Evolution and Human Behavior 37 (2016) 29–34

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

What makes people go to war? Defensive intentions motivate retaliatory and preemptive intergroup aggression☆

Robert Böhm [a,*], Hannes Rusch [b,c,1], Özgür Gürerk [d]

[a] Decision Analysis, School of Business and Economics, RWTH Aachen University, Aachen, Germany
[b] Behavioral Economics, Faculty of Economics and Business Studies, JLU Giessen, Giessen, Germany
[c] Business Ethics, TUM School of Education, TU München, Munich, Germany
[d] Experimental Economics, School of Business and Economics, RWTH Aachen University, Aachen, Germany

ARTICLE INFO

Article history:
Initial receipt 9 April 2015
Final revision received 24 June 2015

Keywords:
Intergroup conflict
Parochial altruism
Intergroup aggression
Attack
Defense

ABSTRACT

Although humans qualify as one of the most cooperative animal species, the scale of violent intergroup conflict among them is unparalleled. Explanations of the underlying motivations to participate in an intergroup conflict, however, remain unsatisfactory. While previous research shows that intergroup conflict increases individually costly behavior to the benefit of the in-group, it has failed to identify robust triggers of aggressive behavior directed at out-groups. Here, we present a controlled laboratory experiment which demonstrates that such aggression can be provoked systematically by manipulating the extent to which the own group is perceived to be on the offensive or the defensive side of a conflict. We find direct and causal evidence that the motivation to protect the in-group not only is a predictor of retaliatory aggression, but also promotes preemptive offensive actions against out-groups if they pose a potential threat. This finding improves our understanding of the escalation of intergroup conflicts and may have important implications for their prevention, as we find in our experiment that removing out-group threat substantially reduced intergroup aggression and led to full peace.

1. Introduction

Intergroup conflict constitutes a complex social dilemma (Bornstein, 2003; Choi & Bowles, 2007). For the individual, participation in intergroup conflict is costly, as it may lead to injury or death, whereas victory benefits all in-group members, e.g., through territorial gains or the establishment of deterrence and safety. Therefore, strong individual incentives to free-ride on the other group members' efforts exist (Bornstein, 2003). Eventually, however, extensive free-riding can cause a breakdown of cooperation within the group, leading to detrimental outcomes for all in-group members. To explain why we observe individually costly engagement in intergroup conflicts, it has been suggested that behavioral traits causing personally costly cooperation with in-group members and aggressiveness toward out-groups might have co-evolved (Choi & Bowles, 2007). The empirical evidence gathered to test this co-evolutionary hypothesis, however, is inconclusive (Rusch, 2014a).

Although the presence of rivaling out-groups increases individuals' engagement in various types of costly behaviors benefiting the in-group (e.g., Bornstein, 1992; Bornstein & Ben-Yossef, 1994; Burton-Chellew, Ross-Gillespie, & West, 2010; Puurtinen, Heap, & Mappes, 2015), it is still unclear which factors trigger costly behavior to the detriment of out-groups (Fry & Söderberg, 2013). Astonishingly, unconditional 'out-group hate' has been rarely observed in previous research. Instead, in-group cooperation often coincides with the avoidance of intergroup competition (De Dreu et al., 2010; Halevy, Bornstein, & Sagiv, 2008), even after the experimental induction of a "history of conflict" (Halevy, Weisel, & Bornstein, 2012), or in interactions between natural groups with a strong enmity (Weisel & Böhm, 2015). In a similar vein, when measuring attitudes towards in-group and out-group members, in-group positivity does not systematically correlate with out-group negativity (e.g., Kosterman & Feshbach, 1989; Struch & Schwartz, 1989). Given the existence of frequent and brutal intergroup aggression and competition in human history (Gat, 2009; Keeley, 1997; Kelly, 2005), however, there must be important motivations causing such persistent and spiteful behaviors.

This study focuses on one arguably very important motivation that has, nevertheless, received remarkably little attention in previous research: protecting the in-group from potential out-group threat. Such defense preparedness, e.g., conquering territories strategically in order to reach an advantageous position against possible attacks, or preemptive strikes against potential aggressors, is considered as a key factor in the explanation of intergroup conflicts in hunter–gatherer societies (Gat, 2009) and wars on the country-level (Levy, 2011; von Clausewitz, 1832).

But why is it that individuals should be more motivated to participate in defensive rather than in offensive intergroup aggression?

☆ Financial support from the "Excellence Initiative (ZUK II)" of the German Research Foundation (DFG) is appreciated.

* Corresponding author. Decision Analysis, School of Business and Economics, RWTH Aachen University, Templergraben 64, 52062, Aachen. Tel.: +49 241 8097000; fax: +49 241 8092167.
E-mail address: robert.boehm@rwth-aachen.de (R. Böhm).
[1] Present address: Corporate Development and Business Ethics, Faculty of Management, Economics and Social Sciences, University of Cologne, Cologne, Germany.

http://dx.doi.org/10.1016/j.evolhumbehav.2015.06.005
1090-5138/

Historical evidence (Glowacki & Wrangham, 2013; Rusch, 2013) and evidence from animal studies (Crofoot & Gilby, 2012; Kitchen & Beehner, 2007; Rusch, 2014a) suggest that the cost/benefit structure of intergroup conflict is different for attacks and defenses. Specifically, across species, the likelihood of aggressive attacks against out-groups decreases with perceived risk of failure (Kelly, 2005), and increases with prospects of individual benefits, including additional reproductive access for the victors (Glowacki & Wrangham, 2013, 2015; McDonald, Navarrete, & Van Vugt, 2012; Rusch, Leunissen, & van Vugt, in press). These observations militate against the typically supposed public good structure of intragroup conflict nested within intergroup conflict (Rusch, 2013, 2014b). Costly involvement in group defense, on the other hand, increases with the individual importance of the resources at stake, including vulnerable relatives and reproductive access (Crofoot & Gilby, 2012; Rusch, 2014a), as well as with perceived fear of out-groups (De Dreu et al., 2010). Additionally, an unsuccessful defense may have negative consequences for any member of the group, e.g., loss of resources and/or relatives, or death, irrespective of whether the individual participated in the defensive action or not, increasing the perception of common fate among group members.

In addition to the structural differences between offensive and defensive intergroup encounters that are likely to influence individual perception and behavior, research in social psychology indicates that the perception of an existential threat, e.g., by being reminded of one's own mortality, increases in-group favoritism and out-group derogation (for reviews, see Burke, Martens, & Faucher, 2010; Jonas & Fritsche, 2013). Particularly a threat to the own social identity has been shown to fuel negative attitudes and emotions towards the threatening out-group (e.g., Brewer, 2007; Stephan & Stephan, 2000). Such evaluations can culminate in "delegitimizing beliefs", i.e., the attribution of extremely negative characteristics to another group, which in turn may justify intergroup violence (Bar-Tal, 2000).

This variety of evidence from multiple scientific disciplines suggests that the motivations for participating in offensive and defensive intergroup aggression are likely to differ significantly. We suppose that the negligence of this important difference might explain the ambiguity of the existing experimental evidence, as in previous experiments it was unclear whether subjects would perceive themselves as being on the offensive or the defensive side of the intergroup conflict. We therefore hypothesize that triggering a clear defensive motivation in individuals, as compared to the ambiguous cues of previous studies, will cause an increase of the readiness to engage in costly behavior to the detriment of the out-group.

To test our hypothesis, we conducted an incentivized behavioral experiment, which differentiates two types of defensive intergroup actions: (1) retaliation of previously experienced harm aiming to reduce the relative disadvantage of the in-group compared to the out-group. We will refer to this type of aggression as *retaliatory* (ex post) intergroup aggression. (2) Preemptive strikes that aim to reduce the likelihood of an absolute and relative disadvantage in beforehand, which we will refer to as *preemptive* (ex ante) intergroup aggression.

2. Materials and methods

In our experiment, participants faced modified versions of an established experimental game modeling intergroup conflict — the Intergroup Prisoner's Dilemma-Maximizing Difference (IPD-MD; e.g., De Dreu et al., 2010; Halevy et al., 2008, 2012). In the IPD-MD, participants are randomly assigned to one of two equal-sized groups. Each participant decides individually and independently how to divide valuable endowment-points between three options: (1) each point KEPT in the personal account gives the individual a benefit of 1 point with no effect on any other player. (2) Each point contributed to the WITHIN pool of the in-group credits every in-group member with 0.5 points. Contribution to WITHIN has no negative consequences for out-group members, hence indicating cooperation with the in-group and a disregard of, or even a positive concern for, the out-group's payoff. (3) Lastly, each point contributed to the BETWEEN pool has the same positive consequences for the in-group as a point contributed to WITHIN, but additionally reduces the payoff of each out-group member by 0.5 points. Thus, contributing to BETWEEN also indicates in-group cooperation but now coupled with the intention to decrease the out-group's absolute or relative payoff. Consequently, the higher the level of contribution to BETWEEN relative to WITHIN, the more aggression and/or competition towards the out-group is shown.[2]

As a baseline, we utilize the classic IPD-MD game with a group size of $n = 3$ and an endowment of $e = 5$ (see the Procedures section for details), in which all individuals in both groups decide simultaneously (subsequently labeled as *SIM*). Hence, in *SIM*, individuals cannot be sure whether their actions are defensive or offensive. We contrast *SIM* with four other conditions utilizing a similar game structure, but with groups playing sequentially (labeled *SEQ*).

Two of these conditions – *SEQ-BETWEEN* and *SEQ-WITHIN* – are designed to investigate *retaliatory* (ex post) intergroup aggression, that is, a response to an aggressive attack from another group. In the *SEQ-BETWEEN* condition, members of one group – the second-movers – allocate points contingent on the number of points allocated to BETWEEN by the opposing first-mover group. That is, we use the strategy vector method to elicit second-movers' complete contribution strategies (Selten, 1967; see the Procedures section for details). Hence, in *SEQ-BETWEEN*, second-movers are able to retaliate against potential attacks by the first-movers and can thus avert a relative loss. The full interaction structure is common knowledge among both first- and second-movers. In the control condition *SEQ-WITHIN*, second-movers distribute their points contingent on the first-movers' contributions to WITHIN. Thus, in contrast to *SEQ-BETWEEN*, second-movers in *SEQ-WITHIN* make their decisions conditional on the level of (peaceful) cooperation shown by the first-movers. As higher levels of contributions to WITHIN by the first-movers indicate that they do not intend to harm the second-movers, we suspect that second-movers' in-group cooperation, particularly contributions to WITHIN, will increase with the amount of first-movers' contributions to WITHIN (Böhm & Rockenbach, 2013; Burton-Chellew & West, 2012).

Two additional conditions – *SEQ-PREEMPTIVE-STRIKE* and *SEQ-SECURE-STRIKE* – test whether the motivation to defend the in-group can also cause an offensive display of *preemptive* (ex ante) intergroup aggression. Such defense-motivated attacks are historically well documented, e.g., the Israeli strike against Egyptian airfields in 1967. Moreover, there is recent evidence that aggression between individuals increases if it serves to protect one's own self (Abbink & de Haan, 2014; Simunovic, Mifune, & Yamagishi, 2013), particularly when the interaction partner is an out-group member (De Dreu et al., 2010).

In the *SEQ-PREEMPTIVE-STRIKE* condition, first-movers can reduce the negative effect of second-movers' intergroup aggression on their in-group's payoff. Each point contributed to the own BETWEEN pool preemptively reduces the negative effect of out-group members' possible subsequent BETWEEN-contributions by 0.05 points. Thus, by contributing to BETWEEN first-movers can reduce the second-movers' BETWEEN contributions' impact on the own payoff, eventually down to 0 (from initially 0.5 point). For example, if the sum of first-movers' contributions to BETWEEN is 5 points, each point potentially contributed to BETWEEN by the second-movers reduces each first-mover's payoff only by $0.5 - (5 \times 0.05) = 0.25$ points. Hence, contributions of 10 or more points to BETWEEN by the first-movers completely remove the potential threat of losing points in this condition (because $0.5 - (10 \times 0.05) = 0$), while still harming the second-movers with maximum severity (in the example case, each second-mover loses $10 \times 0.5 = 5$ points). Note, however, that keeping points and free-riding on the protective contributions of the other in-group members is still payoff

[2] In what follows, we refer to 'intergroup aggression' when behavior reduces both the out-group's absolute and relative payoff, i.e., contributions to BETWEEN are considered as 'intergroup aggression'.