What Makes People Go to War? Defensive Intentions Motivate Retaliatory and Preemptive Intergroup Aggression

Robert Böhm and Hannes Rusch and Özgür Gürerk

RWTH Aachen University, Aachen, Germany, Behavioral Economics, JLU Giessen, Giessen, Germany, Chair of Business Ethics, TU München, Munich, Germany

9. April 2015

What Makes People Go to War?

Defensive Intentions Motivate Retaliatory and Preemptive Intergroup Aggression

Robert Böhm[1]*, Hannes Rusch[2,3], Özgür Gürerk[1]

[1] School of Business and Economics, RWTH Aachen University, Aachen, Germany

[2] Behavioral Economics, JLU Giessen, Giessen, Germany

[3] Peter Löscher Chair of Business Ethics, TU München, Munich, Germany

words: 4,851

* Correspondence to:

Robert Böhm

School of Business and Economics

RWTH Aachen University

Templergraben 64

52062 Aachen

Phone:	+49 241 8097000
Fax:	+49 241 8092167
E-mail:	robert.boehm@rwth-aachen.de
Co-author emails:	hannes.rusch@tum.de (HR)
	ozgur.gurerk@rwth-aachen.de (ÖG)

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 motivation to participate in an intergroup conflict, however, remain unsatisfactory. While previous research shows that intergroup conflict increases ‘in-group love’, it fails to identify robust triggers of ‘out-group hate’. Here, we present a controlled laboratory experiment, which demonstrates that ‘out-group hate’ can be provoked systematically. We find direct and causal evidence that the intention to protect the in-group is not only a crucial motivator of ‘out-group hate’ in defensive reactions, but also promotes preemptive offensive actions against out-group threat. Hence, the strength of ‘out-group hate’ depends on whether the own group is perceived to be on the offensive or the defensive side of the conflict. 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 reduces intergroup aggression, leading to full peace.

Keywords: intergroup conflict, parochial altruism, in-group love, out-group hate, defense

What makes people go to war?

Defensive intentions motivate retaliatory and preemptive intergroup aggression

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 ('in-group love') and aggressiveness toward out-groups ('out-group hate') might have co-evolved (Bernhard, Fischbacher, & Fehr, 2006; Choi & Bowles, 2007). The empirical evidence gathered to test this co-evolutionary hypothesis is inconclusive, though (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 & Ben-Yossef, 1994; Bornstein, 1992; Burton-Chellew, Ross-Gillespie, & West, 2010; Puurtinen, Heap, & Mappes, in press), 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 interaction between natural groups with a strong enmity (Weisel & Böhm, in press). 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 in human history (Gat, 2009; Keeley, 1997; Kelly, 2005), however, there must be important motivations causing persistent and spiteful 'out-group hate'.

This study focuses on an important motivator for 'out-group hate' that has received little attention in previous research: the intention to protect the in-group from potential out-group threat. Defense preparedness, e.g., conquering territories strategically in order to reach an advantageous position against possible attacks, and preemptive strikes against the potential aggressor, are important factors for explaining intergroup conflicts in hunter-gatherer societies (Gat, 2009) and wars on the country-level (Clausewitz, 1832; Levy, 2011).

Why should individuals be more motivated to participate in defensive rather than offensive intergroup actions? 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; McDonald, Navarrete, & Van Vugt, 2012; Rusch, Leunissen, & van Vugt, in press). This militates against the typically supposed public good structure of intragroup cooperation (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 or death, irrespective of whether the individual participated or not in the defensive action, 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 by being reminded about one's own mortality increases in-group favoritism and out-group derogation (for reviews, see Burke, Martens, & Faucher, 2010; Jonas & Fritsche, 2013). Particularly an 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 aggression and violence (Bar-Tal, 2000).

In line with this variety of evidence from various scientific disciplines, the motivations for participating in offensive and defensive intergroup aggression are likely to differ in their structure and strength. 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 hypothesize that triggering defensive intentions in subjects will cause a greater readiness to engage in costly behavior to the detriment of the out-group than offensive intentions. We conducted an incentivized behavioral experiment to test our hypothesis by means of two different types of defensive intergroup actions: (1) retaliation of previously experienced 'out-group hate' in order to reduce the relative disadvantage compared to the other group ('out-group hate' *ex post*), and (2) preemptive strikes in order to reduce the likelihood of an absolute and relative disadvantage beforehand ('out-group hate' *ex ante*).

2. Materials and Methods

Participants face modified versions of an established experimental game modeling intergroup conflict – the Intergroup Prisoner's Dilemma-Maximizing Difference (IPD-MD) (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 peaceful 'in-group love' and a disregard 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. Thus, contributing to BETWEEN again indicates 'in-group love' but now coupled with spiteful 'out-group hate', that is, the concern to minimize the out-group's absolute or relative payoff.

We contrast 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 and can therefore not be sure whether their actions are defensive or offensive, (subsequently labeled *SIM*) with four other conditions utilizing a similar game structure, but with groups playing sequentially (labeled *SEQ*).

Two of these conditions – *SEQ-HATE* and *SEQ-LOVE* – are designed to investigate retaliatory 'out-group hate' *ex post*, that is, a response to an aggressive attack from another group. In the *SEQ-HATE* condition, members of one group – the second-movers – simultaneously allocate points contingent on the number of points allocated to BETWEEN by the opposing first-mover group, i.e., we use the so-called strategy vector method to elicit second-movers' complete contribution strategies (Selten, 1967; see the Procedures section for details). Hence, in *SEQ-HATE*, second-movers are able to retaliate against potential attacks by the first-movers and therefore avert a relative loss, which is common knowledge among both first- and second-movers. In the control condition *SEQ-LOVE*, second-movers distribute their points contingent on the first-movers' contributions to WITHIN. Here, in the absence of any relative threat through the out-group, one may assume that second-movers' cooperation (particularly 'in-group love') increases with a higher amount of contributions to WITHIN by the out-group, as a form of peaceful competition via in-group cooperation (Böhm & Rockenbach, 2013).

Two additional conditions – *SEQ-PREEMPTIVE-STRIKE* and *SEQ-SECURE-STRIKE* – aim to test whether the desire to defend the in-group can also cause an offensive display of 'out-group hate' *ex ante*. 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' 'out-group hate' on themselves. Each point contributed to the own BETWEEN pool preemptively reduces the negative effect of out-group members' subsequent BETWEEN-contributions by 0.05 points until it may reach its minimum of 0 (from initially 0.5 points). 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 first-movers' 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 out-group member 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 maximizing from the individual perspective because the intragroup public good dilemma remains unchanged. Should first-movers' contribute to BETWEEN in *SEQ-PREEMPTIVE-STRIKE*, though, this could be explained by two motivations: either (1) the intention to preemptively defend against absolute and relative harm caused by the out-group, or (2) the desire to attack and harm the out-group without having to fear retaliation. To disentangle these two motivations, we conducted a control condition *SEQ-SECURE-STRIKE* in which we removed the second-movers' BETWEEN pool. Here, second-movers can only choose between keeping points and contributing to WITHIN. Thus, first-movers do not have to fear second-movers 'out-group hate' and should thus only contribute to BETWEEN if they intended to harm the out-group without worrying about retaliation.

2.1 Participants and Design

In total, 216 students (131 male and 85 female) from various academic disciplines of a German university participated in the experiment ($MD_{\text{age}} = 24$, $M_{\text{age}} = 24.47$, $SD = 5.12$). Treatment of participants was in agreements with the ethical guidelines of the German Research Foundation (*Deutsche Forschungsgemeinschaft*) and the German Psychological Society (*Deutsche Gesellschaft für Psychologie*). All participants gave their written informed consent to participate voluntarily and were assured that all statistical analyses and reports would be anonymous. Decisions were incentivized. To avoid potential losses, participants received an additional flat-fee of €4. On average, participants earned €7.40 in the 45-minute experiment.

The sessions of 24 participants each were randomly assigned to one of five experimental between-subjects conditions: *SIM*, *SEQ-HATE*, *SEQ-LOVE*, *SEQ-PREEMPTIVE-STRIKE*, and *SEQ-SECURE-STRIKE*. Because the sequential (*SEQ*) conditions assign two roles randomly to participants (first-mover and second-mover), there was one session assigned to *SIM* and there were two sessions assigned to each of the other conditions in order to have 24 independent observations in each condition. One session was filled with a research assistant, because one subject did not show up; the data of the assistant were excluded from all analyses.

2.2 Procedures

Participants pre-registered for experimental sessions online (Greiner, 2004). On arrival, each person was randomly allocated to a cubicle. All the interactions and corresponding decisions in the experiment were computer-mediated using the software z-Tree (Fischbacher, 2007). Participants were given written instructions (see electronic supplementary material). The experimenter read the instructions aloud. It was explained that each participant is randomly assigned to a group of size $n = 3$ that is matched with another three-person group (labeled the blue and the green group). In all conditions except SIM, each group was randomly given either the role of the first-movers or the second-movers. Each participant was endowed with $e = 5$ points (1 point = €0.50) and had to decide how to allocate these points among three behavioral options (except in the *SEQ-SECURE-STRIKE*/second-mover condition, where subjects only had two options). Allocations affected the personal outcome, as well as the in-group and the out-group members' outcomes. In order to have an equal amount of behavioral observations for each potential response pattern of second-movers without deceiving participants, we used the strategy vector method (Selten, 1967). Accordingly, second-movers indicated their responding contribution for each amount potentially contributed to WITHIN (*SEQ-LOVE*) or BETWEEN (*SEQ-HATE* and *SEQ-PREEMPTIVE-STRIKE*), respectively, by the first-movers. This yields $n \times e + 1 = 3 \times 5 + 1 =$ 16 responses in *SEQ-HATE* (0 to 15 points) and 11 responses in *SEQ-PREEMPTIVE-STRKE* (0 to 10 or more points), but only the stated response to the actual behavior of first-movers was eventually payoff-relevant.[1] Neutral labels were used in the instructions, e.g., 'decision making task', 'project A' (WITHIN pool), and 'project B' (BETWEEN pool), and there was no use of potentially emotion-laden words like 'cooperation', 'competition', 'aggression' etc. Participants could ask clarifying questions before the experiment started. To make sure that they had understood the game structure properly, participants had to correctly answer some test questions before making their allocation decision anonymously. Afterwards, participants answered a post-experimental questionnaire, including demographics. At the end, each participant received individual payoff information, and was paid in private.

3. Results

The dataset containing individual-level behavioral observations used for the following analyses can be found in the electronic supplementary material.

3.1 Retaliatory 'Out-Group Hate'

Figure 1 displays the contributions in *SIM*, as well as first-movers' contributions in *SEQ-HATE* and *SEQ-LOVE*. On average, first-movers overall contributions, i.e., all points not KEPT, in *SEQ-HATE* and *SEQ-LOVE* are similar to the overall contributions of players in *SIM*, *SIM*: $M = 2.63$, $SD = 2.34$, *SEQ-HATE*: $M = 2.92$, $SD = 2.19$, *SEQ-LOVE*: $M = 2.61$, $SD = 1.56$; one-way ANOVA: $F(2, 68) < 1$, $p > .250$, $\eta_p^2 = .005$. In line with previous research, participants in *SIM* show little 'out-group hate', contributing less than 10 % of points to BETWEEN. Importantly, 'out-group hate' does not significantly differ between first-movers in the two sequential conditions and players in the simultaneous condition, *SIM*: $M = 0.42$, $SD = 1.18$, *SEQ-HATE*: $M = 0.50$, $SD = 1.29$, *SEQ-LOVE*: $M = 0.65$, $SD = 0.94$; one-way ANOVA: $F(2, 68) < 1$, $p > .250$, $\eta_p^2 = .007$. Overall, first-movers' contribution patterns look very similar across conditions.

[1] The strategy space in *SEQ-PREEMPTIVE-STRIKE* was limited from 0 to "10 or more" points because the strategic consequences of own contributions were the same if at least 10 points were contributed to

Figure 2 displays allocations of points KEPT (panel A), contributed to WITHIN (panel B), or contributed to BETWEEN (panel C) of second-movers in *SEQ-HATE* and *SEQ-LOVE*. Similarly to first-movers, overall contributions of second-movers do not significantly differ between *SEQ-HATE* and *SEQ-LOVE* (*SEQ-HATE*: M = 3.14, SD = 0.36, *SEQ-LOVE*: M = 2.63, SD = 0.36; repeated-measures ANOVA: $F(1, 46) < 1$, $p > .250$, $\eta_p^2 = .021$). Moreover, the effect of first-movers' contributions to BETWEEN (*SEQ-HATE*) or WITHIN (*SEQ-LOVE*) has no impact on overall contributions of second-movers, $F(15, 690) < 1$, $p > .250$, $\eta_p^2 = .021$. There is a weak interaction effect of first-movers' contributions and condition on second-movers overall contributions, $F(15, 690) = 1.577$, $p = .074$, $\eta_p^2 = .033$. Simple slopes indicate that aggregated contributions to WITHIN and BETWEEN increase as a function of first-movers' contributions to WITHIN (*SEQ-LOVE*), $F(15, 345) = 1.841$, $p = .028$, $\eta_p^2 = .074$, but not as a function of first-movers' contributions to BETWEEN (*SEQ-HATE*), $F(15, 345) < 1$, $p > .250$, $\eta_p^2 = .024$ (see Figure 2, panel A: white bars but not grey bars are decreasing).[2]

More importantly, supporting our hypothesis, we indeed find a significant interaction effect of first-movers' contributions and condition on second movers' 'out-group hate', repeated-measures ANOVA: $F(15, 690) = 6.602$, $p < .001$, $\eta_p^2 = .126$. While there is a significant increase of second-movers' 'out-group hate' in *SEQ-HATE*, $F(15, 345) = 8.817$, $p < .001$, $\eta_p^2 = .277$, we observe no such effect in *SEQ-LOVE*, $F(15, 345) < 1$, $p > .250$, $\eta_p^2 = .027$ (see Figure 2, panel C: grey bars but not white bars are increasing). This result indicates that the second-movers retaliate against the first-movers' (potential) aggression by increasing their 'out-group hate'. In fact, at a certain point aggressive contributions to BETWEEN by second-movers even exceed their peaceful contributions to WITHIN, increasing up to almost the quadruple compared to the *SIM* condition.

[2] The increasing overall contributions in *SEQ-LOVE* are particularly driven by an increase of peaceful contributions to WITHIN, $F(15, 345) = 2.428$, $p < .001$, $\eta_p^2 = .126$.

3.2 Preemptive 'Out-Group Hate'

Although the results presented so far strongly support the hypothesis that spiteful 'out-group hate' increases as a response to the opponents' harmful acts, they do not answer the question as to how conflicts start in the first place. We hypothesize that the desire to defend the in-group is not only important for the participation in vindictive intergroup actions but might also play a crucial role in motivating attacks in the form of preemptive strikes.

Figure 1 displays the contributions of first-movers in *SEQ-PREEMPTIVE-STRIKE* and *SEQ-SECURE-STRIKE*. A comparison of the overall contributions in *SIM*, *SEQ-PREEMPTIVE-STRIKE* and *SEQ-SECURE-STRIKE* reveals a significant difference between conditions, one-way ANOVA: $F(2, 69) = 5.005$, $p = .009$, $\eta_p^2 = .127$. We observe the highest overall contributions in *SEQ-PREEMPTIVE-STRIKE* ($M = 3.08$, $SD = 1.89$), moderate contributions in *SIM* ($M = 2.63$, $SD = 2.34$), and the lowest in *SEQ-SECURE-STRIKE* ($M = 1.42$, $SD = 1.28$). The distribution of points to WITHIN and BETWEEN is particularly important for the focus of the present paper. Supporting our hypothesis, we find that 'out-group hate' significantly differs between these conditions (one-way ANOVA: $F(2, 69) = 11.691$, $p < .001$, $\eta_p^2 = .253$. As Figure 1 shows, 'out-group hate' is more than four times larger in *SEQ-PREEMPTIVE-STRIKE* ($M = 1.75$, $SD = 1.82$) than in *SIM* ($M = 0.42$, $SD = 1.18$), Bonferroni: $p = .001$. In contrast, 'out-group hate' is virtually not present in *SEQ-SECURE-STRIKE* ($M = 0.08$, $SD = 0.28$) – in fact, there are only 2 out of 24 subjects who each contribute just one point to BETWEEN – which does significantly differ from *SEQ-PREEMPTIVE-STRIKE* (Bonferroni: $p < .001$). This strongly supports our conjecture that the intention to protect the in-group may indeed cause offensive actions against out-groups and leads to considerable 'out-group hate'.

4. Discussion

We have argued that the intention to protect the own group may be an important motivator of 'out-group hate' in intergroup conflicts. Our results clearly show that having the option of retaliation in order to decrease relative losses of the own group causes individually costly, off-equilibrium strategy choice, i.e., 'out-group hate'. Furthermore, the findings indicate that preemptive 'out-group hate' is displayed only when it serves to prevent (potential) future aggression against the in-group, i.e., groups use defense-motivated first-strikes to avoid absolute and relative losses. Besides these main findings, the experiment provides further support for the 'intergroup comparison – intragroup cooperation hypothesis' (Böhm & Rockenbach, 2013; see also Burton-Chellew & West, 2012) by showing that intragroup cooperation increases as a function of the out-group's peaceful cooperation, i.e., first-movers' WITHIN-contributions in the condition *SEQ-LOVE*.

Our findings provide direct and causal evidence that 'out-group hate' can be provoked by manipulating the extent to which conflict parties perceive themselves to be in a vulnerable position. However, 'out-group hate' still is collectively destructive and likely to cause retaliation since the complete elimination of competition is often unlikely. Thus, a cascade of strikes and counter-strikes may eventually lead to mutual losses and conflict perpetuation (Halevy et al., 2012; Leibbrandt & Sääksvuori, 2012).

Our results and experimental manipulations are likely to stimulate further research on intergroup conflict and aggression in different directions. For instance, accessing retaliatory 'out-group hate' through responses to expected harmful actions by the out-group using the strategy vector method may be considered as a rather "cold" and deliberate response. We decided to use the strategy vector method in order to avoid deception of participants. However, future research may investigate whether automatic responses due to "hot" emotional states like anger, which have been shown to elicit in-group biased behaviors toward out-groups (e.g., Mackie, Devos, & Smith, 2000), further increase 'out-group hate'.

Moreover, although our experiment provides first behavioral evidence that realistic threat induces ‘out-group hate’, in real-word intergroup relations threats are often rather symbolic (e.g., due differences between group morals, standards, beliefs, and attitudes; see Stephan & Stephan, 2000). Therefore, it is an interesting question for future research whether the manipulation of perceived symbolic threats also increases ‘out-group hate’ in a behavioral setting.

The present results have important implications for peace research and research on group aggression more generally: Once the threat of aggression against the in-group – either currently or in the future – is credibly removed, individuals show virtually no ‘out-group hate’ in our experiment, substantially reducing the risk of violent intergroup conflict. Furthermore, being able to observe (or expect) out-groups’ peaceful and unthreatening cooperation may spark increased peaceful cooperation by the observing group as well.

Competing Interests

None of the authors has competing interests.

Author Contributions

RB, HR, and ÖG designed the study, RB conducted the experiment and analyzed the data; RB, HR, and ÖG wrote the paper and acknowledged the final version of the manuscript prior to submission.

Acknowledgements

We are grateful to Hillie Aaldering, Bernd Irlenbusch, Peter Letmathe, and Shaul Shalvi for helpful comments on an earlier draft. We thank Maik Theelen for assistance in data collection.

Literature

Abbink, K., & de Haan, T. (2014). Trust on the brink of Armageddon: The first-strike game. *European Economic Review*, *67*, 190–196.

Bar-Tal, D. (2000). *Shared beliefs in a society: Social psychological analysis*. Sage Publications.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, *442*, 912–915.

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

Bornstein, G. (1992). The free-rider problem in intergroup conflicts over step-level and continuous public goods. *Journal of Personality and Social Psychology*, *62*, 597–606.

Bornstein, G. (2003). Intergroup conflict: Individual, group, and collective interests. *Personality and Social Psychology Review*, *7*, 129–145.

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

Brewer, M. B. (2007). The importance of being we: Human nature and intergroup relations. *American Psychologist*, *62*, 728–738.

Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of terror management theory: a meta-analysis of mortality salience research. *Personality and Social Psychology Review*, *14*, 155–95.

Burton-Chellew, M. N., Ross-Gillespie, A., & West, S. A. (2010). Cooperation in humans: competition between groups and proximate emotions. *Evolution and Human Behavior*, *31*, 104–108.

Burton-Chellew, M. N., & West, S. A. (2012). Pseudocompetition among groups increases human cooperation in a public-goods game. *Animal Behaviour*, *84*, 947–952.

Choi, J.-K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science*, *318*, 636–40.

Clausewitz, C. von. (1832). *Vom Kriege* (1st ed.). Hinterlassene Werke.

Crofoot, M. C., & Gilby, I. C. (2012). Cheating monkeys undermine group strength in enemy territory. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 501–505.

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., … Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, *328*, 1408–1411.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*, 171–178.

Fry, D. P., & Söderberg, P. (2013). Lethal aggression in mobile forager bands and implications for the origins of war. *Science*, *341*, 270–273.

Gat, A. (2009). So why do people fight? Evolutionary theory and the causes of war. *European Journal of International Relations*, *15*, 571–599.

Glowacki, L., & Wrangham, R. W. (2013). The role of rewards in motivating participation in simple warfare. *Human Nature*, *24*, 444–460.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003* (GWDG Bericht, pp. 79–93). Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

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*, 405–411.

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*, 188–195.

Jonas, E., & Fritsche, I. (2013). Destined to die but not to wage war: How existential threat can contribute to escalation or de-escalation of violent intergroup conflict. *American Psychologist*, *68*, 543–558.

Keeley, L. H. (1997). *War before civilization: The myth of the peaceful savage*. New York, NY: Oxford University Press.

Kelly, R. C. (2005). The evolution of lethal intergroup violence. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 15294–15298.

Kitchen, D., & Beehner, J. (2007). Factors affecting individual participation in group-level aggression among non-human primates. *Behaviour*, *144*, 1551–1581.

Kosterman, R., & Feshbach, S. (1989). Toward a measure of patriotic and nationalistic attitudes. *Political Psychology*, *10*, 257–274.

Leibbrandt, A., & Sääksvuori, L. (2012). Communication in intergroup conflicts. *European Economic Review*, *56*, 1136–1147.

Levy, J. S. (2011). Declining power and the preventive motivation for war. *World Politics*, *40*, 82–107.

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology*, *79*, 606–616.

McDonald, M. M., Navarrete, C. D., & Van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: the male warrior hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 670–679.

Puurtinen, M., Heap, S., & Mappes, T. (in press). The joint emergence of group competition and within-group cooperation. *Evolution and Human Behavior*.

Rusch, H. (2013). Asymmetries in altruistic behavior during violent intergroup conflict. *Evolutionary Psychology*, *11*, 973–993.

Rusch, H. (2014a). The evolutionary interplay of intergroup conflict and altruism in humans: A review of parochial altruism theory and prospects for its extension. *Proceedings of the Royal Society B: Biological Sciences*, *281*, 20141539.

Rusch, H. (2014b). The two sides of warfare : An extended model of altruistic behavior in ancestral human intergroup conflict. *Human Nature*, *25*, 359–377.

Rusch, H., Leunissen, J. M., & van Vugt, M. (in press). Historical and experimental evidence of sexual selection for war heroism. *Evolution and Human Behavior*.

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiment. In H. Sauermann (Ed.), (pp. 136–168). Tübingen: J. C. B. Mohr (Paul Siebeck).

Simunovic, D., Mifune, N., & Yamagishi, T. (2013). Preemptive strike: An experimental study of fear-based aggression. *Journal of Experimental Social Psychology*, *49*, 1120–1123.

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–45). Mahwah, NJ US: Erlbaum Associates.

Struch, N., & Schwartz, S. H. (1989). Intergroup aggression: Its predictors and distinctness from in-group bias. *Journal of Personality and Social Psychology*, *56*, 364–373.

Weisel, O., & Böhm, R. (in press). "Ingroup love" and "outgroup hate" in intergroup conflict between natural groups. *Journal of Experimental Social Psychology*.

Figure Captions

Figure 1. Mean points KEPT, contributed to the WITHIN pool, and contributed to the BETWEEN pool by experimental condition. Each experimental condition with $N = 24$ participants. Error bars represent 95 % confidence intervals.

Figure 2. Second-movers' mean points (**A**) KEPT, (**B**) contributed to the WITHIN pool, and (**C**) contributed to the BETWEEN pool by the expected aggregated contribution to the BETWEEN pool (*SEQ-HATE* condition) or the WITHIN pool (*SEQ-LOVE* condition) of first-movers. Each experimental condition with $N = 24$ participants. Error bars represent 95 % confidence intervals.

Figure 1.

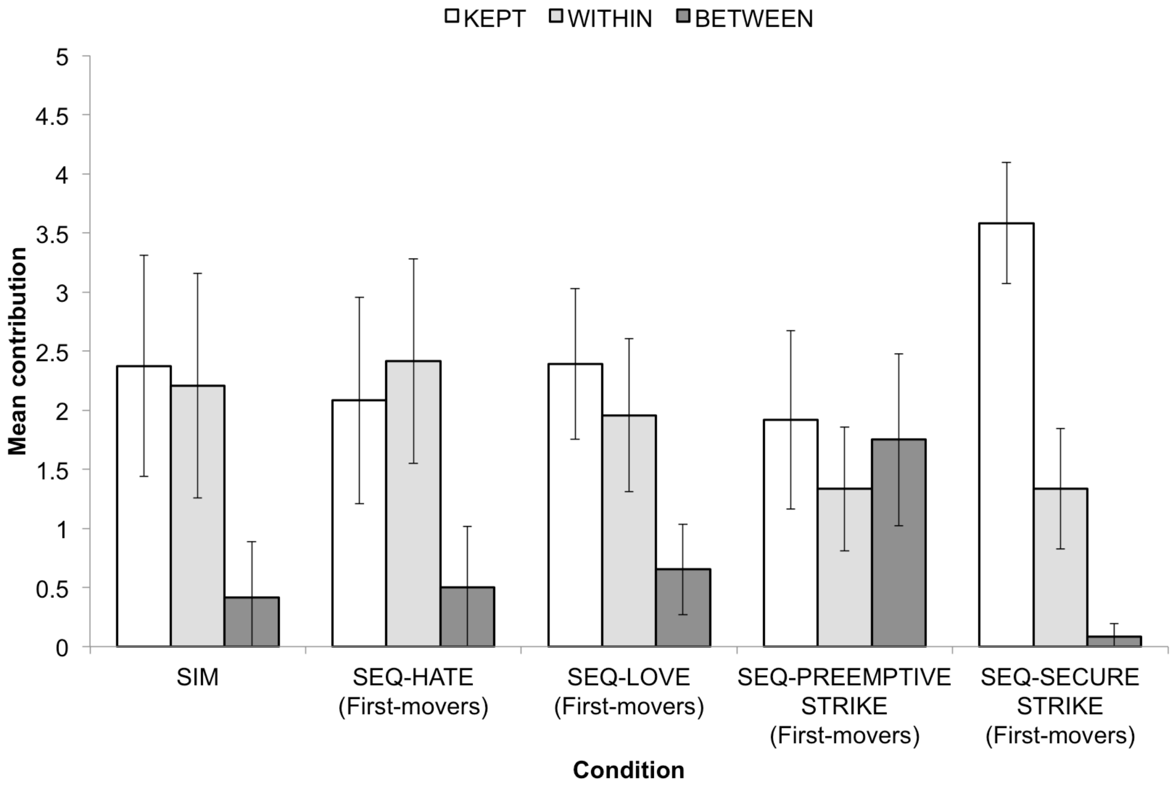
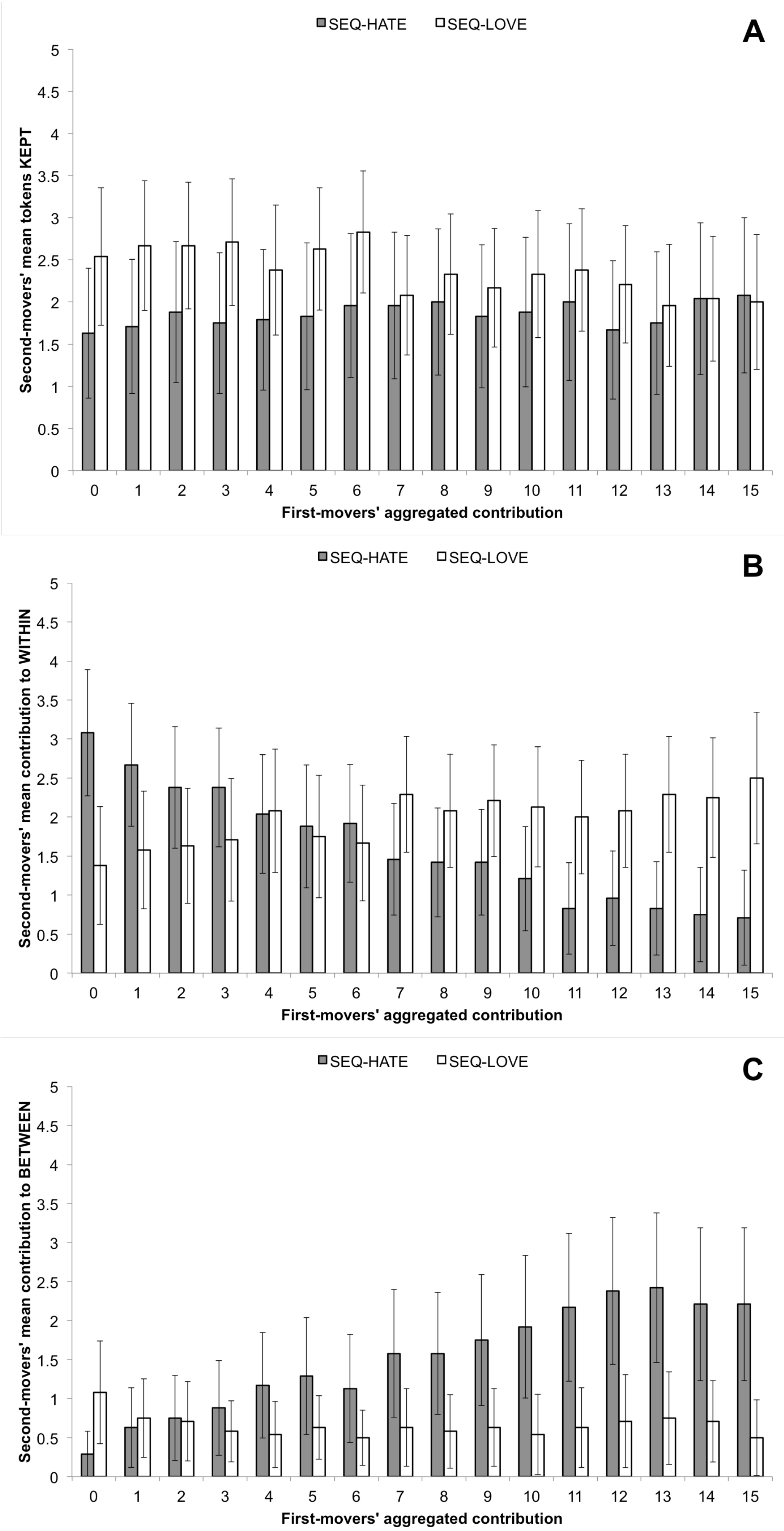

KEPT
WITHIN
BETWEEN
Mean contribution
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
SIM
SEQ-HATE (First-movers)
SEQ-LOVE (First-movers)
SEQ-PREEMPTIVE STRIKE (First-movers)
SEQ-SECURE STRIKE (First-movers)
Condition

Figure 2.


SEQ-HATE
SEQ-LOVE
A
Second-movers' mean tokens KEPT
First-movers' aggregated contribution
B
Second-movers' mean contribution to WITHIN
C
Second-movers' mean contribution to BETWEEN