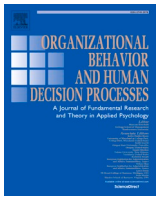
Organizational Behavior and Human Decision Processes 171 (2022) 104152

Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

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



Shadow of conflict: How past conflict influences group cooperation and the use of punishment☆

Jörg Gross [a,*], Carsten K.W. De Dreu [a,b], Lennart Reddmann [a]

[a] Social, Economic and Organizational Psychology, Leiden University, The Netherlands
[b] Center for Experimental Economics and Political Decision Making, University of Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Conflict
Conflict resolution
Cooperation
Behavioral spill-over
Punishment

ABSTRACT

Intergroup conflict profoundly affects the welfare of groups and can deteriorate intergroup relations long after the conflict is over. Here, we experimentally investigate how the experience of an intergroup conflict influences the ability of groups to establish cooperation after conflict. We induced conflict by using a repeated attacker-defender game in which groups of four are divided into two 'attackers' that can invest resources to take away resources from the other two participants in the role of 'defenders.' After the conflict, groups engaged in a repeated public goods game with peer-punishment, in which group members could invest resources to benefit the group and punish other group members for their decisions. Previous conflict did not significantly reduce group cooperation compared to a control treatment in which groups did not experience the intergroup conflict. However, when having experienced an intergroup conflict, individuals punished free-riding during the repeated public goods game less harshly and did not react to punishment by previous attackers, ultimately reducing group welfare. This result reveals an important boundary condition for peer punishment institutions. Peer punishment is less able to efficiently promote cooperation amid a 'shadow of conflict.' In a third treatment, we tested whether such 'maladaptive' punishment patterns induced by previous conflict can be mitigated by hiding the group members' conflict roles during the subsequent public goods provision game. We find more cooperation when individuals could not identify each other as (previous) attackers and defenders and maladaptive punishment patterns disappeared. Results suggest that intergroup conflict undermines past perpetrators' legitimacy to enforce cooperation norms. More generally, results reveal that past conflict can reduce the effectiveness of institutions for managing the commons.

1. Introduction

Intergroup relations can transition from competition and conflict to mutually beneficial exchange and cooperation (Bar-Tal, 2000; Beekman, Cheung, & Levely, 2017; De Dreu, Gross, Fariña, & Ma, 2020; Gross & De Dreu, 2019; Sherif, Harvey, White, Hood, & Sherif, 1961). Political parties compete for votes before an election but have to work together and form coalitions thereafter. Managers from competing firms need to work together following a hostile take-over or merger. And groups that fought each other in a civil war are faced with the problem of how to (re-) establish cooperative relationships and reunite. Transitioning from conflict to cooperation can be difficult. A history of intergroup conflict (a 'shadow of conflict') can lead to prejudices and tensions that perpetuate long after the conflict is settled (Bar-Tal, 2000; Bar-Tal & Halperin, 2011; Cilliers et al., 2016; Gat, 2019; Ross & Ward, 1995; Rouhana & Bar-Tal, 1998). Enduring spite, revenge motives, or mistrust, as well as many other psychological barriers between former conflicting parties can undermine cooperation and reduce social welfare (Bar-Tal, 2000; Bar-Tal & Halperin, 2011; Cilliers et al., 2016; Ross & Ward, 1995).

Yet, even without a history of conflict, group cooperation is difficult to establish and maintain in the first place. Group members can be tempted to free-ride on the cooperation of others or fear that others may free-ride on their cooperative efforts. This inherent free-rider problem can quickly crowd out cooperation even without prior conflict, as many laboratory studies have shown (Burton-Chellew et al., 2015; Fehr & Fischbacher, 2004). Previous research has highlighted that effective

☆ This article is part of the special issue 'Experimental studies of Conflict; Edited by Dr. Julia Minson, Dr. Corinne Bendersky, Dr. Taya Cohen, Dr. Carsten de Dreu, Dr. Eran Halperin and Dr. Juliana Schroeder'.

* Corresponding author at: Institute of Psychology, Leiden University, PO Box 9555, 2300 RA Leiden, The Netherlands.
E-mail address: mail@joerg-gross.net (J. Gross).

https://doi.org/10.1016/j.obhdp.2022.104152
Received 15 June 2021; Received in revised form 4 April 2022; Accepted 5 April 2022
Available online 18 May 2022
0749-5978/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

institutions need to be in place to prevent cooperation breakdown (Fehr & Williams, 2018; Gürerk et al., 2006; Henrich, 2006). In particular, peer punishment institutions have received much attention in the cooperation literature. Punishment can sustain cooperation in groups when it is primarily used by cooperators to sanction free-riders (Fehr & Gächter, 2000, 2002; Gross et al., 2016; Masclet et al., 2003; Yamagishi, 1986). Previous research, however, also showed that punishment is sometimes misused, for example, when free-riders punish cooperators (Herrmann et al., 2008). Such anti-social punishment can undermine the effectiveness of the punishment institution. Likewise, Engelmann and Nikiforakis (2015; Nikiforakis, 2008) showed that punishment can lead to punishment feuds that decrease group welfare. In short, although peer punishment usually prevents the breakdown of cooperation, it can also lose its effectiveness.

Here, we conjecture that a history of conflict within groups can undermine group cooperation and give rise to misuse or underuse of peer punishment. This is because people may use punishment opportunities to 'get back' at previous enemies rather than to deter free-riding. Evidence for this possibility would, first of all, show that previous intergroup conflict makes it more difficult for groups to establish cooperative relationships. It would further reveal an important boundary condition of peer punishment in fostering group cooperation. If peer punishment loses its effectiveness to promote cooperation in the shadow of a previous intergroup conflict, it would qualify previous findings on the circumstances under which punishment fails to foster group cooperation.

We examined these possibilities in a laboratory experiment with four-person groups that engaged in two tasks – an intergroup conflict and a public goods provision task. In our experiment, intergroup conflict was modeled as an asymmetric, step-level contest in which (individuals in) one group invests resources to exploit another group. In turn, the other group can invest resources to defend itself against such out-group aggression (Chowdhury & Topolyan, 2016; Clark & Konrad, 2007; De Dreu & Gross, 2019; Duffy & Kim, 2005). Examples of such intergroup attack-defense contests include those between revisionist states claiming their neighbors' territory (Wright, 2014), companies engaging in hostile take-over of their competitors (De Dreu et al., 2016), political parties challenging the status quo that is defended by the incumbents (De Dreu, Pliskin, Rojek-Giffin, Méder, & Gross, 2021), and tribal raiders and cattle herders (Glowacki et al., 2016). We chose an asymmetric conflict setting for three interrelated reasons. First, asymmetric conflicts divide its members into (former) perpetrators and victims and create a history of exploitation and inequality. This in itself can reduce cooperation (Anderson, Mellor, & Milyo, 2008; Fung & Au, 2014; Gross & Böhm, 2020; Gross, Veistola, De Dreu, & Dijk, 2020; Martinangeli & Martinsson, 2020). Second, many intergroup conflicts, like civil wars or hostile take-over of competitors, have an asymmetric conflict structure followed by an opportunity for groups to reconcile and cooperate with each other. Third, the division of groups into former perpetrators and victims may give rise to different motivations for cooperation and the counterproductive use of punishment, for example, by punishing former perpetrators not for free-riding but for their previous role in the conflict.

Group cooperation in our experiment is investigated by using a standard public goods dilemma with peer punishment. Public good provision creates a social dilemma for individuals, since cooperation maximizes group welfare while defection maximizes personal welfare. This means that individuals have an incentive to cooperate but also to 'free-ride' on the cooperative efforts of other group members (Fehr & Gächter, 2002; Gross et al., 2016; Yamagishi, 1986). In our experiment, peer punishment was induced by giving individuals the option to deduct earnings from other group members at a personal cost to themselves after each investment round.

Our *first pre-registered hypothesis* is that experiencing an asymmetric conflict within the group (vs. not) decreases subsequent cooperation. This hypothesis builds on previous work showing how previous interactions within and between groups can influence future interactions (also referred to as 'spill-over' effects in the literature). However, earlier work mostly examined spill-over of cooperative interactions. For example, Peysakhovich and Rand (2016) exposed participants to a repeated prisoner's dilemma game and manipulated the length of the interaction and the temptation to free-ride. When the temptation to free-ride was relatively low, participants displayed more cooperation and trust in a subsequent and independent follow-up interaction (for similar results, see Cassar et al., 2013; Iacono & Somnez, 2020; and Stagnaro et al., 2017). Similarly, Cason et al. (2012) documented spill-over effects when groups were faced with a coordination problem, with more successful coordination when groups first experienced a coordination problem that is easier to solve (the median effort game) before transitioning to a more difficult coordination problem (the minimum effort game). Finally, several studies showed that more successful coordination in a weakest-link game leads individuals to exhibit higher cooperativeness in a subsequent, unrelated prisoner's dilemma game (Knez & Camerer, 2000, see also Ahn et al., 2001; Brandts & Cooper, 2006).

A few experimental studies have looked at conflict spill-over effects, reporting mixed findings. For example, Ke and colleagues (2013) investigated whether conflict effort between two individuals changes when the prize at stake was (vs. not) jointly created in a previous conflict (i.e., competing for the 'spoils' of conflict). They found no significant spill-over effects: Former conflict-allies competed as much as individuals without a history of conflict. A similar result was obtained by Halevy and colleagues (2012). In their study, individuals first contributed to public goods that would impose a negative externality on the out-group and then transitioned to an environment in which they could (also) contribute without hurting the out-group. Group members quickly switched to 'peaceful' cooperation and this was independent of their history of intergroup conflict. In contrast to these results showing little evidence for conflict spill-over, Cason and Gangadharan (2013) showed that a competitive trading environment decreased cooperation in a subsequent task in which former competitors could work together towards a joint goal. Likewise, Beekman and colleagues (2017) found that individuals exposed to a symmetric conflict subsequently increased parochial cooperation that exclusively benefitted their in-group and reduced universal cooperation that could have benefitted both in- and out-group. These findings and our hypothesis also relate to the so-called 'cutthroat cooperation effect' (see Beersma et al., 2009; Johnson et al., 2006) according to which it is more difficult for groups to transition from a competitive to a cooperative environment than the other way around.

To our knowledge, there is no work that has investigated spill-over effects of asymmetric conflict on cooperation and punishment. As mentioned above, punishment usually increases group cooperation but can also lose its effectiveness to do so in some circumstances. This can be the case when punishment is motivated by, for example, spite or revenge (Bar-Tal, 2000; Bar-Tal & Halperin, 2011; Ross & Ward, 1995). In the shadow of a conflict, punishment could be aimed at previous perpetrators or victims as a means of retaliation. Importantly, this can only happen when previous conflict roles are identifiable. In real-world conflicts, this is the case when members of conflicting parties are distinguishable in terms of surface-level characteristics like language or ethnicity that are difficult to mask or change (Harrison et al., 1998; Van Knippenberg et al., 2004). Oftentimes, however, conflicting parties are separated by political affiliation, religion, or world-views – features that can be masked or adapted (see also Goldenberg et al., 2018; Halperin et al., 2011). For example, González & Clémence (2019) studied former paramilitary group members of Columbia's civil war. They found that many previous perpetrators chose to conceal their past role and group affiliation in their attempts to reintegrate into civil society. This is important, because only when past perpetrators and victims can be identified, individuals can use punishment for other purposes than to enforce cooperation norms. When past memberships are hidden and individuals can no longer identify past perpetrators and victims, peer punishment may regain its effectiveness in inducing and enforcing cooperation norms. Accordingly, our *second pre-registered hypothesis* is

that not being able to identify past perpetrators or victims (vs. being able to) can help to (re-)establish cooperation after a conflict.

Both of our hypotheses rest on the assumption that the asymmetric conflict impacts subsequent group cooperation partly through 'maladaptive' punishment patterns. According to the needs-based model of reconciliation by Shnabel & Nadler (see Nadler & Shnabel, 2015; Shnabel & Nadler, 2008, 2015), conflicts threaten the status, sense of agency, and power of victims whereas they threaten the moral self-image of previous perpetrators (for a general overview of the psychological barriers that impede reconciliation after conflict beyond revenge or punishment use, see also Bar-Tal, 2000; Bar-Tal & Halperin, 2011; Ross & Ward, 1995; Baumeister et al., 1990; Bilali & Vollhardt, 2019). Thus, asymmetric conflicts create different needs depending on the role in the previous conflict that can hinder reconciliation but also the effectiveness of punishment institutions after conflict. Based on this model, victims may punish their previous perpetrators to regain status or power and undo past wrong-doing rather than to combat free-riding. They may also ignore punishment by previous perpetrators for free-riding and be unwilling to adapt selfish behavior. Previous perpetrators, on the other hand, may feel reluctant to punish former victims for free-riding to repair their self- or social image. We therefore expected that peer punishment is less able to prevent a breakdown of cooperation in groups with rather than without a history of conflict, whereas maladaptive punishment patterns disappear when previous conflict roles are hidden and punishment cannot be directed at past perpetrators and victims anymore. Yet, we had no a priori predictions on how peer punishment is exactly applied in the shadow of a conflict. We therefore pre-registered punishment as one of our main dependent variables and as an exploratory analysis.

We tested our hypotheses in small groups that first engaged in a standard attacker-defender contest. Thereafter, individuals engaged in a public good game with peer punishment. Depending on the treatment, identities of individuals in the conflict were kept or masked during the public goods game to test the second hypothesis. To test our first hypothesis, we added a no-conflict baseline treatment in which individuals engaged in a four-person public goods game without experiencing an asymmetric intergroup conflict.

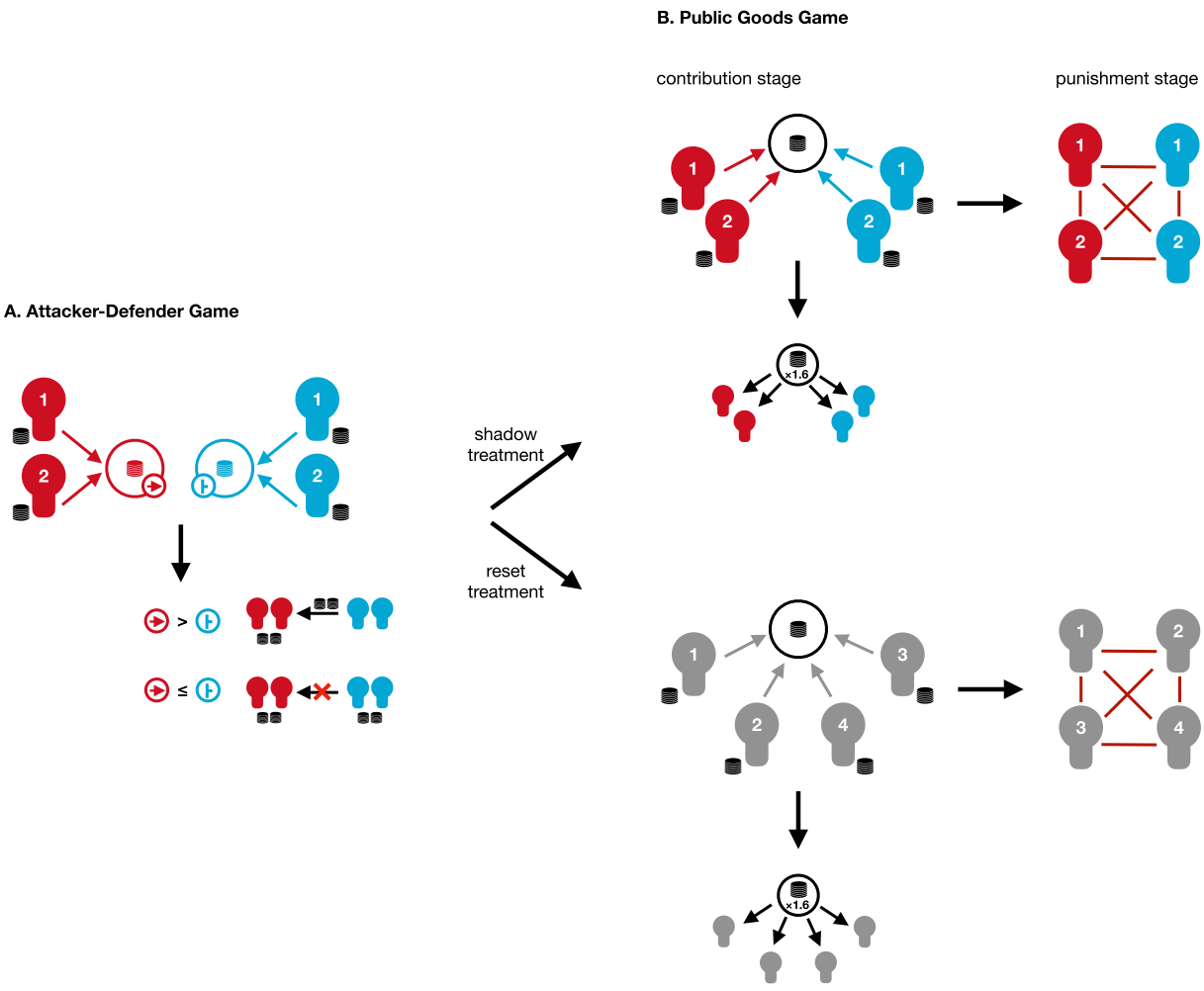


Fig. 1. *Experimental Design.* In the two experimental treatments, participants were either assigned to the role of attackers (red) or defenders (blue) in the attacker-defender game (A). In each round, they decided how many of their units to invest into a conflict pool. If the investments in the attacker pool exceeded the investments of the defender pool, attackers received all remaining units from the defenders and the defenders earned zero. If defenders invested more or equal units, they defended successfully and each group member earned the units that they kept for themselves. After repeatedly playing this game, groups transitioned to the public goods game with punishment (B). In each round, participants decided how many units to invest into a group pool (contribution stage). Each unit in the group pool was multiplied by 1.6 and divided equally among all group members. After observing contributions, participants could punish each other by investing additional units. Each unit assigned for punishment reduced the earnings of the punished group member by 3 while reducing the earnings of the punisher by 1. In the shadow treatment (upper panel), group members retained their group tags and it was, hence, possible to identify who was previously an attacker or defender. In the reset treatment (lower panel), new group tags were assigned, making it impossible to identify the past role in the conflict.

2. Methods

2.1. Research ethics and participants

The study received ethics approval from the Psychology Ethics Board of Leiden University. Subjects ($N = 240$, 186 females, mean age = 23.8, sd = 4.1) provided informed consent and were debriefed upon completion of the studies. The experiment did not involve any deception. Study design, hypotheses, and sample size were pre-registered on asPredicted (aspredicted.org/blind.php?x=4gh365). We did not perform an a-prior power analysis. Instead, we determined the sample size based on previous studies from our own lab and standards in the field (see, e.g., Beekman et al., 2017; Brandts & Cooper, 2006; Cason & Gangadharan, 2013; Gross & Böhm, 2020 for experimental designs similar to ours).

2.2. Experimental design

Participants were randomly assigned to groups of four and to one out of three treatments. In the 'shadow' and 'reset' treatment ($n = 80$ participants / 20 groups per treatment), participants first engaged in a block of a repeated attacker-defender game (first block, Fig. 1A; De Dreu et al., 2016) followed by a block of a repeated public goods game with peer punishment (second block, Fig. 1B; De Dreu, Pliskin, Rojek-Giffin, Méder, & Gross, 2021; Fehr & Gächter, 2000, 2002; Gross, Méder, Okamoto-Barth, & Riedl, 2016). In the control treatment ($n = 80$ participants / 20 groups), participants engaged in two blocks of the repeated public goods game instead. Below, we first describe the general rules of these two economic games with our implemented parameters and then describe the instructions, timeline, experimental procedures, and additional measures in more detail in the *Procedures* section.

2.3. Game design

Attacker-defender game. In the attacker-defender game, two players are in the role of 'attackers' and the other two players are in the role of 'defenders.' Each player receives an endowment of 20 units and has to decide how many of their units they want to spend on attacking the other group or defend against an attack, respectively ('conflict investment'). Players make this decision simultaneously, i.e., without knowing the decisions of the other players. If the two attackers, together, invest more of their units on conflict than the two defenders, the attackers win the conflict and receive all remaining (i.e., non-invested) units of the defender group while defenders end up with a payoff of 0. If defenders spend the same or more units into conflict, defenders successfully defend. In this case, each player simply earns the units that they did to not invest into conflict. The general payoff function of the attacker-defender game is:

$$\pi_{Ai} = \begin{cases} e - c_{Ai} + (n_D e - \sum_{j=1}^{n_D} c_{Dk})/n_D, & if \quad \sum_{j=1}^{n_D} c_{Dj} < \sum_{j=1}^{n_A} c_{Aj} \\ e - c_{Ai}, & if \quad \sum_{j=1}^{n_D} c_{Dj} \geq \sum_{j=1}^{n_A} c_{Aj} \end{cases}$$

$$\pi_{Di} = \begin{cases} 0, & if \quad \sum_{j=1}^{n_D} c_{Dj} < \sum_{j=1}^{n_A} c_{Aj} \\ e - c_{Di}, & if \quad \sum_{j=1}^{n_D} c_{Dj} \geq \sum_{j=1}^{n_A} c_{Aj} \end{cases}$$

where c = contribution to conflict, A = attacker, D = defender, n_A/n_D = attacker/ defender team size, e = individual endowment.

Hence, the attacker-defender game is an asymmetric conflict game in which one party has the possibility to increase their payoff at the expense of the other party. Also note that any unit invested into conflict does not count towards payoff. It follows that the most socially efficient outcome (i.e., the outcome that maximizes social welfare) is achieved if no player would invest anything into conflict and everybody would keep their units instead ('peace'). However, peace is game-theoretically unstable, since the best response of attackers would be to invest just one unit into conflict and thereby take all the units of the defender side in this case.

Public goods game with peer punishment. The public goods game with peer punishment has two stages. In the first stage, each player receives 20 units and simultaneously decides how many of their units to invest into a public good ('cooperation'). Every unit invested into the public good is multiplied by 1.6 and distributed equally across all group members, inducing a public goods dilemma. If everybody would invest all their 20 units to the group pool, they would earn $20 \times 4 \times 0.4 = 32$ units, each. Yet, a player can take advantage of the cooperation of others by withholding units. For example, if three players would decide to invest all their units into the public good while one player is keeping all units to herself, cooperating players only earn $20 \times 3 \times 0.4 = 24$ units, while the fourth, free-riding player earns $20 + 20 \times 3 \times 0.4 = 44$ units.

After the first stage, players learn about each others contributions and receive 18 additional units for punishment. Each player can then assign up to 6 units to punish each of the other players. Each unit assigned for punishment reduces the earnings of the punisher by 1 unit and reduces the earnings of the punished player by 3 units (i.e., a 1:3 punishment device). The public goods game with peer punishment was implemented following previous studies that have shown that peer punishment can prevent the usually observed breakdown of cooperation in public goods games (Fehr & Gächter, 2000, 2002; Gross et al., 2016).

2.4. Procedures

Although originally designed for our behavioral laboratory, the outbreak of the COVID-19 pandemic required us to build an on-line laboratory environment for interactive group decision-making. The experiment was programmed in PHP and jQuery by the first author and the instructions and code is available on OSF (https://osf.io/em653/). Screenshots of the instructions can also be found in the *Supplementary Information*. We invited participants from our local recruitment pool to sign up for an online session at a specific time of the day. Shortly before each session, participants were asked to go to a website using their personal computer. On this website, they first entered a virtual lobby where the general rules and procedure of the experiment were explained to them and where they were greeted by the experimenter over a chat box. The experiment started as soon as all participants entered the lobby. Participants could contact the experimenter over the chat box for questions throughout the entire experiment.

At the start of the experiment, participants were told that the experiment consists of multiple parts, that they would interact in a fixed group of four participants throughout the experiment, that they would earn a fixed amount of 4 euro for their participation, and that they could earn additional money that would depend on their decisions as well as the decision of the other participants in their group. Each part started with extensive instructions presented on the computer screen. Across all treatments, the first part was a measure of social preferences (the social value orientation slider measure; see *additional measures*, below).

Shadow and reset treatment. The second part in the shadow and reset treatment was the attacker-defender game. Participants learned that they were split into two groups of two and that one group could invest resources to take over the remaining resources of the other group, while the other group could invest resources to prevent this from happening. In the instructions, we used neutral labels, avoiding terms like 'attack' or 'defense' to mitigate framing effects. Attackers and defenders were identifiable by a color and were labeled 'group member 1' and 'group member 2' from group A (in one color) and 'group member 1' and 'group member 2' from group B (in another color). Before the start of the attacker-defender game, participants learned which role (attacker or defender) they were assigned to and that this role was fixed throughout this part. Further, they learned that one round from this part would be

randomly chosen for payoff by the computer at the end of the experiment and that 1 unit was worth 0.10 cents. The attacker-defender game was played for 15 rounds. Participants could see in which round they were and the total number of rounds at the top of the screen at the start of each round. Each round consisted of a decision and a feedback stage. In the decision stage, each participant decided how to distribute their 20 units between an 'invest' and a 'keep' pool. After every group member submitted their decision, the group transitioned to the feedback stage. They saw (a) how many units in total group A (the attackers) invested and how many units in total group B (the defenders) invested, (b) the outcome of the conflict, (c) how many units group A and group B earned in this round in total, and (d) individual earnings. If the two participants of the attacker group invested more units in the 'invest' pool than the two participants of the defender group, the attackers won the conflict and all remaining units in the defenders' keep pool were transferred to them. If the attacker group invested equal or less units in the 'invest' pool compared to the defender group, the defenders successfully defended and every group member simply kept their remaining units.

After the attacker-defender game, groups saw a summary screen showing how many units group A and group B accumulated in total, respectively. Then, they transitioned to the next part of the experiment; the public goods game with punishment. There was no time break between parts. They again received instructions followed by comprehension questions. In the instructions, we used neutral labels, avoiding terms like 'cooperation' or 'punishment' to mitigate framing effects. Participants also were told that the payoff of one randomly selected round of this block would be added to their payoff with a conversion rate of 1 unit = 0.10 cents. The public goods game was played for 15 rounds. Participants could see in which round they were and the total number of rounds at the top of the screen at the start of each round. Each round started with the contribution stage in which each group member decided how many of their 20 units to invest into the public good and how many units to keep for themselves. After the contribution stage, group members received feedback on (a) how many units each individual group member invested, (b) the total investment, (c) the return from the public good, and (d) how many units each individual group member would earn for this round, adding up the return from the public good and their kept units. Then, group members entered the punishment stage and had to simultaneously decide how many 'deduction points' to assign to each of the other group members (between 0 and a maximum of 6 for each group member). In the final feedback stage, we provided feedback on (a) how many deduction points each group member received from each other group member (and the received deduction points in total), and (b) a breakdown of the earnings for this round for each group member, summing up the return from the public good, the kept units, and deducting the points spent on punishment and received punishment. After this feedback stage, participants moved to the next round.

Importantly, group members were still identifiable as 'group member 1' and 'group member 2' from group A (in one color) and 'group member 1' and 'group member 2' from group B (in a different color) in the shadow treatment. Hence, it was common knowledge who was in the role of attackers and defenders in the previous game. In the reset treatment, group members were labeled from 'group member 1' to 'group member 4' and we removed the color code that indicated group membership in the previous attacker-defender game (see also Fig. 1B). Further, we shuffled these numbers, which was known to participants, to make sure that it was not possible to identify past 'perpetrators' and 'victims' from the previous attacker-defender block.

Control treatment. In the 'control' treatment, groups did not play the attacker-defender game. Instead, they played the public goods game with punishment for two consecutive blocks of 15 rounds, each. They received the same instructions for the public goods game as the groups in the shadow and reset treatment in block 2. In the second block of the control treatment, we summarized the rules again and simply told them that they would again interact in their group under the same rules for another 15 rounds. To isolate the influence of previous intergroup conflict and allow for a clean comparison to the shadow treatment (Hypothesis 1), participants in the control treatment received the same markers dividing them into two groups. Thus, participants were labeled as 'group member 1' and 'group member 2' from group A (in one color) and 'group member 1' and 'group member 2' from group B (in a different color).

Additional measures. In all treatments, we measured individual-level social preferences with the social value orientation slider task (Murphy et al., 2011) at the beginning of the experiment. In this task, participants have to decide how to distribute points between themselves and another unknown person across six allocation decisions. For example, the participant has to choose one out of nine possible choices ranging from allocating 100 points to oneself and 50 points to the other person (maximal 'pro-self' option) to allocating 50 points to oneself and 100 points to the other person (maximal 'pro-social' option). The total points kept for oneself and received from another randomly selected participant were converted to money at a rate of 100 points = 6 euro cents and added to the final payoff for the study. Each participant was matched with one receiver and was the receiver for another, different (and unknown) participant.

The experiment concluded with a task aimed to explore whether past interactions also influence trust between participants. To this end, we measured trust and reciprocity in a standard trust game followed by a demographics questionnaire. In the trust game, each participant in the role of the 'trustor' received 10 units and was told that they were paired to one of the other group members (without knowing which group member). They then decided how many of these 10 units to transfer to this group member (the 'trustee'). Each unit transferred was multiplied by three. Participants were also in the role of the trustee for another group member. After making their trust decision, they were asked how many units they would transfer back for every given possible transfer in the role of the trustee (so-called strategy method). Transfers in this game in the role of the trustor are interpreted as trust, because participants take the risk that the trustee does not transfer anything back to them. Back-transfers (in the role of the trustee), instead, can be interpreted as reciprocity of trust. Earned points in this game were converted to money at a rate of 1 unit = 10 euro cents and added to their payoff. Exploratory analyses revealed no effects (see Table S1 and S2) and these measures are further ignored. Participants were paid through bank-transfer shortly after the end of the experimental session.

2.5. Statistical analyses

The data is hierarchically structured, such that each data point (i.e., an investment decision) is nested in participants and groups over rounds. We accounted for the resulting violation of independence of individual data points by either aggregating the data on the group level (i.e., one average per group) and using non-parametric tests or fitted multilevel regression models using the 'lme4' package in R (and applying the Satterthwaite's degrees of freedom method to derive *p*-values). In each regression, we estimated two hierarchically clustered random intercepts to model decisions (level 1) nested in subjects (level 2) within four-person groups (level 3), as shown in equation (1). Note that a group here refers to four participants that interacted with each other in the experiment.

$$\begin{aligned} &y_{ijk} = \beta_{0jk} + \beta_1 X_{1ijk} + e_{ijk}, e_{ijk} \sim N(0, \sigma_e^2) && \text{(level-1)} \\ &\beta_{0jk} = \beta_{0k} + e_{0jk},\ e_{0jk} \sim N(0, \sigma_{e_{0jk}}^2) && \text{(level-2)} \\ &\beta_{0k} = \beta_0 + e_{0k}, e_{0k} \sim N(0, \sigma_{e_{0k}}^2) && \text{(level-3)} \\ &\text{where } k = \text{group}, \quad j = \text{subject}, && i = \text{response} \end{aligned} \tag{1}$$

3. Results

The results section is structured as follows: First, we report findings on decisions and outcomes in the attacker-defender game (the first part

of the experiment in the shadow and reset treatment). Second, we test the pre-registered Hypothesis 1 that having experienced a conflict (shadow treatment) vs. not (control treatment) reduces group cooperation. Third, we test the pre-registered Hypothesis 2 that not being able to identify past perpetrators (reset treatment) vs. being able to (shadow treatment) increases cooperation after a conflict. In these analyses, we will also analyze earnings (that is, earnings from the public good interaction minus the cost of punishment) as a measure of group success. We do this because observing high levels of cooperation in a public goods game with punishment does not necessarily mean that group members also benefit from cooperation. If high levels of cooperation are associated with high levels of punishment, groups may actually earn relatively little, since their earnings are reduced by the cost of punishment (as also discussed in previous literature, see Egas & Riedl, 2008; Engelmann & Nikiforakis, 2015; Gächter et al., 2008). Fourth, we explore whether differences in how groups use punishment can explain different levels and trajectories of cooperation across treatments. Lastly, we explore whether conflict dynamics within groups predict subsequent cooperation and punishment patterns.

3.1. Conflict dynamics in the shadow and reset treatment

Before playing the public goods game with punishment, groups in the shadow and reset treatment first interacted in the attacker-defender game. Because the experimental manipulation (identifiability of past roles in the public goods game) was only introduced in the second block of the experiment, groups played the same game under the same rules in the first block. We should therefore not find significant differences in behavior across treatments in terms of conflict intensity or outcomes.

Fig. 2 shows the average investments and earnings per role (attacker vs. defender), as well as the average win-rate of attackers between the reset and shadow treatment. In line with previous research on the attacker-defender game (De Dreu et al., 2016; De Dreu & Gross, 2019), we found that defenders invested more units than attackers, that attackers earned significantly more than defenders, and that attackers had a win-rate of around 26%. On average, participants invested 8 of their 20 units into the conflict and attackers earned 7.8 units more than defenders, indicating that the game induced a conflict and that attackers used their position in the conflict to exploit defenders.

As expected, we did not find significant differences in conflict dynamics between the shadow and the reset treatment (investments: multilevel regression, treatment coefficient = -0.32, se = 1.12, p = 0.78; earnings: treatment coefficient = -0.14, se = 0.99, p = 0.89, win-rate: linear regression, treatment coefficient = -0.01, se = 0.03, p = 0.77). This indicates that our random assignment of participants to treatments was successful and that groups experienced a similarly intense conflict in both treatments before entering the second part in which we introduced our experimental manipulation.

3.2. Cooperation under the shadow of conflict (Hypothesis 1)

To test our first hypothesis, whether the experience of a conflict (shadow treatment) vs. not (control treatment) reduces group cooperation, we compared (1) average group cooperation rates with a Mann-Whitney *U* test (i.e., data aggregated to the group level) and (2) cooperation across rounds with a multilevel regression model (i.e., analysis on the individual level) in the public goods game with punishment across these two treatments (pre-registered analyses). As discussed above, we also analyzed earnings. Remember that in both treatments, groups were divided into two subgroups (group A vs. group B). Hence, the information, rules, and group composition were exactly the same in the public goods game with punishment across these two treatments, but groups either experienced the attacker-defender conflict before (shadow treatment) or not (control treatment).

Fig. 3A illustrates the average cooperation rates across rounds. Contrary to our hypothesis that a previous intergroup conflict impedes cooperation, we did not find a significant difference in cooperation rates on the aggregate level (Wilcoxon signed-rank test, W = 214, p = 0.71). According to the fitted regression model, there was a marginally significant interaction between round and treatment (Table 1). This suggests that, while cooperation did not significantly decline across rounds in the control treatment (multilevel regression, round coefficient = -0.01, se = 0.02, p = 0.79), it declined in the shadow treatment (multilevel regression, treatment × round coefficient = -0.06, se = 0.03, p = 0.09). However, looking at Fig. 1A, the differences in cooperation across rounds were small and average cooperation rates in the last round were very similar (52.2% of the endowment was invested in the public good in the control treatment vs. 50.9% in the shadow treatment). Hence, the marginally significant interaction between round and treatment should be interpreted very cautiously. Groups in the shadow treatment, however, earned significantly less over rounds compared to the control treatment (Fig. 3B, multilevel regression, treatment × round coefficient = -0.13, se = 0.04, p = 0.004, see also Table 1). This already indicates that punishment was used more and differently when groups had (vs. had not) experienced an asymmetric conflict before.

3.3. Mitigating conflict spill-over by removing group-tags (Hypothesis 2)

To test our second hypothesis, that not being able to identify past perpetrators (reset treatment vs. shadow treatment) increases cooperation after a conflict, we compared cooperation in the public goods game between the reset and shadow treatment with a multilevel regression model (pre-registered analysis). As discussed above, we also analyzed earnings. Remember that groups played the same attacker-defender game in the first block in both treatments. In the second block, the rules of the public goods game with punishment were exactly the same with one critical difference: In the shadow treatment, the role that each participant had in the previous game (attacker vs. defender) was still identifiable, whereas these roles were hidden by reshuffling subject numbers and removing group tags in the reset treatment.

Table 2 summarizes results from the regression model. Fig. 4A illustrates the average cooperation rates across rounds. While groups in both treatments had similar cooperation rates in the first round (multilevel regression, treatment coefficient = -0.62, se = 1.55, p = 0.69), cooperation significantly increased over rounds in the reset treatment (multilevel regression, round coefficient = 0.12, se = 0.03, p < 0.001) compared to the shadow treatment (multilevel regression, treatment × round coefficient = -0.23, se = 0.04, p < 0.001). Hence, groups became more cooperative over time when conflict roles were hidden and punishment could not be aimed at previous attackers or defenders. Further, while groups in the shadow treatment earned descriptively (but not significantly) more in the first round (multilevel regression, treatment coefficient = 2.15, se = 1.28, p = 0.10), earnings significantly increased in the reset treatment over rounds (multilevel regression, round coefficient = 0.27, se = 0.04, p < 0.001) compared to the shadow treatment (Fig. 4B, multilevel regression, treatment × round coefficient = -0.32, se = 0.06, p < 0.001)[1].

We did not find any difference in cooperation rates depending on group membership. Hence, the role people had in the previous conflict was not significantly related to cooperation in the public goods game. When comparing cooperation between the reset and the control treatment, we also found that cooperation in the reset treatment significantly increased over rounds compared to the control treatment (multilevel regression, treatment × round coefficient = 0.17, se = 0.03, p < 0.001).

[1] Note that earnings sharply declined in the last round in both treatments. We attribute this to the so-called 'end-game effect' that is often observed in finitely repeated cooperation games. Some participants realize that the game ends and reduce their cooperation in the last round (see also Fig. 4A). Interestingly, group members that do not adapt their cooperation decision in the last round are still willing to punish those that do. Taken together, this leads to a decline in group-earnings in the last round.

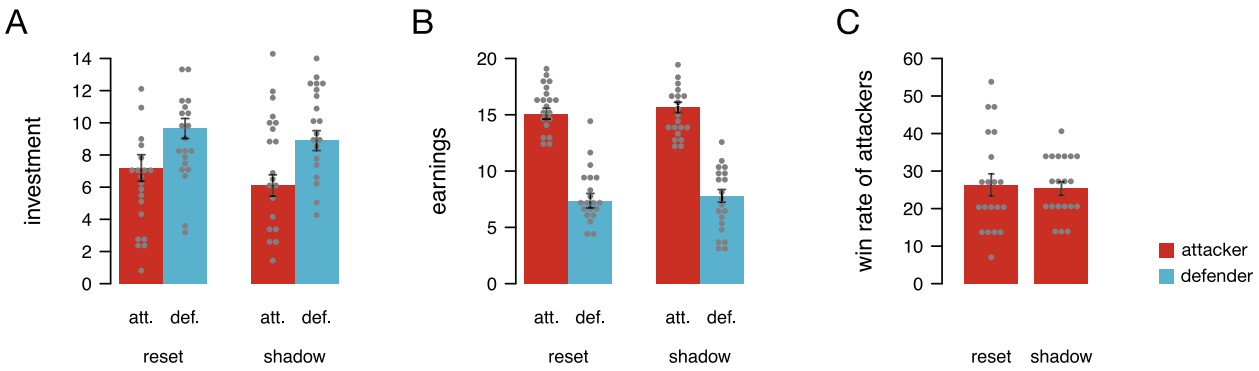


Fig. 2. *Conflict dynamics.* Average investments (A) and earnings (B) of attackers (red) and defenders (blue) in the reset and shadow treatment, as well as average win-rate in percent (C) of attackers across treatments. Error bars show the standard error of the mean. Individual points indicate averages per group.

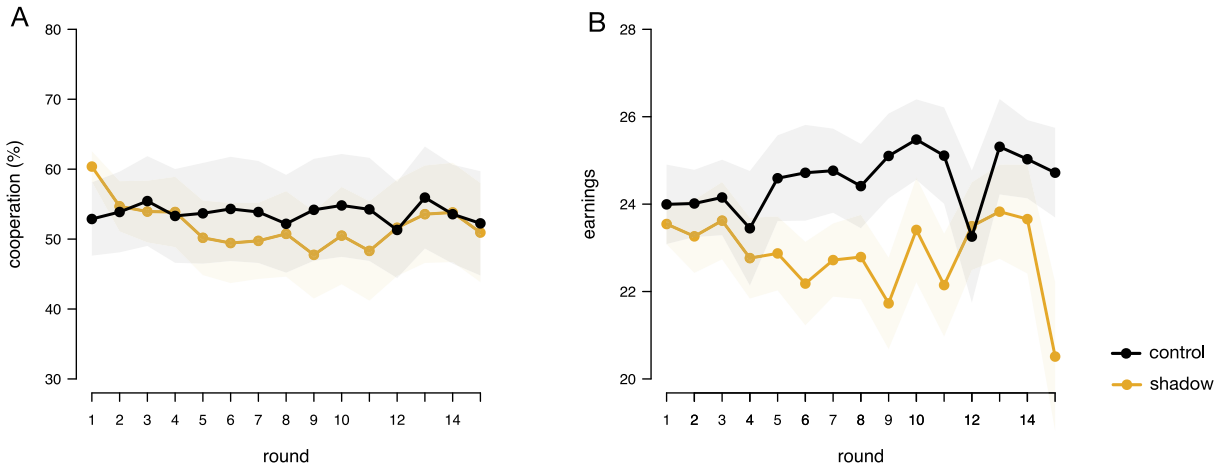


Fig. 3. *Cooperation dynamics with and without a shadow of conflict.* Average cooperation (A, percent of the endowment) and earnings (B) in the shadow treatment (yellow) in which groups first played the attacker-defender game and the control treatment (black) in which groups did not. Bands around the line indicate the standard error of the mean.

Table 1
Cooperation and earnings across shadow and control treatment. Multilevel regression modeling contributions to the public good (left) and earnings (right) in the second block of the experiment as a function of round and treatment.

	cooperation		earnings	
coefficient	estimate (std. error)	p-value	estimate (std. error)	p-value
intercept (*round 1; control*)	10.789 (1.218)	<0.001	24.050 (0.763)	<0.001
shadow treatment	0.043 (1.723)	0.980	−0.824 (1.080)	0.450
round	−0.006 (0.023)	0.789	0.070 (0.030)	0.022
round × treatment	−0.056 (0.033)	0.088	−0.126 (0.043)	0.004
$\sigma_{\text{level 1}}$	3.491		4.564	
$\sigma_{\text{level 2}}$	1.990		1.887	
$\sigma_{\text{level 3}}$	5.287		3.084	

Table 2
Cooperation and earnings across shadow and reset treatment. Multilevel regression modeling contributions to the public good (left) and earnings (right) in the second block of the experiment as a function of round and treatment.

	cooperation		earnings	
coefficient	estimate (std. error)	p-value	estimate (std. error)	p-value
intercept (*round 1; reset*)	11.454 (1.093)	<0.001	21.079 (0.904)	<0.001
shadow treatment	−0.622 (1.546)	0.689	2.148 (1.278)	0.100
round	0.166 (0.026)	<0.001	0.267 (0.039)	<0.001
round × treatment	−0.229 (0.037)	<0.001	−0.323 (0.055)	<0.001
$\sigma_{\text{level 1}}$	3.878		5.873	
$\sigma_{\text{level 2}}$	2.229		2.137	
$\sigma_{\text{level 3}}$	4.664		3.621	

3.4. The use of punishment amid a shadow of conflict

Results so far showed that having played the attacker-defender game did not affect cooperation rates significantly, but did reduce earnings over rounds compared to our control treatment. While this result is not in line with our first hypothesis, results on earnings indicate that punishment was used more frequently and possibly differently in the shadow compared to the control treatment. In line with our second hypothesis, removing group tags (reset treatment) increased cooperation compared to the shadow treatment. This indicates that removing the ability to identify past perpetrators and victims helps to foster cooperation after conflict, possibly also because punishment was used differently across these treatments. To shed light on how a previous intergroup conflict influenced peer punishment, we explored (1) the extent of punishment, (2) the targets of punishment, (3) the likelihood to get punished, and (4) the reaction to punishment across treatments.

Extent of punishment. Participants assigned punishment in 23% of the total opportunities to punish in the control treatment. This number increased to 33% in the reset treatment and 34% in the shadow treatment. While this may indicate a higher prevalence of punishment in the conflict treatments, participants in the control treatment also had a higher punishment prevalence in the first block of the experiment (33%)

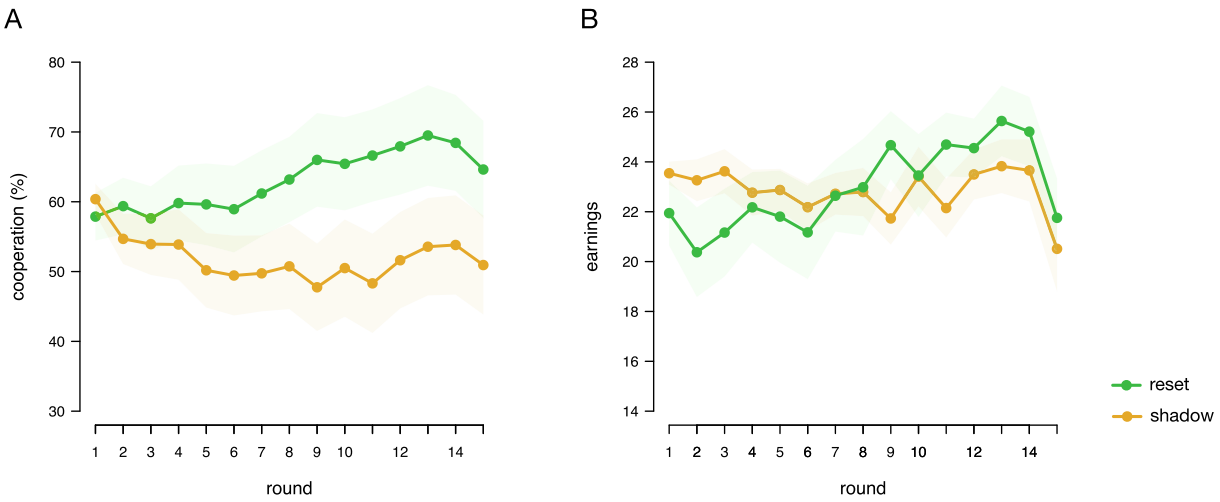


Fig. 4. *Cooperation dynamics with and without group tags.* Average cooperation (A, percent of the endowment) and earnings (B) in the shadow treatment (yellow) in which groups were able to identify previous attackers and defenders and the reset treatment (green) in which groups were not able to identify which role group members had in the previous conflict. Bands around the line indicate the standard error of the mean.

suggesting that this difference is mainly due to being confronted with the public goods problem for the first (shadow and reset treatment) vs. second time (control treatment).

Targets of punishment. Being divided into two groups of previous victims and perpetrators, participants in the shadow treatment may exhibit a bias towards punishing group members of the opposing team ('out-group') more than the own team ('in-group'), i.e. parochial punishment. Especially previous defenders may punish previous attackers more than fellow defenders, which could indicate that punishment is motivated by spite or revenge rather than just punishing free-riders. We therefore ran regressions predicting punishment based on the target of punishment (coded as 'in-group' member or 'out-group' member) and the own role in the previous conflict (attacker vs. defender). The main effect of target tells us whether out-group members were punished more compared to in-group members, while the interaction with the own role (dummy coded as 0 = attacker and 1 = defender), tells us whether defenders more strongly punished out-group members (i.e., previous attackers) compared to in-group members. We should not observe any effects in the reset treatment because previous group affiliation was hidden and the group was not divided into sub-groups anymore in this treatment. The comparison with the control treatment without a previous conflict can reveal whether a bias towards punishing out-group members is, in fact, a consequence of having previously experienced an intergroup conflict or whether it should rather be interpreted as a consequence of dividing the group into two sub-groups.

Table 3 shows the regression results (see also Fig. 5A). As expected, the extent of punishment did not significantly differ within vs. across previous sub-groups in the reset treatment (since previous roles were not identifiable anymore in this treatment). In the shadow treatment, we do observe that out-group members were significantly more punished compared to in-group members (shadow treatment: multilevel regression, target: out-group coefficient = 0.10, se = 0.04, p = 0.01). However, we also observed a significant effect of punishing out-group member more harshly compared to in-group members in the control treatment (control treatment: multilevel regression, target: out-group coefficient = 0.05, se = 0.02, p = 0.03). Hence, while we did find evidence for parochial punishment (i.e., punishing out-group members more than in-group members), this was likely driven by dividing groups into two sub-groups rather than having experienced an intergroup conflict. Defenders were also not significantly more likely to punish the out-group (i.e., previous attackers) in either treatment (target × role interactions in Table 3). Of note is that punishment significantly decreased over rounds according to the fitted model in the control and reset treatment, which was not the case in the shadow treatment (round effect in Table 3). This partly explains the lower earnings across rounds in the shadow treatment compared to both of the other treatments.

Likelihood to get punished. Although the use of punishment was influenced by group membership even when groups did not experience an intergroup conflict (control treatment), the shadow of conflict influenced the likelihood to get punished (Table 4). When a participant deviated from the average cooperation level of the group, the odds of getting punished increased by 20.3 in the control treatment (Table 4, multilevel logistic regression, cooperation deviation coefficient = 3.01, se = 0.28, p < 0.001). These odds did not significantly change in the reset treatment (multilevel logistic regression, cooperation deviation × reset treatment coefficient = -0.48, se = 0.36, p = 0.17), while they significantly decreased in the shadow treatment by 13.7 points (multilevel logistic regression, cooperation deviation × reset treatment

Table 3
Targets of punishment. Multilevel regression modeling punishment decisions as a function of the own and target's role in the past.

	control		shadow		reset	
coefficient	estimate (std. error)	p	estimate (std. error)	p	estimate (std. error)	p
intercept	0.153 (0.043)	<0.001	0.217 (0.059)	<0.001	0.493 (0.104)	<0.001
round	−0.006 (0.002)	0.003	−0.001 (0.003)	0.804	−0.011 (0.004)	0.013
target: out-group	0.054 (0.024)	0.025	0.095 (0.038)	0.012	0.016 (0.053)	0.763
role: defender	0.017 (0.048)	0.728	0.038 (0.065)	0.554	−0.013 (0.097)	0.891
target × role	0.008 (0.034)	0.808	−0.033 (0.054)	0.544	−0.112 (0.074)	0.133
$\sigma_{\text{level 1}}$	0.420		0.655		0.910	
$\sigma_{\text{level 2}}$	0.184		0.234		0.366	
$\sigma_{\text{level 3}}$	0.105		0.132		0.323	

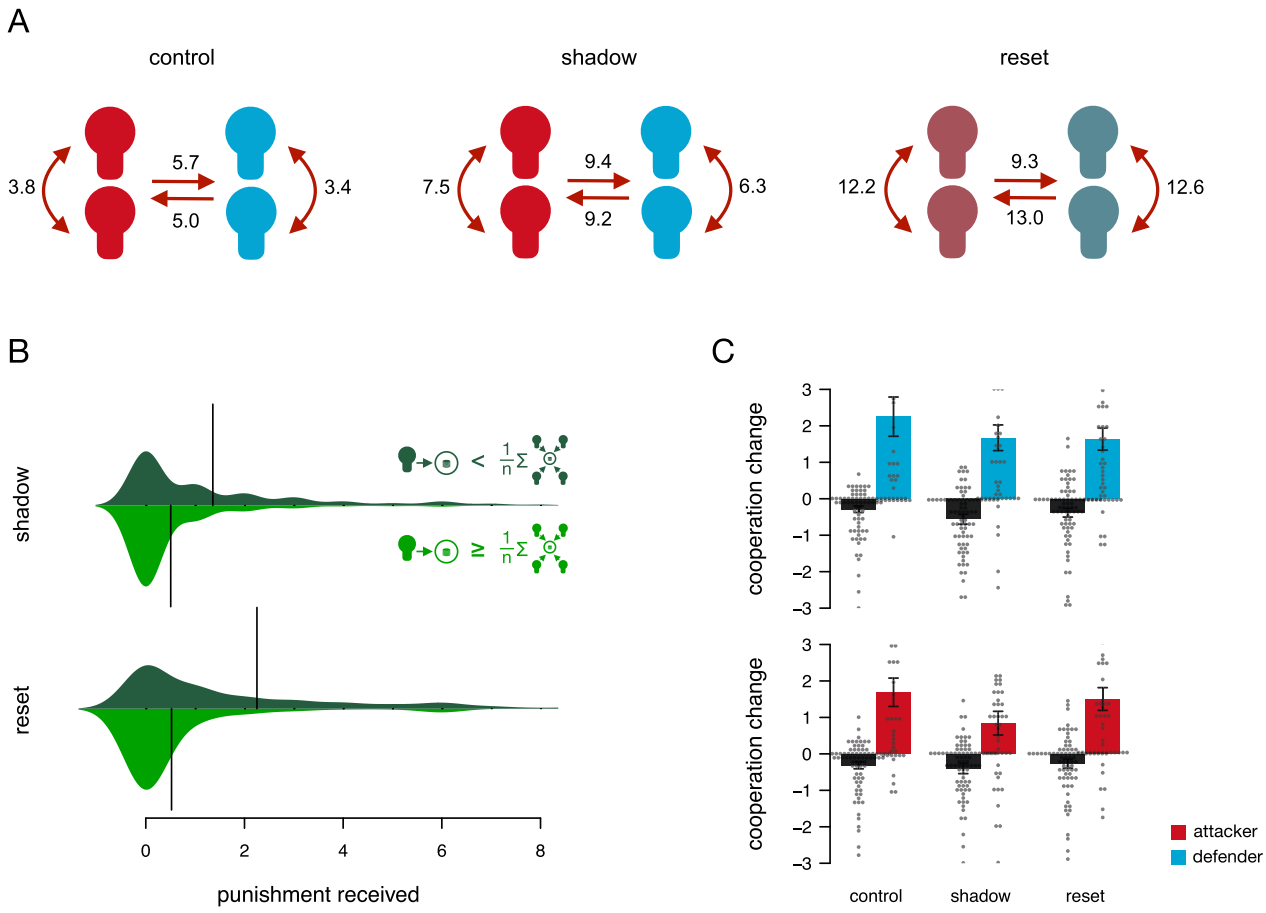


Fig. 5. *Punishment dynamics.* Average total received punishment points from participants belonging to the own group or the opposing group. Punishment across groups (straight arrows) was more frequent than punishment between group members (curved line) in the control and shadow treatment, which was not the case in the reset treatment (A). On average, group members received more punishment when deviating from average group cooperation (dark green) in the reset treatment compared to the shadow treatment, while cooperators are punished similarly (light green). The vertical line indicates the average (B). Receiving punishment by defenders (blue, upper panel) vs. not (grey) generally increased cooperation. The effect of receiving punishment by attackers (red bars, lower panel) on the change of cooperation was lower in the shadow treatment compared to the other treatments (C).

coefficient = -1.22, se = 0.34, $p < 0.001$). In other words, the likelihood to get punished for free-riding was significantly lower in the shadow treatment, as also illustrated in Fig. 5B.

Reaction to punishment. Participants also reacted differently to receiving punishment depending on the source of punishment and the treatment. To analyze how participants reacted to punishment, we calculated the change in cooperation from one round to the next and fitted a model that predicts this cooperation change based on the punishment received by members from the two sub-groups (Table 5). As illustrated in Fig. 5C, group members generally increased their cooperation when they got punished compared to not getting punished. In the reset treatment, group members increased their cooperation both when punished by previous attackers (multilevel regression, punishment by attackers (t-1) coefficient = 0.56, se = 0.09, $p < 0.001$) or defenders (multilevel regression, punishment by defenders (t-1) coefficient = 0.37, se = 0.07, $p < 0.001$). This is of course not surprising, since previous group membership was not visible anymore in this treatment. In the control treatment, the reaction to punishment by members from group A (attackers in the other two treatments) did not significantly differ compared to the reset treatment (multilevel regression, punishment by attackers (t-1) × control coefficient = 0.22, se = 0.18, p = 0.23). Yet, when group membership was still identifiable and associated with a specific role in the previous intergroup conflict (i.e., the shadow treatment), group members reacted significantly less to punishment from previous attackers (multilevel regression, punishment by attackers (t-1) × shadow coefficient = -0.32, se = 0.14, p = 0.03). Participants also increased their cooperation more when punished by previous defenders in the shadow treatment compared to the reset treatment (multilevel regression, punishment by defenders (t-1) × shadow coefficient = 0.69, se = 0.13, $p < 0.001$). However, a similar effect was found for the control treatment (multilevel regression, punishment by defenders (t-1) × control coefficient = 0.66, se = 0.15, $p < 0.001$).

An alternative way to analyze how group members react to

Table 4
Likelihood to get punished. Logistic multilevel regression modeling the likelihood to get punished as a function of the deviation from average cooperation and treatment.

Coefficient	estimate (std. error)	p-value
intercept (*round 1; control*)	−3.124 (0.455)	<0.001
deviation	3.010 (0.282)	<0.001
reset treatment	1.580 (0.606)	0.009
shadow treatment	1.835 (0.603)	0.002
round	−0.059 (0.012)	<0.001
deviation × reset treatment	−0.484 (0.355)	0.173
deviation × shadow treatment	−1.220 (0.336)	<0.001
$\sigma_{\text{level 2}}$	0.644	
$\sigma_{\text{level 3}}$	1.696	

Table 5
Reaction to getting punished. Multilevel regression modeling the change in cooperation as a function of getting punished in round t-1 and treatment.

Coefficient	estimate (std. error)	p-value
intercept (*treatment* = *reset*)	−0.414 (0.130)	0.002
control	−0.026 (0.181)	0.884
shadow	−0.234 (0.185)	0.212
punishment by attackers (t-1)	0.560 (0.094)	<0. 001
punishment by defenders (t-1)	0.366 (0.074)	<0. 001
punishment by attackers (t-1) × control	0.216 (0.181)	0.233
punishment by attackers (t-1) × shadow	−0.322 (0.143)	0.025
punishment by defenders (t-1) × control	0.662 (0.150)	<0.001
punishment by defenders (t-1) × shadow	0.686 (0.130)	<0.001
$\sigma_{\text{level 1}}$	3.77	
$\sigma_{\text{level 2}}$	0.00	
$\sigma_{\text{level 3}}$	0.155	

punishment is to categorize received punishment in round t-1 based on whether it was coming from in-group members or out-group members and their role in the previous intergroup conflict (similar to the setup of the model shown in Table 3). This showed that previous defenders in the shadow treatment increased their cooperation in the next round by 0.9 points for every punishment point they received from their in-group (i.e., a fellow defender, multilevel regression, punishment by in-group (t-1) = 0.91, se = 0.24, p < 0.001), whereas they only increased their cooperation by 0.2 points in the next round for every punishment point they received from the out-group (i.e., previous attackers; multilevel regression, punishment by out-group (t-1) = 0.20, se = 0.12, p = 0.09). This relationship flipped for previous attackers. Previous attackers increased their cooperation in the next round by 1.2 points for every punishment point they received from the out-group (i.e., previous defenders, multilevel regression, punishment by out-group (t-1) = 1.16, se = 0.15, p < 0.001), whereas they only increased their cooperation by 0.1 points in the next round for every punishment point they received from the in-group (i.e., fellow attacker, multilevel regression, punishment by in-group (t-1) = 0.13, se = 0.30, p = 0.678). Thus, the effect of punishment on cooperation was exclusively driven by defenders in the shadow treatment. Punishment by previous attackers did not significantly increase subsequent cooperation (regardless of the previous role of the punished).

Taken together, a shadow of conflict in combination with the ability to identify past perpetrators and victims selectively reduced (a) the likelihood to aim punishment at free-riders, and (b) the sensitivity of punishment executed by (previous) attackers.

3.5. Dynamics within groups

Above, we analyzed how a previous intergroup conflict influences cooperation and punishment patterns based on comparisons across our experimental manipulations, i.e., whether groups experienced an intergroup conflict (shadow treatment) or not (control treatment) and whether previous perpetrators and victims were identifiable (shadow treatment) or not (reset treatment). This allows to draw causal inferences based on our experimental manipulations. Another approach is to look at specific patterns within groups to understand how the previous conflict influences cooperation. While exploratory and only providing correlational evidence, it may further help us to understand why group cooperation in the shadow treatment did not significantly deviate from the control treatment (rejection of Hypothesis 1), even though we did find differences in punishment patterns across treatments, as reported above.

Previous conflict intensity and cooperation. The willingness of attackers and defenders to cooperate (especially in the first round) may depend on the conflict intensity they experienced. Previous conflict dynamics could also influence the extent or use of punishment. We therefore explored whether conflict intensity in the first block (measured as the average investment in conflict or win-rate of attackers) predicts subsequent cooperation and punishment in the public goods game in the shadow and reset treatment. For this, we fitted regression models predicting cooperation and punishment (in the second block) based on the investment in conflict in the attacker-defender game (i.e., average investment in conflict of the group) in interaction with the role in the previous conflict. We could not find statistical evidence that previous conflict intensity significantly predicted cooperation rates nor punishment across rounds or in the first round (all $p > 0.20$). This suggests that hiding previous conflict roles play a more important role for the cooperation trajectory that groups take than the intensity of the previously experienced conflict. Yet, we should also note that these comparisons rely on within-condition comparisons across groups and are less sensitive than across round or across treatment comparisons.

Attackers' initial cooperation. Cooperation across rounds in public goods games is often path-dependent, meaning that first-round decisions can influence the trajectory of cooperation within groups. In our setting, this may give attackers an opportunity for 'reconciliation.' Especially if previous attackers cooperate strongly in the first round of the public goods game, they can signal their willingness to cooperate with the whole group which may then help to foster group cooperation in subsequent rounds of the public goods game. Importantly, this is not possible for attackers in the reset treatment since previous roles are hidden. To test this idea, we looked at first-round decisions of attackers in the public goods game, i.e., decisions that are still unaffected by other group members' cooperation decisions and with the largest potential to influence subsequent group cooperation. We split groups into those in which previous attackers cooperated more than their fellow, previous defenders and those in which previous attackers cooperated less or equal than their fellow, previous defenders in the first round. Based on this split, we analyzed cooperation rates across rounds in these groups.

Fig. 6 shows the cooperation trajectory of groups in which attackers cooperated more in the first round or not for the three treatments. Note that in the control treatment, groups were not split into attackers and defenders but still divided into two groups of two participants (group A vs. group B). In the control treatment, cooperation across rounds was not predicted by whether participants in group A cooperated more than group B (Fig. 6A, multilevel regression, group A > group B × round = 0.06, se = 0.05, p = 0.182). In the shadow treatment, on the other hand, cooperation across rounds decreased significantly when previous attackers cooperated less or equal than defenders compared to the control treatment (multilevel regression, shadow treatment × attacker ≤ defender × round = -0.14, se = 0.05, p = 0.003). On the flipside, when previous attackers cooperated more than defenders in the first round, cooperation stabilized across rounds (Fig. 6C, shadow treatment × attacker > defender × round = 0.17, se = 0.07, p = 0.014). In the reset treatment, in which previous conflict roles were hidden, this relationship even reversed (Fig. 6B). This shows that identifiability of past perpetrators was not necessarily bad for all groups in our setting. High first-round cooperation of previous attackers allowed groups to overcome the 'shadow of conflict' and cooperate, whereas low first-round cooperation of previous attackers was associated with declining group cooperation.

4. Discussion

In asymmetric conflicts, resources are invested towards hurting and exploiting or defending against such attempts. This makes such conflicts not only collectively wasteful, past conflict can also reduce the ability to establish cooperative relationships between groups after a history of conflict (Bar-Tal, 2000; Beekman et al., 2017; Cilliers et al., 2016; Rouhana & Bar-Tal, 1998). Here we investigated this psychological 'shadow of conflict' experimentally and showed how a previous intergroup conflict influences cooperation and the use of punishment. We did

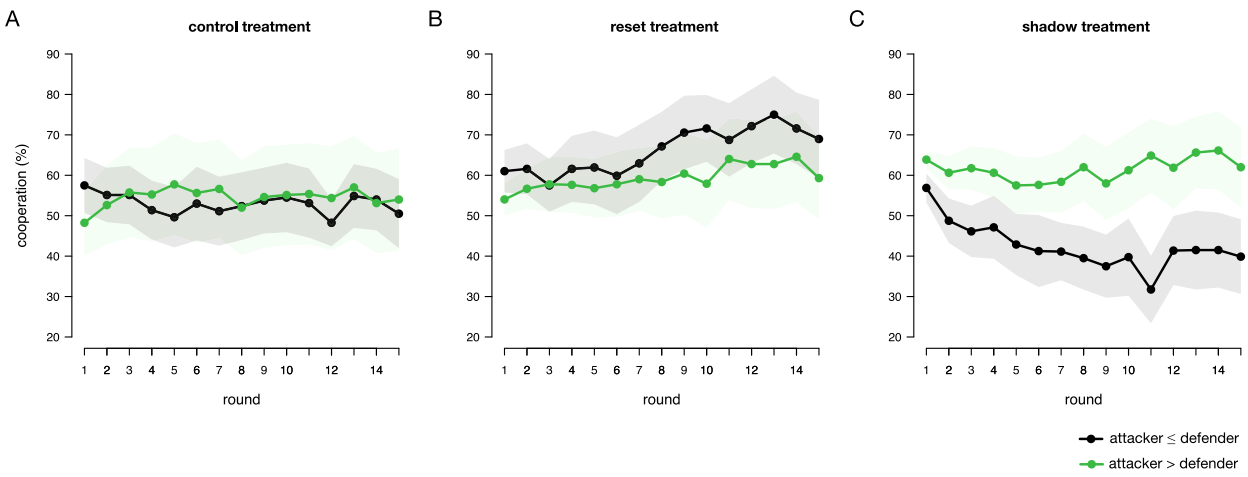


Fig. 6. *Cooperation trajectory.* Average cooperation across rounds in groups in which previous attackers cooperated more (green) or equal/below their fellow defenders (black) in the first round in the control (A) reset (B) and shadow treatment (C). Note that in the control treatment, groups were not split into attackers and defenders but still divided into two groups of two participants. Bands around the line indicate the standard error of the mean.

not observe that experiencing a previous conflict necessarily leads to lower rates of cooperation compared to groups with no conflict history. While the rejection of our first hypothesis could be due to a lack of statistical power to detect a truly existing difference, average cooperation only differed by 0.4 percentage points (Cohen's $d = 0.07$) across our shadow and our control treatment[2]. Hence, even if a shadow of conflict reduces cooperation, the effect would be negligible. Instead, we observed that cooperation after conflict is highly path dependent in our groups. Participants in the role of previous attackers could positively (and negatively) influence group cooperation by highly cooperating in the first round of the public goods game, suggesting that previous attackers can 'repair' the intergroup relations by immediately signaling their willingness to cooperate after conflict.

We further tested whether removing past group affiliations, making it impossible to identify past perpetrators and victims of the conflict, increases cooperation. In line with our second hypothesis, we did observe that groups established higher levels of cooperation compared to our treatment with identifiability. While it is not always possible (or desirable) to hide past conflict roles in real conflicts, our experimental results show that masking past membership can reduce negative spill-over of intergroup conflict on group cooperation. In real conflicts, group members may also try to actively hide past group affiliations to regain status in the group or avoid retaliation, by, for example, denying the role they played in the conflict. While this may decrease the ability to reverse past wrong-doings or engage in active reconciliation (Bar-Tal, 2000; Cilliers et al., 2016), it may also be functional to re-establish cooperative relationships. For example, in the efforts by the Colombian government to reintegrate past members of paramilitary groups into civil society, many guerrilla fighters choose to hide their past role in the civil war, which is associated with higher self-reported identification with civil society (see González & Clémence, 2019).

Across all treatments, we observed that groups with a shadow of conflict earned progressively less and, hence, benefitted less from the cooperation opportunities they had after the conflict episode compared to groups without a previous intergroup conflict (control treatment) and groups in which previous conflict roles were hidden (reset treatment). By analyzing the patterns of punishment, we found that groups that experienced a shadow of conflict did not punish free-riders as harshly compared to the other treatments. Further, punishment by past attackers was less effective in inducing subsequent cooperation, suggesting that attackers lose their legitimacy to enforce norms of cooperation when their past role in the conflict is identifiable (see also Baldassarri & Grossman, 2011; Faillo et al., 2013; Gross et al., 2016 for related findings on the role of legitimacy for the effectiveness of punishment in non-conflict settings). Even previous attackers did not significantly change their subsequent cooperation when having received punishment by their fellow, previous attacker. Hiding previous group affiliations, instead, made punishment by previous attackers as effective in promoting cooperation as in the control treatment.

These results reveal an important boundary condition for peer punishment institutions. While many experiments have shown that peer punishment can stabilize cooperation in groups (Fehr & Gächter, 2000; Masclet et al., 2003; Yamagishi, 1986), other research also showed that peer punishment can be misused or underused and is not always aimed at free-riders. In such cases, the ability to punish group members can have detrimental consequences for cooperation and group earnings (Abbink et al., 2017; Engelmann & Nikiforakis, 2015; Herrmann et al., 2008; Nikiforakis, 2008). We qualify and extend these observations by highlighting that past relationships of groups can change the use and effectiveness of punishment institutions. These findings also resonate with previous research on spill-over effects. Empirical research frequently observed that past interactions can shape expectations, norms, and cooperation (Beekman et al., 2017; Cason et al., 2012; Cassar et al., 2013; Iacono & Sonmez, 2020; Knez & Camerer, 2000; Peysakhovich & Rand, 2016; Stagnaro et al., 2017) and, as shown here, the effectiveness of institutions to promote cooperation. Previous research also has shown that third parties are less likely to intervene when people had a history of conflict (Nakashima et al., 2017).

We also found that group membership alone can induce parochial punishment patterns in which group members punish individuals from the opposing group more than fellow group members. This pattern was observed regardless of whether groups experienced an intergroup conflict previously or not, resonating with findings that show that assigning people to arbitrary groups can already induce parochial behavior (i.e., a 'mere membership' effect, Gummerum et al., 2009; Chakravarty & Fonseca, 2014; Charness & Chen, 2020; McAuliffe & Dunham, 2016). As such, our findings resonate with extant work on recategorization and social behavior. For example, (re-)categorizing individuals from distinct social categories into an overarching collective by emphasizing features that members of both sub-categories share can increase collective identification and cooperative inclinations (Dovidio et al., 2009; Gaertner et al., 1994; Hornsey & Hogg, 2000; Weber et al., 2004), similar to the enhanced cooperation we found in our reset versus control treatment (for a discussion on the role of group identification and joint

[2] To illustrate, it would require a sample size of 3350 groups per treatment to detect this observed difference with a power of 80% on the group level.

action, see also Hasan-Aslih et al., 2020). Furthermore, our public goods game resembles a situation of intergroup contact following a conflict. While conflict can lead to parochial 'bonding' norms that increase parochialism (Bauer et al., 2016; Choi & Bowles, 2007), increased intergroup contact reduces social segregation and can promote the emergence of intergroup cooperation (Bakke et al., 2009; Dyrstad, 2012; Mironova & Whitt, 2016). This reasoning would resonate with the non-negligible extent of cooperation observed in our shadow of conflict treatments.

4.1. Limitations and future outlook

In our study, we were particularly interested in asymmetric conflict situations that divide groups into perpetrators and victims, as is often the case in civil wars or internal societal conflicts that separate groups into fractions that defend a status quo or challenge it (De Dreu & Gross, 2019). Further, by investigating an asymmetric conflict, we could investigate how cooperation and punishment is influenced by the previous role in the conflict. However, conflicts between two parties can also be more symmetric, for example when parties fight for the same resource rather than challenge the resource that one has and the other one wants. In real conflicts, the role of attackers and defenders can also switch over time in which case the concept of attackers and defenders becomes fuzzy and the strict asymmetry that we can induce in the laboratory disappears. A study by Beekman et al. (2017) investigated the influence of symmetric conflict on subsequent cooperation. They divided six participants into two groups that first played a repeated symmetric conflict game. In this symmetric conflict, members of the two groups contributed resources to compete for a fixed prize that only one group could attain. Specifically, if one group invested more of their resources, they won the prize (in a given round). Subsequently, the two groups interacted in a repeated nested public goods game in which they could invest resources towards a public good that benefitted all participants or a group good (also referred to as club good) that only benefitted participants of their own group. They found that experiencing a previous conflict (vs. not) increased parochial cooperation (i.e., cooperation towards the group exclusive club good). Further, playing the public goods game with the same opposing group that they competed with in the symmetric conflict reduced cooperation towards the public good compared to a treatment in which groups were rematched after the symmetric conflict (which may be psychologically similar to our reset treatment). However, Beekman et al. (2017) did not investigate punishment institutions. Based on our results, the asymmetric conflict also asymmetrically influenced the effectiveness of punishment when previous roles in the conflict were identifiable. Punishment by previous defenders could still induce higher levels of cooperation, while the effectiveness of punishment by previous attackers was significantly reduced in the shadow treatment. A straightforward prediction from our results is that we should still expect parochial punishment patterns in symmetric conflicts, since these were even observed in groups without a previous conflict. Yet, the asymmetry in the effectiveness of punishment should disappear when there is no conflict role differentiation anymore and the conflict is symmetric. However, future work is needed to specifically test these predictions in symmetric conflicts.

An important question is why exactly punishment by previous attackers was less effective in inducing higher cooperation (even among attackers). One straightforward interpretation is that attackers lose legitimacy to foster cooperation due to their previous conflict role. However, it may also be more difficult to attribute clear meaning to punishment when previous perpetrators and victims are still identifiable, especially when it comes from attackers that previously used their resources to exploit and hurt the other party. Receiving punishment by past attackers could be interpreted as genuine punishment for free-riding or an attempt to just hurt another person. Such attributional ambiguity resonates with work on punishment as a communication device (Cushman et al., 2021; Ho et al., 2019) that highlights that punishment is not only taken as an incentive to change behavior but that people learn from the inferred meaning or intention behind punishment. Yet, a clear meaning or intention behind punishment (i.e., what is communicated through punishment) may be disrupted in our shadow treatment when it is executed by previous attackers.

It is also important to note that a myriad of psychological factors can impede reconciliation after a conflict. As others point out (Bar-Tal, 2000; Bar-Tal & Halperin, 2011; De Dreu & Carnevale, 2003; Ross & Ward, 1995), intergroup conflict can create psychological barriers for conflict resolution on the cognitive level (like, e.g., distorted information processing, lack of trust, perceiving a situation as zero-sum), emotional level (lingering emotions of anger or spite), and motivational level. Our research focused on the behavioral outcomes of intergroup conflict (i.e., cooperation and punishment use) in a controlled laboratory experiment. Ultimately, we therefore can only speculate about the complex psychological processes that may influence the individual's perception and interpretation of behavior after conflict.

Since punishment was less effective when executed by attackers, only defenders were able to influence subsequent cooperation through punishment in the shadow treatment. Since punishment is costly, this creates an additional burden for participants in the role of previous defenders. Not only were they in the inferior position in the conflict, they subsequently are in a position in which only they can efficiently enforce a norm of cooperation. Previous attackers, instead, while being less able to enforce cooperation through punishment were still able to foster group cooperation by their initial behavior in the public goods game, as our analysis of splitting groups into high cooperating and low cooperating attackers (in the first round) demonstrated. This also resonates with the needs-based model by Nurit Shnabel and Arie Nadler (e.g., Shnabel & Nadler, 2008, 2015; Nadler & Shnabel, 2015) according to which past victims have a deprived need for agency and power and past perpetrators are motivated to restore their moral self-image (see also Baumeister et al., 1990). A signal of previous perpetrators to be willing to invest their resources to benefit the whole group may not only help to restore their own self-image but also partially fulfill the need of defenders to perceive more power or equality and increase their willingness to cooperate in subsequent rounds. An interesting follow-up idea would be to directly test whether a message of apology or another form of active reconciliation immediately after the conflict episode could repair group relations and increase the effectiveness of punishment executed by previous attackers. This may especially help in settings in which the intergroup conflict abruptly ends and the possibility for group-wide cooperation is introduced, as in our setup. It is important to note that parties not necessarily transition from conflict to cooperation immediately in real-world conflicts. Instead, this transition can be associated with a slower integration process in which a longer period of reconciliation is possible. Experimentally, this could be investigated by giving groups an opportunity to exchange costly gifts for multiple periods before being confronted with a public goods dilemma.

5. Conclusions

Conflicts prevail in the organizational context, between competing firms or work-teams, and at the societal level, between different ethnicities or nations. How to (re-)establish cooperative relationships after a merger in the organizational context or the end of intergroup violence between social groups is a pressing question for conflict resolution. Here we specifically investigated the (mal)adaptive function of punishment institutions to foster cooperation after an asymmetric conflict that splits groups into previous perpetrators and victims. While punishment allowed to sustain cooperation even after the intergroup conflict, punishment was also underused and less effective amid a shadow of conflict, in particular when previous conflict parties could be identified. From a practical perspective, our results suggest that punishment institutions should be combined with mechanisms that blur past group affiliations to

avoid counterproductive punishment by, for example, creating a shared cooperate identity after the merging of former competing organizations or highlighting the present commonalities rather than past dividing lines of conflicting parties.

CRediT authorship contribution statement

Jörg Gross: Conceptualization, Formal analysis, Funding acquisition, Software, Writing – original draft. **Carsten K.W. De Dreu:** Conceptualization, Funding acquisition, Writing – review & editing. **Lennart Reddmann:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project has received funding from the Netherlands Science Foundation (VENI 016.Veni.195.078) and the European Union to JG and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (AdG agreement n° 785635) to CKWDD. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.obhdp.2022.104152.

References

Abbink, K., Gangadharan, L., Handfield, T., & Thrasher, J. (2017). Peer punishment promotes enforcement of bad social norms. *Nature Communications, 8*(1), 609. https://doi.org/10.1038/s41467-017-00731-0

Ahn, T. K., Ostrom, E., Schmidt, D., Shupp, R., & Walker, J. (2001). Cooperation in PD games: Fear, greed, and history of play. *Public Choice, 106*(1–2), 137–155. https://doi.org/10.1023/a:1005219123532

Anderson, L. R., Mellor, J. M., & Milyo, J. (2008). Inequality and public good provision: An experimental analysis. *The Journal of Socio-Economics, 37*(3), 1010–1028. https://doi.org/10.1016/j.socec.2006.12.073

Bakke, K. M., Cao, X., O'Loughlin, J., & Ward, M. D. (2009). Social distance in Bosnia-Herzegovina and the North Caucasus region of Russia: Inter and intra-ethnic attitudes and identities. *Nations and Nationalism, 15*(2), 227–253. https://doi.org/10.1111/j.1469-8129.2009.00363.x

Baldassarri, D., & Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences, 108* (27), 11023–11027. https://doi.org/10.1073/pnas.1105456108

Bar-Tal, D. (2000). From intractable conflict through conflict resolution to reconciliation: Psychological analysis. *Political Psychology, 21*(2), 351–365. https://doi.org/10.1111/0162-895x.00192

Bar-Tal, D., & Halperin, E. (2011). Socio-psychological barriers to conflict resolution. In D. Bar-Tal (Ed.), *Intergroup conflicts and their resolution: A social psychological perspective*. Psychology Press.

Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., & Mitts, T. (2016). Can war foster cooperation? *Journal of Economic Perspectives, 30*(3), 249–274. https://doi.org/10.1257/jep.30.3.249

Baumeister, R. F., Stillwell, A., & Wotman, S. R. (1990). Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology, 59*(5), 994–1005. https://doi.org/10.1037/0022-3514.59.5.994

Beekman, G., Cheung, S. L., & Levely, I. (2017). The effect of conflict history on cooperation within and between groups: Evidence from a laboratory experiment. *Journal of Economic Psychology, 63*, 168–183. https://doi.org/10.1016/j.joep.2017.02.004

Beersma, B., Hollenbeck, J. R., Conlon, D. E., Humphrey, S. E., Moon, H., & Ilgen, D. R. (2009). Cutthroat cooperation: The effects of team role decisions on adaptation to alternative reward structures. *Organizational Behavior and Human Decision Processes, 108*(1), 131–142. https://doi.org/10.1016/j.obhdp.2008.07.002

Bilali, R., & Vollhardt, J. R. (2019). Victim and perpetrator groups' divergent perspectives on collective violence: Implications for intergroup relations. *Political Psychology, 40*(S1), 75–108. https://doi.org/10.1111/pops.12570

Brandts, J., & Cooper, D. J. (2006). Observability and overcoming coordination failure in organizations: An experimental study. *Experimental Economics, 9*(4), 407–423. https://doi.org/10.1007/s10683-006-7056-5

Burton-Chellew, M. N., Nax, H. H., & West, S. A. (2015). Payoff-based learning explains the decline in cooperation in public goods games. *Proceedings of the Royal Society B: Biological Sciences, 282*(1801), 20142678. https://doi.org/10.1098/rspb.2014.2678

Cason, T., & Gangadharan, L. (2013). Cooperation spillovers and price competition in experimental markets. *Economic Inquiry, 51*(3), 1715–1730. https://doi.org/10.1111/j.1465-7295.2012.00486.x

Cason, T., Savikhin, A. C., & Sheremeta, R. M. (2012). Behavioral spillovers in coordination games. *European Economic Review, 56*(2), 233–245. https://doi.org/10.1016/j.euroecorev.2011.09.001

Cassar, A., D'Adda, G., & Grosjean, P. A. (2013). Institutional quality, culture, and norms of cooperation: Evidence from a behavioral field experiment. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2263989

Chakravarty, S., & Fonseca, M. A. (2014). The effect of social fragmentation on public good provision: An experimental study. *Journal of Behavioral and Experimental Economics, 53*, 1–9. https://doi.org/10.1016/j.socec.2014.07.002

Charness, G., & Chen, Y. (2020). Social identity, group behavior, and teams. *Annual Review of Economics, 12*(1), 691–713. https://doi.org/10.1146/annurev-economics-091619-032800

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. M., & Topolyan, I. (2016). The attack and defense group contests: Best shot versus weakest link. *Economic Inquiry, 54*(1), 548–557. https://doi.org/10.1111/ecin.12246

Cilliers, J., Dube, O., & Siddiqi, B. (2016). Reconciling after civil conflict increases social capital but decreases individual well-being. *Science, 352*(6287), 787–794. https://doi.org/10.1126/science.aad9682

Clark, D. J., & Konrad, K. A. (2007). Asymmetric conflict. *Journal of Conflict Resolution, 51*(3), 457–469. https://doi.org/10.1177/0022002707300320

Cushman, F., Sarin, A., & Ho, M. (2021). Punishment as communication. In J. Doris, & M. Vargas (Eds.), *Oxford Handbook of Moral Psychology*. Oxford University Press.

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*, Article e116. https://doi.org/10.1017/s0140525x18002170

De Dreu, C. K. W., Gross, J., Fariña, A., & Ma, Y. (2020). Group cooperation, carrying-capacity stress, and intergroup conflict. *Trends in Cognitive Sciences, 24*(9), 760–776. https://doi.org/10.1016/j.tics.2020.06.005

De Dreu, C. K. W., Gross, J., Méder, 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, 113* (38), 10524–10529. https://doi.org/10.1073/pnas.1605115113

De Dreu, C. K. W., Pliskin, R., Rojek-Giffin, M., Méder, Z., & Gross, J. (2021). Political games of attack and defence. *Philosophical Transactions of the Royal Society B, 376* (1822), 20200135. https://doi.org/10.1098/rstb.2020.0135

Dovidio, J. F., Gaertner, S. L., & Saguy, T. (2009). Commonality and the complexity of "We": Social attitudes and social change. *Personality and Social Psychology Review, 13* (1), 3–20. https://doi.org/10.1177/1088868308326751

De Dreu, C. K. W., & Carnevale, P. J. D. (2003). Motivational bases of information processing and strategy in conflict and negotiation. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (vol. 35, pp. 235–291). New York: Academic Press.

Duffy, J., & Kim, M. (2005). Anarchy in the laboratory (and the role of the state). *Journal of Economic Behavior & Organization, 56*(3), 297–329. https://doi.org/10.1016/j.jebo.2003.10.007

Dyrstad, K. (2012). After ethnic civil war: Ethno-nationalismin the Western Balkans. *Journal of Peace Research, 49*(6), 817–831. https://doi.org/10.1177/0022343312439202

Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences, 275*(1637), 871–878. https://doi.org/10.1098/rspb.2007.1558

Engelmann, D., & Nikiforakis, N. (2015). In the long-run we are all dead: On the benefits of peer punishment in rich environments. *Social Choice and Welfare, 45*(3), 561–577. https://doi.org/10.1007/s00355-015-0884-5

Faillo, M., Grieco, D., & Zarri, L. (2013). Legitimate punishment, feedback, and the enforcement of cooperation. *Games and Economic Behavior, 77*(1), 271–283. https://doi.org/10.1016/j.geb.2012.10.011

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences, 8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140. https://doi.org/10.1038/415137a

Fehr, E., & Williams, T. (2018). Social norms, endogenous sorting and the culture of cooperation. *SSRN Electronic Journal*.. https://doi.org/10.2139/ssrn.3062528

Fung, J. M. Y., & Au, W. (2014). Effect of inequality on cooperation: Heterogeneity and hegemony in public goods dilemma. *Organizational Behavior and Human Decision Processes, 123*(1), 9–22. https://doi.org/10.1016/j.obhdp.2013.10.010

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science, 322*(5907). https://doi.org/10.1126/science.1164744

Gaertner, S. L., Rust, M. C., Dovidio, J. F., Bachman, B. A., & Anastasio, P. A. (1994). The contact hypothesis: The role of a common ingroup identity on reducing intergroup

bias. *Small Group Research, 2*(25), 224–249. https://doi.org/10.1177/1046496494252005
Gat, A. (2019). Is war in our nature? *Human Nature, 30*(2), 149–154. https://doi.org/10.1007/s12110-019-09342-8
Glowacki, L., Isakov, A., Wrangham, R. W., McDermott, R., Fowler, J. H., & Christakis, N. A. (2016). Formation of raiding parties for intergroup violence is mediated by social network structure. *Proceedings of the National Academy of Sciences, 113*(43), 12114–12119. https://doi.org/10.1073/pnas.1610961113
Goldenberg, A., Cohen-Chen, S., Goyer, J. P., Dweck, C. S., Gross, J. J., & Halperin, E. (2018). Testing the impact and durability of a group malleability intervention in the context of the Israeli-Palestinian conflict. *Proceedings of the National Academy of Sciences, 115*(4), 696–701. https://doi.org/10.1073/pnas.1706800115
González, O. C., & Clémence, A. (2019). Concealing former identity to be accepted after the demobilization process in Colombia: A real reintegration in a post conflict scenario? *Journal of Social and Political Psychology, 7*(2), 941–958. https://doi.org/10.5964/jspp.v7i2.864
Gross, J., & Böhm, R. (2020). Voluntary restrictions on self-reliance increase cooperation and mitigate wealth inequality. *Proceedings of the National Academy of Sciences, 117*(46), 29202–29211. https://doi.org/10.1073/pnas.2013744117
Gross, J., & De Dreu, C. K. W. (2019). The rise and fall of cooperation through reputation and group polarization. *Nature Communications, 10*(1), 776. https://doi.org/10.1038/s41467-019-08727-8
Gross, J., Méder, Z. Z., Okamoto-Barth, S., & Riedl, A. (2016). Building the Leviathan – Voluntary centralisation of punishment power sustains cooperation in humans. *Scientific Reports, 6*(1), 20767. https://doi.org/10.1038/srep20767
Gross, J., Veistola, S., De Dreu, C. K. W., & Dijk, E. V. (2020). Self-reliance crowds out group cooperation and increases wealth inequality. *Nature Communications, 11*(1), 5161. https://doi.org/10.1038/s41467-020-18896-6
Gummerum, M., Takezawa, M., & Keller, M. (2009). The influence of social category and reciprocity on adults' and children's altruistic behavior. *Evolutionary Psychology, 7*(2). https://doi.org/10.1177/147470490900700212
Gürerk, Ö., Irlenbusch, B., & Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science, 312*(5770), 108–111. https://doi.org/10.1126/science.1123633
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
Halperin, E., Russell, A. G., Trzesniewski, K. H., Gross, J. J., & Dweck, C. S. (2011). Promoting the middle east peace process by changing beliefs about group malleability. *Science, 333*(6050), 1767–1769. https://doi.org/10.1126/science.1202925
Harrison, D. A., Price, K. H., & Bell, M. P. (1998). Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. *Academy of Management Journal, 41*(1), 96–107. https://doi.org/10.5465/256901
Hasan-Aslih, S., Shuman, E., Pliskin, R., Zomeren, M., Saguy, T., & Halperin, E. (2020). With or without you: The paradoxical role of identification in predicting joint and ingroup collective action in intergroup conflict. *European Journal of Social Psychology, 50*(6), 1334–1343. https://doi.org/10.1002/ejsp.2677
Henrich, J. (2006). Cooperation, punishment, and the evolution of human institutions. *Science, 312*(5770), 60–61. https://doi.org/10.1126/science.1126398
Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science, 319*(5868), 1362–1367. https://doi.org/10.1126/science.1153808
Ho, M. K., Cushman, F., Littman, M. L., & Austerweil, J. L. (2019). People teach with rewards and punishments as communication, not reinforcements. *Journal of Experimental Psychology: General, 148*(3), 520–549. https://doi.org/10.1037/xge0000569
Hornsey, M. J., & Hogg, M. A. (2000). Assimilation and diversity: An integrative model of subgroup relations. *Personality and Social Psychology Review, 4*(2), 143–156. https://doi.org/10.1207/s15327957pspr0402_03
Iacono, S. L., & Sonmez, B. (2020). The effect of trusting and trustworthy environments on the provision of public goods. *European Sociological Review, 37*(1), 155–168. https://doi.org/10.1093/esr/jcaa040
Johnson, M. D., Hollenbeck, J. R., Humphrey, S. E., Ilgen, D. R., Jundt, D., & Meyer, C. J. (2006). Cutthroat cooperation: Asymmetrical adaptation to changes in team reward structures. *Academy of Management Journal, 49*(1), 103–119. https://doi.org/10.5465/amj.2006.20785533
Ke, C., Konrad, K. A., & Morath, F. (2013). Brothers in arms – An experiment on the alliance puzzle. *Games and Economic Behavior, 77*(1), 61–76. https://doi.org/10.1016/j.geb.2012.08.011
Knez, M., & Camerer, C. (2000). Increasing cooperation in Prisoner's dilemmas by establishing a precedent of efficiency in coordination games. *Organizational Behavior and Human Decision Processes, 82*(2), 194–216. https://doi.org/10.1006/obhd.2000.2882
van Knippenberg, D., Dreu, C. K. W. D., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology, 89*(6), 1008–1022. https://doi.org/10.1037/0021-9010.89.6.1008
Martinangeli, A. F. M., & Martinsson, P. (2020). We, the rich: Inequality, identity and cooperation. *Journal of Economic Behavior & Organization, 178*, 249–266. https://doi.org/10.1016/j.jebo.2020.07.013
Masclet, D., Noussair, C., Tucker, S., & Villeval, M.-C. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review, 93*(1), 366–380. https://doi.org/10.1257/000282803321455359
McAuliffe, K., & Dunham, Y. (2016). Group bias in cooperative norm enforcement. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1686), 20150073. https://doi.org/10.1098/rstb.2015.0073
Mironova, V., & Whitt, S. (2016). The evolution of prosociality and parochialism after violence. *Journal of Peace Research, 53*(5), 648–664. https://doi.org/10.1177/0022343316648204
Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgement and Decision Making, 6*(8), 771–781. https://doi.org/10.2139/ssrn.1804189
Nadler, A., & Shnabel, N. (2015). Intergroup reconciliation: Instrumental and socio-emotional processes and the needs-based model. *European Review of Social Psychology, 26*(1), 93–125. https://doi.org/10.1080/10463283.2015.1106712
Nakashima, N. A., Halali, E., & Halevy, N. (2017). Third parties promote cooperative norms in repeated interactions. *Journal of Experimental Social Psychology, 68*, 212–223. https://doi.org/10.1016/j.jesp.2016.06.007
Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92*(1–2), 91–112. https://doi.org/10.1016/j.jpubeco.2007.04.008
Peysakhovich, A., & Rand, D. G. (2016). Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science, 62*(3), 631–647. https://doi.org/10.1287/mnsc.2015.2168
Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. *Advances in Experimental Social Psychology, 27*, 255–304. https://doi.org/10.1016/s0065-2601(08)60407-4
Rouhana, N. N., & Bar-Tal, D. (1998). Psychological dynamics of intractable ethnonational conflicts. *American Psychologist, 53*(7), 761–770. https://doi.org/10.1037/0003-066x.53.7.761
Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Intergroup conflict and cooperation: The robbers cave experiment*. University Book Exchange.
Shnabel, N., & Nadler, A. (2008). A needs-based model of reconciliation: Satisfying the differential emotional needs of victim and perpetrator as a key to promoting reconciliation. *Journal of Personality and Social Psychology, 94*(1), 116–132. https://doi.org/10.1037/0022-3514.94.1.116
Shnabel, N., & Nadler, A. (2015). The role of agency and morality in reconciliation processes. *Current Directions in Psychological Science, 24*(6), 477–483. https://doi.org/10.1177/0963721415601625
Stagnaro, M. N., Arechar, A. A., & Rand, D. G. (2017). From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement. *Cognition, 167*, 212–254. https://doi.org/10.1016/j.cognition.2017.01.017
Weber, J. M., Kopelman, S., & Messick, D. M. (2004). A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review, 8*(3), 281–307. https://doi.org/10.1207/s15327957pspr0803_4
Wright, T. M. (2014). Territorial revision and state repression. *Journal of Peace Research, 51*(3), 375–387. https://doi.org/10.1177/0022343314520822
Yamagishi, T. (1986). The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology, 51*(1), 110–116. https://doi.org/10.1037/0022-3514.51.1.110