
Beyond the Payoff Matrix: Emotions, Personality, and Sex in Evolution of 'Irrational' Human Behavior

Victoria V. Rostovtseva¹,
Anna A. Mezentseva¹, Emil R. Imaeev^{1, 2},
Yaroslav O. Sizov¹, Alexandr V. Latanov²,
Marina L. Butovskaya¹

¹ *Institute of Ethnology and Anthropology of the Russian Academy of Sciences, Moscow, Russia*

² *Lomonosov Moscow State University, Moscow, Russia*

ABSTRACT

Human social behavior is a result of complex interplay of numerous factors, personality traits and emotional reaction being among the basic agents for regulation of social relationships and stability of social networks. Fairness and moralistic punishment presumably evolved as instruments for strengthening ingroup cooperation. Contrary to traditional game-theoretic assumptions, humans often deviate from Nash strategies in cooperation. Emotions were thought to be among the sources of such 'irrationality'. We explore how do emotional reactions and participants' personalities interplay to modulate prosocial and punishing behavioral strategies. Participants engaged in a sequence of one-shot dyadic games, interacting face-to-face with the same-sex partner, who was a part of the experiment ('actor'). Emotions were provoked by an extremely unfair and impudent decision of the partner. The tendency to punish unfair partner at own cost was associated with proactive aggression of participants. Counterintuitive,

Recommended citation: Rostovtseva V. V., Mezentseva A. A., Imaeev E. R., Sizov Y. O., Latanov A. V., Butovskaya M. L. Beyond the Payoff Matrix: Emotions, Personality, and Sex in Evolution of 'Irrational' Human Behavior. *Social Evolution & History*, Vol. 25 No. 2, September 2026, pp. 150–176. DOI: 10.30884/seh/2026.02.06.

© 2026 'Uchitel' Publishing House

proactively aggressive men refrained from punishment, while proactively aggressive women, were prone to punish. Men also cooperated more often, than women, despite bad partner's reputation. Similar effects, but related to psychological comfort, were also revealed: 'agreeable' and 'introvert' participants cooperated more often with unfair partner, while 'extraverts' defected. Our results indicate that emotional load in combination with basic personality dispositions generate sex-specific patterns of 'irrational' strategic decisions, prioritizing 'costs' and 'benefits' outside the payoff matrix of the game.

Keywords: cooperation, prosocial punishment, emotions, aggression, Big Five, decision-making, game theory.

INTRODUCTION

Ultrasociality is one of the key human features, evolved, among others, as a result of developed cognitive abilities, language, and culture (Butovskaya and Rostovtseva 2021; Rostovtseva 2026). Advanced cooperative skills have enabled humans to master ecological and evolutionary niche construction, develop unique adaptability through technology (Laland and O'Brien 2011; Laland *et al.* 2016; Leonard 2018; Rainey 2023), organize life in large-scale societies (Turchin 2013). Unlike other species, human cooperation extends beyond kinship to include interactions with unrelated individuals and complete strangers, even in situations with no future opportunity for reciprocity (Melis and Semmann 2010; Henrich and Muthukrishna 2021). This remarkable capacity is further regulated by external enforcement of prosocial behavior, such as cultural norms, morality, ideas of social justice and their control via legal institutions (de Almeida 2020; Butovskaya *et al.* 2022).

In the twentieth century, cooperative processes were widely studied within the framework of economic game theory (von Neumann and Morgenstern 1944; Nash 1950, 1951; Luce and Raiffa 1957; Rapoport and Chammah 1965; Axelrod 1984). One of the most remarkable achievements of game theory of those times was the development of the concept of Nash equilibrium (Nash 1950, 1951). Nash strategies came to be considered 'rational' behavior. Traditional economic game theory assumed that humans are (or should be) 'rational' actors seeking to maximize individual payoff. However, empirical evidence that emerged later clearly demonstrate that humans often deviate from following Nash strategies, both in experiments and in real life (Henrich *et al.* 2004; Gintis 2014). Today we know that fully

‘rational’ human beings (*Homo economicus*) do not exist (Henrich *et al.* 2004; Levitt and List 2008; Urbina and Ruiz-Villaverde 2019). Recognizing that rationality *per se* is a very sophisticated concept, in this study, we explore how inherent human features such as emotions, personality and sex differences impact human strategic behavior.

In our study, we make use of experimental game theory to investigate multiple facets of human prosocial behavior such as trust and trustworthiness (Rostovtseva *et al.* 2023a; Rostovtseva 2024), costly punishment of antisocial behavior (Fehr and Gächter 2002; Balliet *et al.* 2011a; Wu *et al.* 2022), the dynamics of cooperation and prosocial expectations in a series of experimental interactions. We explore the following behavioral tendencies: (1) the extent to which individuals are willing to engage in costly punishment of antisocial behavior when interacting face-to-face with an explicitly unfair partner, (2) an individual's capacity to maintain cooperative intentions towards an interaction partner, even after that partner has broken their trust, (3) an individual's expectations about the future cooperative intentions of a partner with a known negative reputation – after that partner demonstrates willingness to cooperate. To test these behavioral tendencies, we used one-shot dyadic economic games: Prosocial Punishment Game (explores punishing dispositions) (Peysakhovich *et al.* 2014; Rostovtseva *et al.* 2023b), Prisoner's Dilemma game (explores cooperative tendencies) (Rapoport and Chammah 1965; Axelrod 1984; Rostovtseva *et al.* 2020), Ultimatum Game (estimates sense of fairness and prosocial expectations) (Güth *et al.* 1982; Ensminger and Henrich 2014; Rostovtseva *et al.* 2023b).

These behavioral tendencies, while important to human society, highlight a key tension: from classical game theoretic perspective, they appear ‘irrational’, at least from short-term perspective. Consider costly prosocial punishment, a mechanism essential for maintaining long-term cooperation (Fehr and Gächter 2002; Balliet *et al.* 2011a; Frey and Rusch 2012; Butovskaya and Rostovtseva 2021; Wu *et al.* 2022; Rostovtseva 2026). Prosocial punishment has deep evolutionary roots (Raihani *et al.* 2012), and plays important role in large-scale human societies, where selfishness, free-riding, and cheating are spreading easier compared to traditional small-scale communities (Frey and Rusch 2012; Butovskaya and Rostovtseva 2021, although see: Henrich *et al.* 2006). From a strict payoff-maximizing standpoint, sacrificing one's own resources to punish antisocial strangers seems pointless. The same can be said for trying to cooperate with someone known for untrustworthiness, especially under condition where cooperation is

extremely vulnerable to exploitation (*e.g.*, in one-shot Prisoner's Dilemma game, where defection of all players is the only 'rational' equilibrium) (Rapoport and Chammah 1965; Axelrod 1984).

All these aspects of human prosociality have been already extensively investigated in a huge number of studies worldwide (to mention just a few reviews and meta-analyses: Balliet 2010; Balliet *et al.* 2011a, 2011b; Balliet and Van Lange 2013a, 2013b; Thielmann *et al.* 2020; Spadaro *et al.* 2022, 2023). However, most of the studies concentrate on 'neutral' interactions, either examining behavior in completely anonymous settings (with limited, if any, information about the interaction partner) or incorporating some degree of personalization (such as displaying a photo or video of the partner) (Rostovtseva 2026). These experiments offer valuable insights into broad patterns of human trust, punishment, and cooperation, as well as sense of fairness in relatively neutral settings. However, given that humans are deeply emotional creatures, there is no wonder that emotions pervade virtually every facet of human experience. Real-life scenarios involving trust, trustworthiness, fairness, and other prosocial or selfish interactions are often highly tuned with emotions. Consequently, a key issue for understanding cooperation in the real world is to examine how emotions modulate prosocial responses.

The interplay between emotions and reasoning in cooperative contexts is not a novel topic (see, *e.g.*, Frank 1988). Nowadays, the role of emotions in prosocial behavior and economic decision-making is widely studied by various branches of psychology, economics and neuroscience (Serra 2021; Van Kleef and Lelieveld 2022; Al-Shawaf and Shackelford 2024). Closest to our study is the neuroeconomic concept of emotional modulation of decision-making (Lempert and Phelps 2014; Serra 2021; Dennison *et al.* 2022). This field examines how the decision itself, or its outcome, induces an emotional reaction that then modulates behavior in economic contexts (Lempert and Phelps 2014; Serra 2021). These studies typically aim to understand how emotional arousal shifts choices, and predict or modulate decision-making by fixing or manipulating emotional reactions and accompanying physiology (Lempert and Phelps 2014; Bossaerts 2021; Serra 2021).

Our experiment examines prosocial behavior in an emotionally charged environment. We established this setting by directly manipulating participants' experiences in experimental games. Participants' interaction partners in our experiment were actually members of the research team ('actors') making their decisions according to a unified study protocol. 'Actors' were introduced to the participants as fellow

study subjects. We began with the Trust Game (Rostovtseva *et al.* 2023a; Rostovtseva 2024) to provoke negative emotions. Negative affect was induced from the start of interactions by an extremely unfair and impudent decision of the partner ('actor'). Consequently, those participants that entrusted their funds to the partner in the context of the Trust Game were left emptyhanded. They watched as their partner, face-to-face, announced the decision, resulting in a complete loss of the participant's initial points and their corresponding monetary value.

Beyond manipulating the experimental environment, we also explored the role of individual differences in personality. We assessed these using two sets of self-reported measures: Reactive/Proactive Aggression (Raine *et al.* 2006) and the Big Five personality traits (Rammstedt and John 2007; Shchebetenko 2014). Such traits represent baseline behavioral predispositions and reactions of individual, which we expected could influence decision-making in emotionally charged contexts. We anticipated that reactive aggression, which is often considered an emotional aggressive response, would be particularly relevant to punishing behavior. We also expected that the Big Five personality traits, which encompass dimensions related to both emotionality and sociability, could influence behavior across a range of experimental scenarios.

Our study prioritized capturing the complexity of real-life social interactions, despite possible loss of some degree of rough experimental control (internal validity). We believe this emphasis on external validity makes up a key strength of our investigations.

DATA AND METHODS

The aim of our experiment was to explore cooperation and prosocial punishment under high emotional stress. Participants of the study were 173 individuals, young Russian men ($N = 85$) and women ($N = 88$), predominantly students aged 17 to 31 years (mean age 21 ± 2.5 y.) residing in Moscow. All participants were native Russian-speakers. Individual cooperative and punishing tendencies were tested in a series of one-shot dyadic economic games adopted from game theory: Trust Game (TG) (Kreps 1990; Berg *et al.* 1995), Prosocial Punishment Game (PPG) (Fehr and Gächter 2002), Prisoner's Dilemma (PD) (Rapoport and Chammah 1965; Axelrod 1984), and Ultimatum Game (UG) (Güth *et al.* 1982). Each of these games is designed to estimate individual proneness to (1) trust and behave trustworthily, (2) punish unfair behavior; (3) cooperate or free-ride; (4) behave fairly under threat of sanctions. Decisions in the games resulted in payoffs in

points according to certain game rules or payoff matrices. Participants were informed that at the end of experiment all their individually earned points will be transferred into real money, but the exact exchange rate was not disclosed. At the end of experiment each participant received a payoff up to 1000 RUB, but highly dependent on individual performance. The sequence of the games was fixed as listed above, and was the same for each participant. However, participants were not informed on the essence, sequence, and number of the games beforehand, the rules of each game and the payoff structure was explained only prior to each certain game. The games were played face-to-face, that is, each participant made decisions directly facing the same-sex interaction partner. Each couple of decisions was provided with an immediate feedback on the outcome of the interaction.

Experimental Games

The first game in a sequence was the Trust Game (TG), which is a two-player game where one player acts as the trustor and the other as the trustee. While many TG variations exist (Kreps 1990; Berg *et al.* 1995; Breuer *et al.* 2016; Andreozzi *et al.* 2020), we used a version where the trustor makes a binary choice (either keeps or entrusts the whole amount of the endowment to the trustee), while the trustee has a range of options how much to return to the trustor (Jaeger *et al.* 2022; Rostovtseva *et al.* 2023a). The trustor began with an endowment of 100 points and decided whether to keep it or entrust it to the trustee. If entrusted, the points were multiplied by the experimenter by the factor of six. The trustee, now possessing 600 points, then determined how much to return to the trustor: any amount from 0 to 600.

The second game was the Prosocial Punishment Game (PGG) (Fehr and Gächter 2002; Rostovtseva *et al.* 2023b), and was played only if in the first game participant did entrust the endowment to the trustee. Entrusting their funds, participants always received an unfair return. In the subsequent PGG they had an opportunity to punish an unfair behavior of the interaction partner at own cost. For this game, the participant was endowed with a new budget of 100 points, and could either (1) keep all 100 point without imposing any punishment on the unfair partner, (2) invest 50 points out of 100 to take away half of the points (300 points) from the unfair partner ('moderate punishment'), or (3) invest all 100 points out of 100 to take away all points (600 points) from the interaction partner ('extreme punishment'). Importantly, the points taken away from the unfair partner did not go to the participant, but were simply removed from the game.

The third interaction was the Prisoner's Dilemma (PD) (Rapoport and Chammah 1965; Axelrod 1984; Rostovtseva *et al.* 2023b). In this game, two players make simultaneous choices without any verbal or other intended communication: to cooperate or to defect. The subsequent payoffs are structured as follows: mutual cooperation yields 60 points for each player, while mutual defection yields 30 points for each. If one of the interaction partners cooperates and the other defects, the defector receives a substantial 100 points, leaving the cooperator with zero payoff. While mutual cooperation yields the highest total payoff (120 points for pair), game theory suggests that defection is the only 'rational' individual strategy in this game (Nash equilibrium) (Nash 1950, 1951). Cooperation in the PD is risky because a cooperator receives nothing if the other player defects, while defection may bring the highest possible individual payoff (100 points) or at least 30.

The last game was the Ultimatum Game (UG) (Güth *et al.* 1982; Rostovtseva *et al.* 2023b). In the UG, a participant was endowed with a new 100 points, and had to suggest how to distribute them between self and the interaction partner. Any kind of distribution could be suggested, including 0/100, but if the interaction partner would not accept the suggestion (by any personal reasons), both partners end up with zero payoff. If the interaction partner accepts the suggestion, the points are shared according to the proposition of the participant. The suggestion could be made only once. Again, any prior communication was prohibited.

Manipulation of Emotional Load

To ensure an emotionally charged environment of the experiment, participants interacted with partners who were part of the research team ('actors'), who made provocatively unfair decisions from the very beginning of the experiment. There were two 'actors' involved in the study – one male and one female – who were of the same nationality and belonged to the same age cohort as the participants of the study. All male participants interacted with the same male 'actor', and all female participants with the same female 'actor'. The 'actors' made their decisions according to the study protocol, acting independently of participants' decisions. In the first interaction – TG – the participant of the study was always in a role of trustor, and the 'actor' always took the role of trustee. 'Actor' always kept all funds entrusted by a participant, leaving the trustor (participant) emptyhanded. This first decision set the general tone for the experiment, being extremely unfair

to a participant who entrusted all own funds to the partner. In the following PGG the ‘actor’ did not make any decision, while the participant decided whether and how to punish an unfair partner. In the third game – PD – the ‘actor’ always chose to cooperate, and in the last game – UG – the ‘actor’ accepted suggestions of only those participants, who distributed not more than 50 % of the endowment in their own favor. Participants of the study were not aware they were interacting with ‘actors’. Instead they were informed that they are facing the same kind of the study subject as they are.

Personality Measures

Our study involved two self-reported personality questionnaires: (1) the Reactive-Proactive Aggression Questionnaire (RPQ) (Raine *et al.* 2006), and (2) a 10-item short version of the Big Five Inventory (BFI-10) (Rammstedt and John 2007).

The RPQ is used to assess reactive and proactive types of aggression, which are known to be two distinct forms of aggressive behavior (Merk *et al.* 2005; Romero-Martinez *et al.* 2022). Researchers generally describe proactive aggression as more provoking, goal-oriented, and less emotionally-driven, while reactive being more characterized by emotional lability, impulsiveness, as affective aggressive response. The questionnaire consists of 23 items (statements) each rated on a three-point scale (from 0 to 2: 0 – ‘never do this’, 1 – ‘sometimes do this’, 2 – ‘often do this’). It yields separate scores for reactive and proactive aggression, calculated by summing the ratings for individual items (accounting for several reversed items). The RPQ has been used extensively in studies across many cultures (Fung *et al.* 2009; Tharp *et al.* 2011; Cima *et al.* 2013; Pechorro *et al.* 2017; Chen *et al.* 2024). We used a well-established Russian translation of the RPQ, which has been tested on Russian populations before (Lazebny *et al.* 2020). Since both subscales of the RPQ are known to be highly correlated between each other, we followed the developers’ recommendations, and used residualized scores as measures of pure reactive and pure proactive aggression (Raine *et al.* 2006).

The BFI-10 was used as an official Russian adaptation of the original version of the questionnaire (Shchebetenko 2014). It consists of 10 statements each rated on a five-point scale (from 1 – ‘fully disagree’ to 5 – ‘fully agree’). The questionnaire has five-factor structure, assessing such traits as agreeableness, extraversion, conscientiousness, neuroticism, and openness to new experience. Scores for each sub-

scale are calculated by summing up the ratings for individual items (accounting for reversed items).

Experimental Procedure

Upon arrival, participants were seated in a room where they provided written consent, completed demographic and psychological questionnaires, and underwent some unrelated measurements. Subsequently, each participant was taken individually to a separate room and met by the experimenter. Shortly after, the second ‘participant’ (‘actor’) was brought into the same room, and both were seated facing each other at the table. The same experimenter, wearing neutral clothing and a face mask to prevent emotional cues, conducted the entire experiment. Participants were instructed to avoid any verbal or intentional nonverbal communication unless specifically asked to. The experimenter explained the rules of each game to both players prior to each interaction. In each game, participants announced their decisions aloud, and then the experimenter announced the results of the interactions. The only exception was the PD game, where participants secretly wrote down simultaneously their decisions (cooperate/defect) on provided papers, before announcing them aloud. Throughout the experiment, both players maintained full visual contact. At the end of the games, participants received their monetary payoffs, and were walked out of the experimental space.

Participants in this study were recruited from diverse educational Moscow institutions. To maintain the integrity of the experiment, it was crucial that participants who had completed the study did not disclose information about the experimental procedure (especially about the fact that the second player was constantly making extremely unfair decisions). To ensure this, we employed two primary strategies: inviting groups from each institution only once, and ensuring participants waited for their turn in the first experimental room not being able to communicate with those who had already participated. To further minimize possible recognition, the actors changed their hair and clothing styles (changed t-shirts, jackets, accessories, hats, glasses) before each interaction with a new participant.

RESULTS

Only three participants declined to entrust their 100 points in the initial TG. This decision had a cascading effect, preventing them from disappointment in the TG and precluding their participation in the subsequent

PGG (prosocial punishment of the unfair response). Consequently, these participants were removed from the analysis. The remaining 170 participants all trusted their partner with their 100 points. However, in every instance, the partner ('actor') chose to keep the entire multiplied amount of 600 points, demonstrating a lack of reciprocity.

Analysis of Decisions

Figure 1a demonstrates that, following an unfair partner's decision, most participants chose not apply costly punishment in the following PPG. However, as depicted in Figure 1b, females were more inclined to use punishment than males. Since the frequencies of 'moderate' and 'extreme' types of punishment did not differ considerably, we combined these two types into a single category, 'punishment', for the further analysis. A chi-squared test of independence considering two options (not to punish, punish somehow) confirmed that the use of costly punishment in response to the unfair partner differed significantly between the sexes, with women applying it more frequently (43 % of women) than men (25 % of men): $X^2 = 6.141$, $df = 1$, $p = 0.013$.

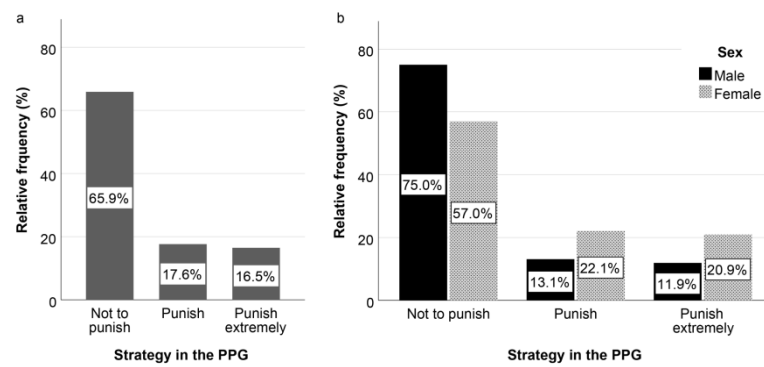


Fig. 1. Distributions of decisions in the Prosocial Punishment Game

Strategy in the PPG: not to punish (kept the 100 points, refrained from punishing the partner), punish (invested 50 points out of 100 to reduce the partner's funds by half; 'moderate punishment'), punish extremely (invested their entire 100 points to take away all of the partner's funds; 'extreme punishment'). (a) Relative frequencies for the total sample ($N = 170$) are displayed; (b) relative frequencies for men ($N = 84$) and women ($N = 86$) are displayed.

As shown in Figure 2a, the distribution of decisions in the subsequent PD game was shifted towards defection, which is not surprising. However, even after the negative experience in the first TG, the over-

all cooperation rate remained relatively high, with 43.5 % of participants choosing to cooperate. A chi-squared test of independence revealed significant sex differences in decisions ($X^2 = 6.812$, $df = 1$, $p = 0.009$), with women defecting more frequently (66.3 % of women) than men (46.4 % of men) (Fig. 2b). No significant associations between decisions in the PPG and the subsequent PD game for either men or women were revealed (based on a chi-squared test of independence). That is, the decision to cooperate in the Prisoner's Dilemma was not related to whether the participant had used prosocial punishment in the preceding game.

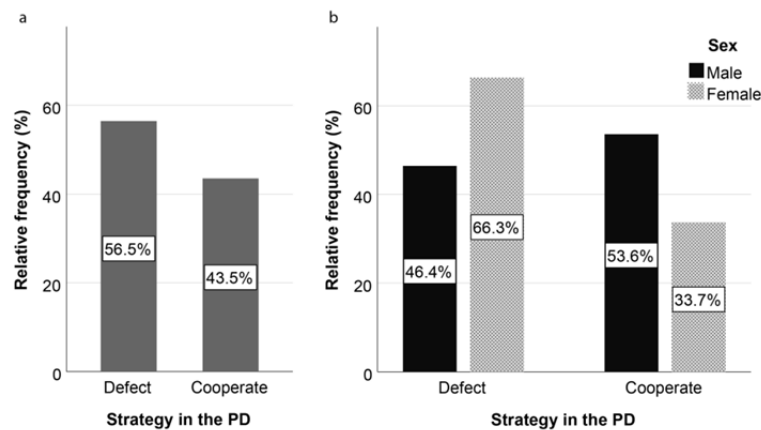


Fig. 2. Distributions of decisions in the Prisoner's Dilemma

(a) Relative frequencies for the total sample ($N = 170$) are displayed; (b) relative frequencies for men ($N = 84$) and women ($N = 86$) are displayed.

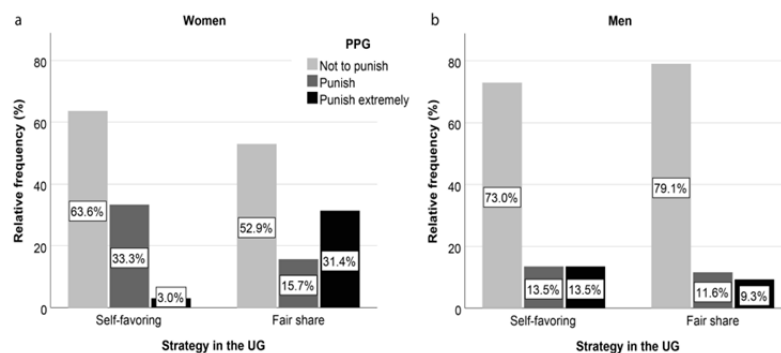


Fig. 3. Association of decisions between UG and PPG

The frequencies of decisions in Prosocial Punishment Game (PPG) per type of decisions in the Ultimatum Game (UG) are presented. Decisions in PPG: not to punish (kept the 100 points, refrained from punishing the partner), punish (invested 50 points out of 100 to reduce the partner's funds by half; 'moderate punishment'), punish extremely (invested their entire 100 points to take away all of the partner's funds; 'extreme punishment'). Decisions in UG: self-favoring (distributed more than 50 % of the endowment in own favor), fair share (proposed equal split). **(a)** Distributions for female sample ($N = 84$); **(b)** distributions for male sample ($N = 80$).

After the PD, participants moved to the UG, where they proposed a split of 100 points. A key feature of the UG is that the partner could reject the proposed split, resulting in both players receiving nothing. Only four men and two women offered their partner more than 50 % of the endowment. A majority of participants (51.2 % of men and 59.3 % of women) proposed an equal split (fair share), while the rest proposed splits that favored themselves. Given a low number of partner-favoring suggestions, we focus further analysis only on two types of decisions: equal split and self-favoring suggestions. As indicated by the percentages of equal split offers mentioned above, there were no significant sex differences in the UG decisions (this was also confirmed by a Mann-Whitney U-test on the raw, continuously scaled decisions). However, sex differences were revealed in associations between decisions in the UG and punishing decisions in the PPG. According to a chi-squared test of independence, women who proposed an equal split (fair share) in the UG were prone to apply 'extreme punishment' in the PPG ($X^2 = 11.112$, $df = 2$, $p = 0.001$) (Fig. 3a). This association was not evident in men (Fig. 3b).

This section of our findings highlights significant sex differences in responses to unfairness and breaches of trust. Under conditions of high negative emotional stress, women were more inclined to punish unfair behavior and less inclined to cooperate following an experience of partner untrustworthiness. Moreover, only among women was the strict equality of distributions in the Ultimatum Game positively related to propensity for extreme costly punishment.

Impact of Personality on Punishing and Cooperative Decisions

In this section, we tested how personality traits affected decisions in the experiment. First, we focus on the proactive and reactive aggression measures. Due to incomplete responses to the questionnaire, only 161 participants (81 men and 80 women) were included in the analy-

sis. Raw proactive and reactive aggression scores were nearly symmetrically distributed. Men on average scored higher on proactive aggression, whereas women scored higher on reactive aggression, however, according to an independent-samples *t*-test, these differences were not significant in either case. The high correlation between the proactive and reactive subscales (Pearson's $r = 0.537$, $p < 0.001$) reinforced the decision to use residualized scores, as measures of pure proactive and reactive aggression in the further analyses (see Methods). In case of residualized measures, men on average scored higher on proactive aggression, however, the sex difference was also non-significant.

We observed that individual aggressiveness only influenced behavior within the context of costly prosocial punishment (decisions in the PPG). Across all other games, aggressiveness had no any impact on behavior. In the PPG, only proactive aggression had a significant effect. Interestingly, the effect of proactive aggression diverged by sex. Men high in proactive aggression showed a decreased propensity to apply costly punishment, while women high in proactive aggression exhibited the opposite pattern, displaying an increased inclination to punish. Figure 4 illustrates this result. The supporting statistics is presented in Table 1.

The BFI-10 personality traits were approximately symmetrically distributed, except for openness to new experience, which was negatively skewed. Independent-samples *t*-tests revealed no significant sex differences in agreeableness, conscientiousness, and extraversion. However, women scored significantly higher than men in neuroticism ($N = 169$, $t = -3.850$, $df = 167$, $p < 0.001$) and also in openness to new experience, according to an independent-samples Mann-Whitney *U*-test ($N = 169$, $U = 4383.0$, $p = 0.009$).

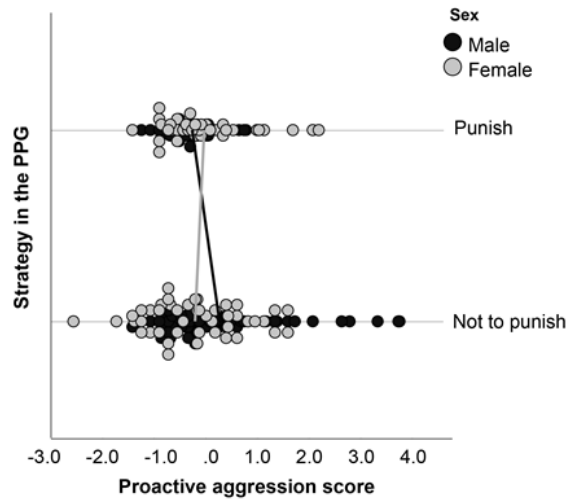


Fig. 4. The effect of proactive aggression on decisions in PPG

Strategy in the Prosocial Punishment Game (PPG): not to punish (kept the 100 points, refrained from punishing the partner), punish (a combination of both 'moderate' and 'extreme' punishment, when participant invested at least half of the own endowment to punish unfair behavior of the interaction partner). Proactive aggression scores are presented as standardized residualized measure (see Methods). $N = 161$.

Table 1

Association between prosocial punishment, proactive aggression and sex of participants

<i>Dependent variable: Propensity to punish in Prosocial Punishment Game</i>				
Predictor	B	p	R ²	p (model)
Sex [female]	0.855	0.015	0.093	0.011
Proactive aggression	-1.559	0.040		
Sex [female] * proactive aggression	0.892	0.042		

Note: The results of a binary logistic regression are presented. Dependent variable: decisions in the PPG (1 = punish, 0 = not to punish). Predictors: participant's sex (test category: female), proactive aggression (continuously scaled residualized scores), and interaction of these two. $N = 161$. B – regression coefficients, R² – Nagelkerke R-squared; p – statistical significance.

Our analysis revealed that only two of the Big Five personality traits – agreeableness and extraversion – significantly affected participants' behavior. These traits had significant impact only on cooperation (decisions in the PD), and were not involved in punishing deci-

sions, or behavior in the context of the UG. Both men and women scoring high on agreeableness and low on extraversion were more prone to cooperate in the Prisoner's Dilemma game, despite previous unfair behavior of the interaction partner. Accordingly, not agreeable and extravert individuals defected in the PD more frequently. This result is illustrated in Figure 5. Supporting statistics is presented in Table 2.

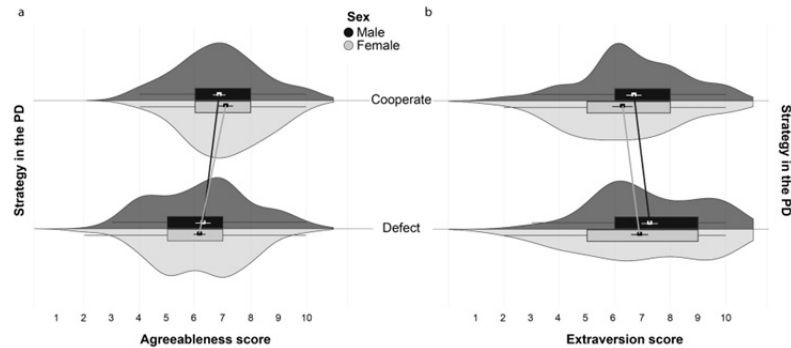


Fig. 5. An impact of agreeableness and extraversion on cooperation in Prisoner's Dilemma

(a) An impact of agreeableness trait on cooperation in Prisoner's Dilemma ($N = 170$); **(b)** an impact of extraversion trait on cooperation in Prisoner's Dilemma ($N = 170$). The visualization was created using *introdaviz* package for R (Nordmann *et al.* 2022).

Table 2

Association between cooperation, extraversion and agreeableness of participants

<i>Dependent variable:</i> Propensity to cooperate in Prisoner's Dilemma				
Predictor	B	p	R ²	p (model)
Sex [female]	−0.942	0.005	0.174	< 0.001
Extraversion	−0.238	0.007		
Agreeableness	0.400	< 0.001		

Note: The results of a binary logistic regression are presented. Dependent variable: decisions in the PD (1 = cooperate, 0 = defect). Predictors: participant's sex (test category: female), extraversion (continuously scaled scores), and agreeableness (continuously scaled scores). $N = 170$. B – regression coefficients, R^2 – Nagelkerke R-squared; p – statistical significance.

DISCUSSION

As it has been mentioned in the introduction, economic games are considered to be one of the effective modern methodologies for testing

models of social evolution in humans, particularly, the evolution of cooperation and sharing of limited resources. Since childhood, children across cultures prefer to act prosocial in distribution of limited food resources toward familiar and unfamiliar children (Butovskaya *et al.* 2022), indicating that prosociality is an immanent and universal feature of human nature. However, human behavior is very flexible and context dependent. Already in children we observe variation, manifested in obvious preference in favor of friends in case of altruistic sharing (Butovskaya *et al.* 2022). Our current experiment was aimed at exploring an impact of such factor as emotions on shifts in general behavioral tendencies. We tested punishing and prosocial behavior in same-sex face-to-face interactions under high negative emotional load. The results revealed that several factors significantly impacted participants' decisions related to punishment and cooperativeness. The distributions of these decisions deviated from those typically observed in emotionally neutral settings, as reported in previous studies.

Only 34.1 % of the total sample chose costly punishment in the PPG, which is below typical expectations (at least 50 %) (Peysakhovich *et al.* 2014; Rostovtseva *et al.* 2023b). Research also suggests that punishment increases with the difference in prosocial contributions between the punisher and the punished (Egas and Riedl 2008), with high contributors punishing at rates of 45–65 %. Furthermore, prior work indicates acute stress, particularly in men with high testosterone, stimulates prosocial punishment (80 % rate reported) (Wang *et al.* 2022). In contrast, our participants, despite stressful experimental setting, did not punish as much as expected. Men applied costly punishment less frequently than women (25 % compared to 43 %). Our results revealed significant sex differences in the resorting to costly punishment, with women punishing more often than men (43 % and 25 % respectively). We found no prior literature reporting sex differences in the propensity for prosocial punishment. While in two studies, one using anonymous design, and another showing partner's facial videos to participants, no significant sex differences were detected (Mieth *et al.* 2017; Rostovtseva *et al.* 2023b). To our knowledge, no other experiment has explored sex differences in prosocial punishment under conditions of face-to-face interactions. We are inclined to conclude that negative emotional load in our experiment caused the decrease in prosocial punishment rate, especially in male participants, producing significant sex differences.

Participants who used the extreme strategy of punishment, giving up all their funds to eliminate the second player's capital, may have

been seeking fairness by equalizing the score after the asymmetric outcome of the first Trust Game. Interestingly, women who used this strategy also demonstrated fairness in their decisions in the context of the Ultimatum Game. This may be interpreted in various ways. For instance, extreme punishment (equalizing the winnings of both players to 0) could bring emotional satisfaction to women who applied it, reduced emotional stress, and further accelerated the feeling of reconciliation, revealing a peaceful preference for ‘fair share’ in the last game. Alternatively, those women who normally prefer equality and symmetrical reciprocity in social interactions, might react more acutely to unfairness and more often use extreme punishment, either due to excessive impulsiveness or just striving for equality at any cost. This behavioral pattern was fully absent in male participants.

Significant sex differences emerged in PD decisions, with women defecting more frequently (66.3 %) and men cooperating more often (53.6 %). This led to an overall defection rate (56 %) exceeding cooperation. In our previous research conducted in an anonymous setting, a much higher cooperation rate (over 70 %) with no sex differences was observed (Rostovtseva *et al.* 2023b). However, two face-to-face studies in Russia (Buryatia and Tula) revealed patterns similar to those we observe in the current study: a defection rate of around 55 %, and women defecting more often than cooperating (Rostovtseva *et al.* 2020; Rostovtseva *et al.* 2024). While results from Russia, using face-to-face design, showed similar trend, the current study had the highest rate of female defection. Apparently, high negative emotional stress leads to decrease of cooperation rate, specifically within female-female interactions, producing significant sex differences in cooperativeness under such condition.

Although the ‘unfair interaction partner’ always cooperated in the PD, potentially offering a sense of reconciliation, fewer than expected proposed highly generous offers in the subsequent UG. Equal split offers were low, at 51.2 % for men and 59.3 % for women, compared to around 75 % in our previous study (Rostovtseva *et al.* 2023b). Interestingly, in both studies, most of the remaining participants proposed self-favoring splits. The low rate of fair share propositions and relatively high rate of self-favoring splits (42.7 % overall) in the current study most probably stem from a long-lasting negative emotional effect imposed by interaction partner from the very start of the experiment.

Our results highlight the nuanced effects of personality traits on decision-making under emotional stress. The analysis revealed that the

decisions to engage in costly prosocial punishment were associated with aggressiveness, whereas the tendency to cooperate was linked to personality traits related to sociability (agreeableness, extraversion). This distinction underscores the idea that fundamentally different motivations drive human punishing and cooperative behaviors (Peysakhovich *et al.* 2014; Rostovtseva *et al.* 2023b). In contrast to our predictions, the tendency to punish unfair partner was not associated with individual reactive aggressiveness, suggesting that prosocial punishment was not a result of impulsive emotional response. However, proactive aggression revealed a significant link. Proactive aggression is defined as less emotional, but rather more intentional, provoking, and goal-oriented type of aggression. Interestingly, male and female participants prone to such kind of aggression, demonstrated opposite behavioral strategies: higher levels of proactive aggression were associated with a lower likelihood of applying costly punishment among men, but with a higher likelihood of applying it among women. We also found that both men and women who were high in agreeableness and low in extraversion were more inclined to cooperate in PD, even after experiencing unfair behavior from their partner. Those who were low in agreeableness and high in extraversion were more likely to defect.

Female tendency to punish, although being slightly lower than expected, tended to normal level. The positive association between female propensity for punishment and proactive aggression is intuitively interpretable, likewise is the very low level of female cooperation with unfair partner in subsequent experimental interactions. The noteworthy results were obtained, particularly, for male behavior (extremely low punishment rate of 25 %, the negative association between proneness to punish and male proactive aggressiveness, unexpectedly high rate of cooperation with the unfair partner – 53.6 % – in a subsequent PD game). Therefore, it would be reasonable to address these results not as sex differences, but rather a notable deviation observed specifically in male behavior. We further speculate on some possible explanations, giving rise to hypotheses for testing in future studies.

From the perspective of classical game theory, refraining from costly punishment is a rational strategy, as punishment imposes certain costs, but zero immediate returns to a person who punishes (Fehr and Gächter 2002; Rostovtseva *et al.* 2023b). From the same theoretical perspective, cooperation in PD, moreover in cases of a known unreliable partner, is a completely irrational strategy, since the risk of exploitation is too high (Rapoport and Chammah 1965; Axelrod 1984;

Rostovtseva *et al.* 2023b). Since male participants were prone to refrain from punishment ('rational' decision), but at the same time were willing to cooperate in PD ('irrational' decision), an attempt to develop a strategy for maximizing payoffs is not a valid, or, at least, not a straightforward, explanation of their behavior in the experiment. Here, we consider two mechanisms as other potential explanations of our findings. First is related to demonstration of contempt or indifference, as a compensatory mechanism of rehabilitation after insulting betrayal, while second is related to conflict avoidance in negatively charged male-male interactions. Both of these behavioral responses are manifested through distancing from the person causing the anxiety, but the purpose of such distancing will be different in each of these scenarios.

In case of demonstration of contempt or indifference to the offender, a participant could seek to restore own social status ('face-saving') through rising above the situation. Such behavioral reaction to betrayal may facilitate negative emotional relief, especially in individuals who prioritize dominant social roles, and are largely involved in social environments with high levels of competition for social status. In repeated interactions, overt displays of contempt or indifference are risky in terms of escalating the conflict. Such behavior does not imply conciliatory actions or expressions, but rather explicitly demonstrate superiority to the interaction partner, devaluing his decisions. Both refraining from punishment and cooperation in subsequent PD in case of males fit well such explanation. Studies consistently show that men exhibit higher levels of intrasexual competition for social status than women, a difference deeply rooted in evolutionary pressures linked to reproductive success (Buss 1988; Buss and Dedden 1990; Buunk *et al.* 2014; Stenstrom *et al.* 2025). Presumably, 'face-saving' in intrasexual male social contexts is also more important than in female intrasexual relationships, and distinct sex-specific coping strategies might also exist (Xu *et al.* 2022).

The literature indicates that men are more likely to engage in direct competition for status, displaying direct aggression and engaging in aggressive face-to-face confrontations, than women (Griskevicius *et al.* 2009; Benenson and Abadzi 2020). This sex-specific feature suggests an alternative explanation for male behavior in our experiment: a strategy for unilateral de-escalation (Osgood 1962) aimed at reducing tension between the parties, and performed explicitly to avoid more acute conflict, which is more likely to arise between men than women. A greater tendency for conflict avoidance among men than women has deep evolutionary roots. Men are generally more massive and physically stronger

than women, they have more capacities for combat, therefore, conflicts among men are more harmful and dangerous than among women (de Looze *et al.* 2019). In case of equal resource-holding potential (*e.g.*, equal size, fighting abilities, and technology possession), spiteful conflict brings nearly equal damage to both parties. Thus, under condition of equality, avoidance of unnecessary conflicts is beneficial from adaptive perspective (necessity of a conflict may be dictated by the high cost of avoidance, for instance, a combat for food resource under condition of starvation). In case conflicting individuals differ in ‘weight’, the dominance relationships are usually established, also preventing parties from direct spiteful confrontations. However, correct assessment of resource-holding potential of a counterpart prior to the conflict is challenging, and under lack of information direct contest is dangerous for both parties (Crowley 2001; Morikawa *et al.* 2002). In our study, an ‘actor’ was first to demonstrate extremely antisocial dispositions. This could already be perceived as a signal of potential individual spitefulness (Morikawa *et al.* 2002). In face-to-face setting, both decisions to impose punishment on a partner and to defect in the PD, may be viewed as a competitive or hostile gesture. Following the above reasoning, it may be suggested that male participants on average were more oriented to conflict suppression by avoidance of direct confrontation with antisocial male partner. Conflict avoidance may generally underlie the elevated level of cooperativeness among men (in male-male interactions, or all-male groups) compared to women. Such tendency has already been demonstrated in our earlier studies using face-to-face experimental design (Rostovtseva *et al.* 2020, 2024).

Which of the suggested explanations of male behavior in our study is more probable (‘face-saving’ or conflict avoidance) can be tested empirically in future studies, for instance, by exploring and specifying exact emotions felt and displayed by participants during experimental interactions.

The results of our study and further considerations illustrate in the simplest way how complex the motivational foundations of human prosocial behavior can be, and how deeply its mechanisms can be rooted in human evolution. Our study provides a broad perspective on the multi-level system of costs and benefits that shapes the complex rationality of human behavior in various social situations. We show, how cooperation may be fostered locally by seemingly unrelated events or consequences (*i.e.*, male cooperation in PD), and how some irrelevant human motivation, that may be even unconscious, can produce an implicit payoff matrix of a higher order, with different weights and

structure of individual costs and benefits. We propose that due to multilayered nature of social contexts, such implicit payoff matrices are numerous, coexistent, and have fluctuating hierarchies, varying both at group and individual levels for subjects involved in a specific social situation. Each payoff matrix is determined by various factors, including (but not limited to) a structure of interaction itself, general environmental and evolutionary conditions, group and personal characteristics of interacting individuals. When being dominant, implicit structure of costs and benefits may produce behavioral strategies that seem ‘irrational’, inconsistent, or even inexplicable to an observer. Our study focused on a tiny aspect: emotions, personality and sex impacts on strategic behavior. But as we uncover more of the underlying hidden mechanisms, the implicit payoff matrices become overt and accountable, enhancing our ability to understand and predict human behavior in its complexity.

Acknowledgements

The study was supported by Russian Science Foundation, project No. 24-18-00457.

REFERENCES

- Al-Shawaf, L., and Shackelford, T. K. (eds.). 2024. *The Oxford Handbook of Evolution and the Emotions*. Oxford University Press.
- Andreozzi, L., Ploner, M., and Saral, A. S. 2020. The Stability of Conditional Cooperation: Beliefs Alone Cannot Explain the Decline of Cooperation in Social Dilemmas. *Scientific Reports* 10: 13610.
- Axelrod, R. 1984. *The Evolution of Cooperation*. New York: Basic Books.
- Balliet, D. 2010. Communication and Cooperation in Social Dilemmas: A Meta-Analytic Review. *Journal of Conflict Resolution* 54: 39–57.
- Balliet, D., and Van Lange, P. A. 2013a. Trust, Conflict, and Cooperation: A Meta-Analysis. *Psychological Bulletin* 139: 1090.
- Balliet, D., and Van Lange, P. A. 2013b. Trust, Punishment, and Cooperation Across 18 Societies: A Meta-Analysis. *Perspectives on Psychological Science* 8 (4): 363–379.
- Balliet, D., Mulder, L. B., and Van Lange, P. A. 2011a. Reward, Punishment, and Cooperation: A Meta-Analysis. *Psychological Bulletin* 137: 594.
- Balliet, D., Li, N. P., Macfarlan, S. J., and Van Vugt, M. 2011b. Sex Differences in Cooperation: A Meta-Analytic Review of Social Dilemmas. *Psychological Bulletin* 137: 881.

- Benenson, J. F., and Abadzi, H. 2020. Contest Versus Scramble Competition: Sex Differences in the Quest for Status. *Current Opinion in Psychology* 33: 62–68.
- Berg, J., Dickhaut, J., and McCabe, K. 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10: 122–142.
- Bossaerts, P. 2021. How Neurobiology Elucidates the Role of Emotions in Financial Decision-Making. *Frontiers in Psychology* 12: 697375.
- Breuer, W., Helduser, C., and Schade, P. 2016. Breaking the Rules: Anticipation of Norm Violation in a Binary-Choice Trust Game. *Economics Letters* 146: 123–125.
- Buss, D. M. 1988. The Evolution of Human Intrasexual Competition: Tactics of Mate Attraction. *Journal of Personality and Social Psychology* 54: 616–628.
- Buss, D. M., and Dedden, L. A. 1990. Derogation of Competitors. *Journal of Social and Personal Relationships* 7: 395–422.
- Butovskaya, M. L., and Rostovtseva, V. V. 2021. *The Evolution of Human Altruism and Cooperation: A Biosocial Perspective*. Moscow: Lenand. Original in Russian (Бутовская М. Л., Ростовцева В. В. Эволюция альтруизма и кооперации человека: Биосоциальная перспектива. Москва: Ленанд).
- Butovskaya, M., Rostovtseva, V., Dronova, D., Burkova, V., and Adam, Y. 2022. Variations in Limited Resources Allocation towards Friends and Strangers in Children and Adolescents from Seven Economically and Culturally Diverse Societies. *Scientific Reports* 12: 15232.
- Buunk, A. P., Stulp, G., and Ormel, J. 2014. Parental Social Status and Intrasexual Competitiveness among Adolescents. *Evolutionary Psychology* 12: 1022–1037.
- Chen, X., Xie, J., Wang, D., Liu, R., and Chen, C. 2024. Reliability and Validity of the Reactive-Proactive Aggression Questionnaire in Chinese Boxers. *International Journal of Sport and Exercise Psychology* 22: 160–176.
- Cima, M., Raine, A., Meesters, C., and Popma, A. 2013. Validation of the Dutch Reactive Proactive Questionnaire (RPQ): Differential Correlates of Reactive and Proactive Aggression from Childhood to Adulthood. *Aggressive Behavior* 39: 99–113.
- Crowley, P. H. 2001. Dangerous Games and the Emergence of Social Structure: Evolving Memory-Based Strategies for the Generalized Hawk-Dove Game. *Behavioral Ecology* 12: 753–760.
- de Almeida, F. P. L. 2020. *Constitution: the Darwinian Evolution of a Societal Structure*. Baden-Baden: Nomos Verlag.
- de Looze, M., Elgar, F. J., Currie, C., Kolip, P., and Stevens, G. W. 2019. Gender Inequality and Sex Differences in Physical Fighting, Physical Ac-

- tivity, and Injury among Adolescents across 36 Countries. *Journal of Adolescent Health* 64: 657–663.
- Dennison, J. B., Sazhin, D., and Smith, D. V. 2022. Decision Neuroscience and Neuroeconomics: Recent Progress and Ongoing Challenges. *Wiley Interdisciplinary Reviews: Cognitive Science* 13: e1589.
- Egas, M., and Riedl, A. 2008. The Economics of Altruistic Punishment and the Maintenance of Cooperation. *Proceedings of the Royal Society B: Biological Sciences* 275: 871–878.
- Ensminger, J., and Henrich, J., (eds.). 2014. *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective*. New York: Russell Sage Foundation.
- Fehr, E., and Gächter, S. 2002. Altruistic Punishment in Humans. *Nature* 415: 137–140.
- Frey, U. J., and Rusch, H. 2012. An Evolutionary Perspective on the Long-Term Efficiency of Costly Punishment. *Biology and Philosophy* 27: 811–831.
- Frank, R. H. 1988. *Passions within Reason: The Strategic Role of the Emotions*. WW Norton and Co.
- Fung, A. L. C., Raine, A., and Gao, Y. 2009. Cross-Cultural Generalizability of the Reactive–Proactive Aggression Questionnaire (RPQ). *Journal of Personality Assessment* 91: 473–479.
- Gintis, H. 2014. *The Bounds of Reason: Game Theory and the Unification of the Behavioural Sciences-Revised Edition*. Princeton, NJ: Princeton University Press.
- Güth, W., Schmittberger, R., and Schwarze, B. 1982. An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization* 3: 367–388.
- Griskevicius, V., Tybur, J. M., Gangestad, S. W., Perea, E. F., Shapiro, J. R., and Kenrick, D. T. 2009. Aggress to Impress: Hostility as an Evolved Context-Dependent Strategy. *Journal of Personality and Social Psychology* 96: 980–994.
- Henrich, J., and Muthukrishna, M. 2021. The Origins and Psychology of Human Cooperation. *Annual Review of Psychology* 72: 207–240.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. (eds.). 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford: Oxford University Press.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... and Ziker, J. 2006. Costly Punishment across Human Societies. *Science* 312: 1767–1770.

- Jaeger, B., Oud, B., Williams, T., Krumhuber, E. G., Fehr, E., and Engelmann, J. B. 2022. Can People Detect the Trustworthiness of Strangers Based on Their Facial Appearance? *Evolution and Human Behavior* 43: 296–303.
- Kreps, D. M. 1990. Corporate Culture and Economic Theory. In Alt, J., and Shepsle, K. (eds.), *Perspectives on Positive Political Economy* (pp. 90–143). Cambridge: Cambridge University Press.
- Laland, K., Matthews, B., and Feldman, M. W. 2016. An Introduction to Niche Construction Theory. *Evolutionary Ecology* 30: 191–202.
- Laland, K. N., and O'Brien, M. J. 2011. Cultural Niche Construction: An Introduction. *Biological Theory* 6: 191–202.
- Lazebny, O. E., Kulikov, A. M., Butovskaya, P. R., Proshakov, P. A., Fokin, A. V., and Butovskaya, M. L. 2020. Analysis of Aggressive Behavior in Young Russian Males Using 250 SNP Markers. *Russian Journal of Genetics* 56: 1118–1128.
- Lempert, K. M., and Phelps, E. A. 2014. Neuroeconomics of Emotion and Decision Making. *Neuroeconomics* 219–236.
- Leonard, W. R. 2018. Centennial Perspective on Human Adaptability. *American Journal of Physical Anthropology* 165: 813–833.
- Levitt, S. D., and List, J. A. 2008. Homo economicus Evolves. *Science* 319: 909–910.
- Luce, R. D. and Raiffa, H. 1957. *Games and Decisions: An Introduction and Critical Survey*. New York: Wiley and Sons.
- Melis, A. P., and Semmann, D. 2010. How is Human Cooperation Different? *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 2663–2674.
- Merk, W., Orobio de Castro, B., Koops, W., and Matthys, W. 2005. The Distinction between Reactive and Proactive Aggression: Utility for Theory, Diagnosis and Treatment? *European Journal of Developmental Psychology* 2: 197–220.
- Mieth, L., Buchner, A., and Bell, R. 2017. Effects of Gender on Costly Punishment. *Journal of Behavioral Decision Making* 30: 899–912.
- Morikawa, T., Hanley, J. E., and Orbell, J. 2002. Cognitive Requirements for Hawk-Dove Games: A Functional Analysis for Evolutionary Design. *Politics and the Life Sciences* 21: 3–12.
- Nash, J. F. 1950. Equilibrium Points in N-Person Games. *Proceedings of the National Academy of Sciences* 36: 48–49.
- Nash, J. 1951. Non-Cooperative Games. *Annals of Mathematics* 54: 286–295.

- Nordmann, E., McAleer, P., Toivo, W., Paterson, H., and DeBruine, L. 2022. Data Visualisation Using R, for Researchers Who Don't Use R. *Advances in Methods and Practices in Psychological Science* 5: 25152459221074654.
- Osgood, C. E. 1962. *An Alternative to War or Surrender*. University of Illinois Press.
- Pechorro, P., Ray, J. V., Raine, A., Maroco, J., and Goncalves, R. A. 2017. The Reactive–Proactive Aggression Questionnaire: Validation among a Portuguese Sample of Incarcerated Juvenile Delinquents. *Journal of Interpersonal Violence* 32: 1995–2017.
- Peysakhovich, A., Nowak, M. A., and Rand, D. G. 2014. Humans Display a ‘Cooperative Phenotype’ that is Domain General and Temporally Stable. *Nature Communications* 5: 4939.
- Raihani, N. J., Thornton, A., and Bshary, R. 2012. Punishment and Cooperation in Nature. *Trends in Ecology and Evolution* 27: 288–295.
- Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., ... and Liu, J. 2006. The Reactive–Proactive Aggression Questionnaire: Differential Correlates of Reactive and Proactive Aggression in Adolescent Boys. *Aggressive Behavior* 32: 159–171.
- Rainey, P. B. 2023. Major Evolutionary Transitions in Individuality between Humans and AI. *Philosophical Transactions of the Royal Society B* 378: 20210408.
- Rapoport, A., and Chammah, A. M. 1965. *Prisoner's Dilemma. A Study in Conflict and Cooperation*. Ann Arbor: University of Michigan Press.
- Rammstedt, B., and John, O. P. 2007. Measuring Personality in one Minute or Less: A 10-item Short Version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41: 203–212.
- Romero-Martinez, A., Sarrate-Costa, C., and Moya-Albiol, L. 2022. Reactive vs Proactive Aggression: A Differential Psychobiological Profile? Conclusions Derived from a Systematic Review. *Neuroscience and Biobehavioral Reviews* 136: 104626.
- Rostovtseva, V. V. 2026. *Understanding Human Cooperativeness*. Groningen: University of Groningen.
- Rostovtseva, V. V. 2024. The Phenomenon of Human Trust: An Interdisciplinary Review. *Chelovek* 35: 27–43. *Original in Russian* (Ростовцева В. В. Феномен человеческого доверия: междисциплинарный обзор. *Человек* 35: 27–43).
- Rostovtseva, V. V., Weissing, F. J., Mezentseva, A. A. and Butovskaya, M. L. 2020. Sex Differences in Cooperativeness – An Experiment with Buryats in Southern Siberia. *PLoS One* 15: e0239129.

- Rostovtseva, V. V., Butovskaya, M. L., Mezentseva, A. A., and Weissing, F. J. 2023a. Effects of Sex and Sex-Related Facial Traits on Trust and Trustworthiness: An Experimental Study. *Frontiers in Psychology* 13: 925601.
- Rostovtseva, V. V., Puurtinen, M., Méndez Salinas, E., Cox, R. F., Groothuis, A. G., Butovskaya, M. L., and Weissing, F. J. 2023b. Unravelling the Many Facets of Human Cooperation in an Experimental Study. *Scientific Reports* 13: 19573.
- Rostovtseva, V. V., Butovskaya, M. L., Mezentseva, A. A., Dashieva, N. B., and Korotkova, A. A. 2024. Male Cooperation is Culture-Specific: Evidence from Cross-Population Experiments. *Social Evolution and History* 23: 124–143.
- Serra, D. 2021. Decision-Making: From Neuroscience to Neuroeconomics – An Overview. *Theory and Decision* 91: 1–80.
- Shchebetenko, S. 2014. ‘The Best Man in the World’: Attitudes toward Personality Traits. *Psychology. Journal of Higher School of Economics* 11: 129–148.
- Spadaro, G., Graf, C., Jin, S., Arai, S., Inoue, Y., Lieberman, E., ... and Balliet, D. 2022. Cross-Cultural Variation in Cooperation: A Meta-Analysis. *Journal of Personality and Social Psychology* 123: 1024.
- Spadaro, G., Jin, S., and Balliet, D. 2023. Gender Differences in Cooperation across 20 Societies: A Meta-Analysis. *Philosophical Transactions of the Royal Society B* 378: 20210438.
- Stenstrom, E. P., Nepomuceno, M. V., and Hingston, S. T. 2025. Testosterone and the Pursuit of Possessions: The Interaction between Baseline Testosterone and Intrasexual Competition on Materialism in Males. *Acta Psychologica* 255: 104937.
- Tharp, A. L. T., Sharp, C., Stanford, M. S., Lake, S. L., Raine, A., and Kent, T. A. 2011. Correspondence of Aggressive Behavior Classifications among Young Adults Using the Impulsive Premeditated Aggression Scale and the Reactive Proactive Questionnaire. *Personality and Individual Differences* 50: 279–285.
- Thielmann, I., Spadaro, G., and Balliet, D. 2020. Personality and Prosocial Behavior: A Theoretical Framework and Meta-Analysis. *Psychological Bulletin* 146: 30.
- Turchin, P. 2013. The Puzzle of Human Ultrasociality: How Did Large-Scale Complex Societies Evolve? In Richerson, P. J., and Christiansen, M. H. (eds.), *Cultural Evolution: Society, Technology, Language, and Religion* (pp. 61–73). Frankfurt am Main, Germany: Massachusetts Institute of Technology and the Frankfurt Institute for Advanced Studies.

- Urbina, D. A., and Ruiz-Villaverde, A. 2019. A Critical Review of Homo economicus from Five Approaches. *American Journal of Economics and Sociology* 78: 63–93.
- Van Kleef, G. A., and Lelieveld, G. J. 2022. Moving the Self and Others to Do Good: The Emotional Underpinnings of Prosocial Behavior. *Current Opinion in Psychology* 44: 80–88.
- von Neumann, J., and Morgenstern, O. 1944. *Theory of Games and Economic Behaviour*. Princeton, NJ: Princeton University Press.
- Wang, H., Zhen, Z., Zhu, R., Yu, B., Qin, S., and Liu, C. 2022. Help or Punishment: Acute Stress Moderates Basal Testosterone's Association with Prosocial Behavior. *Stress* 25: 179–188.
- Wu, J., Luan, S., and Raihani, N. 2022. Reward, Punishment, and Prosocial Behavior: Recent Developments and Implications. *Current Opinion in Psychology* 44: 117–123.
- Xu, Y., Yuan, Y., Xie, X., Tan, H., and Guan, L. 2022. The Impact of Facial Attractiveness on the Processing of Self-Relevant and Friend-Relevant Information after Romantic Rejection: A Study of Sex Differences. *Current Psychology* 41: 4145–4157.