JENA ECONOMIC RESEARCH PAPERS

2013 – 033

Social motives in intergroup conflict

by

Ori Weisel
Ro'i Zultan

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

Jena Economic Research Papers 2013 - 033

Social motives in intergroup conflict*

Ori Weisel
Max Planck Institute of Economics

Ro'i Zultan
Ben-Gurion University of the Negev

Abstract

We experimentally test the social motives behind individual participation in intergroup conflict by manipulating the framing and symmetry of conflict. We find that behavior in conflict depends on whether one is harmed by actions perpetrated by the out-group, but not on one's own influence on the outcome of the out-group. The way in which this harm is presented and perceived dramatically alters participation decisions. When people perceive *their group* to be under threat, they are mobilized to do what is good for the group and contribute to the conflict. On the other hand, if people perceive to be *personally* under threat, they are driven to do what is good for themselves and withhold their contribution. The first phenomenon is attributed to group identity, possibly combined with a concern for social welfare. The second phenomenon is attributed to a novel victim effect. Another social motive—reciprocity—is ruled out by the data.

Keywords
intergroup conflict, intergroup prisoner's dilemma, asymmetric conflict, framing.
JEL classification codes
C72, C92, D03, D62, D74, H41

*We dedicate this paper to the late Gary Bornstein, without whom our understanding of human behavior in group conflict would be greatly impoverished. Financial support from the Max Planck Society is gratefully acknowledged. We thank Klaus Abbink, Werner Güth, Glenn Harrison, Yan Chen, and participants in ESA meetings in Cologne and Tucson, IMEBE, EssexLab inaugural workshop and seminars in Jena and Jerusalem for helpful comments and discussion.

1 Introduction

Conventional knowledge across the social sciences states that intergroup conflict increases intragroup cooperation (Campbell, 1965; Coser, 1956; Sherif, 1961; Stein, 1976). This hypothesis gains support from field studies (e.g., Bauer et al., 2013; Penner et al., 2005) and laboratory experiments (Bornstein and Ben-Yossef, 1994; Halevy et al., 2012). The current paper extends existing research by studying *asymmetric conflict games*, in which only one group harms the other, but not vice versa, under two alternative presentations. In one presentation conflict is presented as a competition between groups, while in the other as a game of harm imposed by (individual) members of one groups on members of the other. The new manipulations allow us to disentangle and test different underlying social motives for cooperation in intergroup conflict and study how they alter depending on the way conflict is presented and perceived. We test three hypothesized social motives: That common fate of an attacked group breeds *group identity*, which in turn leads to cooperation; that *social welfare* concerns lead individuals to refrain from participation in conflict; and that *reciprocal preferences* lead to increased participation in two-sided conflict.

The strategic aspects of conflict between groups involve two levels; the *intergroup* level, in which collective action in one group has adverse effects on the aggregate welfare of the other group, and the *intragroup* level, where collective action is attained through mobilization of individual group members. Collective action in group conflict based on voluntary individual participation is acknowledged to pose a particularly puzzling challenge to rational choice theory (Blattman and Miguel, 2010; Gould, 1999; Olson, 1974). The benefits reaped by the group (e.g., resources, political power) form a classic public good, as they are typically freely available to all group members regardless of their contribution to the group effort. Since the participating individual incurs a personal cost that can be much higher than the marginal benefit, a rational group member has strong incentives to free ride on the contributions of others, leading to an inferior outcome for the group (Bornstein, 1992, 2003; Hardin, 1997; Rapoport and Bornstein, 1987). On the other hand, while collective effort improves the group outcome, it reduces social welfare as resources invested in the conflict by both groups counter each other with destructive results.

These simultaneous intra- and inter-group processes have been modelled and studied experimentally using team games (Bornstein, 2003). In team games, each individual can contribute to her group effort (at a cost to herself). Contributions generally increase payoffs in the individual's in-group and reduces the payoffs in her out-group.[1] The tension between the intra- and inter-group levels of strategic interactions is best captured by the *Intergroup Prisoner's Dilemma* (IPD) game (Bornstein, 1992). The standard two-player *Prisoner's Dilemma* (PD) game has long been used to model both intergroup conflicts (Axelrod, 1984; Brams, 1975) and problems of collective action (Lichbach, 1996; Ostrom, 2000). By embedding an intra-group PD game within an intergroup PD game, the IPD game serves to study issues of collective action in intergroup conflict.

Bornstein and Ben-Yossef (1994) compared behavior in the IPD to behavior in corresponding PD games

[1] Games that have been studied as team games include Voting games (Palfrey and Rosenthal, 1983), the Rent-seeking game (Abbink et al., 2010), and Assurance and Chicken games (Bornstein et al., 1996).

to find that cooperation in the intra-group social dilemma substantially increased when the intra-group game was embedded in intergroup conflict. In explaining this phenomenon, Bornstein and Ben-Yossef (1994) built on social identity theory (Tajfel and Turner, 1979) to argue that intergroup conflict nurtures group identity by creating common fate, leading individuals to act in line with their group's interest. Pervious research established that direct manipulations of common fate affect contributions in social dilemmas. for example, when the payoff level of all members of a group is determined by a single random draw, cooperation is higher than when the payoff levels of individuals or subgroups are determined independently (Kramer and Brewer, 1984; Wit and Wilke, 1992). The collective action of the opposing out-group similarly induces common fate, as it affects the payoffs in the in-group as a whole, and hence is expected to result in high cooperation.[2]

Group identity, however, is not the only social motive involved in group conflict. Reciprocity, for example, has been shown to drive behavior across many social interactions (Fehr and Gächter, 2000).[3] The high level of contributions in the IPD can accordingly be explained as the manifestation of negative reciprocity between individuals belonging to opposing groups. In the appendix we develop a model based on Rabin (1993) that incorporate reciprocal preferences into individuals' utility functions and show that, consequently, expected contribution levels are higher in the IPD than in the PD.

Another important social motive that emerges from the experimental and behavioral literature is a taste for maximizing social welfare or efficiency, as people are willing to incur a small personal cost if it results in a larger benefit to others (e.g., Charness and Rabin, 2002; Engelmann and Strobel, 2004; Kritikos and Bolle, 2001).[4] Perhaps somewhat oddly, the opposite is observed in the context of intergroup conflict, where contributions are higher in the less efficient IPD (where it is necessary to harm the out-group in order to benefit the in-group) in comparison with the PD (where the in-group benefit does not entail harming the out-group). Nonetheless, preferences for social welfare maximization can explain another result in the context of group conflict. In the *Intergroup Prisoner's Dilemma-Maximizing Differences* (IPD-MD) game, players who wish to contribute in order to increase the payoff of their in-group can allocate their contribution between two pools. Contributions to the *within-group pool* do not (negatively) affect the payoff of the out-group, as in the PD, whereas contributions to the *between-group pool* do, as in the IPD (Halevy et al., 2008). Several experiments found that contributions to the inefficient between-group pool are substantially lower than contributions to the efficient within-group pool (De Dreu et al., 2010; Halevy et al., 2008, 2012).

Why do people tend to engage in conflict in the IPD game of Bornstein and Ben-Yossef (1994), but choose to avoid conflict in the IPD-MD game of Halevy et al. (2008)? We see three possible explanations. The explanation put forward by Halevy et al. (2008) is that group identity induced by conflict fosters in-group

[2] Baron (2001) argued that individuals do not contribute more in intergroup conflict because they care about the group outcome but are led to falsely believe that contributions increase their personal payoff.

[3] In the one-shot simultaneous games that we study, players are not able to reciprocate observed actions by others. We nonetheless use the term *reciprocity* to describe reciprocal preferences that sustain cooperation in equilibrium when actions are taken to reciprocate *expected* actions by others. A static equilibrium notion based on a 'principle of reciprocity' was offered by Sugden (1984). Fischbacher and Gächter (2010) have argued, based on experimental evidence, that people choose their contribution to a public good as if they reciprocate the contributions they expect others to make.

[4] In this paper we do not distinguish between concern for efficiency—maximizing aggregate payoffs—and for social welfare—maximizing aggregate utilities—since both notions are essentially equivalent in the conflict games that we study.

love but not out-group hate, such that people have no particular interest in harming the out-group unless it is essential for helping the in-group. Another possible explanation is that people have a preference for social welfare maximization, which is not apparent in the IPD due to the stronger effects of group identity. In the IPD-MD, however, this preference induces aversion to the between-group pool, as individuals can satisfy both their desire to benefit in-group members and to increase social welfare by directing their contributions at the within-group pool.

A third explanation for the discrepancy between the two types of studies hangs on differences in methodology. The experiments that found increased contributions in the IPD compared to the PD presented the game as a comparison between the total contributions made by the members of the two groups. The experiments on the IPD-MD, in comparison, distinguished in their experimental instructions between the direct effects of individual contributions on in- and out-group members, without explicitly comparing contributions in the two groups. In fact, no published study has hitherto compared contributions in the PD and the IPD using the same presentation in the IPD-MD studies. The current study remedies this situation.

1.1 Aims and hypotheses

This paper aims to identify the unique effects of group identity, reciprocity and social welfare considerations (henceforth *GI*, *REC*, and *SW*, respectively) and to explore how these social motives interact with the way the conflict is presented and perceived. To this end, we introduce a one-way conflict version of the IPD game, the *Asymmetric intergroup Prisoner's Dilemma* (AIPD). In the AIPD, contributions made by members of one group, which we label the *Attacker* group, decrease the payoffs of members of the other group, which we label the *Victim* group, as in the IPD. Contributions made by members of the Victim group only affect payoffs within the group, as in the PD. In other words, the payoffs of the members of the Attacker group are determined as in the PD (they are not affected by the out-group), and the payoffs of the members of the Victim group are determined as in the IPD (they are affected by the out-group).

The social motives considered above diverge in their predictions with regard to contributions in the different games. We accordingly draw our main hypotheses, also summarized in Table 1:

Hypothesis 1. Group identity: *Cooperation is higher in the IPD and in the Victim group, in which group members share a common fate due to the joint effect of contributions in the out-group on their payoff.*

Hypothesis 2. Reciprocity: *Cooperation is higher in the IPD, in which there is a reciprocal relationship between the two groups.*

Hypothesis 3. Social welfare: *Cooperation is higher in the PD and the Victim group, in which contributions do not have a negative effect on out-group members.*

These hypotheses can be combined to generate new predictions for combinations of any two motives. Table 2 summarizes the different theoretical predictions.[5]

Table 1: **Effects of social motives on contribution**

Social motive	PD	IPD	Attacker	Victim
GI	$-$	$+$	$-$	$+$
REC	$-$	$+$	$-$	$-$
SW	$+$	$-$	$-$	$+$

Note: The "$+$" signs indicate that the social motive (row) increases contributions in the game (column).

Table 2: **Predicted contribution levels for different social motives**

Social motive(s)	Predicted contribution levels						
GI	IPD	$=$	$Victim$	$>$	PD	$=$	$Attacker$
REC	IPD	$>$	$Victim$	$=$	PD	$=$	$Attacker$
SW	$Victim$	$=$	PD	$>$	IPD	$=$	$Attacker$
GI & REC	IPD	$>$	$Victim$	$>$	PD	$=$	$Attacker$
GI & SW	$Victim$	$>$	IPD	$>=<$	PD	$>$	$Attacker$
REC & SW	IPD	$>=<$	PD	$=$	$Victim$	$>$	$Attacker$

Notes: Contributions in different games are assumed to be identical when the same set of social motives applies ("$=$"); contributions in a given game are assumed to be higher than in another when the relevant social motives in the latter are a (proper) subset of those in the former ("$>$"); we do not assume anything about the relative effect of *different* social motives on contributions ("$>=<$").

Our second objective in this study is to compare the different ways in which the experimental game of conflict has been presented in previous experiments to determine how perceptions of the nature of the conflict affect behavior. Bornstein and Ben-Yossef (1994) used what we refer to as the *Comparison* frame to describe the payoffs of the IPD, and found more cooperation in the IPD as compared to the PD. Under a *Comparison* frame payoffs are determined by a player's choice (to contribute or not) and by the difference in the number of contributors between her group and the other group, as in Table 3. In contrast, Halevy et al. (2008) used an *Individual Harm* frame and found that people refrain from contributing to an IPD-type pool. In an *Individual Harm* frame the payoffs are determined by the direct externalities (negative or positive) that individual contribution has on members of the two groups. We manipulate the framing to explicitly test whether contributions in the IPD depend on the way in which the game is presented. To explain the existing results, we conjecture that the Comparison frame, in which the actions and payoffs are presented at the group level, induces an increased sense of common fate, which triggers group identity. Furthermore, in the Individual Harm frame, in which actions and payoffs are presented at the individual level, the harm inflicted on out-group members is brought to the fore and social welfare considerations are triggered. Our next hypotheses reflect this conjecture.

Hypothesis 4a. Comparison frame: *The effects of group identity described in Hypothesis 1 are stronger under the comparison frame.*

Hypothesis 4b. Individual harm frame: *The effects of social welfare maximization described in Hypothesis 3 are stronger under the individual harm frame.*

The effects of conflict on cooperation depend on individual characteristics. For example, Probst et al. (1999) found that contributions in the IPD are either higher or lower than in the PD, depending on individually held cultural values. To account for individual differences, we measure participants' social value orientation (SVO) using the slider measure (Murphy et al., 2011) and their willingness to punish by eliciting minimum acceptable offers (MAO) in an ultimatum game. SVO increases with the willingness to forgo a personal payoff in order to increase others' welfare, hence, it is expected to be correlated with concern for SW. GI is also a pro-social tendency, and is therefore expected to be moderated by SVO. REC, on the other hand, is conceptually orthogonal to SVO, and is expected to be correlated with willingness to retaliate in the ultimatum game. Accordingly, we draw the following hypotheses:

Hypothesis 5a. Social value orientation and SW: *The effects of social welfare maximization described in Hypothesis 3 are stronger as the SVO increases.*

Hypothesis 5b. Social value orientation and GI: *The effects of group identity described in Hypothesis 1 are stronger as the SVO increases.*

[5]The qualitative prediction of a combination of all three motives cannot be distinguished from that of a combination of just GI and SW due to the conflicting effects in the IPD.

Hypothesis 6. RETALIATION: *Participants who are more willing to reject low offers in the ultimatum game are more likely to contribute (only) in the IPD.*

2 Experimental design and procedure

The computerized experiment was conducted at the experimental economics lab in Jena and programmed in z-Tree (Fischbacher, 2007). Four hundred and forty four participants were recruited using ORSEE (Greiner, 2004). Sessions lasted approximately one hour, and the average payoff was 15€, including a showup fee of 2.50€.

Each experimental session included three stages comprised of the conflict game, an ultimatum game, and the SVO slider measure. The participants were informed at the beginning of the experiment that there will be three independent stages, but not of their content.

2.1 The conflict game (stage 1)

We manipulated group type and the way conflict is framed (presented) in a 2 x 4 between-subjects design with group type (PD / IPD / Attacker / Victim) and framing (Comparison / Individual harm) as independent variables. At the beginning of the stage, participants were randomly allocated to pairs of groups, each with three members. The groups in each pair were labeled Group A and Group B. Each group member was endowed with 140 ECU (Experimental Currency Unit). Contribution carried a fixed cost of 50 ECU and increased the payoff of each group member, including the contributor, by 30 ECU. Contributions made in the IPD and by the Attacker group in the AIPD additionally reduced the payoff of each out-group member by 30 ECU.[6] Accordingly, the payoff function for participants in the PD and in the Attacker group was

$$\pi_i = 140 - 50c_i + 30\sum_{j\in I} c_j, \tag{1}$$

where $c_i \in \{0, 1\}$ is i's contribution decision and I denotes i's in-group. The corresponding payoff function for participants in the IPD and in the Victim group was

$$\pi_i = 140 - 50c_i + 30\sum_{j\in I} c_j - 30\sum_{k\in O} c_k, \tag{2}$$

where O denotes i's outgroup.

The instructions in the Individual harm frame explained the direct effect of contribution on the contributor, her in-group members and her out-group members. The instructions in the comparison frame treatments

[6] To keep the instructions identical across treatments and to control for the mere existence of another group, the instructions in the PD also referred to Group A and Group B.

Table 3: Payoff tables in the Comparison frame

(a) PD and Attacker payoffs

contributions in the group	**0**	**1**	**2**	**3**
Contribute	-	120	150	180
Not contribute	140	170	200	-

(b) IPD and Victim payoffs

Contributions in the group minus contribution in the other group	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**
Contribute	-	30	60	90	120	150	180
Not contribute	50	80	110	140	170	200	-

presented the game payoffs in tables following Bornstein and Ben-Yossef (1994).[7] In the PD and Attacker groups, the payoffs were presented as a function of the *number of contributors* in the group and the individual decision. In the IPD and Victim groups, the payoff was presented as a function of the *difference in the number of contributors* between the in-group and the out-group and of the individual decision (to contribute or not). Table 3 reproduces the payoff tables shown to participants. The instructions for the PD and IPD sessions included and explained the relevant payoff table. Instructions for the AIPD sessions included both payoff tables.

In all treatments, the explanation of the payoff structure was followed by a detailed example. Participants indicated that they understood the instructions and were then required to calculate their payoff in two hypothetical situations correctly before the experiment proceeded.

Next, participants made their contribution decisions, after which they were asked to guess how many members of their in-group and their out-group chose to contribute. Guesses were not paid. This concluded the first stage.

2.2 The Ultimatum game (stage 2)

In the second stage, participants played a strategy-method ultimatum game (UG), in which each player played both roles. First, each participant i played the role of the *responder* in the UG by indicating the minimal amount of ECU, $r_i \in \{10, 20, ...190\}$, she is willing to accept.[8] Second, each participant made a decision in the role of the *proposer* in the UG and suggested a division of 200 ECU between herself and her partner j by selecting $p_i \in \{10, 20, ...190\}$ ECU for herself, with the rest, $200 - p_i$, going to the partner. Participants

[7] See the appendix for an English translation of the instructions. The German original is available upon request.

[8] We also asked participants what is the maximal amount they wish to reject, to ensure proper understanding.

were randomly paired and the payoff relevant roles were determined randomly, with one member of the pair playing as proposer and the other as responder. If the proposer's (i) division was accepted by the responder (j), i.e., $200 - p_i \geq r_j$, the proposer received p_i ECU and the responder received $200 - p_i$ ECU. If it is rejected, i.e., $200 - p_i < r_j$ then both the proposer and the responder received nothing.

The instructions for the second stage explained the rules of the UG, the way participants were paired and selected to the roles of proposer or repeonder, and included an example for illustration. Participants indicated that they understood the instructions and then proceeded to make their decision.

2.3 The slider measure

In the third and final stage participants completed the Social Value Orientation slider measure (Murphy et al., 2011). In this measure, which aims to measure the magnitude of the concern people have for others, each participant makes a series of 15 choices between nine allocations of ECU between herself and another participant (i.e., in each choice there were nine available allocations). Choices were than aggregated to determine a unique value for each participant, expressed as an angle on the self/other two-dimensional space. A value of zero indicates perfect selfishness. Higher values indicate stronger regard to the payoff of others.

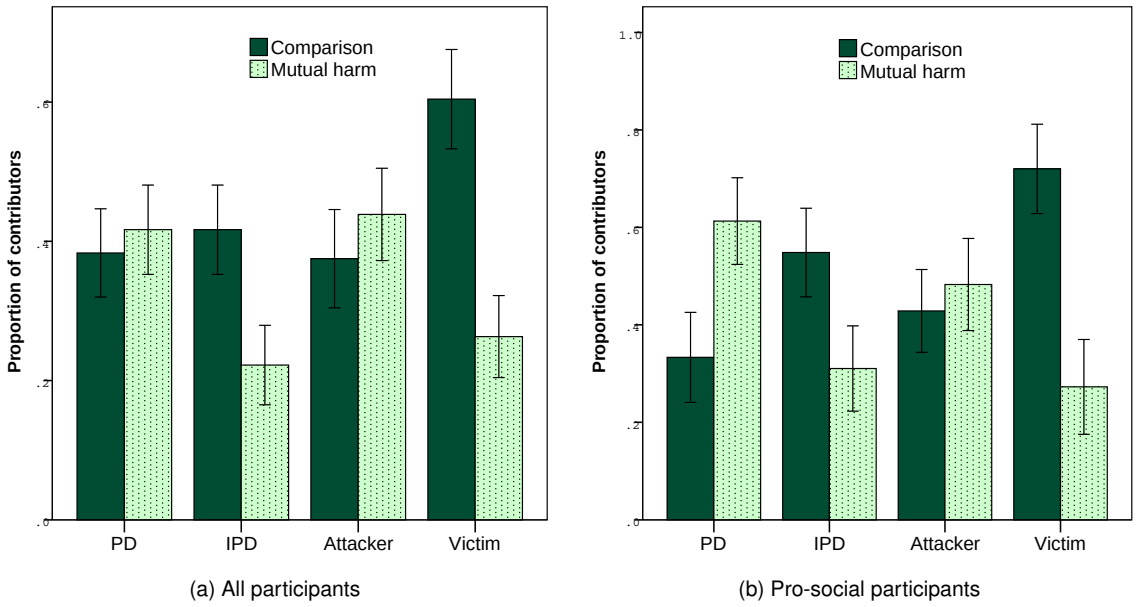
3 Results

Figure 1 presents the proportions of participants contributing in the different treatments, for all participants (Panel 1a) and for pro-social participants only (Panel 1b).[9] Table 4 presents the results of a series of logistic regressions taking the probability of contribution as the dependent variable. Columns (1) and (2) describe separate models for the two frames. Column (3) presents an aggregate model, and Column (4) adds the SVO and its interactions. The MAO data collected in the ultimatum game of Stage 2 did not predict behavior in the conflict games and are therefore omitted from the analysis. Because the results are markedly different for the two frames, we describe the results separately for the Comparison and for the Individual harm frame.

3.1 Comparison frame

Contributions levels were significantly higher in the Victim group (60.4%) than in the PD (38.3%, $z = 2.50, p = 0.012$), the IPD (41.7%, $z = 1.99, p = 0.047$), and Attacker (37.5%, $z = 2.45, p = 0.014$) groups, which were not significantly different from each other ($p > 0.500$ for all three pairwise comparisons).[10] The high proportion of contributors in the Victim group, in particular when compared to our baseline PD treatment, is in line with the Group Identity Hypothesis 1. Recall that group identity was originally invoked to explain the high contributions in the IPD game observed by Bornstein and Ben-Yossef (1994), a pattern not

[9] Individuals are classified as pro-social if their social value orientation lies in the interval $[37.09°, 52.91°]$ (Murphy et al., 2011).
[10] Pairwise z- and p-values were computed based on model (4) in table 4.



Note: error bars indicate 95% confidence intervals.

Figure 1: Proportions of contributors

Table 4: Logistic regressions on the probability of contribution

	(1) Comparison	(2) Individual harm	(3) All data	(4) all data
IPD	0.139		0.139	1.314*
	(0.373)		(0.373)	(0.771)
Victim	0.898**		0.898**	2.413***
	(0.397)		(0.397)	(0.872)
Attacker	-0.0354		-0.0354	0.655
	(0.399)		(0.399)	(0.778)
Individual harm			0.139	2.128***
			(0.373)	(0.802)
Individual harm x IPD		-0.916**	-1.055*	-2.781**
		(0.419)	(0.561)	(1.172)
Individual harm x Victim		-0.693*	-1.591***	-4.125***
		(0.399)	(0.563)	(1.247)
Individual harm x Attacker		0.0896	0.125	-1.616
		(0.374)	(0.547)	(1.122)
SVO				-0.0186
				(0.0209)
IPD x SVO				0.0520*
				(0.0294)
Victim x SVO				0.0663**
				(0.0326)
Attacker x SVO				0.0322
				(0.0321)
Individual harm x SVO				0.0910***
				(0.0318)
Individual harm x IPD x SVO				-0.0782
				(0.0483)
Individual harm x Victim x SVO				-0.118**
				(0.0478)
Individual harm x Attacker x SVO				-0.0837*
				(0.0457)
Constant	-0.475*	-0.336	-0.475*	-0.917
	(0.266)	(0.262)	(0.266)	(0.570)
N	216	228	444	444

Notes: Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. SVO baseline is set at 45°.

replicated in our data. Thus, the lack of significant difference between the PD and IPD treatments deviates both from previous findings and from the predictions of Hypothesis 1. Nonetheless, these deviations can be accounted for when considering the interactions of the treatments with the SVO measure presented in Column (4) of Tables 4. We see that participants who score high on the SVO measure are more likely to contribute in the IPD than in the PD, as in Bornstein and Ben-Yossef (1994).[11] The fact that the behavior observed by Bornstein and Ben-Yossef (1994) was similar to the behavior we observe in highly pro-social participants suggests higher levels of pro-sociality in their sample, which differed from ours not only in generation and country, but also in the recruitment procedure, which can affect the selection of pro-social individuals into participation (see, e.g., Krawczyk, 2011). Individual contributions to conflict, therefore, can be viewed as a manifestation of pro-social tendencies. This interpretation is consistent with the view that human altruism, as manifested in high pro-sociality, is inherently parochial and rooted in intergroup conflict, which gains theoretical support from evolutionary models and simulations (Bowles, 2008; Choi and Bowles, 2007) and empirical support from hormonal studies (De Dreu, 2012; De Dreu et al., 2011).

The test of Hypothesis 5b further supports group identity as the main motive driving behavior in the comparison frame. As can be seen in Figure 1b, pro-socials contribute significantly more in the IPD and Victim group, in which intergroup conflict instills common fate. This result is supported by the regression presented in Column (4) of Table 4, which provides the treatment coefficients estimated for purely pro-social participants.[12]

If group identity is the only social motive at play, we would expect to observe the same levels of contributions in the IPD and in the Victim group. The difference between the two groups can be explained by a concern for social welfare. That is, if we restrict ourselves to the PD, IPD, and Victim groups, the results are perfectly in line with a combination of GI and SW (cf. Table 2).[13] Behavior in the Attacker group is not in line with this explanation, however, as contributions are not significantly lower than in the PD and the IPD as would be predicted according to SW. This disparity can be explained in two ways. One is that members of the Attacker group, unlike players in the IPD, completely ignore the other group as it does not affect them, hence social welfare considerations do not enter their decision. Another possible explanation is that although the members of the attacker group do not share a common fate in the sense that exists in the IPD and the Victim group, they experience enhanced group identity due to having a dependent, strategically linked, out-group. This explanation, although sufficient to fully explain the contribution patterns in the comparison frame, does not explain why contributions in the Attacker group are not moderated by pro-sociality and are not sensitive to the framing of the game (as will be discussed below), two factors that theoretically and empirically should

[11] A similar dependence on individual types was previously observed by Probst et al. (1999), who found higher contributions in the IPD compared to the PD only for *vertical individualists*, classified as participants that agree with competitive statements such as 'competition is the law of nature' and 'winning is everything' (Probst et al., 1999; Singelis et al., 1995).

[12] No significant effects effects were found for participants categorized as pro-self and for the model estimates at SVO=0°, representing purely selfish individuals.

[13] Note, however, that the data in the Individual harm frame, discussed below, strongly rejects the existence of social welfare considerations. As our ex-ante Hypothesis 4b predicted that social welfare plays a larger role in the Individual harm frame, it is not clear that it indeed comes into play in the Comparison frame.

moderate group identity. Therefore, to the extent that group identity is promoted by having a strategically-linked out-group (as in the Attacker group), this effect is separate from the group identity effect that emerges from sharing a common fate (as in the IPD and Victim group), which is moderated by pro-sociality and the framing of conflict.

In sum, we conclude that, when intergroup conflict involves a comparison between groups, the conflict triggers group identity due to common fate and, possibly, also due to imposing a common fate on another group. We find indirect evidence for the existence of social welfare concerns that mitigate the effects of group identity in bilateral conflict. Reciprocity in the sense of retaliation against expected hostilities appears not to play a role in participation in conflict.

3.2 Individual harm frame

The effect of the framing of intergroup conflict is stark. While conflict had a weak positive effect on cooperation within the group in the comparison frame (i.e., slightly more cooperation in IPD relative to PD), conflict framed as individual harm had the opposite effect, with cooperation levels dropping sharply from 41.7% in the PD to 22.2% in the IPD ($z = 2.18, p0.029$). Furthermore, the interactions with SVO follow a different pattern from that observed in the comparison frame, as high SVO scores in the individual harm frame are associated with higher contribution rates only in the PD (and not in the IPD and Victim group), and do not affect the overall contribution pattern.

The Individual Harm framing boosts the saliency of the adverse effect of contributions on out-group members as well as the fact that there is no overall social gain from contributions. According to our hypotheses, the results in the PD and the IPD support a social welfare maximization motive. However, SW and its combinations with any of the other two motives predict that cooperation in the Attacker group would be lower than in the other groups (in particular, lower than the Victim group) and cooperation in the Victim group would be at least as high as in the PD (cf. Table 2). The pattern evident in the results is quite the opposite, with low cooperation levels evident in the Victim (26.3%) rather than in the Attacker group (43.9%). Thus, including the AIPD game in our design allows us to reject all three social motives as driving behavior under the individual harm frame. Rather, we find that being exposed to individual harm inflicted by members of another group led people to withdraw their contribution to the in-group. We conclude that perceiving conflict as a threat to one's self triggers selfishness and leads to low contribution rates.

4 Discussion and conclusion

The literature on in-group cooperation and collective action in conflict has so far focused on group identity as the main mediator of the effects of intergroup conflict on intra-group cooperation. The behavioral literature led us to hypothesize that collective action in conflict is driven by (negative) reciprocity between the members

of the opposing groups. Additionally, we hypothesized that these effects mask a negative effect of conflict on cooperation to the extent that cooperation is driven by social welfare maximization.

The experiment reported in this paper explicitly tested these three hypotheses by introducing an asymmetric conflict game. Our framing manipulation varied whether the harm inflicted by contributions in the out-group is targeted jointly at the in-group or separately at the individuals in the group. Our results can be summed by the following principle: *if people perceive their group to be under threat, they are mobilized to do what is good for the group and contribute to the conflict. On the other hand, if people perceive to be personally under threat, they are driven to do what is good for themselves and withhold their contribution.* These effects are apparent in the two treatments that involve being harmed by contributions in the out-group, namely the IPD and the Victim group. In line with the principle stated above, contributions in these treatments are significantly higher when the threat is presented at the group level than when the threat is presented at the individual level.

In the Comparison frame, previously used in similar laboratory studies of the IPD game, Group identity provided the best explanation for observed behavior across symmetric and asymmetric conflict, lending support to our Hypotheses 1, 4a, and 5b, which focused on group identity. Studies of evolutionary dynamics argue that parochial altruism, the willing to sacrifice for one's fellow group members, evolved as a response to intergroup conflict (Bowles, 2008; Choi and Bowles, 2007). Our results support this view and identify the effect of common fate on group identity as the underlying mechanism.

In consideration of the alternative social motives invoked by the comparison frame of intergroup conflict, the comparison of contributions in the IPD and in the Victim group served to ascertain the role of reciprocal tendencies, as both group types include an element of common fate but differ with respect to the existence of reciprocal relationships between the in-group and the out-group. The experimental results reject reciprocity as an underlying motive, as contributions are strongest in the Victim group, where contributions do not affect out-group members and therefore there is no room for intergroup reciprocity. Thus, related Hypotheses 2 and 6 were not supported.

The evidence for social welfare maximization in the comparison frame data is mixed. The markedly high contribution rates in the Victim group when compared to the IPD suggest that behavior is partially driven by a concern for SW. Hence, the central Hypothesis 3 is partially supported. Related hypotheses 4b and 5a, which predict the effects of SW to increase in the Individual harm frame and with the SVO scores, respectively, receive no support from the data.

Our study of the individual harm framing of the conflict games reveals that the way in which the game is presented and perceived carries dramatic implications for mobilization of individual group members in intergroup conflict. For pro-social individuals in particular, we find a strong interaction of the framing and conflict. When individual payoffs in group conflict are presented as a result of a comparison between groups, conflict increases contributions. Conversely, when the effects of in-group and out-group contributions on *individual* payoffs are presented, conflict reduces contributions. An alternative way to think about this result is to consider the way in which the framing affects contribution decisions in the two games. Emphasizing

the effects of contribution on individuals leads pro-social individuals to contribute more in the PD, where contributions have a positive effect on society. In contrast, this framing leads to fewer contributions in the IPD, in which contributions carry a negative effect on (some) others and on overall social welfare.

In both presentations of the games, behavior in the IPD qualitatively resembles that in the Victim, but not in the Attacker, group. In the Comparison frame, this pattern is predicted by group identity. In the Individual Harm frame, however, none of the three social motives considered is able to explain this pattern. Our interpretation is that we observe a *Victim effect*, referring to the tendency of individuals to behave selfishly when under personal threat. We consider this novel effect to be a generalized principle of *concern withdrawal*. Charness and Rabin (2002) formulated the original principle by arguing that "[subjects] withdraw their willingness to sacrifice to allocate the fair share toward somebody who himself is unwilling to sacrifice for the sake of fairness." Our results suggest that people's unwillingness to sacrifice is not a result of misbehavior on the part of the potential beneficiary of the sacrifice, but is rather a more egocentric reaction to being harmed or being exposed to harm.

Generalized concern withdrawal is linked to the notion of indirect reciprocity, and specifically *upstream reciprocity*, by which an individual who receives favorable treatment from another individual reciprocates by treating a third individual favorably (Boyd and Richerson, 1989). Upstream reciprocity is evolutionary stable in certain conditions, such as when interactions occur in small groups or when direct reciprocity is possible (Nowak and Roch, 2007; Pfeiffer et al., 2005).

Several experimental studies found that people are more likely to help others after being helped themselves (Bartlett and DeSteno, 2006; Dufwenberg et al., 2001; Greiner and Levati, 2005; Güth et al., 2001). However, unlike our experiment, these studies compared behavior conditional on the actual actions of others, whereas our study compared different game structures. This difference allows us to extend the existing knowledge in two important ways. First, the behavioral principle we identify is based on *negative* reciprocity, as in Charness and Rabin's (2002) original formulation of concern withdrawal. That is, out-group actions potentially affected personal welfare in a negative way, as our participants contributed less than in the baseline PD game. Second, the factor that drives people behavior in our setup is not actual harm. but *mere exposure to potential harm*.

The unexpected result under the Individual harm frame raises two testable hypotheses: that concern withdrawal is a special case of a general victim effect, and that exposure to harm has similar effects as actual harm. Future research is needed to test and refine these hypotheses and their implications.

Our results also have significant implications for the study of in-group love and out-group hate as driving motives in intergroup conflict (Brewer, 1999). Halevy et al. (2008) introduced the IPD-MD game as a way to disentangle the two motives associated with group identity. The study of the IPD-MD game has established that people refrain from harming out-group members when given the alternative option of cooperating within the group without affecting the out-group. Consequentially, enhanced group identity as a result of intergroup conflict was considered to induce in-group love but not out-group hate.

The current results show, however, that group identity is not invoked by conflict when framed as individual harm, as was done in the studies of the IPD-MD game. Whether out-group hate will emerge in the IPD-MD

when intergroup conflict is presented in a way that triggers group identity (e.g., using the comparison frame) is up for future experiments to determine. One observation suggests a negative answer to this question. In our experimental games, in-group love implies higher contributions in the IPD and Victim groups, whereas out-group hate should only affect the IPD players. Since the positive effect of conflict on intra-group cooperation was primarily observed in the Victim group, in-group love appears to be the major motive invoked by group identity.

Appendix A: experimental instructions

Welcome and thank you for participating in this experiment. Please remain quiet and switch off your mobile phone. Do not speak to the other participants. Communication between participants will lead to the automatic end of the session with no payment to anyone. Whenever you have a question, please raise your hand and one of the experimenters will come to your cubicle.

Please read the instructions carefully, the better you understand the instructions the more money you will be able to earn. The instructions are the same for all participants.

You will receive 2.50€for having shown up on time. The experiment allows you to earn additional money. Since your earnings during the experiment will depend on your decisions, and the decisions of the other participants.

You will receive no feedback about the decisions of the other participants and your payoff until the end of the experiment (after all three phases have ended).

During the experiment all sums of money are listed in ECU (for Experimental Currency Unit). Your earnings during the experiment will be converted to €at the end and paid to you in cash. **The exchange rate is 150 ECU = 1€.**

Instructions for the first phase

In this phase each participant is a member of a 3-person group. There are two types of groups, **A** and **B**. Each **A** group is paired with a **B** group. You will be randomly assigned to one of the two groups, **Group A** or **Group B**.

Each participant has to decide whether to contribute to his or her group account. The decisions are made independently.

Comparison frame

Instructions for PD

The payoff of each participant is determined by **his or her own decision** and the **decisions of the other members in his or her group**, but not by the decisions of the members of the other group. The exact

payoffs are calculated according to the following table:

Contributions in the group	**0**	**1**	**2**	**3**
Contribute	-	120	150	180
Not contribute	140	170	200	-

Explanation for the table:

The payoffs for a participant who chose to contribute are shown in the row marked **"contribute"**. Similarly, the payoffs for a participant who chose not to contribute are shown in the row marked **"not contribute"**.

The payoff depends on the number of contributors in the group. This number is shown in the top row of the table.

Example:

Two members of **Group A** and one member of **Group B** contributed. Therefore:

- The number of contributors in **Group A** is 2.
- In **Group A**, each contributing member receives 150 ECU (column "2", row "contribute");
- The non-contributing member in **Group A** receives 200 ECU (column "2", row "not contribute").
- The number of contributors in **Group B** is 1.
- In **Group B**, the contributing member receives 120 ECU (column "1", row "contribute");
- Each non-contributing member in **Group B** receives 170 ECU (column "1", row "not contribute").

Instructions for IPD

The payoff of each participant is determined by **his or her own decision**, the **decisions of the other members in his or her group, and the decisions of members of the other group.** The exact payoffs are provided in the following table:

Contributions in the group minus contribution in the other group	**-3**	**-2**	**-1**	**0**	**+1**	**+2**	**+3**
Contribute	-	30	60	90	120	150	180
Not contribute	50	80	110	140	170	200	-

Explanation for the table:

The payoffs for a participant who chose to contribute are shown in the row marked **"contribute"**. Similarly, the payoffs for a participant who chose not to contribute are shown in the row marked **"not contribute"**.

The payoff depends on the difference between the **number of contributors in the participant's group and the number of contributors in the other group.** This number is shown in the top row of the table.

Note that the difference is positive if there are more contributors in the participant's group and negative if there are more contributors in the other group.

Example:

Two members of **Group A** and one member of **Group B** contributed. Therefore:

- There is one more contributor in **Group A** than in **Group B** (the difference is +1).
- Each contributing member in **Group A** receives 120 ECU (column "+1", row "contribute")
- The non-contributing member in **Group A** receives 170 ECU (column "+1", row "not contribute").
- There is one less contributor in **Group B** than in **Group A** (the difference is -1).
- The contributing member in **Group B** receives 60 ECU (column "-1", row "contribute")
- Each non-contributing member in **Group B** receives 110 ECU (column "-1", row "not contribute").

Instructions for AIPD

The payoff of each participant is determined by **his or her own decision**, the **decisions of the other members in his or her group, and possibly by the decisions of members of the other group.** The exact payoffs are provided in the following table:

Group A payoffs

Contributions in the group minus contribution in the other group	**-3**	**-2**	**-1**	**0**	**+1**	**+2**	**+3**
Contribute	-	30	60	90	120	150	180
Not contribute	50	80	110	140	170	200	-

Group B payoffs

Contributions in the group	**0**	**1**	**2**	**3**
Contribute	-	120	150	180
Not contribute	140	170	200	-

Explanation for the table:

The payoffs for a participant who chose to contribute are shown in the row marked **"contribute"**. Similarly, the payoffs for a participant who chose not to contribute are shown in the row marked **"not contribute"**.

The payoff of **Group A** members depends on the difference between the **number of contributors in the participant's group and the number of contributors in the other group.** This number is shown in the top row of the "Group A payoffs" table.

Note that the difference is positive if there are more contributors in **Group A** and negative if there are more contributors in the other group.

The payoff of **Group B** members depends on the number of contributors in the group. This number is shown in the top row of the "Group B payoffs" table.

Example:

Two members of **Group A** and one member of **Group B** contributed. Therefore:

- There is one more contributor in **Group A** than in **Group B** (the difference is +1).
- Each contributing member in **Group A** receives 120 ECU (column "+1", row "contribute")
- The non-contributing member in **Group A** receives 170 ECU (column "+1", row "not contribute").
- The number of contributors in **Group B** is 1.
- In **Group B**, the contributing member receives 120 ECU (column "1", row "contribute");
- Each non-contributing member in **Group B** receives 170 ECU (column "1", row "not contribute").

MUTUAL HARM FRAME

Instructions for PD

The payoff of each participant is determined by **his or her own decision** and the **decisions of the other members in his or her group**, but not by the decisions of the members of the other group. The exact payoffs are calculated in the following manner:

Each participant receives an initial sum of 140 ECU. If a participant contributes, then each member of **his or her group** (including him- or herself) **gains 30 ECU.**

A participant who contributes to the group account has to pay 50 ECU.

Example:

Two members of **Group A** and one member of **Group B** contributed.

Therefore, in **Group A**, each contributing member

Receives 140 ECU as the initial endowment	$+140$
Has to pay 50 ECU for his contribution	-50
Receives 30 ECU for each of the two contributions in Group A	$+(2 \times 30) = +60$
And earns a total of 150 ECU	**150**

The non-contributing member in **Group A**

Receives 140 ECU as the initial endowment	$+140$
Receives 30 ECU for each of the two contributions in Group A	$+(2 \times 30) = +60$
And earns a total of 200 ECU	**200**

In **Group B**, the contributing member

Receives 140 ECU as the initial endowment	$+140$
Has to pay 50 ECU for his contribution	-50
Receives 30 ECU for the single contribution in Group B	$+30$
And earns a total of 120 ECU	**120**

Each non-contributing member in **Group B**

Receives 140 ECU as the initial endowment	$+140$
Receives 30 ECU for the single contribution in Group B	$+30$
And earns a total of 170 ECU	**170**

Instructions for IPD

The payoff of each participant is determined by **his or her own decision**, the **decisions of the other members in his or her group, and the decisions of members of the other group.** The exact payoffs are calculated in the following manner:

Each participant receives an initial sum of 140 ECU. If a participant contributes, then each member of **his or her group** (including him- or herself) **gains 30 ECU**. Additionally, each member of the **other group loses 30 ECU**.

A participant who contributes to the group account has to pay 50 ECU.

Example:

Two members of **Group A** and one member of **Group B** contributed.

Therefore, in **Group A**, each contributing member

Receives 140 ECU as the initial endowment	$+140$
Has to pay 50 ECU for his contribution	-50
Receives 30 ECU for each of the two contributions in Group A	$+(2 \times 30) = +60$
Loses 30 ECU for the contribution in Group B	-30
And earns a total of 120 ECU	**120**

The non-contributing member in **Group A**

Receives 140 ECU as the initial endowment	$+140$
Receives 30 ECU for each of the two contributions in Group A	$+(2 \times 30) = +60$
Loses 30 ECU for the contribution in Group B	-30
And earns a total of 170 ECU	**170**

In **Group B**, the contributing member

Receives 140 ECU as the initial endowment	$+140$
Has to pay 50 ECU for his contribution	-50
Receives 30 ECU for the single contribution in Group B	$+30$
Loses 30 ECU for each of the two contributions in Group A	$-(2 \times 30) = -60$
And earns a total of 60 ECU	**60**

Each non-contributing member in **Group B**

Receives 140 ECU as the initial endowment	$+140$
Receives 30 ECU for the single contribution in Group B	$+30$
Loses 30 ECU for each of the two contributions in Group A	$-(2 \times 30) = -60$
And earns a total of 110 ECU	**110**

Instructions for AIPD

The payoff of each participant is determined by **his or her own decision**, the **decisions of the other participants in his or her group, and possibly by the decisions of members of the other group.** The exact payoffs are calculated in the following manner:

Each participant receives an initial sum of 140 ECU. If a participant in **Group A** contributes, then each member of **his or her group (A)** (including him- or herself) receives an additional 30 ECU. If a participant in **Group B** contributes, then each member of **his or her group (B)** (including him- or herself) receives an additional 30 ECU. Additionally, each member of **Group A** loses 30 ECU.

A participant who contributes to the group account (in both groups) has to pay 50 ECU.

Example:

Two members of **Group A** and one member of **Group B** contributed.+

Therefore, in **Group A**, each contributing member

Receives 140 ECU as the initial endowment	$+140$
Has to pay 50 ECU for his contribution	-50
Receives 30 ECU for each of the two contributions in Group A	$+(2 \times 30) = +60$
Loses 30 ECU for the contribution in Group B	-30
And earns a total of 120 ECU	**120**

The non-contributing member in **Group A**

Receives 140 ECU as the initial endowment	$+140$
Receives 30 ECU for each of the two contributions in Group A	$+(2 \times 30) = +60$
Loses 30 ECU for the contribution in Group B	-30
And earns a total of 170 ECU	**170**

In **Group B**, the contributing member

Receives 140 ECU as the initial endowment	$+140$
Has to pay 50 ECU for his contribution	-50
Receives 30 ECU for the single contribution in Group B	$+30$
And earns a total of 120 ECU	**120**

Each non-contributing member in **Group B**

Receives 140 ECU as the initial endowment	$+140$
Receives 30 ECU for the single contribution in Group B	$+30$
And earns a total of 170 ECU	**170**

If you have read and understood the instructions, please indicate so on your computer screen.
If you have any questions, please raise your hand and an experimenter will come to you.

Instructions for the second phase

In this phase, you will be randomly assigned to be in **Role X** or in **Role Y**. Each participant in **Role X** will be randomly matched with a participant in **Role Y**.

In this phase, there are 200 ECU that can be gained by the two participants according to the following rules:

The Participant in Role **X** chooses a division of the 200 ECU between him- or herself and the participant in role **Y**. The division must be in units of 10 ECU, and allocate at least 10 ECU to each participant. In other words, **X** chooses a sum of 10, 20, 30,. . . , 190 ECU to take out of the 200 ECU and give to him- or herself, with the rest going to **Y**.

The participant in Role **Y** decides which divisions he or she accepts and which he or she rejects, by indicating the minimal amount he or she is willing to accept. By indicating this amount, it automatically follows that **Y** accepts this amount or any higher amount, and rejects any lower amount.

Once the decisions of both participants have been made, the computer will check whether the participant in Role **Y** accepts the division chosen by the participant in Role **X**.

- If **Y** accepts, then both players will receive the payoff according to the division.
- If **Y** rejects, then none of the players receive any payoff in this phase of the experiment.

You will make decisions both in the role of X and in the role of Y. Later, the computer will randomly determine your role, and your payoff in the phase will be calculated according to your decision in this role.

Example:

The participant in role **Y** indicated that he or she accepts at least 110 ECU. This means that divisions that give **Y** less than 110 ECU (10, 20, 30,. . . , 100) are rejected, and divisions that give **Y** 110 ECU or more (110, 110, 120,. . . , 200) are accepted.

Suppose that the role **X** participant chooses to divide the 200 ECU by taking 150 ECU to him- or herself, and leaving 50 ECU to **Y**. Since 50 ECU is less than **Y**'s minimal acceptance amount (110 ECU), the division will be rejected by **Y**, and both **X** and **Y** will not receive any payoff.

Suppose that the role **X** participant chooses to divide the 200 ECU by taking 60 ECU to him- or herself, and leaving 140 ECU to **Y**. Since 140 ECU is more than **Y**'s minimal acceptance amount (110 ECU), the

division will be accepted by **Y**, and both **X** and **Y** will receive the payoff according to the division. In this example, **X** would receive 60 ECU and **Y** 140 ECU.

If you have read and understood the instructions, please indicate so on your computer screen.
If you have any questions, please raise your hand and an experimenter will come to you.

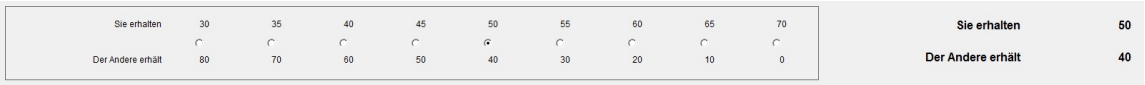
Instructions for the third phase

In this phase you will make a series of decisions about allocating resources (ECU) between yourself and another person. For each of the following items, please indicate the distribution you prefer most by clicking the respective position.

There are no right or wrong answers, this is all about personal preferences.

In the example below, a person has chosen to distribute the resources so that he/she receives 50 ECU, while the other person receives 40 ECU.

Example:



After all participants have made their decisions you will be randomly assigned to be an "Allocator" or a "Recipient". If you are an allocator then one of your decisions (randomly chosen) will determine your payoff and the payoff of another participant. If you are a recipient then your payoff will be determined by one of the other participants.

If you have read and understood the instructions, please indicate so on your computer screen.
If you have any questions, please raise your hand and an experimenter will come to you.

Appendix B: reciprocity model

To analyze reciprocity in the conflict games, we modify the fairness equilibrium concept due to Rabin (1993) in two ways.[14] First, we extend the original model to apply to N players. Although this extension is non-trivial in general, it is straightforward in the conflict games due to the separability of externalities. In all three games, the effect of any player i on the payoff of any player j is a constant and independent of the actions of all other players. Therefore i's reciprocity utility can be taken to be the sum of the reciprocity utilities obtained from all pairwise interactions as defined in Rabin (1993).

Our second extension is to allow for individual reciprocity parameters. This allows us to compare equilibrium behavior of heterogenous populations in the PD and the IPD.[15] In the next section, we define the modified fairness equilibrium. The subsequent sections compare the two games under the assumption that all individuals in the population share a (commonly known) taste for reciprocity and under the assumption that reciprocity preferences vary between individuals.

4.1 Model

Rabin (1993) defined fairness equilibrium as a *psychological Nash equilibrium* (Geanakoplos et al., 1989) in the game resulting from adding a reciprocity term to the players' utility functions. Player i's utility gained from reciprocity to player j is defined as the product of the kindness shown by i to j and the kindness that i believes j is showing to her. Thus, if i believes that j is kind to her, she wishes to return kindness, whereas if she believes that j is unkind to her, she wishes to be unkind in return.

We adopt Rabin's (1993) definition of the kindness function f_{ij}, denoting i's kindness towards j, to the PD and the IPD games. We start by defining the basic terminology. Let S_i be the space of (mixed) strategies for player i. Write $s_i \in S_i$ for player i's strategy, write $s_{ij} \in S_j$ for i's belief regarding j's strategy, and write $s_{ijk} \in S_k$ for i's belief regarding j's belief over k's strategy.[16] Trivially, $s_i = s_{ii} = s_{iii}$.

Let $\pi_j(s)$ be player j's payoff given the profile of strategies $s = (s_1, s_2, \ldots, s_n)$, as defined in the main text. Hence $\pi_j(s_{i\cdot})$ is the payoff of j induced by i's beliefs. Write $\Pi_{ij} \equiv \{\pi_j(s_{i\cdot}) | s_i \in S_i\}$ for the set of feasible payoffs for j if all players other than i play according to i's beliefs. Let π^h_{ij} and π^l_{ij} be the highest and lowest payoffs in Π_{ij}, respectively. Write π^e_{ij} for the "equitable payoff" $(\pi^h_{ij} + \pi^l_{ij})/2$.[17] The kindness

[14] See Dufwenberg and Kirchsteiger (2004) for a similar treatment in the domain of extensive form games.

[15] The AIPD is equivalent in this sense to the PD, as reciprocal relationships only exist within a group.

[16] Note that in the conflict games, $S_i \equiv S_j$ for any i and j.

[17] Rabin (1993) defines the equitable payoff using the midpoint along the Pareto frontier. This distinction is irrelevant in the conflict games as unilateral deviations never lead to Pareto improvement or deterioration.

that i shows j, denoted by f_{ij}, can now be defined in relation to the payoff in Π_{ij} that i chooses for j, taking π^e_{ij} as the reference point and normalizing to the interval $[-1, 1]$. That is,

$$f_{ij} = \frac{\pi_j(s_{i\cdot}) - \pi^e_{ij}}{\pi^h_{ij} - \pi^l_{ij}}.$$

Similarly, the function $\tilde{f}_{ji}$, denoting i's belief about j's kindness to her, is given by

$$\tilde{f}_{ji} = \frac{\pi_i(s_{ij\cdot}) - \pi^e_{iji}}{\pi^h_{iji} - \pi^l_{iji}},$$

where π^e_{iji}, π^e_{iji} and π^e_{iji} are defined with respect to $\Pi_{iji} \equiv \{\pi_i(s_{ij\cdot})|s_{ij} \in S_j\}$. The psychological utility that i receives from her reciprocal interaction with j is defined as in Rabin (1993) to be

$$u_{ij}(s_{i\cdot}, s_{i\cdot\cdot}) = \tilde{f}_{ji}(s_{ij\cdot}) \cdot [1 + f_{ij}(s_{i\cdot})].$$

The utility function is simply the sum of the monetary payoff and the psychological utility obtained from all pairwise interactions. The psychological term in the utility function is weighted by an individual parameter α_i to capture the magnitude and idiosyncracy of reciprocal preferences. Formally, the utility function $U_i(s_{i\cdot}, s_{i\cdot\cdot})$ takes the form

$$U_i(s_{i\cdot}, s_{i\cdot\cdot}) = \pi_i(s_{i\cdot}) + \alpha_i \sum_{j \in N} u_{ij}.$$

This model can be readily applied to the conflict games. Because the choice is binary, for any i and j who are in the same group,

$$\pi_j(s_{i\cdot}) = \begin{cases} \pi^h_{iji} & \text{if } i \text{ contributes.} \\ \pi^l_{iji} & \text{if } i \text{ does not contribute.} \end{cases}$$

Substituting in (B-1), we see that the kindness that i shows to her fellow group member j when she is contributing is 0.5. Similarly, the kindness is equal to 0.5 if i does not contribute. The opposite holds for the kindness towards out-group members in the IPD (and in the Attacker group). In the PD (and in the Victim group), the kindness shown to out-group members is always zero. The utility function in our games is thus

$$U_i(s_{i\cdot}, s_{i\cdot\cdot}) = 140 - 50s_i + 30 \sum_{j \in I} s_{ij} - 30g \sum_{k \in O} s_{ik} + \alpha_i \sum_{l \neq i} u_{il}(s_{i\cdot}, s_{i\cdot\cdot}),$$

where s_i indicates the probability assigned to $c = 1$, I and O indicate i's in-group and out-group, respectively, and g is an indicator for the game, taking the value 1 for the IPD and 0 otherwise. In

psychological game theory, all beliefs are assumed to be correct in equilibrium. Thus, we can expand the psychological utility term to obtain

$$U_i(s) = 140 - 50s_i + 30\sum_{j\in I} s_j - 30g\sum_{k\in O} s_k + \alpha_i\left[(s_i+0.5)\sum_{l\in I\setminus i}(s_l-0.5) + (1.5-s_i)g\sum_{m\in O}(0.5-s_m)\right]. \tag{B-1}$$

4.2 Homogeneous population

We now apply Equation (B-1) to a homogeneous population in which $\alpha_i = \alpha$ for all i.

Proposition 1. *If full cooperation is an equilibrium in the PD, it is also an equilibrium in the IPD. There are (population-wide) reciprocal preferences that support full cooperation in equilibrium in the IPD but not in the PD.*

Proof. To determine the minimal α under which full cooperation can be sustained in equilibrium, fix $s_j = 1, j \neq i$ to obtain

$$U_i(s) = \begin{cases} 180 + 1.5\alpha - g(90 + 0.75\alpha) & \text{if } s_i = 1. \\ 200 + 0.5\alpha - g(90 + 2.25\alpha) & \text{if } s_i = 0. \end{cases}$$

The benefit from contributing is therefore

$$\begin{cases} \alpha - 20 & \text{if } g = 0. \\ 2.5\alpha - 20 & \text{if } g = 1. \end{cases}$$

We find that full cooperation can be sustained in equilibrium in the IPD iff $\alpha \geq 8$. In the other games considered, full cooperation can only be sustained for levels of α greater than 20. Hence, bilateral conflict is predicted to lead to (weakly) more cooperation. □

4.3 Heterogeneous population

We now allow α_i to vary between individuals. Let α_i be distributed iid on some interval $[0, \bar{\alpha})$ with a cumulative distribution function $F(\alpha)$, which is continuous and with positive density everywhere in its

support. We restrict the analysis to symmetric equilibria in which all agents contribute iff $\alpha_i \geq \alpha^*$ for some α^* and all $i \in N$.

Proposition 2. *The maximal proportion of contributors in any symmetric equilibrium is weakly greater in the IPD than in the PD.*

Proof. Write $p = 1 - F(\alpha^*)$ for the proportion of agents in the population who contribute in equilibrium. Applying to Equation (B-1), we obtain for the PD that

$$U_i(s) = \begin{cases} 120 - 1.5\alpha_i(1-p)^2 + 30p(1-p) + (60 + 1.5\alpha_i)p^2 & \text{if } s_i = 1. \\ 140 - 0.5\alpha_i(1-p)^2 + 30p(1-p) + (60 + 0.5\alpha_i)p^2 & \text{if } s_i = 0. \end{cases}$$

The benefit from contributing is thus

$$\pi_i(s_{ij}|s_i = 1) - \pi_i(s_{ij}|s_i = 0) = \alpha_i(2p - 1) - 20. \tag{B-2}$$

Hence a symmetric equilibrium exists if there exists α^* such that

$$F(\alpha^*) = 0.5 - \frac{10}{\alpha^*}. \tag{B-3}$$

Note that if the probability of contribution is positive, it must be larger than 0.5. For example, if α_i is uniformly distributed on the interval [0,100], no player contributes in the unique symmetric equilibrium. If α_i is uniformly distributed on the interval [0,200], there exists a symmetric equilibrium in which each player i contributes iff $\alpha_i \geq 27.64$ with around 86% of the players contributing.

Utility in the IPD is given by

$$U_i(s) = \begin{cases} \begin{aligned} &120 - 1.5\alpha_i(1-p)^2 + 30p(1-p) + (60 + 1.5\alpha_i)p^2 \\ &\quad + 0.75\alpha_i(1-p)^3 - (30 - 0.25\alpha_i)p(1-p)^2 \\ &\quad - (60 + 0.25\alpha_i)p^2(1-p) - (90 + 0.75\alpha_i)p^3 \end{aligned} & \text{if } s_i = 1. \\ \begin{aligned} &140 - 0.5\alpha_i(1-p)^2 + 30p(1-p) + (60 + 0.5\alpha_i)p^2 \\ &\quad + 2.25\alpha_i(1-p)^3 - (30 - 0.75\alpha_i)p(1-p)^2 \\ &\quad - (60 + 0.75\alpha_i)p^2(1-p) - (90 + 2.25\alpha_i)p^3 \end{aligned} & \text{if } s_i = 1. \end{cases}$$

The benefit from contributing is thus

$$\pi_i(s_{ij}|s_i = 1) - \pi_i(s_{ij}|s_i = 0) = \alpha_i(2p^3 - 3p^2 + 6p - 2.5) - 20. \tag{B-4}$$

Substituting for p, we find that a symmetric equilibrium exists if there exists α^* such that

$$-2\alpha^* F(\alpha^*)^3 + 3\alpha^* F(\alpha^*)^2 - 6\alpha^* F(\alpha^*) + 2.5\alpha^* - 20 = 0. \tag{B-5}$$

Returning to the examples above, if α_i is uniformly distributed on the interval [0,100] there exists a symmetric equilibrium in which 89.5% of the players contribute. If α_i is uniformly distributed on the interval [0,200], there exists a symmetric equilibrium with around 95.5% of the players contributing, nearly 10% more than in the PD.

Comparing (B-2) and (B-4), we see that, if the benefit from contributing is positive, it is necessarily larger in the IPD than in the PD.[18] Moreover, since in both games the benefit from contributing is increasing in p and decreasing in α_i, it follows from the existence of an equilibrium cutoff point α^* in the PD that an equilibrium cutoff point $\alpha^{**} < \alpha^*$ exists in the IPD. □

[18] The difference in benefit is given by $2p^3 - 3p^2 + 4x - 1.5$, which is strictly increasing and equals to zero at $p = 0.5$.

References

Abbink, K., J. Brandts, B. Herrmann, and H. Orzen (2010). Intergroup conflict and intra-group punishment in an experimental contest game. *American Economic Review 100*(1), 420–447.

Axelrod, R. M. (1984). The evolution of cooperation.

Baron, J. (2001). Confusion of group interest and self-interest in parochial cooperation on behalf of a group. *Journal of Conflict Resolution 45*(3), 283–296.

Bartlett, M. and D. DeSteno (2006). Gratitude and prosocial behavior helping when it costs you. *Psychological Science 17*(4), 319–325.

Bauer, M., A. Cassar, J. Chytilová, and J. Henrich (2013). War's enduring effects on the development of egalitarian motivations and ingroup biases. *Psychological Science*, forthcoming.

Blattman, C. and E. Miguel (2010). Civil war. *Journal of Economic Literature 48*(1), 3–57.

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.

Bornstein, G. (2003). Intergroup conflict: Individual, group, and collective interests. *Personality and Social Psychology Review 7*(2), 129–145.

Bornstein, G. and M. Ben-Yossef (1994). Cooperation in intergroup and single-group social dilemmas. *Journal of Experimental Social Psychology 30*(1), 52–67.

Bornstein, G., D. Mingelgrin, and C. Rutte (1996). The effects of within-group communication on group decision and individual choice in the assurance and chicken team games. *Journal of Conflict Resolution 40*(3), 486–501.

Bowles, S. (2008). Conflict: Altruism's midwife. *Nature 456*(7220), 326–327.

Boyd, R. and P. J. Richerson (1989). The evolution of indirect reciprocity. *Social Networks 11*(3), 213–236.

Brams, S. J. (1975). *Game theory and politics*. New York: The Free Press.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues 55*, 429–444.

Campbell, D. T. (1965). Ethnocentric and other altruistic motives. In *Nebraska symposium on motivation*, Volume 13, pp. 283–311.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *Quarterly journal of Economics 117*(3), 817–869.

Choi, J.-K. and S. Bowles (2007). The coevolution of parochial altruism and war. *Science 318*(5850), 636–640.

Coser, L. A. (1956). *The Function of Social Conflict*. Glenco, IL: Free Press.

De Dreu, C. K. (2012). Oxytocin modulates cooperation within and competition between groups: an integrative review and research agenda. *Hormones and behavior 61*(3), 419–428.

De Dreu, C. K., L. L. Greer, G. A. Van Kleef, S. Shalvi, and M. J. Handgraaf (2011). Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences 108*(4), 1262–1266.

De Dreu, C. K. W., L. L. Greer, M. J. J. Handgraaf, S. Shalvi, G. A. Van Kleef, M. Baas, F. S. Ten Velden, E. Van Dijk, and S. W. W. Feith (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science 328*(5984), 1408–1411.

Dufwenberg, M., U. Gneezy, W. Güth, and E. E. C. Damme (2001). Direct versus indirect reciprocity: An experiment. *Homo Oeconomicus 18*(1/2), 19–30.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior 47*(2), 268–298.

Engelmann, D. and M. Strobel (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review 94*(4), 857–869.

Fehr, E. and S. Gächter (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives 14*(3), 159–181.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fischbacher, U. and S. Gächter (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review 100*(1), 541–556.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological games and sequential rationality. *Games and Economic Behavior 1*, 60–79.

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

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63*, pp. 79–93. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Greiner, B. and M. Levati (2005). Indirect reciprocity in cyclical networks: An experimental study. *Journal of Economic Psychology 26*(5), 711–731.

Güth, W., M. Königstein, N. Marchand, and K. Nehring (2001). Trust and reciprocity in the investment game with indirect reward. *Homo Oeconomicus 18*, 241–262.

Halevy, N., G. Bornstein, and L. Sagiv (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.

Halevy, N., O. Weisel, and G. Bornstein (2012). "in-group love" and "out-group hate" in repeated interaction between groups. *Journal of Behavioral Decision Making 25*(2), 188–195.

Hardin, R. (1997). *One for all: The logic of group conflict*. Princeton Univ Pr.

Kramer, R. M. and M. Brewer (1984). Effects of group identity on resource use in a simulated commons dilemma. *Journal of Personality and Social Psychology 46*(5), 1044–1057.

Krawczyk, M. (2011). What brings your subjects to the lab? a field experiment. *Experimental Economics 14*(4), 482–489.

Kritikos, A. and F. Bolle (2001). Distributional concerns: equity-or efficiency-oriented? *Economics Letters 73*(3), 333–338.

Lichbach, M. I. (1996). *The cooperator's dilemma*. Ann Arbor: University of Michigan Press.

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

Nowak, M. A. and S. Roch (2007). Upstream reciprocity and the evolution of gratitude. *Proceedings of the Royal Society B: Biological Sciences 274*(1610), 605–610.

Olson, M. (1974). *The logic of collective action: Public goods and the theory of groups*. Harvard University Press.

Ostrom, E. (2000). Collective action and the evolution of social norms. *The Journal of Economic Perspectives 14*(3), 137–158.

Palfrey, T. R. and H. Rosenthal (1983). A strategic calculus of voting. *Public Choice 41*(1), 7–53.

Penner, L., M. T. Brannick, S. Webb, and P. Connell (2005). Effects on volunteering of the september 11, 2001, attacks: An archival analysis. *Journal of Applied Social Psychology 35*(7), 1333–1360.

Pfeiffer, T., C. Rutte, T. Killingback, M. Taborsky, and S. Bonhoeffer (2005). Evolution of cooperation by generalized reciprocity. *Proceedings of the Royal Society B: Biological Sciences 272*(1568), 1115–1120.

Probst, Tahira, M., P. J. Carnevale, and H. C. Triandis (1999). Cultural values in intergroup and single-group social dilemmas. *Organizational Behavior and Human Decision Processes 77*, 171–191.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83*(5), 1281–1302.

Rapoport, A. and G. Bornstein (1987). Intergroup competition for the provision of binary public goods. *Psychological Review 94*(3), 291–299.

Sherif, M. (1961). *Intergroup conflict and cooperation: The Robbers Cave experiment*. University Book Exchange Norman.

Singelis, T. M., H. C. Triandis, D. P. S. Bhawuk, and M. J. Gelfand (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research 29*(3), 240–275.

Stein, A. A. (1976). Conflict and cohesion. *Journal of Conflict Resolution 20*(1), 143–172.

Sugden, R. (1984). Reciprocity: the supply of public goods through voluntary contributions. *The Economic Journal 94*(376), 772–787.

Tajfel, H. and J. C. Turner (1979). An integrative theory of intergroup conflict. In W. G. Austin and S. Worchel (Eds.), *The Social Psychology of Intergroup Relations*, Chapter 3, pp. 33–47. Monterey, CA: Brookes/Coole.

Wit, A. P. and H. A. M. Wilke (1992). The effect of social categorization on cooperation in three types of social dilemmas. *Journal of Economic Psychology 13*(1), 135–151.