Evolution and Human Behavior xxx (xxxx) xxx

Contents lists available at ScienceDirect

Evolution and Human Behavior

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

Parochial reciprocity

Simon Columbus [a,*], Isabel Thielmann [b], Ingo Zettler [a], Robert Böhm [a,c]

[a] Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 København, Denmark
[b] Max Planck Institute for the Study of Crime, Security and Law, Günterstalstraße 73, 79100 Freiburg, Germany
[c] Department of Occupational, Economic, and Social Psychology, University of Vienna, Universitätsstraße 7, 1010 Wien, Austria

ARTICLE INFO

Keywords:
Intergroup conflict
IPD-MD
Parochial altruism
Reciprocity
Social preferences

ABSTRACT

Parochial altruism suggests that humans are intrinsically motivated to harm out-groups, and that this is tightly connected to a preference for benefitting their in-group. Yet, there is little evidence for the kind of unconditional out-group harm suggested by this account, nor for the assertion that it would be associated with in-group cooperation. Instead, humans selectively reciprocate actual, but also potential aggression. We therefore posit a model of parochial reciprocity, according to which individuals retaliate against actual and anticipated harms to their in-group. To test predictions arising from these competing accounts, we manipulated out-group threats and elicited preferences for the welfare of in-group and out-group members, as well as beliefs about in-group and out-group members' behaviours in an incentivised intergroup conflict game with natural groups (online sample; $N = 973$). In this game, individuals could pay to benefit their in-group, but had the option to additionally harm the out-group without incurring any further costs. Individuals who valued their in-group more strongly were no more likely to harm the out-group, thus contradicting parochial altruism. Instead, individuals who expected the out-group to harm their in-group preemptively retaliated the anticipated attack. Importantly, they only did so when the out-group posed an actual threat to the in-group. Taken together, the findings suggest that participation in intergroup conflict is better explained by parochial reciprocity than purely by group-based preferences.

Intergroup conflict is ubiquitous in human history and prehistory, and indeed even in non-human primates (Blattman, 2022; Kelly, 2005; Pinker, 2011; Wilson et al., 2014; Wrangham & Glowacki, 2012). Individual participation in intergroup conflict, however, poses an evolutionary puzzle. Many intergroup conflicts form a multilayered social dilemma (Bornstein, 2003): conflict between groups is destructive, and thus, participation is collectively inefficient, although it may provide benefits to one's in-group. At the same time, defending one's in-group or attacking an out-group is individually costly. Therefore, in the absence of additional rewards or punishments, each individual faces strong incentives to freeride on the efforts of their in-group.

Yet, humans do participate in intergroup conflicts, just as they cooperate with members of their in-group (Balliet, Wu, & De Dreu, 2014). One interpretation is that aggression towards out-group members and cooperation with in-group members are truly costly to the individual. This poses the question how such costly behaviours could have arisen evolutionarily. To answer this question, a number of models suggest that costly out-group aggression and costly in-group cooperation could have co-evolved by mutually reinforcing each other (Bowles, 2009; Choi & Bowles, 2007; García & van den Bergh, 2011).

One corollary of these models is what has been termed *parochial altruism*: preferences over the welfare of in-groups and out-groups which dispose humans to cooperate with in-group members and to harm out-group members (Choi & Bowles, 2007; García & van den Bergh, 2011; Rusch, 2014). These two preferences are thought to be linked due to their shared evolutionary past, such that intragroup cooperation and intergroup aggression go hand-in-hand (Rusch, 2014; Yamagishi & Mifune, 2016). Crucially, neither behaviour is thought to require additional, individual-level incentives (Rusch, 2014). Here, we first review the empirical evidence (a) on the linkage between group-based preferences over the welfare of in-group and out-groups, and (b) on the prevalence of unconditional out-group aggression in humans. Then, we turn to an alternative explanation of individual participation in intergroup conflict, which we term *parochial reciprocity*.

1. Empirical evidence against parochial altruism

Models of parochial altruism suggest that preferences over the welfare of in-groups and out-groups are closely connected (Choi & Bowles, 2007; García & van den Bergh, 2011; Rusch, 2014; Yamagishi & Mifune,

* Corresponding author.
E-mail address: simon@simoncolumbus.com (S. Columbus).

https://doi.org/10.1016/j.evolhumbehav.2023.02.001
Received 2 May 2022; Received in revised form 13 December 2022; Accepted 3 February 2023

2016). This has been interpreted in two ways: On the one hand, individuals who value the welfare of their in-group more strongly, and are consequently more willing to cooperate with in-group members, may be more negatively disposed towards, and more willing to harm, out-group members. On the other hand, differences may exist at the group level, such that groups which are more cooperative internally may also be more aggressive towards other groups.

At the individual level, empirical evidence shows that group-based preferences are positively correlated: Individuals who cooperate more with their in-group are more, not less, cooperative towards out-group members (Brewer, 2007; Chowdhury, Mukherjee, & Sheremeta, 2021; Corr, Hargreaves Heap, Seger, & Tsutsui, 2015; De Dreu, 2010a; Halevy, Bornstein, & Sagiv, 2008; Müller, 2019; Rusch, 2014; Yamagishi & Mifune, 2016). This is also supported by the limited evidence on the relationship between intragroup cooperation and individual aggression against out-groups (Abbink, Brandts, Herrmann, & Orzen, 2012; Thielmann & Böhm, 2016). A similar pattern holds at the group level: A group's cooperativeness or internal coherence does not imply hostility or competitiveness towards other groups (Cashdan, 2001). Thus, there is little evidence that 'parochial altruism' describes the pattern of human preferences for the well-being of in-groups and out-groups.

Models of parochial altruism further imply that individuals will harm out-groups when there is no personal incentive to do so (Choi & Bowles, 2007; García & van den Bergh, 2011; Rusch, 2014). Yet, there is scant evidence that humans will engage in this kind of unconditional aggression against out-groups. Instead, when given the choice, most individuals choose to cooperate selectively with their in-group, but will forego the opportunity to harm the out-group (Böhm, Halevy, & Kugler, 2022; Halevy et al., 2008). This holds true when it costs no additional resources to harm the out-group (Halevy et al., 2008) and even between natural groups characterised by a history of conflict (Halevy, Weisel, & Bornstein, 2012) and strong enmity (Weisel & Böhm, 2015). The purely preference-based account of parochial altruism thus fails to explain individual participation in intergroup conflict.

2. Conditional aggression against out-groups

Several factors can explain why humans may not harm out-groups unconditionally. Aggression against peaceful out-groups can cause highly destructive spirals of retaliatory violence (Beckerman et al., 2009; Benard, Berg, & Mize, 2017; Gat, 2000; Gould, 2000; Walker & Bailey, 2013). Consequently, in-group members may discourage and suppress such indiscriminate attacks (Gould, 2000; Greif, 2006). For example, Beckerman et al. (2009) write that among the Waorani of Ecuador, "the major reputational effect of high participation in raiding was a reluctance of others to live with the fiercest warriors because of the danger of return raids." Thus, individuals who attack out-groups—whether to avenge personal grievances or for private gains—not only bear the risks inherent in fighting abroad, but also the costs of a bad reputation at home.

When the in-group is threatened, however, aggression against out-groups may be rewarded and even expected by other members of the in-group (Nawata & Yamaguchi, 2013; Thrasher & Handfield, 2018). Under threat, individuals willingly attack out-groups (Böhm, Rusch, & Baron, 2020; de Dreu, 2010b; Simunovic, Mifune, & Yamagishi, 2013; Stephan & Stephan, 2020) and vicariously retaliate against harm to in-group members (Lickel, Miller, Stenstrom, Denson, & Schmader, 2006). Importantly, they will even do so preemptively (Böhm, Rusch, & Gürerk, 2016). Conversely, removing the threat by neutering the out-group's ability to harm the in-group drastically reduces aggression against the out-group (Böhm et al., 2016). This suggests that out-group harm may be gratuitous, but not unconditional, and that it is motivated by (anticipated) harm to one's in-group.

Preemptive aggression against threatening out-groups may be promoted because it can garner the in-group a reputation for toughness. A strong showing against a potential attacker may signal to the out-group that members of the group are able and willing to defend against and retaliate attacks (Benard et al., 2017; Delton & Krasnow, 2017; Gould, 1999; Thrasher & Handfield, 2018). Such a group reputation for toughness can deter future aggression (Crescenzi, 2007; Walter, 2006). Individuals who attack threatening out-groups thus provide a public good to their in-group (Thrasher & Handfield, 2018).

3. Ultimate explanations of parochial reciprocity

In conflicts between individuals, a personal reputation for toughness has direct fitness benefits for the individual because it can deter future attacks. Individuals with a reputation for meekness, in contrast, may become targets for exploitation (Gambetta, 2009; McCullough, Kurzban, & Tabak, 2013). Therefore, the individual willingness to fight in order to establish or maintain a tough personal reputation can be selected for, at least as long as individuals stand to lose a lot in fights (McElreath, 2003). Selection for reputation maintenance as a way of preempting attacks can explain why some social ecologies exhibit a culture of honour, in which individuals escalate with aggression in response to perceived sleights or threats (Nisbett & Cohen, 1996; Thrasher & Handfield, 2018).

Fighting for the reputation of one's in-group is, however, prima facie costly to the individual and thus requires a separate explanation. One possibility is that groups with a reputation for toughness become larger or wealthier, and that individuals who fight on behalf of their in-group themselves or their kin benefit from being part of a larger or wealthier group. Group augmentation theory shows that intragroup cooperation can be selected for when individuals benefit from enlarging their group because additional members contribute to their fitness (Kokko, Johnstone, & Clutton-Brock, 2001). When group members are interdependent, intergroup aggression with the goal of maintaining or increasing the size of the in-group may similarly be fitness enhancing. Indeed, evidence suggests that group augmentation can explain why chimpanzees participate in patrols to defend their group's territory (Langergraber, Watts, Vigilant, Mitani, & Cheney, 2017). Similarly, humans may seek to enhance their group's reputation for toughness because their fitness depends on the well-being or number of their peers.

Individuals who enhance their in-group's reputation for toughness and protect it from threatening out-groups may also be rewarded directly. Those who attack out-groups on behalf of their in-group may garner a reputation for being both tough and prosocial (Nawata, 2020). Consequently, both indirect reciprocity and reputation-based partner choice can support the evolution of out-group aggression (Roberts et al., 2021; Rusch, Leunissen, & Van Vugt, 2015). However, reputation-based mechanisms should be sensitive to whether aggression benefits the in-group (e.g., by fostering a reputation for toughness among threatening out-groups) or harms the in-group (e.g., by engendering counterattacks). We therefore propose that aggression against out-groups should be sensitive to whether the out-group poses a threat to the in-group. In other terms, humans should engage in what we term *parochial reciprocity*, attacking threatening out-groups to protect the in-group, but refraining from harming out-groups indiscriminately.

4. The present research

The following hypotheses were preregistered.[1] Parochial altruism suggests that aggression against out-groups is largely driven by group-based preferences: humans value the welfare of in-group members and devalue the welfare of out-group members (Choi & Bowles, 2007; Rusch, 2014). In contrast, our account of parochial reciprocity suggests that individuals harm out-group members specifically when they perceive their in-group to be threatened by the out-group, and thus privileges

[1] Here, the hypotheses are ordered as they appear in the text, which differs from their numbering in the preregistration. The corresponding numbers are provided in the supplementary information.

beliefs about the behaviour of out-group members over group-based preferences. We test this in an experimental game in which participants could peacefully cooperate with their in-group or (additionally) harm the out-group. We experimentally manipulated whether the out-group's behaviour could actually harm the in-group (THREAT condition) or not (NO THREAT condition) and directly elicited preferences and beliefs about the behaviour of both in-group and out-group members. We hypothesise that individuals are less willing to harm the out-group when the out-group is deprived of the possibility to harm the in-group (*H1*). Finally, we test whether removing actual threat completely eliminates out-group harm (as in Böhm et al., 2016), or whether individuals still harm the out-group in the NO THREAT condition, which would indicate aggression not based on perceived threats (*H2*).

We elicited beliefs about the behaviour of the out-group and the in-group. Following our account of parochial reciprocity, we expect that individuals who anticipate greater harm from the out-group will engage in preemptive retaliation and harm the out-group (parochial reciprocity; *H3*). Importantly, individuals who believe that the out-group will attempt to harm their in-group should only act on these beliefs and harm the out-group when the out-group's behaviour poses an actual threat to the in-group (in the THREAT condition) but not when it is ineffectual (in the NO THREAT condition) (*H4*). In addition, our account posits that the willingness to harm out-groups depends on social norms in the in-group which support such attacks. Although we did not directly elicit normative expectations (i.e., beliefs about what in-group members think the individual ought to do), social norms may be reflected in empirical expectations (i.e., beliefs about the behaviour of other in-group members Bicchieri, 2005). We therefore hypothesise that individuals who expect other in-group members to harm the out-group would be more willing to do so as well (*H5*). However, this influence of empirical expectations could also reflect conditional cooperation with in-group members (Falk & Fischbacher, 2006; Fischbacher & Gächter, 2010; Fischbacher, Gächter, & Fehr, 2001).

In contrast to the predictions suggested by parochial altruism, we hypothesise that individuals who value their in-group members more positively will not engage in more out-group harm (*H6*). We do, however, expect that individuals who value the out-group less will engage in more out-group harm (*H7*). Thus, people might harm out-groups because they dislike them, but not merely because they value the welfare of their in-group.

Finally, we are also able to examine why individuals may cooperate with their in-group while foregoing the opportunity to harm the out-group. We motivate the relevant hypotheses and report and discuss the results of these analyses in the supplementary information.

5. Methods

5.1. Ethics and open science statement

This study received ethical approval from the institutional review board of the Department of Psychology, University of Copenhagen, Denmark (#21062021). All participants provided informed consent and the study used no deception.

This study was preregistered, including information about all hypotheses, how the sample size was determined, all data exclusions, all manipulations, and all measures in the study. The preregistration is available here: https://osf.io/mkezp/?view_only=c70008c87dc2480d9e4819601337d8ee. All materials, data, and analysis code are available on the Open Science Framework: https://osf.io/f74be/?view_only=04607cfb8c474f909224ec052d7fd073.

5.2. Sample

Because we sought to study the role of both group-based preferences and beliefs, we draw on natural groups whose members hold significant negative attitudes towards each other: US American supporters of the Democratic and Republican parties (Iyengar, Lelkes, Levendusky, Malhotra, & Westwood, 2019). We aimed to recruit 960 Democrats via Prolific (http://prolific.co), using the built-in screening questions at Prolific.[2] The sample size was based on financial constraints. Because the study took place over two consecutive days, there was some drop-out after the first session ($n = 27$). As preregistered, we excluded participants who indicated a party affiliation other than "Democrat" in the intake survey and recruited additional participants until reaching 960 completed responses to the second part of the study (before performance-based exclusions; see below).

Overall, $N = 983$ participants provided complete responses to both parts of the study. Ten participants failed more than one comprehension question about the intergroup conflict game after two attempts and were excluded as preregistered. The final sample thus consisted of $N = 973$ participants (479 female, 450 male, 7 no information; $M_{age} = 34.11$, $SD_{age} = 11.34$ years).

Participants were paid a flat fee of US$2.80 for an estimated 20 min of effort, plus a potential bonus payment based on their own and other participants' decisions in the study (M = US$1.51, SD = US$3.27). Because incentives for beliefs and decisions in the experimental game required an out-group, we recruited an additional 96 participants who self-identified as Republicans to make these out-group decisions. These participants also acted as the out-group recipients of the decisions made by the Democrat participants. Because the Republicans' knowledge that their decision would be implemented at random only 50% of the time (see experimental manipulation below) may have affected the decisions of these participants, we did not include their data in the preregistered analyses. These participants therefore did not complete the survey part of the study and were paid US$1.40 for an estimated 10 min of effort, plus a decision-based bonus. For details on the payoff calculation, see each task below.

5.3. Experimental setup

5.3.1. Intergroup Game

We study intergroup conflict using the Intergroup Prisoner's Dilemma-Maximising Difference (IPD-MD) game (Halevy et al., 2008). The IPD-MD is regularly employed to study the motivations underlying individual participation in intergroup conflict (for a review, see Böhm et al., 2022). A key advantage of the IPD-MD is that it separates out-group harm from intragroup cooperation: individuals can behave selfishly, benefit their in-group at a cost to themselves, or benefit their in-group while additionally harming the out-group. Thus, individuals who are motivated to benefit their in-group are not forced to harm the out-group, and out-group harm is purely gratuitous (but individually costly).

Participants played a one-shot IPD-MD game with two groups of $n = 3$ players each. Each player received an endowment of $E = 100$ MU. They could contribute any integer amount of the endowment to one of three pools: a private account, in-group cooperation, and out-group harm. Per 1 MU, contributions to the private account yielded 1 MU to self and 0 MU to others; in-group cooperation yielded 0.5 MU to each in-group member (including self); out-group harm yielded 0.5 MU to each in-group member (including self) and − 0.5 MU to each out-group member. Thus, freeriding (i.e., contributing to the private account) increased an individual's personal outcome, irrespective of the other players' contributions; conversely, both in-group cooperation and out-

[2] We initially preregistered to recruit an equal number of Democrat and Republican participants. After recruiting 50 participants of each party affiliation, it became apparent that fulfilling the Republican quota would likely be infeasible given Prolific's user base. We therefore amended the preregistration to recruit only Democrat participants for the main treatment. The original preregistration is available here: https://osf.io/v8j4g/?view_only=4f4744ee1ae04a888fd28e6d6be4d8e6.

group harm maximised the in-group's welfare. Note that we use 'out-group harm' for brevity to describe behaviour that both benefitted the in-group and harmed the out-group.

To avoid negative payoffs, which may occur if out-group members invest more in out-group harm than in-group members invest in in-group cooperation, each participant also received a non-investable base pay of 100 MU. To determine behaviour-contingent payoffs, we randomly selected 96 participants (i.e., 10%) to be matched to the sample of Republicans (the out-group) and have their decisions paid out. Earned MU were paid out at a rate of 1 MU = US$0.05 to selected participants. Selected participants earned US$10.85 (*SD*= US$1.31) on average.

5.3.2. Experimental manipulation

We manipulated whether the out-group posed an actual threat to the in-group. Specifically, participants were informed that the out-group made a decision with the knowledge that their intended harm to participants' in-group would only be implemented in 50% of cases, and that this would be randomly determined. Participants then made their own decision using the strategy method (Selten, 1967), that is, once for the case that the out-group's intended harm to participants' in-group would be implemented and could harm the in-group (THREAT condition), and once for the case that it would not be implemented and could not harm the in-group (NO THREAT condition).

5.4. Measures

5.4.1. Preference measures

We assessed preferences for the welfare of in-group and out-group members separately, using the six primary items of the Social Value Orientation (SVO) slider measure (Murphy, Ackermann, & Handgraaf, 2011). The SVO slider measure assesses social preferences using six trade-offs between one's own and another's welfare. The six items can be summarised as an angle expressing the relative valuation of own and others' outcomes, with higher values indicating more prosocial preferences (see Murphy et al., 2011). We computed one index each for preferences for the welfare of in-group members and for preferences for the welfare of out-group members. Stated preferences for in-group and out-group members' welfare may be affected by the expectation that another in−/out-group member has a choice to benefit or harm the decision-maker (Misch, Paulus, & Dunham, 2021; Rabbie, Schot, & Visser, 1989; Yamagishi, Jin, & Kiyonari, 1999), essentially reflecting reciprocity concerns. To reduce such influences, we explicitly informed participants that both an in-group and an out-group member would make a decision about the participant's outcomes. In addition, we emphasised that participants' decisions could not affect the decisions of any other participant, but only others' outcomes.

We incentivised the SVO slider measure; for each participant, one slider item was randomly selected for payoff at a rate of 100 points = US $0.10. Each participant was also matched to another participant as a beneficiary of the other's decision and paid out at the same rate (actual earnings across both roles, M = US$0.15, SD = US$0.02). Because we sampled only Democrat participants, we matched their decisions to those of 50 Republican participants recruited before the change in sampling plan (see above). For the Republican participants, one decision each as decision-maker and as recipient was paid out at random.

Most participants provided fully transitive responses (93% for Democrat targets, 92% for Republican targets; Bakker & Dijkstra, 2021). Transitivity means that if a participant preferred outcome A over outcome B, and outcome B over outcome C, they also preferred outcome A over outcome C. Democrat participants expressed more positive preferences towards Democrats, SVO angle = 0.47, than towards Republicans, SVO angle = 0.31, paired-samples t-test, $t(972) = 18.55$, $p < .001$, Cohen's $d_{av} = .60$. However, on average, both preferences were in the prosocial range, meaning that participants positively valued the welfare of both in-group and out-group members.

5.4.2. Belief elicitation

We elicited beliefs about the behaviour of in-group and out-group members using an adapted version of the box arrangement task (Fragiadakis, Kovaliukaite, & Rojo Arjona, 2023). For each group, participants were presented with three sliders labelled 'green', 'orange', and 'purple', representing contributions to the private account, in-group cooperation, and out-group harm. Each slider ranged from 0 to 100 points. Moving a slider to the right by one point indicated the belief that a random member of the target group would contribute 1 MU to the respective pool. The participant could distribute a total of 100 points across all three sliders. They received one lottery ticket for each point on the slider that overlapped with the true contributions of the respective group, and no ticket otherwise. For example, if the participant indicated a belief that in-group members would, on average, contribute 20 MU to out-group harm, and in-group members actually contributed 15 MU to out-group harm, the overlap was 15 MU, and the participant received 15 lottery tickets from this belief. The total number of tickets received was q. The corresponding lottery paid US$0.35 with probability $q/100$ and US$0.07 with probability $(100 - q)/100$. One of the two belief elicitation tasks (for beliefs about in-group and out-group members) was paid out at random to each participant (actual earnings, M = US$0.29, SD = US$0.11).

5.4.3. Other measures

We included a number of survey measures for future exploratory analyses: the Honesty-Humility and Agreeableness subscales of the HEXACO-100 (Lee & Ashton, 2018; Thielmann et al., 2020), a 16-item measure of the Dark Factor of Personality (Moshagen, Zettler, & Hilbig, 2020), and the 21 relevance items of the Morality as Cooperation Questionnaire (Curry, Chesters, & Van Lissa, 2019). Basic demographic information was obtained from questions previously answered by participants and provided by Prolific.

5.5. Procedure

Participants were invited to participate in a two-part study on two subsequent days. During the first part, they completed the SVO slider measures and other measures not pertinent to the current investigation (see above). We randomised the order of elicitation of in-group and out-group preferences in the SVO slider measure. On the second day, participants played the IPD-MD game, starting with detailed instructions and comprehension questions. We used three comprehension questions with three answer options each. Participants had two attempts to pass each question; if they failed a comprehension question, they were informed about the correct response. Participants who failed more than two questions after two attempts were excluded from the analyses. Subsequently, participants indicated their beliefs about the behaviour of in-group and out-group members, in counterbalanced order. Finally, participants decided on their contributions to the different pools. At this point, participants were informed that the out-group's decision would be implemented with a probability of 0.5. Each participant made one set of decisions in the THREAT condition and a second set of decisions in the NO THREAT condition. The order of these conditions was counterbalanced. A few days after the study, we randomly selected participants to have their bonus paid out and matched them with an out-group constructed from Republican participants (see above). All bonuses were paid out anonymously through Prolific.

5.6. Software

We analysed the data using R and the tidyverse (R Core Team, 2021; Wickham et al., 2019). Linear mixed models were fitted using the packages lmer and lmerTest (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017). R^2 statistics for linear mixed models were computed using r2mlm (Rights & Sterba, 2019; Shaw, Rights, Sterba, & Flake, 2020). Interaction plots were generated

using interaction (Long, 2019). Average adjusted predictions were computed using marginaleffects (Arel-Bundock, 2022).

6. Results

6.1. Descriptive Statistics

On average, participants kept 44% ($SD = 24\%$) of their endowment for themselves (contributions to the private account). They spent 19% ($SD = 20\%$) to harm the out-group and 36% ($SD = 24\%$) to cooperate with the in-group. Participants spent more of their endowment to harm the out-group when the out-group posed a genuine threat to their in-group (22%) than when it did not (17%). Conversely, they spent less to solely cooperate with the in-group when it was threatened by the out-group (34%) than when it was not (38%). Total contributions across in-group cooperation and out-group harm (which still benefitted the in-group) did not depend on the threat level. Zero-order correlations among preferences, beliefs, and contributions are shown in Fig. 1. Notably, preferences for the welfare of in-group and out-group members showed a strong positive correlation ($r = .52$, $p < .001$). Thus, the more individuals valued the welfare of fellow in-group members, the more they also valued the welfare of out-group members.

6.2. Confirmatory analyses

We used a linear mixed-effects model to test the role of threat, beliefs, and preferences in predicting out-group harm (Table 1). We regressed individual out-group harm on the threat treatment, beliefs about the degree to which both in-group and out-group members cooperated with their in-group and harmed their out-group, as well as in-group and out-group preferences (i.e., SVO). We also included the second-order interactions between the threat treatment and all other variables. All predictors were standardised. The threat treatment was contrast-coded; therefore, main effects are interpretable as in a purely additive model. To account for the repeated responses due to the within-participant design, the model included a random participant factor (i.e., random intercept). In addition to the regression estimates, we report average adjusted predictions (AAP) at one standard deviation below (-1SD) and above (+1SD) the mean of the predictor.[3]

The threat manipulation strongly affected individual aggression against the out-group. As predicted, participants imposed much less harm on the out-group in the NO THREAT condition, $M = 16.6$, $SD = 18.2$, than in the THREAT condition, $M = 22.5$, $SD = 21.4$, $B = -5.86$, $SE = 0.64$, $t(966) = -9.20$, $p < .001$ (*H1*). However, subsetting to the NO THREAT condition and repeating the analysis without the interaction terms showed that even when the out-group posed no threat, participants still imposed significant unexplained harm on them, $B_{intercept} = 16.61$, $SE = 0.52$, $t(966) = 32.09$, $p < .001$, Table S2 (*H2*).

Overall, individuals who expected more severe harm from the out-group also imposed more harm on the out-group, $B = 1.71$, $SE = 0.61$, $t(966) = 2.82$, $p = .005$ (*H3*). Specifically, an individual who believed that the out-group would contribute 8.39 points towards harming the individual's in-group (one standard deviation below the average belief) was predicted to contribute 17.83 points towards harming their out-group, whereas an individual who believed that the out-group would contribute 43.21 points towards harming the individual's in-group (one standard deviation above the average belief) was predicted to contribute 21.25 points towards harming the out-group.

Importantly, as expected, the role of beliefs about the out-group's decision to harm the in-group varied across the THREAT and NO THREAT conditions, $B = -1.67$, $SE = 0.81$, $t(966) = -2.06$, $p = .040$. To further examine this interaction, we subset the data to each treatment and regressed out-group harm on the measured beliefs and preferences (Table S2). In the THREAT condition, participants were more likely to harm the out-group when they believed out-group members would do the same, $B = 2.55$, $SE = 0.79$, $t(966) = 3.20$, $p = .001$, AAP: -1SD = 19.93, +1SD = 25.02. In contrast, this was not the case in the NO THREAT condition, $B = 0.88$, $SE = 0.66$, $t(966) = 1.34$, $p = .181$, AAP: -1SD = 15.73, +1SD = 17.49, Fig. 2 (*H4*).

As hypothesised, individuals who believed that other in-group members would harm the out-group were more likely to do so themselves, $B = 6.75$, $SE = 0.72$, $t(1683.23) = 9.32$, $p < .001$ (*H5*). Specifically, an individual who expected other in-group members to contribute 6.72 points towards harming the out-group (-1SD) was predicted to contribute 12.52 points, whereas an individual who expected others to contribute 32.67 points (+1SD) was predicted to contribute 26.56 points. Neither this nor any further effects were moderated by the threat treatment.

Finally, supporting our predictions about the role of group-based preferences, individuals who valued the in-group's welfare more did not harm the out-group more strongly, $B = .67$, $SE = 0.57$, $t(966) = 1.18$, $p = .240$, AAP: -1SD = 18.87, +1SD = 20.22 (*H6*), whereas individuals who valued the out-group's welfare less were more willing to harm the out-group, $B = -2.79$, $SE = 0.57$, $t(966) = -4.91$, $p < .001$, AAP: -1SD = 22.33, +1SD = 16.67 (*H7*).

6.3. Exploratory analyses

To test whether beliefs or preferences were stronger predictors of out-group harm, we fitted two models including (a) only the four measures of beliefs or (b) only the two measures of preferences, as well as all interactions with the threat treatment. We computed $R_1^{2(f)}$ estimates of the variance explained by the fixed effects in each model and compared the models using a log-likelihood ratio test. This showed that beliefs, $R_1^{2(f)} = .19$, explained more variance in out-group harm than did preferences, $R_1^{2(f)} = .05$, $\chi^2(4) = 26.23$, $p < .001$.

7. Discussion

When and why are people willing to harm the members of other groups? Parochial altruism posits that participation in intergroup conflict is driven by co-evolved preferences for the welfare of in-group members and for harm to out-groups (Choi & Bowles, 2007; Rusch, 2014). We contrast this account with a model of *parochial reciprocity*, according to which individuals attack out-groups which they perceive to threaten their in-group. Such conditional out-group harm may be supported by norms which reward those who maintain their in-group's reputation for toughness (Nawata, 2020; Thrasher & Handfield, 2018). By experimentally varying actual threat emanating from an out-group and measuring both preferences and beliefs in natural groups, we found that out-group harm in an experimental intergroup conflict game was driven to a significant degree by beliefs about the behaviour of out-group members. By contrast, preferences over the welfare of out-group and in-group members played a relatively minor role.

7.1. Out-group threat and beliefs in intergroup conflict

Participants were far more willing to harm the out-group when the out-group posed an actual threat to the in-group than when it did not. Under threat, individuals who expected more harm from the out-group spent more resources to preemptively attack the out-group. Taking away the out-group's threat capacity, while leaving intact their ability to express their ill intentions, reduced the level of out-group harm. However, in contrast to previous findings by Böhm et al. (2016), removing the threat did not completely eliminate out-group harm. One key difference is that Böhm et al. (2016) removed the out-group's decision entirely, whereas in our NO THREAT condition, the out-group made a decision which

[3] In the preregistration, we did not specify the use of average adjusted predictions; however, their inclusion does not change the interpretation of the interaction effects.

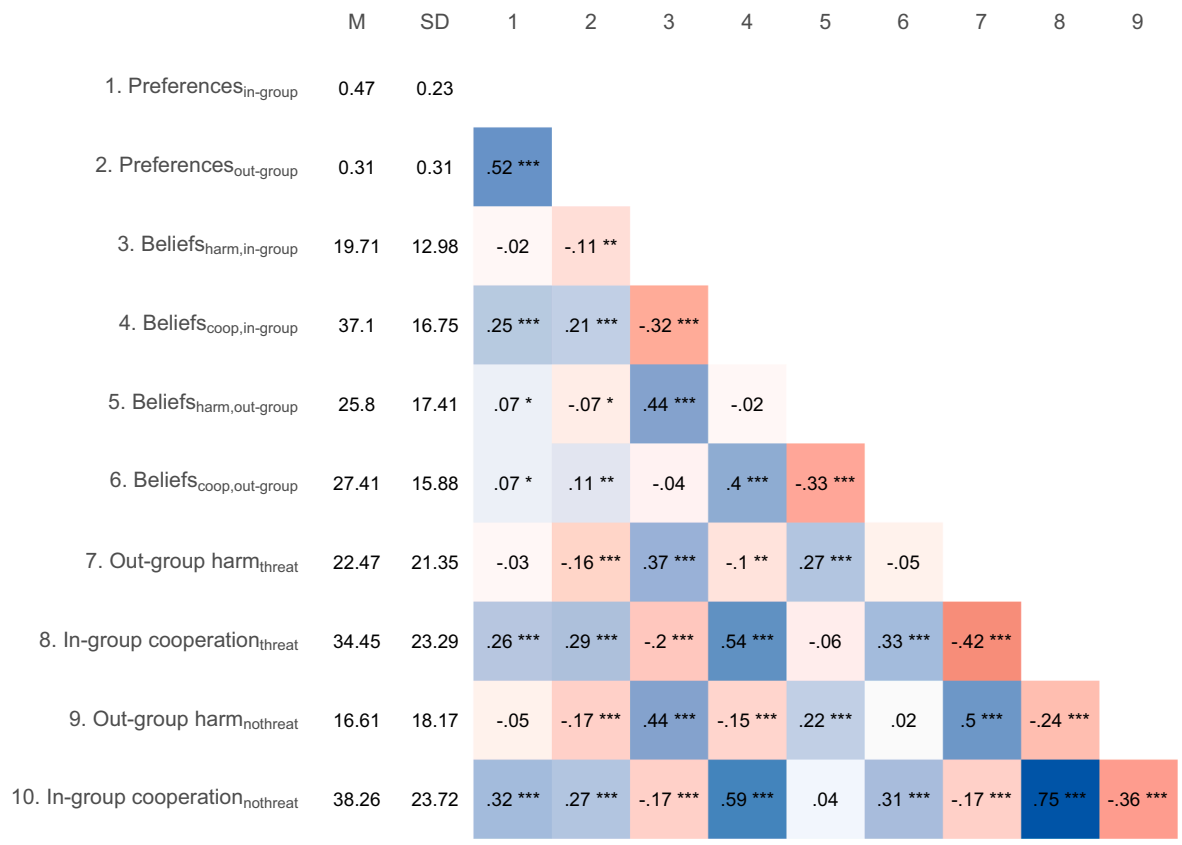

Fig. 1. Means, standard deviations, and zero-order correlations among preferences, beliefs, and contributions in the IPD-MD game.
Note: * $p < .05$, ** $p < .01$, *** $p < .001$. All p-values adjusted using the Bonferroni-Holm procedure.
Note: harm = out-group harm; coop = in-group cooperation.

Table 1
Preferences, beliefs, and threat manipulation as predictors of out-group harm. 'No threat' indicates the contrast-coded treatment variable.

	B	SE	df	*t*	*p*
(Intercept)	19.54	0.48	966	40.88	<0.001
$Preferences_{in-group}$	0.67	0.57	966	1.18	0.240
$Preferences_{out-group}$	−2.79	0.57	966	−4.91	<0.001
$Beliefs_{harm,\ in-group}$	7.02	0.60	966	11.65	<0.001
$Beliefs_{coop,\ in-group}$	0.14	0.61	966	0.23	0.821
$Beliefs_{harm,\ out-group}$	1.71	0.61	966	2.82	0.005
$Beliefs_{coop,\ out-group}$	0.64	0.59	966	1.08	0.279
No threat	−5.86	0.64	966	−9.20	<0.001
No threat×$Preferences_{in-group}$	−0.05	0.76	966	−0.06	0.951
No threat×$Preferences_{out-group}$	0.22	0.76	966	0.29	0.773
No threat×$Belief_{harm,\ in-group}$	0.54	0.80	966	0.67	0.505
No threat×$Belief_{coop,\ in-group}$	−1.04	0.81	966	−1.29	0.198
No threat×$Belief_{harm,\ out-group}$	−1.67	0.81	966	−2.06	0.040
No threat×$Belief_{coop,\ out-group}$	1.32	0.78	966	1.69	0.092

Note: harm = out-group harm; coop = in-group cooperation.

was subsequently rendered ineffectual. In contrast to Böhm et al. (2016), our study thus isolates the effect of the actual threat posed to the in-group.

Out-group harm was even more strongly predicted by the belief that other members of the in-group would engage in out-group harm. This may simply reflect a preference to conform with empirical expectations (Bicchieri, 2005) or to reciprocate the behaviour of in-group members (Falk & Fischbacher, 2006; Fischbacher et al., 2001). However, our theoretical account suggests that individuals may be willing to engage in preemptive or retaliatory aggression against out-groups in part because this is rewarded by other in-group members. The first-order beliefs we elicited may thus be indicative of the influence of normative expectations: individuals may harm out-groups insofar as they believe that others expect them to do so, and will reward them for providing the public good of a group reputation for toughness (Thrasher & Handfield, 2018).

There are several parts of our theoretical account which are not tested in the experiment. In particular, because groups only met in a single one-shot interaction, they could not gain a group reputation for toughness, nor could individual members reap the reputational benefits of having contributed to their group's reputation. Arguably, omitting the functional mechanisms which should drive the role of beliefs in inter-group conflict made it less likely to observe the predicted pattern of behaviour. Our results may thus represent conservatives estimates of the role of beliefs in out-group harm in the IPD-MD. However, future studies may examine whether, as predicted, the ability to form individual and group reputations across interactions exacerbates the initial willingness to attack threatening out-groups. In addition, such studies may explore other responses to out-group threat, such as increased intragroup cooperation to shore up the defensive potential of the in-group (Gould, 1999, 2000; Yamagishi & Mifune, 2009).

7.2. Group-based preferences and intergroup conflict

Parochial altruism suggests that individuals who value their in-group are more willing to harm out-groups. In contrast to this proposition, in our study, individuals who expressed more positive preferences for their in-group's welfare were no more likely to harm the out-group. Indeed, the more willing an individual was to incur a cost to benefit an in-group member (a fellow Democrat), the more they were willing to do the same to benefit an out-group member (a Republican) (see also Thielmann & Böhm, 2016). This also contradicts a basic tenet of social identity theory, according to which positive preferences for the in-group's welfare should

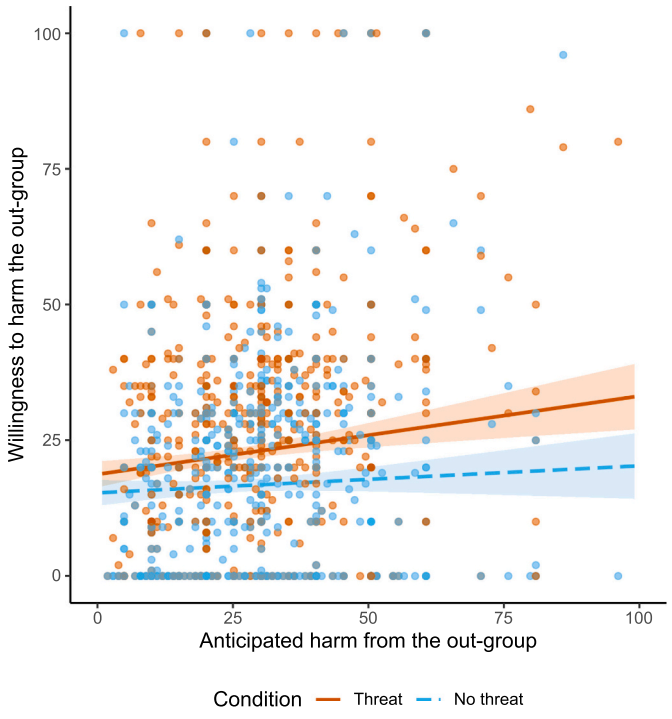

Fig. 2. The relationship between expectations of harm from the out-group and willingness to harm the out-group when the out-group could actually harm the in-group (THREAT condition) and when it could not harm the in-group (NO THREAT condition). The figure shows conditional effects based on the mixed-effects model predicting *out-group harm. Orange and blue dots represent individual responses under the two conditions.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

be tightly connected to derogation of the out-group (Tajfel, Billig, Bundy, & Flament, 1971; Yamagishi & Mifune, 2016). In addition, participants were more likely to simply cooperate with their in-group than to additionally harm the out-group, even though they incurred no additional cost in doing so (see also Böhm et al., 2022; Halevy et al., 2008). This contrast was even more pronounced when the out-group was unable to harm the in-group (see also Böhm et al., 2016). Thus, most participants did not appear to use out-group harm to improve the relative standing of their in-group.

It is worth noting, however, that preferences and beliefs together only explained about 18% of the variance in out-group harm in the THREAT condition (22% in the NO THREAT condition). Thus, we found significant out-group harm in the IPD-MD which could neither be explained by group-based preferences nor by beliefs. We also explored whether including interactions between preferences and beliefs increased the proportion of variance explained, but these only added very limited information (see supplementary information). One explanation for the unexplained variance in behaviour could be additional individual differences such as broad personality traits which are not captured by group-specific preferences or beliefs (Thielmann & Böhm, 2016; Thielmann, Spadaro, & Balliet, 2020). However, there may also be measurement error in the outcome variable (i.e., contributions to out-group harm in the IPD-MD). Yet, estimates may even be inflated due to the game-based assessment of preferences, beliefs, and behaviour (i.e., common method variance). Further research may explore the role of individual differences beyond group-based preferences and beliefs in intergroup conflict, for instance, traits associated with general prosociality (e.g., Honesty-Humility; Aaldering & Böhm, 2020; Ashton, Lee, & de Vries, 2014; Thielmann, Spadaro, & Balliet, 2020), retaliation (e.g., Agreeableness; Ashton et al., 2014; Thielmann, Spadaro, & Balliet, 2020), and preferences for hierarchies between groups (e.g., Social Dominance Orientation; Aaldering & Böhm, 2020; Halali, Dorfman, Jun, & Halevy, 2018; Pratto, Sidanius, Stallworth, & Malle, 1994).

One limitation of our study is that we focused on the inclinations to peacefully benefit the in-group and to harm the out-group in a behavioral game in which participation in intergroup conflict was not aligned with self-interest. Yet, groups may encounter each other in a variety of different situations (Bornstein, 2003; Doğan, Glowacki, & Rusch, 2018; Lopez, 2017). Consequently, intergroup behaviour may differ substantially by the structure of the conflict situation (Aaldering & Böhm, 2020; de Dreu et al., 2016; Halevy et al., 2008). For example, in asymmetric attacker-defender conflicts, contributing to the group's efforts in intergroup conflict can be self-serving if others do so as well (as an attacker) or if they fail to do so (as a defender; Bornstein, 2003; de Dreu & Gross, 2019; Méder, de Dreu, & Gross, 2022). In such a situation, participation in intergroup conflict may be driven more by considerations of individual costs and benefits and less by (preemptive) retaliation. At the same time, individuals may rely even more on their beliefs about the behaviour of in-group and out-group members when this affects their strategic position (de Dreu et al., 2016; de Dreu & Gross, 2019). Similarly, where a group reputation for toughness matters, it may be achieved by other means than attacks against the out-group, such as internal enforcement of in-group norms or displays of military prowess (e.g., in parades). Future research may thus explore the role of beliefs and group-based preferences in different kinds of intergroup conflicts, as well as the role and maintenance of group reputations by different means.

8. Conclusion

History abounds with examples of wars started in the belief of preempting a coming attack. This suggests that when people harm out-groups, they often do so to preemptively retaliate anticipated harms to their in-group. Here, we investigated the influence of beliefs and group-based preferences on participation in intergroup conflict using an experimental game which allows individuals to benefit in-group members and to harm out-group members. By measuring beliefs about the behaviour of in-group and out-group members as well as preferences for the welfare of in-group and out-group members, we found that out-group harm was motivated in large parts by expectations of harm coming from the out-group and the belief that other in-group members would harm the out-group. Conversely, when the out-group posed no actual threat, participants were far less willing to harm its members. This contrasts with the idea that intergroup conflict is primarily a matter of preferences for the welfare of in-group and out-group members. Overall, preemptive retaliation and beliefs about the behaviour of other in-group members played a stronger role than group-based preferences in motivating behaviour in this intergroup context. Our findings support a model of parochial reciprocity, according to which humans are motivated to preempt and retaliate attacks against their in-group rather than to unconditionally harm out-groups.

Author note

Simon Columbus, Department of Psychology, University of Copenhagen, simon@simoncolumbus.com, https://orcid.org/0000-0003-1546-955X; Isabel Thielmann, Department of Criminology, Max Planck Institute for the Study of Crime, Security and Law, https://orcid.org/0000-0002-9071-5709; Ingo Zettler, Department of Psychology and Copenhagen Center for Social Data Science, University of Copenhagen, https://orcid.org/0000-0001-6140-7160; Robert Böhm, Faculty of Psychology, University of Vienna, and Department of Psychology and Copenhagen Center for Social Data Science, University of Copenhagen, https://orcid.org/0000-0001-6806-0374.

The authors thank the editor, two anonymous reviewers, and the audience at the Copenhagen Network for Experimental Economics in Lund for valuable feedback on earlier versions of this paper.

Declaration of Competing Interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2023.02.001.

References

Aaldering, H., & Böhm, R. (2020). Parochial versus universal cooperation: Introducing a novel economic game of within- and between-group interaction. *Social Psychological and Personality Science, 11*(1), 36–45. https://doi.org/10.1177/1948550619841627

Abbink, K., Brandts, J., Herrmann, B., & Orzen, H. (2012). Parochial altruism in inter-group conflicts. *Economics Letters, 117*(1), 45–48. https://doi.org/10.1016/j.econlet.2012.04.083

Arel-Bundock, V. (2022). marginaleffects: Marginal effects, marginal means, predictions, and contrasts [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=marginaleffects.

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18*(2), 139–152. https://doi.org/10.1177/1088868314523838

Bakker, D. M., & Dijkstra, J. (2021). Comparing the slider measure of social value orientation with its main alternatives. *Social Psychology Quarterly, 84*(3), 1–11. https://doi.org/10.1177/01902725211008938

Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin, 140*(6), 1556–1581. https://doi.org/10.1037/a0037737

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

Beckerman, S., Erickson, P. I., Yost, J., Regalado, J., Jaramillo, L., Sparks, C., … K.. (2009). Life histories, blood revenge, and reproductive success among the waorani of Ecuador. *Proceedings of the National Academy of Sciences, 106*(20), 8134–8139. https://doi.org/10.1073/pnas.0901431106

Benard, S., Berg, M. T., & Mize, T. D. (2017). Does aggression deter or invite reciprocal behavior? Considering coercive capacity. *Social Psychology Quarterly, 80*(4), 310–329. https://doi.org/10.1177/0190272517728904

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge, UK: Cambridge University Press.

Blattman, C. (2022). *Why we fight: The roots of war and the paths to peace*. Penguin.

Böhm, R., Halevy, N., & Kugler, T. (2022). The power of defaults in intergroup conflict. *Organizational Behavior and Human Decision Processes, 168*, Article 104105. https://doi.org/10.1016/j.obhdp.2021.104105

Böhm, R., Rusch, H., & Baron, J. (2020). The psychology of intergroup conflict: A review of theories and measures. *Journal of Economic Behavior and Organization, 178*, 947–962. https://doi.org/10.1016/j.jebo.2018.01.020

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*, 29–34. https://doi.org/10.1016/j.evolhumbehav.2015.06.005

Bornstein, G. (2003). Intergroup conflict: Individual, group, and collective interests. *Personality and Social Psychology Review, 7*(2), 129–145. https://doi.org/10.1207/S15327957PSPR0702_129-145

Bowles, S. (2009). Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science, 324*(5932), 1293–1298. https://doi.org/10.1126/science.1168112

Brewer, M. B. (2007). The importance of being we: Human nature and intergroup relations. *American Psychologist, 62*(8), 728–738. https://doi.org/10.1037/0003-066X.62.8.728

Cashdan, E. (2001). Ethnocentrism and xenophobia: A cross-cultural study. *Current Anthropology, 42*(5), 760–764. https://doi.org/10.1086/323821

Choi, J.-K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science, 318*(5850), 636–640. https://doi.org/10.1126/science.1144237

Chowdhury, S., Mukherjee, A., & Sheremeta, R. (2021). In-group versus out-group preferences in intergroup conflict: An experiment (MPRA Working Paper No. 105690) Retrieved from. . https://mpra.ub.uni-muenchen.de/105690/.

Corr, P. J., Hargreaves Heap, S. P., Seger, C. R., & Tsutsui, K. (2015). An experiment on individual 'parochial altruism'revealing no connection between individual 'altruism'and individual 'parochialism'. *Frontiers in Psychology, 6*, 1261. https://doi.org/10.3389/fpsyg.2015.01261

Crescenzi, M. J. (2007). Reputation and interstate conflict. *American Journal of Political Science, 51*(2), 382–396. https://doi.org/10.1111/j.1540-5907.2007.00257.x

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality, 78*, 106–124. https://doi.org/10.1016/j.jrp.2018.10.008

De Dreu, C. K. W. (2010a). Social value orientation moderates ingroup love but not outgroup hate in competitive intergroup conflict. *Group Processes & Intergroup Relations, 13*(6), 701–713. https://doi.org/10.1177/1368430210377332

Delton, A. W., & Krasnow, M. M. (2017). The psychology of deterrence explains why group membership matters for third-party punishment. *Evolution and Human Behavior, 38*(6), 734–743. https://doi.org/10.1016/j.evolhumbehav.2017.07.003

Doğan, G., Glowacki, L., & Rusch, H. (2018). Spoils division rules shape aggression between natural groups. *Nature Human Behaviour, 2*(5), 322–326. https://doi.org/10.1038/s41562-018-0338-z

de Dreu, C. K. W. (2010b). Social conflict: The emergence and consequences of struggle and negotiation. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Vol. 2. Handbook of social psychology* (pp. 983–1023). Wiley.

de Dreu, C. K. W., & Gross, J. (2019). Revisiting the form and function of conflict: Neurobiological, psychological, and cultural mechanisms for attack and defense within and between groups. *Behavioral and Brain Sciences, 42*(E116). https://doi.org/10.1017/S0140525X18002170

de Dreu, C. K. W., Gross, J., Méder, Z. Z., Giffin, M., Prochazkova, E., Krikeb, J., & Columbus, S. (2016). In-group defense, out-group aggression, and coordination failures in intergroup conflict. *Proceedings of the National Academy of Sciences of the United States of America, 201605115*. https://doi.org/10.1073/pnas.1605115113

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*(2), 293–315. https://doi.org/10.1016/j.geb.2005.03.001

Fischbacher, U., & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review, 100*(1), 541–556. https://doi.org/10.1257/aer.100.1.541

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

Fragiadakis, D. E., Kovaliukaite, A., & Rojo Arjona, D. (2023). *Does an individual predict opponents according to cognitive hierarchy?* (under review).

Gambetta, D. (2009). *Codes of the underworld*. Princeton University Press.

García, J., & van den Bergh, J. C. J. M. (2011). Evolution of parochial altruism by multilevel selection. *Evolution and Human Behavior, 32*(4), 277–287. https://doi.org/10.1016/j.evolhumbehav.2010.07.007

Gat, A. (2000). The human motivational complex: Evolutionary theory and the causes of hunter-gatherer fighting, part II. Proximate, subordinate, and derivative causes. *Anthropological Quarterly*, 74–88.

Gould, R. V. (1999). Collective violence and group solidarity: Evidence from a feuding society. *American Sociological Review, 64*(3), 356–380. https://doi.org/10.2307/2657491

Gould, R. V. (2000). Revenge as sanction and solidarity display: An analysis of vendettas in nineteenth-century Corsica. *American Sociological Review, 65*(5), 682–704. https://doi.org/10.2307/2657542

Greif, A. (2006). History lessons: The birth of impersonal exchange: The community responsibility system and impartial justice. *Journal of Economic Perspectives, 20*(2), 221–236. https://doi.org/10.1257/jep.20.2.221

Halali, E., Dorfman, A., Jun, S., & Halevy, N. (2018). More for us or more for me? social dominance as parochial egoism. *Social Psychological and Personality Science, 9*(2), 254–262. https://doi.org/10.1177/1948550617732819

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. https://doi.org/10.1111/j.1467-9280.2008.02100.x

Halevy, N., Weisel, O., & Bornstein, G. (2012). "In-group love" and "out-group hate" in repeated interaction between groups. *Journal of Behavioral Decision Making, 25*(2), 188–195. https://doi.org/10.1002/bdm.726

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science, 22*, 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Kelly, R. C. (2005). The evolution of lethal intergroup violence. *Proceedings of the National Academy of Sciences, 102*(43), 15294–15298. https://doi.org/10.1073/pnas.0505955102

Kokko, H., Johnstone, R. A., & Clutton-Brock, T. H. (2001). The evolution of cooperative breeding through group augmentation. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 268*(1463), 187–196. https://doi.org/10.1098/rspb.2000.1349

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13

Langergraber, K. E., Watts, D. P., Vigilant, L., Mitani, J. C., & Cheney, D. L. (2017). Group augmentation, collective action, and territorial boundary patrols by male chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America, 114*, 7337–7342. https://doi.org/10.1073/pnas.1701582114

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. *Assessment, 25*(5), 543–556. https://doi.org/10.1177/1073191116659134

Lickel, B., Miller, N., Stenstrom, D. M., Denson, T. F., & Schmader, T. (2006). Vicarious retribution: The role of collective blame in intergroup aggression. *Personality and Social Psychology Review, 10*(4), 372–390. https://doi.org/10.1207/s15327957pspr1004_6

Long, J. A. (2019). *Interactions: Comprehensive, user-friendly toolkit for probing interactions*.

Lopez, A. C. (2017). The evolutionary psychology of war: Offense and defense in the adapted mind. *Evolutionary Psychology, 15*(4), 1474704917742720.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences, 36*(1), 1–15. https://doi.org/10.1017/S0140525X11002160

McElreath, R. (2003). Reputation and the evolution of conflict. *Journal of Theoretical Biology, 220*(3), 345–357. https://doi.org/10.1006/jtbi.2003.3166

Méder, Z. Z., de Dreu, C. K. W., & Gross, J. (2022). *Equilibria of attacker-defender games*. https://doi.org/10.48550/arXiv.2202.10072

Misch, A., Paulus, M., & Dunham, Y. (2021). Anticipation of future cooperation eliminates minimal ingroup bias in children and adults. *Journal of Experimental Psychology: General, 49*(89), 2036–2056. https://doi.org/10.1037/xge0000899

Moshagen, M., Zettler, I., & Hilbig, B. E. (2020). Measuring the dark core of personality. *Psychological Assessment, 32*(2), 182–196. https://doi.org/10.1037/pas0000778

Müller, D. (2019). The anatomy of distributional preferences with group identity. *Journal of Economic Behavior and Organization, 166*, 785–807. https://doi.org/10.1016/j.jebo.2019.09.009

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making, 6*(8), 771–781.

Nawata, K. (2020). A glorious warrior in war: Cross-cultural evidence of honor culture, social rewards for warriors, and intergroup conflict. *Group Processes & Intergroup Relations, 23*(4), 598–611. https://doi.org/10.1177/1368430219838615

Nawata, K., & Yamaguchi, H. (2013). Intergroup retaliation and intra-group praise gain: The effect of expected cooperation from the in-group on intergroup vicarious retribution. *Asian Journal of Social Psychology, 16*(4), 279–285. https://doi.org/10.1111/ajsp.12032

Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the south*. Westview.

Pinker, S. (2011). *The better angels of our nature: The decline of violence in history and its causes*. Penguin.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67*(4), 741–763. https://doi.org/10.1037/0022-3514.67.4.741

R Core Team. (2021). *R*. R Foundation for Statistical Computing.

Rabbie, J. M., Schot, J. C., & Visser, L. (1989). Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model. *European Journal of Social Psychology, 19*, 171–202. https://doi.org/10.1002/ejsp.2420190302

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods, 24*(3), 309–338. https://doi.org/10.1037/met0000184

Roberts, G., Raihani, N., Bshary, R., Manrique, H. M., Farina, A., Samu, F., & Barclay, P. (2021). The benefits of being seen to help others: Indirect reciprocity and reputation-based partner choice. *Philosophical Transactions of the Royal Society B, 376*(1838), 20200290. https://doi.org/10.1098/rstb.2020.0290

Rusch, H. (2014). 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*, 20141539. https://doi.org/10.1098/rspb.2014.1539

Rusch, H., Leunissen, J. M., & Van Vugt, M. (2015). Historical and experimental evidence of sexual selection for war heroism. *Evolution and Human Behavior, 36*(5), 367–373. https://doi.org/10.1016/j.evolhumbehav.2015.02.005

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In H. Sauermann (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung* (pp. 136–168). J. C. B. Mohr.

Shaw, M., Rights, J. D., Sterba, S. K., & Flake, J. K. (2020). *r2mlm: R-squared measures for multilevel models*. https://doi.org/10.31234/osf.io/xc4sv

Simunovic, D., Mifune, N., & Yamagishi, T. (2013). Preemptive strike: An experimental study of fear-based aggression. *Journal of Experimental Social Psychology, 49*(6), 1120–1123. https://doi.org/10.1016/j.jesp.2013.08.003

Stephan, W. G., & Stephan, C. W. (2020). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–46). Erlbaum. https://doi.org/10.4324/9781410605634-7.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Thielmann, I., Akrami, N., Babarović, T., Belloch, A., Bergh, R., Chirumbolo, A., … others. (2020). The hexaco–100 across 16 languages: A large-scale test of measurement invariance. *Journal of Personality Assessment, 102*(5), 714–726. https://doi.org/10.1080/00223891.2019.1614011

Thielmann, I., & Böhm, R. (2016). Who does (not) participate in intergroup conflict? *Social Psychological and Personality Science, 7*(8), 778–787. https://doi.org/10.1177/1948550616660160

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin, 146*(1), 30–90. https://doi.org/10.1037/bul0000217

Thrasher, J., & Handfield, T. (2018). Honor and violence. *Human Nature, 29*(4), 371–389. https://doi.org/10.1007/s12110-018-9324-4

Walker, R. S., & Bailey, D. H. (2013). Body counts in lowland south american violence. *Evolution and Human Behavior, 34*(1), 29–34. https://doi.org/10.1016/j.evolhumbehav.2012.08.003

Walter, B. F. (2006). Building reputation: Why governments fight some separatists but not others. *American Journal of Political Science, 50*(2), 313–330. https://doi.org/10.1111/j.1540-5907.2006.00186.x

Weisel, O., & Böhm, R. (2015). "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *Journal of Experimental Social Psychology, 60*, 110–120. https://doi.org/10.1016/j.jesp.2015.04.008

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., , … Yutani, & H.. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilson, M. L., Boesch, C., Fruth, B., Furuichi, T., Gilby, I. C., Hashimoto, C., … others. (2014). Lethal aggression in *Pan* is better explained by adaptive strategies than human impacts. *Nature, 513*(7518), 414–417. https://doi.org/10.1038/nature13727

Wrangham, R. W., & Glowacki, L. (2012). Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers. *Human Nature, 23*(1), 5–29. https://doi.org/10.1007/s12110-012-9132-1

Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: Ingroup boasting and ingroup favouritism. *Advances in Group Processes, 16*, 161–197. https://doi.org/10.2307/2695887

Yamagishi, T., & Mifune, N. (2009). Social exchange and solidarity: In-group love or out-group hate? *Evolution and Human Behavior, 30*(4), 229–237. https://doi.org/10.1016/j.evolhumbehav.2009.02.004

Yamagishi, T., & Mifune, N. (2016). Parochial altruism: Does it explain modern human group psychology? *Current Opinion in Psychology, 7*, 39–43. https://doi.org/10.1016/j.copsyc.2015.07.015