



Research



Cite this article: Pulcu E. 2024 Individualistic attitudes in Iterated Prisoner's Dilemma undermine evolutionary fitness and may drive cooperative human players to extinction. *R. Soc. Open Sci.* **11**: 230867.

<https://doi.org/10.1098/rsos.230867>

Received: 21 June 2023

Accepted: 9 February 2024

Subject Category:

Psychology and cognitive neuroscience

Subject Areas:

psychology, computational biology, evolution

Keywords:

Prisoner's Dilemma, Ultimatum Game, evolutionary game theory, zero-determinant strategy, interpersonal cooperation, extortion

Author for correspondence:

Erdem Pulcu

e-mails: pulcerd@gmail.com; pulcerd@hotmail.com

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7124576>.

Individualistic attitudes in Iterated Prisoner's Dilemma undermine evolutionary fitness and may drive cooperative human players to extinction

Erdem Pulcu

Department of Psychiatry, Psychopharmacology and Emotion Research Lab, Computational Psychiatry Lab, University of Oxford, Oxford, UK

EP, 0000-0002-2170-0677

Inarguably, humans perform the richest plethora of prosocial behaviours in the animal kingdom, and these are important for understanding how humans navigate their social environment. The success and failure of strategies human players devise also have implications for determining long-term socio-economic/evolutionary fitness. Following the footsteps of Press and Dyson (2012), I implemented their evolutionary game-theoretic modelling from Iterated Prisoner's Dilemma (a behavioural economic probe of interpersonal cooperation) and re-analysed already published data on human proposer behaviour in the Ultimatum Game (a behavioural economic probe of altruistic punishment) involving 50 human participants versus stochastic computerized opponents with prosocial and individualistic social value orientations. Although the results indicate that it is more likely to break cycles of mutual defection in ecosystems in which humans interact with individualistic opponents, analysis of social-economic fitness at the Markov stationary states suggested that this comes at an evolutionary cost. Overall, human players acted in a significantly more cooperative manner than their opponents, but they failed to overcome extortion from individualistic agents, risking 'extinction' in 70% of the cases. These findings demonstrate human players might be short-sighted, and social interactive decision strategies they devise while adjusting to different types of opponents may not be optimal in the long run.

1. Introduction

Humans engage in a variety of prosocial behaviours [1–4]. Sometimes people go out of their way to help another person (e.g. interpersonal cooperation [5–7]) or even sacrifice some part of their resources to stand up against unfairness [8,9], which is commonly known as altruistic punishment. Behavioural-economic games such as the Prisoner's Dilemma (PD) [10,11] and the Ultimatum Game (UG) [12,13] are well-established methods of probing prosocial behaviours (interpersonal cooperation and altruistic punishment, respectively) in experimental/laboratory settings. In 2012, Press and Dyson put forward an important mathematical model of strategy setting in the Iterated PD (IPD, a social-economic game in which two parties interact over a predefined matrix of payoffs repeatedly, allowing players to learn about opponent strategies in time), by which a player can establish a one-sided claim to an unfair share of rewards [14]. Press and Dyson suggested that players using these zero determinant strategies (ZD) can enforce an advantageous linear relationship over the payoffs of their opponents, meaning that if the opponents are rewarded, they would be rewarded even more. Note that more recently, in finitely repeated games such as in applications to interactions between two human players or a human player against a computerized agent, it is shown that unconditional strategies that cooperate with fixed independent probabilities can also enforce a linear relationship between two players' payoffs [15]. Other studies suggested that the social dilemma associated with the linear payoff structure enforced by the ZD (and unconditional) strategies that hamper an agent cooperating can be quantified by the concept of universal dilemma strength [16,17]. Advantageous linear relationship in payoffs enforced by ZD strategies is shown to pose a challenge even for evolutionary opponents (i.e. those that can learn and dynamically adjust their behaviours), who are faced with either giving in to extortion or increasing their Theory of Mind (ToM) sophistication [18] in order to break free of a relationship that ultimately distributes rewards unfairly. In support of the latter, Press and Dyson [14] postulated [in their abstract] that:

Only a player with a theory of mind about his opponent can do better, in which case Iterated Prisoner's Dilemma is an Ultimatum Game.

In my opinion, this is a very important aspect of Press and Dyson's work which had been, to the best of my knowledge, overlooked by the scientific community. It raises the possibility of elegantly bridging two key experimental probes of human social-interactive decision making. This approach can reveal important insights about the evolutionary fitness trajectories of strategies humans use in these (cooperation versus competition) scenarios.

Recently, we showed computational model-based evidence to suggest that human players employ second-order or higher [19] degrees of ToM sophistication in the UG (e.g. taking account of their opponent's acceptance probability) while choosing between binary ultimatums [20] (figure 1*a*). This was captured by the best-fitting computational model to our data, as we showed that human players compute the expected values of ultimatums by a multiplicative integration of their potential rewards which were modulated by a nonlinear power utility function accounting for participants' risk attitude [21], and the opponent's inferred acceptance probabilities which were modulated by an exponential-logarithmic weighting function [20](figure 1*b*). More importantly, this overarching value-based social decision-making model applied irrespective of the opponent's social value orientation (SVO, which determines a player's degree of prosociality, e.g. whether the opponents displayed individualistic or prosocial tendencies) [20,22,23]. A concrete prediction arising from our earlier findings related to the involvement of higher-order ToM processes in social-interactive decision making is that human players should be able to escape extortion. Earlier work suggested that human players can be subdued by computerized opponents dictating extortionate terms, but this approach also impairs the extortionate opponent's own reward trajectory, therefore does not go unpunished [24]. Recent reviews of the literature suggested that extortionate rivals generally rely on cooperative strategies [25], and will eventually give rise to more cooperative partners to spread in the population [26], therefore they are prosocial overall. Although ZD strategies are often associated with extortion in social interactive decision making, these recent formulations suggest that they can also be understood within the framework of the dual strategies theory [27,28] which has been influential in evolutionary psychology, for example, extortion by which agents attempt to exert dominance, and generosity [15,25] (usually deployed to counteract altruistic punishment) which is one method of prestige/reputation building. However, as Press and Dyson demonstrated, the ultimate test of any extortion-related prediction demands analysis of interpersonal cooperation/competition not within the confines of an experimental

timeline (irrespective of whether the experiment is a quantitative simulation or with human players), but at the equilibrium point of these systems. So far, how well human players can escape extortion in iterative games has not been investigated by implementing Press and Dyson's algebraic approach. It is possible that general adversarial models can account for imagining alternative courses of action for escaping extortion [29]. Furthermore, no previous work attempted to unify IPD and UG in a way that would allow a game-theoretic analysis of evolutionary social-economic fitness associated with human interpersonal cooperation at the stationary state (i.e. the equilibrium point of a competitive system at the finite evolutionary horizon). In the subsequent sections, based on a reanalysis of a previously published dataset, I demonstrate quantitative evidence in support of Press and Dyson's original claim, followed by some valuable insights into human interpersonal cooperation and human-machine/artificial intelligence interactions [30].

2. Material and methods

2.1. Participants

Fifty non-clinical volunteers were recruited for the ultimatum-giving experiment. Initially, participants were asked to learn the preferences of two computerized opponents in an observational learning session. Opponent strategies were based on social-value functions (q_A) that generated online accept or reject decisions from self-versus-other reward distributions probabilistically, globally agreeing with a prosocial and an individualistic SVO [22]:

$$q_A = 1 / \left(1 + e^{(-\beta(\alpha R_S - \delta R_O + \rho |R_S - R_O|))} \right). \quad (2.1)$$

In the formula above, R_S refers to the share of the reward for the computerized agent, whereas R_O refers to the reward amount to the participant (figure 1c). The hyper-parameters defining the valuation of the agents [α , δ , ρ and β] were set to [1.096, 0.382, -2.512 and 0.037] for the individualistic agent; and [1.368, -0.644, -3.798 and 0.045] for the prosocial agent.

In the subsequent step after adequate learning, participants were asked to choose between two monetary ultimatums to be given to these opponents that makes 'accept' or 'reject' decisions following the stochastic social-value function shown in (also see figure 1a, experiment involving 120 binary decisions against each opponent type). The results from the original experiments were published previously [20] and the current analytic approach concerns the reanalysis of that dataset readily available at: https://osf.io/qy9e6/?view_onlycbcd3eac3cd0744f69ba0828033789e61.

3. Results

3.1. Reformulating an iterative UG as an iterative PD

If we take Press and Dyson's original claim at face value, it should be possible to make quantitative transitions between IPD and UG, or vice versa. Indeed, it is possible to construct PD matrices by indexing trials by (i) the participants' inferred choice probability for each of the [computerized] opponents (i.e. $s \in \{i, p\}$, i : individualistic opponent, p : prosocial opponent, also see §2) generated under the best-fitting computational decision model that accounts for participant choice behaviour [20], where choosing an option with higher inferred acceptance probability represents the participant's intention to cooperate: 'C'; and (ii) the opponent's responses to these offers (e.g. stochastically rejecting the offer: 'CD' or accepting the offer: 'CC'). After repeating this step for all possible combinations of $xy \in \{CC, CD, DC, DD\}$, the average reward amounts R_s (i.e. reward to self) and R_o (i.e. reward to opponent) can be calculated across all trials satisfying these conditions, and the cells of the 2×2 matrix for both players can be constructed (figure 1c). Figure 2a shows an example 2×2 matrix which satisfies the payoff conditions of a PD game.

These transformations demonstrate that it is possible to recover IPD matrices, which are more suitable for evolutionary game-theoretic fitness analysis, directly from an iterative UG experiment. Here, it is important to point out that, although these payoff matrices are introduced with a general reference to IPD, it would be more accurate to define them as ' 2×2 evolutionary game' matrices which were uniquely constructed for each interaction between human participants and computerized opponents (i.e. 50 non-clinical/healthy volunteers \times 2 different computerized opponents (prosocial

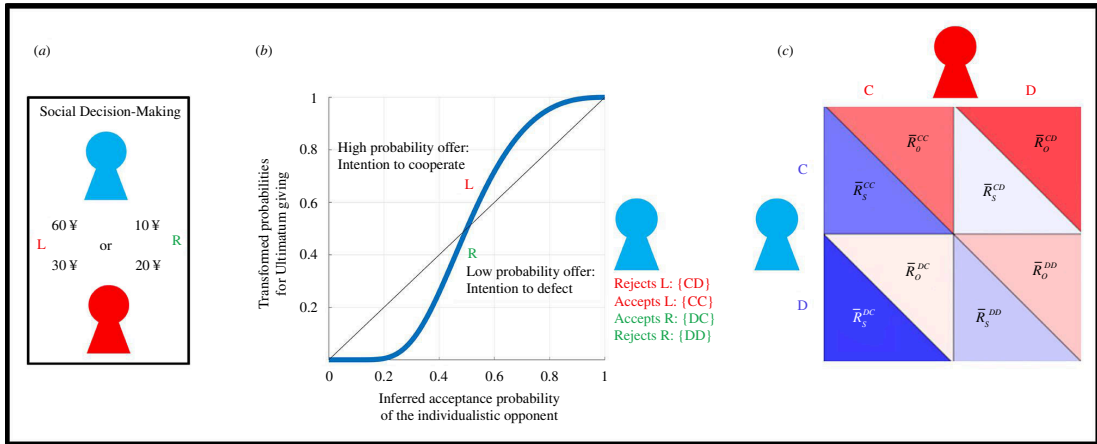


Figure 1. Schematic outline of recovering IPD matrices from an iterative UG. (a) Participants (the red icon) chose between two ultimatums to be given to a computerized opponent (blue icon representing the individualistic opponent) which accepts or rejects them following a probabilistic social-value function (see §2, equation 2.1). At this stage, only the participant can see both offers (L: left offer; R: right offer). Once selected, the offer is presented to the responder/opponent. The opponent can only evaluate the chosen option but cannot know whether it is better or worse than the unchosen one. The UG included 120 trials in total, that participants played twice (one time against each opponent: individualistic or prosocial). Panel adapted from Pulcu and Haruno [20]. (b) Previously, we showed model-based evidence to suggest that participants track their opponents' acceptance probabilities (on x-axis, also q_A in equation 2.1), which are transformed by a nonlinear probability weighting function in the UG. Thick blue line represents the weighted probabilities used while choosing between binary ultimatums, transforming participant's inference about the opponent's acceptance probability. As a result, choosing the offer associated with higher acceptance probability indicates the participant's intention to cooperate with the opponent. Given that opponents make decisions probabilistically following a [social] stochastic value function (q_A), it is possible to index all outcome types of PD. (c) Schematic diagram outlining the recovered PD matrix from the UG experiment. For example, \bar{R}_O^{CC} refers to the participant's reward amounts averaged across all conditions satisfying CC outcomes (calculated from the lower rows of reward magnitudes for the chosen option in panel 1A), whereas \bar{R}_S^{CC} refers to the opponent's average reward amounts for the same condition. Panel adapted from Press and Dyson [14].

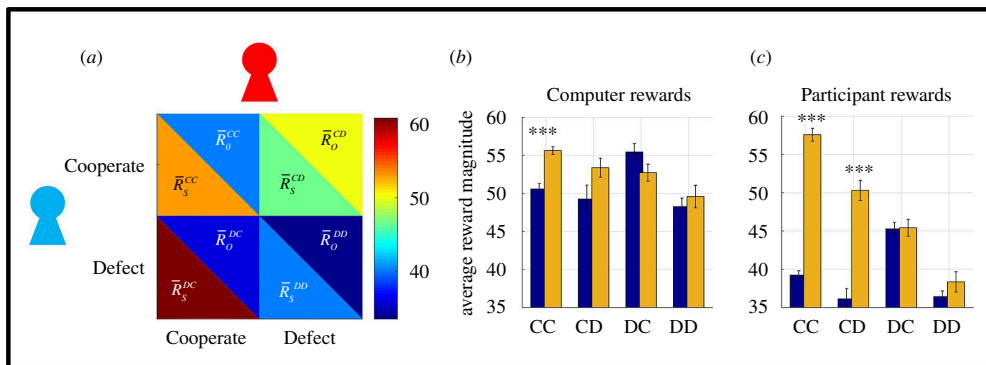


Figure 2. Analyses of payoffs. (a) An example payoff matrix recovered from the UG experiment satisfying the conditions of the PD game (with an uneven payoff structure). As in figure 1c, the red icon represents the participant, and the blue icon represents the opponent following an individualistic value function. Colour bar represents the average reward magnitude in each cell. (b) Computerized opponent with prosocial value function (orange bars) are able to acquire significantly higher rewards establishing mutual cooperation with the human participant relative to the individualistic strategy (blue bars, $***p < 0.001$, Bonferroni corrected for multiple comparisons). In line with labels in figures 1c and 2a, IPD conditions are labelled first with reference to the opponent response (i.e. CD meaning opponent cooperated when human participant defected). (c) Human participants acquired significantly higher rewards under mutual cooperation with a prosocial opponent relative to the individualistic opponent both when they cooperated and defected ($***p < 0.001$, Bonferroni corrected for multiple comparisons). Error bars designate ± 1 s.e.m.

versus individualistic), in total 100 unique interactions). Press and Dyson [14] explicitly stated that their formulations apply to all 2×2 evolutionary matrix games, not only those which satisfy the conditions of the Prisoner's Dilemma Game (e.g. $CC > DD$).

3.2. Human players cannot escape extortion against individualistic opponents

The next step involves an analysis of the payoff amounts shown in [figure 2b,c](#). Here, it is important to clarify that cells corresponding to opponent defection (i.e. the cells in the lower row in [figure 2a](#), based on rejected offers in the UG) actually designate the average amount of forgone rewards for both sides [24], as both sides get nothing when an offer is rejected in the UG. This is exactly what Press and Dyson referred to in their abstract. In this approach, the average rewards acquired under opponent cooperation (i.e. CC and CD conditions in [figure 2b,c](#)) may be more informative and conservative. The results indicate that opponents with a prosocial value orientation were able to acquire significantly more rewards than the individualistic strategy (based on 2×2 multivariate ANOVA, $F(1,98) = 15.046$, $p < 0.001$, [figure 2b](#)). Similarly, the human players were also able to acquire significantly higher rewards playing against a prosocial relative to an individualistic opponent ($F(1,98) = 297.33$, $p < 0.001$, [figure 2c](#)). On the other hand, the individualistic opponent acquired significantly more rewards relative to human participants under participants' unilateral defection (i.e. CD outcomes) or mutual cooperation (comparison of blue bars between [figure 2b,c](#), $F(1,98) = 98.123$, $p < 0.001$). However, when it came to social interactive decision making with the prosocial opponent, there was a significant outcome by player-type interaction ($F(1,98) = 4.585$, $p = 0.035$), whereby human participants were able to acquire higher rewards relative to the prosocial opponent, but only under mutual cooperation ($t(98) = 01.979$, $p = 0.051$). Taken together, these findings indicate that interactions with prosocial opponents lead to mutually rewarding outcomes for human players, whereas the individualistic opponents were able to establish extortionate terms (akin to what ZD strategies can achieve) over the human participants, acquiring significantly higher rewards irrespective of whether the human participants cooperated or defected (i.e. comparison of blue bars for CC and CD conditions between [figure 2b,c](#)). It is important to highlight that numerical components of the stimuli (i.e. offer distributions) in the original UG experiment [20] were identical across all conditions, so the differences reported here cannot be attributed to such factors which might otherwise confound the results (i.e. similar to IPD having a predefined payoff matrix).

3.3. ToM-guided value-based decision making promotes interpersonal cooperation

In the preceding section, we demonstrated that iterative UG can be reformulated in terms of iterative PD, revealing that human players cannot escape extortion against individualistic opponents. In the next step, it is important to demonstrate how transition probabilities can be computed from the raw data. These describe the rates of cooperation following all possible outcomes xy in the IPD. Analysing 2×2 social interactive decision-making games by this approach is computationally efficient because by constructing Markov transition matrices it is possible to treat these games as a competition between two memory-1 strategies (i.e. describing agent cooperation in terms of the probability with which it changes its behaviour from the previous trial after observing the joint outcome, commonly known as *transition probabilities*). To the best of my knowledge, this approach has not been implemented on any data collected from human players.

This approach suggests that the prosocial strategy was significantly more likely to cooperate with the human participants relative to the individualistic strategy ($F(1,98) = 448.34$, $p < 0.001$, [figure 3a](#)). This is only reassuring, as the prosocial strategy was designed to follow a social-value function that treats self-other inequality more negatively (see the details in §2). In IPD terms, this would translate to a lower probability of defection as unilateral defections lead to self-other inequality in payoffs. Although participants were not explicitly informed about the SVO of their opponents (i.e. whether the opponent was prosocial or individualistic, they could only infer their opponents' value functions from 180-trial observational social-learning session they completed prior to the UG in which they were trying to predict whether opponents would accept or reject offers coming from a third party), they acted in a significantly more cooperative manner while playing against the prosocial opponent relative to the individualistic opponent ($F(1,98) = 17.295$, $p < 0.001$, [figure 3b](#)). Furthermore, although participants were asked to bank as much money as possible in the experiment, which was paid to them in real monetary terms, they acted in a significantly more cooperative manner relative to the computerized opponents (i.e. comparison of blue and orange bars between [figure 3a, b](#), $F(1,98) = 229.53$, $p < 0.001$ versus the individualistic opponent; $F(1,98) = 30.152$, $p < 0.001$ versus the prosocial opponent). This suggests that ToM-driven value-based decisions during social-interactive decision making (i.e. incorporating opponents' value functions into one's own expected value computations) promote

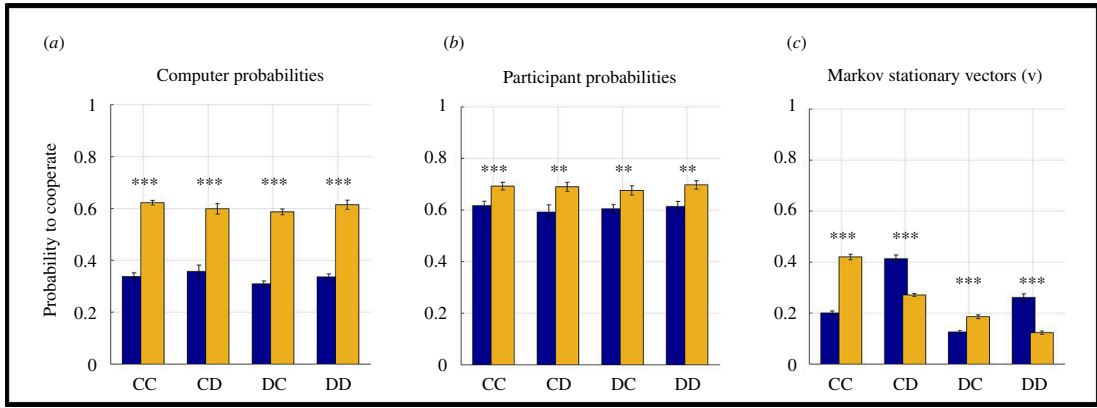


Figure 3. Analyses of transition probabilities. (a) The prosocial opponent was more likely to cooperate relative to the individualistic opponent after all types of outcomes in the IPD (** $p < 0.001$). (b) Similarly, participants also cooperated at a significantly higher rate while playing against the prosocial relative to the individualistic opponent (all $t(98) > 2.91$, all $p < 0.005$, *** $p < 0.001$, ** $p < 0.01$). (c) Summary of Markov stationary vectors computed by equations (3.2 and 3.3). All comparisons were statistically significant (*** $p < 0.001$, Bonferroni corrected for multiple comparisons). Error bars designate ± 1 s.e.m. Orange bars: prosocial opponent; blue bars: individualistic opponent.

interpersonal cooperation. Therefore, interpersonal cooperation emerges naturally as the option with the better-expected value for human players, irrespective of opponent type.

3.4. Interpersonal cooperation erodes in time

An important follow-up question was asking whether interpersonal cooperation could be sustained. In order to address this question, it was necessary to test the stability of these social interactive systems at the equilibrium point. This involves computing the trajectory of the interpersonal competition from the probabilities of cooperation following any outcome type xy in the IPD [31]. Following an identical approach as reported by Press and Dyson and using the vectors of probabilities defining cooperation rate as calculated from the raw data, as shown in figure 3a,b (i.e. two for each participant against each of the opponents (s_p, s_i); and one for each computerized opponent against the participant (p_s, i_s)), Markov transition matrices (\mathbf{M}) can be constructed for each pair accordingly:

$$M(s_i, i_s) = \begin{bmatrix} s_i^1 i_s^1 & s_i^1 (1 - i_s^1) & (1 - s_i^1) i_s^1 & (1 - s_i^1) (1 - i_s^1) \\ s_i^2 i_s^3 & s_i^2 (1 - i_s^3) & (1 - s_i^2) i_s^3 & (1 - s_i^2) (1 - i_s^3) \\ s_i^3 i_s^2 & s_i^3 (1 - i_s^2) & (1 - s_i^3) i_s^2 & (1 - s_i^3) (1 - i_s^2) \\ s_i^4 i_s^4 & s_i^4 (1 - i_s^4) & (1 - s_i^4) i_s^4 & (1 - s_i^4) (1 - i_s^4) \end{bmatrix}. \quad (3.1)$$

Note that the matrix formulated above illustrates an example exchange between a human player and an individualistic opponent. After that the stationary vectors, \mathbf{V} , of the Markov transition matrices (which satisfy)

$$\mathbf{v}^T \mathbf{M} = \mathbf{v}^T, \text{ or } \mathbf{v}^T \mathbf{M}' = 0 \quad (3.2)$$

were computed. Here \mathbf{M}' is the transpose of the matrix \mathbf{M} which holds the transition probabilities of the human-machine interaction, and satisfies $\mathbf{M}' \equiv \mathbf{M} - \mathbf{I}$, where \mathbf{I} is a 4×4 identity matrix. Markov stationary vectors identify the stable transition probabilities which will be achieved for each pair (e.g. a human participant versus the individualistic opponent), if these interactions continue for a very long time (i.e. the finite evolutionary horizon). These analyses were conducted for each pair (i.e. all combinations of 50 human players versus 2 opponents, 100 pairs in total). Here, it is important to highlight the necessity of this approach as the system stability (as shown in figure 3c) cannot be predicted from individual players' transition probabilities (figure 3a,b). This approach demonstrates several significant results. On one hand, systems involving an interaction with a prosocial opponent were significantly more likely to sustain mutual cooperation ($t(98) = 16.37$, $p < 0.001$) and restore cooperation after an opponent's defection (i.e. the DC outcome, $t(98) = 6.25$, $p < 0.001$) relative to those which involved an interaction with an individualistic opponent. On the other hand, systems involving

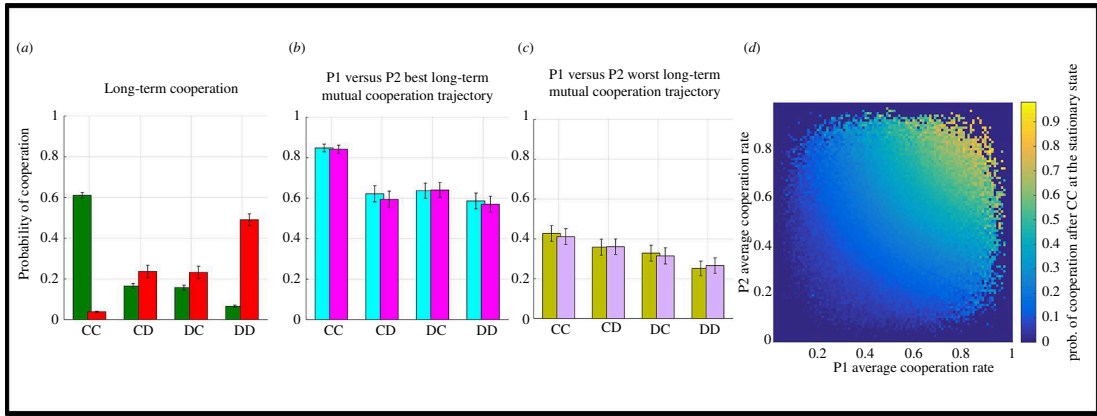


Figure 4. Simulated results for maintaining long-term interpersonal cooperation. (a) System stability for best (green bars) and worst (red bars) faring players. Online cooperation rates for (b) best- and (c) worst-faring players. A total of 1000 simulated agents per group. Error bars designate ± 1 s.e.m. (d) The relationship between P1 versus P2 average cooperation rates across all outcome types and long-term mutual cooperation at the stationary state. Dark blue regions in the parameter space are not covered due to average cooperation rates not pushing to the extreme values.

interaction with an individualistic opponent were significantly more likely to restore cooperation after the participant's unilateral and both players' mutual defection ($t(98) = 8.94$, $p < 0.001$ and $t(98) = 8.80$, $p < 0.001$ for CD and DD outcomes, respectively), relative to those which involved competition with a prosocial opponent.

The results at the stationary state also suggest that interpersonal cooperation will erode in time, considering that cooperation levels at the stationary state (figure 3c) were globally lower than individual rates calculated from the experiment (figure 3a,b).

3.5. Behavioural phenotypes that can safeguard long-term cooperation

Erosion of interpersonal cooperation as demonstrated in the preceding section highlights that it is crucial to identify behavioural phenotypes that can safeguard long-term interpersonal cooperation. Using equations (3.1 and 3.2) across 2×10^4 simulations, it is possible to identify (by reverse engineering) the best versus worst strategies in terms of long-term mutual cooperation. After 2×10^4 simulations, the results described below are robust and fully reproducible even if probabilities stored in the matrix $\mathbf{M}(\mathbf{s}_i, \mathbf{i}_s)$ in equation (3.1) are generated randomly (e.g. using MATLAB's rand function). An important insight arising from these simulations is that strategies that fare best in terms of maintaining long-term mutual cooperation are in fact the worst when it comes to breaking cycles of mutual defection (figure 4a), and vice versa. The comparison of individual player profiles (figure 4b versus figure 4c) suggests that players that are projected to maintain high levels of long-term cooperation are globally more cooperative across the board (i.e. following different outcome types: CC, CD, DC and DD) and likely to establish mutual cooperation online (i.e. during the interactive decision-making phase). Correlations computed within each group in order to prevent Simpson's paradox [32] suggest that online mutual cooperation (CC) rates are the best predictor of long-term mutual cooperation for best-faring players (within top 5% of simulated agent pairs, $r = 0.429$, $p < 0.001$), whereas online cooperation after DD outcomes correlate significantly with long-term mutual cooperation for the worst-faring players (within bottom 5% of simulated agent pairs, $r = 0.151$, $p < 0.001$). The relationship between each player's average cooperation rates and mutual cooperation probability at the stationary state suggests a positive trend (figure 4d; from bottom left to top right increase) albeit with some nonlinearity in the parameter space.

3.6. Human behavioural phenotype risks extinction against individualistic strategies

After computing the transition probabilities at Markov stationary states by equation (3.2), it is relatively straightforward to calculate each player's expected outcome (i.e. evolutionary fitness). The expected payoff of any agent at the stationary state is computed by

$$\mathbf{v} \cdot \mathbf{S}_X = D(\mathbf{p}, \mathbf{q}, \mathbf{S}_X), X \in \{\text{sub}, i, p\}, \quad (3.3)$$

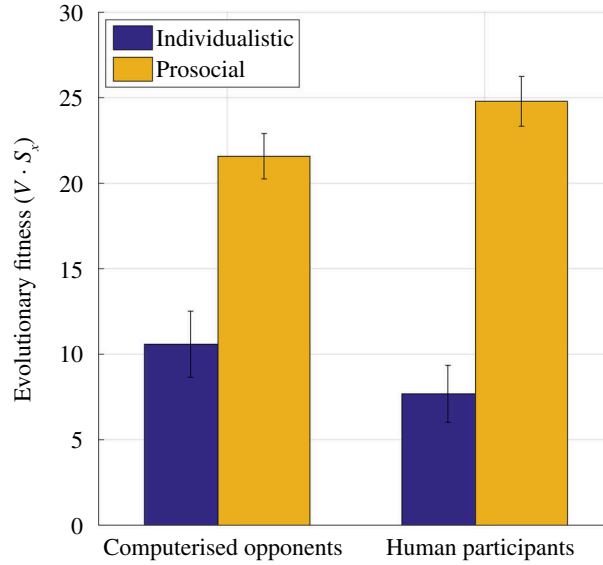


Figure 5. Evolutionary fitness at the Markov stationary states. Irrespective of agent type (i.e. human players or [computerized] opponents) evolutionary fitness is significantly higher in social interactions with a prosocial opponent (main effect of opponent type, *** $p < 0.001$). Error bars designate ± 1 s.e.m.

where $\mathbf{S}_X = \{\bar{R}_O^{CC}, \bar{R}_O^{CD}, \bar{R}_O^{DC}, \bar{R}_O^{DD}\} | X = \text{sub}, \text{ and so forth, contains the numerical reward values of each agent's unique } 2 \times 2 \text{ matrix in a vector format (figure 2b,c); while } \mathbf{p} \text{ and } \mathbf{q} \text{ replace each agent's vector of cooperation probabilities which were previously described (i.e. } \mathbf{s}_i \text{ and } \mathbf{i}_s). \text{ Here, } D(\mathbf{p}, \mathbf{q}, \mathbf{S}_X) \text{ requires no normalization considering } (\mathbf{1} \cdot \mathbf{v}) = 1. \text{ Note that average reward amounts for DC and DD outcomes are negated in this calculation as these represent forgone reward amounts due to the opponent's rejection in the UG (i.e. the reward amounts sacrificed to convey that the proposed offer amount was unfair). Each player's evolutionary fitness at the Markov stationary state was computed, revealing how human players would fare against their opponents (figure 5). The analysis of evolutionary fitness by a } 2 \times 2 \text{ (opponent type (individualistic versus prosocial) } \times \text{ agent type (computer versus human)) multivariate ANOVA suggested a main effect of opponent type } (F(1,98) = 80.377, p < 0.001) \text{ and a marginally significant opponent } \times \text{ agent type interaction } (F(1,98) = 3.805, p = 0.054). \text{ Here, the significant main effect of opponent type highlights the accumulated resource gap between individualistic and prosocial agents, considering that numerical values of the task environment were identical across all pairwise interactions. Although subsequent pairwise comparisons between human participants and computerized opponents were not statistically significant (all } p > 0.105), \text{ the results from the Markov stationary states indicate that humans would prevail over prosocial strategies (extinction rate: } p(\mathbf{v} \cdot \mathbf{S}_p > \mathbf{v} \cdot \mathbf{S}_{\text{sub}}) = 0.14), \text{ whereas they would fare mostly poorly against individualistic strategies } (p(\mathbf{v} \cdot \mathbf{S}_i > \mathbf{v} \cdot \mathbf{S}_{\text{sub}}) = 0.70, \text{ also see figure 4). This would be in line with the preceding results suggesting that human players fail to escape extortion against individualistic opponents.}$

For completeness, note that human players' total evolutionary fitness (f) can be calculated as:

$$f = (\mathbf{v} \cdot \mathbf{S}_{\text{sub}}^i + \mathbf{v} \cdot \mathbf{S}_{\text{sub}}^p) - (\mathbf{v} \cdot \mathbf{S}_i^{\text{sub}} + \mathbf{v} \cdot \mathbf{S}_p^{\text{sub}}). \quad (3.4)$$

3.7. Do ZD strategies exist in the wild?

In the preceding sections, the analysis of average payoff amounts (figure 2b,c) suggested that individualistic opponents can establish extortionate terms over the human players, which is one of the key properties of the ZD strategies discovered by Press and Dyson [14]. Although Press and Dyson's [14] discovery posits the existence of ZD strategies algebraically, this does not necessarily mean that they emerge naturally during social interactive decision making between two competitors. To estimate the probability with which ZD strategies can be observed in a typical lab-based social interactive decision-making experiment, we investigated whether a direct linear relationship is established between the payoffs of the two players (i.e. opponents versus human players) in our recruitment cohort. Note that previous studies implementing the ZD framework to human decision making in the IPD [33]

approximated these linear relationships between payoffs by relying on correlations rather than trying to enforce necessary algebraic conditions in which the determinants would vanish. Following Press and Dyson's notation (also see their eqn 7), I investigated such a possibility, $p(\tilde{\mathbf{q}} = \varphi \mathbf{S}_{\text{sub}}^i + \kappa \mathbf{S}_i^{\text{sub}} + \varepsilon \mathbf{1})$, by numerically estimating the values of $\varphi, \kappa, \varepsilon$ (using MATLAB's *fmincon*) which maximized the function:

$$l = -\ln(1 - [\mathbf{1} \cdot [\tilde{\mathbf{q}}[\varphi \mathbf{S}_{\text{sub}}^i + \kappa \mathbf{S}_i^{\text{sub}} + \varepsilon \mathbf{1}]]]) \quad (3.5)$$

and computed $p(l) = 0$ across all interactions in our recruitment cohort (computerized opponent versus human player and vice versa). If $p(l) = 0$ is satisfied, it would mean that the determinant vanishes and it would suggest that (in this example, opponent's) strategy is truly a ZD strategy against the human player. However, out of 100 combinations between computerized opponents and human players, it was not possible to identify any strategy truly satisfying this condition.

4. Discussion

The majority of human experimental studies investigating interpersonal cooperation by using variants of the PD game [34–36] used well-established, yet somewhat rigid, computerized strategies such as *tit-for-tat* or *tit-for-two-tats*, which were previously thought to be evolutionarily competitive [5]. The main reasoning behind using these strategies in behavioural experiments is to understand how human participants fare against strategies which are shown to have competitive evolutionary fitness, as established by theoretic simulation-based studies. Later expansions of the IPD strategy landscape by Pavlovian [37] and ZD strategies [14] provided highly novel but mostly theoretical insights about the evolution of prosocial behaviours [2]. The current experimental approach, founded in bridging two key behavioural economic paradigms, provides novel insights about human interpersonal cooperation particularly against agents that follow a stochastic social-value function, just as human decision makers do in behavioural economic experiments [20].

By following up on Press and Dyson's [14] proposal that a convergence between IPD and UG can naturally emerge when a player with ToM engages in social interactive decision making, I provided quantitative evidence in favour of this proposal, which essentially transforms human ultimatum bargaining behaviour to be more suitable for analysis using an evolutionary game-theoretic approach. The results communicated in this work suggest that ToM sophistication often exhibited by human players [18] naturally promotes interpersonal cooperation even when players are self-interested, trying to maximize their short-term monetary gains (figure 3). However, the current results also suggest that SVO of the opponents play a significant role, and human players cannot escape extortion by opponents with an individualistic SVO (figure 2b,c), to such degree that they would eventually risk extinction more frequently than not. Irrespective of opponents' SVO, equilibria at the stationary states for these social interactive systems suggest that interpersonal cooperation will erode substantially in time (figure 3c). Exploratory simulations complementing these findings suggested that maintaining online mutual cooperation is the strongest predictor of long-term mutual cooperation, whereas the ability to break cycles of mutual defection (i.e. higher probability of cooperating after DD outcomes) was associated with poor long-term mutual cooperation rates. This is also similar to how individualistic agents behaved in the experiment, overall performing poorly in terms of their reward trajectory (figure 5), yet driving cooperative human players (figure 3b) to (near) extinction.

Results communicated in this work suggest that opponents' SVO (i.e. individualistic versus prosocial) directly influence players' payoffs. For example, sustaining mutual cooperation with prosocial agents yields much better payoffs for human players. At the same time, human players were shown to give in to extortion by individualistic opponents (figure 2b,c). Despite following very different methodologies, these findings are strikingly in line with the results of a previous IPD study in which healthy volunteers played against different computerized agents following stationary conditional probabilistic strategies [24]. Subsequent work suggested that extortionate strategies (i.e. similar to the opponent with individualistic SVO) can outperform generous (i.e. similar to the opponent with prosocial SVO) strategies that play against human players, but only if the experiment duration is reasonably long [33] (500 trials). In our experiment (across 120 trials), human participants tried to bank as much money as possible [20] and they ended up acting in a significantly more cooperative manner than their opponents while doing so (figure 3a,b). A subsequent evolutionary stability analysis at the Markov stationary states (figure 3c) revealed that systems involving interactions with prosocial opponents are significantly more likely to sustain mutual cooperation (i.e. CC outcomes), whereas

systems involving interactions with individualistic opponents are more likely to break cycles of mutual defection (i.e. DD outcomes), albeit this may be at the expense of human participants settling down for less (figure 2c). Over finite evolutionary horizons, these differences converge and show that human participants would prevail over prosocial, but fare poorly against individualistic agents that acted in an extortionate manner (figure 5). This is also in line with findings from previous studies that used longer behavioural protocols [33]. Taken together, these findings suggest that human players focusing on maximizing short-term goals may be oblivious to, and unprepared for the long-term consequences of their decision strategies. This shortsightedness would undermine long-term evolutionary/social-economic fitness in a world in which humans need to interact with artificial agents or even corporate entities which should be quantitatively better equipped to estimate their reward trajectories [38].

Only a limited number of studies investigated interactions between human players and ZD strategies. These studies formulated ZD strategies in terms of stationary conditional probabilities derived such that there would be a correlation between the computerized agents' and the human players' payoffs [24,33], allowing extortionate ZD strategies to acquire higher, whereas generous ZD strategies to acquire lower rewards. This approach satisfies the linear relationship prerequisite defining ZD strategies originally formulated by Press and Dyson. In the current work, I took a different approach and derived the conditional probabilities of the computerized agents from their stochastic decisions to accept or reject ultimatums. As a result, the conditional probabilities of the computerized opponents were varied in the current work for each participant (as these also depended on participant choice behaviour, figure 3a,b) and each opponent, allowing this work to account for wider individual variability and establish a birds-eye view on the long-term trajectory of human interpersonal cooperation. Within this inherent variability in participant and opponent strategies, I numerically estimated the degree to which a stringent linear relationship can be established between the payoffs of humans versus computerized opponents (both players following a stochastic social-value function), as mathematically formulated by Press and Dyson [14], and vice versa. Under these conditions, the estimated likelihood of ZD strategies emerging naturally in cooperative/competitive interactions would be <1%. However, considering the extent of biodiversity on planet Earth and the vast number of interpersonal exchanges taking place between billions of people every day, I do not think these results undermine the ecological validity of ZD strategies or the influence of this framework.

To the best of my knowledge, the current work is the first to follow up Press and Dyson's hypothesis of expressing UG in terms of IPD. Here, I demonstrate that in interactions between human participants with ToM sophistication playing against opponents following a stochastic social-value function, it is possible to bridge these social-economic games. Although the vast majority of the previous literature focused on human responder behaviour in the UG [9,12,36,39,40], this study is based on recent work describing value computations underlying human proposer behaviour [20], which demonstrates that more studies investigating this aspect of social interactive decision making can reveal rich and multidimensional insights about the nature and mechanisms of human prosocial behaviours.

Ethics. The original study was given ethical approval by the Center for Information and Neural Networks (CiNet) Research Ethics Committee. This is a reanalysis of the published data with different analytic approaches.

Data accessibility. The numerical outputs to generate original figures is submitted as Supplementary Material [41].

All analysis scripts and data are available from the servers of the Open Science Framework [42].

Declaration of AI use. I have not used AI-assisted technologies in creating this article.

Authors' contributions. E.P.: conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing—original draft.

Conflict of interest declaration. I declare we have no competing interests.

Funding. No funding has been received for this article.

Acknowledgements. A pre-print version of this manuscript is published in BioRxiv [43].

References

1. Andreoni J. 1990 Impure altruism and donations to public goods: a theory of warm-glow giving. *Econ. J.* **100**, 464–477. (doi:10.2307/2234133)
2. Fehr E, Fischbacher U. 2003 The nature of human altruism. *Nature*. **425**, 785–791. (doi:10.1038/nature02043)
3. Fletcher JA, Doebeli M. 2009 A simple and general explanation for the evolution of altruism. *Proc. R. Soc. B.* **276**, 13–19. (doi:10.1098/rspb.2008.0829)
4. Lehmann L, Keller L. 2006 The evolution of cooperation and altruism—a general framework and a classification of models. *J. Evol. Biol.* **19**, 1365–1376. (doi:10.1111/j.1420-9101.2006.01119.x)

5. Axelrod RM. 1997 *The complexity of cooperation: agent-based models of competition and collaboration*. Princeton University Press. (doi:doi: 10.1515/9781400822300) See <https://www.degruyter.com/document/doi/10.1515/9781400822300/html>
6. Fehr E, Schmidt KM. 1999 A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868. (doi:10.1162/003355399556151)
7. Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, Kilts CD. 2002 A neural basis for social cooperation. *Neuron*. **35**, 395–405. (doi:10.1016/S0896-6273(02)00755-9)
8. Boyd R, Gintis H, Bowles S, Richerson PJ. 2003 The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. USA*. **100**, 3531–3535. (doi:10.1073/pnas.0630443100)
9. de Quervain DJF, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E. 2004 The neural basis of altruistic punishment. *Science*. **305**, 1254–1258. (doi:10.1126/science.1100735)
10. Bazzan ALC, Bordini RH, Campbell JA. 2002 Evolution of agents with moral sentiments in an Iterated Prisoner's Dilemma exercise. In *Game theory and decision theory in agent-based systems* pp. 43–64. (doi:10.1007/978-1-4615-1107-6)
11. Farrell J, Ware R. 1989 Evolutionary stability in the repeated Prisoner's Dilemma. *Theor. Popul. Biol.* **36**, 161–166. (doi:10.1016/0040-5809(89)90027-0)
12. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. 2003 The neural basis of economic decision-making in the Ultimatum Game. *Science*. **300**, 1755–1758. (doi:10.1126/science.1082976)
13. van 't Wout M, Kahn RS, Sanfey AG, Aleman A. 2006 Affective state and decision-making in the Ultimatum Game. *Exp. Brain Res.* **169**, 564–568. (doi:10.1007/s00221-006-0346-5)
14. Press WH, Dyson FJ. 2012 Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proc. Natl. Acad. Sci. USA*. **109**, 10409–10413. (doi:10.1073/pnas.1206569109)
15. Ichinose G, Masuda N. 2018 Zero-determinant strategies in finitely repeated games. *J. Theor. Biol.* **438**, 61–77. (doi:10.1016/j.jtbi.2017.11.002)
16. Tanimoto J, Tanimoto J. 2021 Evolutionary game theory: fundamentals and applications for epidemiology. In *Sociophysics approach to epidemics* pp. 13–60. (doi:10.1007/978-981-33-6481-3)
17. Wang Z, Kokubo S, Jusup M, Tanimoto J. 2015 Universal scaling for the dilemma strength in evolutionary games. *Phys. Life Rev.* **14**, 1–30. (doi:10.1016/j.plrev.2015.04.033)
18. Yoshida W, Dolan RJ, Friston KJ. 2008 Game theory of mind. *PLoS Comput. Biol.* **4**, e1000254. (doi:10.1371/journal.pcbi.1000254)
19. Yoshida W, Seymour B, Friston KJ, Dolan RJ. 2010 Neural mechanisms of belief inference during cooperative games. *J. Neurosci.* **30**, 10744–10751. (doi:10.1523/JNEUROSCI.5895-09.2010)
20. Pulcu E, Haruno M. 2020 Value computations underlying human proposer behavior in the ultimatum game. *J. Exp. Psychol. Gen.* **149**, 125–137. (doi:10.1037/xge0000621)
21. Kahneman D, Tversky A. 1972 *The concept of probability in psychological experiments*. Springer. (doi:10.1007/978-94-010-2288-0)
22. Murphy RO, Ackermann KA, Handgraaf MJJ. 2011 Measuring social value orientation. *Judgm. Decis. Mak.* **6**, 771–781. (doi:10.1017/S1930297500004204)
23. Van Lange PA, Otten W, De Bruin EM, Joireman JA. 1997 Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence. *J. Pers. Soc. Psychol.* **73**, 733–746. (doi:10.1037//0022-3514.73.4.733)
24. Hilbe C, Röhl T, Milinski M. 2014 Extortion subdues human players but is finally punished in the Prisoner's Dilemma. *Nat. Commun.* **5**, 3976. (doi:10.1038/ncomms4976)
25. Milinski M. 2022 Extortion - a voracious prosocial strategy. *Curr. Opin. Psychol.* **44**, 196–201. (doi:10.1016/j.copsyc.2021.08.033)
26. Hilbe C, Chatterjee K, Nowak MA. 2018 Partners and rivals in direct reciprocity. *Nat. Hum. Behav.* **2**, 469–477. (doi:10.1038/s41562-018-0320-9)
27. Maner JK. 2017 Dominance and prestige: a tale of two hierarchies. *Curr. Dir. Psychol. Sci.* **26**, 526–531. (doi:10.1177/0963721417714323)
28. McClanahan KJ, Maner JK, Cheng JT. 2022 Two ways to stay at the top: Prestige and dominance are both viable strategies for gaining and maintaining social rank over time. *Pers. Soc. Psychol. Bull.* **48**, 1516–1528. (doi:10.1177/01461672211042319)
29. Cushing CA, Dawes AJ, Hofmann SG, Lau H, LeDoux JE, Taschereau-Dumouchel V. 2023 A generative adversarial model of intrusive imagery in the human brain. *PNAS Nexus*. **2**, pgac265. (doi:10.1093/pnasnexus/pgac265)
30. Rahwan I et al. 2019 Machine behaviour. *Nature*. **568**, 477–486. (doi:10.1038/s41586-019-1138-y)
31. Hauert C, Schuster HG. 1997 Effects of increasing the number of players and memory size in the iterated Prisoner's Dilemma: a numerical approach. *Proc. R. Soc. Lond. B*. **264**, 513–519. (doi:10.1098/rspb.1997.0073)
32. Simpson EH. 1951 The interpretation of interaction in contingency tables. *J. R. Stat. Soc., Ser. B, Methodol.* **13**, 238–241. (doi:10.1111/j.2517-6161.1951.tb00088.x)
33. Wang Z, Zhou Y, Lien JW, Zheng J, Xu B. 2016 Extortion can outperform generosity in the iterated Prisoner's Dilemma. *Nat. Commun.* **7**, 11125. (doi:10.1038/ncomms11125)
34. Gradin VB, Pérez A, Macfarlane JA, Cavin I, Waiter G, Tone EB, Dritschel B, Maiche A, Steele JD. 2016 Neural correlates of social exchanges during the Prisoner's Dilemma game in depression. *Psychol. Med.* **46**, 1289–1300. (doi:10.1017/S0033291715002834)
35. Mokros A, Menner B, Eisenbarth H, Alpers GW, Lange KW, Osterheider M. 2008 Diminished cooperativeness of psychopaths in a Prisoner's Dilemma game yields higher rewards. *J. Abnorm. Psychol.* **117**, 406–413. (doi:10.1037/0021-843X.117.2.406)
36. Pulcu E et al. 2015 Social-economical decision making in current and remitted major depression. *Psychol. Med.* **45**, 1301–1313. (doi:10.1017/S0033291714002414)
37. Nowak M, Sigmund K. 1993 A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*. **364**, 56–58. (doi:10.1038/364056a0)

38. Russell S, Dewey D, Tegmark M. 2015 Research Priorities for Robust and Beneficial Artificial Intelligence. *AI Magazine*. **36**, 105–114. (doi:[10.1609/aimag.v36i4.2577](https://doi.org/10.1609/aimag.v36i4.2577))
39. Crockett MJ, Clark L, Tabibnia G, Lieberman MD, Robbins TW. 2008 Serotonin modulates behavioral reactions to unfairness. *Science*. **320**, 1739. (doi:[10.1126/science.1155577](https://doi.org/10.1126/science.1155577))
40. Koenigs M, Tranel D. 2007 Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *J. Neurosci.* **27**, 951–956. (doi:[10.1523/JNEUROSCI.4606-06.2007](https://doi.org/10.1523/JNEUROSCI.4606-06.2007))
41. Pulcu E. 2024 Data from: Individualistic attitudes in iterative Prisoner's Dilemma undermine evolutionary fitness and may drive cooperative human players to extinction. Figshare. (doi:[10.6084/m9.figshare.c.7124576](https://doi.org/10.6084/m9.figshare.c.7124576))
42. OSF. 2023 Data from: IPD_ZeroDeterminant. Data and supporting analyses scripts for the project. Open Science Framework. (https://osf.io/rjvk5/?view_only=7562a7d2aa344918b4fa93018cf3df94)
43. Pulcu E. 2023 Individualistic attitudes in Iterative prisoner's dilemma undermine evolutionary fitness and may drive cooperative human players to extinction. *bioRxiv*.