Neurocomputational Mechanism of Reciprocity: The Core and Periphery

Inauguraldissertation

zur Erlangung des akademischen Grades eines Doktors / einer Doktorin der Sozialwissenschaften der Universität Mannheim

submitted by

Huihua Fang

Dean:

Prof. Dr. Michael Diehl

Primary Supervisor:

Prof. Dr. Frank Krüger, University of Mannheim & George Mason University

Secondary Supervisor:

Prof. Dr. Pengfei Xu, Beijing Normal University

Evaluators:

Prof. Dr. Georg W. Alpers, University of Mannheim

Prof. Dr. Gabriele Bellucci, Royal Holloway University of London

Date of Defense:

11th October, 2024

Table of Contents

1. Abstract

Reciprocity is crucial in maintaining cooperative relationships and societal stability. Social representation theory posits that reciprocity decisions comprise a stable core (reciprocity propensity) and a flexible periphery (effect of decision context). Framing techniques allow researchers to differentiate and investigate these components by manipulating the periphery while keeping the core unchanged. In today's increasingly uncertain world, anxiety has become a prevalent concern, potentially impacting social behaviors like reciprocity.

While previous research has identified brain networks that predict reciprocity propensity, these studies were conducted in specific contexts, overlooking the distinction between the core and periphery of reciprocity. Although studies have demonstrated peripheral effects on reciprocity, the underlying neurocomputational mechanisms are not well understood. In addition, previous research has shown that anxiety reduces reciprocity, but the neurocomputational mechanisms by which anxiety modulates the core and periphery of reciprocity remain unexplored. To address these gaps, this dissertation aimed to develop a comprehensive understanding of the core and periphery of reciprocity and to examine how anxiety affects both aspects of reciprocity. Specifically, three studies utilizing multimodal approaches, including task-free (resting-state) and task-based functional magnetic resonance imaging (fMRI), event-related potentials (ERPs), and eye-tracking, were conducted to reveal the neural (both network- and region-level localization and temporal dynamics) and computational mechanisms underlying reciprocity decisions.

Study 1 aimed to identify the neural underpinning that explain the core and periphery of reciprocity. Participants first underwent resting-state fMRI and then completed one-shot Trust Game (TG; give framed context) and Distrust Game (DTG; take framed context) as trustees. Connectome-based predictive modeling (CPM) on resting-state functional connectivity (RSFC) was used to investigate how RSFC predicts the core and periphery of reciprocity. Results showed that inter-network RSFC between the default mode network (DMN) and cingulo-opercular network (CON) contributed to predicting the core of reciprocity under both give and take framed contexts, while DMN intra-network RSFC primarily contributed to predicting the periphery of reciprocity.

Study 2 aimed to delve deeper into the periphery of reciprocity by investigating its neurocomputational mechanisms. Participants engaged in a two-stage binary TG, framed in either gain or loss contexts while undergoing fMRI. Computational modeling revealed that advantageous inequity aversion mediates the contextual effect on reciprocity. Neuroimaging results showed that right amygdala activity negatively correlated with advantageous inequity aversion only in the gain-framed context during overall reciprocity decision-making. During other-oriented inference processes, no significant differences in the neural correlates of advantageous inequity aversion between the two contexts were observed. However, left anterior insula activity modulated this contextual effect, specifically in self-oriented evaluation processes, showing a positive association with advantageous inequity aversion exclusively in the gain frame and reduced activity in the loss-framed context.

Study 3 aimed to understand the neurocomputational mechanism underlying anxiety modulates both the core and periphery aspects of reciprocity. Participants with low and high trait anxiety completed a binary TG framed as gain and loss context while recording eye movement and electroencephalography. Computational modeling, validated by eye-tracking data, revealed that trait anxiety affects both the core and periphery of reciprocity. Regarding the core, trait anxiety reduced both overall reciprocity and specific psychological components of guilt aversion and advantageous inequity liking, regardless of context. ERP findings supported this, showing decreased P2 (selective attention) and increased LPP (cognitive regulation over emotion) amplitudes in anxious individuals. Regarding the periphery, trait anxiety altered contextual perceptions of advantageous inequity aversion and reward. Specifically, trait anxiety reversed the effect of context on advantageous inequity aversion, a pattern reflected in N2 amplitudes (cognitive control).

In summary, the three studies in this dissertation uncover distinct neurocomputational mechanisms underlying the core and periphery of reciprocity and elucidate how reciprocity is affected by anxiety through these mechanisms. These findings provide a new perspective on reciprocity and identify potential psychological and neural targets for interventions aimed at promoting cooperative behavior in individuals with anxiety disorders.

2. General Introduction

2.1. The Dynamics of Reciprocity: Core and Periphery

"There is no duty more indispensable than that of returning a kindness," observed the ancient Roman statesman Cicero, manifesting the vital role of reciprocity in human interactions. Serving as a universal moral norm, reciprocity refers to the mutual exchange of kindness or help among individuals in a society (Gouldner, 1960). On the social level, reciprocity plays a crucial role in building interpersonal trust, promoting fairness, and fostering cooperation (Krueger, 2021). It enhances social harmony and has been pivotal in the evolution of human social systems, influencing daily interactions from personal relationships to broader societal structures (Falk & Fischbacher, 2006; Fehr et al., 2002). On an individual level, reciprocity offers benefits beyond its social impact. It cultivates gratitude and positive emotions, boosting well-being and life satisfaction (Algoe et al., 2008). Additionally, it enhances one's reputation and social standing (Xia et al., 2023), playing a crucial role in both personal friendships and professional collaborations (Lang & Fingerman, 2003).

While reciprocity is a fundamental part of human social behavior, it is not uniformly expressed across individuals or situations, reflecting both dispositional and contextual influences (X. Li et al., 2017; Mahmoodi et al., 2018). Some individuals have a strong inclination toward reciprocity, consistently returning favors regardless of the context under which the decision is made, while others might be more affected by the context where the reciprocity decisions are made (Bowles & Gintis, 2004; Fehr et al., 2002; Tucker & Ferson, 2008). Similarly, contexts that can shape reciprocity behavior often operate without being recognized in daily life. For instance, individuals tend to exhibit higher reciprocity in face-to-face interactions compared to interactions lacking direct personal contact (Behrens & Kret, 2019). People also show higher reciprocity in the context of "give" compared to the context of "take"—although the payoff structure is equivalent in both contexts (Keysar et al., 2008). Reciprocity decision-making is determined by both the stable personal characteristics and the fluctuation of the contexts. Therefore, to get deeper insight into reciprocity decision-making, it is vital to clarify the distinct influences of dispositional propensity and context.

Social Representation Theory

Social representation theory offers a framework for understanding such distinction in reciprocity decision-making (Abric, 1993; Hagen & Hammerstein, 2006; Wagenaar et al., 1988). According to social representation theory, reciprocity decision is composed of both a core and periphery (Abric, 1993). While the core represents the stable aspects of reciprocity which are resistant to contextual changes, the periphery is more flexible and susceptible to immediate contextual modification (Abric, 1993; Hagen & Hammerstein, 2006; Wagenaar et al., 1988). The core is determined by an individual's inherent reciprocity propensity, which is strongly related to their personality and long-standing beliefs. In contrast, the periphery is determined by the context, related to how context is perceived. The same individual might reciprocate more in a social exchange than a business transaction, adapting their behavior to suit the context while their propensity for reciprocity remains constant (Batson & Moran, 1999).

The distinction between the core and periphery structure has often been overlooked in previous studies on reciprocity, potentially limiting our understanding of the sources driving reciprocity behavior. Studies examining the neural mechanisms of reciprocity have typically used one specific context (Baumgartner et al., 2009; Bellucci et al., 2019; Bereczkei et al., 2015; Cáceda et al., 2015; Krueger et al., 2008; J. Li et al., 2009; van den Bos et al., 2009, 2011). Furthermore, without considering the effect brought by contexts, research has investigated how individual characteristics (Dohmen et al., 2008; Gunnthorsdottir et al., 2002; Pelligra, 2011; Perugini et al., 2003) or clinical symptoms (Anderl et al., 2018; Rodebaugh et al., 2011, 2013; Xiao et al., 2022) influence reciprocity. Since decisions always occur within specific contexts, drawing conclusions based on experiments conducted under certain contexts may blur the distinction between effects originating from the core and the periphery of reciprocity decision-making. Therefore, developing a strategy to disentangle this potential confounds is crucial, allowing us to track whether the observed effects are attributable to the core or the periphery of reciprocity.

Framing Reciprocity Decisions

The framing technique appears to be a promising strategy to resolve this challenge. The framing effect is a cognitive bias that impacts how people make decisions based on how information is framed (Tversky & Kahneman, 1981). Framing involves presenting a decision differently while keeping the objective facts the same (De Martino et al., 2006; Tversky & Kahneman, 1981). A well-known example of the framing effect is the phenomenon of loss aversion, as proposed by prospect theory (Kahneman and Tversky, 1979), where individuals tend to avoid losses over obtaining equivalent gains. By manipulating the context in the same decision problem, the choice of the same individual is shifted. The framing technique, well-aligned with the social presentation theory, allows us to manipulate the peripheral context in reciprocity decision-making while leaving the core unchanged.

Research on reciprocity has consistently demonstrated framing effects across various experimental paradigms. In dictator games, where participants alternate between being passive receivers and active dictators, participants tend to reciprocate more generously in the 'give' frames compared to 'take' frames (Keysar et al., 2008). Similarly, another study that systematically compared the Trust Game (TG; 'give' frame) with the Distrust Game (DTG; 'take' frame) found that people reciprocate more in the 'give' frame than in the 'take' frame (Bohnet & Meier, 2005). While individual tendencies are an apparent deciding factor in decision-making, these studies emphasize the critical role of the psychological perceptions of the context in reciprocity decisions. Framing reciprocity decisions in different contexts can further help us to disentangle the context-dependent modifications (periphery) from stable individual tendencies (core) in reciprocity decisions.

2.2. Anxiety's Influence on Reciprocity

While reciprocity internally consists of a stable core and a contextual periphery, it can be affected by various factors. As the global economy slows down (Jung, 2024), geopolitical conflicts escalate (Eberle & Daniel, 2022), public health crises intensify (Collier Villaume et al., 2023), anxiety levels rise, posing a significant threat to social reciprocity and harmony.

Research has shown that individuals with anxiety disorders demonstrate reduced reciprocity in economic exchange tasks (Anderl et al., 2018; Rodebaugh et al., 2011, 2013). However, it remains unclear how anxiety affects reciprocity decisions in terms of its core and peripheral mechanisms.

Anxiety might alter an individual's general tendency to reciprocate (core), perception of different social contexts when reciprocating (periphery), or a combination of both. The observation that anxiety reduces reciprocity across various tasks (Anderl et al., 2018; Rodebaugh et al., 2011, 2013), suggests an attenuation of the core, as this effect persists under different contexts. Conversely, anxiety has also been associated with increased susceptibility to framing effects (Gu et al., 2017; Xu et al., 2013). Anxiety is positively associated with activation of amygdala based “emotional” system when decisions aligned with the framing effect, but negatively associated with activation in the dorsal anterior cingulate cortex based “analytic” system when decisions contradicted this effect (Xu et al., 2013). Heuristic processing seems to dominate individual with high anxiety in decision-making, which leads to their higher sensitivity to the framing effect (Jepma & López-Solà, 2014). These previous studies suggest anxiety modulates not only the core but also the periphery of reciprocity, highlighting the critical need for an empirical study to examine these effects.

2.3. Experimental Paradigms of Reciprocity

To understand the complexity of reciprocity and how anxiety affects it, a suitable experimental paradigm is crucial. Game theory, a mathematical framework for analyzing decision-maker strategies in interactive economic games, provides a theoretical foundation for studying decision-making during social interaction (Morris, 2012; Sanfey, 2007). Interactive economic games based on game theory offer well-controlled settings to study social behavior. Several classical games have been widely used in the investigation of social-economic interaction (van Dijk & De Dreu, 2021), including the Dictator Game (measuring altruism and fairness) (Harsanyi, 1961); the Prisoner's Dilemma (exploring cooperation and mutual benefit versus self-interest) (Rapoport & Chammah, 1965); and the Ultimatum Game (examining fairness) (Güth et al., 1982).

The TG is particularly well-suited for studying reciprocity behavior because it allows researchers to measure both trust and reciprocity, mimicking real cooperative social interactions (Berg et al., 1995). The game involves two players: the trustor was typically endowed with some amount initially while the trustee was not. In TG (continuous version; **Fig. 1 A**), trustors are endowed with some amount initially and decide how much of their endowment to transfer to the trustees (who were not endowed with any amount). The transferred amount is typically tripled by the experimenter, and the trustee then decides how much to return to the trustor. Trust is measured as the ratio between the transferred amount and the initial endowment for the trustor, while reciprocity as the ratio between the returned and received amount. A simpler version is the Binary TG (**Fig. 1B**), in which the trustor decides whether to trust or not (status quo) based on a given payoff structure. If the trustor chooses the status quo, both the trustor and trustee receive an amount directly. However, if the trustor decides to trust, the decision-making shifts to the trustee, who must then choose to either reciprocate the trust or betray it (Kreps & others, 1990). If the trustee chooses to reciprocate, both the trustor and the trustee receive a higher amount than they would under the status quo. Conversely, if the trustee chooses to betray, the trustor receives the lowest possible amount, while the trustee receives the highest amount among all their options. In this game, the levels of trust and reciprocity are typically measured by the frequency with which trustors and trustees choose their respective options.

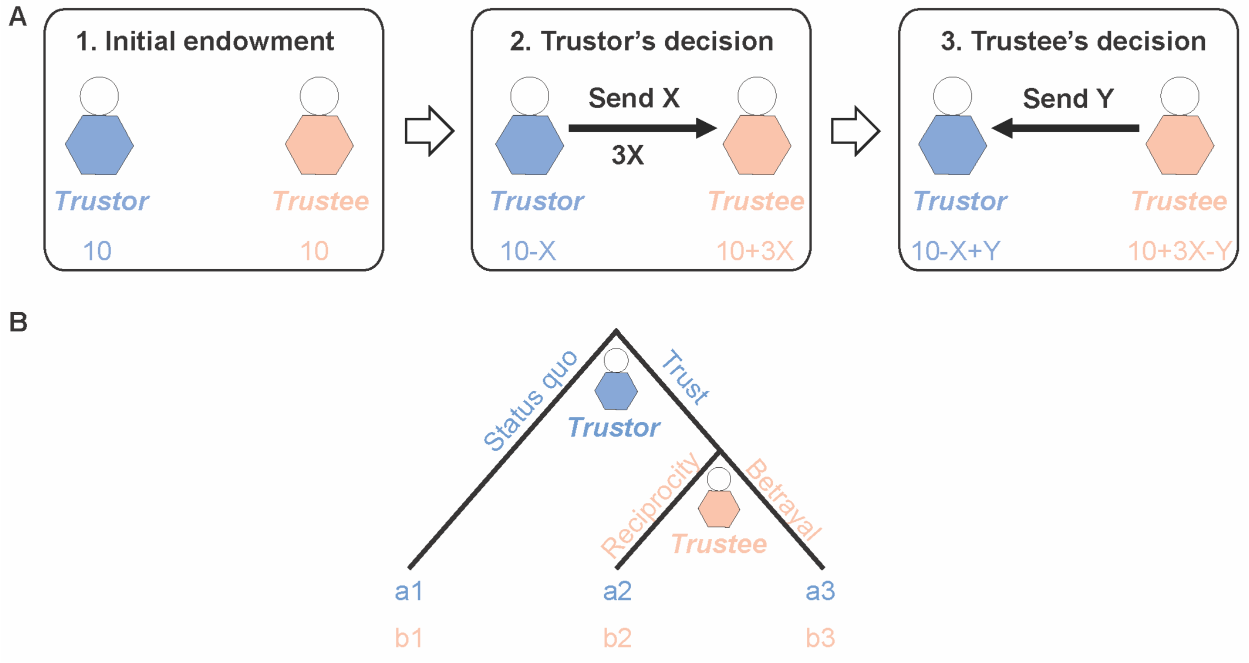


Figure 1. **(A) Trust game with continuous version**. Both the trustor and the trustee are endowed with 10 units. The trustor first decides to send an amount (X) ranging from 0 to 10 to the trustee. Any amount sent is tripled by the experimenter. The trustee then decides to send back an amount (Y) ranging from 10 to 10 + 3X to the trustor. **(B) Trust game with binary version**. Based on the payoff structure presented, the trustor first decides between maintaining the status quo or placing trust. If the status quo is chosen, the trustor receives a1 and the trustee receives b1. If trust is chosen, the decision is delegated to the trustee. If the trustee chooses reciprocity, the trustor receives a2 and the trustee receives b2. If betrayal is chosen, the trustor receives a3 and the trustee receives b3 . Note that a2 > a1 > a3 and b3 > b2 > b1.

Besides different versions, there are two main variants of the TG based on the frequency of interaction with a partner: the one-shot TG and the repeated TG (Kanagaretnam et al., 2010). In the one-shot TG, participants play the game only once with a given partner, although they may engage in multiple rounds of one-shot games with different partners (i.e., multiple one-shots). In contrast, the repeated TG involves participants playing multiple rounds with the same partner(s). Compared to the repeated TG, the one-shot TG minimizes confounding factors such as reputation building, learning, and strategic behavior, which is more suitable for investigating the propensity for reciprocity (Alós-ferrer & Farolfi, 2019, p. 1995; Berg et al., 1995; Xia et al., 2023). Building upon the well-controlled and quantitative economic games in game theory, advanced modeling and neuroimaging techniques can be utilized to uncover the psychological components driving the behavior and the underlying neural mechanisms.

2.4. Psychological Mechanisms of Reciprocity

Computational Modeling

To gain insight into the complex psychological components of decision-making, computational modeling has emerged as a powerful technique and has gained popularity in recent years (Calder et al., 2018; Wilson & Collins, 2019). Traditional behavioral analyses, such as group or condition comparisons, correlation, and regression analyses, focus on inferring relationships between manipulated variables and direct observations, such as responses and reaction times. While these methods provide simple and straightforward results, they often struggle to uncover the underlying psychological processes that drive decisions, especially when these processes are difficult to manipulate or observe directly (Calder et al., 2018). Constructing computational models with hypothesized components to simulate decision-making processes can help map unobservable latent psychological processes onto observable behavior, providing deeper insights into the underlying mechanisms (Wilson & Collins, 2019).

By fitting a computational model that includes these hypothesized components to behavioral data, the relative contributions of different components to decision-making can be estimated (Wilson & Collins, 2019). Generally, several plausible models, each with different hypothesized components, are constructed separately. A model comparison procedure is then employed to select the best-fitting model based on criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). The best-fitting model identified the most effective components (parameters) driving the behavior. To ensure the robustness of the model, a simulation process is conducted using the estimated parameter values to generate simulated data. A parameter recovery procedure is then performed by refitting the model to this simulated data, allowing for the recovery of parameter values. Analyzing the correlation between the estimated and recovered parameters helps verify the model's robustness. The estimated parameter values from the best-fitting model can be extracted and subjected to further statistical analyses. These parameter values can also be further used as regressors in neuroimaging data to explore the neural mechanisms underlying the psychological components.

Computational modeling has been applied to the study of reciprocity in decision-making, identifying psychological components driving reciprocity behavior (Chang et al., 2011; Nihonsugi et al., 2015, 2021a; van Baar et al., 2019). Using a hidden multiplier TG to compare different moral strategies in reciprocity, a computational modeling study has identified components behind reciprocity, including guilt aversion, inequity aversion, and moral opportunism (i.e., adaptive switching between guilt and inequity aversion) (van Baar et al., 2019). Employing the binary TG, guilty aversion and inequity aversion have been identified (Nihonsugi et al., 2015) and guilt aversion has been found to correlate with agreeableness (Nihonsugi et al., 2021). These prior studies emphasize the role of guilt aversion and inequity aversion in reciprocity.

Although these studies did not consider the influence of different context which lead to the ambiguity of core and periphery of reciprocity, they provide valuable insight into the driving force of reciprocity. Guilt aversion is characterized by the discomfort of failing to meet others' expectations, while inequity aversion refers to the discomfort experienced when one receives more or less than others, which can be either advantageous or disadvantageous. Interestingly, advantageous inequity aversion does not always occur. For instance, individuals display advantageous inequity aversion when they actively decide this inequity, but it disappears when they passively receive it (O. Li et al., 2018). Conversely, research suggests that people often engage in social comparisons, striving to improve their relative standing (Festinger, 1954; Fiske, 2011; Starmans et al., 2017). This implies advantageous inequity liking, rather than aversion, might be present in certain contexts (Boyce et al., 2010; Cox, 2013; Dohmen et al., 2011). Thus, individuals may experience advantageous inequity aversion when considering betrayal, which usually leads to a more advantageous outcome and active harm, but presented advantageous inequity liking when contemplating reciprocity.

Reciprocity is a social norm where individuals feel compelled to return the trust or kindness extended to them (Komorita et al., 1992). While reciprocity can alleviate negative feelings such as guilt and inequity aversion, it often comes at a cost, typically requiring the individual to sacrifice their own economic interests. Engaging in reciprocity involves resolving a social dilemma—balancing the motive to reciprocate against the desire to maximize personal gain.

Eye Movement

In addition to computational modeling approaches, eye-tracking technique provides a window to observe the implicit decision-making processes that are not directly observable through behavioral response (Wedel et al., 2023). Eye-tracking offers various basic metrics including fixations, saccades, and pupil size (Skaramagkas et al., 2023). Noteworthily, gaze patterns, which represent transitions between areas of interest, capture eye movements during decision-making and reflect attentional processes (Wedel et al., 2023). These patterns can reveal the decision trade-offs between different payoff structures in economic games (Devetag et al., 2016).

Studies have demonstrated the power of eye-tracking in predicting decision outcomes. For instance, a recent study has shown that gaze patterns can predict consumer product liking decisions with 90%

accuracy (Palacios-Ibáñez et al., 2023). Interestingly, this gaze bias was found to be stronger for "Like" decisions compared to "Dislike" decisions (Mitsuda & Glaholt, 2014). In addition, eye movements also reflect individual characteristics. The gaze patterns of participants correspond to their social value orientations, with prosocial individuals paying more attention to the payoffs of other players compared to individualistic players (Jiang et al., 2016).

Although eye-tracking studies specifically investigating reciprocity decisions have been limited in the literature, this technique has great potential helping us to uncover the myth of reciprocity. Eye-tracking is especially valuable when combined with computational modeling in economic games, as it can help reflect and validate the psychological components underlying the decision-making processes. In binary TG, for instance, by applying eye-tracking while participants are making reciprocity decisions, we can observe how their focus moves through the six payoffs in the game (**Fig. 1 B**). Combining computational modeling, while the model identifies the psychological components, gaze pattern analyses based on eye-tracking data can provide observable validation to confirm the robustness of the according psychological components. For example, if the best-fitting model identifies the psychological component of guilt aversion in reciprocity decisions, with the estimated value indicating how much each participant weights guilt aversion in their decisions, gaze pattern analysis could verify this finding by showing whether participants with higher guilt aversion pay more attention to areas related to guilt estimation. The underlying assumption is that individuals who place greater importance on a component are likely to focus more on the relevant areas associated with that component's estimation.

2.5. Neural Mechanisms of Reciprocity

While computational modeling and eye-tracking provide insights into the psychological mechanisms of decision-making, neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), offer a comprehensive view of the neural basis of social decision-making.

Functional Magnetic Resonance fMRI Imaging

With its high spatial resolution, fMRI has been instrumental in mapping regional and dynamic alterations in brain metabolism (Glover, 2011). fMRI generally falls into two categories: task-based and task-free (resting state) fMRI. Both methods rely on the blood oxygen level-dependent (BOLD) response, the primary signal measured in fMRI (Ogawa et al., 1990; Smith et al., 2009). Task-based fMRI captures these BOLD changes while specific cognitive tasks are performed. In contrast, resting-state fMRI measures spontaneous fluctuations in the BOLD signal when participants are not engaged in any explicit tasks (Smith et al., 2009). While task-based fMRI reflects brain function related to active cognitive processes, resting-state fMRI is useful for understanding baseline brain function, which can be linked to neurological and psychological symptoms and traits.

Task-based fMRI. In studies of reciprocity, task-based fMRI is used when participants are engaged in reciprocity decision-making tasks, such as in TG. Research has shown that prefrontal cortex activity is heightened when participants interact with a human partner, as opposed to a computer partner, in TG, regardless of whether they are in the role of trustor or trustee (McCabe et al., 2001). Specifically, activity in the ventromedial prefrontal cortex (VMPFC) predicts reciprocity behaviors (J. Li et al., 2009), while the anterior medial prefrontal cortex (aMPFC) shows greater activity when participants choose to defect rather than reciprocate (van den Bos et al., 2009). In addition, the role of the dorsolateral prefrontal cortex (DLPFC) for reciprocity was found (van den Bos et al., 2011), but also for exploiting others for profit in TG (Bereczkei et al., 2015). Anterior insula (AI) (Baumgartner et al., 2009; Bellucci et al., 2017, 2018; Krueger et al., 2008; van den Bos et al., 2011) also has a critical role in reciprocity decision-making in TG. Other regions, such as TPJ (Krueger et al., 2008), inferior frontal gyrus (IFG) (Baumgartner et al., 2009; Bereczkei et al., 2015), anterior cingulate cortex (ACC) (Baumgartner et al., 2009; van den Bos et al., 2011) have also been reported to be involved in reciprocity decision-making in TG. In resolving the dilemma between a preference for reciprocity (or avoiding negative feelings associated with not reciprocating) and maximizing self-interest in reciprocity decision-making, brain regions such as DLPFC and ACC (associated with cognitive control to regulate selfish motive), MPFC and TPJ (for mentalizing other's intentions or expectations) and AI (linked to norm compliance or enforcement) are of utmost importance among all regions reported by previous studies.

Combining fMRI with computational modeling can provide researchers with further insights into the neurocomputational mechanisms of reciprocity. Guilt aversion has been linked to the activation of the insula, supplementary motor area, DLPFC, and TPJ (Chang et al., 2011). Building on this, van Baar et al. (2019) showed that guilt aversion was associated with neural activity in the AI, putamen, MPFC, and left DLPFC, while inequity aversion was linked to activity in the AI, VMPFC, and dACC. In contrast, Nihonsugi et al. (2015) demonstrated that right DLPFC activity correlates with guilt aversion, while ventral striatum and amygdala activity correlates with inequity aversion. Importantly, they showed that transcranial direct current stimulation (tDCS; a non-invasive brain stimulation technique that modulates neuronal activity by applying a low electrical current (Gebodh et al., 2019)) targeting the right DLPFC selectively enhances guilt aversion, suggesting a causal role for this region in modulating guilt aversion.

Task-based fMRI studies have provided insights into the neural mechanisms underlying reciprocity decision-making. More recent research that integrates fMRI with computational modeling has advanced our understanding by revealing the neurocomputational mechanisms involved. Building on these findings, further investigation focusing on both the core and periphery of reciprocity is expected to provide a more comprehensive view of the complex dynamics that drive reciprocal behavior.

Resting-state fMRI. In addition to task-based fMRI studies on reciprocity, researchers have also employed resting-state fMRI to examine how baseline brain connectivity patterns relate to reciprocity behavior. Resting-state functional connectivity (RSFC) is robust across sessions and aligns with coactivation patterns elicited by relevant task demands (Cao et al., 2014; Finn et al., 2015; Raichle, 2011, 2015; Zuo & Xing, 2014). It has increasingly been utilized to decode the heterogeneity of individual differences, such as cognitive functions (Finn et al., 2015; Frith et al., 2020; Yang et al., 2021), personality traits (Ren et al., 2021; Wang et al., 2021), and social behaviors (Bellucci et al., 2019; Shen et al., 2017). These individual characteristics are often mapped and interpreted at brain network level. The 160-node brain atlas is widely used (Dosenbach et al., 2010), which defines six networks (**Fig. 2**): cingulo-opercular network (CON), fronto-parietal network (FPN), default mode network (DMN), sensorimotor network (SMN), occipital network (OCCN), and cerebellum network. While the CON (primarily associated with saliency and cognitive control), FPN (cognitive control), DMN (mentalizing), SMN (motor function), and OCCN (visual processing) are widely involved in social behavior, the cerebellum network is less mentioned.

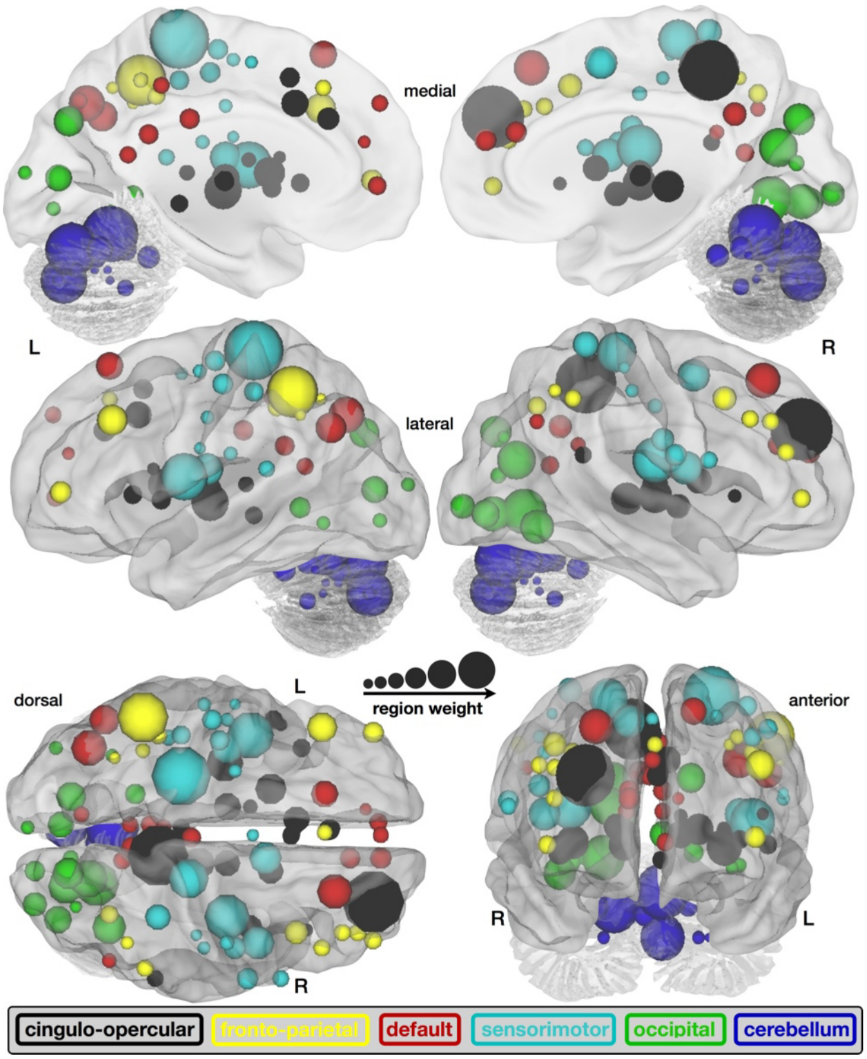


Figure 2. **The 160-node network adapted by** (Dosenbach et al., 2010). Six networks were identified: Cingulo-opercular, fronto-parietal, default, sensorimotor, occipital, and cerebellum network.

Applying independent component analysis (ICA) on RSFC, reciprocity has been associated with increased inter-network RSFC between the cingulo-opercular network (CON) and the frontoparietal network (FPN) (Cáceda et al., 2015). Beyond ICA analysis on RSFC and relating it to reciprocity behavior, another powerful method used to decode individual characteristics based on resting-state fMRI data is connectome-based predictive modeling (CPM). CPM is a data-driven approach that leverages whole-brain RSFC patterns to predict individual differences in behavior or cognitive traits (Finn et al., 2015; Shen et al., 2017). CPM can reveal stable, trait-like neural characteristics that predict individuals, capturing subtle individual differences in brain organization that may not be apparent in task-based studies.

The CPM pipeline involves several key steps (Finn et al., 2015; Shen et al., 2017). Typically, it starts with calculating whole-brain RSFC, where the brain is parcellated into regions of interest (ROIs), and the correlation between the BOLD time series of each pair of ROI is calculated, resulting in a connectivity matrix for each participant. To construct the prediction model, cross-validation procedures, such as leave-one-out-cross-validation (LOOCV) or tenfold-cross-validation are used. For example, during each iteration in LOOCV, the connectivity matrix and the behavior of interest to be predicted (target) of one participant was left out as test set, while all the remaining participants were used as training set. Feature selection is then performed to identify connections (features) that are significantly correlated with the target in the training set. A model, such as support vector regression, is then trained to capture the relationship between these selected features and the target. The trained model is then applied to a new participant to predict the behavior using the selected features in the test set. The training and predicting procedure are repeated N times (N is the number of total participants). At the end of this procedure, each participant receives one predicted value. The model's performance is assessed by metrics such as correlation coefficients between the actual and predicted targets. To validate the significance of the prediction, nonparametric permutation tests are often conducted.

Utilizing a similar prediction framework, Bellucci et al. (2019) found that not only whole-brain RSFC but also intra-network RSFC of the DMN, FPN, or CON can predict reciprocity propensity as measured by a one-shot TG. These findings suggest that these networks serve as neural biomarkers for reciprocity propensity. Building on these findings, applying CPM to resting-state fMRI data to decode the neural mechanisms of reciprocity, while distinguishing between core and periphery, offers a promising approach for advancing our understanding of the neural basis of reciprocity.

Electroencephalography

Complementary to fMRI, EEG captures electrical activity in the brain through electrodes placed on the scalp, offering insight into the high temporal resolution of neural activity (Sur & Sinha, 2009). Event-related potentials (ERPs) are derived from EEG data by averaging signals that are time-locked to specific events or stimuli (Luck, 2014). ERP enables researchers to examine distinct components associated with different cognitive processes.

Regarding decision-making, ERP studies have provided valuable insights into the neural dynamics involved in this process. ERP can help map the neural activity underlying a range of key functions, such as attention deployment, cognitive control, emotional responses, and regulation during decision-making. One early component of interest is the P2, a positive deflection that peaks around 100-250ms post-stimulus. The P2 is thought to reflect early selective attentional allocation in decision-making (Hajcak et al., 2012; Luck et al., 1994; Potts, 2004; Rey-Mermet et al., 2019). The N2, a negative deflection peaking around 200-350ms, is another important ERP component linked to effortful, deliberate cognitive control during decision-making processes (Cavanagh & Shackman, 2015; Folstein & Van Petten, 2008; Hao et al., 2023; McLoughlin et al., 2022). The late positive potential (LPP), a slow wave usually peaking after 400ms, is often considered a marker of emotional reactivity (Hajcak et al., 2010; MacNamara & Proudfit, 2014; Paul et al., 2016; Qi et al., 2016; Thiruchselvam et al., 2011), several studies have also linked it to emotional regulation, showing that enhanced LPP amplitudes reflect increased cognitive effort in managing emotional responses (Bernat et al., 2011; Desatnik et al., 2017; Moser et al., 2014; Shafir et al., 2015). Specifically, in decisions involving moral conflict, larger LPP amplitudes suggest the greater cognitive effort is exerted to resolve these conflicts (Chen et al., 2009; Zhan et al., 2018, 2020).

Although ERP studies focusing on the decision-making process of reciprocity are lacking, this process entails key functions such as selective attention (directing attention to the key components important to the decision-maker), cognitive control (resisting the temptation of exploiting other's trust), emotional regulation (managing negative feeling such as guilt aversion). The relevant ERP components as mentioned above could potentially serve as indicators of how these functions are involved in the decision-making of reciprocity.

2.6. Research Gaps

Despite substantial progress in understanding reciprocity, the neurocomputational mechanism underlying the core and the periphery of reciprocity remains unclear.

Firstly, although RSFC has been shown to predict individual differences in reciprocity propensity (Bellucci et al., 2019; Cáceda et al., 2015), these findings were obtained within specific contexts. This contextual constraint implies that the measured reciprocity propensity may already encompass both the core (reciprocity propensity) and peripheral aspects of reciprocity, thereby potentially confounding the distinct contributions of each.

Secondly, although the literature has informed about the peripheral (contextual) effects on reciprocity (Bohnet & Meier, 2005; Evans & van Beest, 2017; Keysar et al., 2008), the neurocomputational mechanisms underlying the peripheral effect of reciprocity remain unknown.

Lastly, while previous research has demonstrated that anxiety can reduce reciprocity in social interactions (Anderl et al., 2018; Rodebaugh et al., 2011, 2013), the neurocomputational mechanisms underlying the core and periphery of reciprocity which affected by anxiety are unclear.

2.7. Overview of this Dissertation

Research Aims

Grounded in social representation theory and utilizing TG paradigms, the ***overarching aim*** of this dissertation was to investigate the neurocomputational mechanism of reciprocity by employing a multi-method approach that integrates task-free and task-based fMRI, ERP, and eye-tracking (**Fig. 3**). Specifically, this dissertation seeks to delineate the core and periphery of reciprocity and elucidate how anxiety modulates the underlying core and periphery neurocomputational mechanisms.

Study 1 aimed to identify the neural underpinning that represents the core and periphery of reciprocity. Using resting-state fMRI, followed by a one-shot TG (framed as give and take context), this study examined how patterns of RSFC within or between specific networks relate to reciprocity behavior using connectome-based predictive modeling (CPM). This data-driven approach enabled us to pinpoint the neural mechanism that underlies the core and periphery of reciprocity.

Study 2 aimed to delve deeper into the periphery of reciprocity by investigating its neurocomputational mechanisms. By combining task-based fMRI with computational modeling in a multiple one-shot two-stage binary TG (framed as gain and loss context), this study identified the neural substrates associated with specific psychological components that underlie the subprocesses of reciprocity decision-making related to the periphery of reciprocity.

Finally, Study 3 aimed to understand the neurocomputational mechanism underlying anxiety modulates reciprocity on the core and periphery. Integrating EEG and eye-tracking in a multiple one-shot two-stage binary TG (framed as gain and loss context), this study investigated how anxiety influences the neural dynamics and computational mechanism underlying the core and periphery of reciprocity.

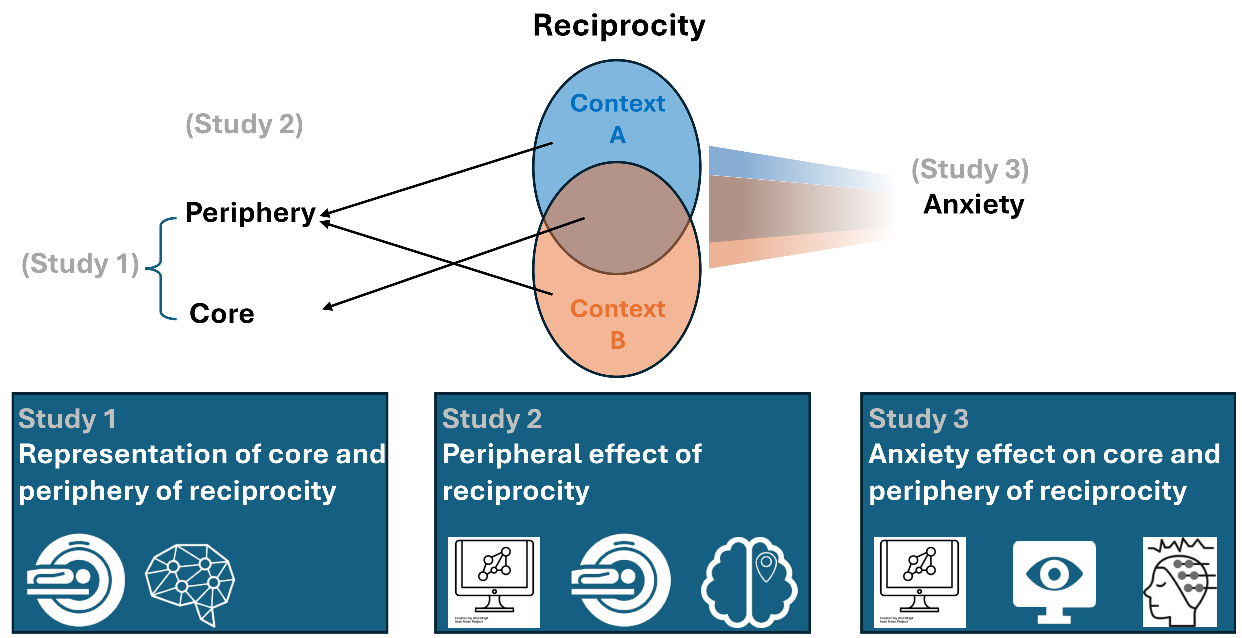

Reciprocity
Context A
Context B
(Study 2)
Periphery
(Study 1)
Core
(Study 3)
Anxiety
Study 1
Representation of core and periphery of reciprocity
Study 2
Peripheral effect of reciprocity
Study 3
Anxiety effect on core and periphery of reciprocity

Figure 3. Overview of research studies on the core and periphery of reciprocity in this dissertation. **Study 1** combined resting-state fMRI (brain network level) with connectome-based predictive modeling to investigate the core (stable propensity) and the periphery (those sensitive to contextual changes) of reciprocity. **Study 2** integrated computational modeling with task-based fMRI (brain region level) to explore the periphery of reciprocity. **Study 3** employed computational modeling, eye-tracking, and ERP to examine how the core and periphery of reciprocity are influenced by anxiety.

Working Hypotheses

In this dissertation, we hypothesized distinct neurocomputational mechanisms for the core and periphery of reciprocity and proposed that anxiety affects reciprocity through these mechanisms.

For study 1, we hypothesized that distinct brain patterns of functional connectivity would predict individual differences in the core and periphery of reciprocity. Specifically, the DMN, FPN, and CON were expected to be the contributing networks for predicting the core, while the DMN was hypothesized to be vital in predicting the periphery of reciprocity.

For Study 2, we anticipated that peripheral manipulation would significantly influence reciprocity behavior through modulation of specific psychological components and subprocesses of reciprocity decision-making. This periphery modulation was expected to engage specific brain regions.

For Study 3, we predicted that trait anxiety would affect both the core and periphery of reciprocity. Specifically, higher trait anxiety was expected to be associated with reduced overall reciprocity, involving altered selective attentional allocation and emotion regulation, reflecting its impact on the core. Additionally, trait anxiety was hypothesized to modulate the peripheral effect, involving cognitive control mechanisms.

3. Study 1: Connectome of Reciprocity's Core and Periphery

“Connectome-based individualized prediction of reciprocity propensity and sensitivity to framing: a resting-state functional magnetic resonance imaging study”

Fang, H., Liao, C., Fu, Z., Tian, S., Luo, Y., Xu, P., & Krueger, F. (2022). Connectome-based individualized prediction of reciprocity propensity and sensitivity to framing: a resting-state functional magnetic resonance imaging study. *Cerebral Cortex*, *33*(6), 3193-3206. https://doi.org/10.1093/cercor/bhac269

Cerebral Cortex, 2022, 1–14

https://doi.org/10.1093/cercor/bhac269

Original Article

Connectome-based individualized prediction of reciprocity propensity and sensitivity to framing: a resting-state functional magnetic resonance imaging study

Huihua Fang[1,2], Chong Liao[1,2], Zhao Fu[1], Shuang Tian[1], Yuejia Luo[1,3], Pengfei Xu[3,*], Frank Krueger[2,4]

[1]Shenzhen Key Laboratory of Affective and Social Neuroscience, Magnetic Resonance Imaging Center, Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen 518060, China,

[2]Department of Psychology, University of Mannheim, Mannheim 68131, Germany,

[3]Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (BNU), Faculty of Psychology, Beijing Normal University, Beijing 100875, China,

[4]School of Systems Biology, George Mason University, Fairfax, VA 22030, USA

*Corresponding author: Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (BNU), Faculty of Psychology, Beijing Normal University, Beijing 100875, China. Email: pxu@bnu.edu.cn

Background: The social representation theory states that individual differences in reciprocity decisions are composed of a stable central core (i.e., reciprocity propensity, RP) and a contextual-dependent periphery (i.e., sensitivity to the framing effect; SFE, the effect by how the decision is presented). However, the neural underpinnings that explain RP and SFE are still unknown.
Method: Here, we employed prediction and lesion models to decode resting-state functional connectivity (RSFC) of RP and SFE for reciprocity decisions of healthy volunteers who underwent RS functional magnetic resonance imaging and completed one-shot trust (give frame) and distrust (take frame) games as trustees.
Results: Regarding the central core, reciprocity rates were positively associated between the give and take frame. Neuroimaging results showed that inter-network RSFC between the default-mode network (DMN; associated with mentalizing) and cingulo-opercular network (associated with cognitive control) contributed to the prediction of reciprocity under both frames. Regarding the periphery, behavioral results demonstrated a significant framing effect-people reciprocated more in the give than in the take frame. Our neuroimaging results revealed that intra-network RSFC of DMN (associated with mentalizing) contributed dominantly to the prediction of SFE.
Conclusion: Our findings provide evidence for distinct neural mechanisms of RP and SFE in reciprocity decisions.

Key words: distrust game; framing effect; prediction; reciprocity; resting-state functional connectivity; trust game.

Introduction

Reciprocity, a social norm intrinsically possessed in human nature that one tends to return the act (positive or negative) given by others, forms the pillar of human interpersonal relationships (Chen et al. 2009; Caliendo et al. 2012). Reciprocity propensity (RP) is regarded as a stable personality trait, indicating how likely an individual reciprocates the actions of others (Li et al. 2017; Schino and Aureli 2017). However, when investigating RP in social decisions, the framing effect—the same objective facts are interpreted differently depending on the framing (Hagen and Hammerstein 2006)—is one of the most prominent factors that account for the instability of decisions (Alós-ferrer and Farolfi 2019). According to the social representation theory, a decision is composed of both a central core (i.e. implicit and stable) and periphery (i.e. explicit and contextual) (Wagenaar et al. 1988; Abric 1993; Hagen and Hammerstein 2006). While the central core (i.e. RP) represents the commonly shared basis of reciprocity, the periphery (i.e. framing of the decisions) is sensitive to and determined by context (Abric 1993). Framing a decision involves manipulating the periphery while keeping the central core unchanged (Abric 1993; Flachaire and Hollard 2008; Columbus et al. 2020).

The framing effect has been documented to influence social decisions. For example, participants are more cooperative when a Prisoners' dilemma game is framed as a "community game" rather than a "wall street game" (Kay and Ross 2003; Liberman et al. 2004) or as a "social exchange" rather than a "business transaction" (Batson and Moran 1999). In reciprocity decisions involving dictator games (where participants play first as the passive receiver and then as the dictator), people reciprocate more in the give compared to the take frame—although the objective outcomes are equivalent in both frames (Keysar et al. 2008). When making social decisions, people rely heavily on their psychological perceptions of social

Received: March 24, 2022. **Revised:** June 11, 2022. **Accepted:** June 12, 2022

actions instead of the objective values (Keysar et al. 2008). The trust game (TG) and distrust game (DTG) assess both trust (or distrust) and reciprocity behaviors. Bohnet and Meier (2005) systematically compared the standard TG (representing a give frame) with the DTG (representing a take frame)—showing that people reciprocate more in the give than take frame. Reciprocators "punish" investors more severely for commission of distrust (i.e. taking back some investment) than for omission of trust (i.e. not giving more investment), suggesting that the inference of intentions plays a vital role in the framing effect during reciprocation (Bohnet and Meier 2005). Though RP is a stable trait rooted in an individual's personality, the framing effect is a non-negligible factor impacting people's reciprocity decisions (Bohnet and Meier 2005; Keysar et al. 2008). Since the evaluation of reciprocity decisions (i.e. core and periphery) in laboratory studies or real-life situations can inevitably involve specific contexts (e.g. describing the game under a give or take frame), individuals' differences in reciprocity decisions depend not only on the unchanging central core (i.e. RP) but also on the sensitivity to the manipulation of the periphery (i.e. sensitivity to the framing effect, SFE). However, studies so far only focused on the neural mechanism of reciprocity decisions under one specific frame [e.g. give frame, (Cáceda et al. 2015; Bellucci et al. 2019)]—leaving the influence of the framing effect on reciprocity decisions unexplored. Therefore, identifying the two components—central core: RP, periphery: SFE—affecting an individual's reciprocity decision, and investigating their underlying neuropsychological substrates can help to better reveal their sources of heterogeneity.

Resting-state functional connectivity (RSFC) has been increasingly used to decode the heterogeneity of individual differences, including cognitive functions (Finn et al. 2015; Frith et al. 2020; Yang et al. 2021), personality traits (Ren et al. 2021; Wang et al. 2021a), and social behaviors (Shen et al. 2017; Bellucci et al. 2019). RSFC assesses the temporal synchronization of the blood-oxygen-level-dependent (BOLD) signal across spatially distributed brain regions at rest (Woodward and Cascio 2015), which is robust across sessions and overlaps with coactivation patterns induced by relevant task demands (Raichle 2011, 2015; Cao et al. 2014; Zuo and Xing 2014; Finn et al. 2015). Studying the heterogeneity of reciprocity decisions in a one-shot TG with RSFC, Bellucci et al. (2019) showed that reciprocity could not only be predicted by whole-brain RSFC but also by single intra-network RSFC of the default-mode network (DMN), frontoparietal network (FPN), or cingulo-opercular network (CON). Combining independent component analysis with linear regression, Cáceda et al. (2015) revealed that increased inter-network RSFC between CON and FPN is associated with reciprocity as measured with a one-shot TG.

While DMN is linked with mentalizing [i.e. the ability to assess and evaluate one's own and others' actions based on internal mental states such as beliefs, thoughts, and emotions (Fonagy et al. 1998, 2018)] for reciprocity decisions (Wang et al. 2021b), FPN and CON are linked with cognitive control (Dosenbach et al. 2006, 2010; Hahn et al. 2015), vital for resolving the social dilemma when reciprocating between implementing prosocial preferences (Li et al. 2009; van den Bos et al. 2009; Cáceda et al. 2017; Bellucci et al. 2019) and maximizing self-benefits (Cáceda et al. 2015; Bellucci et al. 2019). Importantly, the two reciprocity studies reviewed above (Cáceda et al. 2015; Bellucci et al. 2019) may be influenced by the framing effect as they were carried out under a specific context (i.e. the give frame in TG). Hence, the neural mechanism underlying the stable central core (i.e. RP) that resists the change of periphery (i.e. framing) remains elusive. Regarding SFE, a seed-based cross-validated machine learning RSFC study showed that inter-network RSFC between DMN and CON regions predicts SFE under harm and help frames (Cui et al. 2021), while a task-based functional magnetic resonance imaging (fMRI) study found that FC within the DMN is associated with SFE under harm and help frames (Liu et al. 2020). However, the RSFC contribution to SFE in reciprocity decisions remains elusive.

Although previous studies have examined the contributions of RSFC to the prediction of reciprocity decisions (Cáceda et al. 2015; Bellucci et al. 2019), findings based on a single frame (e.g. give frame using the TG) might be biased by specific context and do not allow the differentiation of the central core (i.e. RP) from the periphery (i.e. SFE; Hagen and Hammerstein 2006; Alós–ferrer and Farolfi 2019). Moreover, Bellucci et al. (2019) examined whether intra-network RSFC can predict reciprocity behavior, while inter-network functional connectivity was not examined. Furthermore, Cáceda et al.'s (2015) results were based on a linear regression but not on a cross-validated prediction approach. Therefore, a more comprehensive and robust prediction framework is needed to separately decode the neural mechanism of RP and SFE. Finally, Cui et al. (2021) and Liu et al. (2020) provided evidence for the relationship between FC and SFE in moral decisions; however, the RSFC contribution to SFE in reciprocity decisions remains elusive.

In this study, we employed two anonymous one-shot economic exchange games under two frames [TG (give frame) and DTG (take frame)]. Using cross-validated connectome-based predictive modeling and computational lesion approaches, we examined the predictive role of whole-brain and inter-network in the central core (i.e. RP) and the periphery (i.e. SFE) of reciprocity decisions. In terms of the central core of a reciprocity decision, we hypothesized a positive relationship of reciprocity rates between the give (TG) and take (DTG) frames at the behavioral level since the theory of social representation assumes that the central core of reciprocity (RP) is stable across contexts (Wagenaar et al. 1988; Abric 1993). At the neural level, we assumed a similar predictive RSFC pattern for reciprocity decisions in both the give and take frames. Given that reciprocity requires the engagement of mentalizing

(Li et al. 2009; van den Bos et al. 2009; Cáceda et al. 2017) and the suppression of selfish motives (Cáceda et al. 2015; Bellucci et al. 2019), we postulated that DMN, as well as FPN and CON, are the contributing networks, respectively, for the prediction of reciprocity regardless of frames. In terms of the periphery of a reciprocity decision, we hypothesized that participants reciprocate more in the give than in the take frame at the behavioral level, due to the fact that omission of trust in the give frame is perceived as more benign than commission of distrust in the take frame (Bohnet and Meier 2005; Keysar et al. 2008). At the neural level, we theorized that the DMN plays a vital role in the prediction of SFE—given that the ability to mentalize is necessary to identify the underlying intentions associated with different frames (Bohnet and Meier 2005; Keysar et al. 2008; Liu et al. 2020).

Materials and methods

Subjects

Ninety healthy volunteers (44 females, 46 males, age in years: 19.4 [mean, M] ± 0.16 [standard error of mean, SEM]) were recruited for the study who had no history of neurological and psychiatric disorders or head injury. Two participants were removed from the neuroimaging data analysis because of their head motions (under the criteria of mean frame-wise displacement exceeding 0.2 mm or more than 20% of the total number of volumes exceeding 2 mm maximum translation or 2 degree rotation). As a result, 88 participants (43 females, 45 males, age in years: 19.4 ± 0.16) were included in the neuroimaging data analysis. Our study was conducted according to the Declaration of Helsinki and approved by the local Ethics Committee at Shenzhen University, China. Participants gave written informed consent for the study and received compensation as a fixed show-up fee (¥100, approximately $16) and a variable monetary reward based on their game decisions (ranging from ¥0 to ¥40).

Experimental task

Participants completed two one-shot economic exchange games (TG, give frame; DTG, take frame) as trustees (Player B) (Bohnet and Meier 2005) (Fig. 1A). Each participant interacted with two different anonymous partners in a counterbalanced order, who acted as trustors (Players A). Participants completed the task after the trustors had made their decisions in a previous session (4.87 ± 0.14 days). Before the experiment, participants were familiarized with the game rules and completed a quiz to ensure their understanding of the games. For the give frame, both A and B were endowed with ¥10 (about $1.50). A decided first to send a portion of the endowed money to B (¥X, ranging from ¥0 to ¥10 in the step of ¥1), which was then tripled by the experimenter. Then, B decided to send a portion of the received money to A (send ¥Y from ¥10 + 3 * ¥X). For the take frame, only B was endowed with ¥40 and A with ¥0. First, A decided to take a portion of B's endowed money (¥X, ranging from ¥0 to ¥30 in steps of ¥3). The money taken was divided by three by the experimenter. Then, B decided to send a portion of the remaining money to A (send ¥Y from ¥40 to ¥X). In the view of participants (acting the role of B) in the present study, for example, A gave ¥6 from ¥10 in the give frame to the participant, which is logically and consequently the same as A took ¥12 from ¥30 in the take frame from the participant, because the participant owns ¥28 in both cases (i.e. give frame: ¥18 [trippled from ¥6] + ¥10 [original endowment]; take frame: ¥40 [original endowment] – ¥12 [taken by A]) before he/she makes the reciprocity decision. At the end of the experiment, the monetary reward for both players was determined based on their decisions randomly chosen from one of their games. The experiment included no deception, and the data for the trustor part will be reported in another publication.

Resting-state functional magnetic resonance imaging acquisition

Resting-state functional magnetic resonance imaging (RS-fMRI) images were acquired on the same day before the experimental task with a Siemens Magnetom Prisma 3 Tesla scanner equipped with a 64-channels head coil at the Shenzhen University Magnetic Resonance Imaging Center, Shenzhen, China. While completing the 12-min scan, participants were instructed to open their eyes, keep still, remain awake, and not think about anything systematically. A total of 720 contiguous volumes were acquired with a multi-band Echo planer imaging (EPI) sequence (axial slice, 65; slice thickness, 2.0 mm; multiband slice acceleration factor, 5; repetition time (TR) , 1,000 ms; echo time (TE) , 30 ms; flip angle, 90 degrees; voxel size, 2.0 × 2.0 × 2.0 mm^3; field of view (FOV) , 192 × 192 mm^2). In addition, high-resolution structural images were acquired through a 3D sagittal T1-weighted magnetization-prepared rapid acquisition with gradient-echo (MPRAGE) sequence (sagittal slices, 192; TR, 2,300 ms; TE, 2.26 ms; slice thickness, 1.0 mm; voxel size, 1.0 × 1.0 × 1.0 mm^3; flip angle, 8 degrees; FOV, 256 × 256 mm^2).

Behavioral analysis

The behavioral statistical analyses were performed using the R platform, version 4.1.1 (R Core Team 2020). A P-value less than 0.05 (two-tailed) was considered statistically significant. The reciprocity rate for each frame (give and take) was calculated as the ratio between the amount sent by B, and the amount B had before sending (Bohnet and Meier 2005; Bellucci et al. 2019). In addition, SFE was calculated as the reciprocity rate in the give frame minus the reciprocity rate in the take frame (Liu et al. 2020; Cui et al. 2021). The correlation between the reciprocity rate under the give and take frame was calculated to examine the central core of reciprocity behaviors (i.e. RP). To examine whether reciprocity behaviors were

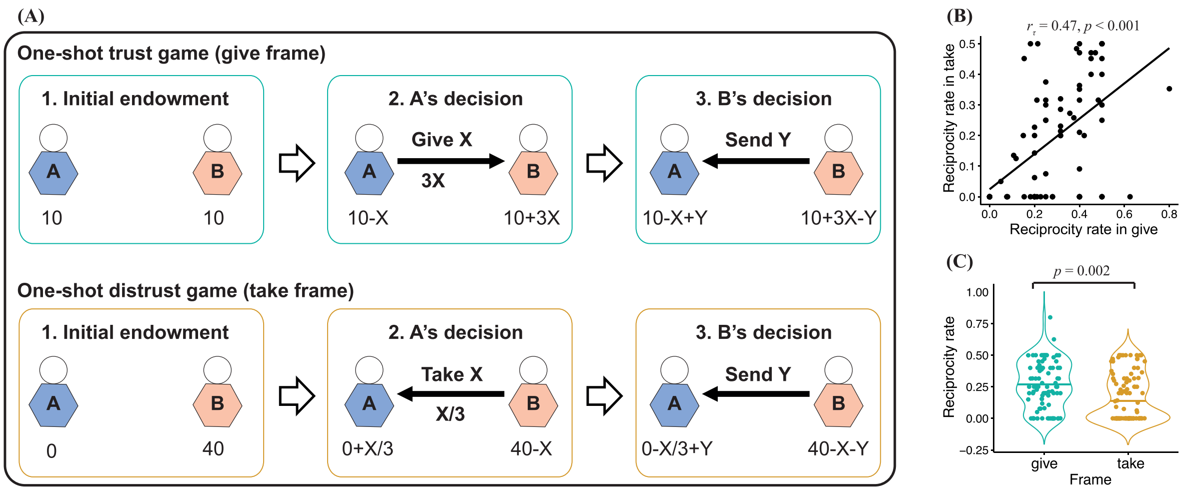


Fig. 1. Experimental task and behavioral results. (A) Acting as a trustee (player B), participants played a one-shot trust game and a one-shot DTG in counterbalanced order with two different anonymous trustors (player A). (B) Participants who reciprocated higher in the give frame tended to reciprocate higher in the take frame, suggesting a stable central core (i.e. RP) for both frames. (C) The reciprocity rate in the give frame was higher than that in the take frame, suggesting that manipulation of the periphery (i.e. frame) can influence reciprocity decisions. Note: X, the amount player a gives or takes; Y, the amount player B sends; RP, reciprocity propensity.

influenced by periphery manipulation, the difference in the reciprocity behavior was compared between the give and take frames. Reciprocity decisions in the give and take frames were tested for normality (Shapiro–Wilk test) and homogeneity (Levene's test). If assumptions for normality or homogeneity were violated, Kendall correlation and paired-sample Wilcoxon test were performed instead of Pearson correlation and paired-sample t-test.

Image analysis

Image preprocessing

Neuroimaging data analyses were performed with the DPABI software plug-in package (http://rfmri.org/dpabi) based on SPM12 (http://www.fil.ion.ucl.ac.uk/spm). The first 10 volumes of the functional images were discarded to allow for signal equilibrium. The images were then slice-time corrected and realigned for head movement correction. Structural brain images were co-registered to their mean functional images and were subsequently segmented. The deformation fields derived from anatomical segmentation were used to normalize each individual's functional images into the standard Montreal Neurological Institute (MNI template, resampling voxel size was 2 × 2 × 2 mm^3) space. Common nuisance variables were regressed out—including white matter signal, cerebrospinal fluid signal (Fox et al. 2005; Snyder and Raichle 2012), and Friston's 24 movement regressors (6 head motion parameters, 6 head motion parameters one time point before and 12 corresponding squared items) (Friston et al. 1996). The linear trends of time courses were removed. Band-pass filtering (0.01–0.1 Hz) was applied to the time series of each voxel to reduce the effect of low-frequency drifts and high-frequency physiological noise (Biswal et al. 1995; Zuo et al. 2010). Finally, functional images were spatially smoothed using a Gaussian filter (4 × 4 × 4 mm^3 full widths at half maximum) to decrease spatial noise.

Analysis procedure

To examine whether RSFC can predict individual differences in RP, whole-brain predictive models for reciprocity decisions in the give and take frames were tested, respectively. Then, lesion analyses were performed to confirm whether both frames exhibit a similar pattern and common key networks contributing to the prediction. Finally, it was tested whether the predictive model in one frame (e.g. give) can predict the other (e.g. take) to further confirm the neural mechanism of the central core. To examine how RSFC can explain the individual difference in SFE, the same procedure was performed in the form of whole-brain predictive models and lesion analyses.

Functional network construction

The workflow for individualized predictive modeling of reciprocity in the give and take frames and SFE was the same (Fig. 2). A 160 node atlas defined by Dosenbach was used for the RSFC feature extraction (Dosenbach et al. 2010). The nodes of the cerebellum network were removed since the image collection of several participants did not cover the entire cerebellum. The reduced atlas of 142 nodes (each node with a 5-mm sphere) consisted of five RSFC networks, including CON, DMN, FPN, sensorimotor network (SMN), and occipital network (OccN). For each participant, the time course of each node was computed by averaging the BOLD signal of all voxels within the node at each time point. RSFC was then computed in the form of a Pearson correlation between the time courses of each pair of nodes. To improve the normality of the correlation coefficients, Fisher's Z transformation was performed—resulting in a 142 × 142 symmetric matrix that represented connections (edges) in the RSFC profile for each participant.

Linear support vector regression modeling

Linear support vector regression (LSVR) models were employed to test the prediction of whole-brain RSFC on reciprocity behavior in the give and take frames and SFE. For each of the prediction models, a leave-one-out-cross-validation (LOOCV) procedure was implemented, where one participant was left out as a test sample in every iteration while all the other participants were used as training samples. For the whole-brain prediction, Kendall correlation between each pair of nodes in the 142 × 142 symmetric matrices (10,011 pairs in total) and target behavior was computed in the training sample. The Kendall correlation was used because of the normality violation of reciprocity rate in both frames and SFE. Edges of the top 1% of the strongest correlation were selected as the most relevant features (Shen et al. 2017; Bellucci et al. 2019). The selected features were rescaled into a range of 0–1 across the training samples, and the same rescaled procedure was performed in the testing sample to avoid some features with greater ranges dominating other features (Cui et al. 2018; Cui and Gong 2018). The LSVR model was trained using the selected features and their associated behaviors from training samples, whereas the trained model was used to predict the behaviors in the testing sample. The training and predicting procedure were repeated 88 times (i.e. the total number of participants included in the imaging analysis) so that each participant was used once as a test sample.

Model validation

To evaluate the performance of the prediction models, Kendall correlation coefficients (r_τ) and mean square error (MSE) were computed between the actual and predicted behaviors. For the significant prediction models, nonparametric permutation tests were further applied to validate the significance of the prediction. The significance of r_τ and MSE was evaluated 1,000 times employing nonparametric permutation tests, i.e. the behavior of participants was permutated 999 times without replacement where the behavioral outcome was shuffled randomly during each permutation, and the LOOCV procedure was performed. As a result, a distribution of 1,000 (true 1 plus shuffled 999) r_τ and MSE was produced, which indicated the null hypothesis. The *P*-values for the permutation tests were calculated as the number of times that the permutated performance is better (higher for r_τ and lower for MSE) than the true performance and then divided by 1,000.

Tenfold cross-validation

Tenfold cross-validation (TFCV) was employed to validate the results since a LOOCV approach biases the prediction (Poldrack et al. 2020). Thus, participants were separated into 10 subsets, 9 of which were utilized for training and the remaining 1 for testing. The training data was scaled before being used to train an LSVR prediction model, which was then used to predict the scaled testing data. This approach was performed 10 times to ensure that each subset was tested just once. Finally, across all individuals, the correlation r_τ and MSE between true and predictive values were determined. Because the entire dataset was divided into 10 subsets at random, performance may have been influenced by data division. Therefore, the TFCV was performed 100 times, and the results were averaged to get a final prediction performance. The significance of the prediction performance was tested 1,000 times using a permutation test.

Computational lesion prediction

To test the importance of intra- and inter-network connectivity in the prediction, computational lesion prediction analyses were performed (Feng et al. 2018; Wang et al. 2021a). For example, to computationally lesion the inter-network of DMN-CON, all edges belonging to the DMN-CON inter-network were set to 0 in the 142 × 142 symmetric matrices (10,011 pairs in total). Then, the lesion prediction procedure was conducted following the same method as in whole-brain prediction. The predictive power was compared between the whole-brain prediction with the accordingly lesion prediction using Steiger's *Z* (Steiger 1980). If the predictive power in lesion prediction was significantly lower than the whole-brain prediction, the lesion network connectivity was considered a significant contributor to the whole-brain prediction (Feng et al. 2018; Wang et al. 2021a).

Commonality validation

To further validate the central core of reciprocity (i.e. RP), it was tested whether the predictive edges (edges that survived all iterations in the whole-brain prediction) within common network connectivity in one frame can predict the reciprocity in the other frame. For example, if RSFC of DMN-CON contributed significantly to the prediction for both the give and take frame, predictive edges in DMN-CON in one frame were extracted to predict the other frame. Then, the model in the give (or take) frame was used to predict the reciprocity rate in the take (or give) frame, and the predictive performance was evaluated by the Kendall correlation between the actual and predicted behaviors.

Results

Behavioral results

Testing for normality and homogeneity, Shapiro–Wilk tests revealed that the reciprocity rates for both frames violated the normality assumption [give: $W(90) = 0.94$, $P < 0.001$, take: $W(90) = 0.83$, $P < 0.001$]. Levene's tests for homogeneity showed that the variance of the reciprocity rates was not significantly different between both frames ($F(1{,}178) = 0.83$, $P = 0.364$). To confirm the central core of reciprocity (i.e. RP), Kendall correlation analysis revealed

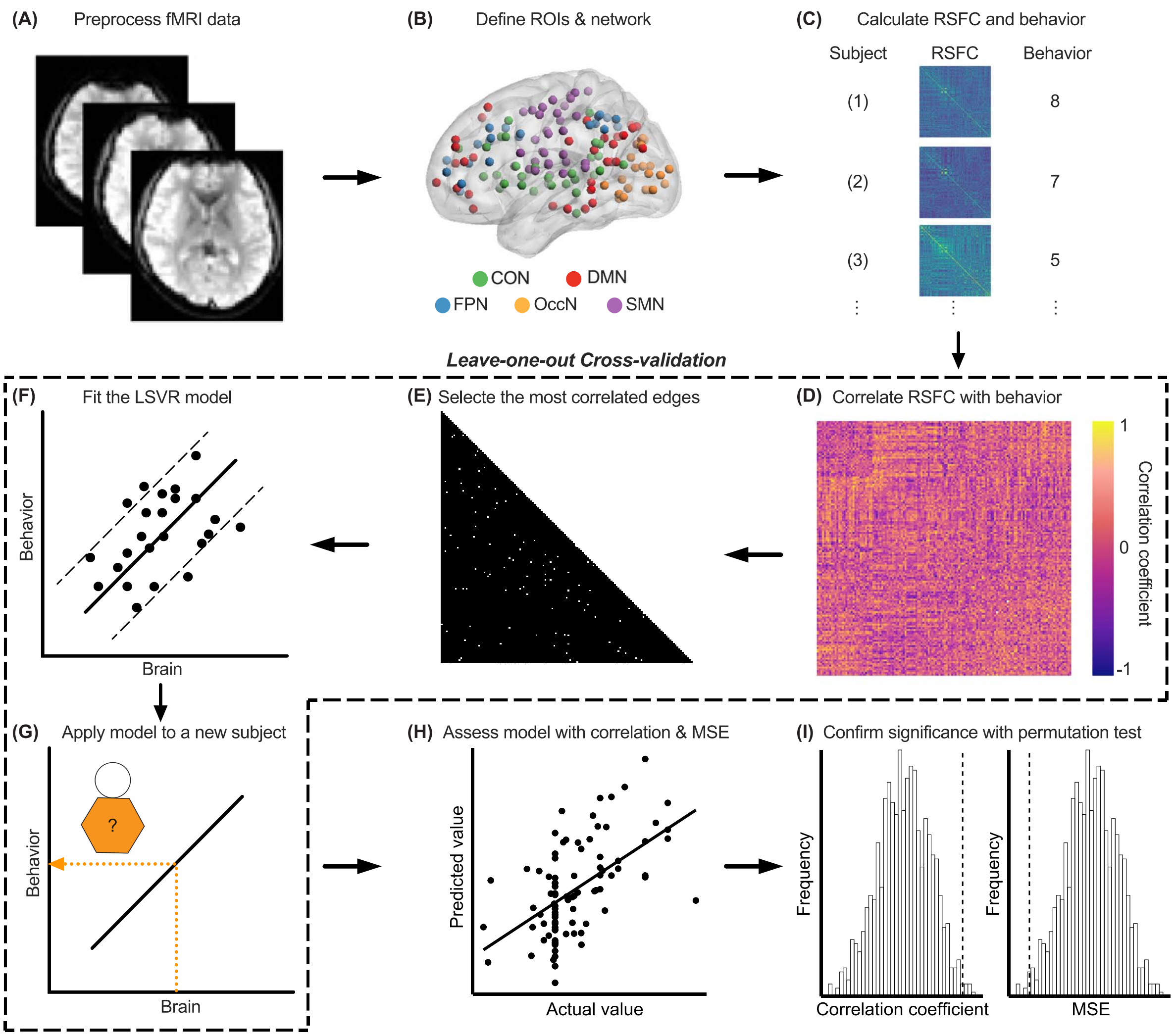

(A) Preprocess fMRI data
(B) Define ROIs & network
CON
DMN
FPN
OccN
SMN
(C) Calculate RSFC and behavior
Subject
RSFC
Behavior
(1)
8
(2)
7
(3)
5
Leave-one-out Cross-validation
(F) Fit the LSVR model
Behavior
Brain
(E) Selecte the most correlated edges
(D) Correlate RSFC with behavior
1
0
-1
Correlation coefficient
(G) Apply model to a new subject
?
Behavior
Brain
(H) Assess model with correlation & MSE
Predicted value
Actual value
(I) Confirm significance with permutation test
Frequency
Correlation coefficient
Frequency
MSE

Fig. 2. The workflow of individual predictive modeling. (A) BOLD RS-fMRI data were preprocessed. (B) RSFC ROIs and the network was defined based on the Dosenbach atlas (five networks: DMN, FPN, CON, SMN, and OccN). (C) RSFC (features) and behavioral scores (targets: reciprocity rate in the give or take frame or SFE) for each participant were calculated. As shown within the area of dash-line, cross-validation (i.e. LOOCV and TFCV) was implemented for prediction (in LOOCV, each participant was used as a test sample once while the remaining participants were used as a training sample; in TFCV, one subset of participants were used as test sample once while the other nine subsets of participants were used as training sample). In cross-validation, (D) the correlation between RSFC and behavior in the training sample was calculated, and (E) the most relevant edges (selected feature, top 1% strongest correlated edges) were selected to (F) train the LSVR model. (G) The trained model was applied to predict the behavior in the test samples. After the cross-validation, (H) the performance of the predictive model was assessed by the Kendall correlation and the MSE between the actual and predictive value. (I) a permutation test was further implemented to confirm the significance of the predictive model, where the P-value is calculated as the number of times that the permutated performance is better (higher for correlation coefficient or lower for MSE) than the true performance (correlation coefficient or MSE as indicated by dash line) and then divided by 1,000. Note: BOLD, blood-oxygen-level-dependent; RS-fMRI, resting-state functional magnetic resonance imaging; RSFC resting-state functional connectivity; ROI, regions of interest; CON, cingulo-opercular network; DMN, default-mode network; FPN, frontoparietal network; OccN, occipital network; SMN, sensorimotor network; SFE, sensitivity to framing effect; LOOCV, leave-one-out-cross-validation; TFCV, tenfold-cross-validation; LSVR, linear support vector regression; MSE, mean square error.

that the reciprocity in both frames was positively correlated ($r_\tau(88) = 0.47$, $P < 0.001$; Fig. 1B), suggesting that participants who reciprocated more in one frame also reciprocated more in the other frame. For the influence of peripheral manipulation (i.e. SFE), paired-sample Wilcoxon tests revealed that participants reciprocated significantly more in the give than in the take frame ($Z = 2{,}997$, $P = 0.002$, effect size $(r) = 0.45$; Fig. 1C). See Supplementary for a comparison of money available before decision-making between the two frames.

Neuroimaging results

Whole-brain RSFC prediction

For the central core, the LOOCV LSVR prediction procedure revealed that whole-brain RSFC significantly predicted reciprocity in the give ($r_\tau(86) = 0.30$, $P_{\text{perm}} = 0.009$; $\text{MSE} = 0.032$, $P_{\text{perm}} = 0.005$; Fig. 3) and take ($r_\tau(86) = 0.33$, $P_{\text{perm}} = 0.015$; $\text{MSE} = 0.028$, $P_{\text{perm}} = 0.004$; Fig. 4) frames. The inter-network RSFC of DMN-CON contributed the highest number of predictive edges in predicting reciprocity behavior under both frames (give: 13 edges, take: 16

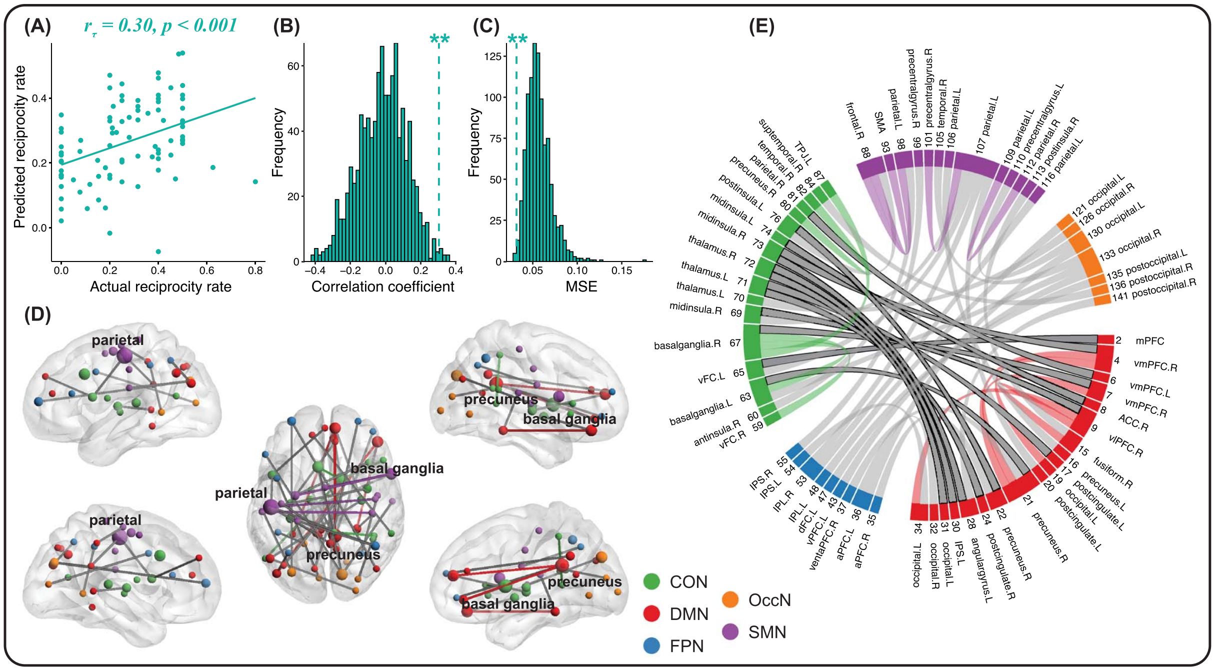


Fig. 3. Whole-brain functional connectivity prediction under the give frame. (A) the predicted model was significant as assessed by Kendall correlation between the actual reciprocity rate and its predictive value. (B) the predicted model was significant as assessed by permutation distribution of correlation coefficient and (C) MSE, where the P-value is calculated as the number of times that the permutated performance is better (higher for correlation coefficient or lower for MSE) than the true performance (correlation coefficient or MSE as indicated by dash line) and then divided by 1,000. (D) Predictive edges that survived all iterations in leave-one-out cross-validation are shown with brain connectomes plot (larger size of the node indicates that higher number of edges are connected to the node), and (E) circle plot (predictive edges in DMN-CON were highlighted using black outlines, which can also predict reciprocity rate under the take frame). Note: see Supplementary Table S1 for the name of abbreviations in plot e. MSE, mean square error; CON, cingulo-opercular network; DMN, default-mode network; FPN, frontoparietal network; OccN, occipital network; SMN, sensorimotor network. *: $P < 0.05$; **: $P < 0.01$.

edges; Fig. 6A, also see Supplementary Fig. S2a for results calculated as the ratio between predictive edges and total edges). For the periphery, the cross-validated LSVR prediction procedure revealed that whole-brain RSFC significantly predicted SFE of reciprocity ($r_\tau(86) = 0.41$, $P_{perm} = 0.004$, $MSE = 0.028$, $P_{perm} = 0.001$; Fig. 5), where networks pairs (e.g. SMN-CON, SMN-SMN, OccN-FPN, DMN-CON, DMN-DMN, DMN-OccN) nearly contributed in terms of predictive edges numbers (Fig. 6B, also see Supplementary Fig. S2b for results calculated as the ratio between predictive edges and total edges).

Tenfold cross-validation

TFCV was conducted to further validate the prediction performance. Consistent with the results of LOOCV, TFCV revealed that whole-brain RSFC can significantly predict reciprocity in the give ($r_\tau(86) = 0.17$, $P_{perm} = 0.009$; $MSE = 0.033$, $P_{perm} = 0.014$) and take ($r_\tau(86) = 0.13$, $P_{perm} = 0.019$; $MSE = 0.034$, $P_{perm} = 0.014$) frame as well as SFE ($r_\tau(86) = 0.24$, $P_{perm} = 0.001$; $MSE = 0.027$, $P_{perm} = 0.001$).

Computational lesion prediction

Lesion prediction analyses were conducted to examine whether specific intra- or inter-networks contributed significantly to the prediction. To examine the significant contributing network connectivity in predicting reciprocity behavior, edges within each intra- or inter-network that paired with DMN, CON, or FPN were lesioned (based on the descriptive plots in Fig. 6A and intra-networks of DMN, CON, and FPN predicting reciprocity (Bellucci et al. 2019). As a result, 12 lesion prediction analyses were performed for each frame. The models' prediction power (r_τ) dropped significantly after removing the inter-network RSFC of DMN-CON (Steiger's $Z = 2.84$, $P = 0.004$) and DMN-FPN (Steiger's $Z = 4.14$, $P < 0.001$) under the give frame, and removing the inter-network RSFC of DMN-CON (Steiger's $Z = 7.99$, $P < 0.001$) under the take frame (Fig. 7A; also see Supplementary Table S2 for the summary of predictive edges in DMN-CON for RP). To examine the significant contribution of intra- or inter-network connectivity in predicting SFE of reciprocity, edges in each network pair were lesioned as suggested by the descriptive plot in Fig. 6B. The results showed that the prediction power of the model dropped significantly after removing the intra-network RSFC of DMN-DMN (Steiger's $Z = 2.84$, $P = 0.004$) (Fig. 7B; also see Supplementary Table S2 for the summary of predictive edges in DMN-DMN for SFE).

Commonality validation

Since inter-network RSFC of DMN-CON significantly contributed to the prediction of reciprocity under both

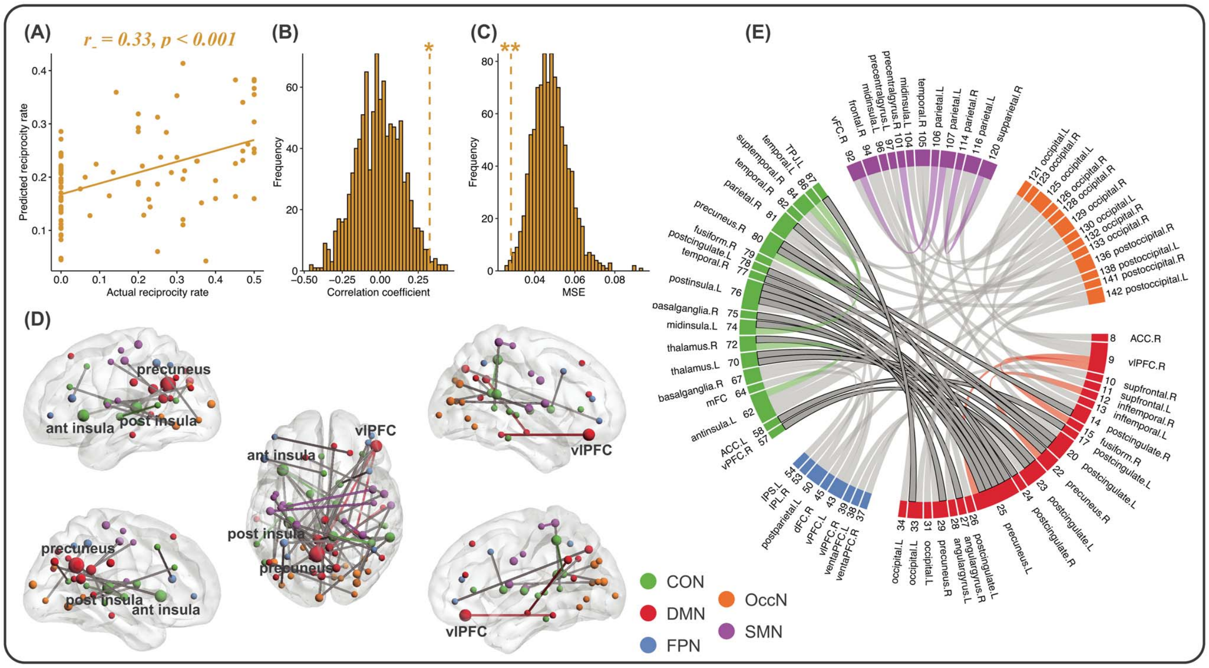


Fig. 4. Whole-brain functional connectivity predicts reciprocity under the take frame. (A) the predicted model was significant as assessed by Kendall correlation between the actual reciprocity rate and its predictive value. (B) the predicted model was further confirmed significant as assessed by permutation distribution of correlation coefficient and (C) MSE, where the *P*-value is calculated as the number of times that the permutated performance is better (higher for correlation coefficient or lower for MSE) than the true performance (correlation coefficient or MSE as indicated by dash line) and then divided by 1,000. Predictive edges that survived all iterations in leave-one-out cross-validation were shown with (D) brain connectomes plot (larger size of the node indicates that higher number of edges are connected to the node), and (E) circle plot (predictive edges in DMN-CON were highlighted using black outlines, which can also predict reciprocity rate under the give frame). Note: see Supplementary Table S1 for the name of abbreviations in plot e,. MSE, mean square error; CON, cingulo-opercular network; DMN, default-mode network; FPN, frontoparietal network; OccN, occipital network; SMN, sensorimotor network. *: $P < 0.05$; **: $P < 0.01$.

frames, it was further investigated whether the predictive edges in one frame can predict the other frame. The results showed that the predictive reciprocity edges within inter-network RSFC of DMN-CON under the give frame can predict reciprocity under the take fame ($r_\tau = 0.23$, $P = 0.004$), and predictive reciprocity edges within inter-network RSFC of DMN-CON under the take frame can also predict reciprocity under the give fame ($r_\tau = 0.33$, $P < 0.001$).

Discussion

Combining economic exchange games measuring reciprocity behavior with a cross-validated connectome-based prediction framework, we investigated the framing effect in reciprocity and how the individual difference of RP (central core) and SFE (periphery) can be predicted by large-scale RSFC networks. Regarding the central core, we found that participants who reciprocated more in one frame also reciprocated more in the other frame, indicating that the decision to reciprocate may share a common process (i.e. RP). Our prediction analyses consistently demonstrated that whole-brain RSFC can predict reciprocity in both frames and the inter-network RSFC of DMN-CON contributed significantly to the prediction in both frames. Commonality validation further confirmed DMN-CON as the common network connectivity for both frames, suggesting the neural mechanism underlying RP. In addition, DMN-FPN contributed significantly to the prediction of reciprocity decisions under the give frame but not the take frame. In terms of the periphery, participants reciprocated more in the give than take frame, and SFE was predicted by whole-brain RSFC, with intra-network RSFC of DMN as the key contributor.

According to social representation theory, the central core of a decision should exhibit consistency while only the periphery is changing (Wagenaar et al. 1988; Abric 1993). Consistent with our first hypothesis, a significant association was found between the reciprocity under the give and take frame; thereby, supporting the social representation theory (Wagenaar et al. 1988; Abric 1993). In both one-shot games (TG and DTG), reciprocity requires one to resolve a social dilemma between engaging prosocial preferences and maximizing self-benefits (Li et al. 2009; van den Bos et al. 2009; Cáceda et al. 2015, 2017; Bellucci et al. 2019). Besides the motive to maximize self-benefit under both frames, the trustee has the prosocial incentive to repay the partner's trust (the partner could not have invested at all) in the TG of the give frame and none-distrust (the partner could have taken all the investment) in the DTG of the take frame.

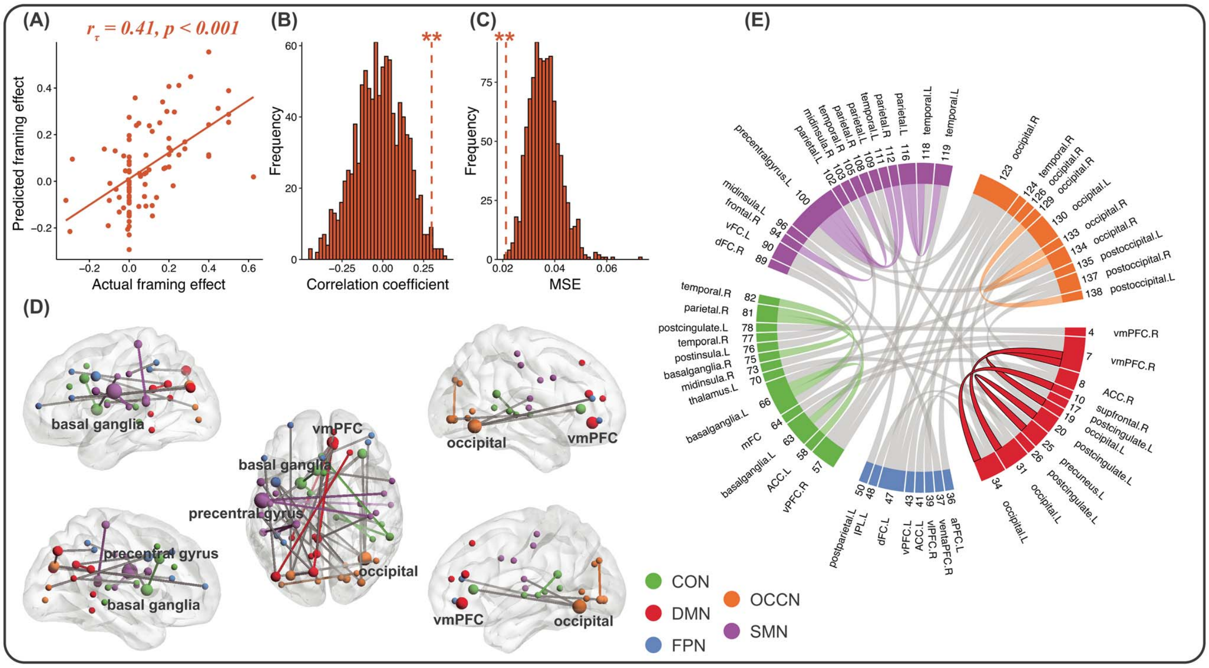
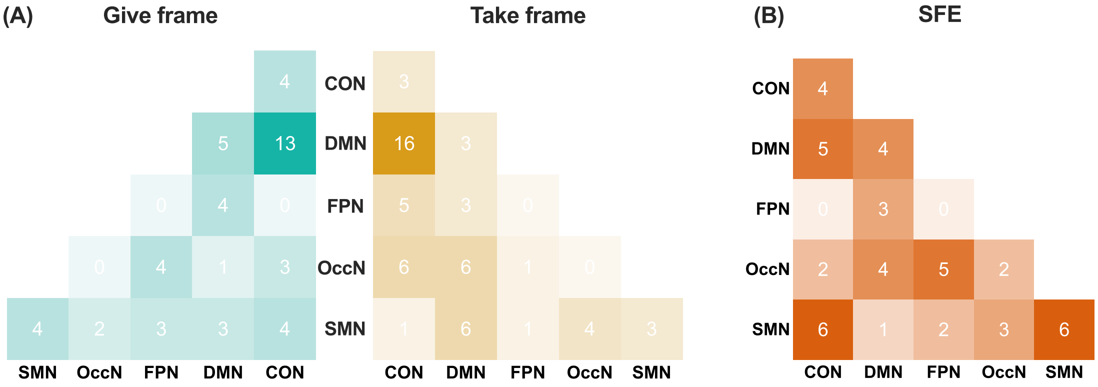


Fig. 5. Whole-brain functional connectivity predicts SFE. (A) the predicted model was significant as assessed by Kendall correlation between the actual SFE and its predictive value. (B) the predicted model was further confirmed significant as assessed by permutation distribution of correlation coefficient and (C) MSE, where the *P*-value is calculated as the number of times that the permutated performance is better (higher for correlation coefficient or lower for MSE) than the true performance (correlation coefficient or MSE as indicated by dash line) and then divided by 1,000. Predictive edges that survived all iterations in leave-one-out cross-validation were shown with (D) brain connectomes plot (larger size of the node indicates that higher number of edges are connected to the node), and (E) circle plot (predictive edges were highlighted using black outlines in DMN-DMN for the significant prediction contribution). Note: see Supplementary Table S1 for the name of abbreviations in plot e. SFE, sensitivity to framing effect; MSE, mean square error; CON, cingulo-opercular network; DMN, default-mode network; FPN, frontoparietal network; OccN, occipital network; SMN, sensorimotor network. *: $P < 0.05$; **: $P < 0.01$.



Give frame	SMN	OccN	FPN	DMN	CON
CON					4
DMN				5	13
FPN			0	4	0
OccN		0	4	1	3
SMN	4	2	3	3	4

Take frame	CON	DMN	FPN	OccN	SMN
CON	3				
DMN	16	3			
FPN	5	3	0		
OccN	6	6	1	0	
SMN	1	6	1	4	3

SFE	CON	DMN	FPN	OccN	SMN
CON	4				
DMN	5	4			
FPN	0	3	0		
OccN	2	4	5	2	
SMN	6	1	2	3	6

Fig. 6. Contributing network connectivity in whole brain prediction. (A) Contribution of network pair in predicting reciprocity under the give and take frame. (B) Contribution of network pair in predicting SFE of reciprocity. Note: the number and color darkness in each cell of the color matrix indicate the number of edges that survived all leave-one-out cross-validation iterations. SFE, sensitivity to framing effect; CON, cingulo-opercular network; DMN default-mode network; FPN, frontoparietal network; OccN occipital network; SMN, sensorimotor network.

Confronting such a dilemma, while mentalizing is needed to infer the intention of others to reciprocate properly, cognitive control is also critical to suppress the temptation of maximizing self-benefit (Fischbacher and Gächter 2010; Gächter et al. 2017; Bellucci et al. 2018). Partially confirming our second hypothesis, we found that reciprocity in both give and take frames can be predicted by whole-brain RSFC and share a similar predictive pattern. Importantly, as revealed by the lesion analysis, the inter-network RSFC of DMN-CON was the key contributor to predicting reciprocity under both frames. Consistent with the positive correlation between behaviors under the give and take frames, our results showed a similar RSFC network predictive pattern

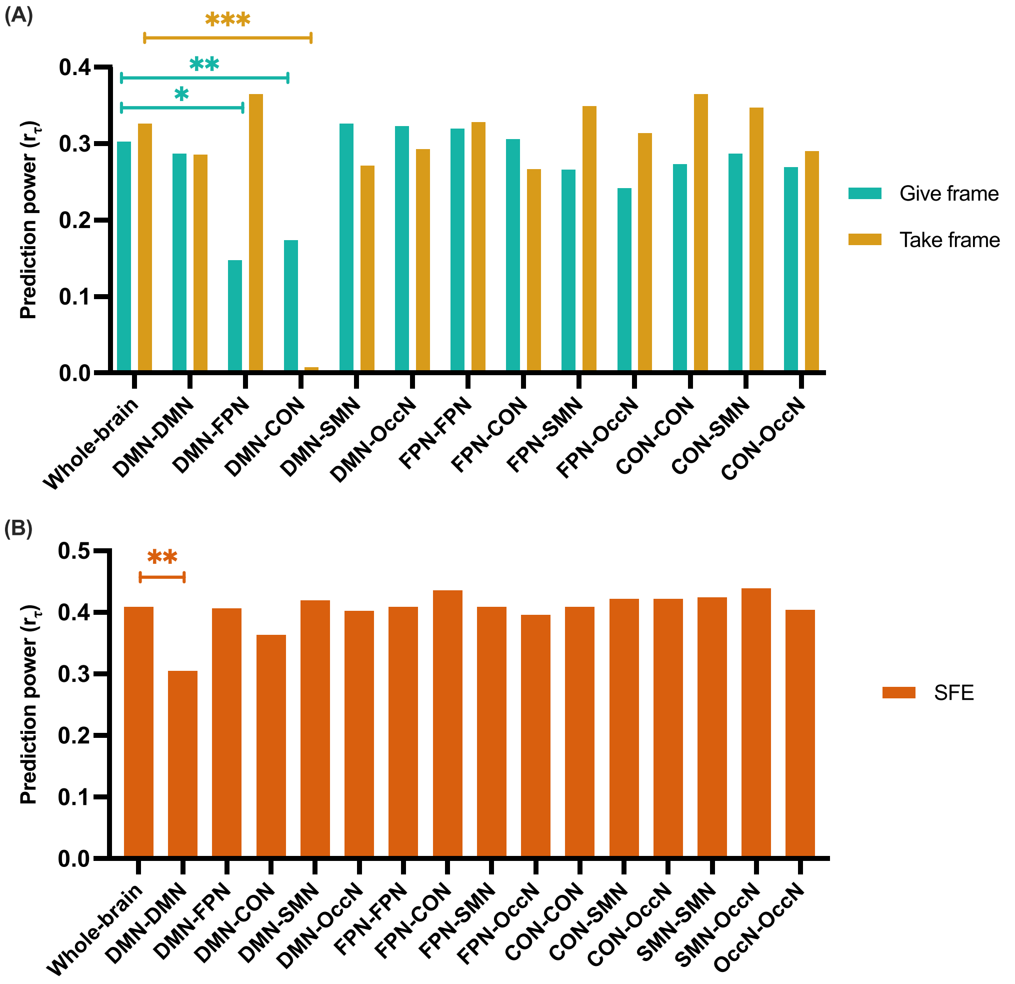


Fig. 7. Prediction power comparison after computational lesion. Comparisons of prediction power between lesion prediction and whole-brain prediction for (A) reciprocity in the give and take frames and (B) SFE. Note: SFE, sensitivity to framing effect; CON, cingulo-opercular network; DMN, default-mode network; FPN, frontoparietal network; OccN, occipital network; SMN, sensorimotor network. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$.

for both frames, with the DMN-CON inter-network as the key contributor. Our prediction results supported the social representation theory; a stable central core (i.e. RP) defines a reciprocity decision, though different peripheries can be presented.

RP was predicted by RSFC of the inter-network of DMN-CON. The DMN has been consistently linked with various social functions, including self-reference, autobiography, moral judgment, theory of mind, perspective-taking, and mentalizing (Whitfield-Gabrieli and Ford 2012; Raichle 2015; Yeshurun et al. 2021), whereas CON has been reliably linked with functions of cognitive control and salience processing (Dosenbach et al. 2006; Seeley et al. 2007; Fischbacher and Gächter 2010). In our study, DMN is possibly involved in mentalizing others' intentions to reciprocate in a proper manner since individuals are intrinsically encoded with prosocial preferences (Li et al. 2009; van den Bos et al. 2009; Cáceda et al. 2017; Bellucci et al. 2019). CON probably acts as a regulator for suppressing self-motives during reciprocation, as the function of cognitive control is essential to overcome the temptation of reaping benefits (Cáceda et al. 2015; Bellucci et al. 2019). Thus, the direct communication between DMN and CON predicting RP may represent the trade-off between the preferences to care for others and self-benefit preferences.

The DMN-CON was involved in reciprocity under both the give and take frames. Specifically, the DMN, including post cingulate cortex (Gobbini et al. 2007; Feldman 2015), precuneus (Gobbini et al. 2007; Feldman 2015), angular gyrus (Perry et al. 2011; Tanaka and Kirino 2019), and occipital cortex (Atique et al. 2011), has been linked with mentalizing, while the CON, consisting of thalamus (Dosenbach et al. 2007, 2008), parietal lobe (Roberts et al. 2010; Luijten et al. 2015), and ventral frontal cortex (Neubert et al. 2014; Loh et al. 2020), has been linked with cognitive control. In addition, the insula (Chang et al. 2011), thalamus (Fourie et al. 2014; Bastin et al. 2016), and basal ganglia (Wagner et al. 2011; Bastin et al. 2016) of CON have also been linked with feelings of guilt.

Therefore, we postulate the role of CON in suppressing self-interest to reciprocate might be driven by the need to avoid feelings of guilt, as guilt aversion is one of the critical components driving reciprocity decisions (Chang et al. 2011; Nihonsugi et al. 2015). Further study is needed to identify the role of these regions in connecting guilt aversion and cognitive control.

Importantly, besides the common regions in both frames, the reciprocity predictive edges in the give frame involve a more distributed DMN regions, including the vmPFC, mPFC, ACC, vlPFC, and IPS, while that in the take frame involves a more distributed CON regions, including precuneus, post cingulate cortex, TPJ, temporal cortex, and ACC. These results suggest that reciprocity under the give frame entails a more distributed mentalizing network, while reciprocity under the take frame entails a more distributed cognitive control network.

Noteworthy, we showed that whole-brain RSFC is predictive of reciprocity, contributed by inter-network RSFC of DMN-CON and DMN-FPN under the give frame. However, Bellucci et al. (2019) showed that intra-network RSFC of DMN, FPN, and CON alone can predict reciprocity under the give frame. One reason for this inconsistency might be that we applied a virtual computational lesion approach based on whole-brain RSFC to test the significant contribution of inter- and intra-network RSFC, whereas Bellucci et al. (2019) only used single intra-network RSFC to test significance predictions without considering the role of inter-network RSFC. Compared to a relatively small sample size ($n = 26$) and a LOOCV approach used in Bellucci et al. (2019), our results were based on a larger sample size ($n = 88$), and the LOOCV findings were confirmed with a more robust TFCV to avoid the overfitting problem in LOOCV (Poldrack et al. 2020). Combining with lesion prediction analyses, these analyses robustly showed the important roles of DMN, FPN, and CON, especially RSFC of the DMN with CON and FPN, in prediction of reciprocity under the give frame.

While both inter-network RSFC of DMN-CON and DMN-FPN contributed significantly to the prediction of reciprocity in the give frame, only RSFC of DMN-CON contributed significantly in the take frame in our study. On the one hand, although CON alongside the FPN is essential for top-down control, their differences have also been documented (Dosenbach et al. 2006, 2007, 2008, 2010; Hahn et al. 2015). The CON controls goal-directed behavior operating on a longer time scale, whereas the FPN monitors ongoing trial-by-trial processes (Dosenbach et al. 2006, 2007, 2008, 2010; Hahn et al. 2015). Since RP is a stable trait that lasts for a long period of time and is resistant to the change in context, the CON may be essential for shaping RP in terms of regulating one's self-motives. This may explain why DMN-CON was consistently found across different frames. On the other hand, RSFC of DMN-FPN only significantly contributed to the prediction of reciprocity in the give but not the take frame. Because people reciprocated more under the give than take frame, these results indicate that more cognitive control is engaged to suppress self-interest under the give frame, supporting the idea that the FPN is specific for monitoring the ongoing trial-by-trial processes (e.g. more context-sensitive).

As opposed to the stable property of the central core, the periphery of reciprocity decisions is conditional upon the context, and manipulation of the periphery can influence those decisions. Confirming our third hypothesis, our findings revealed that reciprocity was higher in the give frame than in the take frame. Despite the similarity between both frames, our result suggested that reciprocity decisions are susceptible to social framing. Framing (contextual manipulation, i.e. the periphery of the decision) significantly shifted individuals' reciprocity decision while RP (stable personality trait, i.e. the central core of the decision) remained well-aligned. Our results support the social representation theory that the framing (representing the periphery) and RP (representing the central core) are independent components in decision-making (Wagenaar et al. 1988; Abric 1993; Hagen and Hammerstein 2006). Though sharing the same payoff structure in both frames, perceived lack of trust when the default is not trusting in TG is different from perceiving distrust when the default is trusting in DTG (Kahneman and Tversky 1979; Thaler 1980; Samuelson and Zeckhauser 1988). For example, the trustor gave ¥5 from his ¥10 to the participant is logically and consequently, the same as the trustor took ¥15 from ¥30 because the trustee received ¥15 in both cases (the ¥5 was tripled in the first case). Nevertheless, the act of giving (trust) is perceived as relatively positive, whereas the act of taking (distrust) is perceived as relatively negative. People tend to "punish" more severely for commissions of distrust than omissions of trust. It has been suggested that mentalizing others' intention plays a key role in the difference in reciprocity between frames (Bohnet and Meier 2005). The more benign (or malign) the reciprocator perceives the partner's intentions of an action to be, the more reciprocator will reward (or punish) the partner (Rabin 1993; McCabe et al. 2003).

To explain why some people are more sensitive to the change of different frames than others, we further investigated how the whole-brain RSFC may predict SFE in reciprocity. Consistent with our fourth hypothesis, our neural network prediction results demonstrated that the whole-brain RSFC can predict SFE. Further, though the number of predictive edges is not the highest among all network pairs, the intra-network RSFC of DMN contributes significantly to the prediction, as revealed by the lesion analysis. It should be noted that a higher number of the predictive edges does not guarantee a significant contribution to the prediction. The brain regions of the DMN, including the occipital cortex (Atique et al. 2011), ventromedial prefrontal cortex (Lombardo et al. 2010; Atique et al. 2011), precuneus (Gobbini et al. 2007; Feldman 2015), post cingulate cortex (Gobbini et al. 2007; Feldman 2015), and superior frontal gyrus (Schneider-Hassloff et al. 2015), have been reported to be involved in mentalizing. Though various functions have been related to DMN as described above, it is suggested to be the core network of the social brain (Mars et al. 2012)—essential for assessment of social contexts before engaging in prosocial actions (Krueger et al. 2009) and moral decision-making (Greene et al. 2001). Consistent with a recent finding that FC within DMN is associated with the social framing effect (help vs. harm frame) (Liu et al. 2020), our results revealed that individuals with stronger intra-network RSFC of DMN tend to exhibit higher SFE. Stronger intra-network RSFC of DMN suggests a better capacity for mentalizing others' intentions, which is related to a higher SFE. Inconsistently, Cui et al. (2021) used a seed-based approach and showed that the framing effect is linearly correlated with RSFC between regions of DMN and CON. Importantly, our results were based on whole-brain RSFC and a more conservative procedure, including cross-validated predictive modeling and lesion approaches. Different results were found possibly because of different task paradigms (i.e. to make a tradeoff between economic benefits and the feeling of others) and framing (i.e. help vs. harm) was used in Cui et al. (2021) compared to the current study. However, further studies are needed to resolve this discrepancy.

Despite our novel findings, the current study had several limitations. First, our sample size was relatively small ($n = 88$), and larger sample sizes are needed to increase the accuracy or robustness of RSFC prediction studies (Cui and Gong 2018; Poldrack et al. 2020). Second, our RSFC prediction results can only be used to infer a correlational predictive relationship between brain connectivity and reciprocity behavior. A potential causal relationship should be examined in future studies, such as investigating the framing effect by stimulating outer-cortex DMN regions using transcranial magnetic stimulation (TMS) or inner-cortex DMN regions using low-intensity transcranial focused ultrasound stimulation (Tyler et al. 2018). Third, our study only manipulated one type of frame (i.e. give vs. take), employing the TG and DTG. Future studies should manipulate other types of frames such as "gain" and "loss" using the same game type to test whether the neural mechanism underlying the central core remains the same, given that the periphery is context-dependent while the central core is stable. Forth, though the current study identified the role of DMN-CON in RP, whether these RP predictive networks can predict other types of reciprocity or prosocial behavior should be tested in future studies.

To summarize, we examined whether the central core (i.e. RP) and the sensitivity to the manipulation of the periphery (i.e. SFE) of reciprocity behavior can be predicted by RSFC using two one-shot economic games (TG [give frame] and DTG [take frame]) that are structurally the same but framed differently. In terms of the central core, we observed a significant association between the reciprocity rate in the give and take frame—suggesting the existence of a central core for reciprocity behaviors. The inter-network RSFC of DMN-CON contributed significantly to the prediction of reciprocity under the two frames, indicating the interplay between mentalizing others' intentions and suppressing one's self-motives in processing the central core (i.e. RP) of reciprocity. Regarding the periphery, we found that manipulation of the periphery can significantly influence reciprocity behavior, leading to a significant framing effect (give vs. take). The intra-network RSFC of DMN contributed significantly to the prediction of SFE, indicating the ability of mentalizing is related to the sensitivity manipulation of the periphery.

In conclusion, our findings support the social representation theory that a central core and periphery constitute reciprocity behavior. We establish the RSFC network model predicting the central core and sensitivity to the periphery in reciprocity behavior. Our results advance the understanding of how large-scale RSFC networks can serve as biomarkers for an individual's social propensity and characteristic of sensitivity to the change in social context.

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Authors' contributions

HF, CL, PX, and FK designed the experiment. HF created the experiment. HF, CL, ZF, and ST collected the data. HF preprocessed and analyzed the data. HF and FK prepared the first draft of the article. All authors contributed to the final version.

Funding

This work was supported by the National Natural Science Foundation of China (31871137 and 31920103009), the Major Project of National Social Science Foundation (20&ZD153), Young Elite Scientists Sponsorship Program by China Association for Science and Technology (YESS20180158), Guangdong International Scientific Collaboration Project (2019A050510048), Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions (2019SHIBS0003), Shenzhen Science and Technology Research Funding Program (JCYJ20180507183500566 and JCYJ20180305124819889).

Conflict of interest statement: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abric J-C. Central system, peripheral system: their functions and roles in the dynamics of social representations. *Pap Soc Represent.* 1993:2:75–78.

Alós-ferrer C, Farolfi F. Trust games and beyond. *Front Neurosci*. 2019:13:1–14.

Atique B, Erb M, Gharabaghi A, Grodd W, Anders S. Task-specific activity and connectivity within the mentalizing network during emotion and intention mentalizing. *NeuroImage*. 2011:55: 1899–1911.

Bastin C, Harrison BJ, Davey CG, Moll J, Whittle S. Feelings of shame, embarrassment and guilt and their neural correlates: a systematic review. *Neurosci Biobehav Rev*. 2016:71:455–471.

Batson CD, Moran T. Empathy-induced altruism in a prisoner's dilemma. *Eur J Soc Psychol*. 1999:29:909–924.

Bellucci G, Feng C, Camilleri J, Eickhoff SB, Krueger F. The role of the anterior insula in social norm compliance and enforcement: evidence from coordinate-based and functional connectivity meta-analyses. *Neurosci Biobehav Rev*. 2018:92:378–389.

Bellucci G, Hahn T, Deshpande G, Krueger F. Functional connectivity of specific resting-state networks predicts trust and reciprocity in the trust game. *Cogn Affect Behav Neurosci*. 2019:19:165–176.

Biswal B, Zerrin Yetkin F, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med*. 1995:34:537–541.

Bohnet I, Meier S. Deciding to distrust. *Federal reserve bank of boston working paper*. 2005:5:4.

Cáceda R, James GA, Gutman DA, Kilts CD. Organization of intrinsic functional brain connectivity predicts decisions to reciprocate social behavior. *Behav Brain Res*. 2015:292:478–483.

Cáceda R, Prendes-Alvarez S, Hsu JJ, Tripathi SP, Kilts CD, James GA. The neural correlates of reciprocity are sensitive to prior experience of reciprocity. *Behav Brain Res*. 2017:332:136–144.

Caliendo M, Fossen F, Kritikos A. Trust, positive reciprocity, and negative reciprocity: do these traits impact entrepreneurial dynamics? *J Econ Psychol*. 2012:33:394–409.

Cao H, Plichta MM, Schäfer A, Haddad L, Grimm O, Schneider M, Esslinger C, Kirsch P, Meyer-Lindenberg A, Tost H. Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *NeuroImage*. 2014:84:888–900.

Chang LJ, Smith A, Dufwenberg M, Sanfey AG. Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron*. 2011:70:560–572.

Chen YR, Chen XP, Portnoy R. To whom do positive norm and negative norm of reciprocity apply? Effects of inequitable offer, relationship, and relational-self orientation. *J Exp Soc Psychol*. 2009:45:24–34.

Columbus S, Münich J, Gerpott FH. Playing a different game: situation perception mediates framing effects on cooperative behaviour. *J Exp Soc Psychol*. 2020:90:104006.

Cui Z, Gong G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage*. 2018:178:622–637.

Cui Z, Su M, Li L, Shu H, Gong G. Individualized prediction of reading comprehension ability using Gray matter volume. *Cereb Cortex*. 2018:28:1656–1672.

Cui F, Yang J, Gu R, Liu J. Functional connectivities of the right temporoparietal junction and moral network predict social framing effect: evidence from resting-state fMRI. *Acta Psychol Sin*. 2021:53:55.

Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE. A Core system for the implementation of task sets. *Neuron*. 2006:50: 799–812.

Dosenbach NUF, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RAT, Fox MD, Snyder AZ, Vincent JL, Raichle ME, et al. Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci U S A*. 2007:104:11073–11078.

Dosenbach NUF, Fair DA, Cohen AL, Schlaggar BL, Petersen SE. A dual-networks architecture of top-down control. *Trends Cogn Sci*. 2008:12:99–105.

Dosenbach NUF, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC, Lessov-Schlaggar CN, et al. Prediction of individual brain maturity using fMRI. *Science*. 2010:329: 1358–1361.

Feldman R. The adaptive human parental brain: implications for children's social development. *Trends Neurosci*. 2015:38: 387–399.

Feng C, Yuan J, Geng H, Gu R, Zhou H, Wu X, Luo Y. Individualized prediction of trait narcissism from whole-brain resting-state functional connectivity. *Hum Brain Mapp*. 2018:39: 3701–3712.

Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci*. 2015:18:1664–1671.

Fischbacher U, Gächter S. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am Econ Rev*. 2010:100:541–556.

Flachaire E, Hollard G. Individual sensitivity to framing effects. *J Econ Behav Organ*. 2008:67:296–307.

Fonagy P, Target M, Steele H, Steele M. *Reflective-functioning manual, version 5.0, for application to adult attachment interviews*. London: University College London; 1998. p. 10

Fonagy P, Gergely G, Jurist EL, Target M. *Affect regulation, mentalization, and the development of the self*. London: Routledge; 2018

Fourie MM, Thomas KGF, Amodio DM, Warton CMR, Meintjes EM. Neural correlates of experienced moral emotion: an fMRI investigation of emotion in response to prejudice feedback. *Soc Neurosci*. 2014:9:203–218.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A*. 2005:102:9673–9678.

Friston KJ, Williams S, Howard R, Frackowiak RSJ. Movement-related effects in fMRI time-series. *Magn Reson Med*. 1996:35:346–355.

Frith E, Elbich DB, Christensen AP, Rosenberg MD, Chen Q, Kane MJ, Silvia PJ, Seli P, Beaty RE. Intelligence and creativity share a common cognitive and neural basis. *J Exp Psychol Gen*. 2020:150: 609–632.

Gächter S, Kölle F, Quercia S. Reciprocity and the tragedies of maintaining and providing the commons. *Nat Hum Behav*. 2017:1: 650–656.

Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV. Two takes on the social brain: a comparison of theory of mind tasks. *J Cogn Neurosci*. 2007:19:1803–1814.

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. *Science*. 2001:293:2105–2108.

Hagen EH, Hammerstein P. Game theory and human evolution: a critique of some recent interpretations of experimental games. *Theor Popul Biol*. 2006:69:339–348.

Hahn T, Notebaert K, Anderl C, Reicherts P, Wieser M, Kopf J, Reif A, Fehl K, Semmann D, Windmann S. Reliance on functional resting-state network for stable task control predicts behavioral tendency for cooperation. *NeuroImage*. 2015:118:231–236.

Kahneman D, Tversky A. 1979. On the interpretation of intuitive probability: a reply to Jonathan Cohen. *Cognition*. 1979:7: 409–411.

Kay AC, Ross L. The perceptual push: the interplay of implicit cues and explicit situational construals on behavioral intentions in the prisoner's dilemma. *J Exp Soc Psychol*. 2003:39:634–643.

Keysar B, Converse BA, Wang J, Epley N. Reciprocity is not give and take: asymmetric reciprocity to positive and negative acts. *Psychol Sci*. 2008:19:1280–1286.

Krueger F, Barbey AK, Grafman J. The medial prefrontal cortex mediates social event knowledge. *Trends Cogn Sci*. 2009:13:103–109.

Li J, Xiao E, Houser D, Montague PR. Neural responses to sanction threats in two-party economic exchange. *Proc Natl Acad Sci U S A*. 2009:106:16835–16840.

Li X, Zhu P, Yu Y, Zhang J, Zhang Z. The effect of reciprocity disposition on giving and repaying reciprocity behavior. *Personal Individ Differ*. 2017:109:201–206.

Liberman V, Samuels SM, Ross L. The name of the game: predictive power of reputations versus situational labels in determining Prisoner's dilemma game moves. *Personal Soc Psychol Bull*. 2004:30: 1175–1185.

Liu J, Gu R, Liao C, Lu J, Fang Y, Xu P, Luo YJ, Cui F. The neural mechanism of the social framing effect: evidence from fMRI and tDCS studies. *J Neurosci*. 2020:40:3646–3656.

Loh KK, Procyk E, Neveu R, Lamberton F, Hopkins WD, Petrides M, Amiez C. Cognitive control of orofacial motor and vocal responses in the ventrolateral and dorsomedial human frontal cortex. *Proc Natl Acad Sci U S A*. 2020:117:4994–5005.

Lombardo MV, Chakrabarti B, Bullmore ET, Wheelwright SJ, Sadek SA, Suckling J, Consortium MA, Baron-Cohen S. Shared neural circuits for mentalizing about the self and others. *J Cogn Neurosci*. 2010:22:1623–1635.

Luijten M, Meerkerk G-J, Franken IHA, van de Wetering BJM, Schoenmakers TM. An fMRI study of cognitive control in problem gamers. *Psychiatry Res Neuroimaging*. 2015:231:262–268.

Mars RB, Neubert FX, Noonan MAP, Sallet J, Toni I, Rushworth MFS. On the relationship between the "default mode network" and the "social brain.". *Front Hum Neurosci*. 2012:6:1–9.

McCabe KA, Rigdon ML, Smith VL. Positive reciprocity and intentions in trust games. *J Econ Behav Organ*. 2003:52:267–275.

Neubert F-X, Mars RB, Thomas AG, Sallet J, Rushworth MFS. Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron*. 2014:81:700–713.

Nihonsugi T, Ihara A, Haruno M. Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. *J Neurosci*. 2015:35: 3412–3419.

Perry D, Hendler T, Shamay-Tsoory SG. Projecting memories: the role of the hippocampus in emotional mentalizing. *NeuroImage*. 2011:54:1669–1676.

Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry*. 2020:77:534–540.

R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020.

Rabin M. Incorporating fairness into game theory and economics. *Am Econ Rev*. 1993:1281–1302.

Raichle ME. The restless brain. *Brain Connect*. 2011:1:3–12.

Raichle ME. The brain's default mode network. *Annu Rev Neurosci*. 2015:38:433–447.

Ren Z, Daker RJ, Shi L, Sun J, Beaty RE, Wu X, Chen Q, Yang W, Lyons IM, Green AE, et al. Connectome-based predictive modeling of creativity anxiety. *NeuroImage*. 2021:225:117469.

Roberts RE, Anderson EJ, Husain M. Expert cognitive control and individual differences associated with frontal and parietal white matter microstructure. *J Neurosci*. 2010:30:17063–17067.

Samuelson W, Zeckhauser R. Status quo bias in decision making. *J Risk Uncertain*. 1988:1:7–59.

Schino G, Aureli F. Reciprocity in group-living animals: partner control versus partner choice. *Biol Rev*. 2017:92:665–672.

Schneider-Hassloff H, Straube B, Nuscheler B, Wemken G, Kircher T. Adult attachment style modulates neural responses in a mentalizing task. *Neuroscience*. 2015:303:462–473.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci*. 2007:27:2349–2356.

Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, Constable RT. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat Protoc*. 2017:12:506–518.

Snyder AZ, Raichle ME. A brief history of the resting state: the Washington University perspective. *NeuroImage*. 2012:62:902–910.

Steiger JH. Tests for comparing elements of a correlation matrix. *Psychol Bull*. 1980:87:245–251.

Tanaka S, Kirino E. Increased functional connectivity of the angular gyrus during imagined music performance. *Front Hum Neurosci*. 2019:13:92.

Thaler R. Toward a positive theory of consumer choice. *J Econ Behav Organ*. 1980:1:39–60.

Tyler WJ, Lani SW, Hwang GM. Ultrasonic modulation of neural circuit activity. *Curr Opin Neurobiol*. 2018:50:222–231.

van den Bos W, van Dijk E, Westenberg M, Rombouts SARB, Crone EA. What motivates repayment? Neural correlates of reciprocity in the trust game. *Soc Cogn Affect Neurosci*. 2009:4:294–304.

Wagenaar WA, Keren G, Lichtenstein S. Islanders and hostages: deep and surface structures of decision problems. *Acta Psychol*. 1988:67: 175–189.

Wagner U, N'Diaye K, Ethofer T, Vuilleumier P. Guilt-specific processing in the prefrontal cortex. *Cereb Cortex*. 2011:21:2461–2470.

Wang Z, Goerlich KS, Ai H, Aleman A, Luo Y, Xu P. Connectome-based predictive Modeling of individual anxiety. *Cereb Cortex*. 2021:31:3006–3020.

Wang Y, Metoki A, Xia Y, Zang Y, He Y, Olson IR. A large-scale structural and functional connectome of social mentalizing. *NeuroImage*. 2021b:236:118115.

Whitfield-Gabrieli S, Ford JM. Default mode network activity and connectivity in psychopathology. *Annu Rev Clin Psychol*. 2012:8: 49–76.

Woodward ND, Cascio CJ. Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiatry*. 2015:72:743–744.

Yang W, Zhuang K, Liu P, Guo Y, Chen Q, Wei D, Qiu J. Memory suppression ability can be robustly predicted by the internetwork communication of frontoparietal control network. *Cereb Cortex*. 2021:31:3451–3461.

Yeshurun Y, Nguyen M, Hasson U. The default mode network: where the idiosyncratic self meets the shared social world. *Nat Rev Neurosci*. 2021:22:181–192.

Zuo X-N, Xing X-X. Test–retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci Biobehav Rev*. 2014:45:100–118.

Zuo XN, Di Martino A, Kelly C, Shehzad ZE, Gee DG, Klein DF, Castellanos FX, Biswal BB, Milham MP. The oscillating brain: complex and reliable. *NeuroImage*. 2010:49:1432–1445.

4. Study 2: Neurocomputational Mechanisms of Contextual Reciprocity

“How context shapes reciprocity: Insights from fMRI and computational modeling”

Fang, H., Liao, C., Fu, Z., Tian, S., Luo, Y., Xu, P., & Krueger, F. (2024). How context shapes reciprocity: Insights from fMRI and computational modeling. [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Title

How Context Shapes Reciprocity: Insights from fMRI and Computational Modeling

Abbreviated title

Contextual Effects on Reciprocity: An fMRI Study

Author names and affiliations

Huihua Fang [1,2], Yuejia Luo [3,4*], Pengfei Xu [4, 5*], Frank Krueger [2,6]

[1] Shenzhen Key Laboratory of Affective and Social Neuroscience, Magnetic Resonance Imaging Center, Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen, China

[2] Department of Psychology, University of Mannheim, Mannheim, Germany

[3] School of Psychology, Chengdu Medical College, Chengdu, China

[4] Faculty of Psychology, Beijing Normal University, Beijing 100875, China

[5] Center for Neuroimaging, Shenzhen Institute of Neuroscience, Shenzhen, China

[6] School of Systems Biology, George Mason University, Fairfax, VA, USA

***Corresponding author**

Yuejia Luo, Ph.D.

Email: luoyj@bnu.edu.cn

Pengfei Xu, Ph.D.

Email: pxu@bnu.edu.cn

Keywords: Reciprocity, frame, advantageous inequity aversion, Functional Magnetic Resonance Imaging (fMRI), computational modeling

Abstract

While reciprocity plays a fundamental role in cooperation, the neurocomputational mechanisms underlying its susceptibility to contextual influences in decision-making remain poorly understood. To examine the influence of contextual framing (gain vs. loss) on reciprocity, we utilized a combination of computational modeling and functional magnetic resonance imaging within the framework of a two-stage interactive binary trust game. Participants, acting as trustees, made reciprocity decisions (reciprocate/betray) and subsequently inferred their partner's prior choice (trust/status quo). Our behavioral findings demonstrate that the loss frame, as opposed to the gain frame, diminishes reciprocity by reducing advantageous inequity aversion. Neuroimaging results revealed distinct neural correlates of advantageous inequity aversion for other-oriented and self-oriented processes under overall reciprocity decision-making. During overall decision-making, right amygdala activity was negatively associated with advantageous inequity aversion in the gain frame, but not the loss frame. During other-oriented inference processes, no frame-related differences were found. During self-oriented evaluation processes, advantageous inequity aversion was positively associated with left anterior insula (lAI) activity in the gain frame but not the loss frame, and lAI activity in the loss frame was reduced compared to the gain frame. In summary, our findings indicate that within a loss-framed context, the mechanism underlying advantageous inequity aversion appears to be attenuated or disrupted during both overall reciprocity decision-making and self-oriented evaluation processes. Notably, advantageous inequity aversion neural correlates during other-oriented inference processes remain unaffected by this contextual framing. In conclusion, our study highlights the pivotal role of self-oriented evaluation in shaping context-dependent reciprocal behavior, pinpointing specific decision-making components and subprocesses susceptible to contextual framing.

Introduction

As a fundamental prosocial behavior, reciprocity has been instrumental in shaping human social interactions and cooperation throughout history (Nowak & Sigmund, 2005). According to social representation theory (SRT) (Abric, 1993; Hagen & Hammerstein, 2006; Wagenaar et al., 1988), reciprocity is determined not only by an individual's inherent disposition towards reciprocity but also by the social context where these decisions are made (Fang et al., 2022).

Manipulating the framing of the decision context, while maintaining the integrity of the payoff structure, allows for the identification of contextual influences on reciprocal behavior (Evans & van Beest, 2017; Fang et al., 2022). Eye-tracking data indicates that individuals framed within a loss context exhibit a heightened focus on their own outcomes, resulting in a greater tendency towards self-interested choices in the dictator game, compared to those framed within a gain context (Fiedler & Hillenbrand, 2020). Research consistently indicates a heightened propensity for reciprocal behavior when individuals are framed within a “give” context, in contrast to a "take" context, even when the underlying payoff structures remain equivalent (Bohnet & Meier, 2005; Fang et al., 2022; Keysar et al., 2008; Y. Zhang et al., 2023). Given that context can influence prosocial behaviors like reciprocity, identifying the underlying psychological mechanisms through which it operates is essential for comprehending the origins of this contextual effect.

Reciprocity decisions encompass multiple psychological components with complex cognitive processing. Reciprocity decisions often involve a conflict between self-interest (e.g., maximizing rewards by securing higher payoffs) and the social expectation to reciprocate kindness (e.g., repaying trust), creating a social dilemma. While reciprocation may necessitate foregoing immediate financial benefits, it can alleviate negative emotions such as guilt aversion (the discomfort of failing to meet reciprocal expectations) and advantageous inequity aversion (the discomfort associated with receiving more than others) (Fehr & Schmidt, 1999; Nihonsugi et al., 2015, 2021). While evidence for advantageous inequity aversion exists, it's inconsistent, appearing when individuals actively choose advantageous inequity but disappearing when they passively receive it (O. Li et al., 2018). In contrast, research suggests that people constantly engage in social comparisons, often striving to improve their relative position comparisons (Festinger, 1954; Fiske, 2011; Starmans et al., 2017). This implies advantageous inequity liking, rather than aversion, might be present in certain contexts (Boyce et al., 2010; Cox, 2013; Dohmen et al., 2011). Consequently, individuals may experience advantageous inequity aversion when considering betrayal, which usually leads to a more advantageous outcome and active harm, but favor advantageous distribution (advantageous inequity liking) when contemplating reciprocity. Furthermore, reciprocity decisions entail complex cognitive processes involving other-oriented inference (e.g., assessing trust bestowed by others or the need of others) (J. Li et al., 2009; van den Bos et al., 2009, 2011), self-oriented evaluation (e.g., gauging willingness to reciprocate or sacrifice) (Knoch et al., 2006; Rilling, 2011), and integrating these considerations to determine whether to reciprocate or betray (Crone, 2018; Fang et al., 2022).

By investigating the psychological components and cognitive processes inherent in reciprocity decisions, we can gain a deeper understanding of how and at which stage of the decision-making process context exerts its influence on the underlying psychological factors driving reciprocal behavior. Although research exploring how context influences psychological components like guilt aversion and advantageous inequity liking in social decisions remains limited, studies have demonstrated that framing outcomes as losses rather than gains in dictator games attenuate advantageous inequity aversion (Boun My et al., 2018). Moreover, a loss-framed context, in contrast to a gain-framed one, appears to diminish moral concerns, resulting in increased dishonest behavior (Schindler & Pfattheicher, 2017) and a decreased preference for equity (De Dreu, 1996).

Given the complex interplay of psychological factors and the potential influence of context on neural activity, an investigation into the neural correlates of reciprocity could yield valuable insights. For example, inequity aversion, encompassing both advantageous and disadvantageous inequity aversion, exhibits a negative correlation with right amygdala activation in prosocial individuals engaged in the dictator game (Haruno & Frith, 2010). However, inconsistent findings exist, with later studies reporting a positive association (Nihonsugi et al., 2015), suggesting that the link between amygdala activity and inequity aversion may be influenced by the interplay between advantageous and disadvantageous inequity aversion. Given the amygdala's established role in emotional processing (Phelps, 2006), and considering that reciprocity or the relinquishment of an advantageous position frequently involves sacrificing personal gains (Fang et al., 2022; Fehr & Schmidt, 1999), it is plausible that advantageous inequity aversion engages top-down cognitive control mechanisms to modulate emotional responses, and this modulation may manifest as reduced amygdala activity (Blair et al., 2007). Furthermore, the loss frame, by diminishing moral concern, may disrupt the association between advantageous inequity aversion and amygdala activity (De Dreu, 1996; Schindler & Pfattheicher, 2017).

A direct investigation into contextual effects on advantageous inequity aversion, comparing self-inflicted pain to other pain conditions, has revealed the involvement of brain regions such as the left anterior insula (lAI), right dorsolateral prefrontal cortex (rDLPFC), and dorsomedial prefrontal cortex (DMPFC) (Xiaoxue Gao et al., 2018). The AI is implicated in reciprocity, norm compliance, and the processing of subjective emotional experiences (Bellucci et al., 2018; Chang & Sanfey, 2011; Xiaoxue Gao et al., 2018). The DLPFC, critical for executive functions and cognitive control, contributes to the regulation of selfish impulses and emotions, while also influencing intention-based decision-making (Knoch et al., 2006; Miller & Cohen, 2001; Nihonsugi et al., 2015; Ruff et al., 2013; Zhu et al., 2014). As a key component of the mentalizing network, the DMPFC plays a vital role in inferring intentions during interpersonal interactions (Isoda & Noritake, 2013; Xiaoxue Gao et al., 2018).

While these studies suggest that context likely influences reciprocity through mechanisms involving advantageous inequity aversion and engages brain regions like the lAI, rDLPFC, DMPFC, and right amygdala, the specific neurocomputational mechanisms and cognitive processing stages underlying this modulation remain elusive. Our study aimed to elucidate the neurocomputational mechanisms by which context influences reciprocity, shedding light on the pertinent psychological components and processing stages involved, through the combined application of fMRI and computational modeling.

To investigate the psychological components and the interplay of self and other considerations in reciprocity decisions, we employed a two-stage interactive economic task based on a binary trust game (Evans & van Beest, 2017). Participants, assigned the role of trustee, were randomly allocated to either a gain- or loss-framed context and interacted with anonymous partners (trustors) in each trial. The task consisted of a decision stage, where participants chose to reciprocate or betray, and an inference stage, where they inferred their partner's prior trust decision. Additionally, participants completed the Interpersonal Reactivity Index (IRI) (Davis, 1980) to assess empathy, reflecting other-oriented tendencies, and the Machiavellianism (Mach-IV) scale (Christie & Geis, 1970) to gauge tendencies towards selfishness or its control, capturing self-oriented tendencies.

Drawing upon prior research, we proposed the following three hypotheses: At the *behavioral level*, we anticipated that individuals in the loss frame would demonstrate lower reciprocity rates compared to those in the gain frame, given the tendency towards increased self-interest in the context of potential losses (Fiedler & Hillenbrand, 2020).

At the *psychological level*, we hypothesized that individuals in the loss frame would exhibit reduced advantageous inequity aversion, which would subsequently impact reciprocity rates, in line with prior findings demonstrating decreased advantageous inequity aversion in loss frames

(Boun My et al., 2018). Furthermore, we posited positive correlations between other-oriented tendencies (as measured by the IRI) and advantageous inequity aversion, and negative correlations with self-oriented tendencies (as measured by the Mach-IV scale).

At the *neural level*, we predicted that the contextual modulation of advantageous inequity aversion during reciprocity decisions would engage specific brain regions, including the right amygdala, lAI, DMPFC, and rDLPFC, consistent with prior neuroimaging findings (Haruno & Frith, 2010; Xiaoxue Gao et al., 2018). In particular, we anticipated the engagement of the DMPFC and rDLPFC during other-oriented inference, considering their well-documented functions in mentalizing (Isoda & Noritake, 2013; Krueger, 2021) and intention-based decision-making (Nihonsugi et al., 2015), respectively. For self-oriented evaluation, contrasting the decision and inference stages, we anticipated the engagement of lAI due to its role in norm compliance (Chang & Sanfey, 2011; Xiaoxue Gao et al., 2018) and the right amygdala given its involvement in personal affect and emotional awareness (Craig, 2009).

Materials and Methods

Participants

Initially, 80 participants with no history of psychoactive medication use, mental disorders, or brain injuries were recruited and randomly assigned to either the gain or loss framed context group. However, the final sample comprised 65 participants (32 in the gain frame, 33 in the loss frame), following exclusions based on pre-defined criteria (**Tab. 1**): lack of belief in the authenticity of real-person interactions within the game (n=10), incorrect identification of their assigned role (n=4; 3 in the gain-framed condition, 1 in the loss-framed condition), and excessive head motion during data acquisition (n=1; criteria: mean frame-wise displacement exceeding 0.2 mm or more than 20% of the total number of volumes exceeding 2 mm maximum translation or 2 degree rotation). One participant met both the incorrect role identification and excessive head motion criteria. The study was conducted in accordance with the ethical guidelines outlined in the Declaration of Helsinki and received approval from the Ethics Committee at Shenzhen University in China. All participants provided written informed consent prior to participation. Participants received a fixed attendance fee of 60 yuan (approximately $8) and a variable monetary reward contingent on their performance in the game, ranging from 40 yuan to 80 yuan (approximately $6 to $11).

Questionnaires

Participants completed two self-report questionnaires through an online survey platform (https://www.wjx.cn). The IRI assessed trait empathy through four distinct 7-item subscales (Davis, 1980): perspective-taking (PT), measuring the ability to understand others' viewpoints; fantasy (FS), evaluating the capacity to immerse oneself in fictional characters' experiences; empathetic concern (EC), measuring sympathy for others' concerns; and personal distress (PD), assessing discomfort when witnessing others' suffering. Participants completed the 28-item questionnaire using a 5-point Likert scale ("Does not describe me well" to "Describes me very well"). Cronbach's α analysis indicated acceptable internal consistency in our sample for all subscales and the total score: PT ($\alpha = 0.682$), FS ($\alpha = 0.706$), EC ($\alpha = 0.658$), PD ($\alpha = 0.621$), and total ($\alpha = 0.805$). The Mach-IV questionnaire (Christie & Geis, 1970) assessed Machiavellianism, reflecting selfish tendencies. Participants responded to 20 items on a 7-point Likert scale, indicating their level of agreement or disagreement with statements about their attitudes and behaviors. The total score was calculated by summing the responses, with ten items reverse-scored. Cronbach's α analysis demonstrated acceptable internal consistency for the Mach-IV in our sample ($\alpha = 0.718$).

Experimental design

Before the experiment, participants attended an orientation session to learn the game's rules and practice their roles within the binary trust game structure, after which they were assigned to either the gain or loss frame (between-subject design) based on their group allocation (**Fig. 1 A**). In this game (programmed with E-Prime 3.0, https://pstnet.com/products/e-prime), Player 1 (P1, trustor) chooses either "status quo" or "trust," followed by Player 2's (P2, trustee) choice of "reciprocate" or "betray" (Evans & van Beest, 2017). If P1 chooses "status quo," both receive immediate payoffs (P1: a1, P2: b1), with P2's decision having no impact. If P1 chooses "trust," final payoffs depend on P2's choice: "reciprocate" (P1: a2, P2: b2) or "betray" (P1: a3, P2: b3). Participants assumed the role of P2 and were informed that their counterparts (P1s) had already made their choices, recorded in the system. Final payoffs would be calculated based on both players' decisions. Participants' responses were collected using a response pad with their right hand. The experiment involved no deception; all counterparts were real participants. Note that data for P1s will be published separately.

The binary trust game featured varying payoff structures across trials, maintaining consistent features: for P1, $a2 > a1 > a3$; for P2, $b3 > b2 > b1$; and in all trials, $a1 > b1$ and $a3 < b3$. The relationship between a2 and b2 varied, with $a2 > b2$ in 32 trials, $a2 = b2$ in 5 trials, and $a2 < b2$ in 43 trials. From a purely rational perspective, P1 should always choose to "trust," anticipating that P2 will "reciprocate" to secure the highest possible payoff (a2). Conversely, P2's rational choice is to consistently "betray" to achieve their highest possible payoff (b3).

The experimental task consisted of 80 trials, divided into two runs (total duration: 25.33 minutes) with short breaks (approximately 3 minutes) interspersed for rest. Each trial followed a structured timeline (**Fig. 1 B, C**): *Fixation stage*: An asterisk was displayed on the screen for an average of 3 seconds (range: 2-4 seconds). *Decision stage*: Participants had 6 seconds to choose between "reciprocate" or "betray." The chosen option was highlighted for 0.5 seconds, and an asterisk filled the remaining time if the decision was made quicker. *Fixation stage*: The asterisk reappeared for an average of 3 seconds (range: 2-4 seconds). *Inference stage*: Participants had 6 seconds to predict their partner's action ("trust" or "status quo"). Similar to the decision stage, the selected choice was highlighted, or an asterisk filled any remaining time. In both stages, if no response was made within the 6-second window, the highlight was omitted, and the asterisk filled the 0.5 seconds.

To mitigate potential spatial biases in decision-making, four versions of the game were created by systematically alternating the on-screen positions of the "status quo"/"trust" and "reciprocate"/"betray" options. These versions were counterbalanced across participants within each frame group. For analysis, all versions were standardized to a single representation (**Fig. 1 A**). The potential impact of these game version variations was assessed to ensure that observed effects were not attributable to differences in presentation.

--- Insert Figure 1 about here ---

The loss frame was adapted from the gain frame using a modified approach based on a previous study (Evans & van Beest, 2017). The sum of the "betray" option values (a3 + b3), rather than just b3, served as the reference point. Loss frame values were then calculated as the difference between each gain frame value and this reference. In the gain frame, participants started with 0 points and accumulated points, while in the loss frame, they began with 9,500 points and lost points based on their decisions. This design guaranteed equivalent outcomes if identical strategies were employed in both frames (Evans & van Beest, 2017). Participants' final earnings were determined by the accumulated points in the gain frame or the remaining points in the loss frame.

After the experiment, participants completed a post-experimental check, evaluating: (1) whether participants believed they had interacted with real partners, and (2) whether they correctly identified their role as P2 in the game. Participants who doubted the interaction's authenticity or misidentified their role were excluded from subsequent data analysis, ensuring that the data for analysis reflected genuine social interactions by including only those who were fully engaged and understood their role.

Computational modeling

A stage-wise model construction procedure was utilized to identify and quantify the latent psychological components influencing reciprocity behavior in our binary trust game (Gagne et al., 2020; Z. Wang et al., 2023; L. Zhang & Gläscher, 2020). The model was sequentially refined based on the performance of the prior best-fitting model. Model comparisons were conducted using Akaike Information Criterion (AIC) (H. Wang et al., 2024), with the lowest values indicating the best model. Parameter estimation was performed using the *optimize* function in SciPy module on Python (van Baar et al., 2019). For each participant, the entire parameter space was explored using 1000 random starting points. The parameters from the first occurrence of the best-fitting model were selected. Five plausible candidate models were tested in total.

Informed by previous studies (Nihonsugi et al., 2015; Xiao et al., 2022), the initial model included components of guilt aversion and inequity aversion (**M1**). It hypothesized that decisions are influenced by the trade-off between socio-emotional factors and potential gains. The utility function (U) was defined as shown in **Eq. 1**:

$$U = \begin{cases} b_3 - \beta_G \cdot (a_2 - a_3) - \beta_I \cdot (b_3 - a_3), & \textit{if betray} \\ b_2 - \beta_I \cdot |b_2 - a_2| & , \textit{if reciprocate} \end{cases} \qquad \text{M1 (1)}$$

Contrary to previous studies (Nihonsugi et al., 2015; Xiao et al., 2022), participants were not informed of their partner's reciprocity beliefs, mimicking real-world social interactions where such knowledge is often lacking. The term $(a_2 - a_3)$ represented the extent of anticipated guilt experienced upon choosing betrayal (where the partner receives a_3), taking into account the partner's assumed trust and expectation of reciprocity (anticipating a_2), with β_G ($0< \beta_G <10$) capturing the participant's subjective aversion to guilt. The terms $(b_3 - a_3)$ and $|b_2 - a_2|$ measure the inequity in choosing betrayal and reciprocity, with β_I ($0< \beta_I <10$) indicating the participant's subjective aversion to inequity. In our binary trust game design, the participant's payoff (b_3) always exceeds the partner's payoff (a3) in the betrayal option. However, for the reciprocity option, the participant's payoff (b_2) may be greater or less than the partner's payoff (a_2). Our baseline model quantifies inequity as the absolute difference between b_2 and a_2 in the reciprocity option, assuming participants give equal weight to both advantageous and disadvantageous inequity aversions, following previous studies (Nihonsugi et al., 2015; Xiao et al., 2022). To calculate the probability of choosing reciprocation in each trial, utility (U) of reciprocity and betrayal were entered into a SoftMax function with an inverse temperature parameter λ ($0 < \lambda < 10$) controlling the participant’s trade-off between randomness and determinism (**Eq. 2**):

$$P(reciprocate) = \frac{1}{1 + e^{-\lambda(U(reciprocate) - U(betray))}} \qquad (2)$$

In **M2** (**Eq. 3**), the Fehr–Schmidt inequity aversion model (Fehr & Schmidt, 1999) was used. This model distinguishes between two types of inequity aversion: advantageous inequity aversion and disadvantageous inequity aversion.

$$U = \begin{cases} b_3 - \beta_G \cdot (a_2 - a_3) - \beta_{Ad-IA} \cdot (b_3 - a_3) & , \textit{if betray} \\ b_2 - p \cdot \beta_{Ad-IA} \cdot (b_2 - a_2) - q \cdot \beta_{DisAd-IA}(a_2 - b_2) & , \textit{if reciprocate} \end{cases} \qquad \text{M2 (3)}$$

Here, β_{Ad-IA} and $\beta_{DisAd-IA}$ represents the participant's subjective aversion to advantageous and disadvantageous inequity, respectively. p and q are conditional indicators: when $b_2 > a_2$, p = 1, q = 0; conversely, when $b_2 > a_2$, p = 0, q = 1. Building on the improvement of model fitting, **M3** was developed by adding a reward parameter (**Eq. 4**) into **M2**, resulting in further improved model performance.

$$U = \begin{cases} \beta_R \cdot b_3 - \beta_G \cdot (a_2 - a_3) - \beta_{Ad-IA} \cdot (b_3 - a_3) & , \textit{if betray} \\ \beta_R \cdot b_2 - p \cdot \beta_{Ad-IA} \cdot (b_2 - a_2) - q \cdot \beta_{DisAd-IA}(a_2 - b_2) & , \textit{if reciprocate} \end{cases} \qquad \text{M3 (4)}$$

Based on **M3**, **M4** posits that participants dislike advantageous inequity in betrayal but favor it in reciprocity. Therefore, the inequity aversion term in the reciprocity option of **M3** were replaced with an advantageous inequity liking component (**Eq. 5**).

$$U = \begin{cases} \beta_R \cdot b_3 - \beta_G \cdot (a_2 - a_3) - \beta_{Ad-IA} \cdot (b_3 - a_3), & \textit{if betray} \\ \beta_R \cdot b_2 + \beta_{Ad-IL} \cdot (b_2 - a_2) & , \textit{if reciprocate} \end{cases} \qquad \text{M4 (5)}$$

In this model, the term $(b_2 - a_2)$ represents the perceived magnitude and direction of advantageous inequity (positive or negative), while β_{Ad-IL} signify the participant's subjective liking to the advantageous inequity in reciprocity. The model comparison showed **M4** outperforming **M3**. **M5** was established upon **M4**, assumes consistent trade-off between randomness and determinism in decision-making ($\lambda = 1$) (Nihonsugi et al., 2021). Ultimately, model comparison confirmed **M4** remained the winning model.

M4 was validated using model prediction and parameter recovery (van Baar et al., 2019; H. Wang et al., 2024). To assess the model's predictive ability, the correlation between predicted and observed reciprocity rates, and the prediction accuracy was calculated. For parameter recovery, **M4** was refitted to the simulated dataset and the correlation between the real and recovered parameters was assessed, thereby verifying the fitting precision of the model.

Statistical analysis

Statistical analyses of behavioral data were performed using Python (Version 3.9.9; https://www.python.org) and its pandas and scipy packages, with statistical significance set at p < 0.05 (two-tailed). Visualizations were created with the *seaborn* package.

The effects of frame on reciprocity rate and psychological components from the winning model (advantageous inequity aversion, guilt aversion, advantageous inequity liking, reward sensitivity) were analyzed using the Mann-Whitney U test. Spearman correlation was employed to investigate the relationship between psychological components and brain activity within each frame. These non-parametric tests were chosen for their robustness to potential outliers and non-normal distributions, commonly observed in psychological data.

To investigate the mediational role of advantageous inequity aversion between context and reciprocity rate, a mediation analysis was performed using the *GLM Mediation Model* within the *MedMod* module in jamovi (Version 2.3.18.0; www.jamoni.org). Confidence intervals were estimated through 1,000 bootstrap resamples using the bootstrap percentile method.

The internal consistency of the questionnaire was assessed using Cronbach's α (Taber, 2018). Mann-Whitney U tests were conducted to compare questionnaire results between the gain and loss frame groups. Bivariate Spearman correlations were used to examine the relationship between scale scores and psychological components.

Imaging data acquisition and preprocessing

MRI data were acquired using a Siemens Trio 3T scanner with a 64-channel phased array head coil. Functional MRI (fMRI) data were collected with a T2*-weighted multiband echo-planar imaging (EPI) protocol, with a multiband slice acceleration factor of 5, employing the following parameters: repetition time (TR) of 1,000 ms, echo time (TE) of 30 ms, 65 axial slices with a 2-mm spacing between slices, a flip angle of 90°, a field of view (FOV) of 192 × 192 mm^2, and a voxel size of 2 × 2 × 2 mm^3. There were two runs in total, each lasting 12.67 minutes and comprising 770 volumes. A high-resolution anatomical image was obtained after the experiment using a T1-weighted MPRAGE protocol, with TR/TE of 2300 ms/2.26 ms, 192 slices, a flip angle of 8°, an FOV of 256 × 256 mm^2, and a voxel size of 1 × 1 × 1 mm^3.

The anatomical and functional MRI data were preprocessed using *fMRIPrep* 21.0.2 (Esteban et al., 2019), which is built on Nipype 1.6.1 (Gorgolewski et al., 2011). For more pipeline details, see the section corresponding to workflows in *fMRIPrep*'s documentation (https://fmriprep.org). Image preprocessing was conducted, including slice-timing correction, realignment, co-registration, spatial normalization to the Montreal Neurological Institute (MNI) space with a spatial resolution of 2 × 2 × 2 mm^3. The preprocessed images (without smoothing) were then smoothed with a Gaussian kernel of 6 mm full width at half maximum (FWHM).

Imaging data analysis

General linear model (GLM) analysis was performed using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/), running within the MATLAB environment. Given that advantageous inequity aversion was the only component in the model exhibiting a significant framing effect, its influence on the decision stage, inference stage, and the contrast between these stages was assessed. In addition, the beta values of identified brain regions related to advantageous inequity aversion were further compared between the gain and loss frames. To ensure the tested hypothesis had maximum variance and minimize potential multicollinearity issues, advantageous inequity aversion was examined using a separate model (Gao et al., 2024).

The onsets of the decision and inference stages were modeled as conditions, with the objective advantageous inequity aversion value for all trials included as a parametric modulator. Six head movement parameters were also incorporated as covariates. The defined regressors were convolved with the canonical hemodynamic response function. Contrast images for the decision stage, inference stage, and their contrast—reflecting the modulator effect of advantageous inequity aversion from the first-level analysis—were entered into the second-level analysis. Here, subjective advantageous inequity aversion served as a regressor, while the game version was included as a covariate to identify the neural correlates of advantageous inequity aversion. Subsequently, beta values were extracted and compared between the gain and loss frames.

The significance level was set to $p < 0.001$ uncorrected at the voxel level, with an FWE-corrected $p < 0.05$ at the cluster level. Small volume correction (SVC) analyses were also performed, focusing on regions of interest (ROIs) previously implicated in advantageous inequity aversion, including lAI, DMPFC, rDLPFC, and right amygdala (Haruno & Frith, 2010; Xiaoxue Gao et al., 2018). For SVC, a threshold of $p < 0.001$ at the voxel level and $p < 0.05$ FWE-corrected at the cluster or peak level was applied. This approach aimed to enhance comparability with prior research while maintaining statistical rigor (H. Wang et al., 2024). Anatomical ROI masks from the AAL atlas in the SPM WFU Pickatlas toolbox (www.ansir.wfubmc.edu, version 3.0) were used for SVC analyses (Tzourio-Mazoyer et al., 2002). These included masks for the lAI, a mask combining Frontal_Sup_Medial_L and Frontal_Sup_Medial_R (covering DMPFC), a mask for Frontal_Mid_R (covering DLPFC), and the right amygdala.

Results

Behavioral and modeling results

Investigating the impact of frame on reciprocity rate, a trend towards lower reciprocity rates in individuals under the loss frame compared to the gain frame was observed ($U = 393$, $p = 0.078$) (**Fig. 2A**).

--- Insert Figure 2 about here ---

To uncover the psychological underpinnings of reciprocity, five plausible computational models were constructed, with model comparison revealing that the proposed model 4 (**M4**)—encompassing components such as advantageous inequity aversion, advantageous inequity liking, guilt aversion, and reward sensitivity—provided the best fit to our data (**Fig. S1, also see supplementary for model validation**).

Examining the impact of frame on the psychological components of reciprocity, individuals exhibited significantly lower advantageous inequity aversion under the loss frame compared to the gain frame ($U = 371$, $p = 0.039$) (**Fig. 2B),** while no significant frame effect was observed on guilt aversion ($U = 450$, $p = 0.281$), advantageous inequity liking ($U = 600.5$, $p = 0.338$), and reward sensitivity ($U = 441.5$, $p = 0.259$). The higher the advantageous inequity aversion was, the higher the reciprocity rate under the gain ($r_s = 0.61$, $p < 0.001$) and loss ($r_s = 0.72$, $p < 0.001$) frame (**Fig. 2C)**. The game version did not affect either the reciprocity decision or the psychological components, *ps* > 0.05.

Mediation analysis revealed a significant complete indirect effect of frame on reciprocity rate via advantageous inequity aversion (a*b = -0.17, SE = 0.05, CI = [-0.196, -0.002]), indicating that the loss frame, compared to the gain frame, decreased reciprocity rate through its effect on advantageous inequity aversion (**Fig. 2D**).

Questionnaire results

Results from the psychometric measures showed no significant differences between the gain and loss frame groups on either the Mach-IV or IRI scales (**Tab. 1)**. Exploratory analyses revealed only a trend-level negative correlation between advantageous inequity aversion and both Machiavellianism ($r_s = -0.24$, $p = 0.058$) and PD ($r_s = -0.22$, $p = 0.073$), suggesting that individuals higher in Machiavellianism or PD tend to exhibit lower advantageous inequity aversion.

Neuroimaging results

To pinpoint brain regions associated with advantageous inequity aversion in overall reciprocity decision-making, parametric modulation analysis was applied to the decision stage. A negative association with advantageous inequity aversion in the right amygdala was observed (peak at [30, 0, -20], k = 12, voxel level $P_{\text{uncorrected}} < 0.001$; cluster level $P_{\text{FWE}} < 0.05$, SVC corrected) (**Fig. 3**), indicating that higher levels of advantageous inequity aversion were linked to decreased activity in the right amygdala. In particular, advantageous inequity aversion showed a significant negative correlation with right amygdala activity under the gain (r_s = -0.48, p = 0.006) but not under the loss (r_s = -0.29, p = 0.102) frame. No significant difference in right amygdala activity was observed between the gain and loss frame (U = 507.0, p = 0.788). The observation that the right amygdala is associated with advantageous inequity aversion exclusively in the gain frame, but not in the loss frame, suggests its involvement in the processing of advantageous inequity aversion is contingent upon the specific context.

--- Insert Figure 3 about here ---

To identify brain regions involved in advantageous inequity aversion in inferencing other's actions, parametric modulation analysis was applied to the inference stage. Brain activities positively related to advantageous inequity aversion were observed in DMPFC ([18, 34, 58], k = 631, cluster level P_{FWE} < 0.05), rDLPFC ([46, 14, 42], k = 368, cluster level P_{FWE} < 0.05) and lSMG ([-52, -48, 58], k = 409, cluster level P_{FWE} < 0.05) (**Fig. 4 A**). Specifically, advantageous inequity aversion showed a significant positive correlation with DMPFC under both gain and loss frame (gain: r_s = 0.52, p = 0.002; loss: r_s = 0.57, p < 0.001), rDLPFC (gain: r_s = 0.47, p = 0.006; loss: r_s = 0.67, p < 0.001), and lSMG (gain: r_s = 0.59, p < 0.001; loss: r_s = 0.53, p = 0.001) (**Fig. 4 B**). No significant differences in DMPFC (U = 507.0, p = 0.788), rDLPFC (U = 498.0, p = 0.699), and lSMG (U = 509.0, p = 0.808) activities were observed between gain and loss frame (**Fig. 4 C**). The presence of the association of advantageous inequity aversion in these regions but the absence of the framing effect suggests their involvement in processing advantageous inequity aversion regardless of context.

--- Insert Figure 4 about here ---

Given that reciprocity decisions entail both self- and other-oriented considerations, parametric modulation analysis was also applied to the contrast between decision and inference stages to identify brain regions associated with advantageous inequity aversion in self-oriented evaluation. lAI activities ([-44, 8, -8], k = 8, voxel level $P_{uncorrected}$ < 0.001; cluster level P_{FWE} < 0.05, SVC corrected) positively, but rDLPFC activity (peak at [46, 14, 44], k = 50, voxel level $P_{uncorrected}$ < 0.001; cluster level P_{FWE} < 0.05, SVC corrected) negatively**,** associated with advantageous inequity aversion was observed (**Fig. 5 A**). Specifically, advantageous inequity aversion showed a significant negative correlation with lAI activity under the gain (r_s = 0.42, p = 0.017) but not under the loss (r_s = 0.17, p = 0.338) frame. Advantageous inequity aversion showed a significant negative correlation with rDLPFC under both gain and loss frame (gain: r_s = -0.43, p = 0.013; loss: r_s = -0.54, p = 0.001) (**Fig. 5 B**). The lAI activity was significantly lower under the loss compared to the gain (U = 685.0, p = 0.040) frame, whereas no significant difference in rDLPFC activity was observed between frames (U = 459.0, p = 0.369). The observation that the lAI is associated with advantageous inequity aversion exclusively in the gain frame, but not in the loss frame, suggests its involvement in the processing of advantageous inequity aversion is contingent upon the specific context. The association of advantageous inequity aversion with rDLPFC in both frames, suggests their role in processing advantageous inequity aversion is context independent.

--- Insert Figure 5 about here ---

Discussion

Our study employed computational modeling and fMRI within an economic binary trust game, incorporating both other-oriented and self-oriented considerations, to examine the neurocomputational mechanisms driving the contextual influence (gain vs. loss frame) of reciprocity. At the behavioral level, our findings indicate that the loss frame, in contrast to the gain frame, showed a decrease trend in reciprocity rate. At the psychological level, this reduction was mediated by a diminished impact of the psychological construct of advantageous inequity aversion on reciprocity behavior. At the neural level, our findings revealed a context-dependent role for the right amygdala in modulating advantageous inequity aversion during overall decision-making, with activity inversely correlated specifically within the gain frame. While brain regions associated with advantageous inequity aversion during other-oriented inference exhibited no significant contextual modulation, the lAI demonstrated heightened activity in the gain compared to the loss frame and a unique association with advantageous inequity aversion exclusively within the gain context during self-oriented evaluation. Overall, our study elucidates the neurocomputational mechanisms and decision-making processes underlying reciprocity, emphasizing the pivotal role of self-oriented evaluation in mediating the contextual modulation of reciprocal behavior.

Our behavioral results aligned with our first hypothesis, revealing a trend towards decreased reciprocity within the loss compared to the gain frame. While the payoff structures may be equivalent, prospect theory posits that individuals demonstrate a greater aversion to losses compared to their desire for equivalent gains (Kahneman & Tversky, 1979). Reciprocity, as opposed to betrayal, necessitates a personal sacrifice to enhance the partner's net payoff. The loss frame may amplify the perceived burden of reciprocation, thus diminishing other-regarding tendencies and reducing the likelihood of reciprocal behavior. Our finding corroborate previous evidence (Bohnet & Meier, 2005; Fang et al., 2022; Fiedler & Hillenbrand, 2020; Keysar et al., 2008; Y. Zhang et al., 2023), demonstrating that negatively framed decisions, particularly those emphasizing potential losses, are frequently associated with decreased altruistic tendencies and an increased propensity for self-protective behaviors. To further investigate the computational mechanism by which context affects reciprocity, we employed computational modeling to unpack the decision-making components. Our analysis revealed that the best-fitting model included psychological components such as advantageous inequity aversion, guilt aversion, advantageous inequity liking, and reward sensitivity.

In line with our second hypothesis, we observed that the contextual effect was specific to advantageous inequity aversion, supporting prior research (Boun My et al., 2018; Xiaoxue Gao et al., 2018). Advantageous inequity aversion was significantly reduced under the loss frame compared to the gain frame, while the remaining three psychological components (guilt aversion, advantageous inequity liking, and reward sensitivity) remained unaffected by the context.

Advantageous inequity aversion, a moral concern for others' welfare, manifests as discomfort when one holds an advantaged position relative to other and has been shown to influence prosocial decision-making (Starmans et al., 2017). Our findings are consistent with research demonstrating that loss contexts can diminish moral concerns, leading to increased dishonesty (Schindler & Pfattheicher, 2017) and reduced equity preferences (De Dreu, 1996). The observed contextual impact on advantageous inequity aversion, but not on reward sensitivity, suggests that the difference in reciprocity rates is not primarily driven by self-interest. Instead, our results indicate that loss frames decrease other-regarding concerns rather than amplify self-protection motives compared to gain frames.

In partial alignment with our third hypothesis, our analysis of the reciprocity decision stage revealed a negative association between right amygdala activity and advantageous inequity aversion exclusively within the gain frame, with no such association observed in the loss frame. Prior studies have reported inconsistent findings regarding the relationship between amygdala activity and inequity aversion, demonstrating both positive and negative associations (Haruno & Frith, 2010; Nihonsugi et al., 2015). This discrepancy may stem from the conflation of advantageous inequity aversion and disadvantageous inequity aversion within the broader concept of inequity aversion. Our study addresses this ambiguity by focusing exclusively on advantageous inequity aversion, uncovering a negative association with amygdala activity. Since concern for others often necessitates personal sacrifice and may engage cognitive control mechanisms. The amygdala plays a well-established role in processing emotional responses, especially in relation to negative emotions and selfish impulses (Phelps, 2006; Scheggia et al., 2022). Our analysis of other-oriented inference revealed the involvement of rDLPFC, a region implicated in both cognitive control (Fang et al., 2022; Fehr & Schmidt, 1999) and emotional down-regulation (Blair et al., 2007), offers a potential explanation for the negative correlation between advantageous inequity aversion and amygdala activity.

Our questionnaire data also revealed a trend towards reduced selfishness and personal distress in individuals exhibiting heightened advantageous inequity aversion, lending further support to the proposed down-regulation mechanism. While this observation diverges from our initial hypothesis, it suggests a potential association between heightened advantageous inequity aversion and enhanced cognitive control especially in stage of inferencing other's need. This enhanced cognitive control could facilitate the effective regulation of self-interest and emotional responses, manifesting as the observed lower selfishness and personal distress. Importantly, the negative correlation between amygdala activity and advantageous inequity aversion was exclusive to the gain frame, indicating a potential disruption of this down-regulation process within the loss frame. This disruption may be attributed to the diminished moral concern often associated with loss-framed decision-making contexts (De Dreu, 1996; Leib et al., 2019; Schindler & Pfattheicher, 2017).

Reciprocity decision-making is a complex social cognitive process involving the integration of other-oriented inference and self-oriented evaluation. To elucidate the contextual effects on these distinct components, we implemented a two-stage design. Participants first made their own reciprocity decision, then immediately inferred their partner's trust decision. This approach enabled us to investigate both other-oriented inference and, through contrast analysis, to indirectly examine self-oriented evaluation.

Our results partially confirmed our third hypothesis, demonstrating distinct neural substrates of advantageous inequity aversion during different processes of the reciprocity decision-making. During the other-oriented inference processes, activity in the DMPFC, rDLPFC, and lSMG was positively associated with advantageous inequity aversion. In contrast, during the self-oriented evaluation processes, we observed a positive modulation of advantageous inequity aversion in lAI and a negative modulation in rDLPFC.

The neural regions identified in our study are well-established within the domains of social cognition and decision-making. The DMPFC, a key component of the mentalizing network, is integral for inferring intentions during social interactions (Isoda & Noritake, 2013; Xiaoxue Gao et al., 2018). The DLPFC, known for its critical role in executive functions and cognitive control (Miller & Cohen, 2001), contributes to the suppression of selfish motives, the regulation of negative emotions (Knoch et al., 2006; Ruff et al., 2013; Zhu et al., 2014), and plays a causal role in intention-based decision-making (Nihonsugi et al., 2015). The SMG, while less frequently implicated in social decision-making research, has also been associated with cognitive empathy (Kogler et al., 2020). The AI, a region commonly observed in reciprocity studies, is associated with norm compliance and the processing of subjective feelings and emotional awareness (Chang & Sanfey, 2011; Craig, 2009; Xiaoxue Gao et al., 2018). Our findings contribute to the existing literature on the neural basis of advantageous inequity aversion (Xiaoxue Gao et al., 2018), by delineating the distinct roles of these regions under other-oriented and self-oriented considerations during the processing of advantageous inequity aversion.

Consistent with prior findings (Xiaoxue Gao et al., 2018), we observed a context-dependent modulation of advantageous inequity aversion specifically within the lAI during self-oriented evaluation processes. Notably, lAI activity was lower and its association with advantageous inequity aversion was absent in the loss frame compared to the gain frame. This attenuation suggests decreased subjective emotional awareness, potentially reflecting diminished moral concern in the loss frame. No contextual effects on the relationship between advantageous inequity aversion and DMPFC or rDLPFC activity were observed during either other-oriented inference or self-oriented evaluation. This divergence from previous findings might be attributable to variations in experimental paradigms (Xiaoxue Gao et al., 2018). Prior research employed stimuli with immediate biological relevance (e.g., pain), whereas our economic gain/loss manipulation might exhibit reduced sensitivity in detecting all brain regions susceptible to contextual modulation. Despite methodological variations between our study and prior research, the implication of the DMPFC, rDLPFC, and lAI underscores their fundamental role in the processing of advantageous inequity aversion. Our findings highlight the pivotal role of self-oriented considerations, particularly the willingness to sacrifice personal gain for the benefit of others, in mediating the contextual modulation of advantageous inequity aversion. The specific involvement of the lAI in this process underscores its significance in self-oriented evaluation, as opposed to other-oriented inference, during the complex decision-making processes involved in reciprocal social interactions.

The current study possesses certain limitations that warrant consideration. Firstly, due to the inherent interconnectedness of other- and self-oriented considerations in reciprocity, we employed an indirect approach to examine the self-oriented evaluation by contrasting the decision stage with the other-oriented inference stage. Future research endeavors should aim to develop and implement more direct measures of self-oriented decision-making processes. Secondly, it is plausible that the contextual effects under investigation may emerge during the integration of other- and self-oriented considerations, rather than solely within each individual process. Consequently, future studies should explicitly explore whether the contextual effect of advantageous inequity aversion manifests during this integrative stage. Finally, although our study has shed light on the neurocomputational mechanisms underlying the framing effect in reciprocity, a more comprehensive approach incorporating additional methodologies, such as eye-tracking, could enhance the identification and validation of differences in advantageous inequity aversion under gain and loss contexts. For instance, analyzing eye-movement trajectories across different payoff distributions during reciprocity decisions could reveal important distinctions (Molter et al., 2022). Addressing these limitations in future research will be crucial to advance our understanding of the framing effect on reciprocity and its underlying mechanisms.

Despite these limitations, our study has shed light on the contextual influences on reciprocity, demonstrating that context modulates reciprocity through its impact on advantageous inequity aversion. We identified distinct neural correlates of this phenomenon, observing context-dependent modulation of right amygdala activity during overall reciprocity decision-making, and lAI activity specifically during the self-oriented evaluation processes. In conclusion, our findings offer valuable insights into the neurocomputational mechanisms underlying the contextual effects on reciprocity. These insights could potentially inform policymakers in their efforts to foster social harmony through the implementation of contextually appropriate frameworks in various social settings.

Author Contributions

Huihua Fang, Pengfei Xu, and Frank Krueger contributed to the study conception and design. Huihua Fang contributed to material preparation and data collection. Huihua Fang analyzed the data and visualized the results. Huihua Fang and Frank Krueger contributed to the original draft of the paper. Pengfei Xu and Yuejia Luo provided the funding support. All authors contributed to and approved the final manuscript.

Acknowledgments

We thank Chong Liao, Zhao Fu, and Shuang Tian for their assistance in data collection.

This study was supported by the National Natural Science Foundation of China (31920103009 to YL, 31871137 to PX), the Major Project of National Social Science Foundation (20&ZD153 to YL), Young Elite Scientists Sponsorship Program by China Association for Science and Technology (YESS20180158 to PX), Shenzhen-Hong Kong Institute of Brain Science Shenzhen Fundamental Research Institutions (2022SHIBS0003 to YL), and Shenzhen Science and Technology Research Funding Program (JCYJ20180507183500566 to PX). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement

Data and material related to this paper are available on request from the corresponding author (Pengfei Xu).

Reference

Abric, J.-C. (1993). Central system, peripheral system: Their functions and roles in the dynamics of social representations. *Papers on Social Representations*, *2*, 75–78.

Bellucci, G., Feng, C., Camilleri, J., Eickhoff, S. B., & Krueger, F. (2018). The role of the anterior insula in social norm compliance and enforcement: Evidence from coordinate-based and functional connectivity meta-analyses. *Neuroscience and Biobehavioral Reviews*, *92*, 378–389. https://doi.org/10.1016/j.neubiorev.2018.06.024

Blair, K. S., Smith, B. W., Mitchell, D. G. V., Morton, J., Vythilingam, M., Pessoa, L., Fridberg, D., Zametkin, A., Nelson, E. E., Drevets, W. C., Pine, D. S., Martin, A., & Blair, R. J. R. (2007). Modulation of emotion by cognition and cognition by emotion. *NeuroImage*, *35*(1), 430–440. https://doi.org/10.1016/j.neuroimage.2006.11.048

Bohnet, I., & Meier, S. (2005). Deciding to distrust. *SSRN Electronic Journal*, *05*. https://doi.org/10.2139/ssrn.839225

Boun My, K., Lampach, N., Lefebvre, M., & Magnani, J. (2018). Effects of gain-loss frames on advantageous inequality aversion. *Journal of the Economic Science Association*, *4*(2), 99–109. https://doi.org/10.1007/s40881-018-0057-2

Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science*, *21*(4), 471–475.

Chang, L. J., & Sanfey, A. (2011). Great expectations: Neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience*, *8 3*, 277–284. https://doi.org/10.1093/scan/nsr094

Christie, R., & Geis, F. L. (1970). *Studies in machiavellianism*. New York, NY: Academic Press.

Cox, C. A. (2013). Inequity aversion and advantage seeking with asymmetric competition. *Journal of Economic Behavior & Organization*, *86*, 121–136. https://doi.org/10.1016/j.jebo.2012.12.020

Craig, A. D. (2009). How do you feel—Now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70.

Crone, K. (2018). Understanding others, reciprocity, and self-consciousness. *Phenomenology and the Cognitive Sciences*, *17*(2), 267–278. https://doi.org/10.1007/s11097-016-9498-3

Davis, M. H. (1980). *A multidimensional approach to individual differences in empathy*.

De Dreu, C. K. W. (1996). Gain–loss-frame in outcome-interdependence: Does it influence equality or equity considerations? *European Journal of Social Psychology*, *26*(2), 315–324.

https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-0992%28199603%2926%3A2%3C315%3A%3AAID-EJSP759%3E3.0.CO%3B2-Z

Dohmen, T., Falk, A., Fliessbach, K., Sunde, U., & Weber, B. (2011). Relative versus absolute income, joy of winning, and gender: Brain imaging evidence. *Journal of Public Economics*, *95*(3), 279–285. https://doi.org/10.1016/j.jpubeco.2010.11.025

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Evans, A. M., & van Beest, I. (2017). Gain-loss framing effects in dilemmas of trust and reciprocity. *Journal of Experimental Social Psychology*, *73*(July), 151–163. https://doi.org/10.1016/j.jesp.2017.06.012

Fang, H., Liao, C., Fu, Z., Tian, S., Luo, Y., Xu, P., & Krueger, F. (2022). Connectome-based individualized prediction of reciprocity propensity and sensitivity to framing: A resting-state functional magnetic resonance imaging study. *Cerebral Cortex*, *33*(6), 3193–3206. https://doi.org/10.1093/cercor/bhac269

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Fiedler, S., & Hillenbrand, A. (2020). Gain-loss framing in interdependent choice. *Games and Economic Behavior*, *121*, 232–251. https://doi.org/10.1016/j.geb.2020.02.008

Fiske, S. T. (2011). *Envy up, scorn down: How status divides us*. Russell Sage Foundation.

Gagne, C., Zika, O., Dayan, P., & Bishop, S. J. (2020). Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife*, *9*, e61387. https://doi.org/10.7554/eLife.61387

Gao, X., Jolly, E., Yu, H., Liu, H., Zhou, X., & Chang, L. J. (2024). The psychological, computational, and neural foundations of indebtedness. *Nature Communications*, *15*(1), Article 1. https://doi.org/10.1038/s41467-023-44286-9

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00013

Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology*, *69*(3), 339–348. https://doi.org/10.1016/j.tpb.2005.09.005

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*(2), 160–161. https://doi.org/10.1038/nn.2468

Isoda, M., & Noritake, A. (2013). What makes the dorsomedial frontal cortex active during reading the mental states of others? *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00232

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Keysar, B., Converse, B. A., Wang, J., & Epley, N. (2008). Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychological Science*, *19*(12), 1280–1286. https://doi.org/10.1111/j.1467-9280.2008.02223.x

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. https://doi.org/10.1126/science.1129156

Kogler, L., Müller, V. I., Werminghausen, E., Eickhoff, S. B., & Derntl, B. (2020). Do I feel or do I know? Neuroimaging meta-analyses on the multiple facets of empathy. *Cortex*, *129*, 341–355. https://doi.org/10.1016/j.cortex.2020.04.031

Krueger, F. (2021). *The Neurobiology of Trust*. Cambridge University Press.

Leib, M., Pittarello, A., Gordon-Hecker, T., Shalvi, S., & Roskes, M. (2019). Loss framing increases self-serving mistakes (but does not alter attention). *Journal of Experimental Social Psychology*, *85*, 103880. https://doi.org/10.1016/j.jesp.2019.103880

Li, J., Xiao, E., Houser, D., & Montague, P. R. (2009). Neural responses to sanction threats in two-party economic exchange. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(39), 16835–16840. https://doi.org/10.1073/pnas.0908855106

Li, O., Xu, F., & Wang, L. (2018). Advantageous inequity aversion does not always exist: The role of determining allocations modulates preferences for advantageous inequity. *Frontiers in Psychology*, *9*, 749. https://doi.org/10.3389/fpsyg.2018.00749

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(Volume 24, 2001), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Molter, F., Thomas, A. W., Huettel, S. A., Heekeren, H. R., & Mohr, P. N. C. (2022). Gaze-dependent evidence accumulation predicts multi-alternative risky choice behaviour. *PLOS Computational Biology*, *18*(7), e1010283. https://doi.org/10.1371/journal.pcbi.1010283

Nihonsugi, T., Ihara, A., & Haruno, M. (2015). Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. *Journal of Neuroscience*, *35*(8), 3412–3419. https://doi.org/10.1523/JNEUROSCI.3885-14.2015

Nihonsugi, T., Numano, S., & Haruno, M. (2021). Functional connectivity basis and underlying cognitive mechanisms for gender differences in guilt aversion. *eNeuro*, *8*(6). https://doi.org/10.1523/ENEURO.0226-21.2021

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), Article 7063. https://doi.org/10.1038/nature04131

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annu. Rev. Psychol.*, *57*(1), 27–53.

Rilling, J. K. (2011). The neurobiology of cooperation and altruism. In *Origins of altruism and cooperation* (pp. 295–306). Springer.

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, *342*(6157), 482–484. https://doi.org/10.1126/science.1241399

Scheggia, D., La Greca, F., Maltese, F., Chiacchierini, G., Italia, M., Molent, C., Bernardi, F., Coccia, G., Carrano, N., Zianni, E., Gardoni, F., Di Luca, M., & Papaleo, F. (2022). Reciprocal cortico-amygdala connections regulate prosocial and selfish choices in mice. *Nature Neuroscience*, *25*(11), 1505–1518. https://doi.org/10.1038/s41593-022-01179-2

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, *69*, 172–177. https://doi.org/10.1016/j.jesp.2016.09.009

Starmans, C., Sheskin, M., & Bloom, P. (2017). Why people prefer unequal societies. *Nature Human Behaviour*, *1*(4), 1–7. https://doi.org/10.1038/s41562-017-0082

Taber, K. S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

van Baar, J. M., Chang, L. J., & Sanfey, A. G. (2019). The computational and neural substrates of moral strategies in social decision-making. *Nature Communications*, *10*(1), 1483. https://doi.org/10.1038/s41467-019-09161-6

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. R. B., & Crone, E. A. (2009). What motivates repayment? Neural correlates of reciprocity in the trust game. *Social Cognitive and Affective Neuroscience*, *4*(3), 294–304. https://doi.org/10.1093/scan/nsp009

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. R. B., & Crone, E. A. (2011). Changing brains, changing perspectives: The neurocognitive development of reciprocity. *Psychological Science*, *22*(1), 60–70. https://doi.org/10.1177/0956797610391102

Wagenaar, W. A., Keren, G., & Lichtenstein, S. (1988). Islanders and hostages: Deep and surface structures of decision problems. *Acta Psychologica*, *67*(2), 175–189. https://doi.org/10.1016/0001-6918(88)90012-1

Wang, H., Wu, X., Xu, J., Zhu, R., Zhang, S., Xu, Z., Mai, X., Qin, S., & Liu, C. (2024). Acute stress during witnessing injustice shifts third-party interventions from punishing the perpetrator to helping the victim. *PLOS Biology*, *22*(5), e3002195. https://doi.org/10.1371/journal.pbio.3002195

Wang, Z., Nan, T., Goerlich, K. S., Li, Y., Aleman, A., Luo, Y., & Xu, P. (2023). Neurocomputational mechanisms underlying fear-biased adaptation learning in changing environments. *PLOS Biology*, *21*(5), e3001724. https://doi.org/10.1371/journal.pbio.3001724

Xiao, F., Zhao, J., Fan, L., Ji, X., Fang, S., Zhang, P., Kong, X., Liu, Q., Yu, H., Zhou, X., Gao, X., & Wang, X. (2022). Understanding guilt-related interpersonal dysfunction in obsessive-compulsive personality disorder through computational modeling of two social interaction tasks. *Psychological Medicine*, *53*(12), 5569--5581. https://doi.org/10.1017/S003329172200277X

Xiaoxue Gao, Gao, X., Hongbo Yu, Yu, H., Ignacio Sáez, Saez, I., Philip R. Blue, Blue, P. R., Lusha Zhu, Zhu, L., Ming Hsu, Hsu, M., Xiaolin Zhou, & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity aversion. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(33), 201802523. https://doi.org/10.1073/pnas.1802523115

Zhang, L., & Gläscher, J. (2020). A brain network supporting social influences in human decision-making. *Science Advances*, *6*(34), eabb4159. https://doi.org/10.1126/sciadv.abb4159

Zhang, Y., Zhang, Y., Wu, Y., & Krueger, F. (2023). Default matters in trust and reciprocity. *Games*, *14*(1), 8. https://doi.org/10.3390/g14010008

Zhu, L., Jenkins, A. C., Set, E., Scabini, D., Knight, R. T., Chiu, P. H., King-Casas, B., & Hsu, M. (2014). Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. *Nature Neuroscience*, *17*(10), 1319–1321. https://doi.org/10.1038/nn.3798

Figure Legends

Figure 1. Experimental design. (A) Payoff structure: Player 1 (P1, trustor) chooses between "status quo" and "trust". If P1 selects "status quo", both players receive immediate payoffs (P1: a1, P2: b1). If P1 chooses "trust", Player 2 (P2, trustee) then decides to either "reciprocate" (P1: a2, P2: b2) or "betray" (P1: a3, P2: b3). Participants completed a binary trust game in the role of trustee (P2), assigned to either the gain **(B)** or loss **(C)** frame. Each trial in both frames consisted of decision and inference stages, separated by fixation periods. First, a fixation asterisk was presented. In the decision stage, participants chose to either "reciprocate" or "betray". In the gain frame, this meant selecting either the second column (P1 gains 40 and P2 gains 37) or the third column (P1 gains 14 and P2 gains 45). In the loss frame, the choice was between the second column (P1 loses 19 and P2 loses 22) or the third column (P1 loses 45 and P2 loses 14). An arrow highlighted the selected choice immediately after it was made. Following the decision stage, another fixation asterisk appeared. In the inference stage, participants inferred whether their partner (P1) had chosen "status quo" or "trust". The "status quo" option corresponded to P1 selecting the first column (in the gain frame: P1 gains 22 and P2 gains 9; in the loss frame: P1 loses 37 and P2 loses 50), while "trust" meant P1 had deferred the decision to P2. An arrow highlighted the selected choice immediately after it was made. The gain frame game began with 0 points, while the loss frame game started with 9,500 points. Identical strategies in both frames would result in equivalent final outcomes.

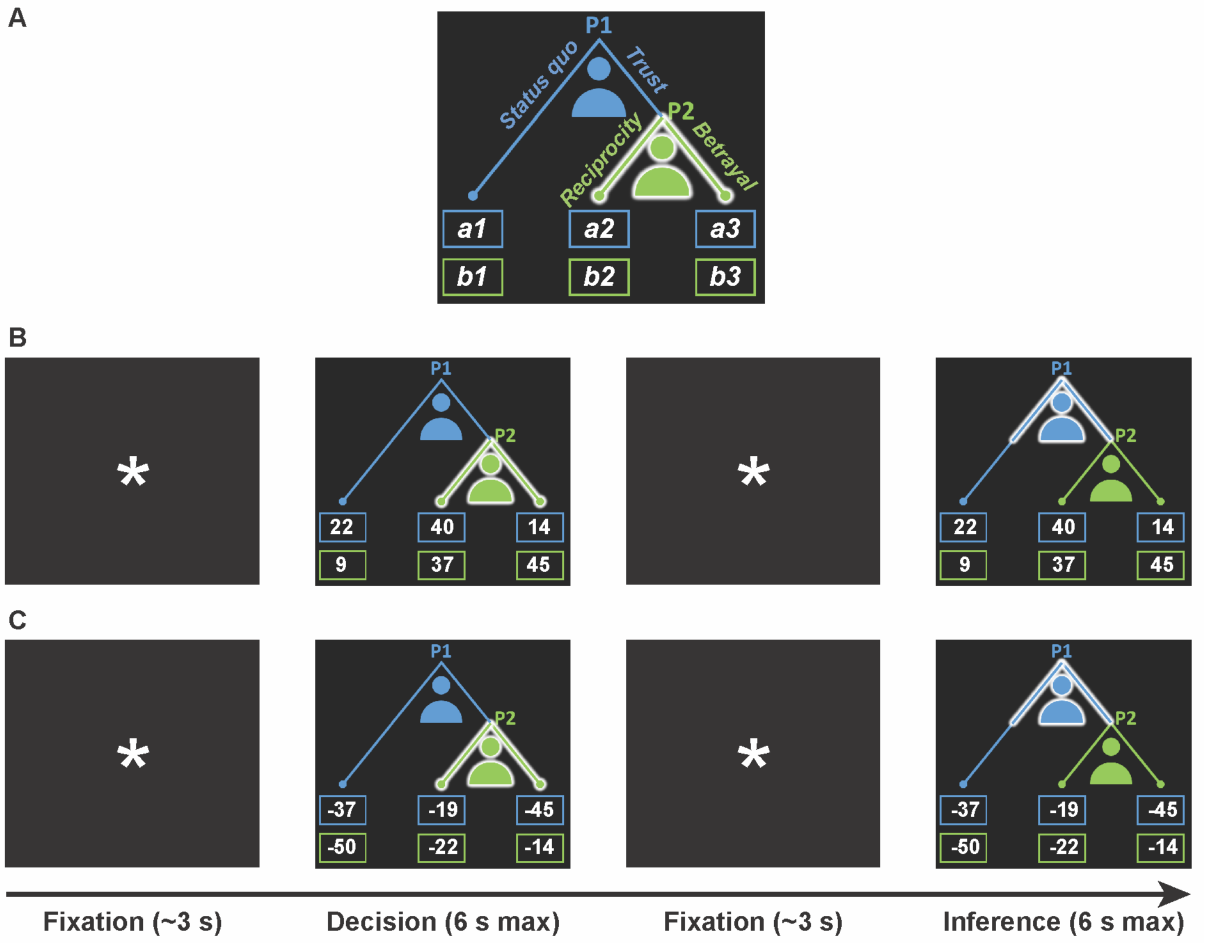

A
P1
Status quo
Trust
P2
Reciprocity
Betrayal
a1
a2
a3
b1
b2
b3
B
*
P1
P2
22
40
14
9
37
45
C
-37
-19
-45
-50
-22
-14
Fixation (~3 s)
Decision (6 s max)
Fixation (~3 s)
Inference (6 s max)

Figure 2. Behavior and modeling results. **(A)** The loss frame tended to decrease reciprocity rates. **(B)** The loss frame significantly reduced advantageous inequity aversion. **(C)** Advantageous inequity aversion positively correlated with reciprocity rate in both gain and loss frame. **(D)** The loss frame indirectly boosted reciprocity by reducing advantageous inequity aversion. ~: $p < 0.1$; *: $p < 0.05$; ***: $p < 0.001$.

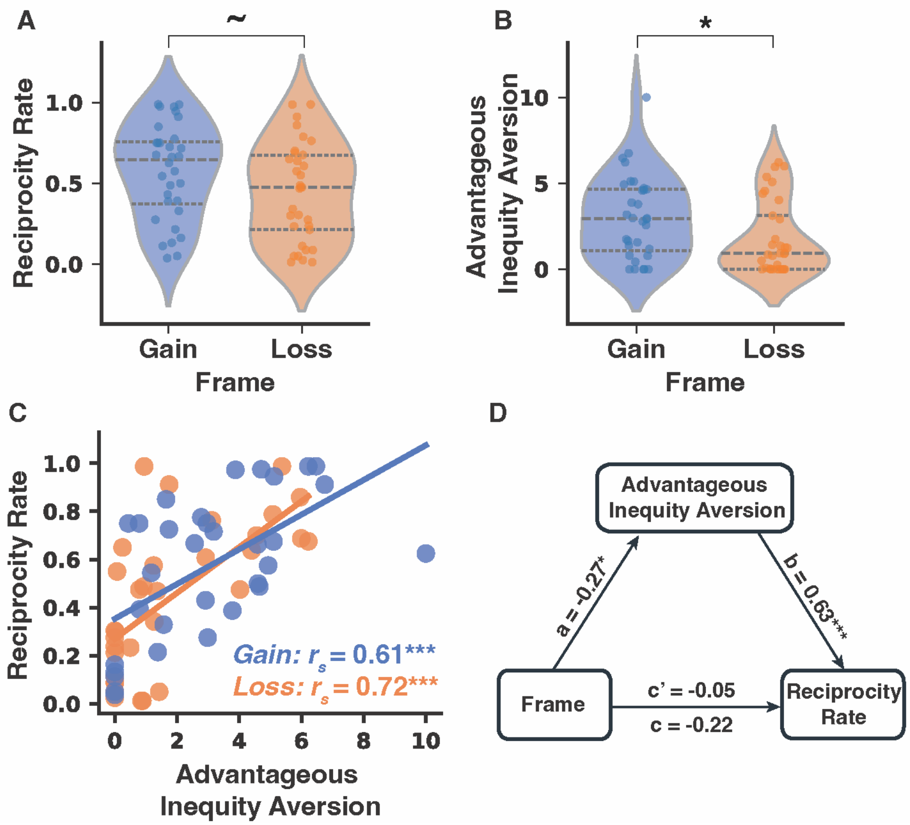


Figure 3. Brain regions associated with advantageous inequity aversion in overall reciprocity decision right amygdala activity. A negative association with advantageous inequity aversion was identified in the right amygdala. Advantageous inequity aversion is negatively associated with right amygdala activity in the gain frame but not in the loss frame. No significant difference in right amygdala activity was found between the gain and loss frames. **: $p < 0.01$.

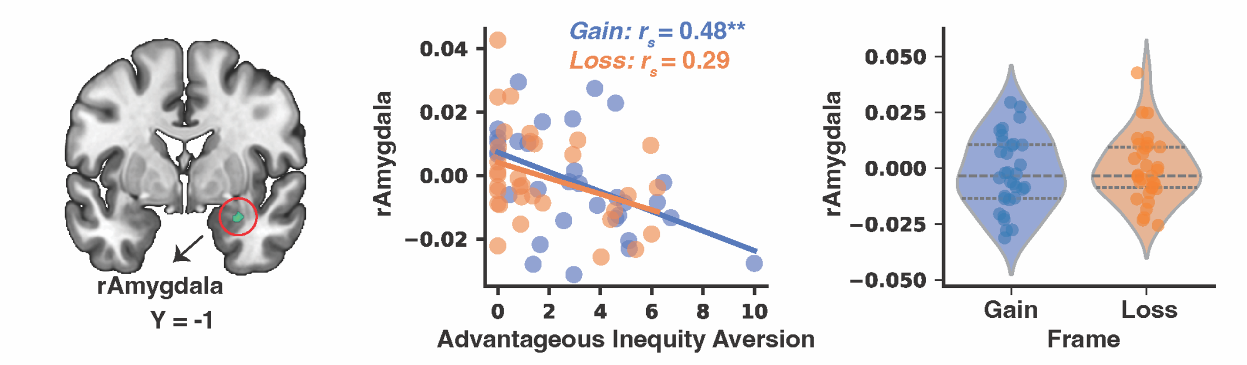


Figure 4. Brain regions associated with advantageous inequity aversion in other-oriented inference. (A) Positive associations with advantageous inequity aversion were identified in the DMPFC, rDLPFC, and lSMG. **(B)** Regardless of context, increased advantageous inequity aversion was associated with increased activity in the DMPFC, rDLPFC, and lSMG. **(C)** No significant differences in DMPFC, rDLPFC, and lSMG activity were observed between the gain and loss frames. **: $p < 0.01$; ***: $p < 0.001$; ns: $p > 0.1$. DMPFC: dorsal medial prefrontal cortex; rDLPFC: right dorsal lateral prefrontal cortex; lSMG: left Supramarginal Gyrus.

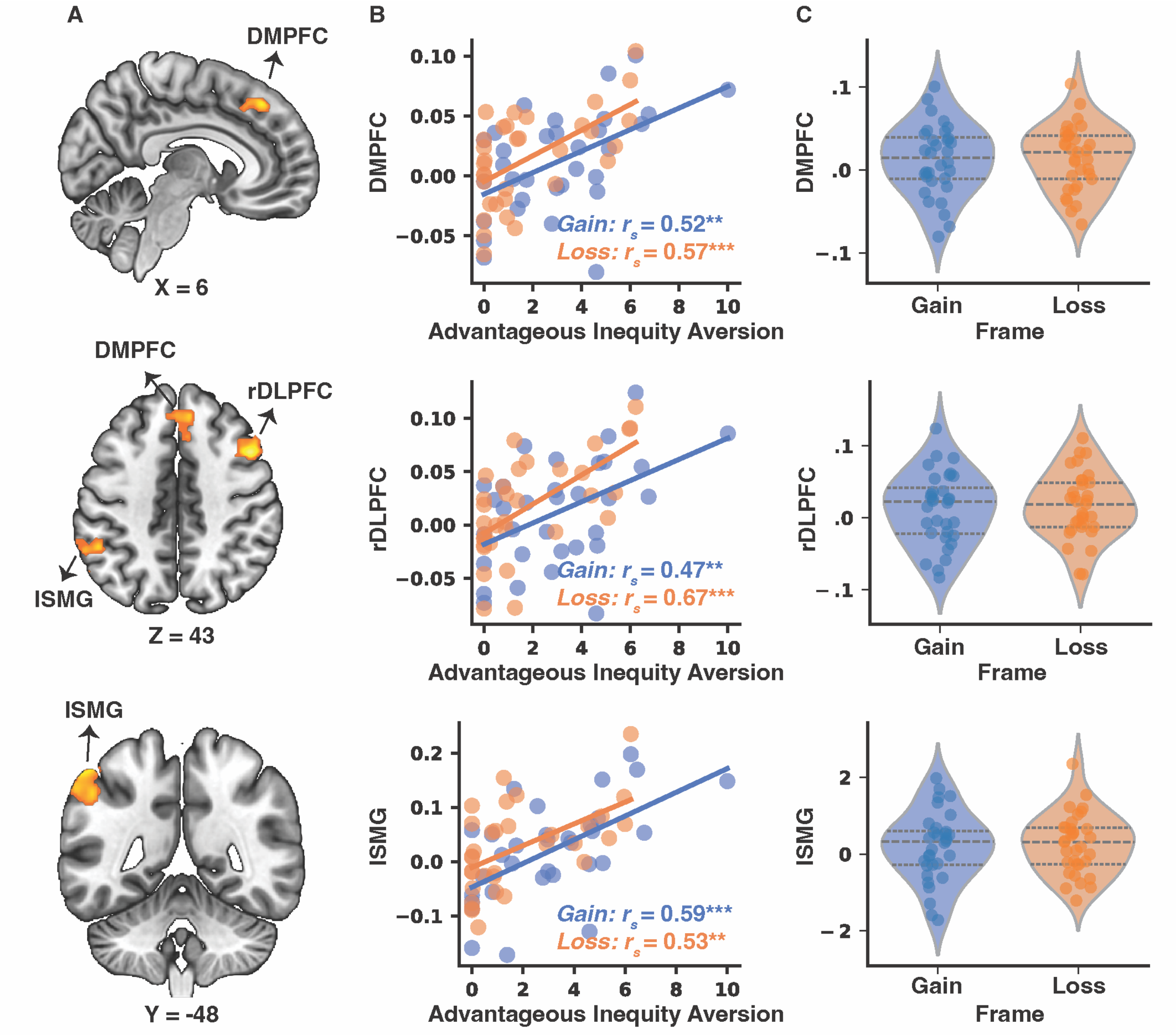


Figure 5. Brain regions associated with advantageous inequity aversion in self-regarding reciprocity. (A) A positive association with advantageous inequity aversion was identified in the lAI. The lAI activity demonstrated a positive correlation with advantageous inequity aversion exclusively in gain contexts, and was significantly elevated in gain compared to loss frame. **(B)** A negative association with advantageous inequity aversion was identified in the rDLPFC. The rDLPFC activity negatively correlated with advantageous inequity aversion under both gain and loss contexts, and did not differ between frames. *: $p < 0.05$; **: $p < 0.01$. lAI: left anterior insula; rDLPFC: right dorsal lateral prefrontal cortex.

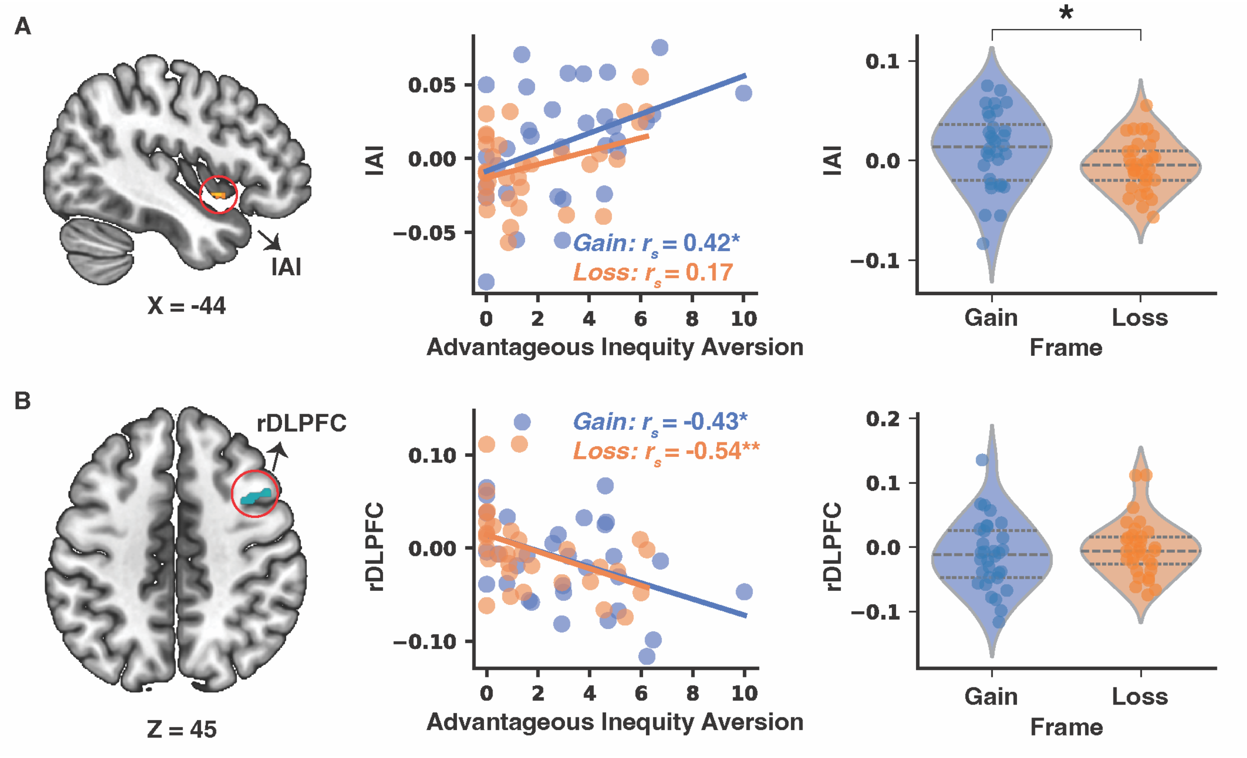


Table Legends

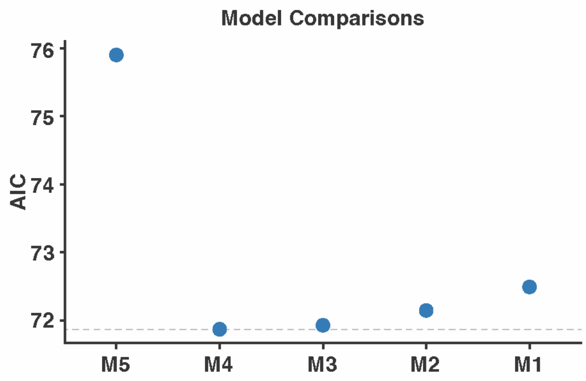
Table 1. Demographic and psychometric measures (mean ± standard error of means).

	Gain	Loss	Group Difference
	(N = 32)	(N = 33)	(χ^2/*U* [p-value])
	Demographic measures		
Gender (male/female)	15/17	16/17	0.01 [0.904]
Age (years)	19.59±0.29	19.27±0.24	477.50 [0.496]
	Psychometric measures		
Machiavellianism	104.97±1.69	105.96±2.04	501.50 [0.733]
Interpersonal Reactivity Index			
Perspective-taking	17.50±0.57	17.45±0.59	512.50 [0.843]
Fantasy	17.44±0.75	17.21±0.78	556.00 [0.717]
Empathic concern	19.38±0.58	19.67±0.58	487.50 [0.597]
Personal distress	15.06±0.76	14.64±0.59	585.50 [0.453]

Supplementary Materials

Model comparison

Figure S1. Model comparisons. Model 4 (M4) outperformed other candidate models. AIC: Akaike Information Criterion.

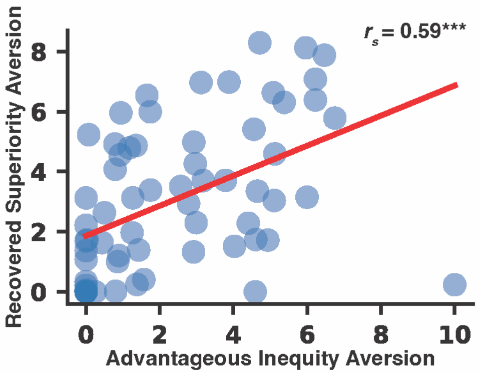


Model validation

Model validation showed that the reciprocity rate predicted by the **M4** was significantly correlated with the true reciprocity rate ($r = 0.98$, 95% CI [0.976, 0.988], $p < 0.001$). Prediction accuracy of **M4** was 0.74, 95% CI: [0.731, 0.750]. Parameter recovery for advantageous inequity aversion also showed that the recovered parameter was significantly correlated with the true parameter ($r_s = 0.59$, $p < 0.001$).

Parameter recovery

Figure S2. Parameter recovery of advantageous inequity aversion. Recovered advantageous inequity aversion was significantly correlated with the true value. ***: p < 0.001.



5. Study 3: Anxiety's Impact on Reciprocity's Core and Periphery

"Trait anxiety impairs reciprocity behavior: A multi-modal and computational modeling study"

Fang, H., Wang, R., Wang, Z., Liu, Q., Luo, Y., Xu, P., & Krueger, F. (2024). Trait anxiety impairs reciprocity behavior: A multi-modal and computational modeling study. [Manuscript submitted for publication]. Department of Psychology, University of Mannheim.

Title

Trait anxiety impairs reciprocity behavior: A multi-modal and computational modeling study

Abbreviated title

Neurocomputational unpacking trait anxiety on reciprocity

Author names and affiliations

Huihua Fang [1,2], Rong Wang [3], Zhihao Wang [4], Qian Liu [5], Yuejia Luo [3,6*], Pengfei Xu [6*], Frank Krueger [1,7]

[1] Department of Psychology, University of Mannheim, Mannheim, Germany

[2] Shenzhen Key Laboratory of Affective and Social Neuroscience, Magnetic Resonance Imaging Center, Center for Brain Disorders and Cognitive Sciences, Shenzhen University, Shenzhen, China

[3] School of Psychology, Chengdu Medical College, Chengdu, China

[4] CNRS—Centre d’Economie de la Sorbonne, Panth ´eon-Sorbonne University, France

[5] Shenzhen Futian Foreign Languages School, Shenzhen, China

[6] Faculty of Psychology, Beijing Normal University, Beijing 100875, China

[7] School of Systems Biology, George Mason University, Fairfax, VA, USA

*Corresponding author

Yuejia Luo, Ph.D.

Email: luoyj@bnu.edu.cn

Pengfei Xu, Ph.D.

Email: pxu@bnu.edu.cn

Abstract

Anxiety significantly impacts reciprocal behavior, crucial for positive social interactions. The neurocomputational mechanisms of anxiety's effects on the core (individual propensity) and peripheral (decision context) factors shaping reciprocity remain unclear. Here, we investigated reciprocity in individuals with low and high trait anxiety using a binary trust game with gain/loss framing, combining computational modeling, eye-tracking, and event-related potentials (ERPs). Our computational model, validated by eye-tracking data, identified four psychological components driving reciprocal behavior: reward, guilt aversion, advantageous inequity aversion, and advantageous inequity liking. Regarding the core of reciprocity, trait anxiety diminished both overall reciprocity and specific psychological components like guilt aversion and advantageous inequity liking, irrespective of context. The reduction in guilt aversion was supported by ERP findings showing decreased P2 (selective attention) and increased LPP (emotion regulation) amplitudes in anxious individuals. Regarding the periphery of reciprocity, trait anxiety altered the contextual perception of both advantageous inequity aversion and reward. Further, trait anxiety reversed the perception of advantageous inequity aversion from gain to loss contexts, a pattern that was linked to the N2 amplitudes (cognitive control). Our findings revealed distinct effects of trait anxiety on core and peripheral factors in reciprocity, offering potential targets for interventions aimed at improving reciprocity in individuals with anxiety disorders.

Introduction

Reciprocity acts as a crucial glue for interpersonal interactions, playing a pivotal role in cultivating a harmonious and thriving society (Schmid et al., 2021). As the global economic slowdown intensifies and geopolitical conflicts escalate, anxiety levels among populations have increased significantly (Collier Villaume et al., 2023). Hence, understanding the impact of anxiety on reciprocity is essential for maintaining social cohesion.

Social representation theory conceptualizes reciprocity decisions as having two main aspects: a core and a periphery (Abric, 1993; Fang et al., 2022; Hagen & Hammerstein, 2006; Wagenaar et al., 1988). The core represents an individual's inherent reciprocity propensity, which remains stable across different contexts. The periphery reflects an individual's contextual perception, shaping the decision by taking into account the situational context. Disentangling reciprocity propensity from contextual dynamics continues to be a challenge in decision-making. One effective solution is to systematically adjust the periphery by reframing the decision context while maintaining the same payoff structure (e.g., framing gains vs. losses) (Evans & van Beest, 2017; Fang et al., 2022). This approach enables the assessment of how trait anxiety influences both the core and periphery of reciprocity, clarifying whether these effects stem from the inherent reciprocity propensity or contextual modifications.

In social interactions, reciprocating rather than betraying a partner’s trust helps mitigate negative emotions such as guilt aversion (failure to meet a partner's reciprocating expectations) and advantageous inequity aversion (feeling discomfort from receiving more than others). However, this often comes at a cost—sacrificing personal benefits, such as financial rewards, which would be the economically optimal choice (Nihonsugi et al., 2015, 2021).

While guilt aversion and reward are widely recognized as key components affecting prosocial behaviors, the concept of advantageous inequity aversion remains controversial. For example, advantageous inequity aversion is not always present; it emerges only when the decision-maker actively chooses an advantageous inequitable distribution but disappears when the distribution is passively received (O. Li et al., 2018). Furthermore, evidence suggests that people constantly compare themselves with others (Festinger, 1954; Fiske, 2011; Starmans et al., 2017), often seeking to increase their own payoff relative to others, demonstrating advantageous inequity liking instead of aversion under certain circumstances (Boyce et al., 2010; Cox, 2013; Dohmen et al., 2011). Since betrayal typically leads to a more advantageous distribution than reciprocity, it is plausible that individuals exhibit advantageous inequity aversion when considering betrayal but show a preference for advantageous inequity when evaluating reciprocity. This scenario creates a complex social dilemma for decision-makers, who must balance moral considerations, social comparison, and personal interests.

Recent research has increasingly utilized computational models to explore the complex psychological components driving reciprocity (Nihonsugi et al., 2015, 2021; Xiao et al., 2022). Those computational models typically suggest that decision-makers aim to maximize the utility of their choices by balancing economic benefits, such as monetary rewards, against the costs of norm violations, including anticipatory negative emotions like guilt and advantageous inequity aversion. However, the extent to which these psychological components reflect the actual psychological processes underlying decision-making has rarely been validated by empirical evidence. Eye-tracking during decision-making, such as in economic games where participants look at payoff structures, can provide direct observation of the decision-maker's focal points (Fiedler et al., 2013; Jiang et al., 2016; Polonio et al., 2015), allowing to validate whether these focuses align with the estimates derived from the proposed model. For example, decision-makers who are more sensitive to guilt aversion are expected to spend relatively more time looking at those payoff structures that trigger guilt aversion. Combining computational modeling and eye-tracking provides deeper insights into the psychological processes underlying reciprocity, including how factors like anxiety influence these processes.

Research on social anxiety and generalized anxiety disorder has shown diminished generosity (Rodebaugh et al., 2016), cooperation (Walters & Hope, 1998), and reciprocity (Anderl et al., 2018; Rodebaugh et al., 2011, 2013), suggesting a broad inhibitory effect of anxiety on prosocial behaviors including reciprocity. Individuals with obsessive-compulsive personality disorder (OCPD), which often co-occurs with high anxiety levels, exhibit less guilt aversion in reciprocity decisions during a binary trust game, suggesting that anxiety might attenuate anticipatory feelings of guilt by restricting affective processing and reducing their sense of moral obligation (Xiao et al., 2022). Additionally, anxious individuals tend to adopt avoidance strategies (Duronto et al., 2005; Raffety et al., 1997; Turner, 1988) and engage in excessive effortful processing when regulating emotion (Aldao et al., 2010; Campbell-Sills et al., 2011; Goldin et al., 2009). Overall, this evidence suggests that anxious individuals may mitigate anticipatory aversive feelings like guilt or advantageous inequity aversion during reciprocity decisions by actively avoiding negative emotions or reducing attention.

Individuals with high trait anxiety are particularly susceptible to contextual effects (Gu et al., 2017; Jepma & López-Solà, 2014; Xu et al., 2013) , which can influence the periphery of reciprocity. Previous studies have mainly focused on non-social aspects of decision-making, like risk evaluation in gambling tasks, rather than on tasks involving social interactions that balance social norms and personal economic benefits (Gu et al., 2017; Jepma & López-Solà, 2014; Xu et al.,

2013) . Although the contextual effects on social reciprocity have been investigated (Evans & van Beest, 2017), the impact of anxiety on these contextual influences and the underlying psychological components of reciprocity remains unexplored. Anxious individuals rely more on heuristics when making decisions, evidenced by increased brain activity in regions associated with cognitive effort during frame-inconsistent decisions (Jepma & López-Solà, 2014; Power & Petersen, 2013) . Further, loss compared to gain frames are heuristically perceived as more harmful to others (Baron, 1995; Evans & van Beest, 2017), and more threatening to one's own payoff (Xu et al., 2013). Therefore, anxiety may alter the perception of other-regarding components, such as guilt and advantageous inequity aversion, and self-beneficial rewards in reciprocity, likely involving cognitive control mechanisms.

To understand how anxiety affects reciprocity, it's essential to investigate the neuropsychological mechanisms by which anxiety influences contextual perceptions and reciprocity propensity. Electroencephalography (EEG), particularly event-related potentials (ERP), can offer significant insights into the temporal dynamics of neural mechanisms. For example, studies have linked P2 with selective attentional allocation (Hajcak et al., 2012; Luck et al., 1994; Potts, 2004; Rey-Mermet et al., 2019) and N2 with effortful top-down cognitive control in decision-making (Cavanagh & Shackman, 2015; Folstein & Van Petten, 2008a; Hao et al., 2023; McLoughlin et al., 2022). Although late positive potential (LLP) is often seen as a marker of emotional reactivity (Hajcak et al., 2010; MacNamara & Proudfit, 2014; Paul et al., 2016; Qi et al., 2016; Thiruchselvam et al., 2011), several studies have linked LPP with emotional regulation and found that an enhanced LPP amplitudes index an increase of cognitive effort in managing emotional responses (Bernat et al., 2011; Desatnik et al., 2017; Moser et al., 2014; Shafir et al., 2015). Specifically, in decisions involving moral conflict, larger LPP amplitudes indicate more cognitive efforts deployed to resolve these conflicts (Chen et al., 2009; Zhan et al., 2018, 2020).

To investigate the neurocomputational mechanisms underlying reciprocity decisions, we formulated four hypotheses. First, based on prior research (Boyce et al., 2010; Cox, 2013; Dohmen et al., 2011; O. Li et al., 2018; Nihonsugi et al., 2015, 2021; Xiao et al., 2022), we hypothesized that the best computational model for reciprocity decisions would consist of four distinct psychological components: reward, guilt aversion, advantageous inequity aversion, and advantageous inequity liking. Second, we predicted that individuals more sensitive to these psychological components would spend relatively more time visually attending to the according payoff structure within the binary trust game, reflecting their evaluation processes during reciprocity decisions (Fiedler et al., 2013; Jiang et al., 2016; Polonio et al., 2015). Third, regarding the core of reciprocity, we expected trait anxiety to attenuate the core of reciprocity across gain and loss frames, given that anxious individuals exhibit reduced reciprocity behavior (Anderl et al., 2018; Rodebaugh et al., 2011, 2013). In particular, we hypothesized that high trait anxiety would reduce guilt and advantageous inequity aversion, as anxious individuals tend to adopt avoidance strategies (Duronto et al., 2005; Raffety et al., 1997; Turner, 1988) and mitigate negative feelings by limiting affective processing and reducing moral obligation (Kantor, 2016; Xiao et al., 2022). We anticipated this process would manifest in the P2 and LPP ERP components, given P2's association with selective attentional allocation (Hajcak et al., 2012; Luck et al., 1994; Potts, 2004; Rey-Mermet et al., 2019) and LPP's connection to emotional regulation (Bernat et al., 2011; Desatnik et al., 2017; Moser et al., 2014; Shafir et al., 2015) and moral conflict (Chen et al., 2009; Zhan et al., 2018, 2020). Fourth, regarding the periphery of reciprocity, we postulated that trait anxiety modulates reciprocity differently under gain and loss contexts, given that anxious individuals are more susceptible to contextual influences (Gu et al., 2017; Xu et al., 2013). We proposed that trait anxiety would differentially modulate other regarding components (i.e., guilt aversion and advantageous inequity aversion) and reward sensitivity across varying contextual frames, given that anxious individuals often rely on heuristics in decision-making (Jepma & López-Solà, 2014) , and that loss frames are generally perceived as more detrimental to others (Baron, 1995; Evans & van Beest, 2017) and to one's own economic interests (Xu et al., 2013). We predicted these anxiety-related alterations would engage N2 cognitive control mechanisms.

To test our hypotheses, we combined EEG, eye-tracking, and computational modeling in a binary trust game to investigate the neurocomputational mechanisms of reciprocity decisions in individuals with low and high trait anxiety under gain and loss frames (**Fig. 1**). Participants also completed in addition two self-report questionnaires, the Interpersonal Reactivity Index (IRI) measuring aspects of empathy (Davis, 1980) and Machiavellianism (Mach-IV) scale measuring the tendency towards selfishness (Christie & Geis, 1970). Our computational modeling analysis identified four psychological components underlying reciprocity decisions—reward, guilt aversion, advantageous inequity aversion, and advantageous inequity liking—which were validated by the eye-tracking analyses. Regarding the core, our findings revealed that trait anxiety reduced reciprocity at the behavioral level and diminished guilt aversion and advantageous inequity liking at the psychological level. At the neural level, trait anxiety attenuated guilt aversion by decreasing P2 amplitude (related to selective attention) and increasing LPP amplitude (associated with effortful emotion regulation). Regarding the periphery, trait anxiety influenced the contextual effects on reward and advantageous inequity aversion at the psychological level, although it did not affect reciprocity behaviorally. Specifically, trait anxiety modified the contextual effect on advantageous inequity aversion through the N2 cognitive control mechanism at the neural level.

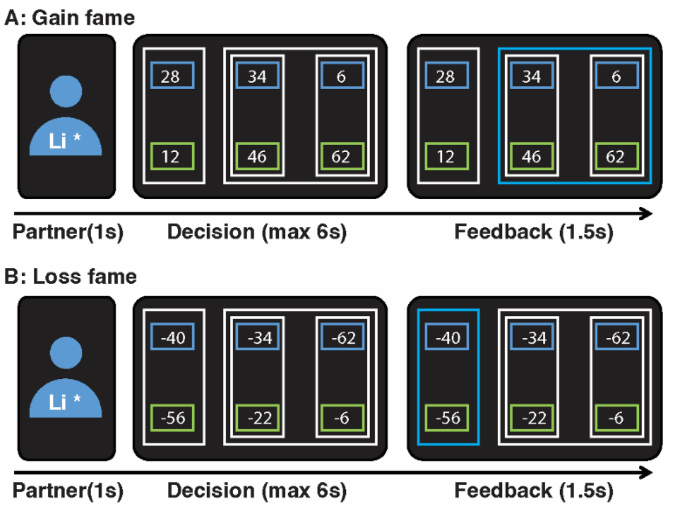


Figure 1. Experimental design. Participants completed a trust game with two distinct frame sessions in a counterbalanced order: **(A)** Gain frame session and **(B)** Loss frame session. The gain frame session started with 0 points while the loss frame session started with 15,000 points. If participants used the same strategies in both frames, their final outcomes would be identical (Evans & van Beest, 2017). For each trial session, participants were introduced to a different anonymous partner, represented by an icon with a partially obscured name (i.e., one word in the first name was blocked). Participants then decided whether to “reciprocate” (e.g., in the gain frame, choosing the middle rectangle resulted in a distribution of 46 points for themselves and 34 points for their partner; in the loss frame, it involved a deduction of 22 points for themselves and 34 points for their partner) or “betray” (e.g., in the gain frame, choosing the right rectangle resulted in a distribution of 62 points for themselves and 6 points for their partner; in the loss frame, it involved a deduction of 6 points for themselves and 62 points for their partner) without knowing whether the partner had chosen “trust” (e.g., the larger rectangle containing both the “reciprocate” and “betray” option) or “distrust” (e.g., in the gain frame, the left rectangle resulted in a distribution of 12 points for themselves and 28 points for their partner; in the loss frame, it involved a deduction of 56 points for themselves and 40 points for their partner). Finally, participants received feedback with a blue highlight indicating whether the partner had initially chosen to "trust" (as shown in **A**) or "distrust" (as shown in **B**). Participants were unaware the feedback was randomly generated by the computer. Participants were informed of the rules before the game that if the partner had chosen to "distrust," the payoff would be distributed accordingly, but if the partner had chosen to "trust," the payoff would be based on the participant's decision.

Results

Computational Modeling of Reciprocity

Computational modeling was employed to unveil the psychological components behind reciprocity as measures with the binary trust game. Seven plausible candidate models (see **Materials and Methods**)—adapted from previous studies (Nihonsugi et al., 2015, 2021; Xiao et al., 2022) with components including reward, guilt aversion, inequity aversion, advantageous inequity aversion, and advantageous inequity liking—were constructed and compared using a stage-wise approach (Gagne et al., 2020; Wang et al., 2023; Zhang & Gläscher, 2020).

The proposed model 4 (**M4**, pseudo $r^2 = 0.359$; **Fig. 2A**; **Tab. 1**) outperformed all other candidate models, as indicated by the leave-one-out information criterion (LOOIC) and widely applicable information criterion (WAIC). **M4** included the components of reward sensitivity, guilt aversion, advantageous inequity aversion, and advantageous inequity liking, explaining over 93% of the variance in reciprocity rates (**Fig. 2B**). Higher subjective sensitivity to guilt aversion, advantageous inequity aversion, and advantageous inequity liking was associated with increased reciprocity rates, while greater sensitivity to reward with lower reciprocity rates (**Fig. 2C**).

Table 1. Model comparison. The winning model M4 outperformed all other candidate models, as indicated by LOOIC and WAIC. **M4** included the components of reward sensitivity (**$\boldsymbol{\beta_R}$),** guilt aversion $\boldsymbol{\beta_G}$, advantageous inequity aversion $\boldsymbol{\beta_{Ad-IA}}$, advantageous inequity liking $\boldsymbol{\beta_{Ad-IL}}$ and Inverse tempareture (**$\boldsymbol{\lambda}$**) for four conditions (Anxiety [*high, low*] × Frame [*gain, loss*]). *ΔLOOIC, leave-one-out information criterion relative to the winning model; ΔWAIC, widely applicable information criterion relative to the winning model.*

Models	*Parameters*	*Number of parameter components*	*ΔLOOIC*	*ΔWAIC*
M4	$\boldsymbol{\beta_R, \beta_G, \beta_{Ad-IA}, \beta_{Ad-IL}, \lambda}$	***20***	***0***	***0***
M3	$\beta_R, \beta_G, \beta_{Ad-IA}, \beta_{DisAd-I}, \lambda$	*20*	*10.9*	*8.2*
M5	$\beta_R, \beta_G, \beta_{Ad-IA}, \beta_{Ad-IL}, \lambda$	*20*	*11.4*	*10.5*
M2	$\beta_G, \beta_{Ad-IA}, \beta_{DisAd-IA}, \lambda$	*16*	*13.5*	*21*
M6	$\beta_R, \beta_G, \beta_{Ad-IA}, \beta_{Ad-IL}, \lambda$	*18*	*23.1*	*27.5*
M1	$\beta_G, \beta_I, \lambda$	*12*	*55.8*	*59.5*
M7	$\beta_R, \beta_G, \beta_{Ad-IA}, \beta_{Ad-IL}$	*16*	*51926.5*	*128630645*

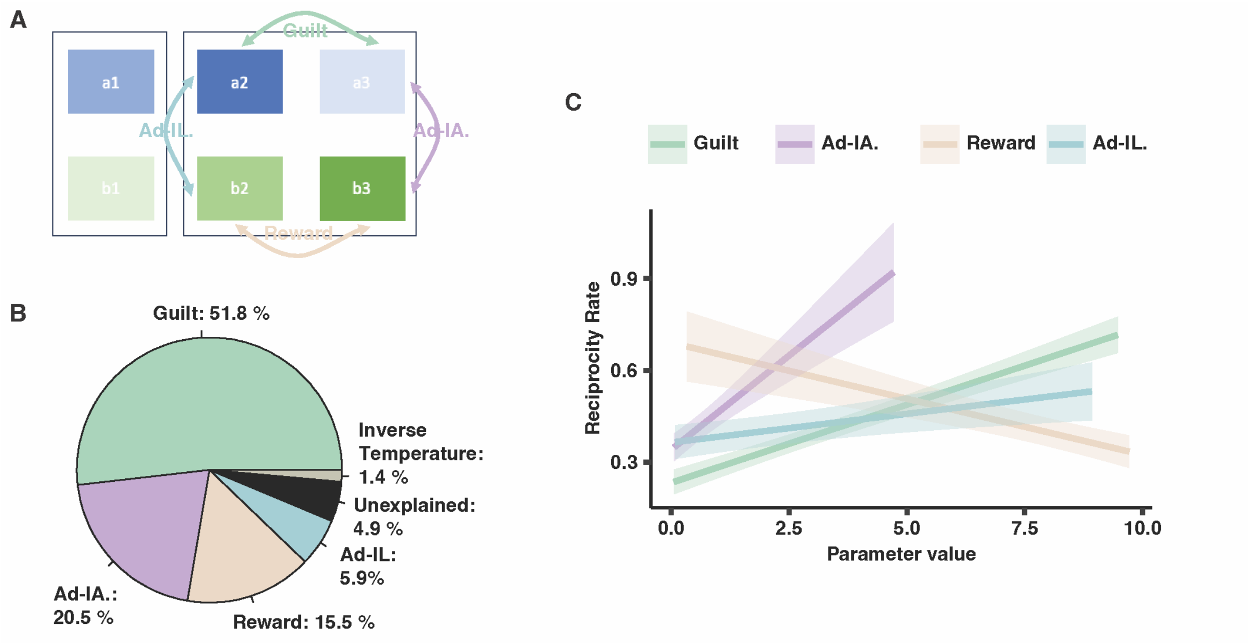


Figure 2. Components in the winning model and their association with reciprocity. (A) Schematic of winning model. Schematic illustration of the payoff structure in the binary trust game with components of the winning model: guilt aversion, advantageous inequity aversion, reward sensitivity, and advantageous inequity liking. The blue-filled areas (a2>a1>a3) were related to the payoff structure of the trustor (partner) and the green-filled areas (b3>b2>b1) for the trustee (participant) in both gain and loss frames (Note that deeper color signifies higher value of gain or lower value of lose). The objective size of guilt aversion was quantified as the difference in the payoff structure between a2 and a3; advantageous inequity aversion between b3 and a3; reward between b3 and b2; and advantageous inequity liking between b2 and a2. **(B) Variance explained by components.** The four components explained more than 93% of the variance of the reciprocity rate. Inverse temperature (1.4%) controls the trade-off between randomness and determinism in decision-making processes and there is a 4.9% variance unexplained by the model. **(C) Association between parameters of components and reciprocity rate.** Higher subjective sensitivity to guilt aversion, advantageous inequity aversion, and advantageous inequity liking were associated with increased reciprocity rates, whereas higher sensitivity to reward was associated with lower reciprocity rates. Guilt: guilt aversion; Ad-IA.: advantageous inequity aversion; Ad-IL.: advantageous inequity liking; Reward: reward sensitivity.

Model prediction from **M4** demonstrated that the true and simulated reciprocity rates were highly correlated (*rs* > 0.96, Fig. S2). Parameter recovery for **M4** also indicated successful recovery of all parameters from **M4** (guilt aversion: *rs* > 0.90; advantageous inequity aversion: *rs* > 0.69; reward: *rs* > 0.64; advantageous inequity liking: *rs* > 0.68; inverse temperature: *rs* > 0.55; **Fig. S3**).

Validation of Winning Model through Eye-Tracking Data

To validate the winning model, the relationship between the estimated parameters of the components and observed eye movements was investigated. Six areas of interest (AOIs) were defined, corresponding to the six rectangles containing payoffs in the binary trust game (**Fig. 2A**). The relative time spent on transitions between each pair of AOIs was calculated based on the fixation sequence of eye movements within each trial (see **Materials and Methods**). As a result, 21 unique transitions were extracted and the proportion of fixation time (normalized with the total reaction time for each trial) on each transition was calculated.

Validating our winning model (**M4**), individuals spent significantly more time on the transitions related to its four identified components compared to all other transitions (**Fig. 3A**). The Linear Mixed Model (LMM) (**Tab. 2**) indicated that greater subjective sensitivity to specific components was associated with more time spent on transitions related to those components (**Fig. 3B**). Furthermore, resembling the relationship between model parameter and reciprocity rate (**Fig. 2C**), increased fixation time on transitions involving guilt aversion, advantageous inequity aversion, and advantageous inequity liking correlated with a higher rate of reciprocity, while increased fixation time on reward transitions was linked to a lower rate of reciprocity (**Fig. 3C, Tab. 2**). These eye-movement results confirmed that the parameter estimated by the winning model reflects the underlying psychological components of reciprocity.

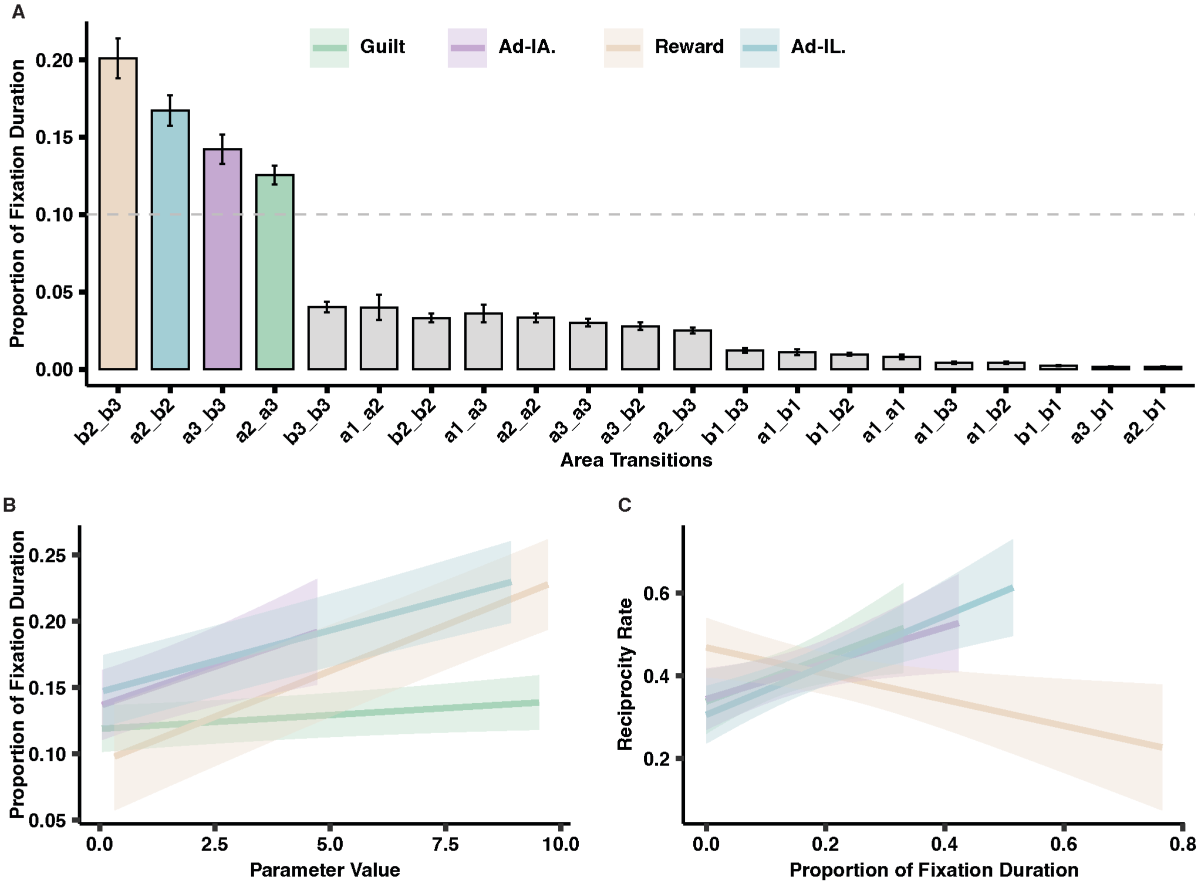


Figure 3. The fixation duration of the transitions and their associations with component parameters and reciprocity rate. (A) Proportion of fixation duration on each transition (Mean ± standard error of mean, SE). The transitions related to the four components (guilt aversion [a2_a3], advantageous inequity aversion [a3_b3], advantageous inequity liking [a2_b2], and reward [b2_b3]) dominated the proportion of fixation duration when making reciprocity decisions. Dash line represent a 10% proportion of fixation duration. **(B) Association between parameter value of components and fixation duration.** Higher sensitivity to the components was associated with an increased proportion of fixation duration on related transitions. **(C) Association between fixation duration and reciprocity rate.** An increased proportion of duration on the transitions of guilt aversion, advantageous inequity aversion, and advantageous inequity liking predicted a higher rate of reciprocity, while an increased proportion of duration on the transitions of reward predicts a lower rate of reciprocity. Guilt: guilt aversion; Ad-IA.: advantageous inequity aversion; Ad-IL.: advantageous inequity liking; Reward: reward sensitivity.

Table 2. Association between fixation duration on components and subjective sensitivity to corresponding components and reciprocity rate. Left columns: An increased proportion of fixation duration of components was associated with higher subjective sensitivity to the corresponding components. **Right columns:** increased proportion of duration on the transitions of guilt aversion [a2_a3], advantageous inequity aversion [a3_b3], and advantageous inequity liking [a2_b2] predicted a higher rate of reciprocity, while an increased proportion of duration on the transitions of reward [b2_b3] predicts a lower rate of reciprocity. SE: standard error of mean; CI (95%): 95% confidence interval.

	Subjective sensitivity to components					**Reciprocity Rate**				
Transitions	*Beta*	*SE*	*CI (95%)*	*t*	*p*	*Beta*	*SE*	*CI (95%)*	*t*	*p*
Guilt aversion [a2_a3]	0.04	0.02	0.00 – 0.07	2.04	0.042	0.04	0.02	0.01 – 0.07	2.29	0.024
Advantageous inequity aversion [a3_b3]	0.04	0.01	0.02 – 0.07	3.24	0.001	0.05	0.02	0.01 – 0.09	2.29	0.024
Reward [b2_b3]	0.11	0.01	0.09 – 0.14	8.80	<0.001	-0.05	0.02	-0.09 – -0.01	-2.46	0.016
Advantageous inequity liking [a2_b2]	0.10	0.01	0.08 – 0.13	7.83	<0.001	0.07	0.02	0.03 – 0.10	3.88	<0.001

Impact of Trait Anxiety on Core and Periphery of Reciprocity

The LMM on the reciprocity rate revealed a significant main effect of Anxiety, $\chi^2(1) = 4.37$, $p = 0.036$, $\eta_p^2 = 0.07$ (**Fig. 4**), where individuals with high trait anxiety showed a lower reciprocity rate compared to those with low trait anxiety. The main effect of Frame was marginally significant, $\chi^2(1) = 3.20$, $p = 0.074$, $\eta_p^2 = 0.05$, where individuals under the Gain Frame reciprocated more than those under the Loss Frame. The Anxiety × Frame interaction effect was not significant, $\chi^2(1) = 0.003$, $p = 0.954$. The presence of the main effect of Anxiety and the absence of Anxiety × Frame interaction effect suggest that trait anxiety attenuated reciprocity regardless of contexts.

The LMM on the reaction time revealed no significant main effect of Anxiety, $\chi^2(1) = 0.21$, p = 0.646, nor a significant Anxiety x Frame interaction, $\chi^2(1) = 1.49$, p = 0.221, but a marginal main effect of Frame, $\chi^2(1) = 3.06$, p = 0.085, with slower decisions under the Loss Frame than the Gain Frame (**Fig. S1**).

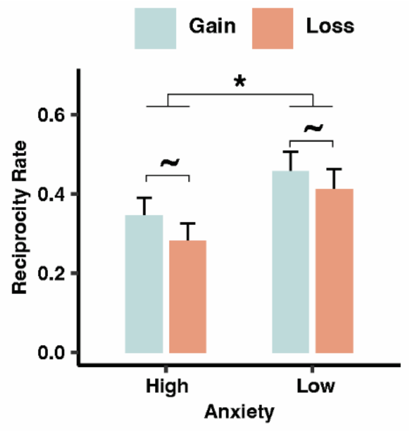


Figure 4. The impact of anxiety on reciprocity rate (Mean ± SE). High Anxiety group exhibited lower reciprocity rates than Low Anxiety group. Further, individuals under a Loss Frame showed a trend toward a lower reciprocity rate compared to those under a Gain Frame. *: $p < 0.05$, ~: $p < 0.1$.

Computational mechanism underlying trait anxiety's impact on core and periphery of reciprocity

The parameter β_G, β_{Ad-IA}, β_R and β_{Ad-IL} in **M4** represent the participant's sensitivity to guilt aversion, advantageous inequity aversion, reward, and advantageous inequity liking, respectively.

Guilt aversion. The LMM on the guilt aversion parameters revealed no significant main effect of Frame, $\chi^2(1) = 0.22$, p = 0.638, but a significant main effect of Anxiety, $\chi^2(1) = 4.23$, $p = 0.039$, $\eta_p^2 = 0.06$, where the High Anxiety group showed lower guilt aversion compared to the Low Anxiety group (**Fig. 5A**). The main effect of Anxiety is significant and the interaction effect of Anxiety × Frame was not, $\chi^2(1) = 0.09$, $p = 0.759$, indicating that trait anxiety attenuated guilt aversion regardless of contexts.

Advantageous inequity aversion. The LMM on advantageous inequity aversion parameters showed no significant main effect of Frame, $\chi^2(1) = 0.01$, p = 0.921, but a significant main effect of Anxiety, $\chi^2(1) = 4.53$, $p = 0.033$, $\eta_p^2 = 0.07$, where the High Anxiety group had a lower advantageous inequity aversion compared to the Low Anxiety group (**Fig. 5 B**). Further, the interaction effect of Anxiety × Frame was significant, $\chi^2(1) = 11.65$, $p < 0.001$, $\eta_p^2 = 0.16$. Planned follow-up post-hoc analyses revealed that Low Anxiety group showed lower advantageous inequity aversion in the Loss than in the Gain Frame, β = 0.41, SE = 0.17, t = 2.46, $p = 0.017$, but a reverse pattern was observed for the High Anxiety group, β = -0.38, SE = 0.16, t = -2.36, $p = 0.021$. The presence of a reversed contextual effect on advantageous inequity aversion between the High and the Low Anxiety group suggests that trait anxiety altered the contextual perception of advantageous inequity aversion. Moreover, High Anxiety group exhibited lower advantageous inequity aversion than those in the Low Anxiety group under the Gain Frame, β = 0.71, SE = 0.19, t = 3.78, $p = 0.003$, but not Loss Frame, β = 0.08, SE = 0.19, t = 0.42, $p = 0.675$. The presence of anxiety effect on advantageous inequity aversion in the Gain Frame but absence in the Loss Frame suggests that the effect of anxiety on advantageous inequity aversion depended on contexts.

Reward sensitivity. The LMM on reward sensitivity parameters revealed no significant main effect of Anxiety, $\chi^2(1) = 1.37$, p = 0.242, but a significant main effect of Frame, $\chi^2(1) = 20.77$, $p < 0.001$, $\eta_p^2 = 0.25$, where individuals under the Loss Frame showed significantly higher reward sensitivity compared to the Gain Frame (**Fig. 5C**). Further, the interaction effect of Anxiety x Frame was significant, $\chi^2(1) = 14.03$, $p < .001$, $\eta_p^2 = 0.19$. Planned follow-up post-hoc analyses revealed that individuals with high anxiety showed greater reward sensitivity in the Loss than in Gain Frame, β = -2.43, SE = 0.41, t = -5.92, $p < 0.001$. However, this pattern disappeared in the Low Anxiety group, β = -0.24, SE = 0.42, t = -0.57, $p = 0.571$. The presence of contextual effect on reward sensitivity in the High but absence in the Low Anxiety group suggests that trait anxiety altered the contextual perception of reward sensitivity. Moreover, High Anxiety group exhibited lower reward sensitivity than those in the Low Anxiety group under the Gain Frame, β = -1.55, SE = 0.49, t = -3.19, $p = 0.002$, but not Loss Frame, β = 0.64, SE = 0.49, t = 1.32, $p = 0.189$. The presence of anxiety effect on reward sensitivity in the Gain Frame but absence in the Loss Frame suggests that the effect of anxiety on reward sensitivity depended on contexts.

Advantageous inequity liking. The LMM on advantageous inequity liking parameters indicated a significant main effect of Anxiety, $\chi^2(1) = 5.34$, p = 0.021, $\eta_p^2 = 0.08$, demonstrating the High Anxiety group showed lower advantageous inequity liking than the Low Anxiety group (**Fig. 5D**). Further, a significant main effect of Frame was observed, $\chi^2(1) = 35.40$, $p < 0.001$, $\eta_p^2 = 0.37$, with a higher advantageous inequity liking in the Gain compared to the Loss Frame. However, the interaction effect of Anxiety x Frame was not significant, $\chi^2(1) = 2.11$, p = 0.147, suggesting that trait anxiety attenuated advantageous inequity liking regardless of contexts.

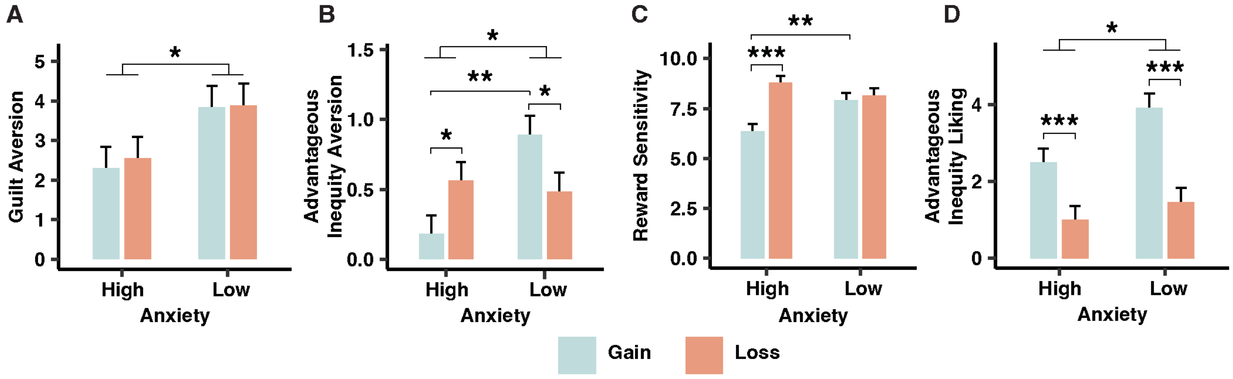


Figure 5. Impact of trait anxiety on the psychological components of reciprocity (Mean ± SE). (A) Guilt Aversion. Trait anxiety attenuated guilt aversion regardless of context. **(B) Advantageous inequity aversion.** Trait anxiety altered contextual perception, with high and low trait anxiety individuals showing reversed contextual effects. In addition, high trait anxiety individuals demonstrated a significant contextual effect, whereas low trait anxiety individuals showed no contextual effect. **(C) Reward sensitivity.** Trait anxiety altered the contextual perception of reward sensitivity, with high trait anxiety individuals displaying a contextual effect, while those with low trait anxiety showed no contextual effect. **(D) Advantageous inequity liking**. Trait anxiety reduced advantageous inequity liking regardless of context, and the loss frame decreased advantageous inequity liking compared to the gain frame. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$.

Brain mechanism underlying trait anxiety's impact on core and periphery of reciprocity

Given that guilt aversion and advantageous inequity liking are psychological components affected by trait anxiety in the core of reciprocity (i.e., Anxiety effect regardless of contexts), and advantageous inequity aversion and reward in the periphery (i.e., Anxiety effect depended on contexts), the brain mechanisms underlying the impact of trait anxiety on core and periphery were investigated. First, ERP components of P2, N2, and LPP were examined, differentiating the impact of trait anxiety on the core and peripheral processes.

P2 component. The LMM analysis on the P2 amplitude showed a significant main effect of Anxiety, $\chi^2(1) = 4.28$, $p = 0.039$, $\eta_p^2 = 0.07$, where the High Anxiety group exhibited a lower P2 amplitude than the Low Anxiety group (**Fig. 6A, B**). No main effect of Frame, $\chi^2(1) = 2.50$, $p = 0.114$, nor an interaction effect of Anxiety × Frame was found, $\chi^2(1) = 0.05$, $p = 0.829$. The presence of the Anxiety main effect and absence of Anxiety × Frame interaction effect suggests that trait anxiety decreased P2 amplitude regardless of contexts.

N2 component. The LMM analysis on the N2 amplitude demonstrated no significant main effect of Frame, $\chi^2(1) = 0.32$, p = 0.573, nor a main effect of Anxiety, $\chi^2(1) = 2.67$, p = 0.102, where the High Anxiety group exhibited a lower N2 amplitude than the Low Anxiety group (**Fig. 6A, C**). Further, a significant Anxiety × Frame interaction effect was observed, $\chi^2(1) = 9.46$, $p < 0.001$, $\eta_p^2 = 0.15$. Planned follow-up post-hoc analyses indicated that the Low Anxiety group showed a greater N2 amplitude under the Loss compared to the Gain Frame, β = -0.46, SE = 0.18, z = -2.56, $p = 0.011$. However, the High Anxiety group showed a reversed pattern, demonstrating a marginally significantly smaller N2 amplitude under the Loss compared to the Gain Frame, β = 0.32, SE = 0.18, z = 1.79, $p = 0.074$. The presence of a reversed contextual effect on N2 amplitude between the High and the Low Anxiety group suggests that trait anxiety altered the contextual effect of N2 amplitude. Moreover, High Anxiety group exhibited a smaller N2 amplitude than Low Anxiety group under the Gain Frame, β = 1.26, SE = 0.54, z = 2.32, $p = 0.020$, but not Loss Frame, β = 0.47, SE = 0.55, z = 0.87, $p = 0.387$. The presence of anxiety effect on N2 amplitude in the Gain Frame but absence in the Loss Frame suggests that the effect of trait anxiety on N2 amplitude depended on contexts.

LPP component. The LMM analysis on the LPP amplitude showed a significant main effect of Anxiety, $\chi^2(1) = 4.19$, $p = 0.041$, $\eta_p^2 = 0.07$, where the High anxiety group exhibited a higher LPP amplitude than the Low Anxiety group (**Fig. 6A, D**). However, no significant main effect of Frame, $\chi^2(1) = 0.12$, $p = 0.716$, nor interaction effect of Anxiety × Frame were found, $\chi^2(1) = 0.23$, $p = 0.633$. The presence of Anxiety main effect and absence of Anxiety × Frame interaction effect suggests that trait anxiety decreased LPP amplitude regardless of contexts.

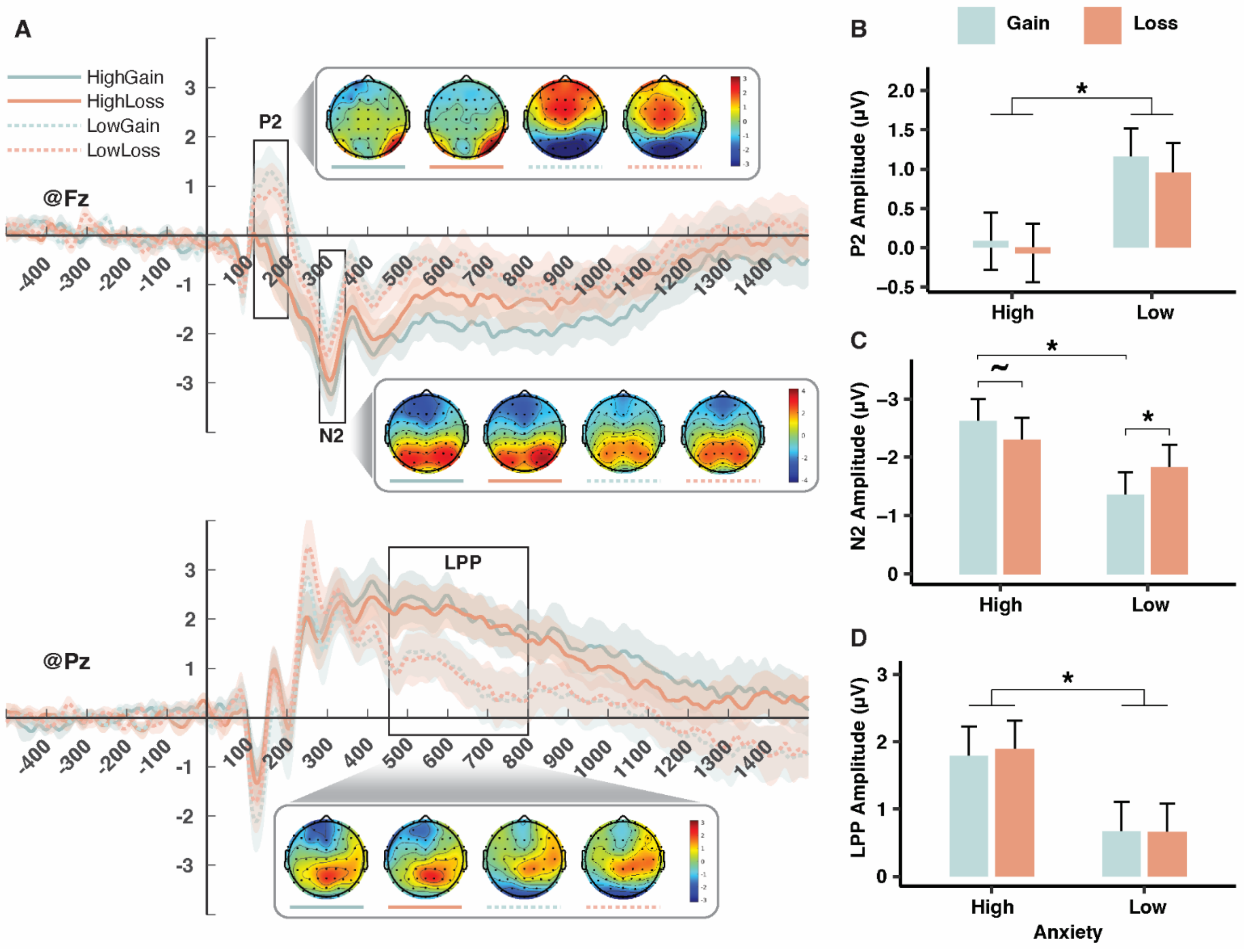


Figure 6. Condition comparisons of event-related potentials on reciprocity decision (Mean ± SE). (A) Grand average potential waveforms. Grand average potential waveforms for the four conditions (Anxiety × Frame) were measured at the Fz electrode and Pz. **(B) Comparison of P2 amplitudes.** P2 amplitudes (measured at F1, Fz, F2) were significantly lower in individuals with high trait anxiety compared to those with low trait anxiety regardless of frame. **(C) Comparison of N2 amplitudes**. N2 amplitudes (measured at F1, Fz, F2) showed a reversed contextual effect for individuals with high and low trait anxiety. **(D) Comparison of LPP amplitudes.** LPP amplitude (measured at P1, Pz, P2) was significantly lower in individuals with high trait anxiety than those with low trait anxiety regardless of frame. *: p < 0.05, ~: p < 0.1.

Second, mediation analyses were employed to investigate how trait anxiety impacted these ERP components and, subsequently, the psychological components underlying reciprocity. Since trait anxiety affected P2 and LPP responses similarly to guilt aversion and advantageous inequity liking regarding the core of reciprocity, serial mediation analysis was used to test if trait anxiety impacts these psychological components through changes in P2 and LPP amplitudes.

The serial mediation analysis revealed that higher trait anxiety was associated with a lower P2 amplitude and a higher LPP amplitude, leading to reduced guilt aversion (indirect effect [a*b*c] = -0.017, SE = 0.003, z = -6.34, $p < 0.001$, 95% CI [-0.022, -0.012]) (**Fig. 7A**). Further, higher trait anxiety attenuated guilt aversion through increased LPP amplitude alone (indirect effect [a2*c] = -0.090, SE = 0.015, z = -6.15, $p < 0.001$, 95% CI [-0.121, -0.063]). Moreover, trait anxiety did not significantly modulate guilt aversion through P2 amplitude alone (indirect effect [a*b2] = 0.019, SE = 0.015, z = 1.23, $p = 0.220$, 95% CI [-0.012, 0.048]). Those results demonstrated that higher trait anxiety led to lower guilt aversion by increased LPP amplitude alone or by reducing P2 amplitude, which then translated into increased LPP amplitude.

However, the serial mediation analysis showed no significant indirect effect of trait anxiety on advantageous inequity aversion through both P2 and LPP amplitude (indirect effect [a*b*c] = 0.001, SE = 0.001, z = 1.01, $p = 0.311$, 95% CI [-0.000, 0.002]), nor effects through P2 amplitude alone (indirect effect [a*b2] = 0.009, SE = 0.006, z = 1.40, $p = 0.163$, 95% CI [-0.003, 0.022]) or LPP amplitude alone (indirect effect [a2*c] = 0.003, SE = 0.003, z = 1.02, $p = 0.309$, 95% CI [-0.003, 0.010]). These insignificant results suggest that P2 or LPP amplitude was not involved in the processing of advantageous inequity liking.

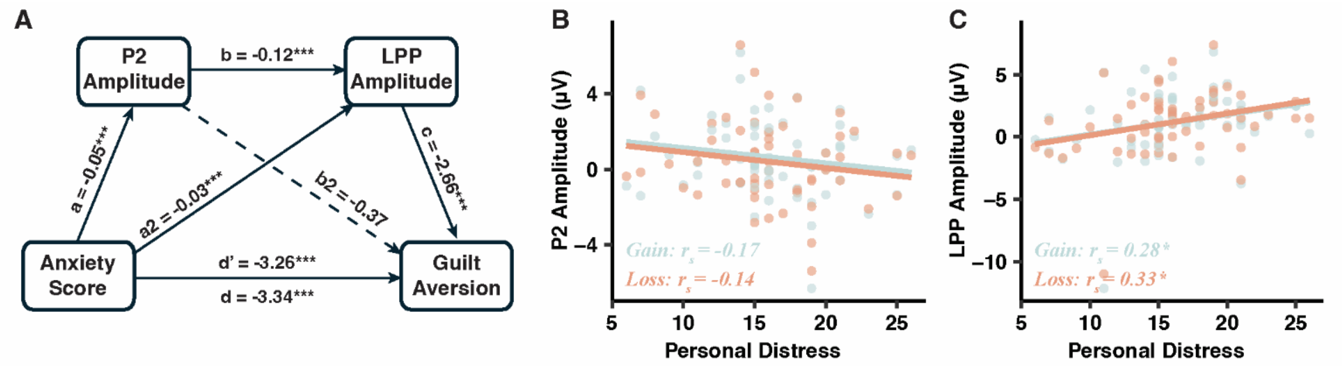


Figure 7. P2 and LPP mediated trait anxiety effects on guilt aversion and their association with personal distress. (A) The attenuating effect of trait anxiety on guilt aversion occurred partially through decreasing P2 amplitude and increasing LPP amplitude. While **(B)** Personal distress is not related to P2 amplitude, **(C)** but positively related to LPP amplitude. ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$.

Third, to better interpret the ERP components, the associations between ERP amplitudes and self-report measures like IRI (empathy) and Machiavellianism (selfishness) were examined. Mann-Whitney U tests showed that the High Anxiety group exhibited higher scores on Machiavellianism and IRI personal distress compared to the Low Anxiety group (**Tab. 3**). Further, Spearman correlation analyses indicated that Machiavellianism was not correlated with P2 amplitude (Gain: r_s = -0.14, p = 0.294; Loss: r_s = -0.074, p = 0.576) nor LPP amplitude (Gain: r_s = 0.17, p = 0.186; Loss: r_s = 0.12, p = 0.373). In contrast, personal distress was not correlated with P2 amplitude (Gain frame: r_s = -0.17, p = 0.196; Loss frame: r_s = -0.14, p = 0.302) (**Fig. 7B**), but positively associated with LPP amplitude (Gain frame: r_s = 0.28, p = 0.027; Loss frame: r_s = 0.33, p = 0.011 (**Fig. 7C**). These results suggest that LPP amplitude may reflect emotional regulation mechanisms for negative emotions, especially in individuals with high trait anxiety. Individuals who experienced higher levels of personal distress, tended to exert more effort to suppress anticipatory guilt (indicated by higher LPP amplitude) as an emotional regulation strategy to avoid further emotional burden during social interactions.

Table 3. Psychometric measures (Mean ± SE). Individuals with high trait anxiety group exhibited higher scores on Machiavellianism and IRI personal distress compared to those with low anxiety. SE: standard error of mean; *U*: Mann-Whitney *U* value; *p: p-value*.

	High trait anxiety	Low trait anxiety	*U*	*p*
Machiavellianism	79.47±2.35	71.87±1.74	604	0.023
Interpersonal Reactivity Index				
Perspective-taking	18.03±0.79	19.35±0.63	357.5	0.172
Fantasy	17.72±0.63	17.77±0.85	457.5	0.917
Empathic concern	19.19±0.64	19.39±0.57	428.5	0.755
Personal distress	17.91±0.63	14.71±0.83	636.5	0.006

Brain mechanism of trait anxiety on contextual effect

Since N2 reflects cognitive control, which is essential for contextual perception and influenced by trait anxiety on the periphery, the relationship between N2 amplitude, advantageous inequity aversion, and reward sensitivity was tested. The contextual effect of advantageous inequity aversion and reward sensitivity was calculated as the difference between loss and gain contexts. The impact of trait anxiety on the contextual effect of N2 amplitude, advantageous inequity aversion, and reward was analyzed.

Mann-Whitney U tests showed that Anxiety significantly altered the contextual effect on advantageous inequity aversion (U = 128, $p < 0.001$ (**Fig. 8A**) and N2 amplitude (U = 316, $p = 0.048$) (**Fig. 8B**) in a similar manner. Spearman correlation showed that changes in the contextual effect on advantageous inequity aversion positively correspond with changes in N2 amplitude ($r_s = 0.38$, $p = 0.003$) (**Fig. 8C**). Mann-Whitney U tests showed that Anxiety significantly altered the contextual effect on reward sensitivity (U = 227, $p < 0.001$). However, Spearman correlation indicated no significant relationship between the contextual effect of reward sensitivity and N2 amplitude ($r_s = 0.15$, $p = 0.245$). These results suggest that N2 amplitude potentially regulates the contextual perception of advantageous inequity aversion.

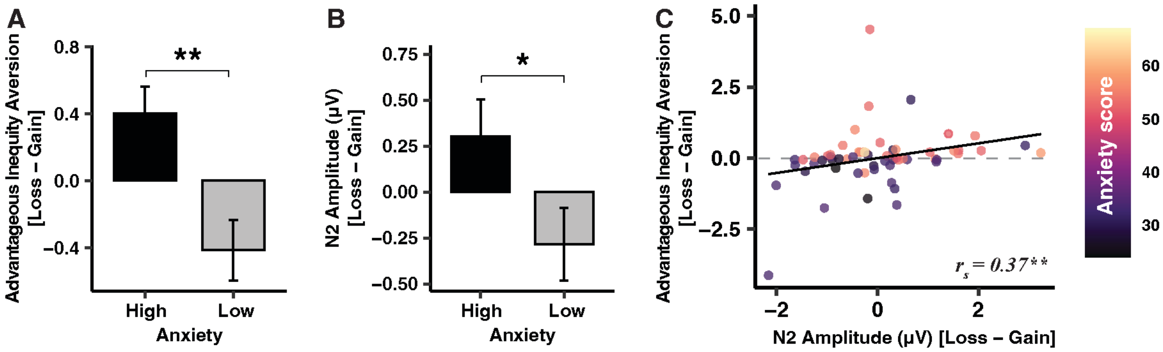


Figure 8. Alteration of trait anxiety on contextual perception of advantageous inequity aversion and N2 mechanism. (A) Alteration of trait anxiety on the contextual perception of advantageous inequity aversion. Trait anxiety significantly altered the contextual effect of advantageous inequity aversion. **(B) Alteration of trait anxiety on the contextual effect of N2 amplitude.** Similarly, trait anxiety affected N2 amplitude in a parallel pattern with advantageous inequity aversion. **(C) Association between the contextual effect of N2 amplitude and those of advantageous inequity aversion**. The contextual effect of N2 amplitude was positively correlated with the contextual effect of advantageous inequity aversion. **: $p < 0.01$; *: $p < 0.05$.

Discussion

Reciprocity is a complex social behavior influenced by both reciprocity propensity (core) and contextual perception (periphery) (Fang et al., 2022; Hagen & Hammerstein, 2006); however, their underlying neurocomputational mechanism remain unknown. In this study, we employed computational modeling, eye-tracking, and ERP within an economic binary trust game under gain and loss frame to examine how trait anxiety affects the core and periphery of reciprocity and its underlying psychological components. We identified four psychological components—validated by eye-tracking analyses—underlying reciprocity decisions: reward, guilt aversion, advantageous inequity aversion, and advantageous inequity liking. For the core of reciprocity decisions, we found that trait anxiety reduced reciprocity at the behavioral level and decreased guilt aversion and advantageous inequity liking at the psychological level across different framed contexts. At the neural level, we found that trait anxiety attenuated guilt aversion by decreasing P2 amplitude related to selective attention and increasing LPP amplitude linked to effortful emotion regulation. For the periphery of reciprocity decisions, we showed that trait anxiety altered the contextual perception of reward and advantageous inequity aversion at the psychological level, but did not affect reciprocity at the behavioral level. In particular, trait anxiety altered the contextual perception of advantageous inequity aversion which involved the N2 cognitive control mechanism at the neural level. Overall, our study added knowledge on the neurocomputational mechanisms of how trait anxiety impacts the core and periphery of reciprocity decisions.

Computational Modeling of Reciprocity Decisions

Using computational modeling, we first analyzed the psychological components of how individuals make reciprocity decisions. In line with our first hypothesis, extending previous findings (Nihonsugi et al., 2015, 2021; Xiao et al., 2022), our best model identified four psychological components of reciprocity decisions: reward, guilt aversion, advantageous inequity aversion, and advantageous inequity liking. Individuals who prioritize reward tended to reciprocate less, while those who emphasize guilt aversion, advantageous inequity aversion, and advantageous inequity liking were more likely to reciprocate. Our model indicated that advantageous inequity aversion occurs when evaluating the betrayal option, while advantageous inequity liking appears when evaluating the reciprocity option in the binary game. Although the reward in the reciprocity option is less appealing compared to the betrayal option, individuals are more likely to reciprocate due to the higher relative payoff, supporting the notion that advantageous inequity aversion is condition-dependent rather than always present (Boyce et al., 2010; Cox, 2013; Dohmen et al., 2011; O. Li et al., 2018).

Next, we validated the model's predicted psychological components for reciprocity decisions using eye-tracking. Consistent with our second hypothesis, in addition to model validation procedures through model prediction and parameter recovery, eye-tracking results also empirically validated our model. Our results showed that the transitions of the four components—guilt aversion, advantageous inequity aversion, reward, and advantageous inequity liking—dominated the top of all transitions, indicating the significant contribution of these components in reciprocity decision-making. Furthermore, higher component parameter values were correlated with increased relative fixation duration on the corresponding transitions, based on the binary trust game's payoff structure. Moreover, resembling the relationship between the four components and reciprocity rate, we found that increased fixation duration on transitions mapping guilt aversion, advantageous inequity aversion, and advantageous inequity liking correlated with a higher reciprocity rate, while increased fixation duration on transition mapping reward correlated with a lower reciprocity rate. Notably, these transitions involve specific comparisons and evaluations during decision-making as a trustee based on the payoff structure in the binary trust game (Fiedler et al., 2013; Jiang et al., 2016; Polonio et al., 2015). For reward, it involves comparing the participant's payoffs between reciprocity and betrayal options, whereas for guilt aversion, it involves comparing the partner's payoffs between these options. Advantageous inequity liking is evaluated by comparing payoffs between the participant and the partner within the reciprocity option, while advantageous inequity aversion is assessed by comparing these payoffs within the betrayal option.

Trait Anxiety's Impact on the Core of Reciprocity

To explore how trait anxiety affects reciprocity, we examined the behavioral and computational mechanisms in both gain and loss contexts. We partially confirmed our third hypothesis that trait anxiety influences reciprocity both at the behavioral and psychological levels regardless of context. At the behavioral level, consistent with previous studies (Anderl et al., 2018; Rodebaugh et al., 2011, 2013), our results indicated that individuals with high trait anxiety exhibited lower reciprocity than those with lower trait anxiety. At the psychological level, our results demonstrated that trait anxiety attenuated guilt aversion as well as advantageous inequity liking independently of context. These effects align with findings in individuals with OCPD, often comorbid with anxiety disorders (Xiao et al., 2022). Anxious individuals tend to use avoidance strategies in social interactions (Duronto et al., 2005; Raffety et al., 1997; Turner, 1988), likely reducing guilt aversion by limiting affective processing and diminishing a sense of moral obligation (Kantor, 2016; Xiao et al., 2022).

Contrary to our hypothesis, we failed to observe an impact of trait anxiety on advantageous inequity aversion as a core psychological component of reciprocity. While anxiety had a main effect on advantageous inequity aversion, the interaction between trait anxiety and frame showed that this effect depends on the context. Specifically, trait anxiety reduced reciprocity in a gain frame but had no effect in a loss frame, suggesting it impacts the periphery of reciprocity rather than its core. Although less discussed in the literature, our results showed that trait anxiety reduces sensitivity to advantageous inequity liking when evaluating reciprocity. Due to the smaller payoff difference in the reciprocity option compared to the betrayal option, individuals with high trait anxiety might overlook minor payoff differences because of impaired attention and increased distractibility (Eysenck et al., 2007; Pacheco-Unguetti et al., 2010). Alternatively, their cognitive bias toward negative outcomes might have reduced their sensitivity to positive incentives (Bar-Haim et al., 2007) like advantageous inequity liking in reciprocity decisions.

Building on our computational findings, we investigated the neural mechanisms of how trait anxiety affects reciprocity. As hypothesized, our study revealed that higher trait anxiety is associated with lower P2 and higher LPP amplitudes, regardless of context. The P2 component is linked to selective attentional allocation (Hajcak et al., 2012; Luck et al., 1994; Potts, 2004; Rey-Mermet et al., 2019), whereas the LPP component is associated with an effortful cognitive regulation over emotion (Bernat et al., 2011; Desatnik et al., 2017; Moser et al., 2014; Shafir et al., 2015), particularly during the resolution of moral conflicts (Chen et al., 2009; Zhan et al., 2018, 2020). These results suggest that individuals with high trait anxiety may allocate fewer attentional resources to evaluating moral conflicts or exert more effort to resolve or disengage from the dilemmas in reciprocity decisions.

Our mediation analysis further supports this interpretation, indicating that trait anxiety attenuated guilt aversion through diminished P2 and elevated LPP amplitudes. To further elucidate the role of these ERP components, we examined relationships between self-report measures (Machiavellianism and IRI) and P2 as well as LPP amplitudes. Both Machiavellianism and IRI personal distress were higher in individuals with high trait anxiety compared to those with low trait anxiety. Further, individuals with higher personal distress exhibited higher LPP amplitudes in both gain and loss contexts. However, no significant relationships were found between personal distress and P2 amplitudes, Machiavellianism and P2 amplitudes, or Machiavellianism and LPP amplitudes. These findings suggest that LPP reflects effortful regulation and disengagement from negative emotions (Bernat et al., 2011; Desatnik et al., 2017; Moser et al., 2014; Shafir et al., 2015), supporting the idea that individuals with higher trait anxiety tend to adopt avoidance strategies (Duronto et al., 2005; Raffety et al., 1997; Turner, 1988). Overall, anxiety's decreasing effect on P2 amplitudes indicates fewer attentional resources allocated to assessing guilt (Hajcak et al., 2012; Luck et al., 1994; Potts, 2004; Rey-Mermet et al., 2019), while anxiety's increasing effect on LPP amplitudes suggests effortful emotion regulation and disengagement from the anticipatory guilt in reciprocity decisions (Chen et al., 2009; Zhan et al., 2018, 2020).

Trait Anxiety's Impact on the Periphery of Reciprocity Decisions

To elucidate the trait anxiety's impact on the periphery of reciprocity decisions, we further investigated the behavioral and computational mechanism underlying how trait anxiety alters the contextual perception between gain and loss context. Partially supporting our fourth hypothesis, trait anxiety did not affect the behavioral aspects of reciprocity but did alter its psychological components. Our results indicated that individuals with low trait anxiety showed no contextual effect on reward, while those with high trait anxiety were more sensitive to reward under a loss than a gain frame. This finding aligns with previous evidence (Gu et al., 2017; Jepma & López-Solà, 2014; Xu et al., 2013) showing that individuals with high trait anxiety are more susceptible to contextual effects on reward. Our results further showed that trait anxiety alters the contextual effect on advantageous inequity aversion. While individuals with high trait anxiety generally exhibited lower advantageous inequity aversion, context modulated this response: a loss frame reduced advantageous inequity aversion in those with low trait anxiety but enhanced it in those with high trait anxiety. This indicates distinct context-dependent processing mechanisms between individuals with low and high trait anxiety.

Anxiety's effect on contextual perception may stem from how individuals with varying trait anxiety levels respond to different contexts. Individuals with high anxiety tend to rely more on heuristic decision-making than those with low anxiety (Jepma & López-Solà, 2014) . Further, loss frames are heuristically perceived as more harmful to others (Baron, 1995; Evans & van Beest, 2017) and more threatening to one's own reward (Xu et al., 2013) compared to gain frames. Therefore, loss framing prompts individuals with high trait anxiety to increase advantageous inequity aversion and reward sensitivity, possibly driven by the "do-no-harm" principle (Evans & van Beest, 2017) and self-protective strategies (Meleshko & Alden, 1993). For example, even with the same payoff structure, making a partner lose more than the decision-maker is seen as more harmful than making the partner gain less. Similarly, losing more of their own payoff is perceived as more harmful than gaining less. The dual increase in advantageous inequity aversion and reward sensitivity in high trait anxiety individuals reflects the complex reality of decision-making with conflicting effects of the latent psychological components. This may potentially explain the lack of a significant contextual effect on behavioral reciprocity in high-anxiety individuals, as the effect of these latent psychological components might cancel each other out. In contrast, low anxiety individuals showed higher advantageous inequity aversion under the gain framing but lower under loss framing, suggesting a shift toward self-protection over other-regarding behavior.

Looking into the neural mechanism underlying trait anxiety’s impact on the periphery of reciprocity, consistent with our fourth hypothesis, our findings indicate that trait anxiety reverses the pattern of N2 amplitude in line with advantageous inequity aversion. When moving from a gain to a loss context, individuals with low trait anxiety showed decreased advantageous inequity aversion and heightened N2 amplitude, while those with high trait anxiety exhibited increased advantageous inequity aversion and attenuated N2 amplitude. The context's influence on advantageous inequity aversion was linked to its effect on N2 amplitude. The N2 component, which emerges during decision-making involving conflict and typically indicates higher cognitive control or more effortful response inhibition (Folstein & Van Petten, 2008b; Nieuwenhuis et al., 2003), is also linked to the framing effect (Zhao et al., 2018). Our results suggest several key insights: Advantageous inequity aversion appears to be an instinctual response, with reducing it requiring cognitive control and enhanced N2 effort, while increasing it involves less cognitive effort. Further, the influence of framing on advantageous inequity aversion is likely modulated by N2 cognitive control, i.e., that perception differences in advantageous inequity aversion due to framing are shaped by the degree of cognitive control exerted during decision-making. Finally, trait anxiety likely alters how framing affects advantageous inequity aversion through mechanisms that regulate cognitive control.

Limitation and Future Direction

Several limitations should be noted in this study. Firstly, while our model identified four components influencing reciprocity in parallel, individuals may evaluate decisions hierarchically, prioritizing some components initially and others later. Future research should model the hierarchical structure of these components (Grover & Vriens, 2006). Secondly, our participants were healthy college students, not clinical patients with anxiety disorders, so generalizing to clinical populations should be done cautiously. Future studies should validate these results in clinical populations. Nonetheless, our findings could aid in diagnosing anxiety disorders. For instance, individuals showing lower guilt aversion, lower P2 amplitudes, and higher LPP amplitudes during reciprocity decisions may be at higher risk for anxiety disorders. Thirdly, we measured trait anxiety but did not assess state anxiety during the experiment. Future studies should measure state anxiety to control for this potential confound. Finally, our study combined eye-tracking and EEG measures, but eye movements can create artifacts in EEG data. We used independent component analysis to minimize these artifacts, enhancing data reliability. Future studies should further refine these methods to improve data quality. Despite these limitations, our work provides valuable insights into how anxiety impacts reciprocity decisions through the lens of eye movement, EEG, and psychological components.

Conclusions

Our study investigated how trait anxiety affects the core and periphery of reciprocity. We found that trait anxiety reduces reciprocity propensity and influences latent psychological components such as guilt and advantageous inequity liking, involving attentional allocation and emotion regulation processes. Additionally, trait anxiety alters the contextual effect of advantageous inequity aversion, involving cognitive control processes. Our findings offer insights into the neurocomputational mechanisms of trait anxiety's impact on reciprocity and may aid in the improvement of reciprocity in individuals with anxiety disorders.

Materials and Methods

Participants

A total of 550 participants completed the online version of the trait subscale of the Chinese State-Trait Anxiety Inventory (STAI) (W. Li & Qian, 1995; Speilberger et al., 1983). The sample's trait anxiety scores ranged with a mean (M) of 41.52 and a standard deviation (SD) of 9.75, with the 25th percentile at 35 and the 75th percentile at 48. Participants were classified into two groups based on their trait anxiety scores (TAS): those with TAS ≤ 35 were defined as the low trait anxiety group, and those with TAS ≥ 48 were defined as the high trait anxiety group. A target sample size of 30 participants for each trait anxiety group was determined based on prior related work (Chen et al., 2009; Zhan et al., 2018, 2020). In total, 69 participants were recruited for the study. Of the recruited participants, 34 were assigned to the low trait anxiety group (16 females; M = 31.65, SD = 2.83), and 35 were assigned to the high trait anxiety group (18 females; M = 53.57, SD = 3.12). None of the participants reported taking psychoactive medications or any history of mental disorder or brain injury. Participants who pressed the same button in 95% of the trials were considered nonresponsive to the task setting or not focused on the task, and were excluded from the analysis. As a result, 63 participants were included in the analysis, with 31 in low trait anxiety group (15 females; M = 31.65, SD = 2.85), and 32 in high trait anxiety group (16 females; M = 53.56, SD = 3.22). The study was conducted according to the Declaration of Helsinki and approved by the local Ethics Committee at Shenzhen University, China. Participants provided written informed consent prior to their involvement in the study. Compensation included a fixed attendance fee of 60 yuan (approximately $10) and a variable monetary reward based on their decisions during the game, which ranged from 40 yuan to 80 yuan (approximately $7 to $13).

Questionnaire

Participants completed two self-report questionnaires: the Interpersonal Reactivity Index (IRI) measuring four empathy subscales (perspective taking, fantasy, empathic concern, and personal distress) (Davis, 1980), and the Machiavellianism (Mach-IV) scale assessing selfish tendencies (Christie & Geis, 1970).

Experimental Procedure and Tasks

Prior to the experiment, participants underwent a comprehensive briefing session on the rules of the game. This session also included practice designed to familiarize participants with their roles and the structure of the binary trust game. Participants were informed that their partners had already participated as the first movers, and their decisions were recorded and stored in the computer system, with these decisions to be shown in the formal experiment. This setup was designed to avoid the potential confound brought by physical interaction with a partner on the measurement of reciprocity and also foster a belief among participants that they were engaging in authentic dynamics of interpersonal interaction.

During the experimental task, participants completed 240 trials in two counterbalanced sessions (Gain frame, Loss frame) to control for order effects, each with two 60-trial blocks separated by short breaks. Each trial paired participants with a new anonymous partner (icon, partially obscured name). Participants then chose to "reciprocate" or "betray" the partner's "trust." Immediate feedback revealed whether the partner (50% chance) initially "trusted" or "distrusted."

The binary trust game, utilized in this study, is an interactive economic game involving two player roles: Player A and B (Evans & van Beest, 2017). In the present study, the participant played the role of B and the partner played the role of A. The rule for the binary game was as follows: A made a first move by choosing "distrust" (the square containing the left column) or "trust" (the combined square containing the middle and right columns). If A chose "distrust", the system distributed the point directly to A (gaining [in the gain frame] or losing [in the loss frame] the points colored in blue) and to B (gaining or losing the points colored in green). In this case, B's subsequent choice did not influence the point distribution. Conversely, if A chose "trust," the distribution of points depended on B's decision. B could either "reciprocate" by selecting the square on the left inside the "trust" rectangle or "betray" by selecting the square on the right inside the same rectangle. Based on B's decision, both players would gain or lose points accordingly.

To mitigate the potential influence of decision spatial location on reciprocity choices, various versions of the binary trust game were employed. These adaptations involved altering the positions of the "trust" and "distrust" options, as well as the "reciprocate" and "betray" options, by switching their placements from left to right and vice versa. Consequently, four distinct versions of the game structure were applied. These variations were counterbalanced across participants in each trait anxiety group to ensure that any effects related to the positioning of choices were minimized, allowing for a more accurate assessment of decision-making behaviors without the bias of spatial location. Note that all four versions were standardized into one version (as displayed in **Fig. 2A**) for statistical analysis. The potential influence of the different versions was examined to ensure that the experimental effects were not due to version differences (**Supplementary**).

Trials were varied by changing the six payoff values. The payoffs structure in all the trials in both gain and loss frames has several features (**Fig. 2A**): **(1)** for A, $a2 > a1 > a3$; **(2)** for B, $b3 > b2 > b1$; and **(3)** while $a1>b1$ and $a3 < b3$ in all trials, a2 can be greater than, equal to, or less than b2 in different trials, with $a2 > b2$ in 124 trials, $a2 = b2$ in 6 trials, and $a2 < b2$ in 110 trials. Thus, rationally, to maximize the outcome, A should "trust" and expect B to "reciprocate", securing the payoff a2, which is economically the most beneficial for A. Conversely, B should consistently choose to "betray" to attain b3, which is economically the most beneficial for B.

The loss frame was constructed from the gain frame using a modified version (Evans & van Beest, 2017), balancing the value scale between frames. Unlike Evans & van Beest (2017), who used b3 (highest gain frame value) as reference, we used the sum of "betray" option values (a3 + b3) to construct the loss frame. The six loss frame values were calculated as the difference between each corresponding gain frame value and the reference value (e.g., 59 [14+45]) (**Fig. 1**). In the gain frame, participants started with 0 points and gained points throughout. In the loss frame, participants started with 15000 points and lost points based on decisions, ensuring equivalent final outcomes if the same strategies were used in both frames (Evans & van Beest, 2017). Final income was randomly determined from either the accumulated points in the gain session or the remaining points in the loss session.

Eye-Tracking Data Acquisition and Processing

Eye movements were recorded at 1000 Hz using an EyeLink 1000 Plus (SR Research Ltd., Ottawa, Ontario, Canada) with head-chin stabilization. Participants were seated 60 cm from a 1280×1024 pixel monitor and instructed to maintain fixation on a central cross. A 9-point calibration/validation was performed before each block. Fixations, saccades, and blinks were classified from raw data using EyeLink Data Viewer software (SR Research Ltd., Ottawa, Canada) with default settings. Subsequent analysis in R ("eyelinker" library) focused on fixations, excluding saccades and blinks. Large stimulus separation and high calibration accuracy allowed for generous AOI margins. Six rectangular AOIs (a1, a2, a3, b1, b2, b3; 164x144 pixels each) were aligned with the six values in the binary trust game (**Fig. 2A**). A central fixation cross AOI ("Center"; 105 x 85.5 pixels) was also defined, with remaining areas labeled "Undefined." Manual examination of fixation areas for each participant and block resulted in the exclusion of four participants due to drift outside AOIs. For each trial, fixation sequences were extracted, and unique transitions between the six value AOIs (excluding "Center" and "Undefined") were identified. These transitions, without considering direction (e.g., a2_a3 is equivalent to a3_a2), reflecting decision-maker comparisons (Devetag et al., 2016), were mapped to model components like guilt aversion (a2_a3), advantageous inequity aversion (a3_b3), reward (b2_b3), and advantageous inequity liking (a2_b2). Relative fixation time for each of the 21 possible AOI combinations was calculated by dividing fixation duration by response time, with absent transitions assigned 0. For example, in a trial with fixation sequence "center_a1_a3_a3_undefined_a2_undefined_b2_b3," only transitions a1_a3 and b2_b3 would be relevant for decision processing.

EEG Data Acquisition and Processing

EEG data was collected during the binary trust game using 64 Ag/AgCl electrodes placed according to the International 10-20 system (Brain Products GmbH). Recordings were made at 1000 Hz with a 0.01-100 Hz passband, using FCz as online reference. Electrode impedance was kept below 5 kΩ. Electrooculographic (EOG) signals were also recorded to identify and remove eye movement artifacts.

EEG data were preprocessed in EEGLAB (Delorme & Makeig, 2004), an open-source toolbox in MATLAB (The MathWorks, Inc., Natick, Massachusetts, USA). Recordings were re-referenced offline to the whole brain common average and band-pass filtered (0.1-30 Hz). Ocular artifacts were identified by their contributions to EOG channels and frontal scalp distribution, and corrected using independent component analysis (Delorme & Makeig, 2004). EEG epochs time-locked to reciprocity decision onsets were extracted using a 2,000 ms window (-500 ms to 1500 ms). Epochs with response times under 800 ms were discarded. After baseline correction using the pre-stimulus interval, epochs were visually inspected for gross movement artifacts, which were then excluded to ensure data quality.

For each participant and each trial, the mean amplitudes of P2, N2, and LPP ERP components were measured within their specific time windows and at their related electrodes. The selection of time windows and electrodes for ERP amplitude measurement was informed by previous studies and by examining the grand average ERP waveforms and scalp topographies. In particular, the P2 amplitude was measured at 110–200 ms (Boudreau et al., 2009; Fang et al., 2021; Liu et al., 2023; Potts et al., 2006), and the N2 amplitude was measured at 280–340 ms (Cavanagh & Shackman, 2015; Folstein & Van Petten, 2008a; Hao et al., 2023; McLoughlin et al., 2022) after stimulus onset at frontal electrodes (F1, Fz, and F2). The LPP amplitude was measured at parietal electrodes (P1, Pz, and P2) 450–800 ms after stimulus onset (Bernat et al., 2011; Desatnik et al., 2017; Moser et al., 2014; Shafir et al., 2015). Scalp topographies of these ERP components were computed by spline interpolation.

Computational Modeling of Reciprocity Decision-Making

A stage-wise model construction procedure was utilized to identify and quantify the latent psychological components influencing reciprocity behavior in our binary trust game (Gagne et al., 2020; Wang et al., 2023; Zhang & Gläscher, 2020). This iterative approach involved sequentially refining the model based on the performance of the previous best-fitting model. Leave-One-Out Information Criterion (LOOIC) and Widely Applicable Information Criterion (WAIC) were used for model comparison to minimize overfitting, with lowest values indicating best fit to the data (Gagne et al., 2020; Wang et al., 2023; Zhang & Gläscher, 2020). The goodness of fit was assessed using Tjur's pseudo R² (Tjur, 2009), with higher values indicating better model fit. Parameters were estimated using hierarchical Bayesian analysis (Gagne et al., 2020; Wang et al., 2023; Zhang & Gläscher, 2020). The posterior inference was conducted via Markov chain Monte Carlo (MCMC) sampling, utilizing four independent chains, each with 4,000 iterations, to draw samples from the posterior distribution and ensure robust parameter estimation (Wang et al., 2023). Seven plausible candidate models were tested in total using Rstan in R (Carpenter et al., 2017) for model-related procedures.

Guided by prior research (Nihonsugi et al., 2015; Xiao et al., 2022), the baseline model (**M1**) incorporated guilt aversion, inequity aversion, and reward, positing that decisions arise from the interplay between these socio-emotional factors and potential gains. The utility function (U) was formulated as **Eq. 1:**

$$U = \begin{cases} b_3 - \beta_G \cdot (a_2 - a_3) - \beta_I \cdot (b_3 - a_3), & \textit{if betray} \\ b_2 - \beta_I \cdot |b_2 - a_2| & \textit{, if reciprocate} \end{cases} \qquad \text{M1 (1)}$$

Departing from previous studies (Nihonsugi et al., 2015; Xiao et al., 2022), participants were not informed of their partner's beliefs regarding reciprocity, thus simulating real-world social interactions where individuals often lack explicit knowledge of others' internal beliefs. The term $(a_2 - a_3)$ approximates the anticipated guilt experienced upon choosing betrayal (where the partner receives a_3), considering the partner's assumed trust and expectation of reciprocity (anticipating a_2), with β_G ($0< \beta_G <10$) capturing the participant's aversion to guilt. The terms $(b_3 - a_3)$ and $|b_2 - a_2|$ quantify the aversion to inequity associated with betrayal and reciprocity, with β_I ($0< \beta_I <10$) representing the participant's aversion to inequity. By design of the binary trust game, what the participant receives (b_3) is constantly larger than what the partner receives (a_3) in the option of betrayal, whereas the participant receives (b_2) can be either greater or less than what the partner receives (a_2) in the option of reciprocity. In the baseline model **M1**, inequity was quantified as the absolute difference between b_2 and a_2 in the option of reciprocity, assuming that participants equally weigh both advantageous and disadvantageous inequity aversions (Nihonsugi et al., 2015; Xiao et al., 2022). Finally, the utility (U) of choosing to reciprocate and betray was entered into a SoftMax function with an inverse temperature parameter λ ($0< \lambda <10$), such that the probability of choosing to reciprocate in each trial was expressed as (**Eq.2**):

$$P(reciprocate) = \frac{1}{1 + e^{-\lambda(U(reciprocate) - U(betray))}} \qquad (2)$$

The initial model, **M1**, expanded upon the established baseline model by incorporating varying sensitivities to guilt and inequity aversion across the four experimental conditions, namely low/high trait anxiety and gain/loss framing. **M2** (**Eq. 3**) employed the Fehr–Schmidt inequity aversion model (Fehr & Schmidt, 1999), which separated inequity aversion into advantageous inequity aversion (advantageous inequity aversion) and disadvantageous inequity aversion.

$$U = \begin{cases} b_3 - \beta_G \cdot (a_2 - a_3) - \beta_{Ad-IA} \cdot (b_3 - a_3) & , \textit{if betray} \\ b_2 - p \cdot \beta_{Ad-IA} \cdot (b_2 - a_2) - q \cdot \beta_{DisAd-IA}(a_2 - b_2) & , \textit{if reciprocate} \end{cases} \qquad \text{M2 (3)}$$

Here, β_{Ad-IA} and $\beta_{DisAd-IA}$ represents the participant's subjective aversion to advantageous and disadvantageous inequity. p and q are conditional indicators: if $b_2 > a_2$, $p = 1$, $q = 0$; if $a_2 > b_2$, $p = 0$, $q = 1$. Given that **M2** outperformed **M1**, **M3** was developed by incorporating a reward parameter (**Eq. 4**) into **M2**, resulting in further improved model performance.

$$U = \begin{cases} \beta_R \cdot b_3 - \beta_G \cdot (a_2 - a_3) - \beta_{Ad-IA} \cdot (b_3 - a_3) & , \textit{if betray} \\ \beta_R \cdot b_2 - p \cdot \beta_{Ad-IA} \cdot (b_2 - a_2) - q \cdot \beta_{DisAd-IA}(a_2 - b_2) & , \textit{if reciprocate} \end{cases} \qquad \text{M3 (4)}$$

Building upon **M3**, **M4** assumes that participants exhibit a dislike for advantageous inequity in betrayal scenarios yet appreciate it in reciprocity scenarios (**Eq. 5**), leading to the replacement of the inequity aversion term in the reciprocity option with an advantageous inequity liking component.

$$U = \begin{cases} \beta_R \cdot b_3 - \beta_G \cdot (a_2 - a_3) - \beta_{Ad-IA} \cdot (b_3 - a_3), & \textit{if betray} \\ \beta_R \cdot b_2 + \beta_{Ad-IL} \cdot (b_2 - a_2) & , \textit{if reciprocate} \end{cases} \qquad \text{M4 (5)}$$

Here, the term $(b_2 - a_2)$ represents the magnitude and direction of the objective advantageous inequity (positive or negative), while and β_{Ad-IL} reflected the participant's sensitivity to this advantageous inequity liking. The model comparison revealed that **M4** outperformed **M3**. Building upon **M4**, **M5** introduces a reward weighting parameter against other components ($0 < \beta_R < 1$) (**Eq. 6**), while **M6** assumes individuals hold a consistent trade-off between randomness and determinism in decision-making, employing a shared inverse temperature parameter (λ) across frames, and **M7** replaces λ with constant 1 as previous suggested (Nihonsugi et al., 2021). Ultimately, model comparison confirmed **M4** as the winning model.

$$U = \begin{cases} \beta_R \cdot (b_3 - b_2) & , \textit{if betray} \\ (1 - \beta_R) \cdot \{\beta_{Ad-IL} \cdot (b_2 - a_2) + \beta_G \cdot (a_2 - a_3) + \beta_{Ad-IA} \cdot (b_3 - a_3)\} & , \textit{if reciprocate} \end{cases} \qquad \text{M5 (6)}$$

M4 was validated using posterior predictive checks and parameter recovery (Wilson & Collins, 2019). For each participant and frame, model parameters were drawn and averaged from the posterior distribution to simulate decision choices and compute a predicted reciprocity rate. The correlation between these predicted and observed reciprocity rates was then used to assess the model's predictive accuracy. Parameter recovery was conducted by refitting **M4** to the simulated dataset and assessing the correlation between the original and recovered parameters, thus validating the model's fitting precision.

Statistical Analysis

All statistical analyses were performed in R (v4.1.1; www.r-project.org). LMM was performed with the lme4 package (Bates et al., 2015). Results were considered statistically significant at the statistical threshold level $p < .05$ (two-tailed). LMM analyses were used to examine the effects of trait anxiety and context (frame) on reciprocity rate, guilt aversion, advantageous inequity aversion, reward sensitivity, and advantageous inequity liking, as derived from the winning model. Fixed effects for trait anxiety, frame, and their interaction were included, along with subject-specific random intercepts. Differences in IRI subscales and Machiavellianism between high and low trait anxiety groups were assessed using the Mann-Whitney U test, chosen for its robustness to potential outliers and non-normal distributions often observed in these psychological measures.

To assess the contribution of each winning model component to reciprocity rate variance, a linear model was employed. Component weights were calculated using the "relaimpo" library in R. To validate the model, LMMs were used to examine the relationship between individual sensitivity to each model component and corresponding eye movement transitions. The LMM included fixed effects for individual component sensitivity, along with subject-specific random intercepts. To control for potential confounding factors, the objective value of the component, frame (gain/loss), and reaction time were included as fixed effects.

To investigate the effects of trait anxiety and frame on ERP P2, N2, and LPP amplitudes, LMMs with trial-level data were employed. The models included fixed effects for trait anxiety, frame, and their interaction, as well as subject-specific random intercepts and slopes for frame. To examine the mediating roles of P2 and LPP amplitudes, serial mediation analysis was performed using the R library "bruceR" (Bao, 2023), an adaptation of SPSS's "PROCESS" (Hayes, 2017). Trial-level P2 and LPP amplitudes were used, along with trial-level guilt aversion and advantageous inequity aversion, calculated as the subjective value (β_G, β_{SAv}) multiplied by the corresponding objective value (a2-a3 for guilt aversion, b2-a2 for advantageous inequity aversion) (Wang et al., 2023).

To investigate the impacts of trait anxiety on the contextual effect of advantageous inequity aversion, reward sensitivity, and N2 amplitude, Mann-Whitney U tests were conducted. To assess the association between the contextual effect of advantageous inequity aversion and N2 amplitude, Spearman correlation analyses were implemented.

Additional information

Funding

This study was supported by the National Natural Science Foundation of China (31920103009 to YL, 31871137 to PX), the Major Project of National Social Science Foundation (20&ZD153 to YL), Young Elite Scientists Sponsorship Program by China Association for Science and Technology (YESS20180158 to PX), Shenzhen-Hong Kong Institute of Brain Science Shenzhen Fundamental Research Institutions (2022SHIBS0003 to YL), and Shenzhen Science and Technology Research Funding Program (JCYJ20180507183500566 to PX). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceptualization: Huihua Fang.

Data curation: Rong Wang.

Data arrangement: Rong Wang.

Formal analysis: Huihua Fang.

Funding acquisition: Yuejia Luo.

Investigation: Huihua Fang.

Methodology: Huihua Fang, Zhihao Wang.

Project administration: Huihua Fang, Rong Wang, Qian Liu.

Supervision: Frank Krueger, Pengfei Xu, Yuejia Luo.

Validation: Huihua Fang, Frank Krueger.

Visualization: Huihua Fang.

Writing – original draft: Huihua Fang, Frank Krueger.

Writing – review & editing: All authors contributed to the final version.

Data Availability Statement

The dataset and code for this manuscript are accessible on the Open Science Framework (OSF) at (link)

Reference

Abric, J.-C. (1993). Central system, peripheral system: Their functions and roles in the dynamics of social representations. *Papers on Social Representations*, *2*, 75–78.

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237. https://doi.org/10.1016/j.cpr.2009.11.004

Anderl, C., Steil, R., Hahn, T., Hitzeroth, P., Reif, A., & Windmann, S. (2018). Reduced reciprocal giving in social anxiety – evidence from the trust game. *Journal of Behavior Therapy and Experimental Psychiatry*, *59*, 12–18. https://doi.org/10.1016/j.jbtep.2017.10.005

Bao, H.-W.-S. (2023). *bruceR: Broadly Useful Convenient and Efficient R Functions*. https://psychbruce.github.io/bruceR/

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin*, *133*(1), 1–24. https://doi.org/10.1037/0033-2909.133.1.1

Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making*, *8*(2), 71–83. https://doi.org/10.1002/bdm.3960080202

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package ‘lme4.’ *Convergence*, *12*(1), 2.

Bernat, E. M., Cadwallader, M., Seo, D., Vizueta, N., & Patrick, C. J. (2011). Effects of instructed emotion regulation on valence, arousal, and attentional measures of affective processing. *Developmental Neuropsychology*, *36*(4), 493–518. https://doi.org/10.1080/87565641.2010.549881

Boudreau, C., McCubbins, M. D., & Coulson, S. (2009). Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. *Social Cognitive and Affective Neuroscience*, *4*(1), 23–34. https://doi.org/10.1093/scan/nsn034

Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science*, *21*(4), 471–475.

Campbell-Sills, L., Simmons, A. N., Lovero, K. L., Rochlin, A. A., Paulus, M. P., & Stein, M. B. (2011). Functioning of neural systems supporting emotion regulation in anxiety-prone individuals. *NeuroImage*, *54*(1), 689–696. https://doi.org/10.1016/j.neuroimage.2010.07.041

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1–32. https://doi.org/10.18637/jss.v076.i01

Cavanagh, J. F., & Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. *Journal of Physiology-Paris*, *109*(1), 3–15. https://doi.org/10.1016/j.jphysparis.2014.04.003

Chen, P., Qiu, J., Li, H., & Zhang, Q. (2009). Spatiotemporal cortical activation underlying dilemma decision-making: An event-related potential study. *Biological Psychology*, *82*(2), 111–115. https://doi.org/10.1016/j.biopsycho.2009.06.007

Christie, R., & Geis, F. L. (1970). *Studies in machiavellianism*. New York, NY: Academic Press.

Collier Villaume, S., Chen, S., & Adam, E. K. (2023). Age disparities in prevalence of anxiety and depression among US adults during the COVID-19 pandemic. *JAMA Network Open*, *6*(11), e2345073–e2345073. https://doi.org/10.1001/jamanetworkopen.2023.45073

Cox, C. A. (2013). Inequity aversion and advantage seeking with asymmetric competition. *Journal of Economic Behavior & Organization*, *86*, 121–136. https://doi.org/10.1016/j.jebo.2012.12.020

Davis, M. H. (1980). *A multidimensional approach to individual differences in empathy*.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desatnik, A., Bel-Bahar, T., Nolte, T., Crowley, M., Fonagy, P., & Fearon, P. (2017). Emotion regulation in adolescents: An ERP study. *Biological Psychology*, *129*, 52–61. https://doi.org/10.1016/j.biopsycho.2017.08.001

Devetag, G., Di Guida, S., & Polonio, L. (2016). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics*, *19*(1), 177–201. https://doi.org/10.1007/s10683-015-9432-5

Dohmen, T., Falk, A., Fliessbach, K., Sunde, U., & Weber, B. (2011). Relative versus absolute income, joy of winning, and gender: Brain imaging evidence. *Journal of Public Economics*, *95*(3), 279–285. https://doi.org/10.1016/j.jpubeco.2010.11.025

Duronto, P. M., Nishida, T., & Nakayama, S. (2005). Uncertainty, anxiety, and avoidance in communication with strangers. *International Journal of Intercultural Relations*, *29*(5), 549–560. https://doi.org/10.1016/j.ijintrel.2005.08.003

Evans, A. M., & van Beest, I. (2017). Gain-loss framing effects in dilemmas of trust and reciprocity. *Journal of Experimental Social Psychology*, *73*(July), 151–163. https://doi.org/10.1016/j.jesp.2017.06.012

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353. https://doi.org/10.1037/1528-3542.7.2.336

Fang, H., Li, X., Zhang, W., Fan, B., Wu, Y., & Peng, W. (2021). Single dose testosterone administration enhances novelty responsiveness and short-term habituation in healthy males. *Hormones and Behavior*, *131*(March), 104963. https://doi.org/10.1016/j.yhbeh.2021.104963

Fang, H., Liao, C., Fu, Z., Tian, S., Luo, Y., Xu, P., & Krueger, F. (2022). Connectome-based individualized prediction of reciprocity propensity and sensitivity to framing: A resting-state functional magnetic resonance imaging study. *Cerebral Cortex*, *33*(6), 3193–3206. https://doi.org/10.1093/cercor/bhac269

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social value orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes*, *120*(2), 272–284. https://doi.org/10.1016/j.obhdp.2012.07.002

Fiske, S. T. (2011). *Envy up, scorn down: How status divides us*. Russell Sage Foundation.

Folstein, J. R., & Van Petten, C. (2008a). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Folstein, J. R., & Van Petten, C. (2008b). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Gagne, C., Zika, O., Dayan, P., & Bishop, S. J. (2020). Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife*, *9*, e61387. https://doi.org/10.7554/eLife.61387

Goldin, P. R., Manber-Ball, T., Werner, K., Heimberg, R., & Gross, J. J. (2009). Neural mechanisms of cognitive reappraisal of negative self-beliefs in social anxiety disorder. *Biological Psychiatry*, *66*(12), 1091–1099. https://doi.org/10.1016/j.biopsych.2009.07.014

Grover, R., & Vriens, M. (2006). *The handbook of marketing research: Uses, misuses, and future advances*. Sage.

Gu, R., Wu, R., Broster, L. S., Jiang, Y., Xu, R., Yang, Q., Xu, P., & Luo, Y.-J. (2017). Trait anxiety and economic risk avoidance are not necessarily associated: Evidence from the framing effect. *Frontiers in Psychology*, *8*, 92. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00092

Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology*, *69*(3), 339–348. https://doi.org/10.1016/j.tpb.2005.09.005

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology*, *35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hajcak, G., Weinberg, A., MacNamara, A., Foti, D., & others. (2012). ERPs and the study of emotion. *The Oxford Handbook of Event-Related Potential Components*, *441*, 474.

Hao, S., Xin, Q., Xiaomin, Z., Jiali, P., Xiaoqin, W., Rong, Y., & Cenlin, Z. (2023). Group membership modulates the hold-up problem: An event-related potentials and oscillations study. *Social Cognitive and Affective Neuroscience*, *18*(1), nsad071. https://doi.org/10.1093/scan/nsad071

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Jepma, M., & López-Solà, M. (2014). Anxiety and framing effects on decision making: Insights from neuroimaging. *Journal of Neuroscience*, *34*(10), 3455–3456. https://doi.org/10.1523/JNEUROSCI.5352-13.2014

Jiang, T., Potters, J., & Funaki, Y. (2016). Eye-tracking social preferences. *Journal of Behavioral Decision Making*, *29*(2–3), 157–168. https://doi.org/10.1002/bdm.1899

Kantor, M. (2016). *Obsessive-compulsive personality disorder: Understanding the overly rigid, controlling person*. Bloomsbury Publishing USA.

Li, O., Xu, F., & Wang, L. (2018). Advantageous inequity aversion does not always exist: The role of determining allocations modulates preferences for advantageous inequity. *Frontiers in Psychology*, *9*, 749. https://doi.org/10.3389/fpsyg.2018.00749

Li, W., & Qian, M. (1995). Revision of the state-trait anxiety inventory with sample of chinese college students. *Chinese Science Abstracts Series B*, *3 Part B*(14), 50.

Liu, S., Duan, M., Sun, Y., Wang, L., An, L., & Ming, D. (2023). Neural responses to social decision-making in suicide attempters with mental disorders. *BMC Psychiatry*, *23*(1), 19. https://doi.org/10.1186/s12888-022-04422-z

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 887–904. https://doi.org/10.1037/0096-1523.20.4.887

MacNamara, A., & Proudfit, G. H. (2014). Cognitive load and emotional processing in generalized anxiety disorder: Electrocortical evidence for increased distractibility. *Journal of Abnormal Psychology*, *123*(3), 557–565. https://doi.org/10.1037/a0036997

McLoughlin, G., Gyurkovics, M., Palmer, J., & Makeig, S. (2022). Midfrontal theta activity in psychiatric illness: An index of cognitive vulnerabilities across disorders. *Biological Psychiatry*, *91*(2), 173–182. https://doi.org/10.1016/j.biopsych.2021.08.020

Meleshko, K. G., & Alden, L. E. (1993). Anxiety and self-disclosure: Toward a motivational model. *Journal of Personality and Social Psychology*, *64*(6), 1000–1009. https://doi.org/10.1037/0022-3514.64.6.1000

Moser, J. S., Hartwig, R., Moran, T. P., Jendrusina, A. A., & Kross, E. (2014). Neural markers of positive reappraisal and their associations with trait reappraisal and worry. *Journal of Abnormal Psychology*, *123*(1), 91–105. https://doi.org/10.1037/a0035817

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(1), 17–26. https://doi.org/10.3758/CABN.3.1.17

Nihonsugi, T., Ihara, A., & Haruno, M. (2015). Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. *Journal of Neuroscience*, *35*(8), 3412–3419. https://doi.org/10.1523/JNEUROSCI.3885-14.2015

Nihonsugi, T., Numano, S., & Haruno, M. (2021). Functional connectivity basis and underlying cognitive mechanisms for gender differences in guilt aversion. *eNeuro*, *8*(6). https://doi.org/10.1523/ENEURO.0226-21.2021

Pacheco-Unguetti, A. P., Acosta, A., Callejas, A., & Lupiáñez, J. (2010). Attention and anxiety: Different attentional functioning under state and trait anxiety. *Psychological Science*, *21*(2), 298–304. https://doi.org/10.1177/0956797609359624

Paul, S., Simon, D., Endrass, T., & Kathmann, N. (2016). Altered emotion regulation in obsessive-compulsive disorder as evidenced by the late positive potential. *Psychological Medicine*, *46*(1), 137–147. https://doi.org/10.1017/S0033291715001610

Polonio, L., Di Guida, S., & Coricelli, G. (2015). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior*, *94*, 80–96. https://doi.org/10.1016/j.geb.2015.09.003

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition*, *56*(1), 5–13. https://doi.org/10.1016/j.bandc.2004.03.006

Potts, G. F., Martin, L. E., Burton, P., & Montague, P. R. (2006). When things are better or worse than expected: The medial frontal cortex and the allocation of processing resources. *Journal of Cognitive Neuroscience*, *18*(7), 1112–1119. https://doi.org/10.1162/jocn.2006.18.7.1112

Power, J. D., & Petersen, S. E. (2013). Control-related systems in the human brain. *Current Opinion in Neurobiology*, *23*(2), 223–228. https://doi.org/10.1016/j.conb.2012.12.009

Qi, S., Luo, Y., Tang, X., Li, Y., Zeng, Q., Duan, H., Li, H., & Hu, W. (2016). The temporal dynamics of directed reappraisal in high-trait-anxious individuals. *Emotion (Washington, D.C.)*, *16*(6), 886–896. https://doi.org/10.1037/emo0000186

Raffety, B. D., Smith, R. E., & Ptacek, J. T. (1997). Facilitating and debilitating trait anxiety, situational anxiety, and coping with an anticipated stressor: A process analysis. *Journal of Personality and Social Psychology*, *72*(4), 892–906. https://doi.org/10.1037/0022-3514.72.4.892

Rey-Mermet, A., Gade, M., & Steinhauser, M. (2019). Sequential conflict resolution under multiple concurrent conflicts: An ERP study. *NeuroImage*, *188*, 411–418. https://doi.org/10.1016/j.neuroimage.2018.12.031

Rodebaugh, T. L., Heimberg, R. G., Taylor, K. P., & Lenze, E. J. (2016). Clarifying the behavioral economics of social anxiety disorder: Effects of interpersonal problems and symptom severity on generosity. *Clinical Psychological Science*, *4*(1), 107–121. https://doi.org/10.1177/2167702615578128

Rodebaugh, T. L., Klein, S. R., Yarkoni, T., & Langer, J. K. (2011). Measuring social anxiety related interpersonal constraint with the flexible iterated prisoner's dilemma. *Journal of Anxiety Disorders*, *25*(3), 427–436. https://doi.org/10.1016/j.janxdis.2010.11.006

Rodebaugh, T. L., Shumaker, E. A., Levinson, C. A., Fernandez, K. C., Langer, J. K., Lim, M. H., & Yarkoni, T. (2013). Interpersonal constraint conferred by generalized social anxiety disorder is evident on a behavioral economics task. *Journal of Abnormal Psychology*, *122*(1), 39–44. https://doi.org/10.1037/a0030975

Schmid, L., Chatterjee, K., Hilbe, C., & Nowak, M. A. (2021). A unified framework of direct and indirect reciprocity. *Nature Human Behaviour*, *5*(10), 1292–1302. https://doi.org/10.1038/s41562-021-01114-8

Shafir, R., Schwartz, N., Blechert, J., & Sheppes, G. (2015). Emotional intensity influences pre-implementation and implementation of distraction and reappraisal. *Social Cognitive and Affective Neuroscience*, *10*(10), 1329–1337. https://doi.org/10.1093/scan/nsv022

Speilberger, C. D., Gorsuch, Rl., Lushene, R., Vagg, P., & Jacobs, G. (1983). Manual for the state-trait anxiety inventory. *Palo Alto, CA: Consulting Psychologists*.

Starmans, C., Sheskin, M., & Bloom, P. (2017). Why people prefer unequal societies. *Nature Human Behaviour*, *1*(4), 1–7. https://doi.org/10.1038/s41562-017-0082

Thiruchselvam, R., Blechert, J., Sheppes, G., Rydstrom, A., & Gross, J. J. (2011). The temporal dynamics of emotion regulation: An EEG study of distraction and reappraisal. *Biological Psychology*, *87*(1), 84–92. https://doi.org/10.1016/j.biopsycho.2011.02.009

Tjur, T. (2009). Coefficients of determination in logistic regression models—a new proposal: The coefficient of discrimination. *The American Statistician*, *63*(4), 366–372. https://doi.org/10.1198/tast.2009.08210

Turner, J. H. (1988). *A Theory of Social Interaction*. Stanford University Press.

Wagenaar, W. A., Keren, G., & Lichtenstein, S. (1988). Islanders and hostages: Deep and surface structures of decision problems. *Acta Psychologica*, *67*(2), 175–189. https://doi.org/10.1016/0001-6918(88)90012-1

Walters, K. S., & Hope, D. A. (1998). Analysis of social behavior in individuals with social phobia and nonanxious participants using a psychobiological model. *Behavior Therapy*, *29*(3), 387–407. https://doi.org/10.1016/S0005-7894(98)80039-7

Wang, Z., Nan, T., Goerlich, K. S., Li, Y., Aleman, A., Luo, Y., & Xu, P. (2023). Neurocomputational mechanisms underlying fear-biased adaptation learning in changing environments. *PLOS Biology*, *21*(5), e3001724. https://doi.org/10.1371/journal.pbio.3001724

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Xiao, F., Zhao, J., Fan, L., Ji, X., Fang, S., Zhang, P., Kong, X., Liu, Q., Yu, H., Zhou, X., Gao, X., & Wang, X. (2022). Understanding guilt-related interpersonal dysfunction in obsessive-compulsive personality disorder through computational modeling of two social interaction tasks. *Psychological Medicine*, *53*(12), 5569--5581. https://doi.org/10.1017/S003329172200277X

Xu, P., Gu, R., Broster, L. S., Wu, R., Van Dam, N. T., Jiang, Y., Fan, J., & Luo, Y. (2013). Neural basis of emotional decision making in trait anxiety. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(47), 18641–18653. https://doi.org/10.1523/JNEUROSCI.1253-13.2013

Zhan, Y., Xiao, X., Li, J., Liu, L., Chen, J., Fan, W., & Zhong, Y. (2018). Interpersonal relationship modulates the behavioral and neural responses during moral decision-making. *Neuroscience Letters*, *672*, 15–21. https://doi.org/10.1016/j.neulet.2018.02.039

Zhan, Y., Xiao, X., Tan, Q., Li, J., Fan, W., Chen, J., & Zhong, Y. (2020). Neural correlations of the influence of self-relevance on moral decision-making involving a trade-off between harm and reward. *Psychophysiology*, *57*(9), e13590. https://doi.org/10.1111/psyp.13590

Zhang, L., & Gläscher, J. (2020). A brain network supporting social influences in human decision-making. *Science Advances*, *6*(34), eabb4159. https://doi.org/10.1126/sciadv.abb4159

Zhao, L., Shi, Z., Zheng, Q., Chu, H., Xu, L., & Hu, F. (2018). Use of electroencephalography for the study of gain–loss asymmetry in intertemporal decision-making. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00984

Supplementary

Reaction time

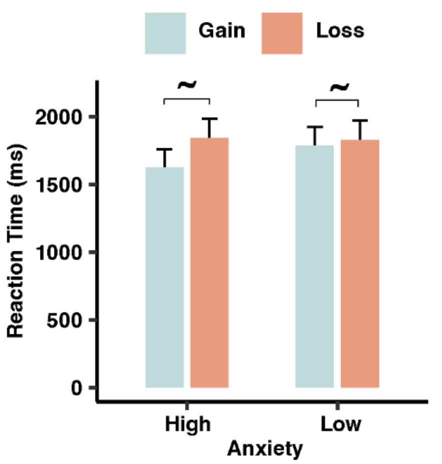


Figure S1. The impact of trait anxiety and context on reaction time in reciprocity decision (Mean ± SE). Individuals under Loss Frame exhibited a trend of longer reaction time than those under Gain Frame. ~: $p < 0.1$.

Model prediction

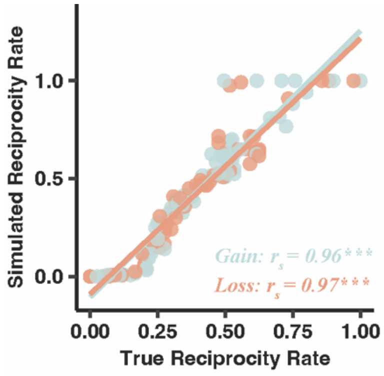


Figure S2. Model prediction from winning model M4. (A) True reciprocity rate was correlated with simulated reciprocity rate in both Gain and Loss frame. ***: $p < 0.001$.

Parameter recovery

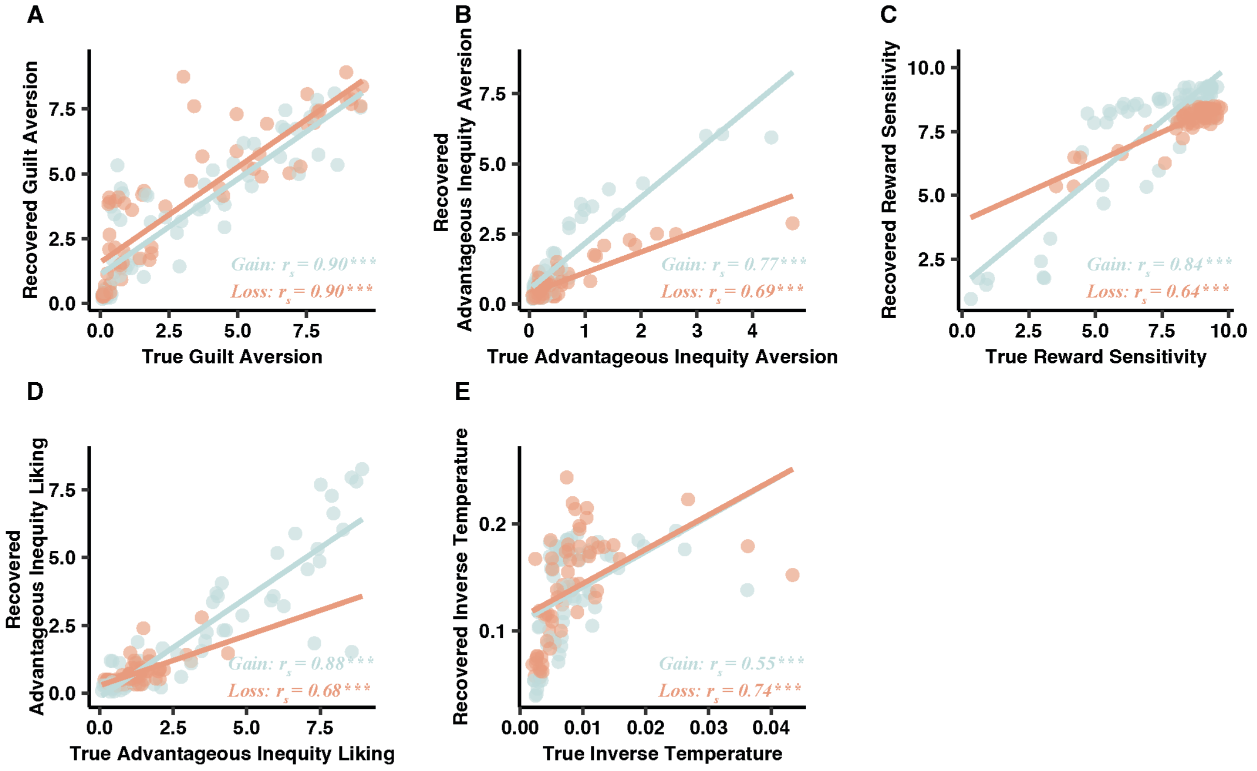


Figure S3. Parameter recovery from winning model M4. True parameters of **(A)** guilt aversion, **(B)** advantageous inequity aversion, **(C)** reward sensitivity, **(D)** advantageous inequity liking, and **(E)** inverse temperature were correlated with recovered parameters in both Gain and Loss frame. ***: $p < 0.001$.

Displaying version effect

The LMMs revealed no significant main effect of displaying version on the recirpocity rate ($\chi^2(1) = 1.64$, $p = 0.650$) and parameters of guilt aversion ($\chi^2(1) = 0.52$, $p = 0.915$), advantageous inequity aversion ($\chi^2(1) = 2.62$, $p = 0.453$), reward sensitivity ($\chi^2(1) = 2.73$, $p = 0.435$), advantageous inequity liking ($\chi^2(1) = 3.16$, $p = 0.368$) and inverse temperature ($\chi^2(1) = 0.758$, $p = 0.860$) from the winning model **M4**.

6. General Discussion

The aim of this dissertation was to investigate the core and periphery of reciprocity within the framework of social representation theory and explore how anxiety influenced reciprocity through these components. The dissertation integrated a multi-modal investigation across three studies, including neuroimaging (EEG, fMRI), eye-tracking, and computational modeling techniques, to comprehensively unveil the complexity of reciprocity. The collective findings provide an understanding of the neural (both brain network- and region-level localization and temporal dynamics) and computational (psychological) mechanisms underlying the core and periphery of reciprocity. The dissertation also elucidated the neurocomputational mechanism underlying how anxiety modulates the core and periphery of reciprocity.

Neural network mechanism for the core and periphery of reciprocity

The aim of Study 1 was to identify the neural networks representing the core and periphery mechanisms of reciprocity. Combining task-free fMRI with CPM in economic games (the one-shot TG [“give” frame] and DTG [“take” frame]), the study elucidated the core and periphery in reciprocity. Regarding the core, it showed that reciprocity under the “give” and “take” frames was positively correlated. CPM results highlighted the contribution of inter-network RSFC between the DMN (associated with mentalizing) and CON (associated with cognitive control) in representing the core of reciprocity. Regarding the periphery, the study showed that the reciprocity rate was higher under the “give” frame than the “take” frame. CPM results indicated the intra-network connectivity of DMN (associated with mentalizing) contributes to the periphery of reciprocity.

Previous studies have elucidated the neural mechanism of reciprocity (Bellucci et al., 2019; Cáceda et al., 2015); however, the effect of context in the measured decision is often neglected. Utilizing a well-controlled framing technique, Study 1 advanced our understanding of reciprocity by considering both the individual propensity for reciprocity and the perception of peripheral context. Previous research has shown that the DMN, CON, and FPN alone can predict reciprocity, underscoring their vital role in resolving the social dilemma of reciprocity (Bellucci et al., 2019). Study 1 extended and refined this by focusing on the core of reciprocity, emphasizing the role of the inter-network RSFC of DMN-CON, which explains the interplay between mentalizing ability for inference of intention and cognitive control for the temptation to betrayal in resolving the dilemma. Consistent with previous findings that functional connectivity within DMN is associated with social framing effects (help vs. harm frame) (Liu et al., 2020). Study 1 also highlights the important role of DMN in the periphery of reciprocity. This aligns with the interpretation that

DMN is the hub of the social brain, which is essential for assessing social contexts (Krueger et al., 2009; Mars et al., 2012). These findings supported social representation theory, providing an empirical behavioral and neural basis for understanding the core and periphery of reciprocity. Although focuses on reciprocity, the revelation of neural bases for these structures can inspire future research to account for the core and periphery in other decision-making domains.

Neurocomputational mechanism for the periphery of reciprocity

The aim of Study 2 was to specifically investigate the neurocomputational mechanism underlying how peripheral manipulation shapes reciprocity. Utilizing task-based fMRI and computational modeling within a two-stage binary TG (framed as either gain or loss contexts), the study provided insights into the neurocomputational mechanism at the brain region level. The results showed that the peripheral manipulation of context influenced reciprocity through advantageous inequity aversion (discomfort associated with receiving more than partners). Neural results highlighted the role of the right amygdala and left anterior insula (lAI) in the periphery of reciprocity. The results also emphasized distinct contributions in the subprocesses (other-oriented inference and self-oriented evaluation) of reciprocity decision-making for the periphery. In overall reciprocity decision-making, right amygdala activity was negatively associated with advantageous inequity aversion in the gain frame only. For the other-oriented inference process, although rDLPFC, DMPFC, and lSMG activity are associated with advantageous inequity aversion, no peripheral effect-related region was found. For the self-oriented evaluation process, reduced lAI association was observed in the loss frame compared to the gain frame, and it was positively associated with advantageous inequity aversion exclusively in the gain frame.

The peripheral effect of advantageous inequity aversion has been demonstrated in a previous study employing stimuli with immediate biological relevance (e.g., pain), showing the involvement of the rDLPFC, lAI, and DMPFC (Xiaoxue Gao et al., 2018). Study 2 is consistent with this prior study, highlighting the important role of advantageous inequity aversion for the peripheral effect, while other components, such as guilt aversion, advantageous inequity liking, and reward sensitivity, do not. In addition, while confirming the important role of the involvement of lAI for the peripheral effect, Study 1 further elucidated that this effect specifically lies in the self-oriented evaluation. Contrary to previous findings (Xiaoxue Gao et al., 2018), Study 2 did not reveal the involvement of DMPFC or rDLPFC in the peripheral effect but showed that these regions are related to the advantageous inequity aversion in other-oriented evaluations across different framed contexts. This discrepancy might be due to the different experimental settings, as previous studies used biologically relevant stimuli, which are thought to be more sensitive in detecting the peripheral effect. Moreover, aligned with precious findings (Haruno & Frith, 2009), Study 2 also indicated the right amygdala's involvement in the periphery effect of advantageous inequity aversion during the overall reciprocity decision-making. These findings provided a more detailed understanding of the periphery of reciprocity, elucidating the specific role of advantageous inequity aversion and the self-oriented evaluation in driving reciprocity.

Neurocomputational mechanism for anxiety effect on the core and periphery of reciprocity

The aim of Study 3 was to understand the neurocomputational mechanism by which anxiety modulates reciprocity in both the core and periphery. To achieve this, a combination of eye-tracking and EEG recording was applied to a binary TG (framed as either gain or loss contexts). As expected, the computational modeling results (validated by eye-tracking data) indicated that both the core and periphery underlying reciprocity were modulated by anxiety. Regarding the core, anxiety impaired reciprocity through reduced guilt aversion and advantageous inequity liking, regardless of context. Specifically, the reduction in guilt aversion due to anxiety was mediated by decreased P2 amplitudes (reflecting reduced selective attention) and increased LPP amplitudes (indicating heightened emotion regulation). Regarding the periphery, anxiety appeared to alter the peripheral perception of advantageous inequity aversion and reward sensitivity. The alteration of anxiety's effect on the peripheral perception of advantageous inequity aversion was associated with changes in N2 amplitudes (indicating modulation of cognitive control).

Following the computational modeling approach of Study 2, the use of eye-tracking techniques in Study 3 provided more observable measurements and successfully validated the computational model. The analysis of gaze patterns in Study 3 focused on transitions between areas of interest, capturing the decision-making process and reflecting the key psychological components underlying reciprocity (Devetag et al., 2016). This validation substantiates the main psychological components identified by the winning computational model, aligning with previous findings that individuals spend relatively more time on their areas of interest (Mitsuda & Glaholt, 2014; Palacios-Ibáñez et al., 2023). Additionally, EEG recordings complemented the fMRI methods used in the previous studies, providing high temporal resolution insights into the neural mechanisms of reciprocity.

Building upon the core and periphery structure of reciprocity established in Study 1 and Study 2, Study 3 took a step further, revealing how anxiety affects reciprocity. According to the literature, anxiety has been documented as detrimental to various prosocial behaviors, including reduced generosity (Rodebaugh et al., 2016), cooperation (Walters & Hope, 1998), and reciprocity (Anderl et al., 2018; Rodebaugh et al., 2011, 2013). Consistent with these broader detrimental effects, Study 3 confirmed that anxiety impairs reciprocity and elucidated how these effects transfer through psychological components and the core and periphery structure.

Extending the previous finding on individuals with the disorder often co-occurring with high anxiety levels, which showed lower guilt aversion (Xiao et al., 2022), findings from Study 3 emphasized the role of guilt aversion in anxiety's effect on the core of reciprocity. This result suggests that higher anxiety may allocate fewer attentional resources to assessing guilt (indicated by P2 amplitude) (Hajcak et al., 2012; Luck et al., 1994; Potts, 2004; Rey-Mermet et al., 2019), and more effortful cognitive regulation over or disengagement from the anticipatory guilt (indicated by LPP amplitude) in reciprocity decisions (Chen et al., 2009; Zhan et al., 2018, 2020). Our finding supports the idea that individuals with higher trait anxiety tend to adopt avoidance strategies (Duronto et al., 2005; Raffety et al., 1997; Turner, 1988).

Interestingly, study 3 demonstrated the distinct and opposite peripheral effects for the advantageous inequity aversion between individuals with low and high trait anxiety. This finding suggests several key points: First, advantageous inequity aversion is an instinctual response, and it take efforts to suppress this aversion. Second, the influence of peripheral manipulation is likely modulated by the N2 cognitive control process. Third, anxiety likely affects the periphery of reciprocity through the N2 cognitive control process. These findings unveiled how anxiety affects reciprocity from the behavioral, psychological, and neural levels and clarified its impact on the core and periphery of reciprocity.

Summary of Studies

While Study 1 provided the core and periphery neural mechanism of reciprocity at the brain network level, Study 2 further delved into the neurocomputational mechanisms, focusing on peripheral reciprocity and examining the overall reciprocity decision-making and its subprocess (other-oriented inference and self-oriented evaluation) at the brain region level. Based on these findings, Study 3 further contributed to identifying the neurocomputational mechanism underlying how anxiety distinctively affects reciprocity on the core and periphery.

Specifically, using computational modeling, both Study 2 and Study 3 identified the same winning (best-fitting) model that included psychological components of guilt aversion, advantageous inequity aversion, advantageous inequity liking, and reward sensitivity. Noteworthily, the employment of the eye-tracking technique in Study 3 validated all these four components, underscoring the robustness of the computational findings in this dissertation. These replication and verification strengthen the validity of the model and suggests the common involvement of these psychological components in reciprocity across different populations, such as groups with varying levels of trait anxiety.

Overall, this work provides convergent evidence for the distinct neurocomputational mechanisms underlying the core and periphery of reciprocity and reveals how anxiety influences reciprocity through these mechanisms.

6.1. Integration of Findings

Reciprocity decisions, while unavoidably influenced by varying peripheral contexts, possess a stable core mechanism that remains consistent. To better understand reciprocity, this dissertation is specifically interested in capturing those core and periphery in reciprocity decision-making.

Core Mechanism of Reciprocity

The core mechanism of reciprocity can be elucidated by examining individuals unaffected by contextual manipulations and identifying commonalities across different contexts. Although Study 2 didn't directly examine the core's psychological components, its results indirectly support Study 3's findings. Study 3 identified guilt aversion and advantageous inequity liking as core components of reciprocity, both attenuated by anxiety regardless of gain or loss contexts. The absence of peripheral effects on these components in Study 2 further suggests their resilience to manipulation, reinforcing the notion that they are part of reciprocity's core rather than its periphery.

Advantageous inequity aversion, identified as a peripheral component of reciprocity in both Studies 2 and 3, revealed complex dynamics between core and periphery across different subprocesses of reciprocity decision-making. In Study 2, the DMPFC, rDLPFC, and lSMG were significantly associated with advantageous inequity aversion during the other-oriented inference subprocess, with no peripheral differences found for these regions. This finding suggests that although specific computational components appear sensitive to the peripheral manipulation overall, the core mechanisms that are resistant to the change of context may also exist, depending on the specific sub-processing stage and brain region involved.

Converging the findings from three studies, the role of mentalizing and cognitive regulation (cognitive control) is of utmost importance for the core of reciprocity, regardless of context. Study 1 demonstrated the vital function of inter-network RSFC of DMN-CON for the core, with this inter-network connectivity contributing commonly to predicting reciprocity across both contexts. The finding highlights the interactive role of DMN and CON in reciprocity, where DMN facilitates mentalizing others' expectations, while CON exerts cognitive control over betrayal temptations. This interaction is crucial for resolving the social dilemma inherent in reciprocity decisions. Extending the findings from Study 1, Study 2 found that during reciprocity, the other-oriented inference subprocess involves DMPFC (a key hub of DMN related to mentalizing) and rDLPFC (associated with cognitive control), which are resistant to the peripheral manipulations. In addition, Study 3 complemented these findings, identifying guilt aversion as a core component underlying anxiety's impact on reciprocity. As the feeling of guilt derived from failing to meet others' expectations, mentalizing and empathy is the basis of guilt aversion (Chang et al., 2011; Xiao et al., 2022). Specifically, LPP amplitude, which mediates guilt aversion, is related to cognitive regulation over emotion (Bernat et al., 2011; Desatnik et al., 2017; Moser et al., 2014; Shafir et al., 2015), especially during the resolution of moral conflicts (Chen et al., 2009; Zhan et al., 2018, 2020). Integrating these findings from different neuroimaging methods (fMRI and EEG) and analytical techniques (CPM, computational modeling), the results from the three studies convergently provided robust evidence for the central role of mentalizing and cognitive regulation in the core of reciprocity.

Periphery Mechanism of Reciprocity

In contrast to the relatively indirect investigation of the core mechanism, a key strength of this dissertation lies in the consistent, direct manipulation of context to investigate the periphery of reciprocity across all three studies. Despite employing different framing (give/take and gain/loss), all studies demonstrated a peripheral effect on reciprocity at behavioral, psychological, and neural levels.

Peripheral Modulation on Behaviors

Specifically, Study 1 examined give and take framing with a one-shot TG and DTG, while Studies 2 and 3 explored then gain and loss framing in a binary TG. The consistent emergence of peripheral effects across different framing contexts highlights the robustness of the finding that reciprocity is susceptible to contextual changes. It suggests that contextual susceptibility is not limited to a specific type of framing but is probably a general property in reciprocity decision-making. Moreover, supporting social representation theory, the observation of peripheral effects across studies strengthens the argument for the existence of separable core and peripheral components in reciprocity decisions (Abric, 1993).

While the contextual effect on reciprocity was significant in Study 1, it showed marginal significance in both Study 2 and Study 3. This discrepancy in the strength of contextual effects can be attributed to these two points: On the one hand, the nature of the experimental paradigms and framing approaches differed between the studies. Study 1 employed a one-shot TG and DTG presented as give and take frames, while Study 2 and Study 3 utilized a multiple one-shot binary TG framed in gain and loss contexts. These different types of framing may have elicited different levels of psychological processing. The give/take framing carries a stronger social-emotional connotation, emphasizing the interpersonal nature of the exchange (Keysar et al., 2008). In contrast, the gain/loss framing could activate broader reward-processing mechanisms, highlighting the economic aspects of the interaction (Xu et al., 2013). As people tend to base their decisions more on emotional factors than on purely calculational ones (Lerner et al., 2015), the give/take framing may elicit a more pronounced effect on reciprocity behavior.

On the other hand, the measurement of reciprocity varied between studies. Study 1 used a continuous measure (the ratio between sent and previous received amounts) in single one-shot TG and DTG, whereas Study 2 and Study 3 employed a multiple categorical measure (the frequency of choosing reciprocity or betrayal) in multiple one-shot binary TG. This difference in measurement could affect the sensitivity to detect contextual effects. Although multiple categorical decisions allow for more stable estimates of behavior, single one-shot continuous measures might capture more subtle, instinctual responses (Rivera-Garrido et al., 2022), which results in relatively higher sensitivity of measurement in Study 1.

Peripheral Modulation on Psychological Components

Delving into the psychological mechanisms underlying reciprocity, the studies revealed intriguing patterns in the peripheral effects on advantageous inequity aversion. Study 2, which involved typically recruited university students, showed lower advantageous inequity aversion in the loss context compared to the gain context. Study 3 replicated this finding in the low anxiety group (trait anxiety score ≤ 35) but revealed an inverse pattern in the high anxiety group (trait anxiety score ≥ 48), where advantageous inequity aversion was higher in the loss context compared to the gain context.

Although trait anxiety score was not explicitly measured in Study 2, it is reasonable to assume that the participant sample represented a middle range of anxiety levels, given the random recruitment from the same university population. Synthesizing the results from both studies, we can infer a potential continuum of anxiety's impact on the peripheral effect of advantageous inequity aversion. Low anxiety and the inferred middle anxiety groups showed lower advantageous inequity aversion in loss compared to gain contexts, while the high anxiety group exhibited higher advantageous inequity aversion in loss compared to gain contexts. This pattern suggests a non-linear relationship between anxiety and the contextual modulation of advantageous inequity aversion. The reversal of contextual effects in high-anxiety individuals, compared to those with low or moderate anxiety, highlights a fundamental shift in how anxiety influences responses to moral concerns, such as other-regarding considerations, across different contexts.

The peripheral effect on reward sensitivity also exhibits interesting variations across Study 2 and Study 3, providing nuanced insights into the role of individual differences in contextual sensitivity. While Study 2 did not observe a peripheral effect on reward sensitivity, Study 3 revealed this effect specifically in the high anxiety group. The finding aligns with previous research demonstrating that individuals with high anxiety are more susceptible to framing effects on reward-related decisions (Gu et al., 2017; Xu et al., 2013). The differential impact of anxiety on contextual perception can be attributed to distinct decision-making strategies employed by individuals with varying levels of anxiety. Those with high anxiety tend to rely more heavily on heuristic decision-making processes compared to their low anxiety counterparts (Jepma & López-Solà, 2014). This tendency interacts with the way different frames are perceived: loss frames are heuristically interpreted as more harmful to others (Baron, 1995; Evans & van Beest, 2017) and more threatening to one's own rewards (Xu et al., 2013) compared to gain frames. Consequently, when faced with a loss frame, individuals with high trait anxiety may increase both their advantageous inequity aversion and reward sensitivity. This dual increase could be driven by two complementary motivations: the "do-no-harm" principle (Baron, 1995), which heightens concern for others in potentially harmful situations, and self-protective strategies (Meleshko & Alden, 1993), which amplifies attention to potential personal losses.

Peripheral Modulation on Neural Correlates

Regarding the neurospacial mechanism, a notable discrepancy for the peripheral modulation emerged between the findings of Study 1 and Study 2:

Study 1 highlighted the role of the intra-network RSFC of DMN in the periphery of reciprocity. In contrast, Study 2 emphasized the involvement of the right amygdala and lAI in the peripheral effect on advantageous inequity aversion. These regions are typically associated with the CON or salience network, but not DMN (Dosenbach et al., 2010; Shen et al., 2017). This discrepancy may be explained by the reason that advantageous inequity aversion may not capture all neural mechanisms involved in reciprocity behavior as it is one of the psychological components underlying reciprocity decisions. The overall reciprocity response likely emerges from the interactive contributions of multiple subcomponents, each potentially influenced by context in distinct ways even if they are not significantly shown under the current analysis method. Nevertheless, further studies are needed to resolve this inconsistent finding.

Regarding neurotemporal mechanisms, Study 3 links the N2 cognitive control mechanism to advantageous inequity aversion, indicating that this aspect of peripheral modulation is sensitive to contextual modification. Anxiety specifically alters both advantageous inequity aversion and the N2 cognitive control mechanism in a similar pattern, further supporting the explanation from Study 2 that advantageous inequity aversion is instinctual, with its modulation requiring cognitive effort—a function that the N2 cognitive process appears to facilitate.

Findings from a different angle highlights the value of a multi-method approach in studying complex social behaviors like reciprocity in this dissertation. By employing different paradigms and analytic methods, we can build a more comprehensive understanding of how peripheral manipulation influences reciprocity.

6.2. Contributions to the Literature

This dissertation makes several significant contributions to the existing literature on reciprocity.

Unlike previous studies that overlooked the role of context in reciprocity decision-making, this dissertation is grounded in social representation theory, unveiling the complexity of reciprocity by delving into its core and periphery. By integrating diverse perspectives from computational modeling, eye-tracking analysis, neural dynamics, and neuroimaging at both regional and network levels, this dissertation offers an unprecedented depth of understanding. The multi-modal and multi-level investigation presented in this dissertation provides a holistic view of reciprocity, spanning from large-scale brain networks to specific regional activities, and from slow hemodynamic responses to rapid electrophysiological changes. Furthermore, it bridges the gap between behavioral manifestations and underlying psychological components, while also linking latent psychological components to their observable validations through eye movement patterns.

Resolving the social dilemma between reciprocating others and maximizing personal interest is the key to reciprocity decision-making. While reciprocating may sacrifice self-interest, it can relieve the negative feelings such as guilt aversion, and advantageous inequity aversion as revealed in Study 2 and Study 3. Through the findings of Study 2 and Study 3, this work has identified core components of reciprocity, such as guilt aversion, as well as peripheral components like advantageous inequity aversion and reward sensitivity. These discoveries make significant contributions to the current literature by elucidating the origins of underlying psychological factors and revealing the fundamental psychological driving force of reciprocity propensity while distinguishing context-dependent elements.

Beyond the psychological implications, these studies also revealed the neural correlates of these psychological components. For example, in Study 2, although serving as a peripheral component overall, neural correlates of advantageous inequity aversion during other-oriented inference in reciprocity decisions exhibited core-like characteristics, localizing to specific brain regions such as the rDLPFC and DMPFC. In contrast, neural correlates of advantageous inequity aversion during overall and self-oriented evaluation in reciprocity decisions were localized to the right amygdala and lAI, demonstrating peripheral effects. In addition, Study 3 showed that the core component of guilt aversion was mediated by P2 and LPP amplitudes, whereas the peripheral component of advantageous inequity aversion was modulated by N2 amplitudes. These findings significantly contribute to the literature by elucidating both the spatial localization and temporal processing underlying the core and peripheral psychological components of reciprocity.

Study 1 provided the core and periphery mechanism of reciprocity by revealing their contributing large-scale neural network. Although not directly examining the psychological processes, the machine-learning prediction method used in this study established a more direct link between reciprocity behavior (core and periphery) and underlying neural mechanisms. Complementary to Study 2 and Study 3, the finding of Study 1 is particularly valuable to the literature given that behavior is often complex and may not perfectly correspond to individual psychological driving forces.

In sum, this dissertation significantly advances the field by providing a theoretically grounded, methodologically diverse, and multi-level analysis of reciprocity. It enhances our understanding of the mechanisms underlying reciprocity behavior and provides insights into the core-periphery framework, which has implications for future research on reciprocity and other aspects of social decision-making.

6.3. Practical Implications

Besides the contribution to the literature, this dissertation has several practical implications.

Study 1 identified connectome-based predictors for both the core propensity and context sensitivity associated with reciprocity. These neural "fingerprints" related to the core (inter-network RSFC of DMN-CON) and periphery (intra-network RSFC of DMN) could serve as biomarkers indexing an individual’s reciprocity propensity and sensitivity to the contexts. Such insights could enable early identification of individuals who may be less inclined to reciprocate, with or without considering their sensitivity to the context, facilitating the implementation of early support or education programs to foster more cooperative tendencies.

Study 2 also identified specific neural regions associated with component of reciprocity (advantageous inequity aversion) that is unaffected by peripheral manipulations, offering potential targets for neural interventions. The brain regions associated with advantageous inequity aversion during other-oriented inference, particularly the DMPFC, rDLPFC, and lSMG, could be potential targets for fMRI-based neurofeedback training or noninvasive stimulation methods like transcranial magnetic stimulation (TMS) on the rDLPFC. These interventions may help to enhance the reciprocity propensity, especially during inferring others' expectations. In contrast, based on the findings in Study 3, EEG-based neurofeedback training focusing on the LPP components (associated with emotional regulation) could potentially reduce guilt aversion, which helps individual with anxiety to relieve their psychological burden although it might also reduce their reciprocity propensity.

Study 3 demonstrated an inverted peripheral effect on advantageous inequity aversion in highly anxious individuals, which could serve as a potential marker for pathological anxiety. The distinct pattern in moral processing may help explain the lack of prosocial behavior often seen in anxiety disorders, suggesting that this behavior may not solely result from dispositional characteristics but rather from the unique response pattern to certain contexts. Moreover, targeting the modulation of advantageous inequity aversion could be a promising approach for interventions to enhance social functioning in individuals with high anxiety.

In summary, these studies provide several psychological and neural indicators that could potentially be used to promote prosocial behavior at both individual and societal levels. By leveraging these insights, we can work towards creating more cooperative and harmonious social environments.

6.4. Limitations and Future Directions

While this dissertation provides significant insights into the neural and computational mechanisms of reciprocity, several limitations should be acknowledged, which in turn suggest promising avenues for future research.

Integration of Spatial and Temporal Neural Measures: While this thesis employed both fMRI (Studies 1 and 2) and EEG (Study 3) techniques, these methods were not combined within a single study. Integrating fMRI and EEG in future research could provide a simultaneous understanding of both the spatial and temporal aspects of neural activity during reciprocity decisions. This combined approach would allow for better elucidation of the rapid temporal dynamics (from EEG) within specific brain regions or networks (from fMRI), offering a more complete picture of the neural processes underlying reciprocity.

Causal Inferences: The correlational nature of the neuroimaging data limits our ability to draw causal conclusions about the relationship between brain activity and reciprocity behavior. Future studies could employ techniques such as TDCS or TMS to establish causal relationships between specific brain regions and reciprocity decisions.

Clinical Applications: The findings regarding anxiety’s impairment of reciprocity suggest potential clinical relevance, but this was not explored in the clinical population. Future research could more directly investigate how clinical anxiety disorder affects reciprocity behavior and its neural underpinnings.

Computational Model Refinement: While the winning model in both Study 2 and Study 3 comprised four psychological components influencing reciprocity in a parallel structure, participants may evaluate decisions hierarchically. For instance, they might evaluate some components first and others later in the reciprocity decision. Future studies could explore more complex models considering the hierarchical structure of these components (Grover & Vriens, 2006).

By addressing these limitations and pursuing these future directions, researchers can build upon the foundation laid by this thesis to develop an even more comprehensive understanding of reciprocity.

6.5. Conclusions

Through a multi-modal and multi-level approach, this dissertation contributes to our understanding of how stable individual tendencies (core) and context-dependent processes (periphery) shapes reciprocity through their underlying neurocomputational mechanism. This research advances our understanding of the complex neurocomputational architecture supporting reciprocity and offers potential avenues for promoting reciprocity. While underscoring the need for further investigation, this work contributes to a nuanced, mechanistic understanding of reciprocity, bridging neuroscience, psychology, and economics. These insights bring us closer to fostering more cooperative social interactions at both individual and societal levels.

7. References

Abric, J.-C. (1993). Central system, peripheral system: Their functions and roles in the dynamics of social representations. *Papers on Social Representations*, *2*, 75–78.

Algoe, S. B., Haidt, J., & Gable, S. L. (2008). Beyond reciprocity: Gratitude and relationships in everyday life. *Emotion*, *8*(3), 425–429. https://doi.org/10.1037/1528-3542.8.3.425

Alós-ferrer, C., & Farolfi, F. (2019). Trust games and beyond. *Frontiers in Neuroscience*, *13*(September), 1–14. https://doi.org/10.3389/fnins.2019.00887

Anderl, C., Steil, R., Hahn, T., Hitzeroth, P., Reif, A., & Windmann, S. (2018). Reduced reciprocal giving in social anxiety – evidence from the trust game. *Journal of Behavior Therapy and Experimental Psychiatry*, *59*, 12–18. https://doi.org/10.1016/j.jbtep.2017.10.005

Baron, J. (1995). Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making*, *8*(2), 71–83. https://doi.org/10.1002/bdm.3960080202

Batson, C. D., & Moran, T. (1999). Empathy-induced altruism in a prisoner's dilemma. *European Journal of Social Psychology*, *29*(7), 909–924. https://doi.org/10.1002/(SICI)1099-0992(199911)29:7<909::AID-EJSP965>3.0.CO;2-L

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron*, *64*(5), 756–770. https://doi.org/10.1016/j.neuron.2009.11.017

Behrens, F., & Kret, M. E. (2019). The interplay between face-to-face contact and feedback on cooperation during real-life interactions. *Journal of Nonverbal Behavior*, *43*(4), 513–528. https://doi.org/10.1007/s10919-019-00314-1

Bellucci, G., Chernyak, S. V., Goodyear, K., Eickhoff, S. B., & Krueger, F. (2017). Neural signatures of trust in reciprocity: A coordinate-based meta-analysis. *Human Brain Mapping*, *38*(3), 1233–1248. https://doi.org/10.1002/hbm.23451

Bellucci, G., Feng, C., Camilleri, J., Eickhoff, S. B., & Krueger, F. (2018). The role of the anterior insula in social norm compliance and enforcement: Evidence from coordinate-based and functional connectivity meta-analyses. *Neuroscience and Biobehavioral Reviews*, *92*, 378–389. https://doi.org/10.1016/j.neubiorev.2018.06.024

Bellucci, G., Hahn, T., Deshpande, G., & Krueger, F. (2019). Functional connectivity of specific resting-state networks predicts trust and reciprocity in the trust game. *Cognitive, Affective* *and Behavioral Neuroscience*, *19*(1), 165–176. https://doi.org/10.3758/s13415-018-00654-3

Bereczkei, T., Papp, P., Kincses, P., Bodrogi, B., Perlaki, G., Orsi, G., & Deak, A. (2015). The neural basis of the Machiavellians' decision making in fair and unfair situations. *Brain and Cognition*, *98*, 53–64. https://doi.org/10.1016/j.bandc.2015.05.006

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Bernat, E. M., Cadwallader, M., Seo, D., Vizueta, N., & Patrick, C. J. (2011). Effects of instructed emotion regulation on valence, arousal, and attentional measures of affective processing. *Developmental Neuropsychology*, *36*(4), 493–518. https://doi.org/10.1080/87565641.2010.549881

Bohnet, I., & Meier, S. (2005). Deciding to distrust. *SSRN Electronic Journal*, *05*. https://doi.org/10.2139/ssrn.839225

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, *65*(1), 17–28. https://doi.org/10.1016/j.tpb.2003.07.001

Boyce, C. J., Brown, G. D., & Moore, S. C. (2010). Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological Science*, *21*(4), 471–475.

Cáceda, R., James, G. A., Gutman, D. A., & Kilts, C. D. (2015). Organization of intrinsic functional brain connectivity predicts decisions to reciprocate social behavior. *Behavioural Brain Research*, *292*, 478–483. https://doi.org/10.1016/j.bbr.2015.07.008

Calder, M., Craig, C., Culley, D., De Cani, R., Donnelly, C. A., Douglas, R., Edmonds, B., Gascoigne, J., Gilbert, N., Hargrove, C., Hinds, D., Lane, D. C., Mitchell, D., Pavey, G., Robertson, D., Rosewell, B., Sherwin, S., Walport, M., & Wilson, A. (2018). Computational modelling for decision-making: Where, why, what, who and how. *Royal Society Open Science*, *5*(6), 172096. https://doi.org/10.1098/rsos.172096

Cao, H., Plichta, M. M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., Esslinger, C., Kirsch, P., Meyer-Lindenberg, A., & Tost, H. (2014). Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *NeuroImage*, *84*, 888–900. https://doi.org/10.1016/j.neuroimage.2013.09.013

Cavanagh, J. F., & Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: Meta-analytic evidence. *Journal of Physiology-Paris*, *109*(1), 3–15. https://doi.org/10.1016/j.jphysparis.2014.04.003

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron*, *70*(3), 560–572. https://doi.org/10.1016/j.neuron.2011.02.056

Chen, P., Qiu, J., Li, H., & Zhang, Q. (2009). Spatiotemporal cortical activation underlying dilemma decision-making: An event-related potential study. *Biological Psychology*, *82*(2), 111–115. https://doi.org/10.1016/j.biopsycho.2009.06.007

Collier Villaume, S., Chen, S., & Adam, E. K. (2023). Age disparities in prevalence of anxiety and depression among US adults during the COVID-19 pandemic. *JAMA Network Open*, *6*(11), e2345073–e2345073. https://doi.org/10.1001/jamanetworkopen.2023.45073

Cox, C. A. (2013). Inequity aversion and advantage seeking with asymmetric competition. *Journal of Economic Behavior & Organization*, *86*, 121–136. https://doi.org/10.1016/j.jebo.2012.12.020

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases and rational decision-making in the human brain. *Science*, *313*(5787), 684–687. https://doi.org/10.1126/science.1128356

Desatnik, A., Bel-Bahar, T., Nolte, T., Crowley, M., Fonagy, P., & Fearon, P. (2017). Emotion regulation in adolescents: An ERP study. *Biological Psychology*, *129*, 52–61. https://doi.org/10.1016/j.biopsycho.2017.08.001

Devetag, G., Di Guida, S., & Polonio, L. (2016). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics*, *19*(1), 177–201. https://doi.org/10.1007/s10683-015-9432-5

Dohmen, T., Falk, A., Fliessbach, K., Sunde, U., & Weber, B. (2011). Relative versus absolute income, joy of winning, and gender: Brain imaging evidence. *Journal of Public Economics*, *95*(3), 279–285. https://doi.org/10.1016/j.jpubeco.2010.11.025

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2008). Representative trust and reciprocity: Prevalence and determinants. *Economic Inquiry*, *46*(1), 84–90. https://doi.org/10.1111/j.1465-7295.2007.00082.x

Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., Nelson, S. M., Wig, G. S., Vogel, A. C., Lessov-Schlaggar, C. N., Barnes, K. A., Dubis, J. W., Feczko, E., Coalson, R. S., Pruett, J. R., Barch, D. M., Petersen, S. E., & Schlaggar, B. L. (2010). Prediction of individual brain maturity using fMRI. *Science*, *329*(5997), 1358–1361. https://doi.org/10.1126/science.1194144

Duronto, P. M., Nishida, T., & Nakayama, S. (2005). Uncertainty, anxiety, and avoidance in communication with strangers. *International Journal of Intercultural Relations*, *29*(5), 549–560. https://doi.org/10.1016/j.ijintrel.2005.08.003

Eberle, J., & Daniel, J. (2022). Anxiety geopolitics: Hybrid warfare, civilisational geopolitics, and the Janus-faced politics of anxiety. *Political Geography*, *92*, 102502. https://doi.org/10.1016/j.polgeo.2021.102502

Evans, A. M., & van Beest, I. (2017). Gain-loss framing effects in dilemmas of trust and reciprocity. *Journal of Experimental Social Psychology*, *73*(July), 151–163. https://doi.org/10.1016/j.jesp.2017.06.012

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315. https://doi.org/10.1016/j.geb.2005.03.001

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*(1), 1–25. https://doi.org/10.1007/s12110-002-1012-7

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Fiske, S. T. (2011). *Envy up, scorn down: How status divides us*. Russell Sage Foundation.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Frith, E., Elbich, D. B., Christensen, A. P., Rosenberg, M. D., Chen, Q., Kane, M. J., Silvia, P. J., Seli, P., & Beaty, R. E. (2020). Intelligence and creativity share a common cognitive and neural basis. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000958

Gebodh, N., Esmaeilpour, Z., Adair, D., Schestattsky, P., Fregni, F., & Bikson, M. (2019). Transcranial direct current stimulation among technologies for low-intensity transcranial electrical stimulation: Classification, history, and terminology. In H. Knotkova, M. A. Nitsche, M. Bikson, & A. J. Woods (Eds.), *Practical Guide to Transcranial Direct Current Stimulation: Principles, Procedures and Applications* (pp. 3–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-95948-1_1

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery Clinics*, *22*(2), 133–139. https://doi.org/10.1016/j.nec.2010.11.001

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, *25*(2), 161–178. https://doi.org/10.2307/2092623

Grover, R., & Vriens, M. (2006). *The handbook of marketing research: Uses, misuses, and future advances*. Sage.

Gu, R., Wu, R., Broster, L. S., Jiang, Y., Xu, R., Yang, Q., Xu, P., & Luo, Y.-J. (2017). Trait anxiety and economic risk avoidance are not necessarily associated: Evidence from the framing effect. *Frontiers in Psychology*, *8*, 92. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00092

Gunnthorsdottir, A., McCabe, K., & Smith, V. (2002). Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology*, *23*(1), 49–66. https://doi.org/10.1016/S0167-4870(01)00067-8

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

Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology*, *69*(3), 339–348. https://doi.org/10.1016/j.tpb.2005.09.005

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology*, *35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hajcak, G., Weinberg, A., MacNamara, A., Foti, D., & others. (2012). ERPs and the study of emotion. *The Oxford Handbook of Event-Related Potential Components*, *441*, 474.

Hao, S., Xin, Q., Xiaomin, Z., Jiali, P., Xiaoqin, W., Rong, Y., & Cenlin, Z. (2023). Group membership modulates the hold-up problem: An event-related potentials and oscillations study. *Social Cognitive and Affective Neuroscience*, *18*(1), nsad071. https://doi.org/10.1093/scan/nsad071

Harsanyi, J. C. (1961). On the rationality postulates underlying the theory of cooperative games. *Journal of Conflict Resolution*, *5*(2), 179–196. https://doi.org/10.1177/002200276100500205

Haruno, M., & Frith, C. (2009). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*, 160–161. https://doi.org/10.1038/nn.2468

Jepma, M., & López-Solà, M. (2014). Anxiety and framing effects on decision making: Insights from neuroimaging. *Journal of Neuroscience*, *34*(10), 3455–3456. https://doi.org/10.1523/JNEUROSCI.5352-13.2014

Jiang, T., Potters, J., & Funaki, Y. (2016). Eye-tracking social preferences. *Journal of Behavioral Decision Making*, *29*(2–3), 157–168. https://doi.org/10.1002/bdm.1899

Jung, S. C. (2024). Economic slowdowns and international conflict. *Journal of Peace Research*, *61*(2), 180–196.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kanagaretnam, K., Mestelman, S., Nainar, S. M. K., & Shehata, M. (2010). Trust and reciprocity with transparency and repeated interactions. *Journal of Business Research*, *63*(3), 241–247. https://doi.org/10.1016/j.jbusres.2009.03.007

Keysar, B., Converse, B. A., Wang, J., & Epley, N. (2008). Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychological Science*, *19*(12), 1280–1286. https://doi.org/10.1111/j.1467-9280.2008.02223.x

Komorita, S. S., Parks, C. D., & Hulbert, L. G. (1992). Reciprocity and the induction of cooperation in social dilemmas. *Journal of Personality and Social Psychology*, *62*(4), 607–617. https://doi.org/10.1037/0022-3514.62.4.607

Kreps, D. M. & others. (1990). Corporate culture and economic theory. *Perspectives on Positive Political Economy*, *90*(109–110), 8.

Krueger, F. (2021). *The Neurobiology of Trust*. Cambridge University Press.

Krueger, F., Barbey, A. K., & Grafman, J. (2009). *The medial prefrontal cortex mediates social event knowledge*. *February*, 103–109. https://doi.org/10.1016/j.tics.2008.12.005

Krueger, F., Grafman, J., & McCabe, K. (2008). Neural correlates of economic game playing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1511), 3859–3874. https://doi.org/10.1098/rstb.2008.0165

Lang, F. R., & Fingerman, K. L. (2003). *Growing together: Personal relationships across the life span*. Cambridge University Press.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, *66*(Volume 66, 2015), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Li, J., Xiao, E., Houser, D., & Montague, P. R. (2009). Neural responses to sanction threats in two-party economic exchange. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(39), 16835–16840. https://doi.org/10.1073/pnas.0908855106

Li, O., Xu, F., & Wang, L. (2018). Advantageous inequity aversion does not always exist: The role of determining allocations modulates preferences for advantageous inequity. *Frontiers in Psychology*, *9*, 749. https://doi.org/10.3389/fpsyg.2018.00749

Li, X., Zhu, P., Yu, Y., Zhang, J., & Zhang, Z. (2017). The effect of reciprocity disposition on giving and repaying reciprocity behavior. *Personality and Individual Differences*, *109*, 201–206. https://doi.org/10.1016/j.paid.2017.01.007

Liu, J., Gu, R., Liao, C., Lu, J., Fang, Y., Xu, P., Luo, Y. J., & Cui, F. (2020). The neural mechanism of the social framing effect: Evidence from fMRI and tDCS studies. *Journal of Neuroscience*, *40*(18), 3646–3656. https://doi.org/10.1523/JNEUROSCI.1385-19.2020

Luck, S. J. (2014). *An introduction to the event-related potential technique*.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 887–904. https://doi.org/10.1037/0096-1523.20.4.887

MacNamara, A., & Proudfit, G. H. (2014). Cognitive load and emotional processing in generalized anxiety disorder: Electrocortical evidence for increased distractibility. *Journal of Abnormal Psychology*, *123*(3), 557–565. https://doi.org/10.1037/a0036997

Mahmoodi, A., Bahrami, B., & Mehring, C. (2018). Reciprocity of social influence. *Nature Communications*, *9*(1), 2474. https://doi.org/10.1038/s41467-018-04925-y

Mars, R. B., Neubert, F. X., Noonan, M. A. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain." *Frontiers in Human Neuroscience*, *6*(JUNE 2012), 1–9. https://doi.org/10.3389/fnhum.2012.00189

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy* *of Sciences of the United States of America*, *98*(20), 11832–11835. https://doi.org/10.1073/pnas.211415698

McLoughlin, G., Gyurkovics, M., Palmer, J., & Makeig, S. (2022). Midfrontal theta activity in psychiatric illness: An index of cognitive vulnerabilities across disorders. *Biological Psychiatry*, *91*(2), 173–182. https://doi.org/10.1016/j.biopsych.2021.08.020

Meleshko, K. G., & Alden, L. E. (1993). Anxiety and self-disclosure: Toward a motivational model. *Journal of Personality and Social Psychology*, *64*(6), 1000–1009. https://doi.org/10.1037/0022-3514.64.6.1000

Mitsuda, T., & Glaholt, M. G. (2014). Gaze bias during visual preference judgements: Effects of stimulus category and decision instructions. *Visual Cognition*, *22*(1), 11–29. https://doi.org/10.1080/13506285.2014.881447

Morris, P. (2012). *Introduction to game theory*. Springer Science & Business Media.

Moser, J. S., Hartwig, R., Moran, T. P., Jendrusina, A. A., & Kross, E. (2014). Neural markers of positive reappraisal and their associations with trait reappraisal and worry. *Journal of Abnormal Psychology*, *123*(1), 91–105. https://doi.org/10.1037/a0035817

Nihonsugi, T., Ihara, A., & Haruno, M. (2015). Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. *Journal of Neuroscience*, *35*(8), 3412–3419. https://doi.org/10.1523/JNEUROSCI.3885-14.2015

Nihonsugi, T., Numano, S., & Haruno, M. (2021). Functional connectivity basis and underlying cognitive mechanisms for gender differences in guilt aversion. *Eneuro*, *8*(6), ENEURO.0226-21.2021. https://doi.org/10.1523/ENEURO.0226-21.2021

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–9872. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC55275/

Palacios-Ibáñez, A., Marín-Morales, J., Contero, M., & Alcañiz, M. (2023). Predicting decision-making in virtual environments: An eye movement analysis with household products. *Applied Sciences*, *13*(12), Article 12. https://doi.org/10.3390/app13127124

Paul, S., Simon, D., Endrass, T., & Kathmann, N. (2016). Altered emotion regulation in obsessive-compulsive disorder as evidenced by the late positive potential. *Psychological Medicine*, *46*(1), 137–147. https://doi.org/10.1017/S0033291715001610

Pelligra, V. (2011). Empathy, guilt-aversion, and patterns of reciprocity. *Journal of Neuroscience, Psychology, and Economics*, *4*(3), 161–173. https://doi.org/10.1037/a0024688

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, *17*(4), 251–283. https://doi.org/10.1002/per.474

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition*, *56*(1), 5–13. https://doi.org/10.1016/j.bandc.2004.03.006

Qi, S., Luo, Y., Tang, X., Li, Y., Zeng, Q., Duan, H., Li, H., & Hu, W. (2016). The temporal dynamics of directed reappraisal in high-trait-anxious individuals. *Emotion (Washington, D.C.)*, *16*(6), 886–896. https://doi.org/10.1037/emo0000186

Raffety, B. D., Smith, R. E., & Ptacek, J. T. (1997). Facilitating and debilitating trait anxiety, situational anxiety, and coping with an anticipated stressor: A process analysis. *Journal of Personality and Social Psychology*, *72*(4), 892–906. https://doi.org/10.1037/0022-3514.72.4.892

Raichle, M. E. (2011). The restless brain. *Brain Connectivity*, *1*(1), 3–12. https://doi.org/10.1089/brain.2011.0019

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, *38*(1), 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation* (Vol. 165). University of Michigan press.

Ren, Z., Daker, R. J., Shi, L., Sun, J., Beaty, R. E., Wu, X., Chen, Q., Yang, W., Lyons, I. M., Green, A. E., & Qiu, J. (2021). Connectome-based predictive modeling of creativity anxiety. *NeuroImage*, *225*(August 2020), 117469. https://doi.org/10.1016/j.neuroimage.2020.117469

Rey-Mermet, A., Gade, M., & Steinhauser, M. (2019). Sequential conflict resolution under multiple concurrent conflicts: An ERP study. *NeuroImage*, *188*, 411–418. https://doi.org/10.1016/j.neuroimage.2018.12.031

Rivera-Garrido, N., Ramos-Sosa, M. P., Accerenzi, M., & Brañas-Garza, P. (2022). Continuous and binary sets of responses differ in the field. *Scientific Reports*, *12*(1), 14376. https://doi.org/10.1038/s41598-022-17907-4

Rodebaugh, T. L., Heimberg, R. G., Taylor, K. P., & Lenze, E. J. (2016). Clarifying the behavioral economics of social anxiety disorder: Effects of interpersonal problems and symptom severity on generosity. *Clinical Psychological Science*, *4*(1), 107–121. https://doi.org/10.1177/2167702615578128

Rodebaugh, T. L., Klein, S. R., Yarkoni, T., & Langer, J. K. (2011). Measuring social anxiety related interpersonal constraint with the flexible iterated prisoner's dilemma. *Journal of Anxiety Disorders*, *25*(3), 427–436. https://doi.org/10.1016/j.janxdis.2010.11.006

Rodebaugh, T. L., Shumaker, E. A., Levinson, C. A., Fernandez, K. C., Langer, J. K., Lim, M. H., & Yarkoni, T. (2013). Interpersonal constraint conferred by generalized social anxiety disorder is evident on a behavioral economics task. *Journal of Abnormal Psychology*, *122*(1), 39–44. https://doi.org/10.1037/a0030975

Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science (New York, N.Y.)*, *318*(5850), 598–602. https://doi.org/10.1126/science.1142996

Shafir, R., Schwartz, N., Blechert, J., & Sheppes, G. (2015). Emotional intensity influences pre-implementation and implementation of distraction and reappraisal. *Social Cognitive and Affective Neuroscience*, *10*(10), 1329–1337. https://doi.org/10.1093/scan/nsv022

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518. https://doi.org/10.1038/nprot.2016.178

Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering*, *16*, 260–277. IEEE Reviews in Biomedical Engineering. https://doi.org/10.1109/RBME.2021.3066072

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., & Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, *106*(31), 13040–13045. https://doi.org/10.1073/pnas.0905267106

Starmans, C., Sheskin, M., & Bloom, P. (2017). Why people prefer unequal societies. *Nature Human Behaviour*, *1*(4), 1–7. https://doi.org/10.1038/s41562-017-0082

Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, *18*(1), 70. https://doi.org/10.4103/0972-6748.57865

Thiruchselvam, R., Blechert, J., Sheppes, G., Rydstrom, A., & Gross, J. J. (2011). The temporal dynamics of emotion regulation: An EEG study of distraction and reappraisal. *Biological Psychology*, *87*(1), 84–92. https://doi.org/10.1016/j.biopsycho.2011.02.009

Tucker, W. T., & Ferson, S. (2008). Evolved altruism, strong reciprocity, and perception of risk. *Annals of the New York Academy of Sciences*, *1128*, 111–120. https://doi.org/10.1196/annals.1399.012

Turner, J. H. (1988). *A Theory of Social Interaction*. Stanford University Press.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

van Baar, J. M., Chang, L. J., & Sanfey, A. G. (2019). The computational and neural substrates of moral strategies in social decision-making. *Nature Communications*, *10*(1), 1483. https://doi.org/10.1038/s41467-019-09161-6

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. R. B., & Crone, E. A. (2009). What motivates repayment? Neural correlates of reciprocity in the trust game. *Social Cognitive and Affective Neuroscience*, *4*(3), 294–304. https://doi.org/10.1093/scan/nsp009

van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A. R. B., & Crone, E. A. (2011). Changing brains, changing perspectives: The neurocognitive development of reciprocity. *Psychological Science*, *22*(1), 60–70. https://doi.org/10.1177/0956797610391102

van Dijk, E., & De Dreu, C. K. W. (2021). Experimental games and social decision making. *Annual Review of Psychology*, *72*, 415–438. https://doi.org/10.1146/annurev-psych-081420-110718

Wagenaar, W. A., Keren, G., & Lichtenstein, S. (1988). Islanders and hostages: Deep and surface structures of decision problems. *Acta Psychologica*, *67*(2), 175–189. https://doi.org/10.1016/0001-6918(88)90012-1

Walters, K. S., & Hope, D. A. (1998). Analysis of social behavior in individuals with social phobia and nonanxious participants using a psychobiological model. *Behavior Therapy*, *29*(3), 387–407. https://doi.org/10.1016/S0005-7894(98)80039-7

Wang, Z., Goerlich, K. S., Ai, H., Aleman, A., Luo, Y., & Xu, P. (2021). Connectome-based predictive modeling of individual anxiety. *Cerebral Cortex*, *31*(6), 3006–3020. https://doi.org/10.1093/cercor/bhaa407

Wedel, M., Pieters, R., & van der Lans, R. (2023). Modeling Eye Movements During Decision Making: A Review. *Psychometrika*, *88*(2), 697–729. https://doi.org/10.1007/s11336-022-09876-4

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

Xia, C., Wang, J., Perc, M., & Wang, Z. (2023). Reputation and reciprocity. *Physics of Life Reviews*, *46*, 8–45. https://doi.org/10.1016/j.plrev.2023.05.002

Xiao, F., Zhao, J., Fan, L., Ji, X., Fang, S., Zhang, P., Kong, X., Liu, Q., Yu, H., Zhou, X., Gao, X., & Wang, X. (2022). Understanding guilt-related interpersonal dysfunction in obsessive-compulsive personality disorder through computational modeling of two social interaction tasks. *Psychological Medicine*, *53*(12), 5569--5581. https://doi.org/10.1017/S003329172200277X

Xiaoxue Gao, Gao, X., Hongbo Yu, Yu, H., Ignacio Sáez, Saez, I., Philip R. Blue, Blue, P. R., Lusha Zhu, Zhu, L., Ming Hsu, Hsu, M., Xiaolin Zhou, & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity aversion. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(33), 201802523. https://doi.org/10.1073/pnas.1802523115

Xu, P., Gu, R., Broster, L. S., Wu, R., Van Dam, N. T., Jiang, Y., Fan, J., & Luo, Y. (2013). Neural basis of emotional decision making in trait anxiety. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(47), 18641–18653. https://doi.org/10.1523/JNEUROSCI.1253-13.2013

Yang, Q., Bellucci, G., Hoffman, M. B., Hsu, K.-T., Lu, B., Deshpande, G., & Krueger, F. (2021). Intrinsic functional connectivity of the frontoparietal network predicts inter-individual differences in the propensity for costly third-party punishment. *Cognitive, Affective, & Behavioral Neuroscience*, *21*, 1222–1232. https://doi.org/10.3758/s13415-021-00927-4

Zhan, Y., Xiao, X., Li, J., Liu, L., Chen, J., Fan, W., & Zhong, Y. (2018). Interpersonal relationship modulates the behavioral and neural responses during moral decision-making. *Neuroscience Letters*, *672*, 15–21. https://doi.org/10.1016/j.neulet.2018.02.039

Zhan, Y., Xiao, X., Tan, Q., Li, J., Fan, W., Chen, J., & Zhong, Y. (2020). Neural correlations of the influence of self-relevance on moral decision-making involving a trade-off between harm and reward. *Psychophysiology*, *57*(9), e13590. https://doi.org/10.1111/psyp.13590

Zuo, X.-N., & Xing, X.-X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: A systems neuroscience perspective. *Neuroscience & Biobehavioral Reviews*, *45*, 100–118. https://doi.org/10.1016/j.neubiorev.2014.05.009

8. Further Publications During My Doctorate

#: Co-first authors

Safari, N.#, **Fang, H.#**, Veerareddy, A., Xu, P., & Krueger, F. (2024). The anatomical structure of sex differences in trust propensity: A voxel-based morphometry study. *Cortex*, *176*, 260-273, https://doi.org/10.1016/j.cortex.2024.02.018.

Veerareddy, A.#, **Fang, H.#**, Safari, N., Xu, P., Krueger, F. (2023). Cognitive empathy mediates the relationship between gray matter size of dorsomedial prefrontal cortex and social network size: A voxel-based morphometry study. *Cortex,* 169, 279-289, https://doi.org/10.1016/j.cortex.2023.09.015.

Yu, F.#, **Fang, H.#,** Zhang, J., Wang, Z., Ai, H., Fang, Y., Guo, Y., Wang, X., Zhu, C., Luo, Y., Xu, P., Wang, K. (2022). Individualized prediction of consummatory anhedonia from functional connectome in major depressive disorder. *Depression and Anxiety, 39(12), 858-869,* https://doi.org/10.1002/da.23292.

Fang, H., Li, X., Zhang, W., Fan, B., Wu, Y., & Peng, W. (2021). Single dose testosterone administration enhances novelty responsiveness and short-term habituation in healthy males. *Hormones and behavior*, *131*, 104963, https://doi.org/10.1016/j.yhbeh.2021.104963.

Fang, H. #, Li, X. #, Wu, Y., & Peng, W. (2020). Single dose testosterone administration modulates the temporal dynamics of distractor processing. *Psychoneuroendocrinology*, *121*, 104838, https://doi.org/10.1016/j.psyneuen.2020.104838.

Zeng, N., Aleman, A., Liao, C., **Fang, H.**, Xu, P., & Luo, Y. (2023). Role of the amygdala in disrupted integration and effective connectivity of cortico-subcortical networks in apathy. *Cerebral Cortex*, 33(6), 3171-3180, https://doi.org/10.1093/cercor/bhac267.

Bayat, D., Mohamadpour, H., **Fang, H.**, Xu, P., & Krueger, F. (2023). The impact of order effects on the framing of trust and reciprocity behaviors. Games, 14(2), https://doi.org/10.3390/g14020021.

Ding, X., Zheng, L., Wu, J., Liu, Y., **Fang, H.**, Xin, Y., & Duan, H. (2023). Performance monitoring moderates the relationship between stress and negative affect in response to an exam stressor. *International Journal of Psychophysiology,* 185, 11-18, https://doi.org/10.1016/j.ijpsycho.2023.01.001.

Ding, X., **Fang, H.**, Liu, Y., Zheng, L., Zhu, X., Duan, H., & Wu, J. (2022). Neurocognitive correlates of psychological resilience: Event-related potential studies. *Journal of Affective Disorders*, 312, 100-106, https://doi.org/10.1016/j.jad.2022.06.023.

Wu, J., Liu, Y., **Fang, H.**, Qin, S., Kohn, N., & Duan, H. (2021). The relationship between childhood stress and distinct stages of dynamic behavior monitoring in adults: neural and behavioral correlates. *Social Cognitive and Affective Neuroscience*, 16(9), 937-949, https://doi.org/10.1093/scan/nsab041.

9. Co-Author’s Statements

Confirmation of Independent Contributions by Huihua Fang for the Publication:

"Connectome-based individualized prediction of reciprocity propensity and sensitivity to framing: A resting-state functional magnetic resonance imaging study"

We hereby confirm that Mr. Huihua Fang has made the following contributions to the above-mentioned publication independently and on his own responsibility:

- Conceptualization and experimental design
- Data curation
- Data analysis
- Writing the first draft of the manuscript
- Manuscript revisions

________________ Mannheim, 04.09.2024
Huihua Fang Place, Date

________________ Shenzhen, 04.09.2024
Chong Liao Place, Date

________________ Shenzhen, 04.09.2024
Zhao Fu Place, Date

________________ Shenzhen, 04.09.2024
Shuang Tian Place, Date

________________ Beijing, 04.09.2024
Yuejia Luo Place, Date

________________ Beijing, 04.09.2024
Pengfei Xu Place, Date

________________ Mannheim, 04.09.2024
Frank Krueger Place, Date

Confirmation of Independent Contributions by Huihua Fang for the Publication:

"How context shapes reciprocity: Insights from fMRI and computational modeling"

We hereby confirm that Mr. Huihua Fang has made the following contributions to the above-mentioned publication independently and on his own responsibility:

- Conceptualization and experimental design
- Data curation
- Data analysis
- Writing the first draft of the manuscript
- Manuscript revisions

―――――――― Mannheim, 04.09.2024

Huihua Fang Place, Date

―――――――― Beijing, 04.09.2024

Pengfei Xu Place, Date

―――――――― Beijing, 04.09.2024

Yuejia Luo Place, Date

―――――――― Mannheim, 04.09.2024

Frank Krueger Place, Date

Confirmation of Independent Contributions by Huihua Fang for the Publication:

"Trait anxiety impairs reciprocity behavior: A multi-modal and computational modeling study"

We hereby confirm that Mr. Huihua Fang has made the following contributions to the above-mentioned publication independently and on his own responsibility:

- Conceptualization and experimental design
- Data curation
- Data analysis
- Writing the first draft of the manuscript
- Manuscript revisions

________________ Mannheim, 04.09.2024
Huihua Fang Place, Date

________________ Shenzhen, 04.09.2024
Rong Wang Place, Date

________________ Paris, 04.09.2024
Zhihao Wang Place, Date

________________ Shenzhen, 04.09.2024
Qian Liu Place, Date

________________ Beijing, 04.09.2024
Yuejia Luo Place, Date

________________ Beijing, 04.09.2024
Pengfei Xu Place, Date

________________ Mannheim, 04.09.2024
Frank Krueger Place, Date

10. Acknowledgments

I would like to express my gratitude to the following individuals who have been very supportive in my PhD journey:

First and foremost, I extend my heartfelt thanks to Professor Frank Krueger for his patient, diligent, and responsible supervision throughout my PhD training. The memories of his warm welcome - picking me up at the airport, training, our countless discussions, and sharing great times with his family - will forever remain cherished moments in my life.

I am grateful to Professor Georg Alpers for his unwavering support during my studies at the University of Mannheim. His guidance has been invaluable to my academic growth.

I sincerely appreciate Professor Yuejia Luo and Professor Pengfei Xu for their continuous support and guidance throughout my academic journey. I also wish to thank Professor Qing Guan for her care and encouragement.

I would like to express my thanks to my Chinese teammates—Zhihao Wang, Rong Wang, Yiman Li, Xuan Yu, Ningning Zeng, Yaner Su, Junyi Zhang, Zhi Guang, Chong Liao, Zhao Fu, and Shuang Tian…—for their help throughout the studies. My PhD journey would have been far less enjoyable without you all!

I also want to thank my colleagues from Professor Alpers's group at Mannheim University for their warm welcome and helping me out during my time in Germany. A special shout-out to Ulrich and Laura, who've been incredibly friendly and supportive while I've been working on my thesis.

I owe so much to my lovely family for always backing my choices, no matter what. A special thank you goes to my wife, Qian Liu, whose unwavering support has been a constant source of strength, regardless of the distance between us. Our travels around Europe together are some of the most cherished memories from my time studying in Germany, and I am incredibly grateful for everything we've shared along this journey.

As I look back on my career, I can see how each step has led me to where I am now. I'm incredibly thankful for the amazing journey and experiences I've had in Germany.