

---

AI translation · View original & related papers at  
[chinaxiv.org/items/chinaxiv-202603.00034](https://chinaxiv.org/items/chinaxiv-202603.00034)

---

## The Influence of Social Value Orientation on Source Memory in Reciprocal Cooperation

**Authors:** Yang Jingtong, Jiang Yingjie, Long Yiting, Jiang Yingjie

**Date:** 2026-03-08T20:53:37+00:00

### Abstract

This study employed the sequential Prisoner's Dilemma paradigm, combining behavioral and EEG experiments to investigate the influence of social value orientation on source memory for cooperative and betraying partners in reciprocal cooperation, as well as its electrophysiological mechanisms. Behavioral results revealed that individuals with a prosocial value orientation exhibited superior source memory for cooperative partners, whereas those with a pro-self value orientation showed better source memory for betraying partners. EEG results indicated that during the early stage of outcome evaluation, individuals with a pro-self value orientation elicited larger FRN amplitudes under betrayal feedback, while those with a prosocial value orientation elicited larger FRN amplitudes under cooperation feedback. In the late stage, only individuals with a pro-self value orientation elicited larger LPP amplitudes and stronger theta oscillations under betrayal feedback. These findings suggest that individuals with different social value orientations exhibit opposing outcome evaluation patterns toward cooperation and betrayal feedback, and only those with a pro-self value orientation experience higher emotional arousal and allocate more attentional resources to betrayal feedback. Furthermore, both FRN amplitudes and theta oscillations were significantly correlated with source memory accuracy. In summary, social value orientation influences individuals' source memory preferences for cooperative and betraying partners by modulating the outcome evaluation process during reciprocal cooperation.

Full Text

Preamble

## The Influence of Social Value Orientation on Source Memory in Reciprocal Cooperation

Yang Jingtong, Jiang Yingjie, Long Yiting (*School of Psychology, Northeast Normal University, Changchun 130024*)

**Abstract:** This study employed the sequential Prisoner's Dilemma paradigm, combining behavioral and electroencephalogram (EEG) experiments to investigate the influence of Social Value Orientation (SVO) on source memory for cooperative and defecting partners in reciprocal cooperation, as well as its underlying electrophysiological mechanisms. Behavioral results revealed that pro-social individuals exhibited superior source memory for cooperative partners, whereas pro-self individuals showed better source memory for defecting partners. EEG results indicated that during the early stage of outcome evaluation, pro-self individuals elicited larger FRN amplitudes in response to betrayal feedback, while pro-social individuals elicited larger FRN amplitudes in response to cooperation feedback. In the late stage, only pro-self individuals exhibited larger LPP amplitudes and stronger theta oscillations under betrayal feedback. These findings suggest that individuals with different SVOs exhibit opposing outcome evaluation patterns toward cooperation and betrayal feedback; specifically, only pro-self individuals experience higher emotional arousal and allocate more attentional resources toward betrayal feedback. Furthermore, both FRN amplitudes and theta oscillations were significantly correlated with source memory accuracy. In conclusion, SVO influences an individual's source memory preference for cooperative versus defecting partners by modulating the outcome evaluation process during reciprocal cooperation.

**Keywords:** reciprocal cooperation, social value orientation (SVO), outcome evaluation, event-related potentials (ERPs), source memory

**Classification Code:** B842

Reciprocal cooperation refers to a form of social interaction in which individuals or groups support one another based on the principle of reciprocity to achieve common goals and interests [?, ?]. This type of reciprocity-based cooperation is a core cooperative mechanism in human society [?, ?]. Memory regarding the targets of reciprocal cooperation and the behavior of partners falls within the domain of episodic memory and significantly influences future cooperative decision-making [?, ?]. Source memory [?, ?, ?] in the context of reciprocal cooperation involves remembering the background information—such as whether a specific partner cooperated or defected in previous interactions—when individuals are faced with different partners who may act either way.

Previous research across various paradigms has consistently demonstrated the impact of source memory on reciprocal cooperative decision-making [?, ?].

*Received Date: 2025-04-09 Supported by the National Natural Science Foundation of China (32271095) and the Natural Science Foundation of Jilin Province (20230101149JC).*

**Corresponding Author:** Jiang Yingjie, E-mail: jiangyj993@nenu.edu.cn

In a study by Bell et al. [?], prior to a cooperative investment decision task, participants were provided with faces of each partner along with their respective reputation information.

**Results indicated that participants were only able to correctly select partners for investment when they possessed source memory for the partners' reputation information.**

Schaper et al. [?] utilized a sequential Prisoner's Dilemma game paradigm and found that the more accurately individuals remembered the source of their interaction partners, the more they were able to make adaptive decisions. Specifically, participants cooperated more with previous cooperators and less with previous defectors in subsequent rounds, demonstrating that individuals form cooperative intentions based on source memory of their partners' prior behavior.

Some researchers argue that there is a source memory advantage for defectors in reciprocal cooperation. Compared to cooperative behavior, defection is thought to receive more cognitive resources [?, ?, ?] and possess a memory advantage [?, ?, ?]. Previous studies have explained this phenomenon based on the cheater-detection hypothesis [?, ?]. Social contract theory [?, ?] suggests that most people have an innate tendency toward cooperation; therefore, individuals have evolved a cheater-detection mechanism to balance this tendency by focusing on and remembering defectors. However, subsequent studies have yielded inconsistent conclusions, finding no significant difference in source memory between cooperators and defectors in the Prisoner's Dilemma [?, ?]. Furthermore, many studies using different experimental paradigms have found that individuals may exhibit superior source memory for both cooperators and defectors [?, ?, ?]. In summary, existing research remains inconsistent regarding whether individuals possess a source memory advantage for defectors in reciprocal cooperation.

In addition to the potential influence of cheater-detection mechanisms, source memory in reciprocal cooperation may also be affected by individual differences in social value orientation. Social Value Orientation (SVO) refers to the stable social preferences individuals exhibit in social interaction contexts that influence their cooperative behavior [?, ?, ?]. Based on behavioral preferences shown in the distribution of outcomes between oneself and others [?, ?], SVO can be categorized into two main types: Prosocial Orientation and Proself Orientation [?, ?]. When allocating outcomes, prosocial individuals (hereafter referred to as prosocials) seek to maximize the joint interests of themselves and others while minimizing differences between them, placing greater emphasis on moral evaluation. In contrast, proself individuals (hereafter referred to as proselfs) tend

to approach problems from a purely rational perspective, aiming to maximize personal gain and emphasizing their own advantage, placing their interests far above those of others [?, ?]. In social interactions, prosocials exhibit a higher tendency toward cooperation and cooperative behavior compared to proselves [?, ?, ?]. For instance, in the Prisoner's Dilemma game, the cooperation rate of prosocials is significantly higher than that of proselves [?, ?]. Prosocials value reciprocity more highly in social interactions and are more willing to maintain long-term reciprocal relationships, even when there is a risk of personal loss. Conversely, the cooperative intentions of proselves usually depend on calculations of personal gain and loss; when facing uncertainty, they are more likely to adopt self-interested strategies [?, ?]. These differences in social decision-making patterns suggest that prosocials and proselves may also have divergent biases in the formation of episodic memory. Evidence for this inference is provided by the distinct neural activities observed when processing social decision feedback [?, ?, ?].

In reciprocal cooperation, the process by which individuals evaluate outcome feedback is known as outcome evaluation [?, ?]. Feedback-related negativity (FRN), which reflects early outcome evaluation processes, typically appears 200–350 ms after feedback. It is also referred to as medial frontal negativity (MFN) and is localized in the anterior cingulate cortex [?, ?, ?]. FRN is thought to reflect the learning and decision-making processes involved in outcome evaluation [?, ?]. Research indicates that FRN amplitude is not influenced by the positive or negative valence of the outcome but is instead sensitive to the discrepancy between expected and actual results [?, ?, ?]. For example, in the Prisoner's Dilemma, both unexpected defection and unexpected cooperation induce larger FRN amplitudes [?, ?]. However, other studies have found that defection feedback generally elicits larger FRN amplitudes regardless of expectations [?, ?]. Consequently, some researchers argue that FRN not only reflects expectancy violation but is also related to the early emotional and motivational assessments triggered by feedback [?, ?]. Furthermore, FRN amplitude is influenced by an individual's SVO; compared to proselves, prosocials show more pronounced FRN differences when facing fair versus unfair feedback in the Ultimatum Game and exhibit higher behavioral rejection rates for unfair distributions [?, ?].

The late positive potential (LPP) is an ERP component elicited during the late stage of outcome feedback, thought to reflect late-stage motivational and emotional processing of feedback information [?, ?]. The LPP typically begins around 400 ms post-feedback and lasts until approximately 1000 ms, primarily distributed over the centro-parietal regions of the scalp [?, ?]. Research shows that, compared to neutral stimuli, positive or negative stimuli typically induce larger LPP amplitudes [?, ?]. Additionally, the LPP is believed to reflect sustained attentional investment and further processing of feedback information during the outcome evaluation process [?, ?]. From a memory processing perspective, the LPP is closely associated with the emotion-enhanced memory effect, where increased amplitude predicts better subsequent memory performance [?, ?]. Studies have found that individuals with different SVOs exhibit

different memory preferences and LPP trends when remembering positive and negative social information [?, ?]. Specifically, prosocials show better recall for positive social information, which also elicits larger LPPs, while proselves exhibit the opposite trend.

In addition to ERP components, neural oscillatory activity is also involved in the outcome evaluation process. Among these, theta (4-8 Hz) oscillation is a common component induced by feedback processing. Research indicates that midfrontal theta oscillations are closely related to feedback processing, typically appearing in the early-to-mid time window (200-500 ms) after feedback [?, ?, ?], and show higher sensitivity to negative feedback [?, ?]. For example, Foti et al. [?] found that theta oscillations significantly increase after receiving negative feedback, suggesting they may reflect the processing of negative outcomes. Furthermore, midfrontal theta oscillations play a crucial role in the formation and retrieval of source memory, as theta synchronization facilitates effective encoding [?, ?]. Therefore, theta oscillations not only reflect the processing of defection feedback during outcome evaluation but may also play a vital role in the source memory encoding process. Overall, FRN, LPP, and theta oscillations reflect the processing mechanisms of outcome evaluation across different temporal stages and levels of neural activity. Existing research suggests that SVO can modulate the processing of social feedback information, manifesting as differences in ERP components such as FRN and LPP. Thus, SVO may further influence source memory in reciprocal cooperation by affecting the outcome evaluation process.

In summary, while previous research consistently shows that source memory guides decision-making in reciprocal cooperation, the existence of a memory advantage for cooperation versus defection remains controversial. This study hypothesizes that SVO may influence source memory in reciprocal cooperation by acting on the outcome evaluation process. The early component of outcome evaluation, FRN, may be related to the expectancy evaluation of feedback, followed by more refined cognitive and emotional processes reflected in the late component LPP, while theta oscillations reflect the processing of defection feedback and support effective source memory encoding. Therefore, this study employs a sequential Prisoner's Dilemma paradigm [?, ?, ?] to examine individual source memory and uses EEG technology to record the outcome evaluation process, aiming to reveal the influence of SVO on source memory in reciprocal cooperation and its underlying mechanisms. Experiment 1 is a behavioral experiment exploring whether individuals with different SVOs exhibit different memory preferences for the source of cooperative versus defecting partners. Experiment 2 utilizes EEG to examine the impact of SVO on the outcome evaluation of cooperative and defection feedback, investigating the mechanisms by which SVO affects source memory. Based on prior research, we hypothesize that behaviorally, proselves—who are more concerned with personal interest and more sensitive to negative information [?, ?]—will have better source memory for defecting partners, whereas prosocials—who value reciprocity and long-term relationships [?, ?, ?]—will have better source memory for cooperative partners.

Electrophysiologically, based on previous findings [?, ?, ?], we predict that individuals with different SVOs will exhibit distinct outcome evaluation processes following defection and cooperation feedback; compared to prosocials, proselves will show stronger FRN and LPP amplitudes under defection feedback, along with more significant theta oscillation responses.

**Experiment 1 utilized the sequential Prisoner' s Dilemma paradigm to create a reciprocal cooperation context [?, ?, ?].**

Participants engaged in two rounds of the Prisoner' s Dilemma game, interspersed with an unexpected memory test, to investigate differences in source memory for cooperative and defecting partners among individuals with different Social Value Orientations (SVO) during reciprocal cooperation.

The experiment consisted of two distinct rounds of the Prisoner' s Dilemma. In the first round, participants were restricted to a cooperative choice, while their partners either cooperated or defected. Each partner' s behavioral strategy remained consistent across both rounds. In the second round, participants encountered the same partners from the first round alongside several new partners; during this stage, participants were required to autonomously decide whether or not to cooperate with each individual partner.

### 2.1.1 Participants

The sample size was determined using G\*Power 3.1. For a two-factor mixed-design analysis of variance (ANOVA), with Social Value Orientation (SVO: prosocial vs. individualistic) as the between-subjects variable and partner behavior (cooperation vs. betrayal) as the within-subjects variable, the power analysis indicated that a total sample size of at least 38 participants was required to achieve a statistical power of 85% at a significance level of  $\alpha = 0.05$  and a medium effect size ( $f = 0.25$ ). A total of 42 undergraduate and graduate students were recruited from a university. Two participants were excluded because their performance on the old/new memory recognition task was below the chance level (0.5).

Ultimately, 40 participants were retained (24 females), with a mean age of  $21.23 \pm 2.42$  years ( $M \pm SD$ ). This sample included 20 prosocial individuals ( $21.3 \pm 2.15$  years, 14 females) and 20 individualistic individuals ( $21.15 \pm 2.67$  years, 10 females). To ensure comparability between the two groups, statistical tests were conducted on the primary demographic variables. The results showed no significant difference in age between the two groups ( $t = 0.19, p = 0.850$ ), and a chi-square test revealed no significant difference in gender distribution ( $t = -0.06, p = 0.950$ ).

One to three days prior to the formal experiment, participants were screened and categorized into prosocial and individualistic groups using the online Social

Value Orientation Slider Measure [?, ?]. This method has been validated in Chinese samples and demonstrates high reliability and validity [?, ?, ?]. The SVO Slider Measure consists of six items, in which participants select their most preferred option from nine different resource allocation ratios between themselves and another person. The SVO angle ( $SVO^\circ$ ) is calculated by taking the inverse tangent of the ratio between the mean score for oneself and the mean score for the other person. Following the classification criteria suggested by Murphy et al. [?] and subsequent studies [?, ?, ?], individuals with an  $SVO^\circ$  score greater than  $22.45^\circ$  were classified as prosocial, while those with an  $SVO^\circ$  score less than or equal to  $22.45^\circ$  were classified as individualistic.

### 2.1.2 Experimental Design

A  $2 \times 2$  two-factor mixed experimental design was employed, with Social Value Orientation (SVO: pro-social vs. pro-self) and partner behavior (cooperation vs. betrayal) serving as the independent variables. The primary dependent variables were source memory performance and decision-making outcomes. Specifically, source memory referred to the participants' ability to remember partner faces and their corresponding behaviors. Utilizing a sequential Prisoner's Dilemma paradigm, the experiment consisted of three distinct phases: the encoding phase (the first round of the game), an unexpected memory recall test phase, and the decision-making phase (the second round of the game).

### 2.1.3 Experimental Materials and Equipment

A total of 180 neutral-expression images of young faces (90 male and 90 female) were selected from the CAS-PEAL-R1 face database [?, ?]. Twenty-one participants (5 males) were recruited to rate these images on a 5-point scale for both facial attractiveness and facial trustworthiness. Based on these ratings, 72 face images (36 male and 36 female) with moderate scores in both attractiveness and trustworthiness were selected as the final experimental stimuli.

The experiment was programmed using E-prime 3.0 on the Windows 10 operating system. All facial stimuli were presented frontally against a white background. Stimuli were displayed on a 27-inch CRT monitor with a resolution of 1920 x 1080 and a refresh rate of 120 Hz. During the experiment, participants were positioned approximately 80 cm from the screen.

**The experimental procedure was divided into an encoding phase, an unexpected recall test phase, and a decision-making phase. The specific flow is shown in Figure 1.**

The experimental paradigm for memory research was adapted from previous studies on reciprocal cooperation [?, ?, ?].

During the instruction phase, participants were informed that in the initial encoding phase of the sequential Prisoner's Dilemma game, they were restricted

to choosing “cooperate” and investing 30 RMB. They were told that their partners could choose either to cooperate or to defect. Participants were clearly informed that clicking “cooperate” meant investing 30 RMB in the other party. If the partner also chose to cooperate, they would invest the same amount as the participant; the resulting bonus was always one-third of the total investment, meaning both the participant and the partner would receive 10 RMB. In the condition where the partner chose not to cooperate (defect), the partner invested 0 RMB, resulting in the partner gaining 20 RMB while the participant lost 10 RMB. In the final decision-making phase, participants could choose whether or not to cooperate with the partner, with the understanding that each partner would behave consistently across both the encoding and decision-making phases [?, ?]. Operationally, “non-cooperative” behavior in the experimental procedure was equivalent to “defection.” Before the experiment, participants were informed that their earnings in the encoding phase would determine 30% of their participation fee, while their earnings in the decision-making phase would determine the remaining 70% [?, ?].

In the encoding phase, participants played the game with 48 partners, whose faces were presented randomly one by one across 48 trials. Each trial began with a fixation point presented for 1s, followed by the presentation of a partner’s face and the question: “Do you want to invest 30 RMB to cooperate with this player?” At this stage, participants were required to choose to cooperate and invest by pressing a key. After the keypress, the text “Your decision: Invest 30 RMB to cooperate with this player” appeared for 3s. This was followed by the text “The player’s decision: …” in red font for 1s, which then updated to “The player’s decision: Cooperate/Not cooperate” for 3s. Finally, the amount the participant received was displayed in black font for 3s: if the partner cooperated, “You gain: 10 RMB” was shown; if the partner did not cooperate, “You lose: -10 RMB” was shown. Before the formal experiment, participants completed four practice trials to ensure full comprehension of the instructions.

Following a break, participants performed a surprise recall test. The 48 old faces (24 cooperative partners and 24 defecting partners) were randomly mixed with 24 new partner faces. Participants first performed an old/new recognition task, followed by a confidence judgment regarding their recall (on a scale of 0–10). If a face was judged as “new,” no further recall regarding that partner’s behavior was required. If a face was judged as “old,” participants were asked to recall whether the partner had cooperated in the previous game to measure source memory (partner behavior), followed by a confidence judgment for the source memory.

In the decision-making phase, the 72 partner faces from the recall test phase were presented randomly in sequence, and participants had to decide whether to cooperate with each partner. Before entering this phase, the rules were clarified: if the participant chose to cooperate, cooperating with a previously cooperative partner would result in a 10 RMB gain for both parties, while cooperating with a previously defecting partner would result in a 10 RMB loss for the participant

and a 20 RMB gain for the partner. If the participant chose not to cooperate and the partner chose to cooperate, the partner would lose 10 RMB while the participant would have no gain or loss. If both parties chose not to cooperate, neither would gain or lose anything. For new partners, the probability of cooperation was set at 50%. During this phase, participants received no feedback regarding the partner's behavior after making their decisions.

## 2.2 Experimental Results

[Figure 1: see original paper]

Figure 1 illustrates the experimental flow for Study 1. Data conversion and exportation were performed using E-prime 3.0, while all subsequent statistical analyses were conducted using SPSS 26.0.

### 2.2.1 Cooperation Rate in the Decision Phase

Following previous research [?, ?], we conducted a one-way analysis of variance (ANOVA) on cooperation rates across different partner types to verify the effectiveness of the experimental manipulation—specifically, whether the experience from the first round of the game influenced decision-making in the second round. The cooperation rate was defined as the proportion of trials in which the participant chose to cooperate out of the total number of decisions. As shown in [Figure 2: see original paper], the results revealed a significant main effect of partner type,  $F(1, 38) = 33.06, p < 0.001, \eta_p^2 = 0.46$ . Participants exhibited a significantly higher cooperation rate when facing cooperative partners ( $M \pm SD = 0.51 \pm 0.17$ ) compared to defecting partners ( $M \pm SD = 0.34 \pm 0.17$ ),  $t(39) = 6.20, p < 0.001$ , Cohen's  $d = 0.98$ , 95% CI = [0.11, 0.22]. Furthermore, participants were more inclined to cooperate with cooperative partners than with new partners ( $M \pm SD = 0.29 \pm 0.25$ ),  $t(39) = 5.75, p < 0.001$ , Cohen's  $d = 0.91$ , 95% CI = [0.14, 0.30]. However, the difference in cooperation rates between defecting partners and new partners was not significant ( $t(39) = 1.40, p = 0.170$ ). These results indicate that during the encoding phase, participants successfully formed source memories of their partners' behaviors, which subsequently influenced their cooperative choices during the decision-making phase. Additionally, participants followed a cautious cooperation strategy in the Prisoner's Dilemma game, selectively cooperating with previously cooperative partners while avoiding cooperation with both defectors and new partners.

[Figure 2: see original paper] Mean cooperation rates for each partner type during the decision-making phase. Note:  $p < 0.001$ ,  $p < 0.01$ ,  $p < 0.05$ , n.s. indicates  $p > 0.05$ . Error bars represent standard error (SE). The same applies below.

### 2.2.2 Source Memory Accuracy

Based on their behavior (cooperation or betrayal), the partners encountered by the participants were categorized into two types: cooperative partners and betraying partners. In the surprise recognition test, the accuracy of old/new judgments was defined as the ratio of the number of correctly identified old and new faces to the total number of faces presented.

The performance metric for source memory was the source memory accuracy rate, calculated as the ratio of the number of trials where the partner's source was correctly identified to the total number of old partners presented, specifically for those trials where the face was correctly recognized as "old."

The accuracy rates for old/new judgments regarding cooperative, betraying, and new partners across different Social Value Orientation (SVO) types during the testing phase are presented in . To investigate whether SVO and partner behavior influenced source memory, a two-way mixed-design analysis of variance (ANOVA) was conducted with SVO and partner type as independent variables. The results revealed that neither the main effect of SVO ( $F(1, 38) = 0.81, p = 0.382$ ) nor the main effect of partner behavior ( $F(1, 38) = 0.18, p = 0.894$ ) was significant. However, a significant interaction between the two was observed (see [Figure 3: see original paper]),  $F(1, 38) = 15.95, p < 0.001, \eta_p^2 = 0.30$ . Further simple effects analysis indicated that for pro-social individuals, source memory for cooperative partners was significantly higher than for betraying partners ( $F(1, 38) = 7.45, p = 0.010, \eta_p^2 = 0.16$ ). Conversely, for pro-self individuals, source memory for betraying partners was significantly higher than for cooperative partners ( $F(1, 38) = 8.52, p = 0.006, \eta_p^2 = 0.18$ ). Descriptive statistics for source memory performance during the recall phase are summarized in . These results suggest that SVO modulates source memory in reciprocal cooperation, with individuals of different SVOs exhibiting inconsistent memory advantages.

[Figure 3: see original paper] illustrates the source memory accuracy for cooperative and betraying partners across different social value orientations. To more intuitively represent the accuracy of source memory under different conditions, one-sample t-tests were performed against the chance level (0.50). The results showed that for pro-social individuals, the source memory accuracy for cooperative partners was significantly higher than chance ( $t(19) = 4.47, p < 0.001$ , Cohen's  $d = 1.00$ , 95% CI = [0.58, 0.72]), while there was no significant difference from chance for betraying partners ( $t(19) = 0.00, p = 0.500$ ). For pro-self individuals, the source memory accuracy for cooperative partners did not differ significantly from chance ( $t(19) = 0.00, p = 0.500$ ), whereas the source memory accuracy for betraying partners was significantly higher than chance ( $t(19) = 3.95, p < 0.001$ , Cohen's  $d = 0.88$ , 95% CI = [0.57, 0.73]).

presents the means and standard deviations of source memory accuracy and confidence judgments under different social value orientations in Experiment 1. For the pro-social orientation, the source memory accuracy and confidence judgments were  $0.65 \pm 0.15$  and  $7.41 \pm 1.03$  for cooperative partners, and  $0.50 \pm$

0.17 and  $7.29 \pm 0.91$  for betraying partners. For the pro-self orientation, the values were  $0.50 \pm 0.18$  and  $7.74 \pm 1.33$  for cooperative partners, and  $0.65 \pm 0.17$  and  $7.39 \pm 1.36$  for betraying partners.

### 2.2.3 Source Memory Confidence

To investigate individual memory confidence regarding cooperative and defecting partners during reciprocal cooperation, a two-way mixed-design Analysis of Variance (ANOVA) was conducted on source memory confidence judgments, using Social Value Orientation (SVO) and partner type as the independent variables.

**The results showed that in the confidence judgments of source memory, neither the main effect of SVO ( $F(1, 38) = 0.03, p = 0.855$ ) nor the main effect of partner behavior was significant.**

The main effect of behavior was not significant ( $F(1, 38) = 0.62, p = 0.434$ ), nor was the interaction between the two variables ( $F(1, 38) = 0.11, p = 0.747$ ). Descriptive statistics for the source memory confidence judgments during the recall test phase are presented in . These results indicate that the impact of Social Value Orientation (SVO) on source memory performance is inconsistent with its impact on confidence (see Figure 3 and [Figure 4: see original paper]). Specifically, both pro-social and pro-self individuals believed they remembered cooperative partners and betraying partners equally well. This finding suggests that, at a subjective level, individuals do not exhibit a memory preference for either cooperative or betraying partners.

Figure 4: Source memory confidence scores for cooperative and betraying partners across different social value orientations.

### 2.2.4 Influence of Source Memory on Decision-Making

To further investigate the influence of source memory on decision-making, a linear regression analysis was conducted with source memory accuracy as the independent variable and decision-making accuracy as the dependent variable. The results indicated that the regression model was significant,  $R^2 = 0.43$ , adjusted  $R^2 = 0.40, F(4, 35) = 13.94, p < 0.001$ . The detailed results of the regression analysis are presented in . These findings demonstrate that source memory significantly predicts decision-making outcomes: higher source memory accuracy was associated with a greater frequency of adaptive decision-making (i.e., more frequent cooperation with cooperators and less frequent cooperation with defectors). This suggests that source memory plays a crucial role in guiding individuals' future cooperative decisions.

Table 3: Regression Analysis Results for Decision-Making Accuracy: Unstandardized Coefficients (B) and Standard Error of B

### 2.3 Discussion of Experiment 1 Results

Standardized Regression Coefficient ( $\beta$ ) Significance ( $p$ ) < 0.001\*\*\*

**Experiment 1 found that SVO modulates the direction of source memory advantage in reciprocal cooperation. Pro-self individuals exhibited a memory advantage for betraying partners.**

Specifically, pro-self individuals exhibited superior source memory for betrayers compared to cooperators, whereas pro-social individuals showed better source memory for cooperators than for betrayers. This discrepancy may stem from differences in how individuals with different Social Value Orientations (SVO) evaluate the outcomes of partner behavior during reciprocal cooperation. Pro-self individuals are more concerned with their own interests [?, ?] and are therefore more likely to expect others to cooperate to maximize their own gains; consequently, they demonstrate enhanced source memory for betrayals that violate these expectations. Furthermore, pro-self individuals exhibit higher sensitivity and adaptive processing tendencies toward negative information [?, ?], leading them to prioritize remembering betrayers to avoid future losses. In contrast, pro-social individuals place greater emphasis on moral evaluation [?, ?] and are more inclined to maintain positive social connections and trust to establish long-term reciprocal relationships [?, ?]. They are thus more likely to remember cooperative partners to facilitate continued cooperation in subsequent interactions [?, ?].

These findings suggest that individuals do not possess a universal source memory preference for betrayers, which contradicts the “cheater detection hypothesis” [?, ?, ?]. Simultaneously, the distinct memory preferences observed across different SVO types refute previous conclusions suggesting that individuals show no difference in source memory for cooperators versus betrayers during reciprocal interactions [?, ?, ?]. The results of the present study indicate that source memory in reciprocal cooperation can favor either cooperators or betrayers depending on the observer’s orientation [?, ?]. Furthermore, consistent with prior research [?, ?], source memory assists individuals in making adaptive decisions—approaching cooperators and avoiding betrayers—thereby highlighting the adaptive significance of source memory within social cooperation.

To further verify the reliability of the results from Experiment 1 and to explore the differences in the outcome evaluation processes through which SVO influences source memory in reciprocal contexts, Experiment 2 of this study utilizes electroencephalography (EEG) technology. This approach aims to investigate the electrophysiological mechanisms underlying the impact of SVO on source memory during reciprocal cooperation.

## Experiment 2 also included encoding, recall testing, and decision-making phases, with EEG activity recorded throughout to validate the findings.

Building upon the results of Experiment 1, this study further investigates the differences in memory mechanisms regarding cooperative and defecting partners among individuals with different Social Value Orientations (SVO) during reciprocal cooperation. By employing time-domain and time-frequency analysis methods, we explore the neural disparities in how various SVO individuals process feedback from partner behavior. This approach aims to reveal the electrophysiological mechanisms through which SVO influences episodic memory within the context of reciprocal cooperation.

### 3.1.1 Participants

The sample size was determined using G\*Power 3.1. For a two-factor mixed-design repeated measures ANOVA—with Social Value Orientation (SVO: pro-social vs. pro-self) as the between-subjects variable and partner behavior (cooperation vs. betrayal) as the within-subjects variable—a total sample size of at least 38 was required to achieve 85% statistical power at a significance level of  $\alpha = 0.05$  and a medium effect size ( $f = 0.25$ ). We recruited 46 undergraduate and graduate students from Northeast Normal University. One participant was excluded because their accuracy in the old/new recognition task was below 0.50. Consequently, valid behavioral data were retained for 45 participants (mean age:  $21.46 \pm 2.3$  years; 10 males), consisting of 22 prosocial individuals and 23 pro-self individuals. There were no significant differences between the two groups in terms of age ( $t = -0.06, p = 0.950$ ) or gender distribution ( $\chi^2 = 2.04, p = 0.284$ ).

One to three days prior to the formal experiment, participants were screened using the online Social Value Orientation Slider Measure [?, ?]. The classification method for social value orientation was identical to that used in Experiment 1.

### 3.1.2 Experimental Design

A  $2 \times 2$  mixed experimental design was employed, utilizing Social Value Orientation (SVO: pro-social vs. pro-self) and partner behavior (cooperation vs. betrayal) as independent variables. The study investigated the dynamic processing of outcome feedback and its underlying mechanisms affecting memory among individuals with different SVOs within reciprocal cooperation contexts.

To achieve this, the Feedback-Related Negativity (FRN) component, which reflects outcome evaluation, and the Late Positive Potential (LPP) component, associated with outcome feedback processing, were utilized as primary temporal indicators. Furthermore, the study focused on theta oscillations—a critical electroencephalographic (EEG) component induced during feedback processing—to provide a comprehensive analysis of the neural dynamics involved.

### 3.1.3 Experimental Materials and Instruments

One hundred and fifty facial images (75 male and 75 female) characterized by moderate levels of attractiveness and trustworthiness were selected from the assessment materials used in Experiment 1. Throughout the experiment, all facial stimuli were presented frontally against a black background. The experimental procedure was programmed using E-Prime 3.0 on the Windows 10 operating system. Stimuli were displayed on a 27-inch CRT monitor with a resolution of 1920x1080 and a refresh rate of 120 Hz. Participants completed the experiment in a dimly lit, soundproof room, maintaining a viewing distance of approximately 80 cm from the computer screen.

**The experiment was divided into an encoding phase, an unexpected recall test phase, and a decision-making phase (see Figure 5).**

During the encoding phase, participants engaged in two repeated rounds of the Prisoner's Dilemma game to enhance memory retention [?, ?]. Each round consisted of 100 trials; while the faces of the players and their associated behaviors remained consistent across both rounds, the order of presentation was randomized. At the beginning of each trial, a fixation point was displayed for a random interval of 0.7-1 s. This was followed by the presentation of a partner's face alongside the question, "Do you want to cooperate with this player?" During this stage, participants were required to choose to cooperate and invest by pressing a key. Subsequently, the text was replaced by a prompt in red font stating, "This player's decision: ...". After 0.7-1 s, the partner's choice—either "Cooperate" or "Defect"—was displayed for 2 s. Finally, the outcome was shown in black font for 1 s: if the partner chose to cooperate, the screen displayed "You gain: 10 yuan"; if the partner chose to defect, it displayed "You lose: -10 yuan." Before the formal experiment, participants completed four practice trials to ensure a full understanding of the instructions.

Following a rest period, participants underwent an unannounced memory recall test. In this phase, 100 "old" faces (consisting of 50 cooperative partners and 50 defecting partners) were randomly mixed with 50 "new" partner faces. Each trial began with a fixation point presented for a random duration of 0.7-1 s, after which participants first performed an old/new recognition task. If a face was judged as "new," no further recall of the partner's previous behavior was required. If a face was judged as "old," participants were asked to recall whether the partner had cooperated or defected in the previous game to measure source memory.

The decision-making phase followed the same procedure as Experiment 1. After a fixation point was presented for a random interval of 0.7-1 s, participants chose whether to cooperate with or defect against the partner. No feedback regarding the partner's behavior was provided after decisions in this phase. The facial stimuli used in this stage consisted of the 150 face images previously presented

during the testing phase.

### 3.1.5 ERPs Recording and Analysis

[Figure 5: see original paper] illustrates the experimental procedure for Experiment 2. ERP data were recorded using a 64-channel Neuroscan ESI-64 system, with electrodes positioned according to the international 10-20 system. During recording, REF served as the online reference and GND as the ground. Impedance for all electrodes was maintained below 10 k $\Omega$ . The continuous EEG signal was recorded with a band-pass filter of 0.05–100 Hz and a sampling rate of 1000 Hz. Behavioral data were collected simultaneously with the EEG signals.

EEG data were analyzed using Matlab R2021b, and preprocessing was conducted using the EEGLAB 2021.0 toolbox.

A band-pass filter of 0.1–40 Hz was applied to the data. For time-domain analysis, the data were segmented into epochs from -200 to 1000 ms relative to the onset of the target stimulus, with baseline correction performed using the average amplitude of the -200 to 0 ms interval. For time-frequency analysis, epochs were extracted from -1000 to 2000 ms, with baseline correction performed using the average amplitude of the -1000 to 0 ms interval.

Independent Component Analysis (ICA) was employed to identify and remove artifacts related to eye blinks and muscle movements [?, ?]. Trials with amplitudes exceeding  $\pm 100\mu V$  were excluded from further analysis. The data were re-referenced to the average of the bilateral mastoids (M1 and M2). Finally, the EEG data were superimposed and averaged for time-domain analysis.

Time-frequency analysis was performed using the Short-Time Fourier Transform (STFT) with a 200 ms Hanning window. For each trial, STFT was applied across the -1000 to 2000 ms time range in 10 ms steps (padding ratio 2:1) for frequencies between 1 and 30 Hz at 1 Hz intervals. The resulting spectrogram represents the joint function of signal power across time and frequency at each point, defined as  $P(t, f) = |f(t, f)|^2$ , which encompasses both phase-locked and non-phase-locked brain responses [?, ?, ?]. Following the methodology of Cohen et al. [?] and Marco-Pallares et al. [?], mean baseline correction was applied by subtracting the average power of the -500 to -200 ms baseline period from the power at each time-frequency point (units:  $\mu V^2/Hz$ ). To control for the False Discovery Rate (FDR) in multiple comparisons, the Benjamini-Hochberg procedure was applied to the p-values of all time-frequency points.

The primary ERP components of interest in this study were the FRN and LPP. Based on prior research and the current dataset, electrodes FCz and Fz in the frontal region were selected for FRN analysis, while Pz and POz in the parietal region were selected for LPP analysis [?, ?, ?]. Consistent with previous literature and observed data characteristics, the time window for the FRN was set at 200–250 ms [?, ?]. The LPP window was set at 400–1000 ms and further divided into two phases: early LPP (400–700 ms) and late LPP (700–1000 ms)

[?, ?]. Finally, based on event-related spectral perturbation (ERSP) maps and relevant studies [?, ?, ?, ?], time-frequency analysis was conducted on the theta (4-8 Hz) band at the Fz electrode during the 200-450 ms encoding stage.

### 3.2 Behavioral Results

Behavioral data were transformed using E-Prime 3.0, and statistical analyses were subsequently performed using SPSS 26.0.

#### 3.2.1 Cooperation Rate in the Decision Phase

Consistent with Experiment 1, a one-way analysis of variance (ANOVA) was conducted to examine whether experience from the first round of the game influenced decision-making in the second round across different partner types. The results (see [Figure 6: see original paper]) revealed a significant main effect of partner type on cooperation rates,  $F(2, 43) = 27.87, p < 0.001, \eta_p^2 = 0.57$ . Participants made significantly more cooperative decisions when facing a cooperative partner ( $M \pm SD = 0.55 \pm 0.02$ ) compared to a defecting partner ( $M \pm SD = 0.34 \pm 0.02; t(44) = 7.50, p < 0.001, \text{Cohen's } d = 1.12, 95\% \text{ CI} = [0.15, 0.26]$ ).

Furthermore, participants were more likely to cooperate with a previous cooperative partner than with a new partner ( $M \pm SD = 0.43 \pm 0.04; t(44) = 3.32, p = 0.002, \text{Cohen's } d = 0.49, 95\% \text{ CI} = [0.04, 0.18]$ ). Conversely, the probability of cooperation was significantly lower when facing a defecting partner compared to a new partner ( $t(44) = 2.47, p = 0.018, \text{Cohen's } d = 0.37, 95\% \text{ CI} = [0.02, 0.16]$ ). These findings suggest that the sequential Prisoner's Dilemma during the encoding phase allowed participants to develop source memory regarding their partners and their associated behaviors, which subsequently influenced their decision-making. Specifically, during the decision phase, participants selectively chose to cooperate with previous cooperators while actively avoiding cooperation with previous defectors.

#### 3.2.2 Source Memory Accuracy

[Figure 6: see original paper] Cooperation rates for each partner type during the decision-making phase. The calculation methods for old/new recognition accuracy and source memory accuracy were consistent with those used in Experiment 1. The old/new recognition accuracy for each SVO type during the testing phase is presented in .

Means and standard deviations of old/new recognition accuracy for different social value orientations in Experiment 2. To investigate whether SVO and partner behavior influenced source memory, a two-way mixed-design repeated measures ANOVA was conducted on source memory performance, with SVO and partner type as independent variables. The results (see [Figure 7: see original paper]) revealed a significant interaction between SVO and partner behavior,  $F(1, 43) = 27.11, p < 0.001, \eta_p^2 = 0.387$ . Simple effects analysis

indicated that for pro-social individuals, the source memory accuracy for cooperative partners was significantly higher than that for betraying partners,  $F(1, 43) = 4.92, p = 0.032, \eta_p^2 = 0.10$ . Conversely, for pro-self individuals, the source memory accuracy for betraying partners was significantly higher than that for cooperative partners,  $F(1, 43) = 26.80, p < 0.001, \eta_p^2 = 0.38$ . Descriptive statistics for source memory accuracy are provided in .

Following the procedure of Experiment 1, source memory accuracy was compared against chance level (0.50) using one-sample t-tests. The results for pro-social individuals [Figure 7: see original paper] showed no significant differences from chance for either cooperative partner source memory ( $t(21) = 0.78, p = 0.222$ ) or betraying partner source memory ( $t(21) = -0.63, p = 0.733$ ). For pro-self individuals, source memory accuracy for cooperative partners did not differ significantly from chance ( $t(22) = -1.28, p = 0.892$ ), whereas source memory accuracy for betraying partners was significantly higher than chance ( $t(22) = 2.24, p = 0.018$ , Cohen's  $d = 0.47$ , 95% CI = [0.50, 0.64]).

In summary, compared to Experiment 1, the overall source memory accuracy of pro-social individuals in Experiment 2 was close to chance level. However, individuals with different social value orientations still exhibited significant relative differences between the cooperation and betrayal conditions.

Means and standard deviations of source memory recall accuracy under different social value orientations. Pro-social orientation:  $0.53 \pm 0.18$  (Cooperate),  $0.48 \pm 0.15$  (Betray); Pro-self orientation:  $0.46 \pm 0.15$  (Cooperate),  $0.57 \pm 0.15$  (Betray).

### 3.2.3 Influence of Source Memory on Decision-Making

To investigate the influence of source memory on decision-making, a linear regression analysis was conducted with source memory accuracy as the predictor for decision-making accuracy. The results indicated that the regression model was significant,  $R^2 = 0.296$ , adjusted  $R^2 = 0.263$ ,  $F(2, 43) = 9.04, p < 0.001$ .

**The analysis results are shown in Table 6. Similar to Experiment 1, the results indicate that source memory significantly predicted decision outcomes.**

Table 6 presents the results of the regression analysis for decision accuracy, detailing the non-standardized regression coefficients (B) and the standard errors of B.

### 3.2.4 Reproducibility Verification of Behavioral Data

Standardized regression coefficients (Beta) were significant at the  $p < 0.001^{***}$  level. The field of psychology has long faced skepticism regarding the reproducibility of research findings [?, ?, ?]. In the behavioral results of the present study, Experiment 2 demonstrated consistency with Experiment 1 regarding

the influence of Social Value Orientation (SVO) on source memory; specifically, SVO moderated individuals' source memory for cooperators and defectors within reciprocal cooperation contexts. Furthermore, the relationship between source memory and decision-making outcomes remained consistent with Experiment 1, indicating that source memory guides individuals in making adaptive decisions during reciprocal interactions.

These behavioral findings exhibit consistency and stability across homogeneous participant groups. The alignment of behavioral data analysis between Experiment 1 and Experiment 2 ensures the stability of SVO's influence on source memory, thereby providing a reliable behavioral foundation for the subsequent electroencephalogram (EEG) research.

### 3.3 EEG Data Results

EEG data were analyzed using Matlab R2021b. Three participants were excluded because they had fewer than 20 valid analysis trials. Consequently, 42 participants were retained for the final time-domain analysis of the EEG data, consisting of 21 individuals in the Prosocial Value Orientation (SVO) group and 21 individuals in the Pro-self Value Orientation group. Time-domain and time-frequency analyses were performed on the EEG data collected during the encoding stage.

**(1) FRN Amplitude Analysis** For the Feedback-Related Negativity (FRN) amplitude, the mean amplitudes for each condition were calculated within the 200–250 ms time window at the Fz and FCz electrode sites. A 2 (SVO: Prosocial, Pro-self)  $\times$  2 (Partner Behavior: Cooperation, Betrayal) mixed-design analysis of variance (ANOVA) was conducted.

The results indicated that the main effect of SVO was not significant ( $F(1, 40) = 0.57, p = 0.813$ ), nor was the main effect of Partner Behavior ( $F(1, 40) = 0.70, p = 0.792$ ). However, a significant interaction between SVO and Partner Behavior was observed,  $F(1, 40) = 11.79, p < 0.001, \eta_p^2 = 0.23$ . Simple effects analysis revealed...

**Analysis found that among prosocial individuals, compared to partner betrayal behavior ( $-3.37 \pm 0.97 \mu\text{V}$ ), partner cooperation behavior elicited a more negative FRN.**

( $-5.19 \pm 1.17 \mu\text{V}$ ) elicited a significantly more negative FRN amplitude compared to cooperative behavior ( $-3.91 \pm 1.50 \mu\text{V}$ ),  $F(1, 40) = 6.84, p = 0.013, \eta_p^2 = 0.15$ . Similarly, among pro-self individuals, partner betrayal ( $-5.48 \pm 1.43 \mu\text{V}$ ) elicited a significantly more negative FRN amplitude than partner cooperation ( $-3.91 \pm 1.50 \mu\text{V}$ ),  $F(1, 40) = 5.02, p = 0.031, \eta_p^2 = 0.11$ . The FRN waveforms, topographic maps, and difference maps during the encoding stage are presented in [Figure 8: see original paper] and [Figure 9: see original paper].

[Figure 8: see original paper] FRN waveforms for different SVOs under cooperation and betrayal feedback during the encoding stage. [Figure 9: see original paper] FRN topographic maps (a) and difference maps (b) (cooperation minus betrayal) for different SVOs during the encoding stage.

**(2) LPP Amplitude Analysis** A 2 (SVO: pro-social, pro-self)  $\times$  2 (Partner Behavior: cooperation, betrayal) mixed-design ANOVA was conducted on the mean amplitudes at the Pz and POz electrodes within the 400–700 ms (early LPP) and 700–1000 ms (late LPP) time windows. For the early LPP, the main effect of SVO was not significant ( $F(1, 40) = 0.75, p = 0.392$ ). However, the main effect of partner behavior was significant,  $F(1, 40) = 4.86, p = 0.033, \eta_p^2 = 0.11$ . A significant interaction between SVO and partner behavior was also observed,  $F(1, 40) = 4.30, p = 0.045, \eta_p^2 = 0.10$ . Further simple effects analysis revealed that only in pro-self individuals did partner betrayal ( $11.58 \pm 1.58 \mu\text{V}$ ) elicit a significantly more positive LPP amplitude than partner cooperation ( $9.45 \pm 1.53 \mu\text{V}$ ),  $F(1, 40) = 9.15, p = 0.004, \eta_p^2 = 0.19$ . In contrast, no significant difference in LPP amplitude was found for pro-social individuals (cooperation:  $8.89 \pm 0.97 \mu\text{V}$ ; betrayal:  $8.96 \pm 1.21 \mu\text{V}$ ),  $F(1, 40) = 0.01, p = 0.927$ . Regarding the late LPP, neither the main effect of SVO ( $F(1, 40) = 1.00, p = 0.324$ ) nor partner behavior ( $F(1, 40) = 2.50, p = 0.122$ ) reached significance, and there was no significant interaction between the two factors,  $F(1, 40) = 2.54, p = 0.119$ .

In summary, the early LPP amplitude following betrayal feedback was significantly more positive than that following cooperation feedback only among pro-self individuals. The LPP waveforms, topographic maps, and difference maps during the encoding stage are shown in [Figure 10: see original paper] and [Figure 11: see original paper].

[Figure 10: see original paper] LPP waveforms for different SVOs under cooperation and betrayal feedback during the encoding stage. [Figure 11: see original paper] LPP topographic maps (a) and difference maps (b) (cooperation minus betrayal) for different SVOs during the encoding stage.

**(3) Theta Oscillations** To analyze the temporal and spectral characteristics of the EEG signals induced by different partner behaviors across SVO types, a time-frequency analysis was performed on the encoding stage data. A repeated-measures ANOVA was conducted on the mean power of theta waves (4–8 Hz) at the mid-frontal Fz electrode within a 200–450 ms time window. The results showed that the main effect of SVO was not significant,  $F(1, 40) = 0.69, p = 0.410$ . However, the main effect of partner behavior was significant,  $F(1, 40) = 11.14, p = 0.002, \eta_p^2 = 0.21$ . Furthermore, a significant interaction between SVO and partner behavior was found,  $F(1, 40) = 5.74, p = 0.021, \eta_p^2 = 0.12$ . Simple effects analysis indicated that only in pro-self individuals did partner betrayal ( $M \pm SD = 1.09 \pm 0.88 \mu\text{V}$ ) elicit significantly stronger theta oscillations than partner cooperation ( $M \pm SD = 0.69 \pm 0.54 \mu\text{V}$ ),  $F(1, 40) = 16.05, p < 0.001, \eta_p^2 = 0.28$ . For pro-social individuals, there was no significant difference in

theta oscillations between cooperation ( $M \pm SD = 0.70 \pm 0.58 \mu V$ ) and betrayal ( $M \pm SD = 0.76 \pm 0.60 \mu V$ ),  $F(1, 40) = 0.45, p = 0.50$ . The theta oscillation frequency plots and topographic maps are presented in [Figure 12: see original paper].

[Figure 12: see original paper] Theta oscillation frequency plots (a) and topographic maps (b) for different SVOs under cooperation and betrayal feedback during the encoding stage.

### 3.4 Correlation Analysis Between EEG and Behavioral Results in the Encoding Phase

To examine the relationship between the outcome evaluation process during the encoding phase and source memory in reciprocal cooperation, we conducted a correlation analysis between the electroencephalogram (EEG) data and behavioral results from the encoding phase. Using Pearson correlation analysis, we investigated the relationship between FRN amplitude and source memory accuracy (see [Figure 13: see original paper]-a). The results revealed a significant negative correlation between FRN amplitude and source memory accuracy ( $r = -0.23, p = 0.034$ ), indicating that more negative FRN amplitudes were associated with better memory performance. In contrast, LPP amplitude did not show a significant correlation with source memory accuracy ( $r = 0.13, p = 0.237$ ). We further examined the correlation between theta oscillations and source memory (see Figure 13-b), finding a significant positive correlation between theta power and source memory accuracy ( $r = 0.27, p = 0.014$ ). These correlation results suggest that early-to-mid-stage outcome evaluation plays a critical role in the impact of Social Value Orientation (SVO) on source memory within reciprocal cooperation. Furthermore, theta oscillations during the encoding phase may facilitate the encoding process of source memory.

Figure 13 Correlation between FRN amplitude (a), theta oscillations (b), and source memory accuracy.

### 3.5 Discussion of Experiment 2 Results

Overall, the behavioral results of Experiment 2 were largely consistent with those of Experiment 1. A minor discrepancy was observed during the decision-making phase: while the difference in cooperation rates toward betraying partners versus new partners was not significant in Experiment 1, participants in Experiment 2 demonstrated a significantly lower cooperation rate toward betraying partners than toward new partners. This divergence may stem from procedural differences during the encoding phase of the two experiments. Compared to Experiment 1, Experiment 2 required participants to engage in two rounds of the game with their partners during the encoding phase. This likely facilitated a higher level of familiarity with the faces of previous partners, making them more distinguishable from new partners during the subsequent decision-making phase [?, ?]. Given that the cooperation probability for new partners was set

at 50%—a strategy superior to that of betraying partners but inferior to that of cooperative partners—the difference in cooperation rates between new and betraying partners reached statistical significance in Experiment 2.

Regarding the EEG results, Social Value Orientation (SVO) influenced both the early Feedback-Related Negativity (FRN) and the late Late Positive Potential (LPP) components, which reflect outcome evaluation during the encoding phase. Furthermore, SVO modulated theta oscillations sensitive to negative feedback. Correlation analyses between behavioral and EEG data revealed that outcome evaluation during the early and middle stages played a crucial role in how SVO influenced source memory within reciprocal cooperation. These findings suggest that SVO affects source memory by modulating the outcome evaluation process of feedback during reciprocal interactions.

## 4 General Discussion

This study focuses on the impact of Social Value Orientation (SVO) on source memory within reciprocal cooperation and its underlying electrophysiological mechanisms. By employing a sequential Prisoner's Dilemma paradigm integrated with an unexpected recall test during EEG recording, we investigated the source memory and electrophysiological differences between individuals with different SVOs when retrieving the behaviors of cooperative versus defecting partners.

The results of the study indicate that:

- 1) SVO modulates source memory in reciprocal cooperation. Specifically, pro-self individuals exhibited superior source memory for defecting partners compared to cooperative partners, whereas pro-social individuals showed better source memory for cooperative partners than for defecting ones.
- 2) SVO influences the outcome evaluation process of decision feedback. Furthermore, the early-to-mid stage of outcome evaluation plays a critical role in how SVO impacts source memory during reciprocal cooperation.

### 4.1 Source Memory as the Foundation for Reciprocal Cooperative Decision-Making

Source memory in episodic memory plays a critical role in the process of reciprocal cooperative decision-making, serving as an essential basis for individual judgment [?, ?]. The decision-making results from Experiment 1 and Experiment 2 of the present study demonstrate that source memory significantly predicts an individual's decision outcomes, with accurate source memory guiding individuals toward making adaptive decisions. Consistent with previous research [?, ?, ?, ?], this study reaffirms that source memory enables individuals to make better adaptive choices during reciprocal cooperation. Taken together, these findings indicate that source memory provides vital guidance during social interactions and reflects the adaptive significance of source memory in social

information processing.

#### **4.2 Differences in Source Memory Advantages Among Individuals with Different Social Value Orientations**

Previous research has yielded conflicting results regarding memory advantages for source memory in reciprocal cooperation. The present study introduces Social Value Orientation (SVO) as a personality variable to explain these discrepancies. In Experiment 1, we utilized a sequential Prisoner's Dilemma paradigm to analyze recall performance during the retrieval phase. The results indicate that SVO moderates source memory for the behaviors of cooperative and defecting partners. Specifically, pro-self individuals exhibited superior source memory for defecting partners compared to cooperative ones, whereas pro-social individuals showed better source memory for cooperative partners than for defectors. Early research proposed the "cheater detection hypothesis," suggesting that individuals allocate more cognitive resources to defecting behaviors [?, ?] and that defection holds a memory advantage over cooperation [?, ?, ?].

However, the results of this study demonstrate that individuals do not exclusively exhibit a memory advantage for defection; rather, they may show a memory advantage for either cooperative or defecting partners depending on their orientation. Furthermore, from the theoretical perspective of adaptive memory [?, ?, ?], source memory in reciprocal cooperation is not simply driven by the negativity of feedback. Instead, it adaptively focuses on information regarding both cooperators and defectors—both of which are useful for subsequent cooperative decision-making. This study further reveals that individuals with different SVOs exhibit distinct adaptive tendencies in social information processing: pro-social individuals focus more on cooperative information, while pro-self individuals focus more on defection. In summary, these findings provide a new perspective for understanding the memory mechanisms underlying reciprocal cooperation and emphasize the critical role of personality variables in memory processing [?, ?], establishing SVO as an important moderator of source memory advantages. Additionally, Experiment 2 provided a replication of the behavioral results from Experiment 1 [?, ?, ?]. The consistent findings across both experiments regarding the influence of SVO on source memory suggest that this effect is stable within homogeneous samples.

**Experiment 2 further explored the electrophysiological mechanisms of SVO's influence on source memory in reciprocal cooperation, finding that SVO affected individual outcome evaluation.**

#### **The Outcome Evaluation Process of Feedback for Cooperation and Defection in Reciprocal Cooperation**

In the study of reciprocal cooperation, the outcome evaluation process regarding feedback for cooperative or defective behavior is a critical mechanism for main-

taining social stability and promoting prosociality. Reciprocal cooperation relies on the ability of individuals to monitor the actions of others and adjust their subsequent behavior based on the perceived value and intent of those actions. When an individual receives feedback—whether it be a reward for cooperation or a sanction for defection—a complex cognitive and neural evaluation process is triggered to determine the significance of this feedback for future interactions.

The evaluation of cooperative outcomes typically involves the integration of social expectations and realized gains. When a partner chooses to cooperate, the feedback is processed as a positive reinforcement signal, strengthening the social bond and increasing the likelihood of future mutual cooperation. This process is not merely a calculation of material gain; it also encompasses the psychological satisfaction derived from mutual trust. Conversely, the evaluation of defection involves processing a violation of social norms. Feedback indicating that a partner has defected triggers a negative evaluation, often associated with feelings of betrayal or unfairness, which serves as a signal to withdraw cooperation or implement punitive measures to enforce reciprocity.

Furthermore, the feedback loop in reciprocal cooperation is influenced by the perceived intentionality behind the behavior. Outcome evaluation distinguishes between accidental failures to cooperate and deliberate defections. If feedback suggests that a negative outcome was unintentional, the evaluation process may mitigate the negative impact on the relationship. However, clear evidence of intentional defection leads to a robust devaluation of the partner's reliability. Ultimately, this continuous evaluation of feedback allows individuals to navigate complex social environments by identifying reliable partners and protecting themselves against exploitation, thereby sustaining the evolutionary viability of cooperation within a group.

**The early stage of outcome evaluation is reflected in the FRN amplitude [?, ?], where individuals perform a rapid evaluation of feedback valence.**

Evaluations are compared with subjective expectations. Pro-self individuals exhibited larger FRN amplitudes in response to betrayal feedback, whereas pro-social individuals showed larger FRN amplitudes following cooperative feedback, demonstrating opposite patterns of expectancy violation. According to existing research, pro-self individuals focus more on self-interest and expect others to cooperate to maximize their own gains [?, ?]; consequently, they exhibit a stronger negative outcome evaluation toward betrayal feedback that violates these expectations, leading to larger FRN amplitudes. However, contrary to the present results, some studies have found that pro-social individuals typically hold higher expectations for cooperation and exhibit more significant FRN effects when facing betrayal [?, ?, ?]. A possible reason for this discrepancy is that the FRN reflects not only expectancy violation but is also influenced by the emotional and motivational relevance of the feedback outcome [?, ?, ?]. Research indi-

cates that pro-social individuals place greater value on reciprocity during social interactions and possess stronger motivations to maintain long-term reciprocal relationships [?, ?, ?]. Furthermore, cooperation itself carries reward value, capable of activating reward-related brain regions and eliciting higher emotional responses [?, ?]. Pro-social individuals specifically exhibit higher reward sensitivity and emotional arousal toward cooperative behavior in social interaction contexts [?, ?, ?]. Therefore, during early outcome evaluation, pro-social individuals experience stronger emotional responses and motivational relevance toward cooperative feedback, thereby inducing larger FRN amplitudes.

In the late stage of outcome evaluation, the LPP reflects an individual's sustained emotional and motivational processing of feedback, as well as the allocation of attentional resources [?, ?]. Only pro-self individuals showed a significant enhancement in early LPP amplitude (400–700 ms) following betrayal feedback, suggesting that they invest more emotional and attentional resources into betrayal, which in turn enhances their memory preference for such information [?, ?]. In contrast, pro-social individuals focus more on long-term reciprocal relationships [?, ?], exhibit higher tolerance for short-term betrayal [?, ?, ?], and tend to employ emotion regulation to reduce the negative affect brought about by betrayal feedback [?, ?, ?]. Consequently, their emotional arousal under betrayal feedback is lower, the change in LPP amplitude is not significant, and they do not show a memory advantage for this condition.

Time-frequency analysis revealed that the enhancement effect of theta oscillations on negative feedback [?, ?] was present only in pro-self individuals, showing a significant difference between cooperative and betrayal feedback. This result may reflect that pro-self individuals are more likely to activate prefrontal regions for conflict detection and behavioral adjustment when faced with betrayal [?, ?]. Pro-social individuals, perhaps due to their pursuit of collective interests and the presence of emotional buffering mechanisms against betrayal, showed no significant difference in theta oscillations across different feedback types. Regarding the correspondence between behavioral and EEG results, Experiment 2 showed that different SVO individuals exhibited source memory differences between cooperative and betrayal conditions; however, the effect size for pro-self individuals ( $\eta_p^2 = 0.38$ ) was larger than that for pro-social individuals ( $\eta_p^2 = 0.10$ ) [?, ?]. EEG results similarly showed that pro-self individuals exhibited more sustained enhancement effects across FRN, LPP, and theta oscillations, while pro-social individuals only showed significant differences during the FRN stage. This suggests that the memory advantage pro-social individuals have for cooperative partners relies more on early selective processing [?, ?]. Furthermore, compared to the source memory pro-social individuals have for cooperative partners, pro-self individuals may possess a stronger memory preference for betrayers.

Additionally, the source memory of pro-social individuals for cooperative partners in Experiment 2 was close to chance level in absolute terms. This may be due to the more complex interaction process in Experiment 2, which increased the difficulty of the source memory task and reduced the stability of judgments

regarding the source of a partner' s behavior during the retrieval phase [?, ?]. EEG results also indicated that the enhanced processing of cooperative partners by pro-social individuals was primarily confined to the early stages and did not extend to late-stage processing, suggesting that their memory advantage is less stable under high-load contexts. In contrast, pro-self individuals exhibited more sustained emotional and cognitive enhancement effects under betrayal feedback, making their source memory advantage more stable across different experiments. In summary, although the source memory for cooperative partners among pro-social individuals in Experiment 2 was not significantly above chance level in absolute terms, the clear behavioral results from Experiment 1 combined with the EEG evidence from Experiment 2 suggest that Social Value Orientation (SVO) influences source memory preferences for cooperative and betraying partners by modulating the outcome evaluation process.

Regarding the selection of ERP components, although feedback processing research often focuses on the P300 [?, ?, ?], no

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv –Machine translation. Verify with original.*