

Social Psychological Bulletin

Intergroup Aggressors: When and Why Do They Acquire a Positive Reputation in the In-Group?

Kai Otsubo¹ , Hiroyuki Yamaguchi² 

[1] Graduate School of Human-Environment Studies, Kyushu University, Fukuoka, Japan. [2] Faculty of Human-Environment Studies, Kyushu University, Fukuoka, Japan.

Social Psychological Bulletin, 2026, Vol. 21, Article e18745, <https://doi.org/10.32872/spb.18745>

Received: 2025-07-07 • Accepted: 2026-02-25 • Published (VoR): 2026-09-02



Handling Editor: Tomasz Besta, University of Gdansk, Gdansk, Poland

Corresponding Author: Kai Otsubo, Graduate School of Humanities and Sociology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan. E-mail: k.otsubo686@gmail.com

Supplementary Materials: Code, Data, Materials [see [Index of Supplementary Materials](#)]



Abstract

It is necessary to understand the causal mechanism of aggressive behavior toward out-group members (out-group aggression) to prevent tragic intergroup conflicts. Recent research suggests that individuals sometimes engage in out-group aggression to maintain their reputation. However, it remains unclear how people actually evaluate those who engage in out-group aggression (intergroup aggressors). Across three experiments—a vignette-based study (Study 1) and two economic game studies (Studies 2, 3)—we examined how situational factors influence the reputation of intergroup aggressors. The results demonstrated that intergroup aggressors' reputation improved under conflicting interests between groups. This effect was caused by the fact that intergroup aggressors in such situations benefit their in-group, rather than simply by harming the out-group. In addition, the perceived prosocial motivation of intergroup aggressors was shown to be an important determinant of their reputation. Our findings are consistent with the idea that reputation plays an important role in the emergence and escalation of intergroup conflict and provide important implications for intergroup conflict prevention.

Keywords

intergroup conflict, out-group aggression, reputation, social dilemma, cooperation



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Highlights

- Previous literature has suggested that out-group aggression can serve as a strategy for maintaining an aggressor's reputation, but how people actually evaluate those who engage in such aggression remains underexplored.
- The present research revealed that people evaluate intergroup aggressors more favorably when there are conflicting interests between the in-group and the out-group.
- This effect occurs *only when* out-group aggression benefits the in-group.
- Participants' intergroup aggression was positively correlated with their favorable attitudes toward aggressors, whereas their within-group cooperation showed no correlation with such attitudes.

In intergroup conflict, members of opposing groups often exhibit aggressive or competitive behavior toward each other, defined as out-group aggression (De Dreu et al., 2016; Mifune, 2022). Out-group aggression is a core element of intergroup conflict and can trigger its infinite escalation. Illuminating the causal mechanism of out-group aggression has been one of the most essential missions for social scientists. So far, numerous studies have investigated various factors that influence the emergence of out-group aggression. For instance, perception of competitive intergroup relationships (Jing et al., 2017; Sherif et al., 1961), extreme identification with the in-group (Chinchilla et al., 2022), emotions such as anger and fear (Spanovic et al., 2010), out-group threat (Böhm et al., 2016), and unstable environments (De Dreu, Gross, et al., 2022) have been proposed as factors associated with out-group aggression. In the present research, we focus on reputation as one of these factors, and aim to address a previously underexplored question: *How are those who engage in out-group aggression evaluated by other in-group members?*

Reputation and Out-Group Aggression

Reputation is defined as “that set of judgments a community makes about the personal qualities of one of its members” (Emler, 1990, p. 171). Reputation plays a crucial role in social life, as it influences whether people choose to cooperate with an individual (e.g., Milinski et al., 2002), and shapes that person's position within the social hierarchy (i.e., status; Van Vugt & Tybur, 2015).

Some recent literature has highlighted the importance of status and reputation as drivers of out-group aggression (De Dreu, Fariña, et al., 2022; De Dreu & Triki, 2022; Glowacki & Wrangham, 2013). De Dreu and Triki (2022) listed non-material benefits such as status and positive reputation among in-group members as one of the five predictors of participation in intergroup conflict, along with material benefits, social preferences for in-group members, social preferences for out-group members, and perceived threat of out-group. Indeed, a variety of research suggests that individuals sometimes aggress

out-group members to maintain their reputation in the in-group. [Nawata et al. \(2016\)](#) demonstrated that Chinese participants' intention to express anti-Japanese attitudes in public places was affected by their perception of other Chinese people's anti-Japanese attitudes. Furthermore, some anthropological evidence revealed a positive relationship between social rewards (e.g., status) to warriors within groups and the frequency ([Nawata, 2020](#)) and mortality rate ([Glowacki & Wrangham, 2013](#)) of intergroup conflict engaged in by the group. These studies consistently suggest that reputation plays an important role in the emergence of out-group aggression in intergroup conflicts.

However, it is not self-evident that out-group aggression can serve as the aggressor's reputation maintenance strategy in the in-group. People generally dislike those who exhibit aggressive behaviors ([Kiyonari & Barclay, 2008](#); [Niimi & Goto, 2023](#); [Rocca & McCroskey, 1999](#)), probably because they are concerned that aggressors will turn their attacks on them. Thus, those who exhibit out-group aggression (i.e., intergroup aggressors) are predicted to be negatively evaluated by others. Indeed, [Halevy et al. \(2012\)](#) showed that intergroup aggressors in an economic game more easily became targets of exclusion than within-group cooperators. Given that, specific situational conditions should be required for intergroup aggressors to obtain a good reputation.

The Current Study

We aim to investigate how situational factors affect the reputation of intergroup aggressors. Elucidating under what situations people confer a positive reputation upon intergroup aggressors can give us a hint about when reputation-induced out-group aggression emerges. Research has shown that people strategically decide whether or not to engage in cooperative ([Barclay & Willer, 2007](#); [Milinski et al., 2002](#); [Semmann et al., 2004](#)) and aggressive ([Benard, 2015](#)) behavior considering expected reputations. Thus, people will likely exhibit reputation-induced out-group aggression in situations where out-group aggression actually brings reputational benefit. Accordingly, examining situational factors that influence the reputation of intergroup aggressors will lay the groundwork for investigating boundary conditions of such out-group aggression in an experimental setting.

Predictions

First, we hypothesized that people evaluate intergroup aggressors more favorably when conflicting interests between the in-group and out-group exist (H1). Although aggressive people are generally disliked, individuals who exhibit aggression toward their enemies in intergroup conflict are sometimes praised as "war heroes." Indeed, prominent warriors in pre-state ([Chagnon, 1988](#); [Glowacki & Wrangham, 2015](#)) and modern ([Rusch et al., 2015](#)) warfare gained positive reputations in their in-groups. One possible explanation is that intergroup conflict has a social dilemma structure ([Bornstein, 1992](#)). Although each

group's chance of winning the conflict increases as the number of participants increases, participation in intergroup warfare is costly for individuals. In addition, some resources earned as a result of conflict (e.g., safety) are distributed equally to group members (i.e., public goods). Thus, individuals can maximize their interest by freeriding on other members' contributions. Out-group aggression (participation in the conflict) in such a situation is an altruistic behavior that increases the benefits of other in-group members. Given that such self-sacrifice in social dilemmas is evaluated favorably (Willer, 2009), we inferred that intergroup aggressors under conflicts between groups would acquire a good reputation.

Second, we predicted that people evaluate intergroup aggressors more favorably when aggressors do not have the option of within-group cooperation (H2). Given the unfavorable characteristic of out-group aggression as an aggressive behavior, it is plausible that the perception that the individual *willingly* engages in out-group aggression can dampen their reputation because it highlights their aggressive traits. Rather, intergroup aggressors may acquire better reputations when the aggression is perceived to be a "last resort" to help the in-group. For example, imagine a situation in which an imminent threat is posed by out-group members, when a police officer discovers that a gang is attempting to raid their police station. In such a situation, the officer would not be perceived as having an aggressive personality even if they fought back, because they had no other option to defend their colleagues. Indeed, using economic game paradigms, Halevy et al. (2012) showed that participants evaluated intergroup aggressors more favorably when the within-group cooperation option was not available than when it was available. Furthermore, in the context of altruistic punishment, a form of aggression against non-cooperators, some researchers argue that not having behavioral options that can signal cooperative intentions other than altruistic punishment is crucial for punishers to obtain a good reputation (Ozono & Watabe, 2012; Raihani & Bshary, 2015). Thus, intergroup aggressors may acquire a better reputation when there is no option of within-group cooperation.

Lastly, we hypothesized that perceived prosocial in-group motivation of intergroup aggressors mediates the effects of the above-mentioned situational factors (H3). Past studies have demonstrated that aggressors are evaluated favorably when they have a prosocial motivation (Nesdale et al., 1975; Rule et al., 1974). There is a similar tendency regarding altruistic punishers (Tateishi et al., 2021). Moreover, contributors to public goods in social dilemmas acquire a positive reputation through the estimation of pro-group motivation (Willer, 2009). These results indicate that perceived prosocial motivation directed toward the aggressor's in-group is crucial for the reputations of intergroup aggressors.

Study 1

Objectives and Experimental Design

In Study 1, we aimed to examine the hypotheses using experimental vignettes. The experiment used a 2 x 2 x 2 between-subjects design. We manipulated conflicting interests (yes/no), an option of within-group cooperation (yes/no), and the type of scenario (Scenario 1/2). Regarding the type of scenario condition, we used two scenarios depicting different real-life group contexts—a workplace setting (Scenario 1) and a sports team setting (Scenario 2)—to examine the generalizability of the results. The study protocol was pre-registered on the Open Science Framework (see [Otsubo, 2023a](#)). The experiment was conducted after being approved by the ethical committee of Kyushu University [approval no. 2023-017]. Informed consent was obtained from all participants prior to their participation in the experiment.

Hypotheses

In Study 1, H1–3 were operationalized as follows:

H1a: Conflicting interests between groups make the evaluation of intergroup aggressors more positive.

H1b: Conflicting interests between groups heighten the willingness of trust and cooperative behavior toward intergroup aggressors.

H2a: Not having an option of within-group cooperation makes the evaluation of intergroup aggressors more positive.

H2b: Not having an option of within-group cooperation heightens the willingness of trust and cooperative behavior toward intergroup aggressors.

H3: Perceived prosocial motivation mediates the effects of H1a–H2b.

Participants

We recruited 1,009 adult native Japanese speakers as participants through a Japanese crowdsourcing platform, Yahoo! crowdsourcing (<https://crowdsourcing.yahoo.co.jp/>). Respondents received 30 Japanese yen (approximately 0.2 US dollars) for their participation. We determined the sample size based on a power analysis and a previous study. Regarding H1 and H2, the analysis using G*Power 3.1 ([Faul et al., 2009](#)) showed that 787 participants were needed to detect small effects ($f = 0.1$) with $1 - \beta = .80$ and $\alpha = .05$ in three-way analyses of variance (ANOVAs). Regarding H3, the previous study showed that 148 participants were needed to detect small to medium indirect effects (indirect effect = .26 * .26) with $1 - \beta = .80$, $\alpha = .05$ in the mediation analysis ([Fritz & MacKinnon, 2007](#)). Although we recruited 900 participants at first, we added 100 more because the number of valid samples was below the target sample size ($N = 787$). This procedure follows the pre-registration.

Procedure

The experiment was conducted online. We created the experimental program using jsPsych (de Leeuw, 2015). Participants were randomly assigned to one of the eight conditions. After providing informed consent and reporting demographic information, they read a scenario depicting out-group aggression in real-life situations. The contents of the scenario differed depending on the condition. We manipulated (a) conflicting interests between the in-group and out-group and (b) availability of an option for within-group cooperation.

We used two types of scenarios, Scenario 1 and Scenario 2, to verify the generalizability of the results across situations. In Scenario 1, participants imagined that they were working at a fictitious company, A, and that their colleague named P decided to sue another company, B, to disturb their product development. The relationship between the two companies and P's options differed depending on the condition. In the conflict condition, the two companies competed in the development race, while there was no existing relationship between them in the no-conflict condition. In the cooperation-available condition, P chose to sue Company B from three options—(1) doing nothing (working as usual), (2) developing a new system to improve efficiency in Company A, and (3) suing Company B to delay Company B's development—while choosing to do so from among two options without option (2) in the cooperation-unavailable condition. The scenario read that the amount of Company A's gain from the development of a new system and Company B's damage from the lawsuit were almost the same in the cooperation-available condition. Scenario 2 presented a similar situation in a different context: a university soccer team. Details of the scenarios and questionnaires are provided in the Supporting Information (see Otsubo & Yamaguchi, 2026).

Afterward, participants answered two comprehension checks about the scenario and estimated P's prosocial motivation (three items: e.g., "I think P acted for their fellows"; 7-point scale: "1. Strongly disagree" to "7. Strongly agree"). They then rated their evaluation of P (six items: e.g., "I think P is trustworthy"; 9-point scale: "1. Strongly disagree" to "9. Strongly Agree"; Kiyonari & Barclay, 2008) and willingness to trust and cooperate with P (referred to as trust/cooperation; six items: e.g., "I want to cooperate with P when I work with P in the same group"; 7-point scale: "1. Strongly disagree" to "7. Strongly agree"; Tateishi et al., 2021). Finally, we asked participants to imagine another individual, Q, who decided to do nothing in the same situation and report their evaluation and trust/cooperation of Q.

Data Analysis

We used R version 4.3.1 (R Core Team, 2023) for the analyses. The R packages "psych" (Revelle, 2023) and "mediation" (Tingley et al., 2014) were used for calculating Cronba-

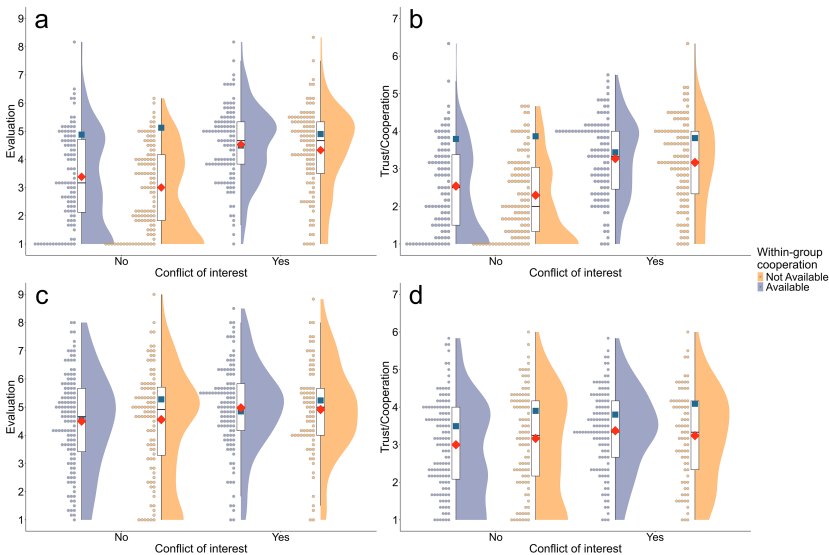
ch’s α and testing indirect effects in mediation analyses, respectively. We used the “anovakun” package (Iseki, 2023) for performing ANOVAs and calculating generalized η^2 .

Data Handling

We excluded data from (1) 31 participants who failed any of the attention check items (Directed Questions Scale: DQS; Maniaci & Rogge, 2014), (2) 133 participants who incorrectly answered any of the comprehension checks about the scenario, and (3) seven participants who reported that they had participated in the same experiment before. Thus, the final sample size was 838 ($M_{\text{age}} = 50.28$; 624 men, 202 women, one other, 11 preferred not to answer).

After confirming the internal consistency, responses to question items for each dependent variable were aggregated¹. Table 1 shows the means and standard deviations of measured variables for each condition. Figure 1 shows distributions of the evaluation (a and c) and trust/cooperation (b and d) of intergroup aggressors in Scenarios 1 (a and b) and 2 (c and d).

Figure 1
Distribution of the Intergroup Aggressors’ Reputation in Study 1



Note. The upper row shows distributions of evaluation (a) and trust/cooperation (b) in Scenario 1. The lower row shows distributions of evaluation (c) and trust/cooperation (d) in Scenario 2. Red diamonds indicate means. Blue squares, corresponding to evaluations of non-aggressors (Q), are included for comparison.

1) See Supporting Information (see Otsubo & Yamaguchi, 2026) for the reliability coefficients.

Table 1
Means (Standard Deviations) of Measured Variables in Study 1

Conflicting Interests	Scenario 1				Scenario 2			
	Yes		No		Yes		No	
	Available	Not Available	Available	Not Available	Available	Not Available	Available	Not Available
Within-Group Cooperation	<i>n</i> = 108 3.98 (1.11)	<i>n</i> = 108 3.86 (1.18)	<i>n</i> = 100 2.83 (1.32)	<i>n</i> = 112 2.36 (1.12)	<i>n</i> = 113 4.60 (1.31)	<i>n</i> = 78 4.49 (1.20)	<i>n</i> = 123 4.06 (1.23)	<i>n</i> = 96 3.90 (1.45)
Perceived Prosocial Motivation of P (Intergroup Aggressor)	4.53 (1.33)	4.33 (1.49)	3.38 (1.64)	3.00 (1.47)	4.97 (1.54)	4.91 (1.64)	4.50 (1.68)	4.56 (1.85)
Trust/Cooperation of P	3.27 (1.05)	3.17 (1.14)	2.54 (1.18)	2.30 (1.07)	3.37 (1.09)	3.24 (1.22)	3.00 (1.21)	3.17 (1.31)
Evaluation of Q (Non-Aggressor)	4.49 (1.23)	4.90 (1.15)	4.88 (1.15)	5.13 (0.89)	4.85 (1.31)	5.24 (1.09)	4.56 (1.30)	5.28 (0.99)
Trust/Cooperation of Q	3.44 (0.92)	3.82 (1.01)	3.80 (0.95)	3.87 (0.84)	3.80 (0.96)	4.09 (0.98)	3.49 (1.06)	3.90 (0.87)

Results and Discussion

To test H1a-H2b, we conducted a three-way ANOVA on the evaluation and trust/cooperation with conflicting interests, availability of within-group cooperation, and the type of scenario as independent variables. The results showed significant main effects of conflicting interests (Evaluation: $F(1, 830) = 56.00, p < .001$, generalized $\eta^2 = .063$, 95% CI [.034, .096]; Trust/cooperation: $F(1, 830) = 40.52, p < .001$, generalized $\eta^2 = .047$, 95% CI [.022, .077]) and type of scenario (Evaluation: $F(1, 830) = 70.67, p < .001$, generalized $\eta^2 = .078$, 95% CI [.045, .118]; Trust/cooperation: $F(1, 830) = 21.64, p < .001$, generalized $\eta^2 = .025$, 95% CI [.009, .052]), and their interaction (Evaluation: $F(1, 830) = 13.83, p < .001$, generalized $\eta^2 = .016$, 95% CI [.004, .037]; Trust/cooperation: $F(1, 830) = 12.86, p < .001$, generalized $\eta^2 = .015$, 95% CI [.003, .037]). Intergroup aggressors were evaluated more positively in the conflict condition than in the no-conflict condition, and more positively in Scenario 2 than in Scenario 1. Regarding the interaction effect, post-hoc analyses demonstrated that the effects of conflicting interests were (marginally) significant both in Scenario 1 (Evaluation: $F(1, 830) = 65.06, p < .001$, generalized $\eta^2 = .073$, 95% CI [.042, .103]; Trust/cooperation: $F(1, 830) = 51.35, p < .001$, generalized $\eta^2 = .058$, 95% CI [.030, .089]) and in Scenario 2 (Evaluation: $F(1, 830) = 6.84, p = .009$, generalized $\eta^2 = .008$, 95% CI [.001, .027]; Trust/cooperation: $F(1, 830) = 3.73, p = .054$, generalized $\eta^2 = .004$, 95% CI [.000, .020]). In sum, participants evaluated intergroup aggressors more positively when there were conflicting interests between the in-group and out-group, but the magnitude of this effect differed depending on the situation. In contrast, the availability of cooperative options did not affect their evaluation. These results mostly supported H1a and H1b but not H2a or H2b.

To test H3, we conducted mediation analyses with conflicting interests as an independent variable, the evaluation and trust/cooperation of intergroup aggressors as dependent variables, and perceived prosocial motivation of intergroup aggressors as a mediator. We adopted the bias-corrected accelerated (BCa) bootstrap method with 10,000 resamples. The results demonstrated that the effects of conflicting interests on the evaluation (Scenario 1: indirect effect = .37, 95% CI [.30, .43], $p < .001$; Scenario 2: indirect effect = .16, 95% CI [.09, .24], $p < .001$) and on the trust/cooperation (Scenario 1: indirect effect = .32, 95% CI [.26, .39], $p < .001$; Scenario 2: indirect effect = .13, 95% CI [.07, .20], $p < .001$) were completely mediated by their perceived prosocial motivation (Figure S1). As H2a and H2b were not supported, we did not investigate the mediating effect of perceived prosocial motivation on the effect of the cooperative option. Thus, H3 was partly supported.²

In a vignette experiment (Study 1), we examined the effects of conflicting interests between groups and the availability of within-group cooperation on the reputations

2) We report the results of exploratory analyses focusing on gender differences (see Otsubo & Yamaguchi, 2026).

of intergroup aggressors. Participants evaluated intergroup aggressors more positively and were more willing to trust and cooperate with them when conflicting interests existed (i.e., intergroup conflict). The effects of conflicting interests were completely mediated by perceived prosocial motivation. In intergroup conflicts, people are more likely to perceive prosocial motivation of intergroup aggressors and intend to treat them favorably. However, the availability of the option of within-group cooperation did not influence the reputations of intergroup aggressors. Even when one harms out-group members with no option of within-group cooperation, their in-group members do not think that they did so to help in-group members. This result seems inconsistent with the argument that altruistic punishers are evaluated more favorably when they cannot signal their cooperative intention other than through punishment (Ozono & Watabe, 2012; Raihani & Bshary, 2015). This may be because there are some qualitative differences between altruistic punishment and out-group aggression. Out-group aggression might be perceived to be a stronger manifestation of competitiveness, as its target is an ordinary member of the out-group, not a defector. Another possibility is that our manipulation of the availability of the cooperative option was not fully effective. Although we explained that out-group aggression and within-group cooperation would have the same effect—that cooperation could serve as an alternative to out-group aggression—it is still possible that participants did not perceive this as sufficiently realistic. In addition, we found a significant main effect of scenario type and a significant interaction between scenario type and the presence of conflicting interests. Intergroup aggressors were generally evaluated more positively, and the effect of conflicting interests on reputations was weaker, in the sports match context than in the business context. This pattern may reflect the fact that out-group aggression is more socially acceptable in sports matches. These results indicate that the reputations of intergroup aggressors also depend on the social norms in the specific context in which the aggression occurs.

Study 1 has an important limitation: we cannot identify what caused the effect of conflicting interests. The results of Study 1 indicate that intergroup aggressors gained more positive reputations under intergroup conflict. However, as we discussed in the rationale for H1, intergroup aggressors also acted as within-group cooperators in the conflict condition. Put differently, intergroup aggressors in the conflict condition benefited their in-group members by engaging in out-group aggression. For example, the intergroup aggressor in the conflict condition of Scenario 1, who decided to sue the rival company, increased their own company's chances of winning the competition. Therefore, we cannot distinguish whether intergroup aggressors were evaluated favorably under intergroup conflict because they harmed out-group members or because they benefited their in-group members. Conducting an experiment that can distinguish these two possible explanations is needed.

Study 2

Objectives and Experimental Design

Do intergroup aggressors acquire a more positive reputation under intergroup conflict even if the aggression does not benefit their in-group members? Study 2 aimed to answer this question. By using an experimental game, we separated out-group aggression from within-group cooperation. In this setting, we compared the reputation of intergroup aggressors who were not cooperators to the in-group in situations where conflicting interests between groups are present and absent. We call the condition with and without conflicting interests the conflict-non-cooperative (CN) condition and the no-conflict (N) condition, respectively. The experiment utilized a one-factor between-participants design. Participants were randomly assigned to either the CN or the N conditions. We pre-registered the study protocol on the Open Science Framework (see [Otsubo, 2023b](#)). The experiment was conducted after being approved by the ethical committee of Kyushu University [approval no. 2023-025]. Informed consent was obtained from all participants prior to their participation in the experiment.

Hypotheses

The hypotheses for Study 2 were as follows:

H1a: Conflicting interests between groups make the evaluation of intergroup aggressors more positive.

H1b: Conflicting interests between groups heighten the willingness of trust and cooperative behavior toward intergroup aggressors.

H2: Perceived prosocial motivation mediates the effects of H1a and H1b.

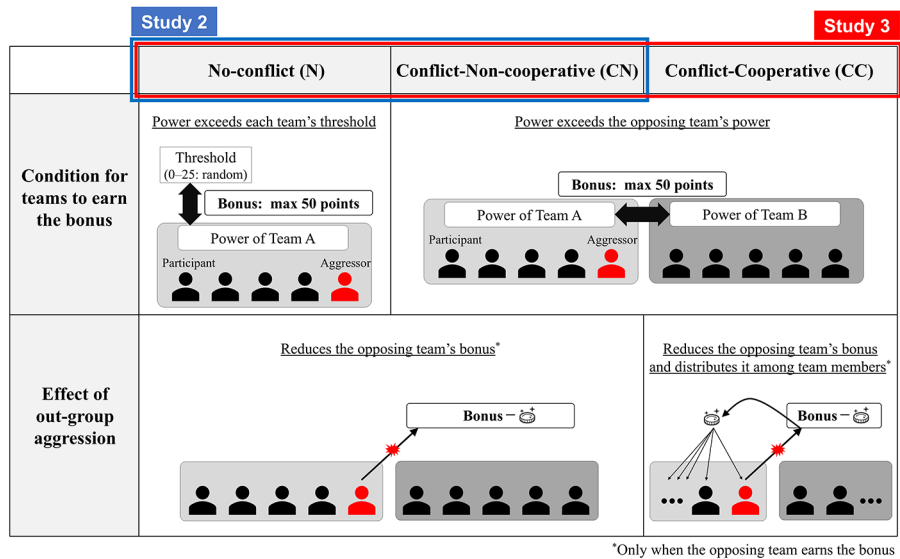
Participants

The participants were 1,002 Japanese adults recruited from Yahoo! Crowdsourcing (<https://crowdsourcing.yahoo.co.jp/>). They received 30 yen (approximately USD 0.20) as a participation fee, and they could earn extra rewards of 20 to 100 yen (approximately USD 0.13 to 0.67) depending on the results of the experiment. The sample size was determined based on a power analysis using G*Power 3.1 (Faul et al., 2009). The analysis showed that 788 participants were needed to detect small differences ($d = 0.2$) between two independent samples with a power of .80 under $\alpha = .05$. Although we stated that we would not recruit over 900 participants in the pre-registration, we recruited 950 participants in total because we modified the program after collecting data from the first 50 participants. We did not include their data in the analysis.

Procedure

The online experiment was programmed using oTree (Chen et al., 2016). We developed a new team game paradigm (Figure 2). In the game, 10 players participated in one session and were divided into two teams of five (teams A and B). Each team had to earn as much “power” as possible. They had an opportunity to receive a bonus point depending on the power amount. In the CN condition, the bonus was a zero-sum resource: the team that had accumulated more power received a bonus of 50 points while the opposing team received 0 points. In the case of a tie, both received 25 points. In the N condition, the bonus was not a zero-sum resource: the team that earned more power than the pre-determined threshold received a bonus of 50 points. Each team’s threshold was randomly set between zero and 20.³ If the power amount was the same as the threshold, the team received a bonus of 25 points.

Figure 2
Overview of the Game Settings in Studies 2 and 3



Notes. All three conditions were used in Study 3, while only the N and CN conditions were used in Study 2.

In the game, players first received an endowment of 10 points (1 point worth JPY 5 or USD 0.03) and had three options regarding how to allocate it: (1) within-group

3) Due to the behavior of Bot N, the highest possible value of Power was 20. Therefore, although we instructed participants that the threshold was randomly set in the range 0–25, in practice we set the maximum threshold to 20.

cooperation, (2) out-group aggression, and (3) keep. Players were not allowed to allocate more than 5 points to each of (1) within-group cooperation and (2) out-group aggression. For every point allocated to the within-group cooperation option, the power of their group increased by 1. For every point distributed to the out-group aggression option, the amount of the bonus for the opposing team decreased by 1 (only if the opposing team earned the bonus). Participants were also allowed to keep their endowment (up to 10 points) for their personal reward. We used neutral terms when explaining these rules.⁴

Under these settings, out-group aggression was separated from within-group cooperation. This was because players were not able to increase their teams' chance of winning by allocating to the out-group aggression option, which only harmed the opposing team, and because the 5-point caps on the aggression and cooperation options made the two types of behaviors completely independent (players could simultaneously be maximally cooperative and aggressive, if they wished).

Participants were told that the 10 players were all real participants. In fact, four bots (two in each team) participated in each session, and the number of real participants was six. There were two types of bots: A and N. Bot A was an intergroup aggressor, who always allocated 5 points to the option of out-group aggression. How Bot A allocated points to the other two options was randomly determined. Bot N was a non-aggressor, who always allocated all 10 points to the "keep" option.

The experiment was divided into the game and evaluation phases. In the game phase, participants provided informed consent and then played a single round of this game after two practice rounds and rule-check quizzes. They were anonymized and could not interact with each other during the experiment. In the evaluation phase, participants evaluated two teammates in the game based on the information about their decisions. Two teammates were not actual participants but Bots A and N. While participants rated them, information about their decisions in the game was displayed at the top of the screen. Specifically, when participants rated Bot A, a screenshot of the decision-making part checking the out-group aggression option as "5 points" was displayed. Bot A's distributions to the other two options were not indicated. When they rated Bot N, a screenshot showing an allocation of 10 points to the "keep" option was displayed. Dependent variables and question items were the same as those used in Study 1. After the evaluation phase, participants were debriefed and thanked for their participation.

Data Analysis

We used R version 4.3.1 (R Core Team, 2023) in the analyses. We used the "psych" (Revelle, 2023) and "effectsize" (Ben-Shachar et al., 2020) packages for calculating Cronbach's α and Cohen's d , respectively.

4) Details on the instructions are available in the Supporting Information (see Otsubo & Yamaguchi, 2026).

Data Handling

Data obtained from (1) 121 participants who did not complete the task, (2) 45 participants who failed any of the DQS items (Maniaci & Rogge, 2014), (3) 552 participants who answered any of the rule-check quizzes incorrectly, and (4) 50 participants who reported that they were completely aware that two persons whom they evaluated in the evaluation phase were not real participants were excluded from the analyses. The final sample size was 234 ($M_{age} = 47.91$, 173 men, 59 women, two unknown). These exclusion criteria were all pre-registered.

We aggregated three items for the prosocial motivation of Bot A, six items each for the evaluation and trust/cooperation of Bots A and N before the analyses¹. Tables 2 and 3 show the means and standard deviations (Table 2) and correlation matrix (Table 3) of measured variables for each condition. Figure S2 shows the distribution of Bot A's evaluation (a) and trust/cooperation (b).

Table 2

Means (Standard Deviations) of Measured Variables in Each Condition in Study 2

	CN condition	N condition
	<i>n</i> = 141	<i>n</i> = 93
Perceived Prosocial Motivation of Bot A(Intergroup Aggressor)	4.26 (1.72)	4.40 (1.80)
Evaluation of Bot A	5.94 (2.10)	5.94 (2.21)
Trust/Cooperation of Bot A	4.09 (1.56)	4.03 (1.47)
Evaluation of Bot N (Non-Aggressor)	2.17 (1.28)	2.39 (1.43)
Trust/Cooperation Bot N	1.87 (1.03)	1.99 (1.06)
Within-Group Cooperation	3.67 (1.81)	3.33 (1.82)
Out-Group Aggression	1.66 (1.86)	1.41 (1.75)
Keep	4.67 (2.77)	5.26 (2.80)

Note. The bottom three rows show participants' decisions in the game.

Table 3

Correlation Matrix of Measured Variables in Each Condition in Study 2

	1	2	3	4	5	6
1. Out-Group Aggression	—	.23*	.50***	.50***	-.02	-.01
2. Within-Group Cooperation	.14	—	-.02	-.05	-.47***	-.47***
3. Evaluation of Bot A	.36***	.14	—	.86***	-.04	-.06
4. Trust/Cooperation of Bot A	.36***	.03	.85***	—	.08	.09
5. Evaluation of Bot N	.06	-.30***	.07	.07	—	.92***
6. Trust/Cooperation of Bot N	-.05	-.39***	-.09	.00	.78***	—

Note. Coefficients in the CN condition are on the lower side, those in the N condition are on the upper side.

p* < .05. **p* < .001.

Results and Discussion

To test H1a and H1b, we performed Welch's *t*-tests on the means of the evaluation (H1a) and trust/cooperation (H1b) of Bot A. The results showed no significant difference in the evaluation ($t(190.23) = -0.01, p = .995$, Cohen's $d = 0.00$, 95% CI $[-0.26, 0.26]$) and trust/cooperation ($t(204.49) = -0.26, p = .793$, Cohen's $d = -0.03$, 95% CI $[-0.30, 0.23]$) between the two conditions. We did not perform any analyses testing H2 because the results did not support H1a and H1b.

In Study 2, we examined whether conflicting interests between in-group and out-group enhance the reputation of intergroup aggressors, even when out-group aggression does not benefit in-group members. We separated out-group aggression from within-group cooperation by developing a new team game paradigm. The results demonstrated that the reputation of intergroup aggressors did not differ between the CN and N conditions. This suggests that intergroup aggressors do not gain better reputations in intergroup conflict over zero-sum resources if their actions do not benefit the in-groups. The more favorable reputations of intergroup aggressors observed in Study 1 under intergroup conflict may be attributable to the fact that their behavior benefited the in-group.

Although it was beyond our primary interest, we also identified an interesting pattern in the correlations between participants' behavior and their evaluations of intergroup aggressors. While evaluations of intergroup aggressors were positively correlated with the monetary unit spent on the out-group aggression option, they were not correlated with those spent on the in-group cooperation option (Table 3). In other words, intergroup aggressors tended to favor other intergroup aggressors, and within-group cooperators appeared to tacitly approve of intergroup aggressors, even when the aggression did not benefit the in-group. This finding has important implications for understanding who grants reputational benefits to intergroup aggressors.

Several limitations of Study 2 should also be considered. Due to the complexity of the game paradigm, an unexpectedly large proportion of participants (more than half) answered the rule-check quizzes incorrectly, resulting in insufficient statistical power. More importantly, we did not manipulate whether out-group aggression provided benefits to in-group members. Therefore, the idea that intergroup aggressors gain better reputations under intergroup conflict because they contribute to their in-groups remains a suggestion rather than a confirmed finding. A follow-up study that simultaneously manipulates the presence of conflicting interests and the extent to which out-group aggression benefits the in-group is needed to test this notion.

Study 3

Objectives and Experimental Design

Study 3 aimed to test the idea, suggested by Studies 1 and 2, that intergroup aggressors acquire a better reputation under intergroup conflict when their actions benefit their in-groups. Specifically, we introduced a new condition—the conflict-cooperative (CC) condition—alongside the two conditions used in Study 2 (CN and N conditions). In the CC condition, intergroup aggressors both harm the out-group and provide benefits to the in-group. We predicted that the reputation of intergroup aggressors would be more positive in the CC condition than in the CN and N conditions. The experiment employed a one-factor between-participants design. Participants were randomly assigned to one of the three conditions (CC, CN, or N). The study protocol was pre-registered on the Open Science Framework prior to data collection (see [Otsubo, 2025](#)). The experiment was conducted after being approved by the ethical committee of The University of Tokyo [approval no. UTSP-24050]. Informed consent was obtained from all participants prior to their participation in the experiment.

Hypotheses

H1a: The evaluation of intergroup aggressors is more positive in the CC condition than in the N condition.

H1b: The willingness of trust and cooperative behavior toward intergroup aggressors is higher in the CC condition than in the N condition.

H2a: The evaluation of intergroup aggressors is more positive in the CC condition than in the CN condition.

H2b: The willingness of trust and cooperative behavior toward intergroup aggressors is higher in the CC condition than in the CN condition.

H3: Perceived prosocial motivation mediates the effects of H1a–H2b.

Participants

The participants were 402 Japanese adults recruited through a Japanese crowdsourcing platform, Lancers (<https://www.lancers.jp/>). They received 143 yen (approximately USD 0.98) for their participation. The sample size was determined based on a power analysis conducted using G*Power 3.1 ([Faul et al., 2009](#)) and informed by a previous study. The analysis showed that 246 participants were required to detect a small to medium main effect ($f = .20$) in ANOVA and that 363 participants in total were needed to detect a small to medium-sized difference ($d = 0.4$) in an independent sample t -test with a Bonferroni-corrected alpha. Additionally, a previous study ([Fritz & MacKinnon, 2007](#)) indicated that 222 participants in total were needed to detect a small to medium indirect

effect (indirect effect = $.26 \times .26$) with $1-\beta = .80$ and $\alpha = .05$ in a mediation analysis. Based on these estimates, we aimed to collect data from 400 participants.

Procedure

The online experiment was programmed using jsPsych (de Leeuw, 2015) and employed the same game paradigm as Study 2 (Figure 2). Several changes were made from Study 2. First, as noted above, a new condition (CC condition) was added to the two conditions used in Study 2 (CN and N). In the CC condition, the requirement for teams to earn the bonus was the same as in the CN condition, which was that the team's power had to exceed that of the opposing team. However, the effect of out-group aggression differed from the other two conditions. In the CC condition, for each point allocated to the out-group aggression option, the possible bonus for the opposing team decreased by 1 point. This reduction was transferred to the aggressor's own team and distributed equally among team members. For example, if members of the losing team allocated a total of five points to the out-group aggression option, the winning team's bonus would be reduced by five points, resulting in 45 points. The deducted five points would then be equally distributed among the losing team members, with each receiving one point. Under this setup, out-group aggression had a cooperative aspect: it not only reduced the bonus of the opposing group but also increased the payoff for the aggressor's own group. Second, Study 3 employed a vignette-based method. Unlike Study 2, participants in Study 3 did not actually play the game; instead, they read the instructions of the game, made hypothetical decisions, and evaluated fictitious players. Third, we improved the instructions and rule-check quizzes. As a result, the proportion of participants who answered the rule-check quizzes incorrectly was substantially lower in Study 3 (13.2%) compared to Study 2 (55.1%).

After providing informed consent, participants read the game instructions and answered two rule-check quizzes. They then reported their hypothetical decisions as if they were actually playing the game. Next, participants responded to questions about two fictitious players, P and Q, based on descriptions of their behavior. P corresponded to Bot A in Study 2 and was described as an intergroup aggressor who allocated five points to the out-group aggression option. Q corresponded to Bot N and was described as a non-aggressor who allocated ten points to the keep option. As in Study 2, P's allocations to the other two options were not specified. Participants first rated P's perceived prosocial motivation, overall evaluation, and trust/cooperation, and estimated P's allocation to the within-group cooperation option. Finally, they rated Q's evaluation and trust/cooperation, and were thanked for their participation.

Data Analysis

All analyses were conducted using R version 4.3.3 (R Core Team, 2024). We used the “psych” (Revelle, 2023) package to calculate Cronbach’s alpha, “anovakun” (Iseki, 2023) to perform ANOVAs, and “mediation” (Tingley et al., 2014) to estimate the indirect effects in the mediation models.

Data Handling

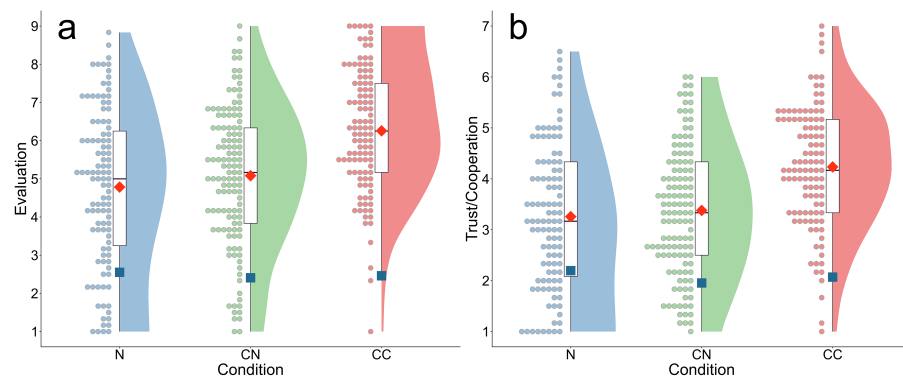
We excluded data from (1) seven participants who failed any of the attention check items, (2) 53 participants who answered any of the rule-check quizzes incorrectly, and (3) two participants who reported having participated in the same experiment previously. As a result, the final sample size was 340 ($M_{\text{age}} = 44.10$; 214 men, 126 women). All exclusion criteria were pre-registered.

Responses to the three items measuring P’s prosocial motivation, and the six items each for the evaluation and trust/cooperation of P and Q, were aggregated prior to analysis¹. Table 4 presents the means and standard deviations of measured variables by condition. Figure 3 shows the distributions of P’s evaluation (a) and trust/cooperation (b).

Table 4
Means (Standard Deviations) of Measured Variables in Each Condition in Study 3

	CC condition	CN condition	N condition
	<i>n</i> = 108	<i>n</i> = 125	<i>n</i> = 107
Perceived Prosocial Motivation of P (Intergroup Aggressor)	4.87 (1.25)	3.99 (1.52)	3.69 (1.67)
Evaluation of P	6.26 (1.62)	5.08 (1.82)	4.78 (2.04)
Trust/Cooperation of P	4.23 (1.19)	3.38 (1.28)	3.26 (1.42)
Evaluation of Q (Non-Aggressor)	2.46 (1.24)	2.41 (1.21)	2.55 (1.28)
Trust/Cooperation of Q	2.07 (0.98)	1.95 (0.85)	2.20 (1.08)
Estimated Cooperation of P	3.54 (1.76)	3.02 (1.97)	2.50 (1.87)
Within-Group Cooperation	3.89 (1.55)	4.08 (1.49)	3.93 (1.44)
Out-group Aggression	1.16 (1.57)	0.63 (1.34)	0.59 (1.24)
Keep	4.95 (2.37)	5.29 (2.07)	5.48 (1.94)

Note. The bottom three rows show participants’ hypothetical decisions in the game.

Figure 3*Distribution of the Intergroup Aggressors' Reputation in Study 3*

Note. The left plot (a) shows the distribution of evaluation of P (intergroup aggressor); The right plot (b) shows trust/cooperation of P. Red diamonds indicate means. Blue squares, corresponding to ratings of Q (non-aggressor), are included for comparison.

Results and Discussion

To test H1a-H2b, we conducted a one-way ANOVA on the evaluation and trust/cooperation ratings of P. The analysis revealed a significant main effect of condition (Evaluation: $F(2, 337) = 19.56, p < .001$, generalized $\eta^2 = .104$, 95% CI [.051, .164]; Trust/cooperation: $F(2, 337) = 18.38, p < .001$, generalized $\eta^2 = .098$, 95% CI [.046, .159]). Multiple comparisons using Shaffer's method showed that evaluations in the CC condition were significantly higher than those in the N condition (Evaluation: $t(337) = 5.88$, adjusted $p < .001$; Trust/cooperation: $t(337) = 5.51$, adjusted $p < .001$) and the CN condition (Evaluation: $t(337) = 4.87$, adjusted $p < .001$; Trust/cooperation: $t(337) = 5.01$, adjusted $p < .001$). In contrast, the difference between CN and N conditions was not significant (Evaluation: $t(337) = 1.23$, adjusted $p = .219$; Trust/cooperation: $t(337) = 0.71$, adjusted $p = .477$). These results support H1a-H2b.

To test H3, we examined whether perceived prosocial motivation mediated the effects observed in H1a-H2b using four mediation models (Figure S3). In Models a and c, we included data from the CC and N conditions. The independent variable was condition (dummy-coded: CC = 1, N = 0), the dependent variable was either evaluation (Model a) or trust/cooperation (Model c), and the mediator was perceived prosocial motivation. In Models b and d, we included data from the CC and CN conditions. The independent variable was condition (CC = 1, CN = 0), the dependent variable was either evaluation (Model b) or trust/cooperation (Model d), and the mediator was perceived prosocial motivation. Mediation analyses using the BCa bootstrap method with 10,000 resamples revealed significant indirect effects in all four models, supporting H3 (Model a: indirect

effect = .33, 95% CI [.21, .44], $p < .001$, Model b: indirect effect = .24, 95% CI [.14, .35], $p < .001$, Model c: indirect effect = .31, 95% CI [.20, .42], $p < .001$, Model d: indirect effect = .22, 95% CI [.13, .32], $p < .001$).

In addition, the pattern of correlations between participants' hypothetical decision and their attitudes toward the intergroup aggressor largely replicated the results of Study 2 (Table S1). More specifically, participants' allocations to the out-group aggression option were positively correlated with their evaluation (CC: $r = .35$, $p < .001$, CN: $r = .27$, $p < .01$, N: $r = .48$, $p < .001$) and trust/cooperation (CC: $r = .31$, $p < .01$, CN: $r = .28$, $p < .01$, N: $r = .56$, $p < .001$) towards intergroup aggressors, whereas allocations to the within-group cooperation option were not significantly correlated with either evaluation (CC: $r = -.13$, CN: $r = -.02$, N: $r = .07$, *n.s.*) or trust/cooperation (CC: $r = -.05$, CN: $r = .02$, N: $r = .03$, *n.s.*). These results highlight the robustness of the findings.

In Study 3, we examined the reputation of intergroup aggressors under three conditions: no conflicting interests between groups (N), conflicting interests with non-cooperative out-group aggression (CN), and conflicting interests with cooperative out-group aggression (CC). Consistent with our prediction, intergroup aggressors' reputation was significantly more positive in the CC condition than in the CN and N conditions. By contrast, their reputation did not differ significantly between the CN and N conditions, replicating the result of Study 2. These findings suggest that conflicting interests between groups improved intergroup aggressors' reputation in Study 1 because they benefited the in-group, rather than because they harmed the out-group. Thus, the decisive factor for their reputation is not the mere presence of intergroup conflict, but whether out-group aggression is cooperative.

General Discussion

Although the view that aggressing against out-group members can serve as a reputation maintenance strategy for intergroup aggressors has been widely accepted (De Dreu, Fariña, et al., 2022; De Dreu & Triki, 2022; Glowacki & Wrangham, 2013), past research has largely overlooked a critical question: how people actually evaluate aggressors. Except for a few pioneering studies (e.g., Benard & Doan, 2020; Halevy et al., 2012), the reputation of individuals who engage in out-group aggression remains underexplored. The present study is one of the first empirical attempts to investigate the reputation of intergroup aggressors. We examined the effects of situational factors—the presence of conflicting interests and the availability of a within-group cooperation option—on the reputational outcomes of intergroup aggression. To this end, we conducted three pre-registered experiments using different methodologies, including experimental vignettes and an economic game.

In Study 1, participants rated intergroup aggressors more favorably when there were conflicting interests between groups than when there were not. This effect was fully

mediated by perceptions of the aggressors' prosocial motivation. These results align with the status theory of collective action (Willer, 2009), which posits that individuals who make self-sacrifices in social dilemmas gain a positive reputation through perceived group-oriented motivation. Another prediction of the theory is that cooperators who receive higher status are more likely to engage in further cooperative behavior toward their in-group. Based on this, it is possible that a chain reaction occurs: people attribute a positive reputation to intergroup aggressors in the context of intergroup conflict, which in turn heightens aggressors' motivation for out-group aggression, potentially leading to a self-reinforcing cycle of intergroup hostility. Although we did not empirically test whether reputations of intergroup aggressors actually promote out-group aggression, future research should explore this process.

Studies 2 and 3 found that participants evaluated intergroup aggressors favorably under conditions of intergroup conflict only when the aggressors benefited their in-group. In other words, conflicting interests did not enhance the reputation of individuals who engaged in "wasteful" out-group aggression that did not contribute to the in-group. This suggests that people grant reputational rewards to intergroup aggressors not for harming the out-group *per se*, but for increasing the welfare of their in-group. The reputation of intergroup aggressors thus depends heavily on whether they contribute to the in-group. This finding is consistent with previous empirical studies showing that people exhibit in-group favoritism but not out-group aggression in minimal group settings (Halevy et al., 2008; Yamagishi & Mifune, 2016). Out-group aggression may confer higher social status on aggressors only when it benefits in-group members.

Nevertheless, it is possible that there are situations in which intergroup aggressors who do not bring immediate material benefit to the in-group may still gain positive reputations. One such situation might involve retaliation. Benard and Doan (2020) showed that individuals who engaged in out-group aggression in response to insults directed at their in-group were evaluated more favorably than interpersonal retaliators and, in some cases, even more favorably than those who chose not to act. In such contexts, intergroup aggressors may be perceived as cooperators for upholding the reputation of the group as a whole. When considering retaliatory out-group aggression, it is important to take into account the relative strength of the opposing group. Retaliating against a stronger out-group may provoke further retaliation, which could ultimately result in serious harm to the in-group. Indeed, past research has shown that the relationship between out-group threat and attitudes toward out-groups is moderated by the relative social status of the out-group (Riek et al., 2006). Future research should examine this possibility.

In this study, we focused on two aspects of intergroup aggressors' reputation: evaluation and willingness to trust and cooperate (trust/cooperation). Those two aspects are likely to be strongly interrelated, as suggested by the nearly identical patterns observed across the three studies. However, they may play distinct roles in intergroup contexts. For example, trust/cooperation can provide direct material incentives to intergroup ag-

gressors, whereas evaluation may contribute more indirectly by enhancing their status within the in-group. Moreover, the reputation of intergroup aggressors is likely to be more multifaceted, with different dimensions exerting distinct influences on reputational outcomes (Macfarlan & Lyle, 2015). For example, Horita (2010) demonstrated that altruistic punishers were not favored as recipients but were preferred as allocators of resources. This suggests that research on the reputation of intergroup aggressors should also examine more diverse dimensions of reputation.

The findings of this study also have important implications for preventing intergroup conflict. The results suggest that people's perceptions of the cooperativeness of out-group aggression enhance the reputation of intergroup aggressors. Therefore, the spread of narratives that frame out-group aggression as benefiting the in-group—such as claims that rejecting foreigners or immigrants serves national interests—may serve as a canary in the coal mine for the escalation of intergroup conflict. This prediction aligns with recent research showing a substantial impact of exposure to right-wing alternative media on attitudes toward immigrants (e.g., Theorin & Strömbäck, 2020). Although conflicts of interest between the in-group and the out-group do exist in some cases (e.g., competition over land), they may be minimal or even nonexistent in other cases. For example, recent research by Snijder et al. (2025) demonstrated that people can readily believe false defense rhetoric disseminated by political leaders. Future research could investigate the dynamic interplay between people's perceptions of intergroup relations and social norms regarding out-group aggression. Although our experiments did not involve realistic social groups, specific characteristics of real social groups, such as in-group cohesiveness and members' in-group identification, may also influence this process. Past research has indicated that in-group identification can promote the perception of antagonistic relationships with out-groups (Greijdanus et al., 2015). It is possible that illusory perceptions of intergroup conflict spread more easily in groups whose members strongly identify with their group, leading to more positive reputations for intergroup aggressors.

To summarize, the experiments revealed that intergroup aggressors were evaluated more positively when there were conflicting interests between the in-group and out-group than when there were not. This effect was caused by the fact that out-group aggression under intergroup conflict increased benefits for in-group members. These findings highlight the pivotal role of reputation in the emergence and escalation of intergroup conflict.

Limitations

This study has several limitations. First, as mentioned above, we did not examine actual aggressive behavior in this research; rather, we identified the conditions under which intergroup aggressors are evaluated favorably. Therefore, it remains an open question whether individuals actually engage in out-group aggression in such situations. Second,

the two dependent variables we used as indicators of reputation were both self-reported rather than behaviorally observed. As such, the present findings do not reveal how intergroup aggressors are treated within their in-groups in real interactions. Future research could address this by employing sequential game paradigms (e.g., [Milinski et al., 2002](#)) to investigate the dynamic interaction processes linking out-group aggression and cooperative behavior toward aggressors. Third, we could not address potentially important factors that affect the reputation of intergroup aggressors, such as their pre-existing status within the in-group. Whether out-group aggressors are high- or low-status individuals at the time of the aggression is likely to influence how they are evaluated. Although this factor was controlled in the game settings of Studies 2 and 3, participants in Study 1 may have inferred that the intergroup aggressors depicted in the scenarios had more power than other in-group members. Fourth, as discussed earlier, data from a relatively large number of participants were excluded across the three studies, which may have reduced statistical power. We interpret this high drop-out rate as being due to the design of the comprehension quizzes: participants were not allowed to re-answer questions that they answered incorrectly. Therefore, we believe that the high drop-out rate does not necessarily indicate poor data quality. Fifth, although we conducted exploratory analyses of gender effects (reported in the Supporting Information, see [Otsubo & Yamaguchi, 2026](#)), we were unable to draw firm conclusions because our data contained a high proportion of men. Given that past research has documented gender differences in behavior under intergroup competition ([Van Vugt et al., 2007](#)), it is possible that gender differences also exist in perceptions of intergroup aggressors. Finally, we used Japanese samples across all three experiments. Although out-group aggression was depicted within artificial groups rather than real-world social groups, there may be cultural differences in how individuals who engage in aggressive behavior are treated (e.g., culture of honor: [Nisbett & Cohen, 1996](#)). Cross-cultural research is needed to verify the generalizability of our findings.

Conclusion

This study found that the reputation of intergroup aggressors improves under conditions of conflicting interests between the in-group and out-group, and that this effect arises because out-group aggression in such situations benefits their in-group. Our findings contribute to a scientific understanding of the mechanisms underlying the emergence and escalation of intergroup conflict, such as warfare. Research on the reputation of intergroup aggressors remains underdeveloped. However, this topic is essential for understanding and preventing intergroup conflict in human societies. Reputational factors influence not only the micro-level processes of individuals' participation in conflict, but also the macro-level processes that drive the cumulative intensification of intergroup conflict. Future investigation into this area is needed.

Funding: This research was supported by JSPS KAKENHI [grant number 22K03033], JST SPRING [grant number JPMJSP2108], and the Foundation for the Fusion of Science and Technology.

Acknowledgments: We are so grateful to Yohsuke Ohtsubo, Hiroshi Ikeda, Kengo Nawata, and Hirotaka Imada, for their helpful advice. We acknowledge Keiko Mizuno, Soichiro Kashiara, and Daiichiro Kuroki for their technical support. We would like to acknowledge Editage (www.editage.jp) for English language editing.

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: *Kai Otsubo*—Conceptualization | Funding acquisition | Investigation | Methodology | Writing – original draft | Writing – review & editing. *Hiroiyuki Yamaguchi*—Funding acquisition | Project administration | Supervision | Writing – review & editing.

Ethics Statement: All experiments in the study were conducted after being approved by the ethical committee of the first author's university (Study 1: Kyushu University [2023-017]; Study 2: Kyushu University [2023-025]; Study 3: The University of Tokyo [UTSP-24050]). Informed consent was obtained from all participants prior to their participation in the experiments.

Data Availability: The data and analysis codes used in this study are available on the OSF (see [Otsubo, 2023c](#)).

Supplementary Materials

For this article, the following Supplementary Materials are available:

- Preregistration Study 1 (see [Otsubo, 2023a](#))
- Preregistration Study 2 (see [Otsubo, 2023b](#))
- Preregistration Study 3 (see [Otsubo, 2025](#))
- Data and analysis code (see [Otsubo, 2023c](#))
- Supplementary results and materials (see [Otsubo & Yamaguchi, 2026](#))

Index of Supplementary Materials

Otsubo, K. (2023a). *Investigating situational factors affecting the reputation of out-group aggressors ver.2* [Preregistration Study 1]. OSF Registries. <https://doi.org/10.17605/OSF.IO/BX8PT>

Otsubo, K. (2023b). *Do conflicts of interest between groups improve the reputation of out-group aggressors?* [Preregistration Study 2]. OSF Registries. <https://doi.org/10.17605/OSF.IO/SGWDZ>

Otsubo, K. (2025). *Do intergroup conflicts improve the reputation of intergroup aggressors only if they benefit the in-group?* [Preregistration Study 3]. OSF Registries. <https://doi.org/10.17605/OSF.IO/AC95Z>

Otsubo, K. (2023c). *Investigating the reputation of out-group aggressors* [Data, Code]. OSF. <https://osf.io/r5acz>

Otsubo, K., & Yamaguchi, H. (2026). *Supplementary materials to "Intergroup aggressors: When and why do they acquire a positive reputation in the in-group?"* [Materials]. PsychOpen GOLD.
<https://doi.org/10.23668/psycharchives.22426>

References

- Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences*, 274(1610), 749–753.
<https://doi.org/10.1098/rspb.2006.0209>
- Benard, S. (2015). The value of vengeance: Reputational incentives for initiating versus reciprocating aggression. *Rationality and Society*, 27(2), 129–160.
<https://doi.org/10.1177/1043463115576135>
- Benard, S., & Doan, L. (2020). When is retaliation respected? Status and vengeance in intergroup and interpersonal contexts. *Socius: Sociological Research for a Dynamic World*, 6, Article 2378023120967199. <https://doi.org/10.1177/2378023120967199>
- Ben-Shachar, M., Lüdtke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56), Article 2815.
<https://doi.org/10.21105/joss.02815>
- Böhm, R., Rusch, H., & Gülerk, Ö. (2016). What makes people go to war? Defensive intentions motivate retaliatory and preemptive intergroup aggression. *Evolution and Human Behavior*, 37(1), 29–34. <https://doi.org/10.1016/j.evolhumbehav.2015.06.005>
- Bornstein, G. (1992). The free-rider problem in intergroup conflicts over step-level and continuous public goods. *Journal of Personality and Social Psychology*, 62(4), 597–606.
<https://doi.org/10.1037/0022-3514.62.4.597>
- Chagnon, N. A. (1988). Life histories, blood revenge, and warfare in a tribal population. *Science*, 239(4843), 985–992. <https://doi.org/10.1126/science.239.4843.985>
- Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88–97.
<https://doi.org/10.1016/j.jbef.2015.12.001>
- Chinchilla, J., Vazquez, A., & Gómez, Á. (2022). Identity fusion predicts violent pro-group behavior when it is morally justifiable. *The Journal of Social Psychology*, 162(6), 701–715.
<https://doi.org/10.1080/00224545.2021.1948813>
- De Dreu, C. K. W., Fariña, A., Gross, J., & Romano, A. (2022). Prosociality as a foundation for intergroup conflict. *Current Opinion in Psychology*, 44, 112–116.
<https://doi.org/10.1016/j.copsyc.2021.09.002>
- De Dreu, C. K. W., Gross, J., Méder, Z., Giffin, M., Prochazkova, E., Kriek, J., & Columbus, S. (2016). In-group defense, out-group aggression, and coordination failures in intergroup conflict. *Proceedings of the National Academy of Sciences of the United States of America*, 113(38), 10524–10529. <https://doi.org/10.1073/pnas.1605115113>

- De Dreu, C. K. W., Gross, J., & Reddmann, L. (2022). Environmental stress increases out-group aggression and intergroup conflict in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1851), Article 20210147. <https://doi.org/10.1098/rstb.2021.0147>
- De Dreu, C. K. W., & Triki, Z. (2022). Intergroup conflict: origins, dynamics and consequences across taxa. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1851), Article 20210134. <https://doi.org/10.1098/rstb.2021.0134>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Emler, N. (1990). A social psychology of reputation. *European Review of Social Psychology*, 1(1), 171–193. <https://doi.org/10.1080/14792779108401861>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18(3), 233–239. <https://doi.org/10.1111/j.1467-9280.2007.01882.x>
- Glowacki, L., & Wrangham, R. (2015). Warfare and reproductive success in a tribal population. *Proceedings of the National Academy of Sciences of the United States of America*, 112(2), 348–353. <https://doi.org/10.1073/pnas.1412287112>
- Glowacki, L., & Wrangham, R. W. (2013). The role of rewards in motivating participation in simple warfare. *Human Nature*, 24(4), 444–460. <https://doi.org/10.1007/s12110-013-9178-8>
- Greijdanus, H., Postmes, T., Gordijn, E. H., & van Zomeren, M. (2015). Steeling ourselves: Intragroup communication while anticipating intergroup contact evokes defensive intergroup perceptions. *PLoS One*, 10(6), Article e0131049. <https://doi.org/10.1371/journal.pone.0131049>
- Halevy, N., Bornstein, G., & Sagiv, L. (2008). “In-group love” and “out-group hate” as motives for individual participation in intergroup conflict: A new game paradigm. *Psychological Science*, 19(4), 405–411. <https://doi.org/10.1111/j.1467-9280.2008.02100.x>
- Halevy, N., Chou, E. Y., Cohen, T. R., & Livingston, R. W. (2012). Status conferral in intergroup social dilemmas: Behavioral antecedents and consequences of prestige and dominance. *Journal of Personality and Social Psychology*, 102(2), 351–366. <https://doi.org/10.1037/a0025515>
- Horita, Y. (2010). Punishers may be chosen as providers but not as recipients. *Letters on Evolutionary Behavioral Science*, 1(1), 6–9. <https://doi.org/10.5178/lebs.2010.2>
- Iseki, R. (2023). *anovakun* (4.8.9) [Computer software]. <https://riseki.cloudfree.jp/?ANOVA%E5%90%9B>
- Jing, Y., Gries, P. H., Li, Y., Stivers, A. W., Mifune, N., Kuhlman, D. M., & Bai, L. (2017). War or peace? How the subjective perception of great power interdependence shapes preemptive defensive aggression. *Frontiers in Psychology*, 8, Article 864. <https://doi.org/10.3389/fpsyg.2017.00864>
- Kiyonari, T., & Barclay, P. (2008). Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *Journal of Personality and Social Psychology*, 95(4), 826–842. <https://doi.org/10.1037/a0011381>

- Macfarlan, S. J., & Lyle, H. F. (2015). Multiple reputation domains and cooperative behaviour in two Latin American communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1683), Article 20150009. <https://doi.org/10.1098/rstb.2015.0009>
- Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, 48, 61–83. <https://doi.org/10.1016/j.jrp.2013.09.008>
- Mifune, N. (2022). Null results for the steal-framing effect on out-group aggression. *Scientific Reports*, 12, Article 686. <https://doi.org/10.1038/s41598-021-04729-z>
- Milinski, M., Semmann, D., & Krambeck, H.-J. (2002). Reputation helps solve the “tragedy of the commons.” *Nature*, 415, 424–426. <https://doi.org/10.1038/415424a>
- Nawata, K. (2020). A glorious warrior in war: Cross-cultural evidence of honor culture, social rewards for warriors, and intergroup conflict. *Group Processes & Intergroup Relations*, 23(4), 598–611. <https://doi.org/10.1177/1368430219838615>
- Nawata, K., Huang, L., & Yamaguchi, H. (2016). Anti-Japanese public attitudes as conformity to social norms in China: The role of the estimated attitude of others and pluralistic ignorance. *Japanese Journal of Applied Psychology*, 42, 16–24.
- Nesdale, A. R., Rule, B. G., & McAra, M. (1975). Moral judgments of aggression: Personal and situational determinants. *European Journal of Social Psychology*, 5(3), 339–349. <https://doi.org/10.1002/ejsp.2420050307>
- Niimi, R., & Goto, M. (2023). Good conduct makes your face attractive: The effect of personality perception on facial attractiveness judgments. *PLoS One*, 18(2), Article e0281758. <https://doi.org/10.1371/journal.pone.0281758>
- Nisbett, R. E., & Cohen, D. (1996). *Culture of honor: The psychology of violence in the South*. Westview Press.
- Ozono, H., & Watabe, M. (2012). Reputational benefit of punishment: Comparison among the punisher, rewarder, and non-sanctioner. *Letters on Evolutionary Behavioral Science*, 3(2), 21–24. <https://doi.org/10.5178/lebs.2012.22>
- Raihani, N. J., & Bshary, R. (2015). The reputation of punishers. *Trends in Ecology & Evolution*, 30(2), 98–103. <https://doi.org/10.1016/j.tree.2014.12.003>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.3.1) [Computer software]. <https://www.R-project.org/>
- R Core Team. (2024). *R: A language and environment for statistical computing* (4.3.3) [Computer software]. <https://www.R-project.org/>
- Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research* (R package Version 2.3.9) [Computer software]. <https://CRAN.R-project.org/package=psych>
- Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, 10(4), 336–353. https://doi.org/10.1207/s15327957pspr1004_4

- Rocca, K. A., & McCroskey, J. C. (1999). The interrelationship of student ratings of instructors' immediacy, verbal aggressiveness, homophily, and interpersonal attraction. *Communication Education*, 48(4), 308–316. <https://doi.org/10.1080/03634529909379181>
- Rule, B. G., Nesdale, A. R., & McAra, M. J. (1974). Children's reactions to information about the intentions underlying an aggressive act. *Child Development*, 45(3), 794–798. <https://doi.org/10.2307/1127848>
- Rusch, H., Leunissen, J. M., & van Vugt, M. (2015). Historical and experimental evidence of sexual selection for war heroism. *Evolution and Human Behavior*, 36(5), 367–373. <https://doi.org/10.1016/j.evolhumbehav.2015.02.005>
- Semmann, D., Krambeck, H.-J., & Milinski, M. (2004). Strategic investment in reputation. *Behavioral Ecology and Sociobiology*, 56(3), 248–252. <https://doi.org/10.1007/s00265-004-0782-9>
- 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.
- Snijder, L. L., Gross, J., & De Dreu, C. K. W. (2025). How defense rhetoric escalates intergroup conflict. *iScience*, 28(11), Article 113749. <https://doi.org/10.1016/j.isci.2025.113749>
- Spanovic, M., Lickel, B., Denson, T. F., & Petrovic, N. (2010). Fear and anger as predictors of motivation for intergroup aggression: Evidence from Serbia and Republika Srpska. *Group Processes & Intergroup Relations*, 13(6), 725–739. <https://doi.org/10.1177/1368430210374483>
- Tateishi, W., Onoda, R., & Takahashi, N. (2021). How do the estimated motives of punishers affect their reputation? An examination using multiple punishment types. *Japanese Journal of Social Psychology*, 36(3), 96–103. <https://doi.org/10.14966/jssp.1916>
- Theorin, N., & Strömbäck, J. (2020). Some media matter more than others: Investigating media effects on attitudes toward and perceptions of immigration in Sweden. *International Migration Review*, 54(4), 1238–1264. <https://doi.org/10.1177/0197918319893292>
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*, 59(5), 1–38. <https://doi.org/10.18637/jss.v059.i05>
- Van Vugt, M., De Cremer, D., & Janssen, D. P. (2007). Gender differences in cooperation and competition: The male-warrior hypothesis. *Psychological Science*, 18(1), 19–23. <https://doi.org/10.1111/j.1467-9280.2007.01842.x>
- Van Vugt, M., & Tybur, J. M. (2015). The evolutionary foundations of status hierarchy. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (2nd ed.) (pp. 788–809). John Wiley & Sons. <https://doi.org/10.1002/9781119125563.evpsych232>
- Willer, R. (2009). Groups reward individual sacrifice: The status solution to the collective action problem. *American Sociological Review*, 74(1), 23–43. <https://doi.org/10.1177/000312240907400102>
- Yamagishi, T., & Mifune, N. (2016). Parochial altruism: Does it explain modern human group psychology? *Current Opinion in Psychology*, 7, 39–43. <https://doi.org/10.1016/j.copsyc.2015.07.015>

