eLife

Reviewed Preprint
v1 • June 12, 2026
Not revised

For correspondence:
dingxw3@mail.sysu.edu.cn
edsgao@mail.sysu.edu.cn

* These authors contributed equally to this work.

Competing interests: No competing interests declared

Funding: See page 18

Reviewing editor: Ryszard Auksztulewicz, Maastricht University, Netherlands

Betrayal is worse than loss during cooperation

Rumeng Tang[a,*], Jingbin Tan[a,*], Yi Gao[b], Chen Lin[a], Jing Gan[a], Xiaowei Ding[a] ✉, Dingguo Gao[a] ✉

[a]Guangdong Provincial Key Laboratory of Social Cognitive Neuroscience and Mental Health, and Department of Psychology, Sun Yat-sen University, Guangzhou, China • [b]Shenzhen Key Laboratory for Systems Medicine in Inflammatory Diseases, School of Medicine, Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University, Shenzhen, China

eLife Assessment

This **valuable** study examines whether reduced cooperation is driven by betrayal aversion beyond nonsocial loss aversion, using matched social and nonsocial risky decision-making tasks combined with computational modeling and EEG. The authors provide **solid** empirical evidence that social risk is processed differently from matched nonsocial risk, offering a meaningful contribution to the study of cooperation and decision-making under uncertainty. However, further justification of the computational modeling approach would strengthen some of the conclusions. This work will be of interest to researchers studying social decision-making, cooperation, trust, and the neural and computational mechanisms underlying risk and betrayal aversion.

https://doi.org/10.7554/eLife.111043.1.sa4

Abstract

Cooperative behavior is a cornerstone of human interaction. Although both “betrayal aversion” (the affective cost of being betrayed) and “loss aversion” (the financial detriment incurred from betrayal) are established determinants of cooperative behavior, their relative potency remains undetermined. Here, we investigated these effects by integrating computational modeling and event-related potential (ERP) techniques. In two tasks involving risk and cooperation, participants decided whether to take financial risks or to cooperate under possible betrayal. Our results showed that betrayal aversion had a stronger effect on reducing cooperation compared to loss aversion. Furthermore, ERP data demonstrated sequential processing: betrayal was encoded early in decision-making, reflected by increased P3 with weaker betrayal aversion, whereas loss aversion manifested later, marked by increased LPP. By dissociating the contributions of betrayal and loss, our finding provides novel insights into the cognitive and neural mechanisms underlying cooperative behavior.

1 Introduction

From collaborating on school projects to managing household responsibilities, humans often engage in cooperative acts that require personal costs (e.g., time, effort, money) for mutual benefit (Tomasello et al., 2005). Such cooperation is central to social life and has shaped human evolution (Nowak, 2006; West et al., 2021). Reciprocity theory views cooperation through the lens of a cost-benefit framework, where people cooperate because the future benefits outweigh the present costs (Nowak & Sigmund, 2005; Panchanathan & Boyd, 2004). It is widely known that betrayal, as an emotional cost in cooperation, provokes negative emotions and promotes punishment (Fehr & Gächter, 2000). While much prior research has only focused on “betrayal aversion”, less attention has been paid to the financial loss from unmet expectations during social interactions (Bohnet et al., 2008; Fehr & Fischbacher, 2003). Indeed, mutual cooperation maximizes social welfare, whereas betrayal benefits the trustee but comes at the trustor’s expense

in the Trust Game (Joyce et al., 1995). Here, we compare how "betrayal aversion" and "loss aversion" (i.e., avoiding cooperation to avoid incurring financial losses due to betrayal) differently shape cooperation. Understanding these distinctions is critical for linking computational and neural explanations of cooperative behavior (Krakauer et al., 2017; Lockwood et al., 2020).

Individuals are more sensitive to potential losses than to equivalent gains, a phenomenon known as loss aversion (Kahneman & Tversky, 1979), which explains why investors typically avoid placing all their eggs in one basket. However, risk aversion alone may not fully account for willingness to take risks when the uncertainty arises from another person's actions (i.e., social risk) rather than nature (i.e., financial risk). Numerous studies demonstrate that betrayal imposes an additional emotional cost beyond monetary loss (Aimone & Houser, 2012). In these studies, participants faced two distinct scenarios: one in which they could keep their money or triple it by giving it to someone else, who might either split it fairly or return only a small portion (less than the original amount), and another involving a lottery-like scenario with identical payoffs, where the return outcome was determined purely by chance. Analyses compared the minimum probability at which people accepted risks between social and non-social domains. A consistent finding is that participants are less willing to take risks in the trust game than in the lottery game (Bohnet et al., 2008; Bohnet & Zeckhauser, 2004), an effect termed betrayal aversion (Bohnet & Zeckhauser, 2004). This aversion is modulated by a variety of factors, including individual characteristics such as gender (Croson & Gneezy, 2009; Finkel et al., 2002) and broad personality traits (Komiya & Mifune, 2015; Thielmann & Hilbig, 2015; Yamagishi, 2011), as well as socio-environmental factors like social status (Hong & Bohnet, 2007), and geographic region (Bohnet et al., 2010). Additionally, psychophysiological factors, such as emotional states (Kugler et al., 2010) and oxytocin (Kosfeld et al., 2005), also shape betrayal aversion. Neuroimaging studies reveal that the differences between betrayal costs and loss costs are tracked by neural activity in the amygdala (van Honk et al., 2013) and insula (Aimone et al., 2014). These findings strongly suggest that social risk and financial risk have some fundamental distinction (Fehr, 2009a; Lauharatanahirun et al., 2012).

While existing studies have focused on how people choose to cooperate when faced with social and financial risks, they have largely overlooked an essential aspect of betrayal cost: the financial loss inherent in betrayal. Specifically, when investors decide to trust trustees, who may either reciprocate or betray, betrayal simultaneously inflicts both emotional harm and financial loss on the trustor (Aimone & Houser, 2011). However, prior studies have treated betrayal aversion and loss aversion as separate phenomena by examining them under distinct uncertain conditions. This methodological limitation leaves unresolved whether the observed reluctance to cooperate stems from the psychological impact of betrayal itself or the direct financial loss it causes, even within the same cooperative context. To address this, we constructed a value-based computational model defining the subjective value of cooperative versus non-cooperative choice, allowing us to dissociate betrayal from loss. We assumed that the best predictive logistic model would incorporate both betrayal- and loss-related components. Moreover, due to the poor temporal resolution of fMRI, previous studies may have conflated neural activities associated with cognitive processes that occur closely in time but are psychologically distinguishable. Indeed, social decision-making can be decomposed into different stages, including early automatic information processing and late-stage elaborative cognitive evaluation, each with distinct ERP components (Fan & Han, 2008; Wang et al., 2023; Yu et al., 2022).

To address these challenges, we employed the event-related potential (ERPs), which provide excellent temporal resolution to characterize the neural dynamics underlying distinct cognitive processes engaged in social decision-making (Luck, 2014). Cooperative choices are mainly associated with two ERP components: the P3 and the late positive potential (LPP). The P3 is a parietally distributed positive deflection that peaks around 350–500 ms after stimulus onset, which reflects a rapid encoding of motivational significance in prosocial actions (Li et al., 2020; Wang et al., 2017), with larger P3 amplitudes signifying greater prosocial motivation (Chiu Loke et al., 2011; Li et al., 2023; Ma et al., 2011). Using a donation task, Carlson et al. (2016) found that that high- versus low-empathy targets increased P3 under intuitive but not reflective decision-

making. Directly following the P3, the LPP is a sustained positive deflection over centroparietal areas following the stimulus onset (Hajcak et al., 2009). Unlike the P3, the LPP is thought to reflect the extended, elaborative emotion and motivation processes (Glazer et al., 2018) and might be sensitive to social information evaluation. Crucially, mounting evidence suggests that prosocial actions may stem from fast, automatic, and intuitive processes, while reflective control reflective control can reduce them (Rand et al., 2012; Righetti et al.; Zaki & Mitchell, 2013). Together, these findings illustrate the possibility that the P300 may provide an early motivational signal that fosters intuitive prosocial behavior, while LPP may represent a later, more elaborate evaluation that promotes reflexive prosocial actions. However, previous ERP research has never considered the joint impact of betrayal and loss aversion on cooperation.

To assess the motivations for cooperation and provide computational and neural accounts of betrayal and loss, we combined computational modelling and ERP with two decision-making tasks: a risk task and a cooperation task. They shared identical probabilities (10%-90%) and principal amounts (¥0.2-¥1.0). Participants could opt for the risky choice, which offered the possibility of doubling the principal but also carried the risk of losing half of the principal, or choose the safe option to retain their initial stake. The primary distinction between the tasks lies in the social context: in the cooperation task, a successful outcome resulted in mutual gain with a partner, while failure meant the partner took half of the participant's principal. In the risk task, losses were solely impersonal. Based on previous studies, we hypothesized that the early P3 reflects intuitive prosocial responses, encoding betrayal aversion, with larger P3 linked to lower betrayal aversion. In contrast, late LPP integrates betrayal and loss aversion after more deliberate evaluation, shown by a more positive LPP as betrayal aversion decreased and loss aversion increased. To provide a comprehensive understanding of cooperation, we also examined participants' decision-making tendencies across the tasks. We hypothesized that participants would be less willing to take risks facing social risk than financial risk.

2 Results

3.1 Individuals are also less willing to take risks in the cooperation task than in the risk task

As a manipulation check, we examined participants' rating data as a function of task type and outcome feedback with a series of repeated-measures analyses of variance (ANOVAs). 42 participants were retained for the final analysis due to data recording issues. As shown in Figure 1E, the cooperation task was perceived as less pleasant than the risk task, reflected by a significant main effect of task type, $F(1, 41) = 297.01$, $p < 0.001$, $\eta p^2 = 0.88$. As expected, we also found a significant interaction between task type and outcome feedback, $F(1, 41) = 16.75$, $p < 0.001$, $\eta p^2 = 0.29$. Post hoc comparisons revealed that the cooperation task elicited greater happiness than the risk task in response to positive feedback (8.29 ± 0.17 vs. 7.76 ± 0.19, $p = 0.005$, Cohen's d = 0.30). Similarly, the cooperation task elicited greater unhappiness than the risk task with negative feedback (3.12 ± 0.22 vs. 3.83 ± 0.20, $p = 0.002$, Cohen's d = 0.41). Together, our rating data suggest that the manipulation of the cooperation task was effective.

We fitted decision-time data using a linear mixed-effects regression with probability, principal, task type, and their interactions as predictors. As illustrated in Figure 1A–C, decision time decreased as probability ($b = -21.51$, $p < 0.001$) and principal magnitude ($b = -26.02$, $p < 0.001$) increased. Additionally, a significant interaction between probability and principal magnitude was observed ($b = 12.97$, $p < 0.001$). Simple slopes analyses revealed that decision time became shorter as probability level increased at low principal magnitude ($M - 1SD$: $b = -69.00$, 95% CI = [-87.50, -50.45], $p < 0.001$), but not at high principal magnitude ($M + 1SD$: $b = -17.10$, 95% CI = [-35.60, 1.43], $p = 0.071$). Decisions were also faster in the cooperation task than the risk task ($b = -73.89$, $p < 0.001$). We also observed significant interactions between probability and task type ($b = 26.55$, $p < 0.001$), as well as principal and task type ($b = -13.33$, $p = 0.046$). Simple slopes analyses revealed that increased probability level decreased the decision time in the risk task ($b = -34.78$, 95% CI = [-44.00, -25.53], $p < 0.001$), but not in the cooperation task ($b = -8.24$, 95% CI = [-17.5, 1.02], $p = 0.081$).

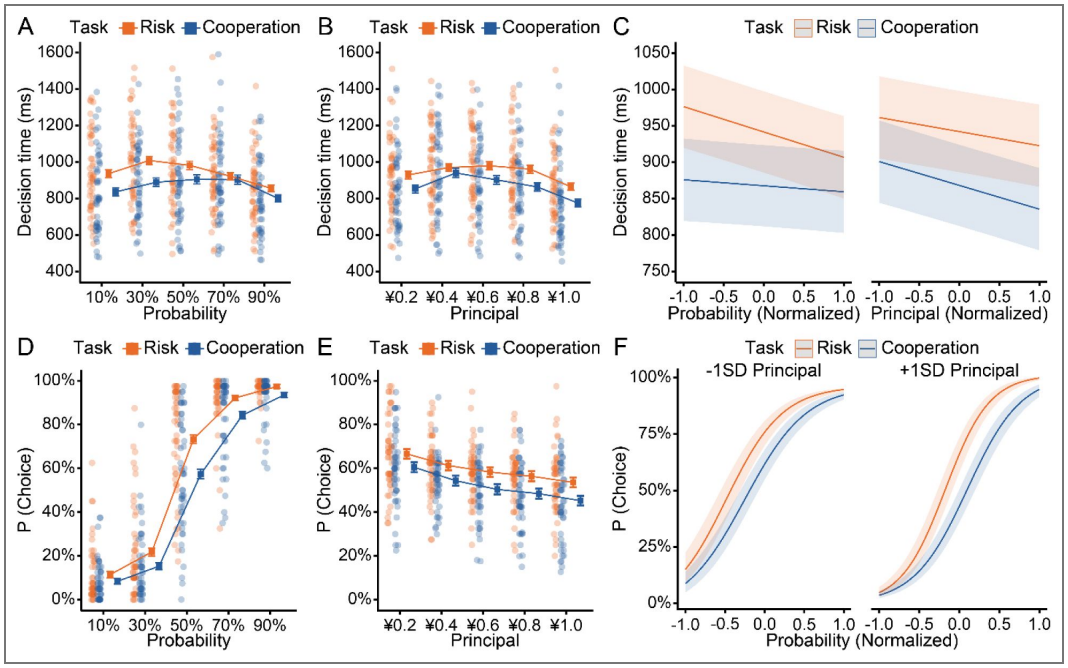

Figure 1. Behavioral results.

(A) Participants took shorter to make decisions as the probability level increased in risk task but not in cooperation task. (B) Increased principal magnitude decreased the decision time more pronouncedly in cooperation task than in risk task. (C) Fixed effects of probability and principal on the decision time as a function of task type. (D–E) Participants were less willing to take risks in the cooperation task than in the risk task, with the effect being more pronounced at high principal magnitude. (F) Fixed effects of probability on the acceptance rate as a function of task type and principal level (low vs. high). Shaded areas depict the 95% confidence intervals.

However, increased principal magnitude decreased the decision time more pronouncedly in the cooperation task (b = -32.70, 95% CI = [-41.90, -23.40], p < 0.001) than in the risk task (b = -19.40, 95% CI = [-28.60, - 10.10], p < 0.001). Full regression estimates are shown in Supplementary Table S1.

Participants' decision preference was fitted by a mixed-effects logistic regression with probability, principal, task type, and their interactions as predictors. As shown in Figure 2D–E, risk-taking increased as the probability level increased (b = 20.18, p < 0.001) and principal magnitude decreased (b = 0.66, p < 0.001). These effects were further qualified by a significant interaction between probability and principal (b = 1.36, p < 0.001). Simple slopes analyses revealed that participants were more willing to take risks as probability level increased at high principal magnitude (M + 1SD: b = 6.63, 95% CI = [6.00, 7.26], p < 0.001) than at low reward magnitude (M – 1SD: b = 5.39, 95% CI = [4.77, 6.01], p < 0.001). Moreover, they exhibited lower risk-taking in the cooperation task than the risk task (b = 0.44, p < 0.001), further modified by task type × probability (b = 0.64, p < 0.001) and task type × principal interaction (b = 0.90, p < 0.001). Simple slopes analyses revealed that the inhibition effect of the principal on willingness to take risks was more pronounced during the cooperation task (b = -0.47, 95% CI = [-0.53, -0.41], p < 0.001) compared to the risk task (b = -0.37, 95% CI = [-0.44, -0.30], p < 0.001). Similarly, the facilitation effect of probability on willingness to take risks was less pronounced during the cooperation task (b = 2.78, 95% CI = [2.47, 3.09], p < 0.001) compared to the risk task (b = 3.23, 95% CI = [2.91, 3.54], p < 0.001).

The probability × task type interaction was further qualified by a significant three-way interaction among probability, principal, and task type (b = 0.82, p = 0.002). Post hoc simple slope analyses (Figure 1F) showed that the facilitative effect of probability on risk-taking was attenuated in the cooperation task than the risk task, particularly at high principal magnitude (risk: b = 3.63, 95% CI = [3.29, 3.97], p < 0.001; cooperation: b = 2.99, 95% CI = [2.67, 3.32], p < 0.001) and to a lesser extent at low principal magnitude (risk: b = 2.82, 95% CI = [2.50, 3.14], p < 0.001; cooperation: b = 2.57, 95% CI = [2.26, 2.89], p < 0.001). Full regression estimates are shown in Supplementary Table S1. Overall, participants were less motivated to take risks in the cooperation task than the risk task, as evidenced by lower accepted probabilities and longer decision times in the cooperation task.

3.2 Betrayal aversion suppresses cooperation willingness to a higher extent than loss aversion

We constructed a series of model to test effects of loss aversion and betrayal aversion on cooperation, including (a) a model with linear formulations of utilities for gain and loss values (Model 1); (b) a model that additionally incorporate betrayal aversion into the loss utility (Model 2); (c) a model that assumes participants consider other's value in addition to gain and loss values (Model 3); and (d) models with task-specific parameters in the above three models (Models 4, 5, and 6) (see Method for more details). Model comparison decisively favored Model 3 (AIC = 238.8), which incorporated CPT, betrayal aversion, and prosocial utility without task-specific parameters (Figure 2A). Parameter recovery analysis showed the parameters of M3 were well identifiable (α_gain = 87%, α_loss = 98%, γ = 99%, λ = 90%, β = 92%, κ = 83%, ω = 88%; Figure 2B and C; details of model comparison and parameter recovery see Supplementary Table S2 & S3).

To examine how betrayal aversion and loss aversion influence cooperation behavior, we fitted decision time using a linear mixed-effects regression and decision preference using a mixed-effects logistic regression with a binomial link. As shown in Figure 2D, both loss aversion and betrayal aversion significantly predicted cooperation, such that participants cooperated less as loss aversion (b = 2.22, p < 0.001) and betrayal aversion (b = 6.96, p < 0.001) increased. Moreover, we employed β coefficient tests to compare the predictive power of loss and betrayal aversion on cooperation willingness, revealing that betrayal aversion had a stronger impact on cooperation than loss aversion (z = -6.97, p < 0.001). Both aversions also robustly predicted decision times, with decision time becoming shorter as loss aversion increased (b = 69.97, p < 0.001) and betrayal aversion decreased (b = -94.04, p < 0.001), again with betrayal aversion showing greater predictive capacity (z = 7.31, p < 0.001). Moreover, as shown in Figure 2E, a greater preference for others' benefits was associated with a higher betrayal aversion (r = -0.50, p < 0.001), but not with loss aversion (r = 0.01, p = 0.997), indicating that individual differences influence cooperative

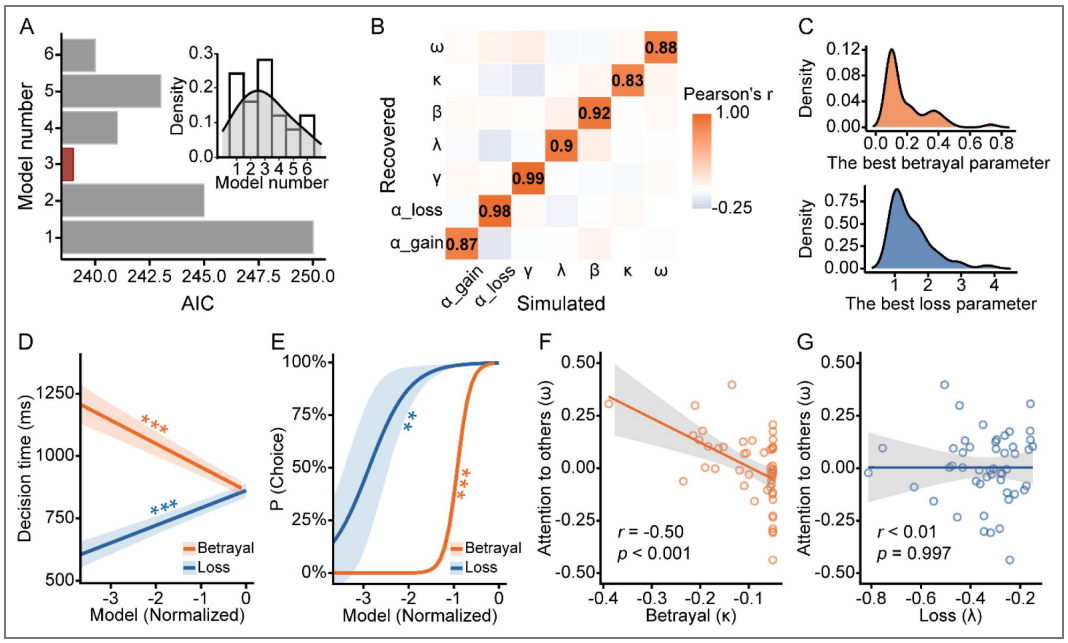

Figure 2. Computational results.

(A) AIC values (bars; lower values indicate better fit) from the comparison of the computational models. The insert shows the distribution of the model selection across participants. Model 3 (incorporating CPT, betrayal aversion, and prosocial utility) was selected as the winning model. (B) Parameter recovery of Model 3: gain sensitivity (α_gain), loss sensitivity (α_loss), probability sensitivity (γ), loss aversion (λ), noise (β), betrayal aversion (κ), and prosocial preference (ω). The correlation matrix shows the Pearson correlation between the actual recovered and simulated data from the winning model. The high correlations (all $r > 0.80$) indicate excellent parameter recovery. (C) The distribution of the best loss (top) and defect (bottom) parameters for the winning model, which maximizes log-likelihood during fitting. (D–E) Decision time was becoming shorter as loss aversion increased, and betrayal aversion decreased, and the willingness to cooperate became lower as loss aversion and betrayal aversion increased. Betrayal aversion significantly predicts both decision time (D) and acceptance rate (E). Shaded areas depict the ± 1 standard errors. (F–G) Focusing more on others' benefits was associated with increased betrayal aversion (F), but did not affect loss aversion (G). Loss represents the subject value for the best-fitting financial loss aversion ($-\lambda * (0.5 * V)^{\alpha_{loss}} * P_{Weight}$) and defect is the subject value of betrayal aversion $\kappa * P_{Weight}$). ***$p < 0.001$, **$p < 0.01$.

tendencies. Full regression estimates are provided in Supplementary Table S4. Together, our computational modeling indicates that betrayal aversion has a stronger suppressive effect on cooperative choices than loss aversion, reflected in lower cooperation willingness and longer decision times.

3.3 Betrayal information was encoded during the early P3 period, whereas loss information was further integrated during the late LPP period

The P3 and LPP components were observed as a relative positivity over centroparietal areas (Figure 3). Betrayal and loss parametric were derived from the winning model (Model 3). To investigate their effects on neural dynamics underlying cooperation, we fitted the amplitude of the P3 and LPP using a linear mixed-effect regression with z-scored betrayal and loss parametric as predictors. As shown in Figure 3, the P3 became less positive as betrayal aversion increased (b = 0.44, p = 0.017), but was unaffected by loss aversion (b = -0.09, p = 0.642), indicating a betrayal inhibition effect. For the LPP, both betrayal inhibition and loss inhibition effects were observed, with reduced positivity as betrayal aversion (b = 0.71, p < 0.001) and loss aversion (b = -0.48, p = 0.009) increased. Further β coefficient tests showed comparable predictive power of betrayal and loss on LPP amplitude (z = -0.85, p = 0.198). Full regression estimates are shown in Supplementary Table S5. Together, our ERP data suggest that betrayal and loss information were processed serially, such that betrayal information was encoded during the P3 periods, whereas loss information was not tracked until the time window of the late positive potential.

3 Discussion

Betrayal suppresses cooperation and trust, yet the associated financial losses it incurs are often overlooked, complicating the reasons behind cooperative behavior. Here, we explore how betrayal and its financial consequences influence cooperation alongside their underlying neural mechanisms, addressing a gap in previous research. We found that betrayal exerts a significantly stronger suppression on cooperation than loss. Moreover, neural processing unfolded sequentially: betrayal modulated early-stage automatic processing indexed by the P3, while loss appeared in later evaluative stages captured by the LPP.

In our study, participants reported a stronger preference for positive feedback and greater aversion to negative feedback during the cooperation task, suggesting that betrayal imposes an additional emotional cost beyond monetary loss. They were also less willing to take risks as the cooperation risk increased, especially when the principal magnitude was high, compared to the financial risk. Moreover, in the cooperation task, decision times were shorter and driven more by principal amounts rather than probability levels, indicating that participants prioritized potential gains and losses and adopted a generally cautious approach, regardless of the cooperative probability. These findings are largely consistent with previous studies, which compared individuals' willingness to take risks when the outcome depends on another player's trustworthiness, versus situations where the outcome is determined by a chance with identical odds and payoffs, demonstrating a stronger avoidance motivation in social contexts (Bohnet et al., 2008; Bohnet et al., 2010; Bohnet & Zeckhauser, 2004). However, these studies overlook that betrayal simultaneously incurs emotional and financial losses.

By employing computational modeling to separately examine how betrayal and its associated loss parameters influence cooperation, our most important finding is a greater substantial inhibitory effect of betrayal on cooperation than loss. Participants exhibited lower cooperation rates and longer decision times as both loss aversion and betrayal aversion increased. Notably, betrayal aversion had a stronger impact on choice or decision time. Several theoretical frameworks can help interpret this effect. First, individuals care about others' payoffs in cooperative situations (Fehr, 2009b). In our study, when partners betrayed by taking half of the investor's money, participants are less likely to accept the social risk compared to the financial risks. This possibility was further supported by our modeling results regarding the correlation between attention to

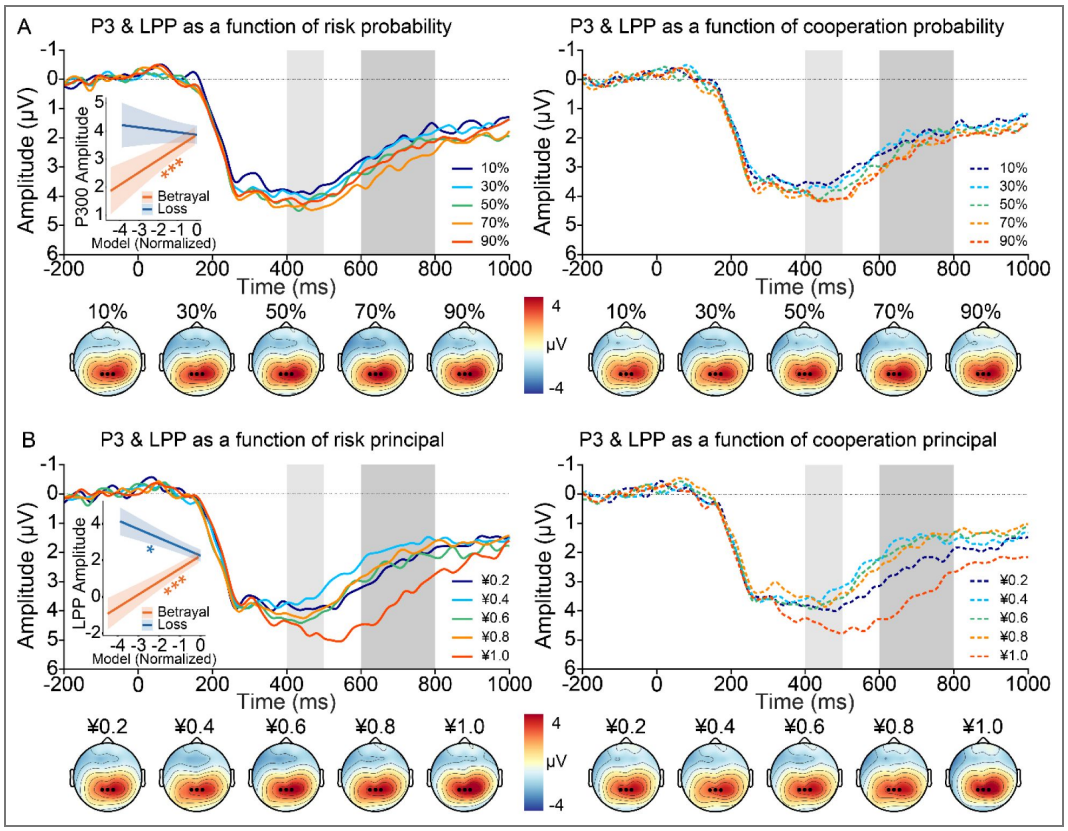

Figure 3. Grand-average ERP waveforms and topographic maps of the P3 and LPP as a function of task type (risk task vs. cooperation task) separately for probability (A) and principal (B) trials.

Gray shaded bars represent time windows used for quantification. The insets show the fixed effects of betrayal and loss aversion on the P3 (A) and LPP (B). A betrayal-inhibition effect emerged during the early P3 phase and persisted through the late LPP phase, while the loss-inhibition effect appeared solely during the late LPP phase. Shaded areas depict the 95% confidence intervals.

others, betrayal, and loss aversion. Specifically, those who focused more on others' benefits show higher betrayal aversion, but not loss aversion. Second, people care not only about outcomes but also the intentions behind them (Charness & Rabin, 2002; Rabin, 1993). Trust builds honor and integrity, promoting cooperation, while betrayal erodes these qualities, discouraging future cooperation. Third, betrayal often inflicts emotional harm, such as damaged trust, which can outweigh material loss. Aimone et al. (2014) found stronger activation in the anterior insular cortex when playing with a human versus a computer, indicating heightened negative emotions (Kuhnen & Knutson, 2005; Wu et al., 2012), supporting betrayal aversion as a desire to avoid negative emotions. Finally, people perceive risks from others' actions as less controllable than natural risks (Slovic, 1987). For example, shareholders would prefer a 1% chance of losing half their value due to a natural disaster than a slightly smaller chance from betrayal by corporate executives.

Interestingly, we observed distinct neural mechanisms for betrayal and loss. Specifically, the early P3 amplitude decreased with increasing betrayal aversion, while no significant change was observed in response to loss aversion. In contrast, the LPP amplitude decreased as both betrayal and loss aversion increased. These findings suggest that betrayal and loss information are processed sequentially, with betrayal being encoded during the P3 period, followed by further integration of loss information during the LPP period. Cooperation is often considered an intuitive behavior (Gao et al., 2020; Koch et al., 2020; Levine et al., 2018; Rand et al., 2012; Shank et al., 2019), which is indexed by P3, a neural signal that rapidly encodes prosocial motivation (Li et al., 2020; Wang et al., 2017). Betrayal, a key factor inhibiting prosocial behavior (Fehr & Gächter, 2000), is therefore prioritized for encoding at the early stages of cooperation decision-making. According to a dual-process framework, intuition is fast, automatic, and effortless, while reflection is often slow, deliberate, and characterized by the rejection of emotional influence. (Chaiken & Trope, 1999; Frederick, 2005; Kahneman, 2011; Plessner et al., 2008; Sloman, 1996; Stanovich & West, 1998). Hence, during the later phase of decision-making, participants further engage in a trade-off between costs and benefits associated with specific actions. This process is indexed by the LPP, a neural marker reflecting sustained and elaborative information processing (Glazer et al., 2018). In this stage, loss-related information is integrated into the decision-making process.

P300 amplitude may serve as a neural marker of aversion to defection in cooperative interactions. Specifically, the P300 amplitude increases when individuals are faced with decisions that challenge cooperation, particularly when they are prompted by intuitive, emotionally-driven responses. This suggests that P300 could signal the underlying motivational processes that discourage defection and promote prosocial behavior, especially when individuals feel a social or emotional connection to others involved in the interaction. In this context, the P300 response not only provides insight into how cooperative behaviors are initiated but also offers a predictive measure of one's commitment to maintaining cooperation under varying conditions of social engagement.

A limitation of this study is the ecological validity of the cooperative scenario. Participants were informed that they would cooperate with others, but some (N=6) expressed doubts about the authenticity of this interaction in a post-experiment check. Although this did not affect the results, future studies could investigate neural synchronization during multi-player cooperation to better understand how co-players synchronize their neural activity in such contexts (Jiang et al.; Yang et al., 2020; Zhang et al., 2023). Additionally, future studies could provide evidence to distinguish the aversion to being personally betrayed and witnessing another's betrayal. This approach will help determine whether a betrayal has a stronger inhibitory effect and is processed with priority from a third-party perspective, where the outcome is unrelated to oneself.

In conclusion, this study demonstrates that betrayal has a stronger suppressive effect on cooperation compared to the associated cost of losses. Additionally, this study provides preliminary evidence that the neural dynamics involved in these two forms of aversion occur sequentially, with betrayal aversion influencing early-stage decision-making and loss aversion

affecting later-stage evaluations. These findings enhance our understanding of how emotional and financial considerations influence cooperative behavior and provide important insights into the neural mechanisms that underlie cooperative decision-making.

4 Materials and methods

All data and code used for this study are available on OSF at *https://osf.io/zw2ra/*. This study was not preregistered.

4.1 Participants

Fifty right-handed university students were recruited via local advertisements for this study. One participant was excluded from data analysis due to a misunderstanding of instructions. The final sample thus consisted of 49 participants (24 females; *M* = 22.00 years, standard deviation [*SD*] = 2.03). We performed a sensitivity analysis using the *simr* v1.0.6 package(Green & MacLeod, 2016) (Green & MacLeod, 2016) to compare the regression weight for each effect of interest with the smallest detectable effect size at a power of 80% based on the current sample. The results showed that most of the significant effects observed were larger than the smallest detectable effect, suggesting that our sample size provided adequate statistical power. All participants had normal or corrected-to-normal vision as determined by self-report and no psychiatric or neurological disorders. Each received a bonus of ¥70–¥80 based on their task performance. This study was approved by the Institutional Review Board of Sun Yat-sen University.

4.2 Procedure

Upon arrival at the lab, participants completed a probabilistic risk task and a single-shot cooperation task while their EEG was recorded. After the EEG tasks, participants rated their perceived happiness on a 9-point Likert scale (ranging from 1 = not at all to 9 = very much), based on the outcomes (gain or loss) associated with each decision.

The risk task

This task was designed to assess neural correlates of loss aversion. On each trial (Figure 4A), participants first viewed a probability pie chart for 1500 ms, followed by a jittered interval (1200–1500 ms). The pie chart illustrated probability levels (10%, 30%, 50%, 70%, or 90%, corresponding to 1, 3, 5, 7, or 9 segments of the chart), indicating the likelihood of doubling the principal versus losing half of it. Afterward, a number appeared for 1500 ms, showing the principal amount (¥0.2, ¥0.4, ¥0.6, ¥0.8, or ¥1.0, corresponding to 1–5 principal levels). The five probability levels were fully crossed with the five principal levels, resulting in 25 unique combinations. Following another jittered interval (900–1100 ms), participants entered the choice phase, choosing between the "reject" option (R), which allowed them to keep the current principal and proceed to the next trial, and the "accept" option (A), which offered a chance to double the principal with the given probability or lose half in case of failure.

Participants had up to 4000 ms to decide using their left or right index finger. The chosen option was highlighted with a green border for 300 ms. Failure to respond within the time limit resulted in ¥0 and a 1000-ms warning. After another jittered interval (1200–1500 ms), an outcome feedback stimulus was presented for 1000 ms: a tick denoted gain, a cross denoted loss, followed by a jittered interval of 900–1100 ms. Subsequently, a reward feedback stimulus was shown for 1000 ms, indicating the final payoff. Each trial ended with an interval varying between 900 and 1100 ms.

The cooperation task

This task was designed to measure participants' neural responses to both loss aversion and defect aversion, which employed a similar structure but incorporated social decision-making elements. One key distinction was that each trial involved an anonymous co-player, who was distinct from the ones in any other trials and would not reappear in any subsequent trial. Another key difference was the presence of a probability cue reflecting the likelihood that the co-player would

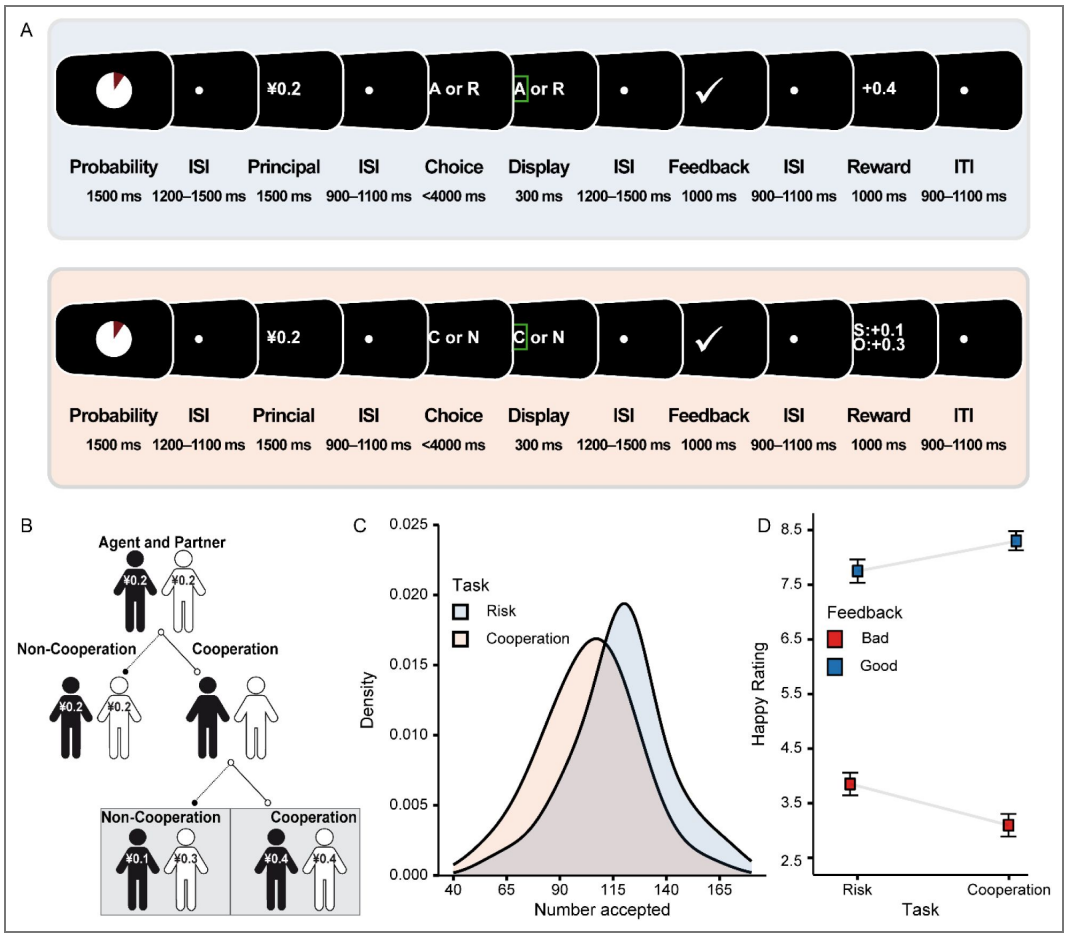

Figure 4. Experimental tasks and rating results.

(A) The risk task (top). Participants chose a reject option, keeping their current principal (¥0.2, ¥0.4, ¥0.6, ¥0.8, or ¥1.0), and an "accept" option, where they gambled to double their principal, with a varying probability (10%, 30%, 50%, 70%, 90%) of success, or lose half the principal if they fail. The cooperation task (bottom). Participants as agents chose whether to cooperate with their partner, with the same probabilities and payoffs as the risk task. If successful, both agent and partner doubled their principal; if it failed, the partner took half of the agent's principal. ISI = interstimulus interval; ITI = intertrial interval. (B) The decision tree depicts an example of a decision to cooperate, and the partner could choose cooperation or betrayal. (C) The distribution of the accepted numbers during the risk task and the cooperation task. (D) Rating data. Participants felt more liking for positive outcomes and more disliking of negative outcomes during the cooperation task than the risk task. Error bars represent the within-subject standard error of the mean.

cooperate. Participants were informed that the probabilities as derived from prior behavioral data, but in reality, they were randomly generated. Critically, unlike the risk task, where outcomes depended solely on chance, the cooperation task introduced interpersonal contingencies. During the choice phase, participants decided whether to cooperate (C) or not (N). Feedback then revealed the co-player's decision (a tick for cooperation, a cross for defection), followed by a reward screen showing the monetary outcome for both players. Mutual cooperation resulted in the principal being doubled and split equally. If the participant cooperated but the co-player defected, the participant incurred a trust cost by forfeiting half of their principal, which was transferred to the co-player in addition to the co-player's own principal. If the participant chose not to cooperate, the principal remained unchanged regardless of the co-player's choice. The risk task and cooperation decision-making task each consisted of 200 trials, divided into eight blocks of 25 trials, with a self-determined break between blocks. The task order was counterbalanced across participants. To maintain attention, each task included 10 catch trials in which participants were required to confirm the probability and principal levels for the current trial. Before the experiment, participants completed 7 practice trials for each task for familiarization. After the task, participants were asked to report whether they believed they were playing with the third-party players. Three subjects had an accuracy rate of less than 60% on catch trials, and six subjects expressed doubts about the cooperation manipulation. However, since this did not affect the results, they were included in the analysis.

4.3 EEG recording and processing

EEG data were recorded using 64 Ag/AgCl channels placed on an elastic cap based on the international 10–20 system. Two additional channels were positioned on the left and right mastoids. The reference electrode was located at Cz. Horizontal and vertical electrooculograms were recorded from two pairs of channels over the external canthi of each eye and the left suborbital and supraorbital ridges, respectively. EEG signals were amplified using a Neuroscan SynAmps^2 amplifier with a low-pass filter of 100 Hz in DC acquisition mode and digitized at a rate of 1000 samples per second. Channel impedances were maintained below 10 KΩ.

The EEG data were analyzed using EEGLAB v2021.0 (Delorme & Makeig, 2004) and ERPLAB v8.10 (Lopez-Calderon & Luck, 2014) toolboxes in MATLAB 2020b (MathWorks, US). The signals were rereferenced to the average of the left and right mastoids and filtered with a bandpass of 0.1–35 Hz using a zero phase-shift Butterworth filter (12 dB/octave roll-off). Channels with poor quality or excessive noise were interpolated using the spherical interpolation algorithm, and portions of EEG containing extreme voltage offsets or break periods were removed. Ocular artifacts were removed using an infomax independent component analysis on continuous EEG with the help of the ICLabel algorithm (Pion-Tonachini et al., 2019). Epochs were then extracted from -200 to 1000 ms relative to principal stimulus onset, with the prestimulus average activity as the baseline. An automatic artifact detection algorithm was applied to remove epochs with a voltage difference exceeding 50 μV between sample points or 200 μV within a trial, a maximum voltage difference less than 0.5 μV within 100-ms intervals, or a slow voltage drift with a slope greater than ± 100 μV. On average, 96.30% of trials were retained for statistical analysis. Measurement parameters were determined by the grand-averaged ERP waveforms and topographic maps collapsed across all conditions (Luck & Gaspelin, 2017). Specifically, the single-trial P300 amplitude and LPP amplitude were measured as the mean activity from 400 to 500 ms and 600 to 800ms, respectively, after the onset of the principal cues over centroparietal areas (P1, Pz, P2). These single-trial data were exported into R v4.2.2 for statistical analyses.

4.4 Data analysis

Statistical analysis

Our key statistical analyses were based on mixed-effects regression models with random intercepts and slopes (unstructured covariance matrix), implemented in the *lme4* package v1.1.31 (Bates, Maechler, et al., 2015). For decision time and choice data, we separately used a linear mixed-effects regression model and a mixed-effects logistic regression model with a binomial link

function. Both models included probability, principal, task type, and their interactions as predictors. For all models, categorical regressors (task type: -0.5 for risk and +0.5 for cooperation) were contrast-coded, and continuous regressors (probability and principal) were z-scored before inclusion in the regression models. We determined random effects for each model using singular value decomposition to report the maximal possible random effects structure (Barr, 2013; Bates, Kliegl, et al., 2015). Follow-up pairwise comparisons of significant interactions were tested on estimated marginal means using the *emmeans* package v1.8.3 (Lenth, 2022). To examine the relationship between behavioral indices of loss aversion (λ) and defect aversion (κ) and neural responses to principal cues, both λ and κ were entered as fixed effects in separate P300 and LPP models. All model predictors were scaled (e.g., λ/sqrt(sum(λ^2)/length(λ)-1)) without mean-centering to ensure comparability of β coefficients. We analyze post-experimental rating data using a repeated-measures analysis of variance (ANOVA), with feedback (good, bad) and task type (risk, cooperation) as within-subjects factors. We excluded trials with no responses (0.15%) in the risk and cooperation task from statistical analyses.

Computational modeling

To quantitatively dissociate the contributions of loss aversion and betrayal aversion to cooperative decision making, we employed a computational framework grounded in Cumulative Prospect Theory (CPT; Tversky & Kahneman, 1992). This approach models risky decisions through two core transformations: a value function that asymmetrically weights losses versus gains, and a probability weighting function that nonlinearly distorts objective probability. A family of seven models (M0-M6) was systematically constructed to incorporate socially specific factors and test hypotheses about how financial losses and betrayal aversion independently affect cooperation choices. The simplest model (M0) served as a random choice baseline. Model 1 was adapted from the cumulative prospect theory, and defined subjective value (SV) as the net expected utility of choosing investment or cooperation. Specifically, for a given option with principal amount X and success probability p, the SV was computed as

$$SV = w(p) * V\left(\Delta X_{\text{success}}\right) + w(1-p) * V\left(\Delta X_{\text{failure}}\right)$$

Where the net utility of success $\Delta X_{success} = X$ and net utility of *failure* $\Delta X_{failure} = -0.5 * X$ as the setting of the experiment. The subjective utility function $V(x)$ was defined as

$$V(x) = \begin{cases} x^{\alpha\_\text{gain}}, \text{ if } x \geq 0(\text{ gain }) \\ -\lambda * |x|^{\alpha\_\text{loss}}, \text{ if } x < 0(\text{ loss }) \end{cases}$$

Here, the parameter α_{gain} and α_{loss} ($0 < \alpha < 1$) represents the diminishing sensitivity to increasing magnitudes of gains and losses, respectively; λ quantifies financial loss aversion, $\lambda \geq 1$ indicates that the reduced utility of losses is greater than the utility of gains at the same magnitude.

Besides, the subjective probability weighting function $w(p)$ was defined as:

$$w(p) = \frac{p^{\gamma}}{\left(p^{\gamma}+(1-p)^{\gamma}\right)^{1/\gamma}}$$

Parameter γ ($\gamma < 1$) governing the curvature and sensitivity of probability for the outcomes. For the simplicity of the model, we estimated a single γ parameter across gains and losses. Lower values on γ indicate greater curvature and lower sensitivity to probabilities.

Building on M1, Model 2 introduced betrayal aversion as an additive emotional cost term independent of loss aversion. The SV of M2 was computed as

$$SV = w(p) * V\left(\Delta X_{\text{success}}\right) + w(1-p) * \left[-\kappa + V\left(\Delta X_{\text{failure}}\right)\right]$$

Here, k ($k > 0$) captures betrayal aversion independent of the monetary loss magnitude, reflecting the hypothesis that social betrayal carries intrinsic disutility beyond monetary consequences.

Model 3 further incorporated explicit valuation of the co-player rewards to capture prosocial preferences independent of self-interest.

$$V_{\text{other}} = \omega * [\text{p} * 2 * \text{X} + (1-\text{p}) * 1.5 * \text{X})]$$

Here, ω is a scaling factor, and the term inside the parentheses reflects the expected reward for others set by experiment. Then, the SV of M3 was computed as

$$SV = w(p) * V\left(\Delta X_{\text{success}}\right) + w(1-p) * \left[-\kappa + V\left(\Delta X_{\text{failure}}\right)\right] + V_{\text{other}}$$

The derived SV of the option was transformed into choice probability through a softmax function incorporating a stochasticity parameter β:

$$P_{\text{risk or coop}} = \frac{1}{1+e^{-\beta * SV}}$$

Where β ($\beta \geq 0$) represents the slope of the logistic function and reflects the sensitivity to SV differences between options. The lower β is, the greater the stochasticity of the choices; with $\beta = 0$, choices are random (Collins & Shenhav, 2022).

To account for potential task-dependent variations, we tested three additional variants (M4-M6), allowing the sensitivity of gain α_{gain}, the sensitivity of probability γ, and the stochasticity parameter β in M1-M3 vary across tasks.

We estimated the parameters of each model for each participant, using maximum likelihood estimation implemented by fmincon function in MATLAB, with multiple initialization points to avoid local minima. Model comparison employed the Akaike Information Criterion (AIC) to balance model fitting and complexity. A lower AIC value indicates a better-fitting and concise model. Moreover, the parameters' identifiability was validated through recovery analyses. For each model, 500 plausible parameter sets were randomly sampled from uniform distributions bounded by empirically informed ranges. Using these ground-truth parameters, we simulated trial-wise choice data (500 trials/set) under experimental conditions matching the original task design. The model was subsequently refitted to the simulated datasets using 20 randomized initialization points per recovery to mitigate local minimum convergence. Parameter recovery similarity was quantified through Pearson correlations between true and estimated parameter values across all simulations.

Supplementary Materials

Predictors	Decision times			Choices		
	Estimates	95% CI	*p*	*Log odds*	95% CI	*p*
Intercept	904.64	853.45 to 955.84	**<0.001**	1.79	1.35 to 2.37	**<0.001**
Probability	-21.51	-28.05 to -14.96	**<0.001**	20.18	14.87 to 27.41	**<0.001**
Principal	-26.02	-32.57 to -19.48	**<0.001**	0.66	0.63 to -0.69	**<0.001**
Task type	-73.89	-118.16 to -29.63	**0.001**	0.44	0.34 to 0.56	**<0.001**
Probability:Principal	12.97	6.42 to 19.51	**<0.001**	1.36	1.28 to 1.45	**<0.001**
Probability:Task type	26.55	13.45 to 39.64	**<0.001**	0.64	0.56 to -0.74	**<0.001**
Principal:Task type	-13.33	-26.42 to -0.24	**0.046**	0.90	0.82 to 0.99	**0.029**
Probability: Principal: Task type	5.16	-7.93 to 18.25	0.440	0.82	0.72 to 0.93	**0.002**
Observations	19571			19571		

Notes: The final model for decision times as: Decision times ~ Probability * Principal * Task type+ (Task type | Participant), and for the proportion accepted as: Proportion accepted ~ Probability * Principal * Task type+ (Probability + Task type | Participant). Probability and principal level were standardized before being entered into the model. Statistically significant *p* values (<0.05, two-sided) are shown in bold. CI = confidence interval.

Table S1. Results of linear mixed-effects models predicting decision times (left) and choices (right) as a function of probability, principal, and task type.

Table S2. **Model comparison**

Model	Model description		Number of parameters	Log likelihood	Average AIC	ΔAIC
Model 0	Random model		1	-13300	534	295.2
Model 1	Cumulative Prospective Theory (CPT)	Shared parameters	5	-5995	250	11.0
		Independent parameters	10	-5515	240.6	1.8
Model 2	CPT + defect aversion	Shared parameters	6	-5830	245.2	6.4
		Independent parameters	9	-5610	242.4	3.6
Model 3	CPT + defect aversion + opponent's value	Shared parameters	7	-5621	238.8	0
		Independent parameters	10	-5478	239.1	0.3

Note: Average AIC = (2 * num_subjects * num_params - 2 * Loglikelihood) / num_subjects; $\Delta AIC = AIC - \min(AIC)$

Table S3. **Parameter recovery analysis**

Model	Model description		*Mean r*	*p*
Model 0	Random model		1	<.001
Model 1	Cumulative Prospective Theory (CPT)	1a. Shared parameters	.922	<.001
		1b. Independent parameters	.916	<.001
Model 2	CPT + defect aversion	2a. Shared parameters	.921	<.001
		2b. Independent parameters	.912	<.001
Model 3	CPT + defect aversion + opponent's value	3a. Shared parameters	.909	<.001
		3b. Independent parameters	.893	<.001

Note: Pearson correlation analysis, two-tailed.

Table S4. **Results of linear mixed-effects models predicting decision times (left) and choices (right) as a function of loss aversion and betrayal aversion.**

	Decision times			Choices		
Predictors	*Estimates*	95% CI	*p*	*Log odds*	95% CI	*p*
Intercept	861.87	806.45 to 917.29	**<0.001**	6.39	5.61 to 7.17	**<0.001**
Loss aversion	69.97	45.23 to 94.70	**<0.001**	2.22	1.36 to 3.09	**<0.001**
Betrayal aversion	-94.04	-131.44 to -56.64	**<0.001**	6.96	5.95 to - 7.98	**<0.001**
Observations	9787			9787		

Notes: The final model for Decision times as: Decision times ~ Loss aversion + Betrayal aversion + (Loss aversion | Participant), and for the proportion accepted as: proportion accepted ~ Loss aversion + Betrayal aversion + (Loss aversion + Betrayal aversion | Participant). Loss aversion and betrayal aversion were standardized before being entered into the model. Statistically significant *p* values (<0.05, two-sided) are shown in bold. CI = confidence interval.

Table S5. Results of linear mixed-effects models predicting P300 (left) and LPP (right) as a function of loss aversion and betrayal aversion.

	P300			LPP		
Predictors	*Estimates*	95% CI	*p*	*Estimates*	95% CI	*p*
Intercept	3.87	3.23 to 4.51	**<0.001**	2.26	1.60 to 2.91	**<0.001**
Loss aversion	-0.09	-0.46 to 0.28	0.640	-0.48	-0.84 to –0.12	**0.009**
Betrayal aversion	0.44	0.08 to 0.81	**0.016**	0.71	0.33 to -1.08	**<0.001**
Observations	9425			9425		

Notes: The final model for P300/ LPP as: P300 amplitude/ LPP amplitude ~ Loss aversion + Betrayal aversion + (Loss aversion | Participant). Loss aversion and betrayal aversion were standardized before being entered into the model. Statistically significant *p* values (<0.05, two-sided) are shown in bold. CI = confidence interval.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32171073).

Additional information

Funding

Funder	Grant reference number	Author
MOST | National Natural Science Foundation of China (NSFC)		Dingguo Gao

Author ORCID iDs

Rumeng Tang: https://orcid.org/0009-0005-5243-5772
Xiaowei Ding: https://orcid.org/0000-0002-0583-291X
Dingguo Gao: https://orcid.org/0000-0003-4078-9942

References

Aimone JA, Houser D (2011) Beneficial Betrayal Aversion. *PLoS ONE* **6**:e17725 https://doi.org/10.1371/journal.pone.0017725 | PubMed

Aimone JA, Houser D (2012) What you don't know won't hurt you: a laboratory analysis of betrayal aversion. *Experimental Economics* **15**:571-588 https://doi.org/10.1007/s10683-012-9314-z

Aimone JA, Houser D, Weber B (2014) Neural signatures of betrayal aversion: an fMRI study of trust. *Proceedings of the Royal Society B: Biological Sciences* **281**:20132127 https://doi.org/10.1098/rspb.2013.2127 | PubMed

Barr DJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**:255-278 https://doi.org/10.1016/j.jml.2012.11.001

Bates D, Kliegl R, Vasishth S, Baayen H (2015) Parsimonious mixed models. *arXiv* https://doi.org/10.48550/arxiv.1506.04967

Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**:1-48 https://doi.org/10.18637/jss.v067.i01

Bohnet I, Greig F, Herrmann B, Zeckhauser R (2008) Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. . *American Economic Review* **98**:294-310 https://doi.org/10.1257/aer.98.1.294

Bohnet I, Herrmann B, Zeckhauser R (2010) Trust and the Reference points for Trustworthiness in Gulf and Western. *Quarterly Journal of Economics* **125**:811-828 https://doi.org/10.1162/qjec.2010.125.2.811

Bohnet I, Zeckhauser R (2004) Trust, risk and betrayal. *Journal of Economic Behavior & Organization* **55**:467-484 https://doi.org/10.1016/j.jebo.2003.11.004

Carlson RW, Aknin LB, Liotti M (2016) When is giving an impulse? An ERP investigation of intuitive prosocial behavior. *Social Cognitive and Affective Neuroscience* **11**:1121-1129 https://doi.org/10.1093/scan/nsv077 | PubMed

Chaiken S, Trope Y (1999) *Dual-process theories in social psychology* The Guilford Press.

Charness G, Rabin M (2002) Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics* **117**:817-869 https://doi.org/10.1162/003355302760193904

Chiu Loke I, Evans AD, Lee K (2011) The neural correlates of reasoning about prosocial–helping decisions: An event-related brain potentials study. *Brain Research* **1369**:140-148 https://doi.org/10.1016/j.brainres.2010.10.109 | PubMed

Collins AGE, Shenhav A (2022) Advances in modeling learning and decision-making in neuroscience. *Neuropsychopharmacology* **47**:104-118 https://doi.org/10.1038/s41386-021-01126-y | PubMed

Croson R, Gneezy U (2009) Gender Differences in Preferences. *Journal of Economic Literature* **47**:448-474 https://doi.org/10.1257/JEL.47.2.448

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**:9-21 https://doi.org/10.1016/j.jneumeth.2003.10.009 | PubMed

Fan Y, Han S (2008) Temporal dynamic of neural mechanisms involved in empathy for pain: An event-related brain potential study. *Neuropsychologia* **46**:160-173 https://doi.org/10.1016/j.neuropsychologia.2007.07.023 | PubMed

Fehr E (2009a) On the Economics and Biology of Trust. *Journal of the European Economic Association* **7**:235-266 https://doi.org/10.1162/JEEA.2009.7.2-3.235

Fehr E (2009b) Social Preferences and the Brain. In: Glimcher PW, Camerer CF, Fehr E, Poldrack RA (Eds). *Neuroeconomics* Academic Press. pp. 215-232 https://doi.org/10.1016/B978-0-12-374176-9.00015-4

Fehr E, Fischbacher U (2003) The nature of human altruism. *Nature* **425**:785-791 https://doi.org/10.1038/nature02043 | PubMed

Fehr E, Gächter S (2000) Cooperation and Punishment in Public Goods Experiments. *American Economic Review* **90**:980-994 https://doi.org/10.1257/aer.90.4.980

Finkel EJ, Rusbult CE, Kumashiro M, Hannon PA (2002) Dealing with betrayal in close relationships: Does commitment promote forgiveness?. *Journal of Personality and Social Psychology* **82**:956-974 https://doi.org/10.1037/0022-3514.82.6.956 | PubMed

Frederick S (2005) Cognitive reflection and decision making. *Journal of Economic Perspectives* **19**:25-42 https://doi.org/10.1257/089533005775196732

Gao Q, Jia X, Liu H, Wang X, Liu Y (2020) Attachment style predicts cooperation in intuitive but not deliberative response in one-shot public goods game. *International Journal of Psychology* **55**:478-486 https://doi.org/10.1002/ijop.12584 | PubMed

Glazer JE, Kelley NJ, Pornpattananangkul N, Mittal VA, Nusslock R (2018) Beyond the FRN: Broadening the time-course of EEG and ERP components implicated in reward processing. *International Journal of Psychophysiology* **132**:184-202 https://doi.org/10.1016/j.ijpsycho.2018.02.002 | PubMed

Green P, MacLeod CJ (2016) SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* **7**:493-498 https://doi.org/10.1111/2041-210X.12504

Hajcak G, Dunning JP, Foti D (2009) Motivated and controlled attention to emotion: Time-course of the late positive potential. *Clinical Neurophysiology* **120**:505-510 https://doi.org/10.1016/j.clinph.2008.11.028 | PubMed

Hong K, Bohnet I (2007) Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology* **28**:197-213 https://doi.org/10.1016/j.joep.2006.06.003

Jiang J, Chen C, Dai B, Shi G, Ding G, Liu L, Lu C (2015) Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences* **112**:4274-4279 https://doi.org/10.1073/pnas.1422930112 | PubMed

Joyce B, John D, Kevin M (1995) Trust, Reciprocity, and Social History. *Games and Economic Behavior* **10**:122-142 https://doi.org/10.1006/game.1995.1027

Kahneman D (2011) *Thinking, fast and slow* Farrar, Straus and Giroux.

Kahneman D, Tversky A (1979) Prospect Theory: An Analysis of Decision under Risk. *Econometrica: Journal of the Econometric Society* **47**:263-291 https://doi.org/10.2307/1914185

Koch A, Dorrough A, Glöckner A, Imhoff R (2020) The ABC of society: Perceived similarity in agency/socioeconomic success and conservative-progressive beliefs increases intergroup cooperation. *Journal of Experimental Social Psychology* **90**:103996

https://doi.org/10.1016/j.jesp.2020.103996 | PubMed

Komiya A, Mifune N (2015) An individual difference in betrayal aversion: Prosociality predicts more risky choices in social but not natural domains. *Letters on Evolutionary Behavioral Science* **6**:5-8 https://doi.org/10.5178/lebs.2015.33

Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. *Nature* **435**:673-676 https://doi.org/10.1038/nature03701 | PubMed

Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA, Poeppel D (2017) Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93**:480-490 https://doi.org/10.1016/j.neuron.2016.12.041 | PubMed

Kugler T, Connolly T, Ordóñez LD (2010) Emotion, Decision, and Risk: Betting on Gambles versus Betting on People. *Journal of Behavioral Decision Making* **25**:123-134 https://doi.org/10.1002/bdm.724

Kuhnen CM, Knutson B (2005) The neural basis of financial risk taking. *Neuron* **47**:763-770 https://doi.org/10.1016/j.neuron.2005.08.008 | PubMed

Lauharatanahirun N, Christopoulos GI, King-Casas B (2012) Neural computations underlying social risk sensitivity. *Frontiers in Human Neuroscience* **6** https://doi.org/10.3389/fnhum.2012.00213 | PubMed

Lenth RV (2022) Estimated Marginal Means, aka Least-Squares Means. https://CRAN.R-project.org/package=emmeans

Levine EE, Barasch A, Rand D, Berman JZ, Small DA (2018) Signaling emotion and reason in cooperation. *Journal of Experimental Psychology: General* **147**:702-719 https://doi.org/10.1037/xge0000399 | PubMed

Li J, Sun Y, Li M, Li H, Fan W, Zhong Y (2020) Social distance modulates prosocial behaviors in the gain and loss contexts: An event-related potential (ERP) study. *International Journal of Psychophysiology* **150**:83-91 https://doi.org/10.1016/j.ijpsycho.2020.02.003 | PubMed

Li M, Li J, Zhang G, Zhong Y, Li H (2023) Influence of social distance and promise levels on trust decisions: an ERPs study. *Acta Psychologica Sinica* **55**:1859-1871 https://doi.org/10.3724/SP.J.1041.2023.01859

Lockwood PL, Apps MAJ, Chang SWC (2020) Is There a 'Social' Brain? Implementations and Algorithms. *Trends in Cognitive Sciences* **24**:802-813 https://doi.org/10.1016/j.tics.2020.06.011 | PubMed

Lopez-Calderon J, Luck SJ (2014) ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience* **8** https://doi.org/10.3389/fnhum.2014.00213 | PubMed

Luck SJ (2014) *An introduction to the event- related potential technique* MIT press.

Luck SJ, Gaspelin N (2017) How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology* **54**:146-157 https://doi.org/10.1111/psyp.12639 | PubMed

Ma Q, Shen Q, Xu Q, Li D, Shu L, Weber B (2011) Empathic responses to others' gains and losses: An electrophysiological investigation. *NeuroImage* **54**:2472-2480 https://doi.org/10.1016/j.neuroimage.2010.10.045 | PubMed

Nowak MA (2006) Five Rules for the Evolution of Cooperation. *Science* **314**:1560-1563 https://doi.org/10.1126/science.1133755 | PubMed

Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. *Nature* **437**:1291-1298 https://doi.org/10.1038/nature04131 | PubMed

Panchanathan K, Boyd R (2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**:499-502 https://doi.org/10.1038/nature02978 | PubMed

Pion-Tonachini L, Kreutz-Delgado K, Makeig S (2019) ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* **198**:181-197 https://doi.org/10.1016/j.neuroimage.2019.05.026 | PubMed

Plessner H, Betsch C, Betsch T (2008) *Intuition in Judgment and Decision Making (1st ed.)* Psychology Press.

Rabin M (1993) Incorporating Fairness into Game Theory and Economics. *The American Economic Review* **83**:1281-1302 https://www.jstor.org/stable/2117561

Rand DG, Greene JD, Nowak MA (2012) Spontaneous giving and calculated greed. *Nature* **489**:427-430 https://doi.org/10.1038/nature11467 | PubMed

Righetti F, Finkenauer C, Finkel EJ (2013) Low Self-Control Promotes the Willingness to Sacrifice in Close Relationships. *Psychological Science* **24**:1533-1540 https://doi.org/10.1177/0956797613475457 | PubMed

Shank DB, Kashima Y, Peters K, Li Y, Robins G, Kirley M (2019) Norm talk and human cooperation: Can we talk ourselves into cooperation?. *Journal of Personality and Social Psychology* **117**:99-123 https://doi.org/10.1037/pspi0000163 | PubMed

Sloman SA (1996) The empirical case for two systems of reasoning. *Psychological Bulletin* **119**:3-22 https://doi.org/10.1037/0033-2909.119.1.3

Slovic P (1987) Perception of risk. *Science* **236**:280-285 https://doi.org/10.1126/science.3563507 | PubMed

Stanovich KE, West RF (1998) Individual differences in rational thought. *Journal of Experimental Psychology: General* **127**:161-188 https://doi.org/10.1037/0096-3445.127.2.161

Thielmann I, Hilbig BE (2015) Trust: An Integrative Review from a Person–Situation Perspective. *Review of General Psychology* **19**:249-277 https://doi.org/10.1037/gpr0000046

Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* **28**:675-691 https://doi.org/10.1017/s0140525x05000129 | PubMed

Tversky A, Kahneman D (1992) Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* **5**:297-323 https://doi.org/10.1007/BF00122574

van Honk J, Eisenegger C, Terburg D, Stein DJ, Morgan B (2013) Generous economic investments after basolateral amygdala damage. *Proceedings of the National Academy of Sciences* **110**:2506-2510 https://doi.org/10.1073/pnas.1217316110 | PubMed

Wang H, Ao L, Gao Y, Liu Y, Zhang X (2023) Empathy for pain in individuals influenced by moral identity: Evidence from an ERP study. *Physiology & Behavior* **266**:114202 https://doi.org/10.1016/j.physbeh.2023.114202 | PubMed

Wang Y, Jing Y, Zhang Z, Lin C, Valadez EA (2017) How dispositional social risk-seeking promotes trusting strangers: Evidence based on brain potentials and neural oscillations. *Journal of Experimental Psychology: General* **146**:1150-1163 https://doi.org/10.1037/xge0000328 | PubMed

West SA, Cooper GA, Ghoul MB, Griffin AS (2021) Ten recent insights for our understanding of cooperation. *Nature Ecology & Evolution* **5**:419-430 https://doi.org/10.1038/s41559-020-01384-x | PubMed

Wu CC, Sacchet MD, Knutson B (2012) Toward an affective neuroscience account of financial risk taking. *Frontiers in Neuroscience* **6** https://doi.org/10.3389/fnins.2012.00159 | PubMed

Yamagishi T (2011) *Trust: The evolutionary game of mind and society* Springer. https://doi.org/10.1007/978-4-431-53936-0

Yang J, Zhang H, Ni J, De Dreu CKW, Ma Y (2020) Within-group synchronization in the prefrontal cortex associates with intergroup conflict. *Nature Neuroscience* **23**:754-760 https://doi.org/10.1038/s41593-020-0630-x | PubMed

Yu M, Lu J, Wang S (2022) Forgiveness weakens counterempathy at the early and late stage of empathic responses to opponents' expressions. *Behavioral Neuroscience* **136**:114-125 https://doi.org/10.1037/bne0000496 | PubMed

Zaki J, Mitchell JP (2013) Intuitive Prosociality. *Current Directions in Psychological Science* **22**:466-470 https://doi.org/10.1177/0963721413492764

Zhang H, Yang J, Ni J, De Dreu CKW, Ma Y (2023) Leader–follower behavioural coordination and neural synchronization during intergroup conflict. *Nature Human Behaviour* **7**:2169-2181 https://doi.org/10.1038/s41562-023-01663-0 | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

The non-social task was a classic risky decision-making task with a binary choice between an option with a sure gain and a risky option with a probabilistic gain or loss. In the social task, the sure option was an individual gain (as in the non-social option) and the probabilities in the risky option, which were shown to participants, were framed as probabilities of other previous participants (i.e., "partners") to cooperate or not; a probabilistic gain (when the partner cooperated) also led to a gain of the partner, while a probabilistic loss meant that the partner would receive the amount lost by the participant. This loss was framed as "betrayal." The authors show differences in how probabilities and amounts (of gains/losses) affected choices, RTs, and ERPs (P3 and LPP).

Strengths:

Since participants faced decisions with the same individual payoffs in a non-social and a social condition, this setup made it possible to use identical standard analyses for choices, RTs, and ERPS as well as (almost) identical economic models for the two conditions.

Weaknesses:

(1) The task does not include many components that are usually considered central for cooperation or "betrayal" and this is not discussed appropriately. At the same time, the "emotional aspects" of the operationalized "betrayal" are not directly assessed.

a) The standard economic game for cooperation is the prisoner's dilemma, in which participants make independent choices at the same time without getting any explicit information on the cooperation probability of their partner before they make their decisions. Furthermore, most of the time the interactions are repeated. Actually, the trust game as one other frequently used economic game, also includes a back and forth of transfers between the partners. So, here, I am not so convinced by the operationalization of a low cooperation probability, which is shown before the decision, as "betrayal." The authors should motivate and explain their rationale more clearly in reference to such other tasks.

b) The setup of the task, especially the fake interaction with the fake partners, should be made clearer in the main text (before reporting the results). I would argue for including the task picture in the main text.

c) In general, I am in favour of taking participants' choice behaviour as the main outcome measure. But given the strong implications of "emotional costs" made by the authors, I would have expected some ratings of "betrayal" on a trial-by-trial basis. I would at least include this as a shortcoming.

d) Also, given the framing of the study, I would have expected some exploratory analyses regarding individual differences with respect to, e.g., social value orientation, etc. I would at least include this as an outlook.

(2) The standard statistical analyses could be improved.

a) It is good that the authors have rather long sections using standard regression analyses. But they are a bit lengthy, and the modelling should be more prominent.

b) In a couple of places, the authors say something like "this is significant, but that is not." Here, it has been made very clear that the interaction term needs to be looked at. As far as I can see, this has not always been done.

c) For this binary choice, the difference in expected value (EV) between the sure and the risky options is one crucial comparison. But the authors never take that into account. This difference does not depend on the amount, which the authors dub "principal." That is, the sure option simply has an EV of x, i.e., the amount. The risky option has the EV = p2x + (1-p)0.5x, with p being the probability of gain/cooperation. That is, the two options have the same EV at p=1/3, independent of x. This should be made clear.

d) Relatedly, RTs should depend on the differences in EV (and not so much on p or on x per se). This can be seen by the more or less quadratic relationship between p and RTs (Fig 1A), with a peak around a p of 1/3.

e) RTs are often log-transformed. It should be briefly mentioned why this was not done here.

(3) The modelling evidence is relatively weak. This is my main point.

a) (Cumulative) prospect theory should be introduced.

b) The models seem overly complicated with many free parameters. I would have expected some simpler versions and more comparisons between models that differ in just one parameter.

- e.g., it is really nice that the authors used a probability weighting function. BTW: Please describe this more clearly in the introduction and in the results. But for this limited range of probabilities, this might be too much.

- e.g., why directly assume two different exponents in the utility function for gains and losses, and in addition a loss aversion parameter lambda? Only lambda would be a better starting point here.

c) The differences in AIC (Figure 2A) seem rather minuscule, and the distribution of winning models is not very peaked. I am not convinced that Model 3 is the winning model.

d) Crucially, and related to the previous points, judging from Fig 2C, the "betrayal" parameter kappa seems to be zero for about half of the participants. The authors should look into this.

- Would a model just like model 3 but without kappa (i.e., kappa set to zero) perform better? Is this just model 2?

- How is kappa set in the non-social condition?

- This massive skew, to say the least, is never discussed.

- A correlation is definitely not warranted.

(4) The ERP results seem to me rather superficial. But I am not an EEG expert.

a) The authors do not seem to look at the outcome phase, which could be interesting for differences in reward/loss processing in the two task versions.

b) Again, differences in EV seem to be more important from a conceptual point than probabilities or amounts; see my comment 2d.

c) Also, the authors report ERPs for the two task types separately but do not seem to run proper comparisons between them, see my comment 2b.

(5) Preregistration: It should be made very clear early on that this study was not preregistered.

(6) Quality checks: The authors should check if some participants are outliers in terms of the number of missed trials, always choosing the same option, etc. It is notoriously difficult to find good post hoc reasons for excluding participants (one reason why replications and preregistrations are important). In any case, the data quality should be checked and described a bit more.

https://doi.org/10.7554/eLife.111043.1.sa3

Reviewer #2 (Public review):

Summary:

This paper investigates risk and cooperation decisions by integrating computational modeling with event-related potential (ERP) measures. Participants completed two tasks involving financial risk and cooperation under possible betrayal. The comparison between social and non-social decision-making is interesting and potentially valuable. However, the conceptual framing, theoretical grounding, and modeling rationale require substantial clarification.

Strengths:

(1) The paper introduces comparable tasks to probe social vs. non-social decision making.

(2) The authors use a model to identify a psychological distinction and test its validity using neural data.

Weaknesses:

(1) Conceptual framing and theoretical clarity

The primary theoretical contribution of the paper is currently unclear. Specifically, it is not clear what key difference the authors hypothesize between risk and cooperation conditions. This distinction should be grounded in prior literature.

The manuscript states: "Indeed, mutual cooperation maximizes social welfare, whereas betrayal benefits the trustee but comes at the trustor's expense in the Trust Game (Joyce et al., 1995)." However, the authors do not discuss the substantial literature on the Trust Game, which is used here but not explicitly acknowledged.

• The original Trust Game framework and behavior in one-shot settings (e.g., Berg et al., 1995).

• The persistence of cooperation even when defection is economically optimal (e.g., Berg et al., 1995; Fehr & Fischbacher, 2003).

• The influence of trustworthiness of the partner on cooperation decisions has been previously studied (Ma et al., 2022).

• Differences between social and non-social decision-making contexts have also been reported with matched tasks (Liu et al., 2024).

(2) Distinction between constructs (risk, loss aversion, betrayal aversion)

The introduction introduces multiple related constructs-risk aversion, loss aversion, and betrayal aversion-but does not clearly differentiate them. A theoretically grounded distinction is needed.

In particular:

• The manuscript introduces multiple related constructs, or maybe the terms are used interchangeably? The distinction between risk aversion, loss aversion, defection aversion, and betrayal aversion should be clearly defined.

• Betrayal aversion versus loss aversion is introduced but not clearly differentiated. Importantly, it should be clarified that this distinction is not experimentally manipulated but instead inferred through computational modeling. This point is currently not made explicit, which leads to confusion in the introduction

• The computational model should be introduced clearly in the introduction. Without explaining how these constructs are operationalized in the model, the framework is difficult to follow.

The statement "In the risk task, losses were solely impersonal" is also unclear. It seems the authors may mean "personal or non-social" rather than "impersonal" as rewards are always personally relevant.

(3) Hypotheses and preregistration

The manuscript would benefit from more theoretical rationale for hypotheses. For example:

• What is the basis for hypothesizing that financial loss aversion and betrayal aversion independently affect cooperation choices?

• Why should these constructs be separable and modeled independently?

• Additionally, the absence of preregistration is a limitation that should be acknowledged even more.

• Given the flexibility of the modeling approach and number of parameters, this is particularly important.

• For instance, the rationale for focusing on decision times is also not clearly explained and should be better motivated.

(4) Computational modeling

There are several concerns regarding the modeling approach:

• The choice of model comparison metric should be justified. Why is AIC used rather than BIC, which penalizes model complexity more strongly? This is particularly relevant given the inclusion of additional parameters to capture processes not directly measured by the task.

• Full model recovery analyses are missing. A full model recovery is necessary to demonstrate that competing models produce distinguishable behavioral patterns. This needs to be shown in order to justify the specificity of the winning model

• How correlated are the parameters across participants, particularly loss and betrayal parameters?

• More broadly, it is unclear how well loss aversion and betrayal aversion can be differentiated based on behavior alone. If these constructs are separable, they should predict

distinct aspects of behavior.

(5) ERP analyses

The ERP results (e.g., P300 and LPP) seem to suggest that betrayal aversion is relevant in both time periods and similarly.

• Do neural signals differentially reflect betrayal aversion versus loss aversion earlier and later on?

• Are there significant interaction effects between betrayal and loss aversion for each ERP component?

https://doi.org/10.7554/eLife.111043.1.sa2

Reviewer #3 (Public review):

Summary:

In this study, the authors aim to address two questions. First, do people avoid cooperation primarily because of betrayal aversion beyond loss aversion? Second, can the effects of betrayal aversion and loss aversion be dissociated at the behavioral and neural levels? To address these questions, the authors compared individuals' choices of taking risks in a nonsocial risk task with those in a social cooperation task, with the two tasks matched in success probability and principal amount. They fitted computational models that include betrayal-aversion and loss-aversion terms and related the model parameters to ERP measures. Based on these analyses, the authors concluded that betrayal aversion has a stronger effect on cooperation than loss aversion and that betrayal is encoded earlier than loss in the brain. This is an important research question, and the attempt to combine computational modeling with ERP analysis is valuable. However, the current data analyses may not be able to support all the conclusions the authors made. For instance, the claims concerning the dissociation between betrayal aversion and loss aversion are not yet sufficiently supported by the evidence.

Strengths:

(1) The research question is theoretically important. Distinguishing betrayal aversion from loss aversion is important for research on trust, cooperation, and risky decision-making.

(2) The approach of integrating behavioral measures, self-report ratings, computational modeling, and ERP data is valuable and gives the study significance.

(3) The behavioral findings are broadly consistent. Participants reported stronger emotional responses in the cooperation task and were less willing to accept risk in the cooperation condition. These findings are generally in line with previous work on betrayal aversion and provide a reasonable manipulation check for the contrast between social and nonsocial risk.

Weaknesses:

(1) The manuscript states that the two tasks are matched in probability and principal amount, but the cooperation task additionally introduces partner outcomes, betrayal, and prosocial components. The Methods section states that, in the cooperation task, if both players cooperate, the principal is doubled and then split equally; if the partner betrays, half of the participant's principal is transferred to the partner. The model also includes an expected-other-reward term, namely, $V_{other}=\omega[p\cdot 2X+(1-p)\cdot 1.5X]$. This raises an interpretive concern: if the two tasks differ not only in whether the source of uncertainty is social, but also in partner outcome, intentionality, and potential inequity structure, then the fitted "betrayal aversion" parameter may in fact reflect multiple motives rather than betrayal aversion alone.

In the current experimental design, the "betrayal aversion" parameter may not be uniquely interpretable as a pure betrayal-specific construct, and the current evidence is insufficient to support such a specific interpretation.

(2) Participants were informed that the cooperation probabilities were derived from previous real participants, whereas in fact these probabilities were randomly generated. In addition, six participants explicitly expressed doubts about the authenticity of the social interaction, yet the authors retained these participants with only the brief statement that this "did not affect the results." For such a critical manipulation, this explanation is too brief. I recommend that the authors report robustness analyses excluding skeptical participants. Since six participants reportedly doubted the authenticity of the social interaction, and some participants also performed poorly on the catch trials, it would be important to show whether the main behavioral, modeling, and ERP findings remain after excluding these participants. This is especially important because the manuscript's central interpretation depends on the assumption that the cooperation task was genuinely experienced as social.

(3) The descriptions of the sample size are inconsistent across sections. The Participants section states that, after excluding one participant for misunderstanding the instructions, the final sample consisted of 49 participants; however, the behavioral results section later states that only 42 participants were included in the final analyses due to recording problems. This discrepancy is important because readers need to know clearly which sample was used for the behavioral analyses, which for the model fitting, and which for the ERP analyses; whether these analyses were conducted on the same participants; and whether the exclusion criteria were consistent across analyses. The manuscript needs a more transparent description of sample size and exclusion criteria.

(4) The authors need to do more thorough analyses to validate their models. In addition to AIC and parameter recovery, I would encourage the authors to include other model comparison metrics where possible, such as BIC and exceedance probability, as well as model-recovery analyses. The authors should also do model-based simulation analyses to show that the winning model can capture the contextual effects observed in real data.

(5) The authors should explain the rationales for the choice of ERP time windows and component selection in more detail. The current ERP analyses are time-locked to principal onset, and P3/LPP are extracted from fixed time windows. The authors should explain why this is the most appropriate time-locking point for examining betrayal- and loss-related computations, and why alternative time-locking points, such as probability-cue onset or other key task events, were not used. More importantly, the time windows of P3 and LPP are defined arbitrarily in the current analyses. The authors need to apply a more principled approach to define ERP components. It looks like the P3 and LPP are from the same ERP component in Figure 3.

(6) The manuscript has several internal inconsistencies in terminology, figure references, and result descriptions. These issues weaken the clarity of the arguments and reduce the readability of the manuscript.

(7) The authors partially achieved their aims. The study does provide evidence that social risk and nonsocial risk are not treated equivalently, and it also offers a computational framework that is informative for the field. This is an important topic, and the overall approach is promising.

https://doi.org/10.7554/eLife.111043.1.sa1

Author response:

We agree that the manuscript would benefit from a more clearly articulated conceptual framing, stronger model validation, more explicit statistical and ERP comparisons, and improved transparency regarding task design, sample inclusion, and preregistration. In the revised manuscript, we plan to address these points through substantial revision of the Introduction and Discussion, along with additional robustness and validation analyses, and more cautious interpretation of the main findings.

Reviewer #1 raised important points about the framing of the cooperation task, the interpretation of betrayal, the standard statistical analyses, the modelling, and the ERP analyses. In response, we plan to clarify that the present task captures betrayal-related social risk or anticipated partner defection, rather than betrayal in its full interpersonal and emotional sense, and to better motivate this operationalization with reference to the betrayal-aversion and trust-game literature. We will moderate our claims regarding "emotional costs," incorporate a more explicit task overview and accompanying schematic into the main text, and frame individual differences as a key avenue for future research. In addition, we will streamline the standard behavioral analyses, make the expected-value structure of the task explicit, add EV-based analyses of choice and reaction time, strengthen the ERP analyses, clarify that the study was not preregistered, and provide a complete report of data-quality checks. For the modelling section, a central revision will be to simplify the model structure and refit the models using a Bayesian hierarchical approach.

Reviewer #2 emphasized the need for stronger theoretical framing and more specific distinctions between related constructs. In the revised manuscript, we will substantially revise the Introduction to better situate the present task in relation to the Trust Game literature and prior work comparing social and non-social decision-making under matched payoff structures. We will also define risk aversion, loss aversion, anticipated partner defection, and betrayal-related aversion more explicitly, and clarify that the distinction between betrayal-related aversion and loss aversion is inferred through computational modelling rather than directly manipulated as separate experimental factors. We also plan to introduce the computational model earlier in the manuscript, clarify how the key constructs are operationalized, replace unclear wording such as "impersonal losses," strengthen the rationale for our hypotheses, and acknowledge the lack of preregistration more clearly.

Reviewer #3 highlighted the need to align our conclusions more closely with the current evidence. In the revised manuscript, we will moderate the interpretation of the betrayal-related parameter, acknowledging that the cooperation task differs from the non-social risk task not only in social versus non-social uncertainty, but also in partner outcome, intentionality, and potential inequity structure. We therefore plan to avoid treating this parameter as a pure betrayal-specific construct and to describe it more cautiously as capturing betrayal-related social risk or aversion to anticipated partner defection. We also plan to report robustness analyses excluding participants who expressed doubts about the social interaction, as well as participants with poor catch-trial performance or otherwise low-quality data, and to clarify the sample sizes and exclusion criteria used for behavioral, modelling, and ERP analyses. Finally, we will strengthen model validation and ERP reporting, including broader validation analyses and more cautious interpretation if the evidence for temporal dissociation between betrayal-related aversion and loss aversion proves weaker than currently stated.

Across these revisions, we also intend to simplify the model structure and use Bayesian hierarchical fitting to strengthen model validation, while avoiding overly strong claims if the additional analyses provide only modest support for a single preferred model.

https://doi.org/10.7554/eLife.111043.1.sa0