Journal of Behavioral and Experimental Economics 105 (2023) 102034

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Intragroup punishment and intergroup conflict aversion weaken intragroup cooperation in finitely repeated games

Jonathan H W Tan [a,*], Friedel Bolle [b,1]

[a] Department of Economics, School of Social Sciences, Nanyang Technological University, 48 Nanyang Avenue, 639818, Singapore
[b] European University Viadrina, Frankfurt, (Oder), Germany

ARTICLE INFO

JEL classification:
C72
C91
D74
D91

Keywords:
Cooperation
Conflict
Altruism
Quantal response
Finitely repeated games

ABSTRACT

We experimentally analyze group-specific social preferences and dynamic strategies in finitely repeated games where helping partners hurts rivals. Cooperation with partners decreases when it inflicts losses on rivals, even when it is socially efficient. Cooperation decreases with rivalry, while it increases with synergy and previous cooperation by partners but not rivals. Our structural model of bounded rationality estimates preferences and strategies. It shows that participants anticipate future payoffs from partner cooperation and punish partner defection with unyielding defection, while further averting intergroup conflict due to altruism towards rivals. Intragroup cooperation and intergroup conflict weaken over time through these two channels.

1. Introduction

Cooperation often occurs in groups amidst conflict with other groups. Intragroup cooperation involves *synergy* that increases the welfare of ingroup *partners*, whereas intergroup conflict involves *rivalry* that decreases the welfare of outgroup *rivals*. This applies to various economic interactions. For example, when firms compete in divisible goods markets, market shares and profits are deterministically affected by the actions of partners and rivals. In the course of time, decision makers face a recurring dilemma of benefitting partners at a personal cost and loss to rivals. Such conflicts occur repeatedly within or between firms (e.g. competing project teams; Birkinshaw, 2001) and networks (e.g. countervailing airline alliances; Gimeno, 2004).[2] We seek to further understand the motives of behavior in such interactions.

Previous studies show that parochial altruism motivates acts that benefit the ingroup at a cost to the self and outgroup (Böhm et al., 2002; Choi & Bowles, 2007). This implies group-specific social preferences (Chen & Li, 2009).[3] However, there is also evidence of prosocial concerns for rivals, e.g. participants prefer to avoid hurting the outgroup when helping the ingroup (Halevy et al.;, 2008), and prosocials will help both ingroup and outgroup (Aaldering et al., 2013). This implies social welfare concerns (Engelmann & Strobel, 2002; Kritikos & Bolle, 2001). Moreover, repeated play allows individual profits to be made via cooperation strategies entailing the punishment of co-player defection (Dal Bó & Fréchette, 2018; Embrey et al., 2018; Fudenberg & Maskin, 1986; Kreps et al., 1982). The use of such strategies in repeated intergroup games is suggested by evidence of how cooperation is conditioned on feedback (Tan & Bolle, 2007) and beliefs (Bornstein et al., 1996) of co-player behavior. The belief that co-players use punishment strategies in finitely repeated games support cooperation and the use of such strategies, which consistently increase with the "value" (i.e. incentive) of cooperation in a supergame (Embrey et al., 2018).

The purpose of our study is to test the alternative hypotheses of group-specific preferences and strategies that drive intergroup conflict in finitely repeated games. The context-sensitivity of behavioral patterns across previous experiments, as the next section further surveys, shows

* Corresponding author.
E-mail address: j.tan@ntu.edu.sg (J.H.W. Tan).
1 In loving memory of Friedel Bolle, a co-author of this paper, who passed away on October 10, 2021.
2 From a broader view, when groups fight for scarce resources, dominance, and survival (e.g. territory or food), the spoils are shared between group members while competing groups are excluded (Rusch, 2014a), and fights can last several generations (Bernhard et al., 2006; Choi and Bowles, 2007; Norenzayan et al., 2016; Rusch, 2014a).
3 For example, Chen and Li (2009) find more charity and less envy towards the ingroup than the outgroup.

https://doi.org/10.1016/j.socec.2023.102034
Received 16 November 2022; Received in revised form 5 May 2023; Accepted 13 May 2023
Available online 18 May 2023

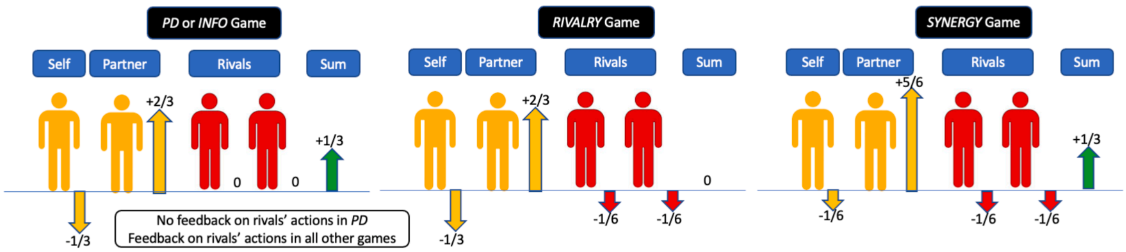

Fig. 1. Marginal payoff implications of cooperation in the experimental games.

the plausibility of both motives. However, less is known about whether – and if so how – they drive repeated intergroup conflict in tandem. We contribute towards filling this gap by testing the sensitivity of cooperation strategies and conflict behavior to variations in the payoff structure and information feedback, and analyzing the underlying structure of bounded rationality, utility, and strategic interaction. We shall show that the intersection of cooperation strategies entailing the unyielding punishment of defecting partners and altruism towards rivals weakens intragroup cooperation over time.

Fig. 1 illustrates our experimental games. They are framed as interactions between alliance firms, which accords with our cited applications (similar examples include Ke et al., 2015; Konrad et al., 2020). This can help induce group identity (Tajfel et al., 1971; Tajfel and Turner, 1979) and a competitive mindset (Sheremeta, 2018) essential to settings where parochialism is expected. We elicit a benchmark of parochialism with our *Rivalry* game, where partners gain and rivals lose while social welfare stays constant when one cooperates. Cooperation is "welfare-neutral" in that it deterministically results in a direct transfer from the self and rivals to the partner, thus revealing parochialism. We compare *Rivalry* with two other games to elicit group-specific motives. Ingroup motives are revealed in the *Synergy* game where cooperation increases when partners (rivals) gain more than (lose as much as) in *Rivalry*. Outgroup motives are revealed in the *Info* game where cooperation increases when partners gain as much as in *Rivalry* but "rivals" do not incur losses while feedback on their actions is still provided. It controls for feedback effects germane to repeated conflict between individuals (Fallucchi et al., 2013) or groups (Böhm & Rockenbach, 2013; Bornstein et al., 2002; Tan & Bolle, 2007), which can be estimated by comparing it to the *PD* (prisoner's dilemma) with the same payoff structure but without feedback on rivals' actions.

Cooperation decreases with rivalry, while it increases with synergy and previous cooperation by partners but not rivals. This indicates prosocial concerns for the outgroup, and that dynamic strategies play a greater role within group than between groups. To explain these group-specific behaviors, we structurally estimate preferences and strategies. *Logit quantal responses* (McKelvey & Palfrey, 1995) to expected payoffs determine the probabilities of actions, which predicts that cooperation increases with synergy, i.e. the marginal per capita return (Andersen et al., 1998). Partner cooperation is best explained by a combination of exogenous beliefs that depend on previous actions and limited look-ahead in anticipation of future payoffs (Johnson et al., 2002), fostered by the belief that partners punish defection. Ingroup motives are internalized in cooperative strategies toward mutual benefit, but mutual cooperation is fragile and is broken by acts of defection. In contrast, outgroup altruism explains conflict aversion to mutual harm in the long run. The tendency for groups to mutually acquiesce de-escalates intergroup conflict over time and is thus self-limiting.[4] Thus, intragroup cooperation and intergroup conflict diminish with repeated play through two channels: ingroup punishment and outgroup altruism.

Section 2 relates our approach to the literature. Section 3 describes the experiment. Section 4 reports the behavioral patterns and structural analysis. Section 5 discusses and concludes.

2. Background

2.1. Group conflict

Dark clouds. Increased cooperation is observed in intergroup conflicts where cooperation reduces social welfare, relative to isolated PDs where it improves welfare (e.g. Bornstein, 1992; Bornstein & Ben-Yossef, 1994; Bornstein et al., 1996). Observed effort is also notably higher than predicted in intergroup Tullock contests (e.g. Abbink et al., 2012; Ahn et al., 2011; Cason et al., 2017; Eisenkopf, 2014; Ke et al., 2015), where effort is inefficient as it wastefully goes towards increasing the probability that a group wins a prize, which either group can instead effortlessly but randomly win. Such findings indicate that intergroup conflict is driven by competitive, cooperative, or parochial concerns (Böhm et al., 2020; Sheremeta, 2018).

Repeated play. However, the gap between cooperation levels with and without welfare-reducing intergroup conflict closes with repeated play (Bornstein et al., 1996), as does the gap between observed and predicted effort levels in Tullock contests (Abbink et al., 2012). In contrast, increased effort is sustained in team competitions where effort is efficient as it increases social welfare through joint production. For example, effort increases when the team that exerts the most effort wins a fixed prize in addition to the benefits from team production (Nalbantian & Schotter, 1997) or receives a higher return to team effort (Tan & Bolle, 2007).[5] This indicates that intergroup conflict weakens over time when it is socially inefficient.

Silver linings. The negative effect of social inefficiency on intergroup conflict is found in past experiments. For example, Halevy et al. (2008) find that participants cooperate more to help the ingroup without hurting the outgroup, which is efficient, compared to when this option is unavailable and cooperation also hurts the outgroup, which is inefficient. Also, cooperation increases when competing groups can *all* win if they are tied compared to when one's gain is another's loss, i.e. without negative externalities for each other (Reuben & Tyran, 2010). Further, prosociality can motivate personal sacrifice in the interest of the outgroup (Aaldering et al., 2013). Such efficiency concerns suggest prosocial motives beyond parochial norm-driven beliefs of duty to the ingroup (Böhm et al., 2020), i.e. prosocial concerns for the outgroup.

Beliefs. The relevance of dynamic cooperation strategies in intergroup conflict is indicated by post-experimental questionnaire responses on how decisions are based on the beliefs that participants hold on the cooperativeness of ingroup and outgroup (Bornstein & Ben-Yossef, 1994). It is also reflected by the negative relationships between cooperation and previous round behavior by partners or rivals (Tan & Bolle,

[4] Böhm et al. (2016) and Rusch (2014b) find that intragroup cooperation is driven by attack and more likely defense.

[5] In Tullock contests and team competitions, outcomes are indeterministic and indirectly inflicts losses on rivals. To help identify group-specific social preferences in our games, the direct tradeoff between ingroup and outgroup payoffs is made salient with outgroup losses deterministically resulting from effort.

2007). Anticipatory and responsive strategies are evident. For example, Böhm et al. (2016) observe retaliatory and pre-emptive motives not spite-driven unilateral strikes in a sequential move game where first movers can mitigate losses by striking first. Consistently, participants avoid conflict by coordinating through communication (Leibbrandt & Sääksvuori, 2012). Given preference heterogeneity, beliefs of future behavior in Tullock contests between participants are updated based on the self-projection of one's type and observations of others' actions (Ke et al., 2015; Konrad & Morath, 2020). This indicates dynamic strategies between partners and rivals.

2.2. Social preferences and strategic reasoning

The collection of experimental results above suggest that behavior is sensitive to welfare implications and repeated play. However, the structure of preferences and reasoning that generates these results remains unclear. The purpose of our structural estimates is to ascertain which group-specific motives best explain the data with a potentially – albeit not necessarily – pluralistic model. We therefore first elicit parochialism without efficiency confounds using a game where social welfare stays constant with cooperation, and then compare it to games that elicit prosocial concerns for the outgroup. The combination of repeated games controls for efficiency across games, allowing for a unified structural analysis of motives. Our structural analysis estimates group-specific social preferences and strategies in the data. As prompted by the literature (e.g. Sheremeta, 2018), we consider a broad range of social preferences including altruism, spite, inequity aversion, efficiency and reciprocity concerns. Each model has separate parameters for ingroup and outgroup co-players j, following Chen and Li's (2009) approach.

$$u^{Alt}(\mathbb{X}) = x_i + \sum_{j \neq i} \alpha_j x_j \tag{1}$$

Eq. (1) presents the simplest model of linear *altruism* (*spite*) utilities that increase (decrease) in the payoffs x of self i and other j, and $\alpha_j > 0$ ($\alpha_j < 0$) implies a *positive* (*negative*) dependency on others payoffs, implying biases such as "ingroup love" ("outgroup hate"). This model is labeled as *Alt*. Parochial altruism can thus be described by different altruism or spite parameters α_j for the respective co-players. We also consider group biases in *inequity aversion* (*FS*; Fehr & Schmidt, 1999) and *reciprocity* (*CR*; Charness & Rabin, 2002).

$$u_i^{FS}(\mathbb{x}) = x_i - \sum_{j \neq i} \alpha_j (x_j - x_i) \cdot I_{x_i < x_j} + \sum_{j \neq i} \beta_j (x_j - x_i) \cdot I_{x_i \geq x_j} \tag{2}$$

$$u_i^{CR}(\mathbb{x}) = x_i - \sum_{j \neq i} (\alpha_j + \theta_j q_j)(x_j - x_i) \cdot I_{x_i < x_j} + \sum_{j \neq i} (\beta_j - \theta_j q_j)(x_j - x_i) \cdot I_{x_i \geq x_j} \tag{3}$$

Inequity averse utility (Eq. (2)) decreases with the payoff differences between self and others, and this effect is greater if one is envious when $x_j > x_i$ than if one is guilty $x_j < x_i$, i.e. $\alpha_j > \beta_j$. The envy parameter α_j and guilt parameter β_j are also found in the *CR* reciprocity model (Eq. (3)), which additionally penalizes a co-player for making an inefficient move by increasing envy and decreasing guilt towards the co-player by θ_j, whereby the reciprocity term is active with $q_j = 1$ when the co-player has made an inefficient move, and otherwise nullifies the effect of θ_j in the case that the co-player did not make an inefficient move.

$$u_i^{RDA}(\mathbb{x}) = x_i + \sum_{j \neq i} \alpha_j x_j \cdot I_{x_i < x'} + \sum_{j \neq i} \beta_j x_j \cdot I_{x_i \geq x'} \tag{4}$$

Altruism and fairness concerns can also be modeled by *reference dependent altruism* (Eq. 4; *RDA*; Breitmoser & Tan, 2013, 2020), which is a good (out-of-sample) predictor for diverse static and dynamic games. It also explains behavior in three-player demand bargaining and majority bargaining games, where behavior is incompatible with quasi-concave utility functions such as *FS, CR*, and CES altruism. RDA allows us to test the nature of conditionality on a reference point x', which is one's ex ante expected payoff (*absolute* RDA) or on co-players' payoffs (*relative* RDA). Relative RDA allows for fairness concerns. The strength of altruism depends on whether one's payoff is less or more than the expected payoff or co-players' payoff, with a discontinuity at the point where payoff equals the reference point, which is given by α_j and β_j respectively. We also extend RDA to consider the dynamic nature of our data by testing its conditionality on historical payoffs.

Further, we consider social preferences jointly with dynamic strategies that could play a role in finitely repeated games under incomplete information of co-player types (Kreps et al., 1982). Punishment strategies can support cooperation in finitely repeated games especially in the earlier part of the supergame as long as there is some probability and belief that others are playing such strategies, as shown theoretically by Kreps et al. (1982) and experimentally with a meta-analysis of finitely repeated game experiments and a new experiment by Embrey et al. (2018). Embrey et al. find that participants cooperate in the earlier rounds of the supergame and use trigger strategies as part of "threshold strategies" (of when to defect first) played under the incomplete information of their co-players' thresholds. Participants are heterogeneous and aware of this inherent feature of experimental environments. Thus, incomplete information on the dynamic strategies or social preferences of co-players and heterogeneity is present in our data and considered in our analysis.[6]

We analyze three main classes of dynamic strategies. First, one could follow *rule-based* strategies such as those described by Kreps et al. (1982). With "tit-for-tat" (*TfT*), one's action follows that of the co-player's in the previous round. Alternatively, one cooperates until the co-player defects, and then defects ever after; this "punishment" strategy (*Punish*) in our finitely repeated game (Fudenberg & Maskin, 1986) is akin to "Grim" strategies in infinitely (Dal Bó & Fréchette, 2018) or finitely repeated games (Embrey et al., 2018). Such rules have no explicit time horizon, and do not involve utility maximization. Implicitly, the belief of partners playing such strategies drives cooperation. Second, we assume explicit exogenous belief formation based on the past behavior of co-players. Third, dynamics are endogenously driven due to reciprocity or the use of past payoffs as reference points.

We allow for bounded rationality: quantal responses relax the best response assumption by letting the profitability of an option determine its choice probability. The exogeneous formation of beliefs is combined with logit quantal responses to these beliefs, and reciprocity is combined with the endogenous formation of beliefs by assuming logit *quantal response equilibrium* (QRE; McKelvey & Palfrey, 1995). Players expect others to mix strategies according to QRE, and expect others to expect the same. Further, players who operate with a two-period time horizon *limited lookahead* (LLA; Johnson et al., 2002) also expect their future selves ("agents") to mix strategies, yielding an *agent QRE* (AQRE; McKelvey & Palfrey, 1998). Structural models allow the juxtaposing of social preferences and bounded rationality. We evaluate which explanation fits the data best, and estimate the strengths of each motive.

3. The experiment

3.1. Games

A game has four players $i = A, B, C, D$ who form two groups $J = \{A, B\}$ and $K = \{C, D\}$. The game is repeated for ten periods. In each period, a player has an endowment e, which can be interpreted as a resource constraint. One has two possible choices: (1) cooperate by contributing

[6] Embrey et al. (2018) further find that subjects learn to delay their defection before co-players do to later in the supergame – in support of the learning of such nonequilibrium strategies.

Table 1
Games, parameters, and predictions*.

1a) Intragroup cooperation						
	Cooperate	Defect				
Cooperate	200, 200	100, 250				
Defect	250, 100	150, 150				
1b) Intragroup cooperation amidst intergroup conflict						
	0 rival cooperates: π−0		1 rival cooperates: π−25		2 rivals cooperate: π−50	
	Cooperate	Defect	Cooperate	Defect	Cooperate	Defect
Cooperate	200, 200	100, 250	175, 175	75, 225	150, 150	50, 200
Defect	250, 100	150, 150	225, 75	125, 125	200, 50	100, 100
1c) Parameters, Nash equilibria, and social optima of the four games.						
	PD	*Info*	*Rivalry*	*Synergy*		
r_{in}	2/3	2/3	2/3	5/6		
r_{out}	0	0	−1/6	−1/6		
Information	No	Yes	Yes	Yes		
Individual NE	e_i=0	e_i=0	e_i=0	e_i=0		
Group NE	e_i=$\boldsymbol{e}$	e_i=$\boldsymbol{e}$	e_i=$\boldsymbol{e}$	e_i=$\boldsymbol{e}$		
Social optimum	e_i=$\boldsymbol{e}$	e_i=$\boldsymbol{e}$	indifferent	e_i=$\boldsymbol{e}$		

* Player 1 (2) receives the payoff stated on the left (right) of each cell, and chooses strategies by row (column). The outcome is the determined by the intersection of both players' choices. In 1b, payoffs depend also on whether 0, 1, or 2 rivals cooperate within the other group, resulting in the payoff π from intragroup cooperation (given by the payoffs in Table 1a) minus losses 0, 25, or 50, respectively. The payoffs are for partners in the group, payoffs of rivals are determined respectively.

$e_i = e$ to a group project, or (2) not cooperate by allocating effort to an individual project, i.e. $e_i = 0$.[7] The total effort by group J is $e_J = e_A + e_B$, and $e_K = e_C + e_D$ for group K. Choosing the individual project (i.e., defecting) does not affect co-players' payoffs, i.e. $e - e_i = e$ where $e_i = 0$. Cooperating on the group project has a positive impact on partners $r_{in}e_J$, where r_{in} captures the degree of synergy between partners and $0.5 < r_{in} < 1$. Cooperation in the rival group has a negative impact $r_{out}e_K$, where r_{out} captures the degree of rivalry between groups and $r_{out} \leq 0$. Players are symmetric in e, r_{in}, and r_{out}. The payoff to player i from group J is given by

$$x_i = e - -e_i + r_{in}e_J + r_{out}e_K. \quad (5)$$

If $r_{in} < 1$ and players are purely self-interested and rational, they would prefer for partners to incur the cost of effort. When $r_{in} > 0.5$, profits accrue for both players if they cooperate, i.e. both put effort into the group project. When $r_{out} = 0$ ($r_{out} < 0$), groups are independent (rivalrous) because players do not (do) suffer losses due to the rivals' efforts.[8]

Assuming players maximize individual payoffs, the Nash equilibrium (NE) is $e_i = 0$ for all players, i.e. no player ever invests in the group project because every such strategy is strictly dominated by zero effort. Alternatively, if players maximize group payoffs, i.e. each group acts as one player, then intragroup cooperation and intergroup conflict described by $e_i = e$ for all i is the unique NE. Finally, if all four players collude to maximize social welfare, then they will put effort in the individual project if $2(r_{in} + r_{out}) < 1$ and the group project if $2(r_{in} + r_{out}) > 1$. By backward induction of the finitely repeated game, the stage game equilibrium holds. However, social preferences or dynamic strategies can cause deviations from this benchmark prediction (Andreoni, 1995; Keser & van Winden, 2000).

Our experiment has four games: *PD*, *Info*, *Rivalry*, and *Synergy*.*PD* is presented in Table 1a and *Rivalry* in Table 1b. Table 1c summarizes parameters and equilibrium predictions for individual, group, and society across games. This combination of games systematically isolates the treatment effects of gains, losses, and information. The parameter values stretch the variance of gains and losses across games, maintaining defection as the unique individual equilibrium prediction, allowing for direct comparability and causal inference. For all games, $e = 150$ to avoid decimals in payoffs and to ease the calibration of marginal incentives.

In *PD*, pairs of participants play in isolated groups ($r_{out} = 0$) with synergy between partners ($r_{in} = 2/3$), with information on partner choices. Responses to feedback on partner indicate cooperation strategies. *Info* differs from *PD* by providing information on the actions of rivals after each period. Such information is germane to intergroup interactions and can be used for social comparisons or to evaluate relative performance and efficacy in reaching goals (Festinger, 1954; Suls & Miller, 1977).[9] This information is extraneous as it bears no pecuniary consequence. Observed responses to this information in *Info* thus reflect concerns for non-pecuniary payoffs from comparison outcomes.

Rivalry is implemented by decreasing r_{out} to $-1/6$ from 0 in *Info*. This manipulation allows us to test the effect of potential losses from rivalry, while keeping the potential gains from cooperation and information feedback similar to *Info*. The value for r_{out} is chosen such that actions have the same efficiency, i.e. welfare-neutral. At a personal cost, the ingroup gains by the same amount as the outgroup loses. An decrease in cooperation relative to that in *Info* therefore reflect ingroup motives. Cooperation in *Rivalry* benchmarks parochialism.[10]

Synergy introduces, relative to *Rivalry*, an increase in r_{in} from 2/3 to 5/6, to test the effect of increased ingroup benefits. The degree of losses inflicted on rivals, r_{out}, is the same as that in *Rivalry*. An increase in cooperation relative to that in *Rivalry* therefore reflect ingroup motives. Cooperation is equally efficient in *Info* and *Synergy*, as each unit

[7] Eliciting binary choices simplifies the task and reduces noise from confusion in the varying impact of actions on different co-players, and allows for a clearer evaluation of responses to previous co-player behavior. It also represents the class of collective action scenarios such as those involving participation or market entry and contest.

[8] A positive (negative) [zero] difference in ingroup-outgroup effort implies that groups have won (lost) [tied] and that the group payoff will increase (decrease) [stay constant].

[9] Information on rivals' actions is essential in the performance comparisons that determine competitive outcomes. It could drive retaliation, enhance group salience and biases (Tan & Zizzo, 2008), or serve as a reference point. Also, individuals might derive utility from the joy of winning (Sheremeta, 2018).

[10] This simplification removes the uncertainty of how an action impacts the payoffs of partners and rivals, such as via success functions in Tullock contests, and thus reduces noise from cognitive complexity. It differs from previous conflict experiments where cooperation yields social welfare gain (e.g. intergroup social dilemmas) or loss (e.g. Tullock contests; Sheremeta, 2018). Unlike the IPD-MD of Halevy et al. (2008), *Rivalry* is welfare-neutral and does not involve an option of helping the ingroup while not hurting the outgroup, which we instead compare with *Info*.

contributed to the group increases social welfare by 0.5.[11] This allows us to control for efficiency while varying cooperation motives, which is stronger in *Synergy* than in *Rivalry*. Responses to feedback on rival actions indicate conflict strategies. In summary, we test these hypotheses.

H1. Positive ingroup motives increase intragroup cooperation in *Synergy* relative to *Rivalry*.
H2. Positive outgroup motives decrease intergroup conflict in *Rivalry* relative to *Info*.
H3. Cooperation strategies motivate intragroup cooperation conditional on partner behavior.
H4. Conflict strategies motivate intergroup conflict conditional on rival behavior.

3.2. Logistics and procedure

The experiment was conducted at the CeDEx experimental economics laboratory at the University of Nottingham. We recruited from a database of potential experimental participants via the recruitment software ORSEE (Greiner, 2004) and email. They were undergraduates and postgraduates from various disciplines. The experiment was programmed in z-Tree (Fischbacher, 2007), and the profit calculator and questionnaire described below were programmed in Visual Basic 6. Appendix 1Appendix 2Experimental instructions (see Online Appendix 1) and control questionnaires were provided in print. Online Appendix 2 shows the screenshots. The online appendices are found on OSF (https://osf.io/z4epn/). Pairs of sessions involving the same game and four participants each were conducted simultaneously in the laboratory to enhance anonymity. Participants knew that there were two sessions running simultaneously, and that participants for each session were randomly seated in the room. We ran six sessions for each of the four games involving a total of 96 participants (52% male, SD = 0.50). Games were rotated across session pairs. Computer terminals were partitioned to prevent communication and seeing others' screens.

Pre-experimental instructions and control questionnaire. After being seated, participants read the experimental instructions and answered a control questionnaire to check their understanding of the rules. The experimental supervisor individually advised participants with queries. The text consistently employed an alliance frame to increase group salience. Participants were told that they were managers of firms in an alliance, and that they could choose between Project I and Project II (implying defection and cooperation, respectively, within a group). This was to contextualize the task consistently with one of its main applications to make it more intuitive and engaging.

Training stage. This next stage allowed participants to engage in an undirected use of a "profit calculator" for five minutes before playing the game. The training stage was included primarily to ensure that participants had the chance to familiarize themselves with the payoff structure of the game, e.g. the consequences of different combinations of choices made by each co-participant. It can be interpreted as a form of pre-operational strategic evaluation stage, where executives are able to familiarize themselves with different aspects of the business setting.

Game play. After the training phase, the computer program randomly assigned participants to their respective groups. They were told their group assignments, and that the participants they were matched with would remain constant throughout the experiment. Participants did not know the group assignments of the other participants in the room. Participants played ten periods of one of four games. We continued to provide the generic profit calculator, because it is a natural way for participants to gather information about the consequences of alternative decisions. The profit calculator was on a computer to the participant's right, and the game was played on a computer to the participant's left.

After receiving feedback in each period, participants then proceeded to the next period until all ten periods were done. In these games they could accumulate earnings in experimental points, each worth £0.01.[12] To minimize noise from participant confusion, experimental points were scaled for cognitive simplicity and straightforward to convert to experimental earnings. This conversion rate was calibrated to yield an expected payment rate comparable with economics experiments ran in the UK.

Final questionnaire. After playing the experimental games, we collected self-reported data on gender, course of study, professional experience, Social Value Orientation (SVO; Messick and McClintock, 1968) and Regulatory Focus (RFQ; Higgins et al., 2001). The relationships between conflict behavior and are SVO or RFQ are reported in our working paper (Tan & Bolle, 2019). Participants entered their responses on the computer.

Payments. After completing the final questionnaire, the computer then reported to each participant how much one had earned, and this information was kept private and unknown to other participants. Participants were paid one at a time by the experimental supervisor. Payments were cumulative across periods. Participants could earn a minimum of £10 and a maximum of £27.50 depending on game, their choices and those of their co-participants. Sessions lasted up to 1 hour. The average payment was £16.53.

Data. We confirm that, for all experiments, we have reported all measures, conditions,[13] data exclusions, and how we determined their sample sizes.[14] The data and experimental materials are available on OSF (https://osf.io/z4epn/).

4. Results

4.1. Treatment effects

Fig. 2 presents the cooperation rates for each game. Our initial test for treatment effects compares games using non-parametric Mann Whitney U tests (p-values are two-tailed). The first of two steps that we take to account for the potential non-independence of data due to feedback and interactions between participants is to treat the average cooperation per session as an independent observation. The second step is to use regression analysis, which further allows us to control for experience from repeated interactions and the responses to the behavior of others. We use binary logit regressions because of the nature of our dependent variable "cooperate" or "not cooperate" and errors are clustered at the session level to account for the non-independence of observations across repeated interaction.

Table 2 presents the regression results. The dependent variable is *Coop* (= 1 if one chose Project II, i.e. to cooperate, or 0 if otherwise). All models include the independent variables of *Period* (=1 to 10) to control

[11] *Synergy* connects our study to those on intergroup conflict where cooperation is efficient. The net effects of r_{in} and r_{out} determine the difference between *Info* and *Synergy* in cooperation rates. Strong outgroup hate can even elicit more cooperation in *Rivalry* relative to *INFO*.

[12] Payments were not based on a randomly chosen round as Stahl and Haruvy (2006) showed how this may result in overstating the measured extent of non-self-interest motivated behavior. It is also consistent with the assumptions of finitely repeated games.

[13] Three other treatments with asymmetric costs within groups, communication and revenue transfers were also conducted at that time for another project on promises by leaders as part of the broader research program.

[14] We determined our sample size following the conventions of related experiments, e.g. a meta-study by Fiala and Suetens (2017) reports that experiments on oligopolies and public goods games have an average of 7.74 (SD = 3.63) and 9.96 (SD = 7.08) independent observations per treatment, respectively. As a gauge, the power achieved based on t-tests of differences in the observed means (SD = 0.477) with 24 participants per treatment is 0.65 for *Info-Rivalry*, 0.2 for *Rivalry-Synergy*, and 0.24 for *Info-Synergy* comparisons. However, our nonparametric tests use session means as independent observations and regressions with robust standard errors clustered by session.

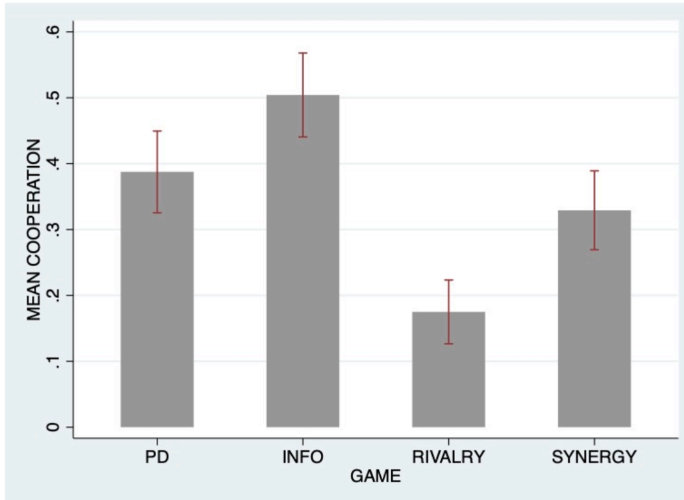

Fig. 2. Mean cooperation by game.

Table 2
Binary logit regressions of cooperation*.

	(1)	(2)	(3)	(4)
Period	−0.151***	−0.146***	−0.151***	−0.146***
	(0.020)	(0.032)	(0.020)	(0.032)
LPartnerCoop	1.185***	1.092***	1.185***	1.092***
	(0.216)	(0.258)	(0.216)	(0.258)
LRivalCoop		0.008		0.008
		(0.240)		(0.240)
PD	−0.558		−0.558	
	(0.411)		(0.411)	
Rivalry	−1.287***	−1.299***		
	(0.372)	(0.493)		
Synergy	−0.714**	−0.717*		
	(0.343)	(0.410)		
r_{in}			3.438**	3.492**
			(1.558)	(1.622)
r_{out}			7.723***	7.794***
			(2.231)	(2.958)
Constant	0.258	0.268	−2.034*	2.060*
	(0.345)	(0.663)	(1.108)	(1.127)
Pseudo R^2	0.1346	0.1359	0.1346	0.1359
N	864	648	864	648

*The dependent variable is *Coop* (= 1 if one cooperated, or 0 if otherwise). Standard errors clustered by session are found in parentheses. The p-values reported here are two-tailed, with * denoting $p < 0.10$, ** denoting $p < 0.05$, and *** denoting $p < 0.01$.

for experience, and a lagged variable *LPartnerCoop* (= 1 if the partner cooperated in the previous period, and = 0 if otherwise). Model 1 compares cooperation in each game relative to *Info* using dummy variables *PD*,*Rivalry*, and *Synergy* (= 1 for the respective treatments, and = 0 for *Info*). Model 2 includes *LRivalCoop* (= rivals' effort in previous period, i.e. 0 1, or 2), to test for responses to partner and rival behavior in the previous period, respectively, but for this data from *PD* is dropped due to the lack of feedback on behavior in the other group. For robustness, models 3 and 4 corroborate by replacing game dummies with treatment variables that capture the degree of synergy with r_{in} (= 2/3 for *PD*,*Info*, and *Rivalry*, or = 5/6 for the *Synergy*) and of rivalry with r_{out} (= 0 for *PD* and *Info*, or = −1/6 for the *Rivalry* and *Synergy*). As before, model 3 tests for the absence of feedback on the other group's actions with *PD* (= 1 for *PD*, or = 0 for *Info*,*Rivalry*, and *Synergy*), while model 4 replaces *PD* with *LRivalCoop*.

We use *Info* as a benchmark to control for the effect of information on rival behavior. Compared to *PD* where all game parameters are otherwise the same, cooperation rates are not significantly higher in *Info* ($z = 1.366$, $p = 0.172$). We check if participants cooperate to generate ingroup gains even though it results in equal outgroup losses, which reveals a positive bias for the ingroup relative to the outgroup keeping social welfare constant. The positive cooperation rate in *Rivalry* of 0.18 is evidence for parochialism. Further, cooperation increases when we increase ingroup gains r_{in}, keeping outgroup losses r_{out} constant. There is significantly more cooperation in *Synergy* at 0.33 than in *Rivalry* ($z = 2.082$, $p = 0.037$). Ingroup gains therefore increase cooperation at the cost of the outgroup.

However, there is less cooperation in these two games where $r_{out} < 0$, compared to *Info* where $r_{out} = 0$, which has a higher cooperation rate of 0.50. Compared to *Info*, cooperation is lower in *Rivalry* ($z = -2.169$, $p = 0.030$) but not significantly so in *Synergy* ($z = -1.604$, $p = 0.109$). However, controlling for responses to previous play, *Rivalry* is negative and significant in models 1 and 2, showing that cooperation decreases when it yields outgroup losses while social welfare stays constant. Therefore, cooperation decreases in the presence of intergroup conflict and this happens not only when there is a decrease in efficiency. This indicates that conflict aversion is a product of group-specific preferences rather than universal efficiency concerns. Models 3 and 4 show that r_{in} (as is *Synergy* in Models 1 and 2) and r_{out} are positive and significant, confirming that cooperation increases with partner gains but decreases with partner losses, respectively.[15]

Next, we consider the effect of feedback and experience on partner and rival behavior on cooperation over time. Fig. 3 shows how cooperation rates change over time in each game. Cooperation rates are lowest in *Rivalry*, which decreases to almost zero by the final round. Across treatments, cooperation decreases over time, as shown by the negative and significant *Period* coefficient in all models. Fig. 4 illustrates the responses to feedback by showing cooperation rates averaged across all games and periods as a function of own, partner, and rival actions in the previous period. Each line depicts the group's state in the previous period. The leftmost to rightmost line respectively depicts cooperation rates in response to mutual defection, unilateral defection, unilateral

[15] We collected data on work experience, and there are no significant differences in behavior. Males made up 42% in *PD*, 71% in *Info*, 54% in *Rivalry*, and 42% in *Synergy*. Regressions in Appendix 1 Table A1 show that males cooperate more and experienced participants cooperate less, and the main results are robust to these controls. They are also robust to controls for Economics students, which were not significant and dropped.

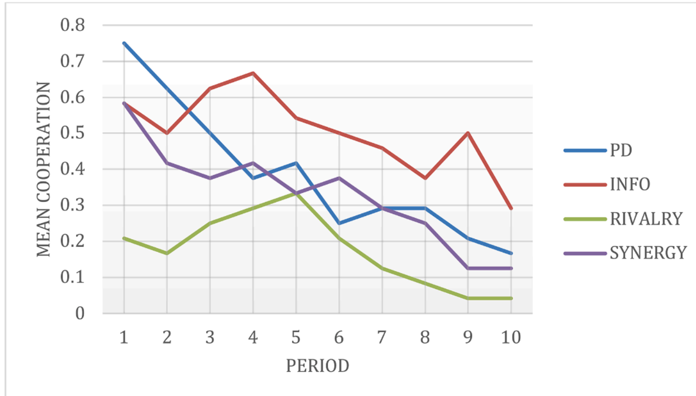

Fig. 3. Mean cooperation over time.

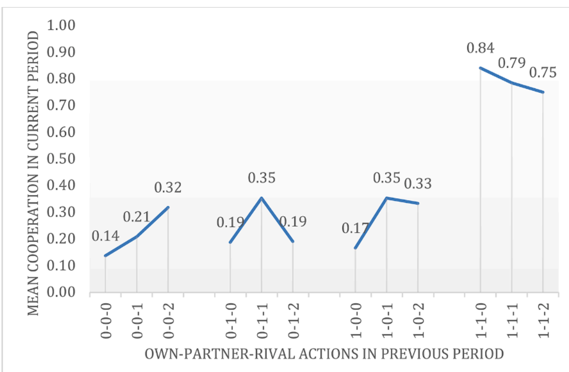

Fig. 4. Cooperation as a function of own, partner, rival actions in the previous period.

cooperation, and mutual cooperation in the previous round.

The use of a trigger strategy where partner defection is punished by defection is suggested by how cooperation rates are highest after mutual cooperation but drop sharply after unilateral defection by either partner or self and stay low after mutual defection. Regressions show that *LPartnerCoop* is positive and significant across models, confirming that participants match the previous defection of partners with defection. This suggests that partner cooperation decreases over time because of dynamic strategies. The increase in cooperation in SYNERGY relative to RIVALRY as incentives to cooperate increases and the responsiveness to partner cooperation indicates the use of a cooperative strategy within groups. This lends support to H1 and H3. In comparison, responses to rivals are however much smaller in magnitude. This suggests that participants cooperate less when it inflicts hurt on rivals because of an intrinsic concern for rivals, which by symmetry averts conflict.[16] To corroborate, cooperation rates are not significantly higher in *INFO* compared to *PD* as the MWU test above shows, *PD* is not significant in models 1–4, and *LRivalCoop* in models 2 and 4 is not significant. This lends support to H2 but not H4. These contrasting patterns of ingroup and outgroup dynamics show that static preferences per se cannot fully explain behavior in our games, thus motivating our structural analysis to follow.

[16] Figure 3 shows that the relatively weak responses to rivalry counteract with previous cooperation within the group, suggesting the escalation (de-escalation) of conflict: cooperation rates increase (decrease) with cooperation in the rival group if partners had mutually defected (cooperated), and increase (decrease or hold) with unilateral (mutual) cooperation in the rival group if the participant or partner had unilaterally defected.

Table 3
Distribution of current actions as a function of own and partner actions in the previous period*.

Condition	*LPartnerCoop*	*LCoop* = 1 *Coop* = 1	 *Coop* = 0	*LCoop* = 0 *Coop* = 1	 *Coop* = 0	Fisher's p-value
PD	0	9	26	16	76	0.322
PD	1	43	11	7	28	<0.001
INFO	0	9	21	20	52	0.814
INFO	1	74	10	4	26	<0.001
RIVALRY	0	6	25	18	126	0.386
RIVALRY	1	5	5	8	23	0.241
SYNERGY	0	17	39	18	66	0.240
SYNERGY	1	12	10	18	38	0.077

* *LCoop* = lagged own cooperation, *LPartnerCoop* = lagged partner cooperation, *Coop* = current period cooperation. Two-sided Fisher Exact Test between *Coop* = 1 and *Coop* = 0.

4.2. Structural analysis of strategies and preferences

Table 3 shows the distribution of current actions disaggregated by game as a function of own and partner actions in the previous period (see Appendix 2, Table A2 for conditionality on rivals' actions). Fisher's Exact Tests comparing *Coop* = 0 and 1 in Table 3 confirm that defection (*LCoop* = 0) is unconditionally repeated, while cooperation (*LCoop* = 1) is conditionally repeated only if the partner has also cooperated ((*LPartnerCoop* = 1: n.s.; *LPartnerCoop* = 1: $p < 0.1$ or better except RIVALRY where cooperation is low overall).[17] The decaying pattern of cooperation over time shapes our structural model. We evaluate theories that predict behavior in period t based on behavior in $t - 1$. Our evaluation is therefore based on the frequencies of cooperating in 2 × 2 classes (*LCoop* by *LPartnerCoop*) for *PD*, and in the 3 × 2 × 2 × 3 = 36 classes (game by *LCoop*, by *LPartnerCoop*, by *LRivalCoop*) for the other three games.

We run structural analyses of the social preferences and boundedly rational dynamic strategies that best explain the data. Table 4 summarizes the results of our structural analysis of various models and their fit based on the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC).[18] We combine altruism or inequity aversion with exogenous beliefs (depending on previous period actions), or reciprocity that has changing coefficients of altruism or inequity aversion with endogenous beliefs (determined by QRE). Combining static utilities with

[17] Table A2 in the Appendix 2 shows that previous rival decisions have weak overall impact.

[18] The Akaike Information Criterion is AIC = $2d - 2ln(\text{LL})$ with number of parameters d. The Bayes Information Criterion is BIC = ln(O)*d-2ln(LL) with number of observations O (Schwarz, 1978).

Table 4
Overview of the performance of structural models*.

Time horizon		Static: Present period				Limited Lookahead: Present and next period			
Characterization	Theory	#Par	-LogL	AIC	BIC	#Par	-LL	AIC	BIC
Mixture of rules	*Punish, …*	7	443.8	901.6	934.9	–	–	–	–
Static utility[§]	*Altruism*	4	525.4	1059.0	1071.0	5	441.4	892.8	916.6
& exogenous beliefs	*FS*	5	464.9	939.8	944.5	6	453.0	918.0	943.0
	RDAexp	6	444.2	900.4	929.0	7	441.1	896.2	929.5
	RDAhist	6	455.9	923.8	952.4	7	439.7	893.4	926.7
Dynamic utility &	*DynAlt*	5	517.7	1044.0	1069.0	6	480.6	973.2	1002.0
endogenous beliefs	*DynFS*	7	467.6	949.2	982.5	8	467.6	951.2	989.3

[*] BIC is computed with 864 data points.
[§] *RDAhist* has a dynamic component but is estimated under the same assumptions as *RDAexp* in order to allow comparisons. *DynFS* = *CR*.

endogenous beliefs predicts the same behavior in every period, which is obviously different from observed behavior as shown in Figs. 3 and 4. Combining dynamic utility with exogenous beliefs introduces two competing sources of dynamic behavior. For static utilities with exogenous beliefs, we evaluate decisions from the viewpoint of player A, whose partner is player B and rivals are players C and D.

We denote utility with $u_i = u_A$ following the social preference models described in Eqs. (1) to (4). One's payoff x_i is determined by the expected payoff x_A, and co-players' expected payoffs x_j are x_B for ingroup player B as well as x_C and x_D for outgroup players C and D. The preference parameters (when applicable in a model) are α_B, β_B, and θ_B for ingroup player B and α_{out}, β_{out}, and θ_{out} for outgroup players C and D. For *PD*, $\alpha_{out} = 0$, $\beta_{out} = 0$, and $\theta_{out} = 0$ due to the lack of feedback on outgroup incomes. Expected payoffs are determined by the following *exogenous beliefs*. If both players have contributed or defected, they are expected to repeat their actions. If a player has defected and the partner has cooperated, she continues defecting, and the partner is expected to cooperate with probability η. The estimated mixture of behavioral rules shows that the majority of participants play the dynamic strategy of *Punish*,[19] where co-players are expected to cooperate if η=1, and defect for the remaining rounds, i.e. unyielding defection if η=0. The same holds for the beliefs of rival strategies.

We first assume that players maximize $u_A(t)$ in every period. Following the definition of RDA in Section 2.2, for the absolute RDA models, the reference payoff x_i' in *RDAexp* is the *expected payoff*, while *RDAhist* conditions on the *historical payoff*, i.e. one's previous period payoffs.[20] The payoff x_A is always the expected payoff, depending on whether one cooperated (*Coop* = 1) or not (*Coop* = 0). Coefficients in the relative RDA models depend on payoffs relative to those of partners and rivals as follows.

- α_B applies only if $x_A < x_B'$, otherwise it is substituted by β_B,
- α_{out} applies only if $x_A < \frac{x_C' + x_D'}{2}$, otherwise it is substituted by β_{out}.

Finally, we extend the models by assuming that players have limited lookahead (*LLA*), i.e. present period actions shape expectations of expected payoffs $u_A(t+1)$ and optimal actions for the next period. In our models, looking ahead strengthens incentives for partner cooperation because one player's defection in period t implies high probabilities of both players defecting in $t + 1$ and therefore in all further periods. Letting players look ahead more than one period increases this incentive, but such an effect does not differ much from an increased discount rate (which is allowed to be larger than 1). Players maximize

$$w_A(t) = u_A(t) + \delta^* u_A(t+1). \tag{6}$$

For comparability with the other approaches, we include the final period of data in the estimations. For all models, we assume errors in the sense of QRE, such that the action that maximizes utility is not chosen with probability 1 but that the probability of cooperation is determined by logit quantal responses

$$p_A(Coop = 1) = \frac{e^{\lambda u_A(Coop=1)}}{e^{\lambda u_A(Coop=1)} + e^{\lambda u_A(Coop=0)}}. \tag{7}$$

For LLA models, u is substituted by w. Table 4 shows that models with exogenous beliefs outperform those with endogenous beliefs, and within the class of models with exogenous beliefs those with altruism outperform the *FS* model. Estimates for the other models (mixture of rules, dynamic utility and endogenous beliefs) are found in Appendix 2, Table A4. Extending the time horizon by one future period (*LLA*) does not improve both *FS* models while all altruism models improve significantly. The best-fitting models are *Alt-LLA* and *RDAhist-LLA*.

The absolute quality of each theory is evaluated by two measures. The first is a χ^2 test, which shows that the predicted frequencies are not significantly different from the empirical frequencies (see Table 5). The second is the best possible log-likelihood score, which is reached when a theory predicts exactly the empirical frequencies.[21] This prediction implies LL = –418.7. The prediction of the same average frequency for all cases has LL = –598.9. Theories can improve on the latter, but can never reach the former. With LL = –439.7 for *RDAhist-LLA* and LL = –441.4 for *Alt-LLA*, the two models are closer to the upper bound than other theories.

The parameter estimates of the four best models, all of which involve altruism, are reported in Table 5. Appendix 2, Table A5 reports the estimates for the other reciprocity models such as with *FS* inequity aversion and *CR* reciprocity. The ingroup altruism coefficients α_B and β_B, as well as the probability of cooperating with a previously defecting partner η, are not significantly different from zero implying a strategy of unyielding defection in response to partner defection. The outgroup altruism coefficients α_{out} and β_{out} are significantly different from zero. By and large, the objective function of the first three models in Table 5 can be written as

$$u_A(t) = x_A(t) + 0.6^* x_{out}(t) \tag{8}$$

$$w_A(t) = u_A(t) + 1.25^* u_A(t+1). \tag{9}$$

Participants expect partners' actions to be rule-based (η = 0, i.e. unyielding punishment of partner defection) and not utility-based, and this cooperative strategy accounts for the absence of altruism towards partners in Eq. (8). $\delta = 1.25$ in Eq. (9) indicates that, on average

[19] Appendix 2, Table A3 reports the population shares of rules: 71% of participants play *Punish*, 14% maintain their previous actions, 1% play *TfT* with partners, 0.5% play *TfT* with rivals, and 16% randomly deviate from these rules.

[20] Thus, *RDAhist* has a dynamic component, but as we want to compare it with *RDAexp* we evaluate this model under the same assumption on belief formation.

[21] This is possible only as a population mixture of deterministic behavior. Actual behavior is different because players often respond differently to the same previous period pattern of decisions.

Table 5
Parameter estimates and standard deviations for the best altruism models*.

	λ	β_B	α_B	β_{out}	α_{out}	η	δ	$p(\chi^2)$
Alt-LLA	2.924	−0.075		0.601		0.041	1.370	0.187
S.D.	1.017	0.204		0.170		0.089	0.403	
RDAhist-LLA	2.661	0.152	−0.005	0.705	0.732	0.099	1.172	0.268
S.D.	1.076	0.209	0.210	0.223	0.237	0.111	0.369	
RDAexp-LLA	4.520	−0.434	−0.217	0.417	0.422	0.017	1.223	0.135
S.D.	2.282	0.501	0.264	0.222	0.226	0.099	0.236	
RDAexp	3.735	1.000	0.602	0.697	0.724	0.055	–	0.044
S.D.	0.504	0.010	0.035	0.141	0.145	0.086	–	

* Standard deviations (*S.D.*) are computed from the Hessian of the log-likelihood function. χ^2 is determined for the predictions of contribution frequencies in the 40 classes defined by the actions in the four treatments by self, partner, and rivals in the previous period. Models and parameters are defined in Section 2.2. We denote λ = precision parameter of QRE, α_B (β_B) = altruism parameter for A's in-group partner B in the case he has less (at least as much) income than (as) A, α_{out} (β_{out}) = altruism parameter for A's outgroup rivals C and D in the case they have less (at least as much) income than (as) A, η = probability of continuing to cooperate after the partner has defected, and δ = discount rate.

participants look ahead more than one period.[22] Cooperation is fostered by one's increased expected payoff in the next and the following periods, but hindered by one's decreased expected payoff in the present period and by altruism towards rivals, which explains H1, H2, and H3.

5. Discussion and conclusion

We analyzed intragroup cooperation in finitely repeated intergroup conflict games, where helping partners hurts rivals, and found evidence of conflict aversion. Whereas there was a willingness to cooperate when it would yield outgroup loss, it was significantly less than with intragroup gains (H1) or without intergroup conflict (H2). Participants anticipated the benefits from cooperation, but also considered the losses that it would inflict on rivals. Indeed, participants responded to feedback on the lagged actions of partners (H3) but not of rivals. Our structural analysis showed that these patterns are explained by the belief that partners will punish defection with unyielding defection and by "altruism" to the outgroup.

Cooperation improves partner welfare and is observationally compatible with altruism. However, "parochialism" is explained by self-serving strategies rather than altruism. The incentive to cooperate is derived from expected profits from partner cooperation, rather than a self-sacrificing preference for the welfare of partners, as is evident from the unyielding punishment that follows from partner defection. Limited lookahead reinforces cooperation but is unstable as it breaks down upon defection under the expectation of unyielding punishment. Thus, cooperation was observed in the initial rounds, but it diminished over time.[23]

Outgroup altruism further motivates the de-escalation of conflict and acquiescence between groups, which serves to avert mutually destructive conflict. It limits hostility when the incentive to engage in intergroup conflict is weakened with less synergy, implying a greater cost and risk of ingroup cooperation, and with more rivalry, implying a lower cost of outgroup altruism and greater expected cost of rival retaliation. The hesitation to retaliate, reflected in the lack of lagged responses to rivals, therefore implies altruism rather than dynamic strategies. Outgroup "altruism" can thus be interpreted as a proximate (i.e. not fueling a war) rather than an ultimate (i.e. pure altruism) cause for intergroup conflict aversion.

Our study demonstrated the limits of intergroup conflict as a motivation for intragroup cooperation in the context of finitely repeated games. It operates through two channels: intragroup cooperation is driven by dynamic cooperative strategies that entail punishment of defection, but attenuated by conflict aversion driven by altruism. Intragroup cooperation in repeated intergroup conflicts therefore diminishes over time. Future research could test the generalizability of these results to conflict behavior in infinitely repeated games with controlled random termination rules and sequences with a minimum number of periods (Frechette & Yuksel, 2017), other dynamic strategies or preferences, larger samples including non-student ones, neutral and other loaded frames, and different group identity inducements.

Data availability

The data and experimental materials are available on OSF (https://osf.io/z4epn/).

Acknowledgements

We thank the Associate Editor María Paz Espinosa, two anonymous reviewers, Pak Hung Au, Yves Breitmoser, Corentin Curchod, Simon Gächter, Ken Kamoche, Alexander Kritikos, George Kuk, Rajesh Kumar, Lars Lefgren, Lim Wee Kiat, Daniele Nosenzo, Yohanes Eko Riyanto, David Saal, Steve Thompson, Wenjie Wang, Jonathan Yeo, Zichen Zhao, Daniel Zizzo and participants of talks in Ningbo and Berlin for the advice and encouragement. We are grateful to Daniele Nosenzo for programming the experiment, to Shahin Bhugaloo and Ruslan Kabalin for the experimental assistance, and to CeDEx and CRIBS for the logistical support. Tan gratefully acknowledges the financial support from Nottingham University Business School under the Spark Fund, Nanyang Technological University under the Start-Up Grant, and the Ministry of Education Singapore under the AcRF Tier 1 Grant (RG126/20).

Appendix 1. Robustness checks

Table A1

[22] While participants may evaluate the expected income of only the next period (limited lookahead), they know that with more periods ahead cooperating with partners will increase all future expected incomes. A discount rate larger than 1 takes approximately into account an extended time horizon.

[23] We do not claim that dynamic strategies can sustain perpetual cooperation as it could in infinitely repeated or indefinite horizon games, but that it can motivate cooperation in the earlier parts of finitely repeated games. Indeed, the cited theoretical literature was originally motivated by the need to explain experimental observations that deviate from subgame perfection in finitely repeated games.

Table A1
Binary logit regressions of cooperation with covariates*.

	(5)	(6)	(7)	(8)
Period	−0.151***	−0.140***	−0.151***	−0.140***
	(0.022)	(0.034)	(0.022)	(0.034)
LPartnerCoop	1.147***	1.066***	1.147***	1.066***
	(0.200)	(0.245)	(0.200)	(0.245)
LRivalCoop		0.042		0.042
		(0.235)		(0.235)
PD	−0.434		−0.434	
	(0.434)		(0.434)	
Rivalry	−1.149***	−1.126**		
	(0.402)	(0.500)		
Synergy	−0.503	−0.484		
	(0.339)	(0.391)		
r_{in}			3.876**	3.853**
			(1.526)	(1.567)
r_{out}			6.896***	6.757**
			(2.411)	(3.000)
Gender	0.589***	0.640***	0.589***	0.640***
	(0.199)	(0.203)	(0.199)	(0.203)
Experience	−0.601**	−0.530*	−0.601**	−0.530*
	(0.291)	(0.321)	(0.291)	(0.321)
Constant	0.012	−0.112	−2.572**	−2.681**
	(0.367)	(0.667)	(1.051)	(1.043)
Pseudo R^2	0.147	0.147	0.147	0.147
N	837	621	837	621

* The dependent variable is *Coop* (= 1 if one cooperated, = 0 if otherwise). *Gender* = 1 if male, = 0 if female. *Experience* = 1 if participant had work experience, = 0 if otherwise. Standard errors clustered by session are found in parentheses. The p-values reported here are two-tailed, with * denoting $p < 0.10$, ** denoting $p < 0.05$, and *** denoting $p < 0.01$.

Appendix 2. Further dynamic analysis

Taking all four games together, we have forty different decision situations with forty frequencies of contributions in the next round. The number of degrees of freedom for chi-square goodness-of-fit tests is forty minus the number of estimated parameters. Table A2 shows the empirical distribution of contributions in the forty decision situations.

Next, we consider other behavioral theories used in analyzing the repeated prisoner's dilemma. Dal Bó and Fréchette (2018) survey the performance of simple behavioral rules as Tit-for-Tat (*TfT*), *Punish*, Win-Stay-Lose-Shift (*WSLS*), "never contribute" (0), "always contribute" (*1*) and some other less often played rules. We extend these five rules by *Inertia* = "repeat your previous choice" and *TfToth* = "contribute if ***LRivalCoop*** = 2 "or, alternatively, ***LRivalCoop*** ≥ 1. In addition, we assume a "trembling hand" probability ε which describes random deviations from the rules (same for all rules). With these components we estimate population shares (*PopSh*, the share of *TfToth* is the residual share) for the behavioral rules. The result is presented in Table A3.

Table A2
Frequency distribution of decisions *Coop* = 0 or 1 dependent on the vector of previous decisions of self, partner, and rivals.

Game	LPartnerCoop	LRivalCoop	LCoop = 1 Coop = 1	 Coop = 0	LCoop = 0 Coop = 1	 Coop = 0
PD	0		9	26	16	76
PD	1		43	11	7	28
Info	0	0	2	8	3	17
Info	0	1	2	4	6	14
Info	0	2	5	9	11	21
Info	1	0	31	1	0	10
Info	1	1	24	4	2	4
Info	1	2	19	5	2	12
Rivalry	0	0	2	15	10	94
Rivalry	0	1	4	8	7	27
Rivalry	0	2	0	2	1	5
Rivalry	1	0	3	3	4	13
Rivalry	1	1	2	2	4	8
Rivalry	1	2	0	0	0	2
Synergy	0	0	4	17	9	27
Synergy	0	1	11	19	7	35
Synergy	0	2	2	3	2	4
Synergy	1	0	3	3	5	16
Synergy	1	1	7	3	11	19
Synergy	1	2	2	2	2	3

Table A3
Estimation of a mixture of behavioral rules. *PopSh* = population share*.

	1	*0*	*TfT*	*Punish*	*WSLS*	*Inertia*	*TfToth*	*ε*	*LogL*
PopSh., ε	0.012	0.044	0.133	0.535	0.049	0.167	0.060	0.134	443.79
PopSh., ε	–	–	0.099	0.708	–	0.138	0.055	0.160	443.84
S.D.	–	–	0.058	0.103	–	0.059	–	0.018	

* The share of *TfToth* is the residual share. Standard deviations in this and all following tables are computed from the inversion of the Hessian of the loglikelihood function. The Hessian of the seven-parameter model is almost singular.

Table A4
Estimated parameters and standard deviations for altruism and *FS* models with exogenous beliefs and constant coefficients of utility functions*.

	λ	β_B	α_B	β_{out}	α_{out}	η	δ	AIC
Alt	3.761	0.358		1		0	–	1059
*S.D.**	–	–		–		–	–	
Alt-LLA	2.926	−0.075		0.601		0.041	1.367	892.8
S.D.	1.017	0.204		0.170		0.089	0.403	
RDAhist	2.607	1.000	*0.461*	0.968	0.995	0.208	–	923.8
S.D.	0.402	0.015	*0.066*	0.200	0.223	0.092	–	
RDAexp	3.735	1.000	0.602	0.697	0.724	0.055	–	900.4
S.D.	0.504	0.010	0.035	0.141	0.145	0.086	–	
RDAhist-LLA	2.631	0.152	−0.005	0.705	0.732	0.099	1.183	893.4
S.D.	1.076	0.209	0.210	0.223	0.237	0.111	0.396	
RDAexp-LLA	4.520	−0.434	−0.217	0.417	0.422	0.017	1.223	896.2
S.D.	2.282	0.501	0.264	0.222	0.226	0.099	0.236	
FS	1.591	1.000	1.000	1.000	1.108	0.101	–	939.8
S.D.	0.164	0.007	0.011	0.011	0.718	0.118	–	
FS-LLA	3.312	−9.631	0.153	0.168	0.167	0.008	1.027	918.0
S.D	–	–	–	–	–	–	–	

* $\alpha_{out} = \alpha_C = \alpha_D$ and $\beta_{out} = \beta_C = \beta_D$. The Hessians of the loglikelihood functions of *Alt* and *FS*-LLA are singular. Models and parameters are defined in Section 2.2. We denote λ = precision parameter of QRE, α_B (β_B) = altruism parameter for A's ingroup partner B in the case he has less (at least as much) income than (as) A, α_{out} (β_{out}) = altruism parameter for A's out-group members C and D in the case they have less (at least as much) income than (as) A, η = probability of continuing to cooperate after the partner has defected, and δ = discount rate. For the *FS* models, the coefficients α and β refer to the difference between own and others' income.

Table A5
Estimated parameters and standard deviations for reciprocity models*.

	λ	β_B	α_B	θ_B	β_{out}	α_{out}	θ_{out}	δ	AIC
DynAlt	9.980	−0.182		−3.504	0.360		−0.137	–	517.7
*S.D.**	–	–		–	–		–	–	
DynFS=CR	17.21	0.506	0.287	0.169	2.526	0.894	0.000	–	467.6
S.D.	9.051	0.081	0.073	0.189	0.337	0.111	0.005	–	
DynAlt-LLA	1.575	0.069		1.696	1.096		0.525	0.443	480.6
S.D.	0.228	5.486		132.4	0.377		0.287	0.111	
CR-LLA	17.24	−0.507	0.287	0.168	2.527	0.894	0.000	0.000	467.6
S.D.	1.307	0.027	0.020	0.223	0.144	0.039	0.005	0.009	

* *DynAlt* with changing altruism coefficients and *DynFS* = *CR* with changing *FS* coefficients with endogenous beliefs (QRE). $\alpha_{out} = \alpha_C = \alpha_D$, $\beta_{out} = \beta_C = \beta_D$, and $\theta_{out} = \theta_C = \theta_D$. The Hessian of the loglikelihood function of *DynAlt* is singular. For the definition of the other parameters, see the note for Table A4. θ_B and θ_{out} define the rate of change for coefficients due to reciprocity (see Section 2.2).

References

Aaldering, H., Greer, L. L., Van Kleef, G. A., & De Dreu, C. K. (2013). Interest (mis) alignments in representative negotiations: Do pro-social agents fuel or reduce inter-group conflict? *Organizational Behavior and Human Decision Processes, 120*(2), 240–250.

Abbink, K., Brandts, J., Herrmann, B., & Orzen, H. (2012). Parochial altruism in inter-group conflicts. *Economics Letters, 117*(1), 45–48.

Anderson, S. P., Goeree, J. K., & Holt, C. A. (1998). A theoretical analysis of altruism and decision error in public goods games. *Journal of Public Economics, 70*(2), 297–323.

Andreoni, J. (1995). Cooperation in public-goods experiments: Kindness or confusion? *The American Economic Review*, 891–904.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature, 442*(7105), 912.

Birkinshaw, J. (2001). Strategies for managing internal competition. *California Management Review, 44*(1), 21–38.

Böhm, R., & Rockenbach, B. (2013). The inter-group comparison–intra-group cooperation hypothesis: Comparisons between groups increase efficiency in public goods provision. *PloS one, 8*(2), e56152.

Böhm, R., Rusch, H., & Gürerk, Ö. (2016). What makes people go to war? Defensive intentions motivate retaliatory and preemptive intergroup aggression. *Evolution and Human Behavior, 37*(1), 29–34.

Böhm, R., Rusch, H., & Baron, J. (2020). The psychology of intergroup conflict: A review of theories and measures. *Journal of Economic Behavior & Organization, 178*, 947–962.

Bornstein, G. (1992). The free-rider problem in intergroup conflicts over step-level and continuous public goods. *Journal of Personality and Social Psychology, 62*(4), 597.

Bornstein, G., & Ben-Yossef, M. (1994). Cooperation in intergroup and single-group social dilemmas. *Journal of Experimental Social Psychology, 30*(1), 52–67.

Bornstein, G., Winter, E., & Goren, H. (1996). Experimental study of repeated team-games. *European Journal of Political Economy, 12*(4), 629–639.

Bornstein, G., Gneezy, U., & Nagel, R. (2002). The effect of intergroup competition on group coordination: An experimental study. *Games and Economic Behavior, 41*(1), 1–25.

Breitmoser, Y., & Tan, J. H. W. (2013). Reference dependent altruism in demand bargaining. *Journal of Economic Behavior & Organization, 92*, 127–140.

Breitmoser, Y., & Tan, J. H. W. (2020). Why should majority voting be unfair? *Journal of Economic Behavior & Organization, 175*, 281–295.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117*(3), 817–869.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review, 99*(1), 431–457.
Choi, J. K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science, 318*(5850), 636–640 (New York, N.Y.).
Dal Bó, P., & Fréchette, G. R. (2018). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature, 56*(1), 60–114.
Embrey, M., Fréchette, G. R., & Yuksel, S. (2018). Cooperation in the finitely repeated prisoner's dilemma. *The Quarterly Journal of Economics, 133*(1), 509–551.
Engelmann, D., & Strobel, M. (2002). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Reply. *American Economic Review, 96* (5), 1918–1923.
Fallucchi, F., Renner, E., & Sefton, M. (2013). Information feedback and contest structure in rent-seeking games. *European Economic Review, 64*, 223–240.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*(3), 817–868.
Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140.
Fiala, L., & Suetens, S. (2017). Transparency and cooperation in repeated dilemma games: A meta study. *Experimental Economics, 20*(4), 755–771.
Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.
Frechette, G. R., & Yuksel, S. (2017). Infinitely repeated games in the laboratory: Four perspectives on discounting and random termination. *Experimental Economics, 20*, 279–308.
Fudenberg, D., & Maskin, E. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica : Journal of the Econometric Society, 54*(3), 533–554.
Gimeno, J. (2004). Competition within and between networks: The contingent effect of competitive embeddedness on alliance formation. *Academy of Management Journal, 47*(6), 820–842.
Greiner, B. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. *University of Cologne, Working Paper Series in Economics, 10*(23), 63–104.
Halevy, N., Bornstein, G., & Sagiv, L. (2008). In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm. *Psychological Science, 19*(4), 405–411.
Johnson, E. J., Camerer, C., Sen, S., & Rymon, T. (2002). Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory, 104*(1), 16–47.
Ke, C., Konrad, K. A., & Morath, F. (2015). Alliances in the shadow of conflict. *Economic Inquiry, 53*(2), 854–871.
Keser, C., & Van Winden, F. (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics, 102*(1), 23–39.
Kritikos, A., & Bolle, F. (2001). Distributional concerns: Equity-or efficiency-oriented? *Economics Letters, 73*(3), 333–338.
Konrad, K. A., & Morath, F. (2020). Escalation in conflict games: On beliefs and selection. *Experimental Economics, 23*, 750–787.
Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory, 27*(2), 245–252.
Leibbrandt, A., & Sääksvuori, L. (2012). Communication in intergroup conflicts. *European Economic Review, 56*(6), 1136–1147.
McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior, 10*(1), 6–38.
McKelvey, R. D., & Palfrey, T. R. (1998). Quantal response equilibria for extensive form games. *Experimental Economics, 1*(1), 9–41.
Nalbantian, H. R., & Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review, 87*(3), 314–341.
Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., et al. (2016). Parochial prosocial religions: Historical and contemporary evidence for a cultural evolutionary process. *Behavioral and Brain Sciences*, 39.
Reuben, E., & Tyran, J. R. (2010). Everyone is a winner: Promoting cooperation through all-can-win intergroup competition. *European Journal of Political Economy, 26*(1), 25–35.
Rusch, H. (2014a). The evolutionary interplay of intergroup conflict and altruism in humans: A review of parochial altruism theory and prospects for its extension. *Proceedings of the Royal Society B: Biological Sciences, 281*(1794), Article 20141539.
Rusch, H. (2014b). The two sides of warfare: An extended model of altruistic behavior in ancestral human intergroup conflict. *Human Nature, 25*, 359–377.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.
Sheremeta, R. M. (2018). Behavior in group contests: A review of experimental research. *Journal of Economic Surveys, 32*(3), 683–704.
Suls, J.M. & Miller, R.L. (1977). Social comparison processes: Theoretical and empirical perspectives. Hemisphere.
Stahl, D. O., & Haruvy, E. (2006). Other-regarding preferences: Egalitarian warm glow, empathy, and group size. *Journal of Economic Behavior & Organization, 61*(1), 20–41.
Tan, J. H. W., & Bolle, F. (2007). Team competition and the public goods game. *Economics Letters, 96*(1), 133–139.
Tan, J. H. W., & Zizzo, D. J. (2008). Groups, cooperation and conflict in games. *The Journal of Socio-Economics, 37*(1), 1–17.
Tan, J.H.W., and Bolle, F., (2019). Intergroup conflict aversion weakens intragroup cooperation. Discussion Paper Report No. 2019/04, Economic Growth Centre, Nanyang Technological University.