Structured Clinical Interview for DSM-5 (SCID-5). The Positive and Negative Syndrome Symptom Scale (PANSS), a semi-structured assessment, was administered to measure recent symptom intensity and patient functionality. Cognitive and social functioning were evaluated with the Schizophrenia Cognition Rating Scale (ScoRS) and the Functional Remission of General Schizophrenia (FROGS).

| **Variable** | **Schizophrenia** | **Control** | **p** | **Stat** |
| --- | --- | --- | --- | --- |
| ***Demographics*** |  | | | |
| Number of Subject | 29 | 49 |  |  |
| sex = M (%) | 17 (58.6) | 17 (34.7) | 0.068 | Chi-square [X^2(1)=4.24] |
| Age (mean (SD)) | 39.86 (11.03) | 38.33 (12.38) | 0.583 | Welch t-test [t(64.4)=-0.57] |
| Education Years (mean (SD)) | 13.00 (3.22) | 13.00 (3.85) | 1.000 | Welch t-test [t(67.4)=0.00] |
| ***Clinical measures*** |  | | | |
| Age of Onset (mean (SD)) | 26.45 (7.11) |  |  |  |
| Duration of Illness (mean (SD)) | 13.72 (10.23) |  |  |  |
| CBZ (mean (SD)) | 502.25 (353.34) |  |  |  |
| PANSS Total (mean (SD)) | 43.58 (10.26) |  |  |  |
| PANSS Positive (mean (SD)) | 8.16 (1.49) |  |  |  |
| PANSS Negative (mean (SD)) | 12.20 (4.99) |  |  |  |
| PANSS General (mean (SD)) | 22.84 (5.85) |  |  |  |
| SCORS-GA (mean (SD)) | 3.74 (2.00) |  |  |  |
| FROGS (mean (SD)) | 70.80 (8.07) |  |  |  |

**Table 1.** Demographics and clinical characteristics of the subjects

**2.2.Task Paradigm**

In the proposed fMRI task, the experimental paradigm is based on the "Trust Game," a widely used game-theoretical model in the literature that captures interpersonal trust32. Participants met the trustee in person and were made to believe they were interacting with a real partner during the task, although the trustee’s responses were preprogrammed. The introduction with the trustee was standardized not to affect the basal trust level.

In each trial, the participant starts with an initial endowment of 40 Turkish Lira (TL) and must decide whether to keep the money or invest it by transferring it to a trustee. If the participant keeps the money, it is evenly split between them and the trustee. If the participant chooses to invest, the amount is tripled to 120 TL, and the trustee then decides whether to share this amount equally with the participant or keep all of it. The central task involves the participant evaluating whether to trust the trustee in each round. The trust game consists of three blocks, each with 20 trials, making a total of 60 trials. Each trial follows a standardized sequence presented across six screens (See Figure 1A). Each trial of the Trust Game followed a structured sequence of six consecutive screens, designed to capture the participant's decision-making and trust behavior under controlled experimental conditions (Figure 1B).

In the first decision or anticipation screen, participants saw a photo of the trustee alongside their initial endowment of 40 Turkish Lira (TL). They were instructed to decide whether to keep the money or invest it by transferring it to the trustee. No action or response was needed during this phase, and the screen remained visible for 6 seconds. On the second-choice screen, participants were asked to confirm their earlier decision (“invest” or “keep”) by pressing the appropriate button within a 3-second window. Before starting the task, they were informed that failing to respond would result in zero earnings for that trial (0 TL) and that missing responses on five trials—whether in a row or not—would lead to the termination of the entire task. A 3-second waiting screen followed, during which participants anticipated the trustee's response. The fourth outcome screen showed the trustee’s decision—whether to share the tripled amount or keep it entirely—along with the outcome of the trial, including the participant’s earnings. This screen was displayed for 3 seconds. In the fifth jitter screen inter-trial interval followed, lasting randomly between 3 and 6 seconds. During this time, a blank screen was displayed to introduce temporal variability between trials. In the last sixth fixation screen, to maintain a total inter-trial interval of 9 seconds, a fixation cross (“+”) was shown for the remaining time—ranging from 3 to 6 seconds—depending on the length of the preceding jittered interval.

Additionally, after every 10 trials, participants were prompted to rate their level of trust in the trustee using a scale from 1 (not at all) to 7 (completely). This rating screen remained visible for 9 seconds. Each trial lasted 24 seconds, leading to a total task duration of approximately 24 minutes and 54 seconds.

A diagram of a group of people

AI-generated content may be incorrect.

**Figure 1.** Trust Game Flow Model

A diagram of a diagram

AI-generated content may be incorrect.

**Figure 2.** fMRI task model for each trial. Each participant plays the game as sets of 20 trials, i.e., 60 trials in total.

**2.3.Computational Learning Model Based on Rescorla-Wagner**

We implemented a reinforcement learning (RL) model based on the Rescorla-Wagner rule to estimate participants’ trial-by-trial reward predictions and learning dynamics during a sequential investment task. The model was applied separately to three experimental blocks (phases, each with 20 trials) featuring varying gain probabilities (80%, 50%, and 80%).

On each trial , the predicted reward value was updated according to the standard delta rule:

Here, is the actual reward and is the trial-specific learning rate. Unlike classical fixed-𝛼 Rescorla-Wagner models, we introduced an adaptive learning rate mechanism that updated 𝛼 based on the social partner’s behavior with a fixed step size (:

If the participant invested and the trustee shared:

If the participant invested and trustee keeps:

If the participant did not invest:

**2.3.1 Prediction Error Acquisition**

Prediction error (PE) was defined as the discrepancy between the obtained reward and the model’s predicted reward on each trial. Formally, at trial t the signed prediction error is

The value estimated is updated via delta rule:

Here, includes direction, positive values indicate outcomes better than expected (positive PE) and negative values indicate worse-than expected values (negative PE). Block-wise initialization followed the task structure with the first block and . At the start of the second and third blocks, the learning rate was reset to 0.5 while the prior predicted value carried over from the last trial of the preceding block. The resulting trial-by-trial sequence of formed a prediction error time series, which was entered as a parametric regressor in the BOLD analysis to examine how neural activity tracked both the magnitude and direction (valence) of deviations from expectation.

**2.4.fMRI Analysis**

**2.4.1. fMRI Preprocessing**

All preprocessing was carried out in SPM running under MATLAB R2024a. Each subject’s functional runs were realigned to the first volume using the “Estimate & Reslice” routine to correct for head motion. The realigned time‑series were then slice‑timing corrected (interpolated to the middle slice) to compensate for interleaved acquisition delays.

Next, the mean functional image was rigid‑body coregistered to the participant’s T1‑weighted structural scan. The structural image was segmented into gray matter, white matter and cerebrospinal fluid via SPM’s unified segmentation (using the tissue‑probability maps). The resulting deformation fields were applied to warp the functional volumes into MNI space (2×2×2 mm isotropic voxels). Finally, the normalized functional images were smoothed with an 8 mm FWHM Gaussian kernel to enhance signal-to-noise ratio and accommodate inter-subject anatomical variability.

**2.4.2.First Level**

For each participant, the design matrix included task regressors and trial-by-trial prediction error (PE) estimates derived from a computational model. Specifically, three separate parametric regressors were included: , and each corresponding to trial-by-trial learning signals within one of three blocks with varying reward contingencies. These regressors were modeled as parametric modulators without orthogonalization and were convolved with the canonical hemodynamic response function (HRF). Motion parameters were not included as additional regressors in the final model, and a high-pass filter with a 128 s cutoff was applied to remove low-frequency drifts.

The GLM was estimated for each subject, and contrast images were generated for each condition of interest. Contrasts included each individual PE regressor (e.g , , ) as well as their average. These contrast images were then carried forward to the second-level analysis for group-level inference.

**2.4.3.Second Level**

Second-level statistical analyses were conducted using SPM to assess within-group and between-group effects of prediction error (PE)–related activation. First-level contrast images representing trial-by-trial PE modulation were entered into second-level models for group-level inference.

For the within-group one-sample t-tests were conducted separately for the HC and SZ groups to identify brain regions showing significant PE-related activation within each population. For the HC group, sex was included as a covariate. For the SZ group, both sex and duration of illness (DoI) were included as covariates to account for individual differences that may influence PE-related brain responses. Covariates were mean-centered before inclusion. All second-level models were estimated using classical (maximum likelihood) estimation, and results were corrected for multiple comparisons at the cluster level (FWE-corrected p < 0.05) using a cluster-forming threshold of p < 0.001 uncorrected.

Additionally, between-group analysis, a two-sample t-test was used to compare individuals with schizophrenia (SZ) to healthy controls (HC). To control for confounding variables, sex was included as a covariate of no interest. Contrast images for each subject were grouped accordingly, and individuals with missing or incomplete data were excluded from the analysis. Group comparisons (e.g., SZ > HC and HC > SZ) were tested while accounting for sex-related variance. Anatomical locations were identified using the Neuromorphometrics atlas provided in SPM.

**2.5 Hypothesis Testing**

For investment decisions, prediction errors, and the learning parameter, we used an appropriate general linear model based on our assumptions. Model diagnostics are detailed in the Supplementary Material.

**2.5.1. Binomial Generalized Linear Mixed-Effects Model for Investment Decisions**

We analyzed binary investment responses using a generalized linear mixed-effects model (GLMM) with a binomial distribution and a logit link function. The aim was to investigate the effects of the experimental group and task phase on the likelihood of making an investment decision, while accounting for repeated measures within participants. The model included fixed effects for group and task phase (three levels: phase 1, phase 2, and phase 3). Participant-level variability was modeled with a random intercept for subject.

Here, the term is the probability that the participant on trial makes a positive response (e.g invests). ​ follows the assumption that the response variable is Bernoulli distributed, with success probability defined by the fixed and random effects through a logistic link. is the fixed intercept, where is fixed effect for Group (patient vs control) and indicates effect of Task phase relative to first phase as reference. Sex and Duration of Illness (DOI) are included in the model, with sex being a categorical factor and DOI a continuous variable. denoting random intercept for the subject with denoting residual error.

**2.5.2.Lognormal Linear Mixed-Effects Model for Prediction Errors**

Because prediction error (PE) values included zero or negative values, we applied a constant shift to ensure all observations were strictly positive. This transformation preserved the relative structure of the data while enabling the use of log-transformation. The log-transformed PE values were then modeled using a linear mixed-effects model, assuming a lognormal distribution. This approach is appropriate for positively skewed data with heteroscedasticity and allows for multiplicative interpretation of fixed effects. Residual diagnostics indicated no issues with dispersion or heteroscedasticity; however, tests for residual uniformity and outliers revealed minor deviations, suggesting the model may be sensitive to extreme values (see Supplementary Material).

Based on the data characteristics and model diagnostics, the final model was defined as:

Here is the PE for the subject in the trial . is the shifted PE value, where the constant is added to make values strictly positive. follows the assumption that PE is lognormal distributed. is the fixed intercept, where is fixed effect for Group (patient vs control) and indicates the effect of the Task phase relative to the first phase as reference. As with the binomial model, Sex and Duration of Illness (DOI) are included in the model, with sex being a categorical factor and DOI is a continuous variable. denotes random intercept for subject with denoting residual error. Model implemented in R 4.3.234 with lme4 package35.

**2.5.3. Gaussian Mixed-Effects Model for Trial-by-Trial Learning Rates**

To examine changes in trial-by-trial learning rate (α) across task phases and clinical groups, we fitted a generalized linear mixed-effects model assuming a Gaussian distribution with a logit link function. This approach is appropriate for continuous outcomes constrained to the (0, 1) interval—such as learning rates derived from the Rescorla-Wagner model—while enabling linear modeling on the logit-transformed scale. All α values were strictly within the (0,1) range, avoiding the boundary values (0 or 1) that are undefined under the logit transformation.

The model included fixed effects for Group, Task phase, and Sex, as well as their interaction (Group × Sex), and Duration of Illness (DoI) as a continuous covariate similar to the PE model above. A random intercept for Subject was included to account for repeated measurements within individuals. This structure allowed us to estimate how α varied across phases and between groups, while accounting for between-subject variability and clinical covariates. The final model was defined as:

**3.Results**

**3.1. Behavioral Results**

A binomial generalized linear mixed-effects model (GLMM) with a logit link was used to examine the effects of group, sex, task phase, and duration of illness (DoI) on investment decisions, while accounting for repeated measures with a random intercept for participants. Results are reported as odds ratios (OR) with 95% confidence intervals (see Methods – Hypothesis Testing).

The schizophrenia (SZ) group showed no significant difference in investment behavior compared to healthy controls (HC) (OR = 0.85, p = .590) in any of the phases, and sex was also not a significant predictor (OR = 1.05, p = .820). However, investment odds were significantly lower in phase 2 compared to phase 1 (OR ≈ 0.76, 95% CI ≈ [0.65, 0.89], p < .001) for both groups, suggesting a reduction in cooperative behavior during this phase. No significant difference was observed between phase 3 and phase 1 (OR ≈ 1.03, p = .750). Duration of illness (DoI) was not significantly associated with investment likelihood (OR ≈ 0.99, p = .503).

These findings indicate that while group and sex did not significantly influence investment decisions, phase 2 consistently decreased the odds of investing across participants. There is no evidence that group or sex moderated this phase effect in the current model specification.

**3.2. Prediction Error**

A linear mixed-effects model was fitted to examine the effects of group, sex, task phase, and duration of illness (DoI) on prediction error (PE), using the natural log of PE as the dependent variable. The model included fixed effects for group, sex, task phase (with phase 1 as the reference), their interaction (group × sex), and duration of illness (DoI), and a random intercept to account for repeated measures across subjects (see Methods – Hypothesis Testing).

The schizophrenia group exhibited a multiplicative effect of 0.96 (95% CI ≈ [0.904, 1.026], p = .22) compared to healthy controls, indicating a non-significant 4% reduction in PE. Male participants had significantly higher PE than females, with a multiplicative effect of 1.08 (95% CI ≈ [1.028, 1.124], p = .001), reflecting an 8% increase. A significant interaction between group and sex was observed (multiplicative effect = 1.12, 95% CI ≈ [1.044, 1.207], p = .001), indicating that males with schizophrenia had 12% higher PE than expected from the additive effects of group and sex alone.

Compared to phase 1, other task phases had no significant effect on PE. Duration of illness was positively associated with PE, with a multiplicative effect of 1.003 (95% CI ≈ [1.000, 1.006], p<.05), reflecting a 0.3% increase in PE per unit increase in illness duration.

We compared psychometric assessments by sex but found no statistically significant differences using Welch’s t-test. None of the assessments correlated with PE. These results suggest that sex effects are independent of symptom intensity, cognitive symptoms, and patient functionality.

**3.3. Learning Rate**

To facilitate interpretation, model coefficients were exponentiated to reflect multiplicative effects on the raw learning rate (), estimated from a Rescorla-Wagner model. A multiplicative effect greater than 1.0 indicates an increase in α relative to the reference level, while values below 1.0 indicate a decrease.

Relative to healthy control (HC) females in Task 1, schizophrenia (SZ) females showed a 1.10-fold increase in learning rate (multiplicative effect = 1.10, 95% CI ≈ [1.07, 1.14],

p < .001). In contrast, HC males exhibited a 0.91-fold learning rate (95% CI ≈ [0.88, 0.94],

p < .001), indicating reduced learning relative to HC females. A significant negative interaction between group and sex revealed that SZ males had the lowest learning rate overall, with a 0.90-fold effect relative to the expected additive effects of SZ and male sex (95% CI ≈ [0.86, 0.95], p < .001).

Task phase did not significantly affect learning rate: Phase 2 yielded a 1.01-fold change (95% CI ≈ [0.98, 1.03], p = .62), and Phase 3 a 1.01-fold change (95% CI ≈ [0.98, 1.04],

p = .79), both statistically equivalent to Phase 1. Duration of illness (DoI) was also not significantly associated with α (multiplicative effect = 0.99, 95% CI ≈ [0.98, 1.00], p = .111), though the effect was slightly negative.

These results indicate that alterations in learning rate are sex-specific in schizophrenia, with increased in SZ females and decreased in SZ males, independent of task phase or illness duration.

**3.4.PE-related Activations**

In the HC group, significant activation was observed in left superior parietal lobule and angular gyrus (peak MNI: [-30, -64, 26], T = 7.46, p < .05 FWE-corrected, k = 58 voxels) with right middle and superior occipital gyrus (peak MNI: [30, -76, 11], T = 7.52, p < .05 FWE-corrected, k = 28 voxels) and in response to PE controlling for sex.

However, in the SZ group, significant activation was observed in right occipital fusiform and inferior occipital gyrus in response to PE, controlling for sex and DoI (peak MNI: [27, -85, -7], T = 4.98, p < .001 uncorrected, k = 48 voxels).

In the between-group analysis, the SZ group exhibited greater PE-related activation than the HC group in the left cerebellum exterior and occipital fusiform gyrus (peak MNI: [−18, −85, −25], T = 4.22, p < .001 uncorrected, k = 26 voxels). No significant clusters were observed for the reverse contrast (HC > SZ).

In summary, the results indicate that the temporal-parietal junction (TPJ) shows increased PE-related activation in healthy controls, whereas this effect is absent in individuals with schizophrenia. In contrast, the schizophrenia group exhibited PE-related activation predominantly in lower-level sensory regions, including the occipital cortex and cerebellum. These findings may reflect a deficit in mentalizing processes in schizophrenia, whereby prediction errors are processed primarily in unimodal sensory regions rather than in higher order transmodal areas such as the TPJ.

**A group of brain images

AI-generated content may be incorrect.**

**Figure 6***. Prediction‐error‐related activations identified in SPM.*A) For visualisation purposes, uncorrected global p-values (p < .001) indicated activation in both right and left TPJ regions among healthy controls, controlling for sex (peak MNI: [30, −76, 11], T = 7.52, p < .001 uncorrected, k = 433 voxels; peak MNI: [-30, −64, 26], T = 7.46, p < .001 uncorrected, k = 359 voxels). Following FWE correction, significant clusters were identified in the left superior parietal lobule/angular gyrus (peak MNI = [–30, –64, 26], T = 7.46, p < .05 FWE-corrected, k = 58), as well as in the right middle and superior occipital gyrus (peak MNI = [30, –76, 11], T = 7.52, p < .05 FWE-corrected, k = 28).   B) In the schizophrenia group (controlling for sex and duration of illness), a significant cluster emerged in the right occipital fusiform and inferior occipital gyrus (peak MNI = [27, –85, –7], T = 4.98, p < .001 uncorrected, k = 48).C) In the between‐group contrast (SZ > HC), greater prediction‐error‐related activation was found in the left cerebellum exterior and occipital fusiform gyrus (peak MNI = [–18, –85, –25], T = 4.22, p < .001 uncorrected, k = 26 )

**4.Discussion**

This study examined behavioral and neural correlates of social prediction error (PE) processing during a trust game in individuals with schizophrenia (SZ) compared to healthy controls (HC). While behavioral investment patterns were broadly comparable across groups, pronounced group differences emerged in PE processing and associated neural activations.

**4.1.Behavioral Adaptation and Trust Decisions**

Statistical results yielded no significant differences in the total number of investments across the task phases, contrary to several previous reports of diminished trust behavior in schizophrenia31,36. Building on our behavioral results, it is striking that both patients and controls invested similar amounts in our computer-driven trust game. This result partially aligns with previous work in first-episode psychosis (FEP) and clinical high-risk populations, where trust adaptation remains preserved22,37. While sex and duration of illness (DoI) did not predict trust, both SZ and HC participants reduced investment during Phase 2, suggesting a sensitivity to dynamic changes in social uncertainty. However, as shown with both increased prediction errors and decreased learning rate in schizophrenia, we replicated the impaired decision making and learning from feedback in schizophrenia38,39. King-Casas and Chiu’s research revealed that individuals with schizophrenia display reluctance to trust others, even when their partners act cooperatively10. Supporting this, Fett et al. found that individuals in the early stages of psychosis exhibit a generalized distrust that is not influenced by partner behavior. Yet, meta-analytic reviews remind us that when social ambiguity is minimized, as in algorithmic trust paradigms, group differences in trust behavior often diminish or disappear40.

By using a pre-programmed trustee with a fixed 80% reciprocation probability, we effectively reduced social uncertainty, which may explain why patients showed similar willingness to invest as healthy controls, despite having different learning processes. Our results indicate that under certain structured and repetitive task conditions, individuals with schizophrenia may exhibit behaviors similar to healthy controls in response to social feedback. Although investment patterns appear equal, this equality may be achieved through different neural pathways. However, the relatively high remission rates and low PANSS scores in our cohort, compared to the general schizophrenia population, may have attenuated behavioral manifestations.

**4.2.Prediction Error Processing and Sex-Specific Modulation**

Despite behavioral parity, group differences emerged in the neural encoding of prediction errors. A key finding was a significant group and sex interaction, wherein male patients exhibited increased PE relative to female patients and controls This effect was not attributable to the main effects of sex or group alone. It may reflect increased volatility or aberrant salience attribution in male patients—a group disproportionately affected by early-onset SZ and severe symptom burden41–43. That PE was positively associated with duration of illness further suggests that prolonged exposure to illness processes may alter PE computations, potentially reinforcing maladaptive belief updating.

Mismatch negativity (MMN) is a well‑established event‑related potential (ERP) component that is consistently found to be attenuated in individuals with schizophrenia44,45. This MMN attenuation—thought to index aberrant prediction error (PE) signaling in auditory cortex—is not only observed in patients but also in their first‑degree relatives, and it has been shown to predict the transition to psychosis in individuals at clinical high risk. These findings support the hypothesis that MMN reduction reflects disruptions in underlying predictive coding mechanisms. Importantly, MMN impairment has been associated with poor prognosis: remitted patients tend to show less MMN attenuation46,47 , suggesting a link between MMN deficits and learning abnormalities such as higher prediction errors and reduced learning rates. However, Erickson et al. reported that MMN deficits are not associated with illness duration and do not appear to worsen over the lifespan48. In contrast, our findings indicate that poorer prognostic factors—such as male gender and longer illness duration—are associated with higher PE and impaired learning. This supports the notion that disrupted prediction error signaling may represent a core and persistent feature of schizophrenia.

**4.3.Neural Correlates of Prediction Errors**

Our results align with hierarchical predictive‑coding accounts of schizophrenia: healthy controls (HC) showed robust prediction‑error (PE) signaling in the temporo‑parietal junction (TPJ), a trans‑modal hub implicated in mentalizing and belief updating49. HC group demonstrated activations in both low-level (occipital) and higher-order (TPJ) for belief updating in social context. Whereas individuals with schizophrenia failed to engage this higher‑order region and instead displayed PE‑related activity confined to lower‑level visual cortices and the cerebellum.

This downward shift suggests that, when higher-level predictions are imprecise or under‑weighted in schizophrenia50–53, unimodal sensory circuits assume a disproportionate role in registering discrepancies between expected and incoming information. The concomitant SZ > HC activation in cerebellar–occipital circuitry may represent a compensatory attempt to refine sensory timing and error calibration in the face of deficient TPJ involvement. Together, these findings reinforce the view that disrupted hierarchical weighting—and not merely reduced dopaminergic signaling—underlie impaired social‑cognitive inference in schizophrenia.

Beyond the hierarchical reweighting evident in our task, recent meta‐analytic evidence suggests that altered cerebellar involvement is a hallmark of prediction‐error dysfunction in schizophrenia. Xun Yang and colleagues (2024) synthesized 14 fMRI studies and reported a broad pattern of mesolimbic hypoactivity—including the striatum, thalamus, amygdala, hippocampus, anterior cingulate, insula, superior temporal gyrus, and notably the cerebellum—when patients processed prediction error49. In contrast, our finding of increased cerebellar activation in SZ > HC likely reflects a task‐specific compensatory recruitment of cerebello–thalamo–cortical loops, whereby patients lean on timing and error‐calibration functions of the cerebellum to support deficient transmodal signaling.

Moreover, theories of cerebellar sequencing propose that the cerebellum “predicts the future” by encoding temporal patterns and refining error signals54, a function that could be up-regulated to compensate for under-weighted top-down predictions in schizophrenia, highlighting the TPJ and cerebello-cortical loops as potential targets for intervention.

**5.Conclusion**

These findings advance our knowledge of the computational and circuit-level mechanisms underlying social dysfunction in schizophrenia.

These results show that, despite similar trust‑based investment behavior in schizophrenia and healthy controls, computational modeling revealed sex‑specific alterations: male patients exhibited elevated prediction errors and reduced learning rates, whereas female patients showed the opposite pattern. Prediction error magnitude also increased slightly with longer illness duration, underscoring a persistent disruption in learning from social feedback.

Neuroimaging demonstrated that healthy controls recruit both sensory cortices and the transmodal TPJ for prediction‑error processing, while individuals with schizophrenia fail to engage the TPJ and instead rely on lower‑level occipital and cerebellar regions. Together with attenuated auditory MMN, these findings point to a core deficit in hierarchical predictive coding in schizophrenia. Enhancing high-order belief updating—through targeted cognitive training or neuromodulation—may therefore help restore balanced error signaling and improve social functioning.

**6.References**

1. Guo, J. Y., Ragland, J. D. & Carter, C. S. Memory and cognition in schizophrenia. *Mol Psychiatry* **24**, 633–642 (2019).

2. Kochunov, P. *et al.* Association of White Matter With Core Cognitive Deficits in Patients With Schizophrenia. *JAMA Psychiatry* **74**, 958 (2017).

3. Caqueo-Urízar, A., Boyer, L., Baumstarck, K. & Gilman, S. E. Subjective perceptions of cognitive deficits and their influences on quality of life among patients with schizophrenia. *Qual Life Res* **24**, 2753–2760 (2015).

4. Lesh, T. A., Niendam, T. A., Minzenberg, M. J. & Carter, C. S. Cognitive Control Deficits in Schizophrenia: Mechanisms and Meaning. *Neuropsychopharmacol* **36**, 316–338 (2011).

5. Tripathi, A., Kar, S. K. & Shukla, R. Cognitive Deficits in Schizophrenia: Understanding the Biological Correlates and Remediation Strategies. *Clin Psychopharmacol Neurosci* **16**, 7–17 (2018).

6. Prouteau, A. *et al.* Cognitive predictors of psychosocial functioning outcome in schizophrenia: A follow-up study of subjects participating in a rehabilitation program. *Schizophrenia Research* **77**, 343–353 (2005).

7. Pascal De Raykeer, R. *et al.* Effects of depression and cognitive impairment on quality of life in older adults with schizophrenia spectrum disorder: Results from a multicenter study. *Journal of Affective Disorders* **256**, 164–175 (2019).

8. Green, M. F. & Leitman, D. I. Social Cognition in Schizophrenia. *Schizophrenia Bulletin* **34**, 670–672 (2007).

9. Robson, S. E., Repetto, L., Gountouna, V.-E. & Nicodemus, K. K. A review of neuroeconomic gameplay in psychiatric disorders. *Mol Psychiatry* **25**, 67–81 (2020).

10. King-Casas, B. & Chiu, P. H. Understanding Interpersonal Function in Psychiatric Illness Through Multiplayer Economic Games. *Biological Psychiatry* **72**, 119–125 (2012).

11. Berg, J., Dickhaut, J. & McCabe, K. Trust, Reciprocity, and Social History. *Games and Economic Behavior* **10**, 122–142 (1995).

12. Fett, A.-K. J. *et al.* To trust or not to trust: the dynamics of social interaction in psychosis. *Brain* **135**, 976–984 (2012).

13. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews* **42**, 9–34 (2014).

14. Van Overwalle, F. Social cognition and the brain: A meta‐analysis. *Human Brain Mapping* **30**, 829–858 (2009).

15. Carter, R. M., Bowling, D. L., Reeck, C. & Huettel, S. A. A Distinct Role of the Temporal-Parietal Junction in Predicting Socially Guided Decisions. *Science* **337**, 109–111 (2012).

16. Jáni, M. & Kašpárek, T. Emotion recognition and theory of mind in schizophrenia: A meta-analysis of neuroimaging studies. *The World Journal of Biological Psychiatry* **19**, S86–S96 (2018).

17. Miall, R. C. & Wolpert, D. M. Forward Models for Physiological Motor Control. *Neural Networks* **9**, 1265–1279 (1996).

18. Schultz, W. & Dickinson, A. Neuronal Coding of Prediction Errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).

19. Friston, K. & Kiebel, S. Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 1211–1221 (2009).

20. Schultz, W. Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience* **18**, 23–32 (2016).

21. Park, B., Fareri, D., Delgado, M. & Young, L. The role of right temporoparietal junction in processing social prediction error across relationship contexts. *Social Cognitive and Affective Neuroscience* **16**, 772–781 (2021).

22. Lemmers-Jansen, I. L. J., Fett, A.-K. J., Hanssen, E., Veltman, D. J. & Krabbendam, L. Learning to trust: social feedback normalizes trust behavior in first-episode psychosis and clinical high risk. *Psychol. Med.* **49**, 780–790 (2019).

23. Yaple, Z. A., Tolomeo, S. & Yu, R. Abnormal prediction error processing in schizophrenia and depression. *Hum Brain Mapp* **42**, 3547–3560 (2021).

24. Chung, Y. S. & Barch, D. M. Frontal-striatum dysfunction during reward processing: Relationships to amotivation in schizophrenia. *Journal of Abnormal Psychology* **125**, 453–469 (2016).

25. Millard, S. J., Bearden, C. E., Karlsgodt, K. H. & Sharpe, M. J. The prediction-error hypothesis of schizophrenia: new data point to circuit-specific changes in dopamine activity. *Neuropsychopharmacol.* **47**, 628–640 (2022).

26. Corlett, P. R. *et al.* Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. *Brain* **130**, 2387–2400 (2007).

27. Griffiths, O., Langdon, R., Le Pelley, M. E. & Coltheart, M. Delusions and prediction error: re-examining the behavioural evidence for disrupted error signalling in delusion formation. *Cognitive Neuropsychiatry* **19**, 439–467 (2014).

28. Howes, O. D. & Nour, M. M. Dopamine and the aberrant salience hypothesis of schizophrenia. *World Psychiatry* **15**, 3–4 (2016).

29. Heinz, A. & Schlagenhauf, F. Dopaminergic Dysfunction in Schizophrenia: Salience Attribution Revisited. *Schizophrenia Bulletin* **36**, 472–485 (2010).

30. Kapur, S. Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia. *AJP* **160**, 13–23 (2003).

31. Fett, A.-K. J. *et al.* To trust or not to trust: the dynamics of social interaction in psychosis. *Brain* **135**, 976–984 (2012).

32. Alós-Ferrer, C. & Farolfi, F. Trust Games and Beyond. *Front. Neurosci.* **13**, 887 (2019).

33. Hartig, F. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. The R Foundation https://doi.org/10.32614/cran.package.dharma (2016).

34. R Core Team. R: A Language and Environment for Statistical Computing. (2023).

35. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using**lme4**. *J. Stat. Soft.* **67**, (2015).

36. Gromann, P. M. *et al.* Trust versus paranoia: abnormal response to social reward in psychotic illness. *Brain* **136**, 1968–1975 (2013).

37. Wisman‐van der Teen, A., Lemmers‐Jansen, I. L. J., Oorschot, M. & Krabbendam, L. Exploring the association between social behaviour, trust, and its neural correlates in first episode psychosis patients and in individuals at clinical high risk for psychosis. *British J Clinic Psychol* **61**, 629–646 (2022).

38. Saperia, S. *et al.* Reward-driven decision-making impairments in schizophrenia. *Schizophrenia Research* **206**, 277–283 (2019).

39. Betz, L. T. *et al.* Deciphering reward-based decision-making in schizophrenia: A meta-analysis and behavioral modeling of the Iowa Gambling Task. *Schizophrenia Research* **204**, 7–15 (2019).

40. Tzieropoulos, H. The Trust Game in neuroscience: A short review. *Social Neuroscience* **8**, 407–416 (2013).

41. Waltz, J. A. *et al.* Motivational Deficits in Schizophrenia Are Associated With Reduced Differentiation Between Gain and Loss-Avoidance Feedback in the Striatum. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **3**, 239–247 (2018).

42. Grossman, L. S., Harrow, M., Rosen, C., Faull, R. & Strauss, G. P. Sex differences in schizophrenia and other psychotic disorders: a 20-year longitudinal study of psychosis and recovery. *Comprehensive Psychiatry* **49**, 523–529 (2008).

43. Grossman, L. S., Harrow, M., Rosen, C. & Faull, R. Sex Differences in Outcome and Recovery for Schizophrenia and Other Psychotic and Nonpsychotic Disorders. *PS* **57**, 844–850 (2006).

44. Näätänen, R., Todd, J. & Schall, U. Mismatch negativity (MMN) as biomarker predicting psychosis in clinically at-risk individuals. *Biological Psychology* **116**, 36–40 (2016).

45. Kirihara, K. *et al.* A Predictive Coding Perspective on Mismatch Negativity Impairment in Schizophrenia. *Front. Psychiatry* **11**, (2020).

46. Kim, J. S. *et al.* Mismatch Negativity Indices as a Prognostic Factor for Remission in Schizophrenia. *Clin Psychopharmacol Neurosci* **18**, 127–135 (2020).

47. Kim, M., Lee, T. H., Yoon, Y. B., Lee, T. Y. & Kwon, J. S. Predicting Remission in Subjects at Clinical High Risk for Psychosis Using Mismatch Negativity. *Schizophrenia Bulletin* **44**, 575–583 (2018).

48. Erickson, M. A., Ruffle, A. & Gold, J. M. A Meta-Analysis of Mismatch Negativity in Schizophrenia: From Clinical Risk to Disease Specificity and Progression. *Biological Psychiatry* **79**, 980–987 (2016).

49. Yang, X., Song, Y., Zou, Y., Li, Y. & Zeng, J. Neural correlates of prediction error in patients with schizophrenia: evidence from an fMRI meta-analysis. *Cerebral Cortex* **34**, (2024).

50. Stuke, H., Weilnhammer, V. A., Sterzer, P. & Schmack, K. Delusion Proneness is Linked to a Reduced Usage of Prior Beliefs in Perceptual Decisions. *Schizophrenia Bulletin* (2018) doi:10.1093/schbul/sbx189.

51. Haarsma, J. *et al.* Influence of prior beliefs on perception in early psychosis: Effects of illness stage and hierarchical level of belief. *Journal of Abnormal Psychology* **129**, 581–598 (2020).

52. Fletcher, P. C. & Frith, C. D. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci* **10**, 48–58 (2009).

53. Sterzer, P. *et al.* The Predictive Coding Account of Psychosis. *Biological Psychiatry* vol. 84 634–643 (2018).

54. Leggio, M. & Molinari, M. Cerebellar Sequencing: a Trick for Predicting the Future. *Cerebellum* **14**, 35–38 (2015).

Supplementary Material

Model Diagnostics

Binary Model

Model diagnostics were performed using the DHARMa package33. Simulated residuals showed no significant deviation from uniformity (KS test: p = 0.77), no overdispersion (p = 0.96), and no evidence of outlier inflation (p = 0.19). Residuals were homogeneously distributed across the range of predicted values, indicating that model assumptions were adequately met (see Supplementary Material). Multicollinearity among fixed effects was assessed using generalized variance inflation factors (GVIF). All adjusted GVIF values were well below 2 (range: 1.00–1.46), indicating no evidence of multicollinearity among predictors.

The intraclass correlation coefficient (ICC) was calculated to assess the proportion of variance in investment decisions attributable to between-subject differences. The ICC was approximately 0.167, indicating that 16.7% of the variance was due to differences between participants, thereby justifying the inclusion of a random intercept in the model.

A comparison of a diagram

AI-generated content may be incorrect.

PE model

To account for repeated measures and individual differences in the task, a random intercept for subject was included in the linear mixed-effects model. As with the Binomial model, all adjusted GVIF values were well below 2 (range: 1.00–1.46), indicating no evidence of multicollinearity among predictors. The intraclass correlation coefficient (ICC = 0.053) indicated that approximately 5.3% of the variance in log-transformed prediction error was attributable to between-subject variability, supporting the inclusion of a random effect for Subject.

A diagram of a model

AI-generated content may be incorrect.

Alfa Model

Residual diagnostics indicated that the model adequately captured the variance structure in the data. The dispersion test was non-significant (p = .92), suggesting no evidence of over- or underdispersion. However, the Kolmogorov-Smirnov (KS) test indicated a significant deviation from uniformity (p < .001), and the outlier test revealed a mild excess of extreme residuals (p = .022), suggesting the presence of a small number of influential observations (Supplementary Figure). Despite these minor deviations, the overall residual pattern remained acceptable for inferential purposes.

The intraclass correlation coefficient (ICC = 0.16) indicated that approximately 16% of the variance in learning rate (α) was attributable to between-subject differences, supporting the inclusion of a random intercept for Subject. To assess multicollinearity, generalized variance inflation factors (GVIFs) were computed using a linear model including only fixed effects. All adjusted GVIF were well below 2.0, with values ranging from 1.00 to 1.45, indicating no problematic collinearity among predictors.

A graph of a test results

AI-generated content may be incorrect.

A graph with black lines and white text

AI-generated content may be incorrect.

**Figure 3**. *Odds ratios for investment decisions from a binomial GLMM*. Dots represent estimated odds ratios and horizontal lines indicate 95% confidence intervals. The model included fixed effects for diagnostic group (schizophrenia vs. healthy control), sex, task phase (Phase 2 and Phase 3 each compared to Phase 1), and duration of illness (in years), plus a random intercept for participants. The odds ratio for groups (SZ vs. HC) was 0.85 (p = .590), indicating no reliable difference in investment odds. Sex (male vs. female) had an OR of 1.05 (p = .820), showing no significant effect. Investments were significantly lower in Phase 2 compared to Phase 1 (OR = 0.76, 95% CI [0.65, 0.89], p < .001), while Phase 3 did not differ from Phase 1 (OR = 1.03, p = .750). Duration of illness was not associated with investment likelihood (OR = 0.99 per year, p = .503). Only the shift from Phase 1 to Phase 2 significantly altered investment odds.

A graph with black lines and white text

AI-generated content may be incorrect.

**Figure 4**. *Multiplicative effects for prediction error (PE) from a linear mixed‐effects model on PE*. Dots represent estimated multiplicative effects and horizontal lines indicate 95% confidence intervals. The model included fixed effects for diagnostic group (schizophrenia vs. healthy control), sex (male vs. female), task phase (Phase 2 and Phase 3 each compared to Phase 1), the group × sex interaction, and duration of illness (DoI, in years), plus a random intercept for participants. The schizophrenia group showed a non-significant 4% reduction in PE compared to controls (multiplicative effect = 0.96, 95% CI [0.904, 1.026], p = .22). Male participants exhibited an 8% higher PE than females (multiplicative effect = 1.08, 95% CI [1.028, 1.124], p = .001). A significant group × sex interaction indicated that males with schizophrenia had 12% higher PE than expected additively (multiplicative effect = 1.12, 95% CI [1.044, 1.207], p = .001). Neither Phase 2 nor Phase 3 differed significantly from Phase 1, and each additional year of illness was associated with a 0.3% increase in PE (multiplicative effect = 1.003, 95% CI [1.000, 1.006], p < .05). These results suggest that elevated prediction errors are driven specifically by male patients with schizophrenia and are not modulated by task phase.

A graph with black lines and white text

AI-generated content may be incorrect.

**Figure 5.** *Linear mixed-effects model predicting raw learning rate ()*. Relative to healthy control females in Phase 1, schizophrenia females showed a 1.10-fold increase in learning rate (95% CI [1.07, 1.14], p < .001), whereas healthy control males showed a 0.91-fold rate (95% CI [0.88, 0.94], p < .001). The group × sex interaction revealed that male patients with schizophrenia exhibited the lowest learning rates, with a 0.90-fold effect relative to additive expectations (95% CI [0.86, 0.95], p < .001). Task Phase 2 (1.01-fold, 95% CI [0.98, 1.03], p = .62) and Phase 3 (1.01-fold, 95% CI [0.98, 1.04], p = .79) did not differ significantly from Phase 1, and each additional year of illness was associated with a 0.99-fold change (95% CI [0.98, 1.00], p = .111). These findings demonstrate that learning-rate alterations in schizophrenia are sex-specific and independent of task phase or illness duration.