Received: 7 June 2022 | Revised: 11 January 2023 | Accepted: 21 January 2023
DOI: 10.1002/hbm.26223

RESEARCH ARTICLE

WILEY

Intention affects fairness processing: Evidence from behavior and representational similarity analysis of event-related potential signals

Qiang Xu[1] | Jiali Hu[1] | Yi Qin[1] | Guojie Li[1] | Xukai Zhang[2] | Peng Li[1]

[1]Brain Function and Psychological Science Research Center, Shenzhen University, Shenzhen, China

[2]Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

Correspondence
Peng Li, Brain Function and Psychological Science Research Center, Shenzhen University, No 3688, Nanhai Road, Nanshan, Shenzhen 518060, China.
Email: peng@szu.edu.cn

Abstract

In an ultimatum game, the responder must decide between pursuing self-interest and insisting on fairness, and these choices are affected by the intentions of the proposer. However, the time course of this social decision-making process is unclear. Representational similarity analysis (RSA) is a useful technique for linking brain activity with rich behavioral data sets. In this study, electroencephalography (EEG) was used to measure the time course of neural responses to proposed allocation schemes with different intentions. Twenty-eight participants played an ultimatum game as responders. They had to choose between accepting and rejecting the fair or unfair money allocation schemes of proposers. The schemes were offered based on the proposer's selfish intention (monetary gain), altruistic intention (donation to charity), or ambiguous intention (unknown to the responder). We used a spatiotemporal RSA and inter-subject RSA (IS-RSA) to explore the connections between event-related potentials (ERPs) after offer presentation and intention presentation with four types of behavioral data (acceptance, response time, fairness ratings, and pleasantness ratings). The spatiotemporal RSA results revealed that only response time variation was linked with the difference in ERPs at 432–592 ms after offer presentation on the posterior parietal and prefrontal regions. Meanwhile, the IS-RSA results found a significant association between inter-individual differences in response time and differences in ERP activity at 596–812 ms after the presentation of ambiguous intention, particularly in the prefrontal region. This study expands the intention-based reciprocal model to the third-party context and demonstrates that brain activity can represent response time differences in social decision-making.

KEYWORDS
EEG, intention, RSA, unfairness aversion

1 | INTRODUCTION

Fairness is a fundamental tenet of human social behavior and resource allocation, and people generally show a strong preference for an equal—rather than unequal—distribution of resources (Kahneman et al., 1986; Polezzi et al., 2008). Research on fairness often involves the classical ultimatum game (UG) task, wherein a player (the proposer) divides a monetary amount between the two players, and the other player (the responder) can choose to either accept the proposer's offer or reject it (Güth et al., 1982). The responder's rejection would mean that both players receive nothing. As such, a rejection by the responder is often thought to reflect an aversion to inequity (Güroğlu et al., 2010). Prior studies found that unfair treatment elicits a negative emotion in the proposer, and they struggle to balance self-interest and unfairness aversion (Gabay et al., 2014). Several psychological processes are reflected in a range of behavioral indicators. For example, rejection behavior in the responder has been proposed to reflect resistance to inequity (Falk et al., 2003); the response time (RT) captures the cognitive processes that accompany the struggle between the player's payoff and a fair distribution (Rubinstein, 2007), fairness ratings reflect the player's perception of how fair the distribution scheme was (Knoch et al., 2006), and the ratings of pleasantness are associated with the negative emotional experiences arising from the offer (Aoki et al., 2015). These behavioral indicators reflect different aspects of inequity aversion and can vary independently. For example, previous research has found that a disruption of the left dorsolateral prefrontal cortex (DLPFC) reduced a player's rate of rejecting their partner's unfair offers but did not affect their perception of the unfairness of the offers (Knoch et al., 2006).

Intentions of others are critical for certain social decision-making, such as moral judgment, judicial decision-making, and social interaction (Frith & Frith, 1999; Buckholtz et al., 2008; Young et al., 2007), but researchers disagree on its influence on fairness decisions. The equity aversion model maintains that only unfairly distributed outcomes affect behavior; however, it does not account for intention (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999). On the contrary, the intention-based reciprocity model reveals that intention plays a major role in decision-making. Past research has suggested that people reciprocate perceived (un)benevolent intentions with (un)benevolent behaviors (Falk & Fischbacher, 2006; Rabin, 1993). Thus, the first aim of this study was to examine the effect of intentions on people's inequity aversion from all aspects. The literature on this subject has primarily focused on the acceptance rate. Moreover, studies investigating the effect of intention on direct reciprocity have shown that good intentions do not result in worse outcomes for the responder compared to bad intentions (Güroğlu et al., 2010; Ma et al., 2015). However, the players' reactions in circumstances where there is a mismatch between the valence of intentions and the distribution outcomes have rarely been examined.

In this study, we considered an adapted UG paradigm, wherein the proposer can keep the money from the UG task or donate it to charity. We recorded the responders' acceptance rate, RT, fairness ratings, and pleasantness ratings for the different offers. The intention-based reciprocity model suggests that the responder's actions would be contingent upon the proposer's intention, whereas the inequity aversion model suggests that the responder's actions would be independent of the proposer's intention.

In addition to testing these predictions, we evaluated how the ambiguity of intentions affects a player's decision-making process. In most cases, it is not entirely possible to determine the intentions of others, whereas behavioral outcomes are much easier to observe. This can lead to an "outcome bias," that is, people who observe bad outcomes tend to think that the actor's decisions are of poor quality or that there are procedural errors (Mazzocco et al., 2004). Some studies indicate that people may also have an outcome bias for estimating intentions (Charness & Levine, 2007). For example, when people do not have access to complete information about others' intentions, they may be more likely to link good behavioral outcomes to good intentions and negative consequences to bad intentions. We tested these assumptions in this study.

Event-related potentials (ERPs) were widely applied to study the temporal progression of brain processing during inequity aversion. Fairness processing is generally thought to be associated with two ERP components: feedback-related negativity (FRN) and P3 (Ma et al., 2015). FRN is a negative component that appears at 250–350 ms over the frontocentral areas of the brain after feedback presentation (Miltner et al., 1997). It is commonly associated with reward prediction error (Holroyd & Coles, 2002; Sambrook & Goslin, 2015). P3 is a positive deflection that peaks at 300–600 ms after stimulus presentation, and its amplitude is maximal at the centro–parietal sites (Luo et al., 2014). Typically, researchers compare these ERP components by calculating the differences between average group responses to different kinds of feedback. While ERP components can disclose information about conditional changes in timing or amplitude caused by various forms of feedback, standard ERP components depend on averages from certain electrodes and time windows that are arbitrarily selected by researchers across different studies. Moreover, the functional significance of these components remains nebulous and can be altered by experiment-specific factors (Donchin et al., 1978).

Unlike traditional ERP component methods, a new multivariate pattern analysis—representational similarity analysis (RSA)—can be used to calculate the similarity between different representations in the brain under different stimuli or experimental conditions (Kriegeskorte et al., 2008). The objective of RSA is to identify activation patterns that reliably differentiate between stimuli or experimental conditions. If differences in patterns of neural activation can be reliably mapped to differences in the stimulus's properties or differences in the behavior elicited by the stimulus, then we can infer that these temporal or spatial neural patterns support the discrimination of the stimulus's properties or drive the corresponding behavior. Compared with fMRI data, scalp-recorded ERP data have the advantages of high temporal resolution and low cost, and the RSA of these data allows us to reveal the rapid neurodynamic processes underlying cognition. Although the RSA approach is well suited to social neuroscience research (Popal et al., 2019), its usefulness for ERP data in this field is still limited. Recently, RSA methods for ERP data have become popular (He et al., 2022; Hu et al., 2022; Kiat et al., 2022). For example, Kiat et al. (2022) examined the temporal course of the brain representation of saliency and high-level meaning-based representations by applying RSA to ERP responses to natural scenes. They found that the link between the saliency-based representational space and ERP data first appears slightly earlier (at about 80 ms) than the link in the significance-based representation space (which appears at about 100 ms). RSA can be used for both single-trial ERP data and averaged

ERP data, but averaging ERP data across conditions is useful for increasing the signal-to-noise ratio of the ERP data. For example, He et al. (2022) performed a single-trial RSA analysis and found a much weaker representative similarity than average data. Therefore, in our study, RSA was applied to average ERP waveforms rather than to single-trial data.

The second objective of our study was to use RSA to investigate the temporal dynamics of fairness decision-making processes related to intentions. Therefore, we recorded the ERP activity for offers received from proposers with different intentions in a UG. To make a final decision, the responders had to consider the fairness of the distribution scheme while taking intention into account. Through the RSA of this ERP data, we examined whether the representations of the ERP activity correlated with the diverse behavioral representations (acceptance rate, RT, fairness rating, and pleasantness rating) for each time point. As discussed earlier, the FRN and P3 components are related to outcome processing for the reward allocation scheme. FRN may reflect the prediction error associated with violating social norms and the associated negative emotions in the ultimatum game, such as unfair aversion (Zheng et al., 2017); meanwhile, the P3 reflects the accumulation of evidence in a decision-making task (O'Connell et al., 2012). Based on these studies, we predicted that the link between the emotion ratings representation and the ERP representation would appear during the early stage (250–350 ms) and that the link between the RT and the ERP data would become evident after that (350–600 ms). Decision-making involves the evaluation of options, with final decisions requiring the combination of accumulated evidence from different brain regions and timescales (Lin et al., 2020; Pinto et al., 2022). As a result, the decision to accept or reject something should take place after the accumulation of evidence. Therefore, we predicted that the link between acceptance and the ERP data would emerge after the link between RT and ERP data (>350–600 ms).

In addition, the inter-subject RSA (IS-RSA), which combines inter-subject correlation (ISC; Hasson et al., 2004) and RSA methods, allows us to explore the connection between brain activity and behavioral performance by individual differences. Unlike the general RSA, which highlights the connections between the differences across various experimental stimuli and ERP activity, the IS-RSA is concerned with whether differences in the responses of different individuals to the same stimuli respond to differences in their brain activity. There were large individual differences in the participants' responses to ambiguous stimuli (Kim et al., 2022). Therefore, in our experiment, we focused on whether differences in the decision-making behavior of different participants regarding ambiguous intention were related to differences in ERP activity when the proposer's intention was presented. Previous studies have shown that ERP activity in frontal and parietal scalp regions during a later time window (from 500 to 1000 ms) after experimental stimulus presentation reflects specific cognitive processes for judging others' intentions or beliefs (Geangu et al., 2013; Liu et al., 2009; McCleery et al., 2011), Therefore, we predict that the connection between ERP activity and behavior should occur within a time window later than 500 ms.

2 | METHODS

2.1 | Participants

Thirty-one neurologically and psychiatrically healthy volunteers were recruited from Shenzhen University. Three participants were excluded because of excessive recording artifacts. The data of the remaining 28 participants (mean = 19.07 years, SD = ±1.02 years; 11 females) were analyzed. All participants were right-handed, had a normal or corrected-to-normal vision, and signed an informed consent form. All the participants received 50 Chinese Yuan (¥50) for their participation, along with the additional money (0–10 Yuan) acquired during the UG. The experiment was approved by the Medical Ethical Committee of Shenzhen University.

2.2 | Experimental procedure

The participants arrived at the lab where they were asked to read the instructions for the experiment. They were asked to play a UG with the other participants, who would divide a given amount of money as part of the game (e.g., ¥10). In our modified version of the UG, the responders were also briefed about a charity project, including its name, primary goal, beneficiaries, and an example of its activities. The responders were then told that the proposer could decide to either keep the money received in the subsequent UG for themselves or donate it to charity. In the first case, any money the proposer received in the subsequent UG would be added to their total payoff and kept by themselves; otherwise, the money is also handed to the proposer, but the proposer must donate it to charity. Given this information, the proposer put forward an allocation scheme for the money (e.g., both the proposer and the responder receive ¥5 from the initial ¥10). The allocation outcome was revealed to the responder, but the intention (keeping for oneself or donating to charity) was not always revealed. The responder could accept the offer or reject it; acceptance would result in both the proposer and the responder gaining the amount they were proposed to receive, whereas rejection would lead to neither player receiving anything. All participants acted as responders in the UG and were informed that their proposers' offers had been collected in the previous experiment. To exclude the possible influence of sex on fairness consideration, the proposer's gender was not shown.

Figure 1 shows that in each trial, the identification number of the proposer was initially presented (duration: 1500 ms), followed by a blank screen (duration: 1000–1200 ms). Next, the proposer's intention, represented by a yellow ball, appeared on the screen. A yellow ball in the box above the letter "S" (meaning "self") indicated that the proposer had chosen to keep their share of the money (selfish intention condition). A yellow ball in the box above the letter "C" (meaning "charity") indicated that the proposer had chosen to donate their share of the money to charity (altruistic intention condition). A yellow ball in the middle of the two boxes indicated that the proposer's choice was unknown to the responder (ambiguous intention

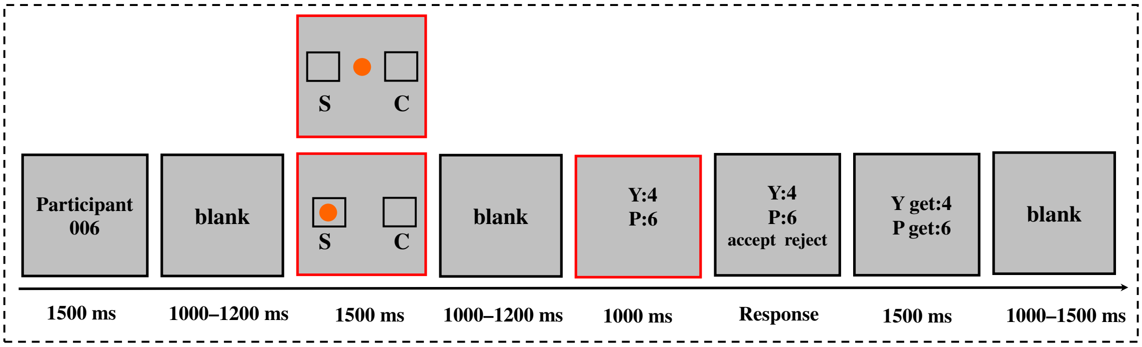

FIGURE 1 Experimental procedure for a single trial. The participants (responders) are shown the identification number of the proposer, followed by their choice regarding the money (whether to keep it for themselves or donate it to charity). In the ambiguous condition, the choice is not displayed. After seeing the offer made by the proposer, the participants make their choice by pressing "F" to accept the offer or "J" to reject it. The final results are displayed at the end of the trial. "Y" indicates the responder (participants), and "P" indicates the proposer (partner). The red borders represent instances where the EEG activity was recorded.

condition). Following the display of another blank screen (duration: 1000–1200 ms), the proposer's allocation scheme for the money was displayed to the participant. The letters "Y" and "P" were used to denote the participant and proposer, respectively. After 1000 ms, a decision cue was presented below the offer, and the participant (responder) was given an option between pressing the "F" key to accept the offer or the "J" key to reject it, without any time restriction. The correspondence between the accept/reject decision and pressing the "F/J" key was counterbalanced between subjects. Once the participant had decided, the payouts for the proposer and responder were presented for 1500 ms, and the trial ended. The subsequent trial began after an interval of 1000–1500 ms.

This experiment included 10 practice trials and 300 formal trials divided into three blocks which lasted ~40 min in total. It included five types of offers with different levels of fairness (two fair offers: 6/4 and 5/5; two unfair offers: 9/1 and 2/8; one filler offer: 7/3). Prior researchers suggested that UG offers were considered unfair if they were <3 out of 10 yuan and fair if ≥4 out of 10 yuan, while 3 out of 10 was difficult to classify as "fair" or "unfair." Therefore, 4/6 was considered a fair offer, and 7/3 was often used as a filler trial (Güth, 1995; Henrich et al., 2006; Peterburs et al., 2017). In our study, we followed this setting. The offers of 6/4, 5/5, 9/1, and 2/8 were each repeated 72 times, and the filler offer of 7/3 was repeated 12 times. The proposer chose to keep the money for themselves in one-third of the trials and to donate it to charity in another one-third of the trials. The proposer's choice was ambiguous in the remaining one-third of the trials. Therefore, we had a three (intention: selfish vs. altruistic vs. ambiguous) by two (offer fairness: fair vs. unfair) design with the following six experimental conditions: selfish–fair, altruistic–fair, ambiguous–fair, selfish–unfair, altruistic–unfair, and ambiguous–unfair.

After the participants had completed all the trials, a trial was randomly selected as the final result. The participants received a ¥50 base fee along with the money they received from the UG. Finally, the participants rated both the pleasantness and fairness of each experimental condition on a 7-point Likert scale. (1 = very unfair, 7 = very fair or 1 = very unpleasant, 7 = very pleasant).

We performed a two-way repeated-measures analysis of variance (ANOVA) to compare the acceptance rates, RTs, fairness ratings, and pleasantness ratings across six conditions: 3 (intention: selfish vs. altruistic vs. ambiguous) × 2 (offer fairness: fair vs. unfair). The Greenhouse–Geisser correction was applied if the sphericity assumption for the ANOVA was violated. We used Bonferroni corrections for our post hoc analyses.

2.3 | Electroencephalogram recordings and data preprocessing

The electroencephalographic (EEG) data were recorded by Brain Products System (Brain Products GmbH, Munich, Germany) at a sampling rate of 1000 Hz with a pass band of 0.01–100 Hz. The 64 electrodes are arranged in accordance with the international 10–20 system. The ground electrode was placed on the medial frontal line between Fz and FPz. The recordings were referenced online to the FCz mastoid and referenced offline to the average of the left and right mastoids. The vertical electrooculogram was recorded with one electrode located below the right eye. All electrode resistances were kept below 10 kΩ.

The electrophysiological signals were analyzed using MATLAB (R2020b; MathWorks). These signals were filtered with a band-pass filter of 0.1–30 Hz. To correct electrooculographic artifacts, an independent component analysis procedure (Jung et al., 2001) was employed. The EEGs were segmented into 1200 ms periods (200 ms before and 1000 ms after the offer onset) and were baseline corrected by subtracting the mean prestimulus signal (200–0 ms before the offer onset). Any artifacts in any channel exceeding ±120 μV were removed. Finally, the EEG epochs of the offer presentation were separately averaged for the selfish–fair, altruistic–fair, ambiguous–fair, selfish–unfair, altruistic–unfair, and ambiguous–unfair conditions.

2.4 | Spatiotemporal RSA

For the ERPs of the offer presentation, we used a spatiotemporal RSA to explore whether—and at what time points—it was possible to discriminate between the six experimental conditions based on neural activity. First, we constructed the neural representational dissimilarity matrix (RDM) and downsampled the EEG signal to 250 Hz. For each participant, we extracted the average ERP amplitude after reducing the noise through a noise normalization procedure (Guggenmos et al., 2018). The spatial features included scalp voltages from 60 channels, and the temporal features were chosen from the epoched EEG data utilizing a 100 ms (25 time points) sliding window. The step size of the sliding window was one time point (Lu et al., 2015). In each channel, a data vector was extracted from the 25 time points as the spatiotemporal pattern of ERP activity of the middle time point. We then calculated the dissimilarity of the spatiotemporal pattern between all the possible pairs of conditions (e.g., selfish_fair and selfish_unfair, selfish_fair and altruistic_fair, and so on) at each channel and per time point (t). To do this, we computed the (1-Spearman's *r*) values between the vectors, and then created the final 6×6 neural RDM per channel and per time point for each pair.

During the next step of our analysis, we constructed RDMs from the behavioral data. For each participant, we calculated the absolute differences in the acceptance rate, RT, fairness ratings, and pleasantness ratings for all the possible pairs of conditions to create four 6×6 behavior RDMs. All the behavioral data were normalized to the Z score before subtraction. The neural and behavioral RDMs were then compared. The diagonals and the lower triangular part of the RDMs were omitted (Ritchie et al., 2017). For each participant, the four behavioral RDMs were quantitatively compared with the neural RDMs per time point and per channel using Spearman's correlation analysis. The correlation coefficients were transformed into Fisher's *Z*-scores for further analysis.

Finally, we evaluated the four behavioral models to deduce which of them could be used to explain the brain's electrical activity in a specific brain representation. For each participant, we obtained a similarity time series at each channel and stored the data in a matrix. As a baseline group, we generated an all-zero matrix of the size of (subject $\times$ channels $\times$ time point). A nonparametric cluster-based permutation test was used to search for significant correlations between the neural and behavior patterns relative to the baseline. The procedure was applied in the FieldTrip toolbox (Oostenveld et al., 2011) as follows: (1) The Fisher's *Z*-scores at each time point in each channel were tested against an equally sized baseline group by dependent-sample *t*-tests. (2) The time points and channels whose statistical values exceeded a threshold ($p = .05$) were selected and combined into a spatiotemporal cluster based on spatial and temporal adjacency. (3) The sum of the t-values inside each detected cluster was used to construct cluster-level statistics. (4) The distribution of the cluster-level statistics under the null hypothesis was approximated using 10,000 random draws. (5) The uncorrected *t*-test value was then compared against the permuted distributions, and any clusters that exceeded a one-tailed significance threshold ($p = .05$) were considered significant. The significant clusters were determined for the first and last time points in at least three channels (Wang et al., 2020). To improve the spatial resolution, channels containing at least 15 adjacent time points were selected. A noise ceiling was computed based on the degree of noise in the data. Noise ceilings indicate how well the best possible model would perform and consists of a range that has both a lower bound and an upper bound. For the upper bound, we calculate the correlation between each participant's RDM with the group-averaged RDM, which included all participants. The lower bound estimate calculates the correlation between each participant's RDM with the group-averaged RDM, which included all other participants after excluding the participant's own data (Nili et al., 2014).

2.5 | Inter-subject RSA

The RSA was performed using a custom script written for our study. For the ERPs of the ambiguous intention presentation, we used an IS-RSA to explore whether the individual differences in brain activity for ambiguous intention were linked to individual differences in behavioral performance. First, we constructed the neural representational dissimilarity matrix (RDM) and downsampled the EEG signal to 250 Hz. For each participant, we extracted the average ERP for ambiguous intention after reducing the noise using a noise normalization procedure (Guggenmos et al., 2018). The voltage maps at each time point were first redefined as 5-point average maps using a moving window (window size: 20 ms; window center moving in 1-time-point steps) (Cecere et al., 2017). Next, we separately calculated (1-Spearman's *r*) between all possible pairwise participants' scalp distributions at each time point to generate a 28×28 subject correlation matrix to estimate the dissimilarity of individual variability in brain activity in the ambiguous intention condition in each time point.

Second, we constructed RDMs from the behavioral data. For the ambiguous intention condition, we calculated the absolute differences in the acceptance rate, RT, fairness ratings, and pleasantness ratings for all possible subject pairs to create four 28×28 behavior RDMs. All the behavioral data were normalized to the Z score before subtraction. The correlations between the neural and behavioral RDMs at each time point were then calculated for the lower triangle of the matrix using Spearman's correlation analysis and the significance levels of the correlation were assessed using a Mantel permutation test (Mantel, 1967; Nummenmaa et al., 2012). The correlation coefficients were transformed into Fisher's *Z*-scores for further analysis.

Finally, we corrected for multiple comparisons using cluster-based correction. Adjacent time points with significance levels exceeding the threshold ($p < .05$) were selected and combined into a cluster. The sum of the z-values inside each detected cluster was used to construct cluster-level statistics. For each time point in the cluster, we recomputed the Spearman correlation coefficient (Fisher's *Z*-scores) between the neural RDM and behavioral RDM after shuffling the order of the observations in one of the two matrices and then summed these correlation coefficients. This procedure was repeated

10,000 times to generate a null distribution of the correlation cluster sum. We then sorted these values and determined the 0.95 quartiles. Next, we used this quartile as our threshold based on clustering and compared the sum of each original cluster to this threshold; any clusters that exceeded the significance threshold were considered significant. We then mapped the topography of the significant time clusters following the leave-one-out method developed by Kiat et al. (2022). In IS-RSA, the RDM was constructed by inter-individual differences among participants; there was no separate RDM for each participant. Therefore, the method for calculating the noise ceiling is inapplicable to IS-RSA.

3 | RESULTS

3.1 | Behavioral results

Figure 2a demonstrates the average acceptance rates for the six conditions. A 3 (intention) × 2 (offer fairness) within-subjects ANOVA showed a significant main effect of intention (F [2, 54] = 41.46, $p < .001$, $\eta^2 = 0.60$). Post hoc analysis indicated that acceptance rates were significantly greater for altruistic intention (mean ± SE: 0.91 ± 0.03) than for ambiguous (0.67 ± 0.04; $p < .001$) or selfish (0.54 ± 0.04; $p < .001$) intentions. Moreover, the acceptance rates were significantly higher for ambiguous intention than for selfish intention ($p < .001$). The offer fairness had a significant main effect (F [1, 27] = 89.60, $p < .001$, $\eta^2 = 0.77$), and the acceptance rates of fair offers (0.94 ± 0.02) were significantly higher than those for unfair offers (0.47 ± 0.05). A significant interaction effect was observed between intention and fairness (F [2, 54] = 39.92, $p < .001$, $\eta^2 = 0.60$). The analysis of simple effects revealed that intention had a significant effect on unfair conditions (all $p < .001$). The acceptance rates were significantly higher for altruistic intention than for the other two types of intentions and also for ambiguous intention compared to selfish intention. However, this was not so in fair condition (all $p > .08$).

Figure 2b shows the averaged RT for the six conditions. An ANOVA of RT revealed a significant main effect of intention (F [2, 54] = 5.264, $p < .01$, $\eta^2 = 0.16$). The post hoc analysis indicated that the RTs were higher for ambiguous intention (1013 ± 66 ms) than for altruistic (895 ± 39 ms; $p = .083$) and selfish (886 ± 42 ms; $p = .063$) intentions and the differences were marginally significant. The difference in the RTs between altruistic and selfish intentions was not significant ($p = 1.00$). Offer fairness had a significant main effect on RT (F [1, 27] = 9.20, $p < .01$, $\eta^2 = 0.25$). The RTs were significantly longer for the unfair condition (991 ± 50 ms) than for the fair one (871 ± 46 ms). Additionally, the interaction of intention and fairness was not significant (F [2, 54] = 0.45, $p = .64$, $\eta^2 = 0.02$).

For the fairness rating, a two-way repeated-measures ANOVA revealed significant main effects of both intentions (F [2, 54] = 51.46, $p < .001$, $\eta^2 = 0.66$) and offer fairness (F [1, 27] = 125.68, $p < .001$, $\eta^2 = 0.82$). A post hoc comparison indicated that fairness ratings were significantly higher for altruistic intention (5.50 ± 0.20) than for ambiguous (4.05 ± 0.15; $p < .001$) or selfish (3.62 ± 0.15; $p < .001$) intentions, and there was a marginally significant difference in the fairness ratings between ambiguous and selfish intentions ($p = .052$). Participants gave a higher fairness rating for fair offers (5.79 ± 0.18) than for unfair offers (2.99 ± 0.18). The interaction between intention and offer fairness was also significant (F [2, 54] = 31.92, $p < .001$, $\eta^2 = 0.54$). An analysis of simple effects revealed that when the offer was fair, the fairness ratings were significantly higher for altruistic intentions than for selfish or ambiguous intentions (both $p < .01$), and

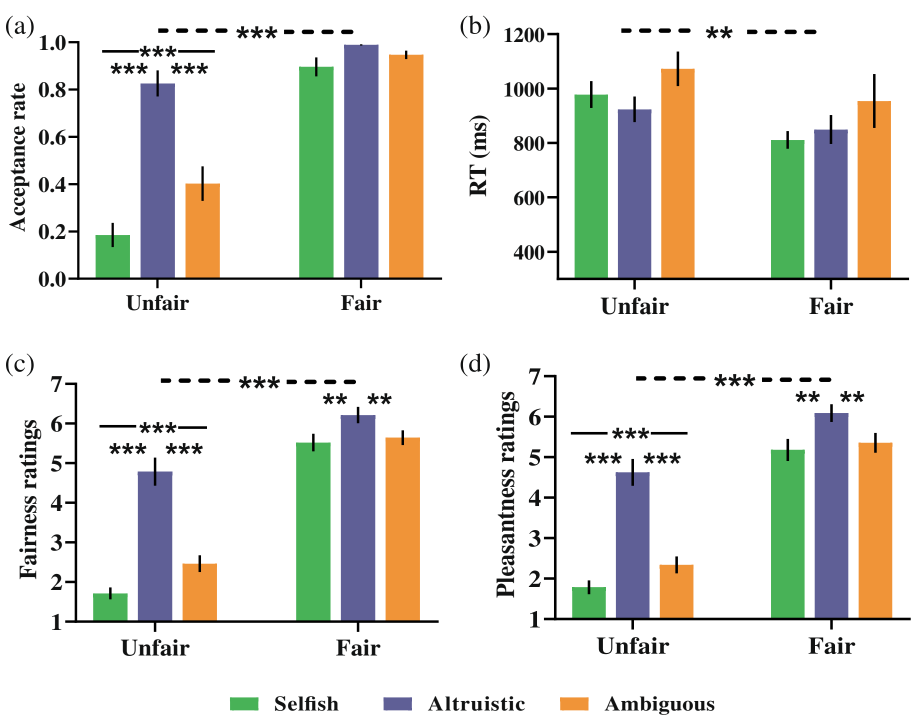

FIGURE 2 Behavioral results of the experiment. (a) Acceptance rates, (b) response times, (c) fairness ratings, and (d) pleasantness ratings of the six offer conditions. The error bars represent the standard errors of the means. RT = response time; * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

the difference in the fairness ratings between selfish and ambiguous intentions was insignificant ($p = 1.00$). In contrast, when the offer was unfair, the fairness ratings were significantly higher for altruistic intention than for selfish or ambiguous intentions (both $p < .01$). Moreover, the fairness ratings were higher for ambiguous intention (2.46 ± 0.21) than for selfish intention (1.71 ± 0.15) when the offer was unfair (Figure 2c).

For the pleasantness rating, a two-way repeated-measures ANOVA revealed significant main effects of both intentions (F [2, 54] $= 46.04$, $p < .001$, $\eta^2 = 0.63$) and offer fairness (F [1, 27] $= 84.66$, $p < .001$, $\eta^2 = 0.76$). A post hoc comparison indicated that pleasantness ratings were significantly higher for altruistic intention (5.36 ± 0.20) than for ambiguous (3.85 ± 0.16; $p < .001$) or selfish intentions (3.48 ± 0.18; $p < .001$). The difference in pleasantness ratings between ambiguous and selfish intentions was marginally significant ($p = .054$). The participants gave a higher pleasantness rating for fair offers (5.54 ± 0.22) than for unfair offers (2.92 ± 0.18). A significant interaction effect was observed between intention and fairness (F [2, 54] $= 27.79$, $p < .001$, $\eta^2 = 0.48$). The analysis of simple effects revealed that when the offer was fair, the fairness ratings were significantly higher for altruistic intentions than for selfish or ambiguous intentions (both $p < .01$), though the difference between selfish and ambiguous intentions was insignificant ($p = 1.00$). However, when the offer was unfair, the pleasantness ratings were significantly higher for altruistic intentions than for selfish or ambiguous intentions (both $p < .01$). Moreover, the pleasantness ratings were higher for ambiguous intention (2.34 ± 0.21) than for selfish intention (1.79 ± 0.17) when the offer was unfair (Figure 2d).

3.2 | Intention propensity

We used each participant's acceptance rate (AR) of unfair offers to describe their intention propensity when the intention was ambiguous. Intention propensity was calculated using the following formula:

$$\frac{(\text{ARambiguous} - \text{ARaltruistic}| - |\text{ARambiguous} - \text{ARselfish})}{(\text{ARambiguous} - \text{ARaltruistic}| + |\text{ARambiguous} - \text{ARselfish})}.$$

We obtained an intention propensity index with values ranging from -1 to 1 for 27 participants. One participant was excluded because he rejected all unfair offers. When the ARs for ambiguous and altruistic intentions are equal, the intention propensity index is -1, indicating a complete propensity for altruistic intentions; when the acceptance rates for ambiguous and selfish intentions are equal, the intention propensity index is 1, indicating a complete propensity for selfish intentions. When the acceptance rate for ambiguous intentions equals the average of those for altruistic and selfish intentions, the intention propensity index is 0, indicating no tendency and a neutral state. Larger absolute values of the intention propensity index indicate a more obvious tendency. As shown in Figure 3, most participants have an obvious tendency (absolute value >0.5).

3.3 | ERP waveforms

Figure 4a illustrates the grand-averaged ERP waveforms separately for the average of the six experimental conditions at the midline electrode sites. Based on the previous ERP studies on FRN and P3 (Ma et al., 2015; Polezzi et al., 2008) and a visual inspection of ERP waveforms, we measured the FRN and P3 components in the 250–350 ms and 350–550 ms time windows. As a traditional ERP analysis is not the focus of this study, the results are reported in the Data S1.

3.4 | Results of the RSA

Figure 5a illustrates the RDM constructed from the behavioral data. The RSA results revealed a significant correlation between the ERP data and RT (Figure 5b; 432–592 ms after stimulus onset), with a peak at 528 ms (Fisher Z: 0.08) after stimulus onset. Channels that significantly

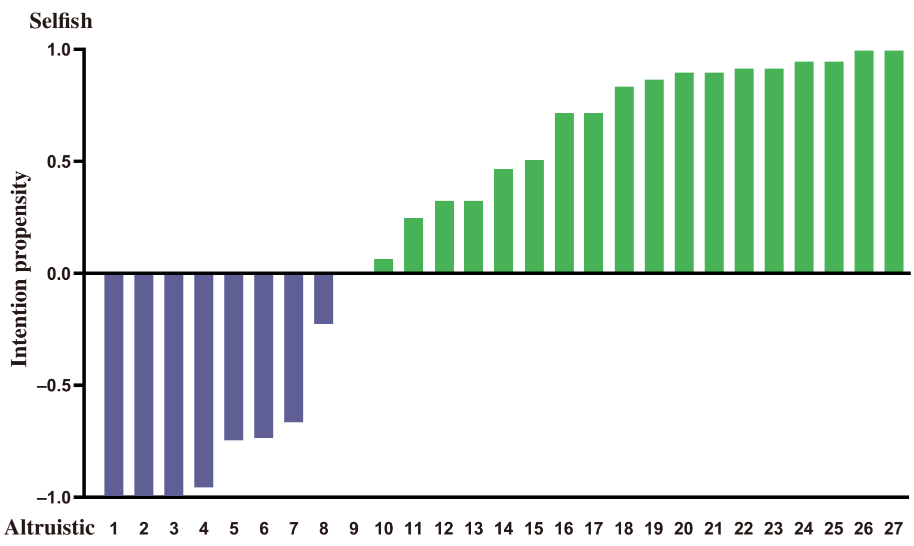

FIGURE 3 Intention propensity indexes of each participant for ambiguous intentions. 1 indicates a complete propensity for selfish intentions, and − 1 indicates a complete propensity for altruistic intentions.

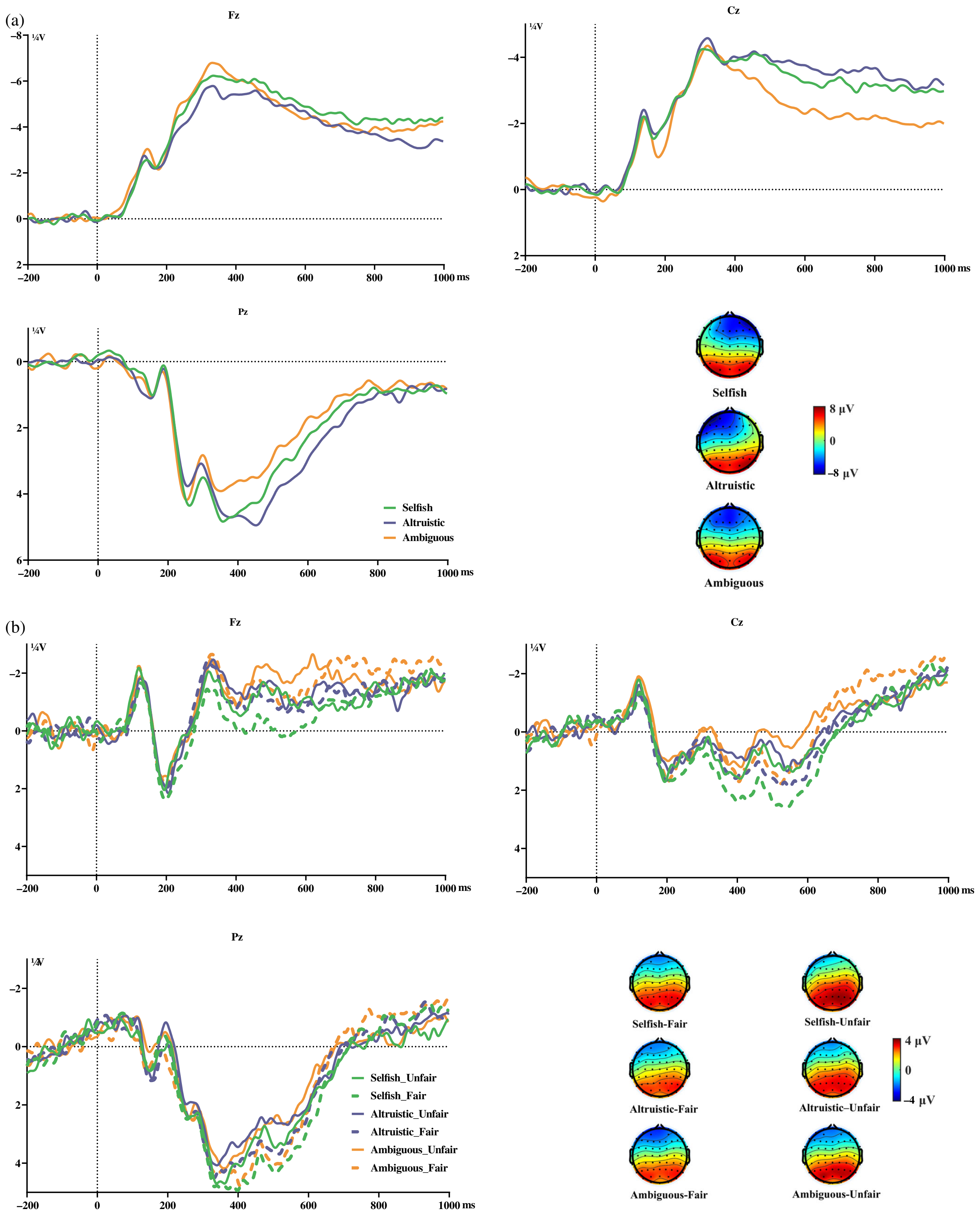

FIGURE 4 Event-related potential (ERP) waveforms. (a) Grand-averaged ERP waveforms at midline electrode sites for intention presentation and the topographic map of 596–812 ms window. (b) Grand-averaged ERP waveforms at midline electrode sites for offer presentation and the topographic map of 432–592 ms window.

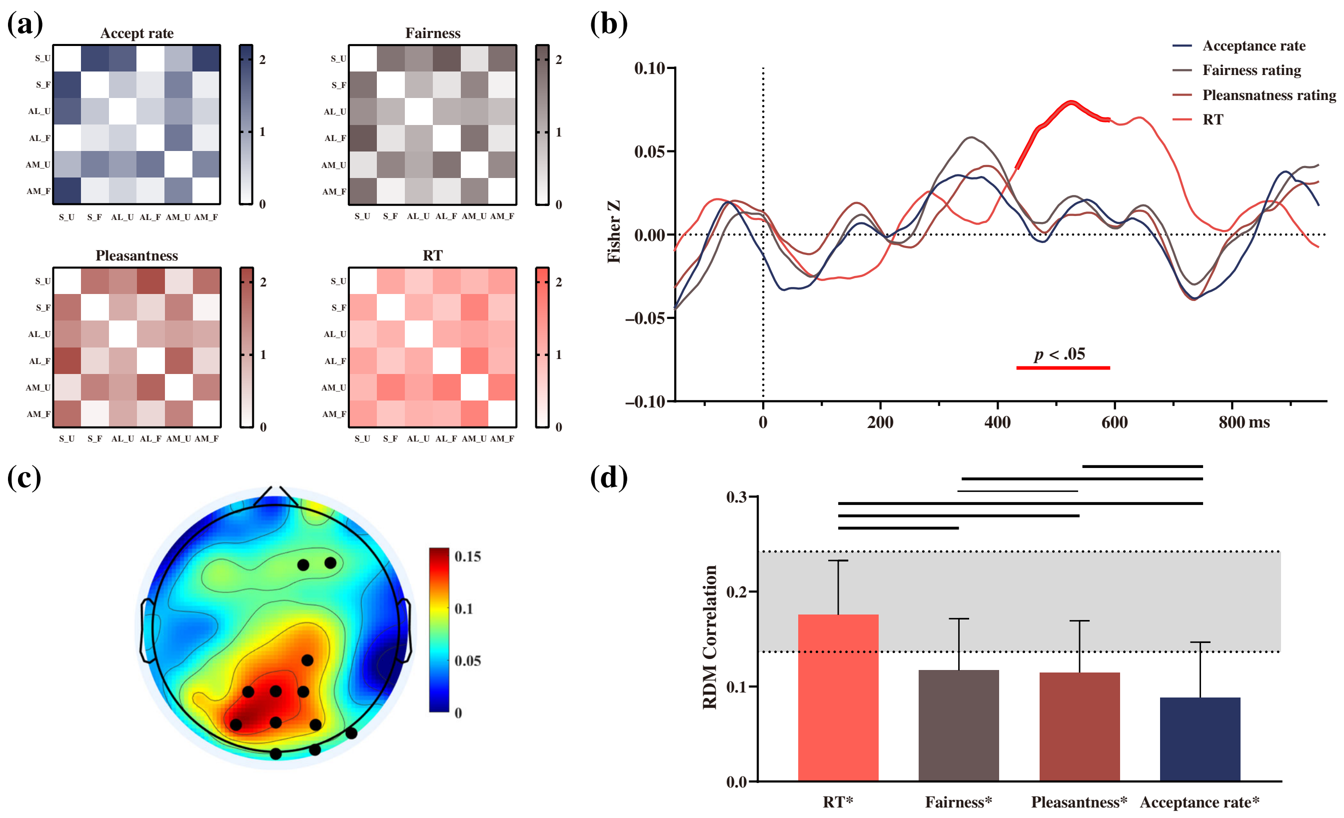

FIGURE 5 Results of the representational similarity analysis. (a) The four behavioral dissimilarity matrices. AL_F, altruistic_fair; AL_U, altruistic_unfair; AM_F, ambiguous_fair; AM_U, ambiguous_unfair; S_F, selfish_fair; S_U, selfish_unfair; (b) Similarity correlation between the four types of behavioral data and the ERP data; (c) The topographies of similarity correlation for RT_neural correlation in the 432–592 ms time windows; (d) Inferential comparisons of the four behavioral neural model representations. The upper and lower boundaries of the shaded region represent the upper and lower noise ceilings. Thick lines represent a significant difference between two representational dissimilarity matrices (RDMs), and the thin line represents a nonsignificant difference; RT = response time; * = $p < .05$.

exceeded the chance-level threshold in the permutation test were primarily in the posterior parietal regions in the early stage and spread to the prefrontal scalp in the late stage (Figure 5c). The noise ceiling analysis showed that only RT_RDM reached the noise ceiling. RT showed a high neural correlation and provided a greater level of explanation than other types of behavioral data in 432–592 ms (Figure 5d).

3.5 | Results of inter-subject RSA

Figure 6a illustrates the inter-subject RDM constructed from the behavioral data on ambiguous intention. The IS-RSA results revealed a significant correlation between the ERP data and RT (Figure 6b; 596–812 ms after intention onset), with a peak at 608 ms (Fisher Z: 0.24). As shown in Figure 6c, the results of the topography show that electrode sites in the prefrontal region have a greater effect on the similarity between ERP data and behavioral data, especially in the left prefrontal region. In addition, to verify whether there was greater behavioral variability in subjects' RT in the ambiguous intention condition, we calculated the overall RT dissimilarity of the three intention conditions. We first computed the average RT distance between each pair of participants. After obtaining the distance between each pair of participants, a pairwise subject-by-subject (28×28) inter-subject behavioral dissimilarity matrix was obtained. Then we calculated the average inter-subject dissimilarity of three discrete intentions to examine the overall dissimilarity. The results revealed that the ambiguous condition produced a larger inter-subject dissimilarity (376 ms) than selfish intention (240 ms) or altruistic intention (262 ms). This indicated that the inter-individual variation is relatively large under ambiguous conditions.

4 | DISCUSSION

In this study, we explored how individuals respond to offers from others with different intentions and evaluated the time courses of social decision-making in a UG. The behavioral results showed that unfair offers with altruistic intention had the greatest ARs, fairness ratings, and pleasantness ratings. Unfair offers with ambiguous intention had moderate acceptance rates, whereas those with selfish intention had the lowest. However, the RT was the longest for ambiguous intention. The RSA of the ERP data showed that the ERP activity at ~430–590 ms after the offer presentation was significantly correlated with the participants' RT.

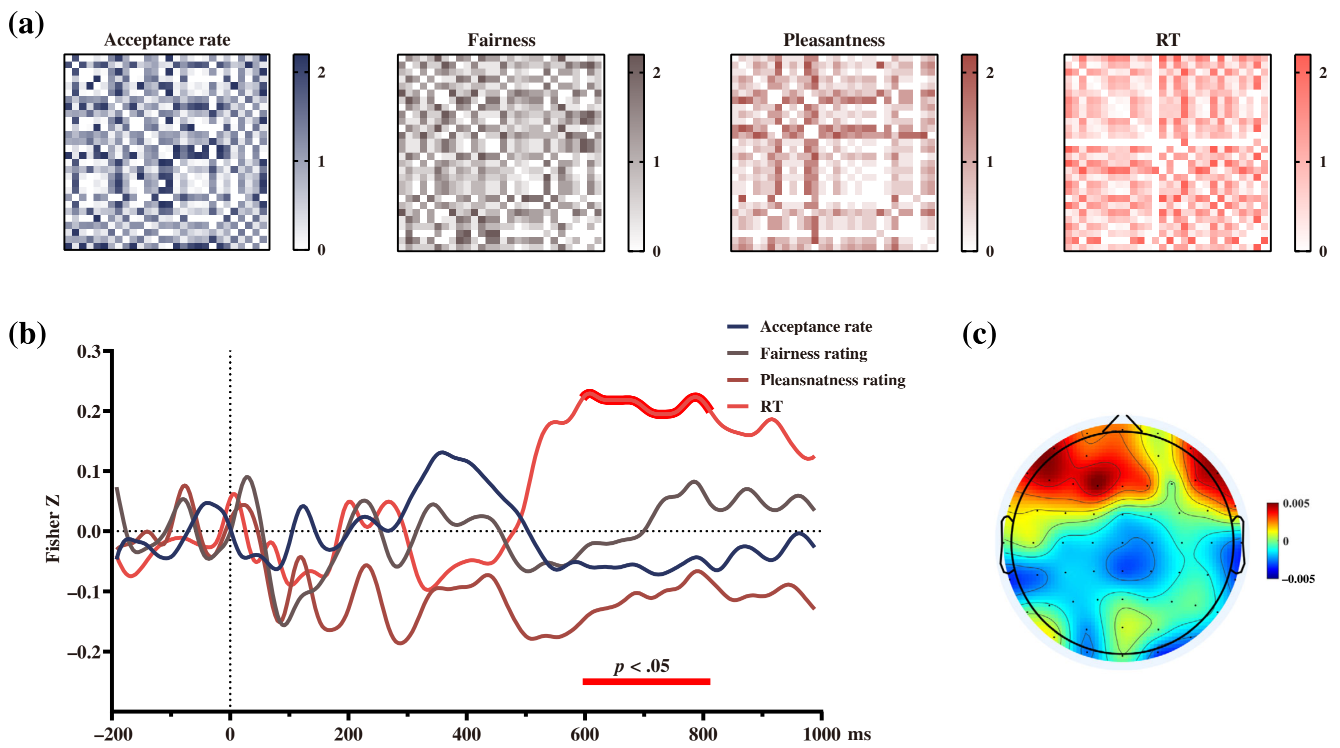

FIGURE 6 Results of the inter-subject representational similarity analysis. (a) The subject dissimilarity matrices of acceptance, fairness, pleasantness, and RT for ambiguous intention; (b) The similarity correlation between the four types of behavioral data and the ERP data; (c) The topographies of the similarity correlation for the RT_neural correlation in the 596–812 ms time windows. Thick lines represent significant differences between two representational dissimilarity matrices (RDMs) and thin lines represent nonsignificant differences. RT = response time. * = p < .05.

Our behavioral analysis revealed that indirect reciprocity-based intention had an impact on participants' unfairness aversion. Altruistic intentions not only increased the acceptance rate of unfair offers but also changed the individual's perception of fairness and their emotional experience of unfair offers. Previous studies have indicated that people are more prone to accepting unfair offers because of direct reciprocity to good intentions than to bad intentions (Falk et al., 2008; Ma et al., 2015). Our results showed that when the responders observed the proposer's selfish action (to keep the money for themselves), they were more likely to reject the unfair scheme. However, when they observed the proposer's altruistic action (donating one's own allotted money to charity to help others), they were more likely to accept the same unfair offer, even if they only received a small amount of money. This suggests that good intentions can reduce an individual's resistance to unfair treatment. This finding is consistent with the forecast of the intention-based reciprocity model. Altruistic behavior can send signals of friendliness and generosity (Cronk, 2005; Smith & Bliege Bird, 2005), and people reward intentions of altruism even when they do not benefit them. These findings validate the profound influence of intention on fairness and decision-making. They also resonate with the intention-based reciprocity model of fair processing, which suggests that people reward friendly behaviors and punish unfriendly behaviors (Falk & Fischbacher, 2006; Guala, 2012).

The behavioral results also showed that for unfair offers, the acceptance rates, fairness ratings, and pleasantness ratings of ambiguous intention were intermediate between those of selfish and altruistic intentions. In addition, the RT of ambiguous intentions was longer than those of the other two intentions. Our findings do not seem to support the notion of an outcome bias in intention inference, as acceptance rates were significantly higher for unfair offers with ambiguous intentions than for those with selfish intentions. This is in line with the results of Friedrichsen et al. (2022), who also found no evidence of an outcome bias in a reciprocating decision. We further calculated each participant's intention propensity to unfair offers based on the acceptance rates. The results showed that some participants tended to respond to ambiguous intentions as they would to altruistic ones, whereas others tended to respond to ambiguous intentions as they would to selfish ones (outcome bias). A majority of participants showed an obvious tendency, indicating that there are considerable individual differences in decision-making under the conditions of ambiguous information. Previous studies have found that several factors can influence decision-making with ambiguous information. These factors include individual abilities such as knowledge, experience, and skill components (Kahneman & Tversky, 1996), control (Goodie, 2003), as well as their anxiety level (Gu et al., 2010). This may indicate that the effect of outcome bias in the process of intentional reasoning is not pronounced and is more related to personal traits. In addition, the RTs were longer for ambiguous intention than for the other two intentions. Previous research has revealed that participants take longer to respond when accumulating evidence than when they are provided with explicit information (Maksimenko et al., 2021). Ambiguous intention provides insufficient information, which makes the decision-making process more complicated and time-consuming.

The RSA results revealed a significant connection between the ERP data and behavioral data. In the UG, individuals make value trade-offs to maximize their self-interest while also controlling the negative feelings induced by unequal reward distribution (Li et al., 2020). This study also takes the proposers' intentions into account. Behavioral data may point to various facets of an individual's neural processing. For example, individuals' emotional responses to unfair offers may be reflected in their ratings of fairness and pleasantness. Researchers have previously been unable to fully identify and characterize the mental processes revealed by specific ERP components. We discovered that only RT data were strongly linked with EEG data. RT is a conventional behavioral metric, and our long-term objective is to connect it to the decision-making process itself. RT is often associated with task difficulty and confidence ratings (Baranski & Petrusic, 1998; Bell et al., 1982). More recently, it has also been used to describe the process of evidence accumulation during decision-making (Milosavljevic et al., 2010; Ratcliff & Smith, 2004; Usher & McClelland, 2001). This is the first study to report evidence showing that the evidence accumulation process for intention and fairness takes place during 432–592 ms post-stimulus onset in the parietal region. As predicted, this time window is close to that of the P3 component (350–550 ms). This is consistent with previous research on this subject. Philiastides et al. (2006) found that the evidence accumulation process started after about 300 ms following the visual perception process, and the processing time relied on the intensity of the evidence. Mostert et al. (2015) established that people's decision-making sensory processing occurs in the 130–320 ms window after stimulus onset, but evidence accumulation for decision-making takes longer. Some studies have also indicated that ERP activity in a conventional P3 time window is responsive to information accumulation (Darriba & Waszak, 2018; Kelly & O'Connell, 2013; O'Connell et al., 2012). These findings indicate that participants may have combined multiple pieces of information (such as fairness, intention, and amount of money, among others) to make a decision and that the ERP activity reflects this evidence accumulation process before making a final decision in a specific period (~432–592 ms). This suggests that the differences in ERP activity in this time window induced by different offers may reflect differences in evidence accumulation processes rather than other psychological processes (e.g., emotional arousal).

The spatial analysis indicated that the process of evidence accumulation occurs mostly in the posterior parietal and prefrontal regions. Several studies have demonstrated that the activities in these regions are closely related to the evidence accumulation for perceptual decision-making (FitzGerald et al., 2015; Liu & Pleskac, 2011). Pereira et al. (2021) recorded single-neuron activity in the posterior parietal cortex of participants and proposed that these neurons are in charge of evidence accumulation. Moreover, Lin et al. (2020) suggested that evidence accumulation for value computation over time was represented in the DLPFC. Our results indicate that the posterior parietal and prefrontal regions are also related to the accumulation of neurological evidence during social decision-making. However, because of the diffuse and macroscopic character of scalp potentials, inferences about the origin of the activity should be considered speculative.

Interestingly, the IS-RSA results also revealed that greater intersubject dissimilarity in ERP activity within 596–812 ms elicited by intention information was associated with a larger dissimilarity in decision times. Accurate comprehension of the intentions of others is essential for efficient social interaction. Previous studies have shown that ERP activity in frontal and parietal scalp regions during a later time window (from 500–1000 ms) after experimental stimulus presentation reflects specific cognitive processes for judging others' intentions or beliefs (Geangu et al., 2013; Liu et al., 2009; McCleery et al., 2011). McCleery et al. (2011) proposed that the difference in ERP activity between 600 and 800 ms may reflect the differential recruitment of executive processes for inhibitory control of intention. Our results provide new evidence that brain activity between 596 and 812 ms in the prefrontal region plays an important role in processing intention, and that this phase of ERP activity could predict subsequent decision-making time. Previous studies have found that frontal cortical regions integrate input information over a longer time window than parietal regions and that the inactivation of the prefrontal region allows rats to make decisions based only on the last few hundred milliseconds of accumulated evidence (Erlich et al., 2015). In our experiments, participants perceived the proposer's intention first and made the final decision in combination with the fairness of the subsequent allocation scheme. Hence, the late ERP activity in frontal areas may reflect the accumulation and storage of information about intentions which can be later used to make decisions.

This study reports the first attempt at exploring the process of fairness-related decision-making using the ERP-based RSA. Our study provides several contributions. First, our data support the intention-based reciprocity model rather than the predictions of the inequality model. In previous intention-based decision-making studies, researchers have tended to focus on the individual's behavioral response to being treated with hostility or benevolence. Our study extends the intention-based reciprocity model to a third-party perspective by finding that as people observe someone showing good intentions towards a third-party individual, they will also show more kindness towards that friendly person. This also reflects the reputation effect of altruistic behavior, in which altruistic intentions allow individuals to gain a good reputation within a group, leading to immediate or potential future rewards. Second, we used the spatiotemporal RSA and IS-RSA methods to associate brain activity after intention presentation and offer presentation with response time, which demonstrates the time course of evidence accumulation in social decision makings. Our findings can inspire researchers to conduct an extensive RSA on EEG data to achieve a deeper comprehension of the brain processes underlying human social behavior.

5 | CONCLUSIONS

This study analyzed the impact of intention on unfairness aversion and demonstrated that people's behaviors and emotions are altered by intention. Compared to selfish and ambiguous intentions, altruistic intentions reduced the participants' unfairness aversion, resulting in higher acceptance rates and a heightened sense of fairness and pleasantness even for unfair allocation proposals. Some individual differences were also apparent in the responses to ambiguous intentions. Importantly, the RSA revealed that the ERP activity at 596–812 ms elicited by intention information and at 430–590 ms following offer presentation represented response time differences, which perhaps reflect the evidence accumulation process before social decision-making.

AUTHOR CONTRIBUTIONS

Qiang Xu: Conceptualization; software; investigation; formal analysis; visualization; writing - original draft. **Jiali Hu**: Investigation; visualization; writing - original draft. **Yi Qin**: Software; investigation; writing - original draft. **Guojie Li**: Formal analysis; visualization; writing - review and editing. **Xukai Zhang**: Formal analysis; writing - review and editing. **Peng Li**: Conceptualization; formal analysis; supervision; funding acquisition; writing - review and editing.

ACKNOWLEDGEMENTS

We thank the editors and three anonymous reviewers for their helpful comments.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data in this study are available upon request by contact with the corresponding author, in consideration of data protection, a formal data sharing agreement is needed when the data are requested.

ORCID

Qiang Xu https://orcid.org/0000-0001-8778-3957
Peng Li https://orcid.org/0000-0002-9330-5713

REFERENCES

Aoki, R., Yomogida, Y., & Matsumoto, K. (2015). The neural bases for valuing social equality. *Neuroscience Research*, *90*, 33–40.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 929–945.

Bell, P. A., Loomis, R. J., & Cervone, J. C. (1982). Effects of heat, social facilitation, sex differences, and task difficulty on reaction time. *Human Factors*, *24*(1), 19–24.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, *90*(1), 166–193.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, *60*(5), 930–940.

Cecere, R., Gross, J., Willis, A., & Thut, G. (2017). Being first matters: Topographical representational similarity analysis of ERP signals reveals separate networks for audiovisual temporal binding depending on the leading sense. *Journal of Neuroscience*, *37*(21), 5274–5287.

Charness, G., & Levine, D. I. (2007). Intention and stochastic outcomes: An experimental study. *The Economic Journal*, *117*(522), 1051–1072.

Cronk, L. (2005). The application of animal signaling theory to human phenomena: Some thoughts and clarifications. *Social Science Information*, *44*(4), 603–620.

Darriba, Á., & Waszak, F. (2018). Predictions through evidence accumulation over time. *Scientific Reports*, *8*(1), 1–15.

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. *Event-Related Brain Potentials in Man*, *349*, 411.

Erlich, J. C., Brunton, B. W., Duan, C. A., Hanks, T. D., & Brody, C. D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on an accumulation of evidence task in the rat. *eLife*, *4*, e05457.

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry*, *41*(1), 20–26.

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior*, *62*(1), 287–303.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

FitzGerald, T. H., Moran, R. J., Friston, K. J., & Dolan, R. J. (2015). Precision and neuronal dynamics in the human posterior parietal cortex during evidence accumulation. *NeuroImage*, *107*, 219–228.

Friedrichsen, J., Momsen, K., & Piasenti, S. (2022). Ignorance, intention and stochastic outcomes☆. *Journal of Behavioral and Experimental Economics*, *100*, 101913.

Frith, C. D., & Frith, U. (1999). Interacting minds--a biological basis. *Science*, *286*(5445), 1692–1695.

Gabay, A. S., Radua, J., Kempton, M. J., & Mehta, M. A. (2014). The ultimatum game and the brain: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *47*, 549–558.

Geangu, E., Gibson, A., Kaduk, K., & Reid, V. M. (2013). The neural correlates of passively viewed sequences of true and false beliefs. *Social Cognitive and Affective Neuroscience*, *8*(4), 432–437.

Goodie, A. S. (2003). The effects of control on betting: Paradoxical betting on items of high confidence with low value. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 598–610.

Gu, R., Ge, Y., Jiang, Y., & Luo, Y. J. (2010). Anxiety and outcome evaluation: The good, the bad and the ambiguous. *Biological Psychology*, *85*(2), 200–206.

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, *35*(1), 1–15.

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, *173*, 434–447.

Güroğlu, B., van den Bos, W., Rombouts, S. A., & Crone, E. A. (2010). Unfair? It depends: Neural correlates of fairness in social context. *Social Cognitive and Affective Neuroscience*, *5*(4), 414–423.

Güth, W. (1995). On ultimatum bargaining experiments—A personal review. *Journal of Economic Behavior & Organization*, *27*(3), 329–344.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, *303*(5664), 1634–1640.

He, T., Boudewyn, M. A., Kiat, J. E., Sagae, K., & Luck, S. J. (2022). Neural correlates of word representation vectors in natural language processing models: Evidence from representational similarity analysis of event-related brain potentials. *Psychophysiology*, *59*(3), e13976.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–1770.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709.

Hu, X., Wang, F., & Zhang, D. (2022). Similar brains blend emotion in similar ways: Neural representations of individual difference in emotion profiles. *NeuroImage*, *247*, 118819.

Jung, T. P., Makeig, S., McKeown, M. J., Bell, A. J., Lee, T. W., & Sejnowski, T. J. (2001). Imaging brain dynamics using independent component analysis. *Proceedings of the IEEE*, *89*(7), 1107–1122.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, *76*, 728–741.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*(3), 582–596.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *Journal of Neuroscience*, *33*(50), 19434–19441.

Kiat, J. E., Hayes, T. R., Henderson, J. M., & Luck, S. J. (2022). Rapid extraction of the spatial distribution of physical saliency and semantic informativeness from natural scenes in the human brain. *Journal of Neuroscience*, *42*(1), 97–108.

Kim, K. I., Jung, W. H., Woo, C. W., & Kim, H. (2022). Neural signatures of individual variability in context-dependent perception of ambiguous facial expression. *NeuroImage*, *258*, 119355.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829–832.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Li, X., Xiong, G., Dong, Z., Cai, S., Zhao, J., She, Z., & Guo, Y. (2020). Causal role of the right dorsolateral prefrontal cortex in organizational fairness perception: Evidence from a transcranial direct current stimulation study. *Frontiers in Behavioral Neuroscience*, *14*, 134.

Lin, Z., Nie, C., Zhang, Y., Chen, Y., & Yang, T. (2020). Evidence accumulation for value computation in the prefrontal cortex during decision making. *Proceedings of the National Academy of Sciences*, *117*(48), 30728–30737.

Liu, D., Meltzoff, A. N., & Wellman, H. M. (2009). Neural correlates of belief and desire reasoning. *Child Development*, *80*(4), 1163–1171.

Liu, T., & Pleskac, T. J. (2011). Neural correlates of evidence accumulation in a perceptual decision task. *Journal of Neurophysiology*, *106*(5), 2383–2398.

Lu, Y., Wang, C., Chen, C., & Xue, G. (2015). Spatiotemporal neural pattern similarity supports episodic memory. *Current Biology*, *25*(6), 780–785.

Luo, Y., Wu, T., Broster, L. S., Feng, C., Zhang, D., Gu, R., & Luo, Y. J. (2014). The temporal course of the influence of anxiety on fairness considerations. *Psychophysiology*, *51*(9), 834–842.

Ma, Q., Hu, Y., Jiang, S., & Meng, L. (2015). The undermining effect of facial attractiveness on brain responses to fairness in the ultimatum game: An ERP study. *Frontiers in Neuroscience*, *9*, 77.

Maksimenko, V., Kuc, A., Frolov, N., Kurkin, S., & Hramov, A. (2021). Effect of repetition on the behavioral and neuronal responses to ambiguous Necker cube images. *Scientific Reports*, *11*(1), 1–13.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220.

Mazzocco, P. J., Alicke, M. D., & Davis, T. L. (2004). On the robustness of outcome bias: No constraint by prior culpability. *Basic and Applied Social Psychology*, *26*(2–3), 131–146.

McCleery, J. P., Surtees, A. D., Graham, K. A., Richards, J. E., & Apperly, I. A. (2011). The neural and cognitive time course of theory of mind. *Journal of Neuroscience*, *31*(36), 12849–12854.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for value-based choice response times under high and low time pressure. *Judgment and Decision making*, *5*(6), 437–449.

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798.

Mostert, P., Kok, P., & de Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, *5*(1), 1–13.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology*, *10*(4), e1003553.

Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., & Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*, *109*(24), 9599–9604.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*(12), 1729–1735.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869.

Pereira, M., Megevand, P., Tan, M. X., Chang, W., Wang, S., Rezai, A., Seeck, M., Corniola, M., Momjian, S., Bernasconi, F., Blanke, O., & Faivre, N. (2021). Evidence accumulation relates to perceptual consciousness and monitoring. *Nature Communications*, *12*(1), 1–11.

Peterburs, J., Voegler, R., Liepelt, R., Schulze, A., Wilhelm, S., Ocklenburg, S., & Straube, T. (2017). Processing of fair and unfair offers in the ultimatum game under social observation. *Scientific Reports*, *7*(1), 1–12.

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience*, *26*(35), 8965–8975.

Pinto, L., Tank, D. W., & Brody, C. D. (2022). Multiple timescales of sensory-evidence accumulation across the dorsal cortex. *eLife*, *11*, e70263.

Polezzi, D., Daum, I., Rubaltelli, E., Lotto, L., Civai, C., Sartori, G., & Rumiati, R. (2008). Mentalizing in economic decision-making. *Behavioural Brain Research*, *190*(2), 218–223.

Popal, H., Wang, Y., & Olson, I. R. (2019). A guide to representational similarity analysis for social neuroscience. *Social Cognitive and Affective Neuroscience*, *14*(11), 1243–1253.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, *83*, 1281–1302.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367.

Ritchie, J. B., Bracci, S., & de Beeck, H. O. (2017). Avoiding illusory effects in representational similarity analysis: What (not) to do with the diagonal. *NeuroImage*, *148*, 197–200.

Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times. *The Economic Journal*, *117*(523), 1243–1259.

Sambrook, T. D., & Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, *141*(1), 213–235.

Smith, E. A., & Bliege Bird, R. (2005). Costly signaling and cooperative behavior. *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, *6*, 115.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Wang, L., Wlotko, E., Alexander, E., Schoot, L., Kim, M., Warnke, L., & Kuperberg, G. R. (2020). Neural evidence for the prediction of animacy features during language comprehension: Evidence from MEG and EEG representational similarity analysis. *Journal of Neuroscience*, *40*(16), 3278–3291.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235–8240.

Zheng, Y., Yang, Z., Jin, C., Qi, Y., & Liu, X. (2017). The influence of emotion on fairness-related decision making: A critical review of theories and evidence. *Frontiers in Psychology*, *8*, 1592.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Xu, Q., Hu, J., Qin, Y., Li, G., Zhang, X., & Li, P. (2023). Intention affects fairness processing: Evidence from behavior and representational similarity analysis of event-related potential signals. *Human Brain Mapping*, 1–14. https://doi.org/10.1002/hbm.26223