Conflict, sticks and carrots: war increases prosocial punishments and rewards

Ayelet Gneezy\(^1,*,†,‡\) and Daniel M. T. Fessler\(^2,†\)

\(^1\)Rady School of Management, University of California, San Diego, CA 92093, USA
\(^2\)Department of Anthropology and Center for Behavior, Evolution and Culture, University of California, Los Angeles, CA 90095, USA

Unlike most species, humans cooperate extensively with group members who are not closely related to them, a pattern sustained in part by punishing non-cooperators and rewarding cooperators. Because internally cooperative groups prevail over less cooperative rival groups, it is thought that violent intergroup conflict played a key role in the evolution of human cooperation. Consequently, it is plausible that propensities to punish and reward will be elevated during intergroup conflict. Using experiments conducted before, during and after the 2006 Israel–Hezbollah war, we show that, during wartime, people are more willing to pay costs to punish non-cooperative group members and reward cooperative group members. Rather than simply increasing within-group solidarity, violent intergroup conflict thus elicits behaviours that, writ large, enhance cooperation within the group, thereby making victory more likely.

**Keywords:** evolution of cooperation; punishment; reward; intergroup conflict

1. INTRODUCTION

In non-human species, cooperation is largely explicable in terms of common genetic interests, mutualism or, more debatably, reciprocity. Humans, however, routinely cooperate with individuals to whom they are not closely related; they often incur costs in cooperating; and they often cooperate in non-repeated or large-scale interactions that preclude reciprocity. Ironically, our species’ unique cooperativeness may derive from a history of violent conflict between groups [1–6]. If intergroup conflict played an important role in the evolution of human cooperation, then it is possible that such conflict elicits conditional changes in individuals’ behaviour that enhance cooperation within the group.

Human groups characterized by comparatively greater levels of cooperation grow faster and endure longer than less cooperative groups, replacing them through conquest, recruitment or the diffusion of beliefs and practices. This favours the cultural evolution of norms promoting cooperation within the group [4,6,7]. Furthermore, models reveal that, by increasing the costs of defection, punishment can stabilize cooperation within a group [8]. Analogous considerations suggest that, by enhancing the payoffs of cooperators, rewards may also bolster cooperation [9]. In experimental economic games in which aggregate welfare is increased through cooperation, yet individual welfare is maximized through defection, the presence of defectors degrades cooperation. However, when game parameters allow participants to punish defectors, cooperation is stabilized [10–13], and when intergroup competition is added to the mix, groups with punishment outperform groups without [14]. Likewise, when parameters allow participants to reward cooperators, cooperation is often enhanced [13,15,16]. In conjunction with cultural evolution, the positive effects of punishment and reward on cooperation explain the ubiquity of norms prescribing such sanctions [4].

Under many circumstances, punishing non-cooperators and rewarding cooperators are costly acts, raising the question of how the propensities to engage in these behaviours evolved, as such public goods invite higher-order free-riding (benefiting from cooperation while not paying its maintenance costs). Debate surrounding this question has produced three classes of solutions. One approach holds that intergroup competition led to the evolution of both group-functional norms and, via genetic group selection (mathematically equivalent to kin selection [17]), the propensity to cooperate with in-group members and punish non-cooperators [2,3,18]; by extension, this position could also conceivably capture the propensity to reward cooperators. A second approach holds that intergroup competition led to the cultural evolution of group-functional norms. This cultural environment then favoured the genetic evolution of the capacity and propensity to internalize local norms, resulting in fewer punishments and greater returns [4,6,7,19]. Building on the insight that costly signals of cooperativeness can benefit the group and the signaler [20], a third approach holds that genetic individual-level selection favoured psychological mechanisms that strategically regulate cooperation and/or norm adherence, facultatively adjusting behaviour by altering motivations largely independent of conscious calculations. In this account, actors possess adaptations that evaluate future behaviour in light of its payoffs for them as individuals, as partly indexed by cues of reputational consequences [21–25]. Within this framework, intergroup conflict can be seen as having favoured the evolution of cooperation, either by changing individual payoffs directly or by doing so in conjunction with the cultural evolution of group-functional norms that regulate cooperation.

*Author for correspondence (agneezy@ucsd.edu).
†All authors contributed equally to this work.
Although none of the above approaches addresses the direct effects of intergroup conflict on individual behaviour, each can generate the prediction that such conflict should enhance individuals’ willingness to punish non-cooperative group members and reward cooperative members. The genetic group selection approach entails the corollary that individuals should increase cooperation-promoting behaviours when the group is threatened. The cultural evolution and internalization approach entails the corollary that cultures should evolve meta-norms that prioritize such behaviours during intergroup conflicts. The strategic reputation approach entails several possibilities. First, because intergroup conflict increases the need for allies, individuals should engage in more behaviours that promote cooperation, including rewarding cooperators and punishing non-cooperators. In doing so, individuals signal to others that recruiting or retaining them will enhance the welfare of a coalition. Relatedly, intergroup conflict should cause coalitions to place increased value on toughness (i.e. the willingness to incur costs in order to inflict costs on antagonists). Punishing non-cooperators can signal toughness, and can therefore enhance recruitment or retention in a coalition. Finally, the above considerations should be amplified if norms prescribe punishing non-cooperators and rewarding cooperators, as conforming to such prescriptions advertises both predictability and prosociality—a valuable combination in a coalition member.

Observers have long noted that intergroup conflict appears to increase in-group solidarity [26], punishment of deviants, and valorization of self-sacrificial heroes [18]. Synthetic contexts of intergroup conflict, consisting of a temporary conflict of pecuniary interests between arbitrarily constructed groups, have been used to demonstrate that intergroup conflict increases first-order prosocial behaviour towards fellow group members [27,28]. However, it is difficult to interpret the relevance of such results for the predictions at issue here, as minor pecuniary conflicts differ substantially from real intergroup violence. More compellingly, archival analysis indicates that the 9/11 terrorist attacks in the USA were followed by increased charitable volunteering [29], and a recent investigation in Burundi, notable for its use of experimental methods in a real-world context [30], reveals that exposure to interethnic violence correlates with increased first-order generosity in economic games [31].

While the above results are important, because cooperation breaks down in the absence of sanctions, studies of first-order behaviour do not examine the factors required for success during sustained intergroup conflicts. The development of experimental economic games played for real stakes provides powerful tools for measuring individuals’ propensity to engage in such costly prosocial behaviour. Recent laboratory investigations document increased punishment of uncooperative in-group members during intergroup conflict [14,32]. Though revealing, these studies are limited by their reliance on a conflict of pecuniary interests between arbitrary groups, and by the absence of opportunities to reward as well as punish. In sum, although existing findings are consonant with the prediction that intergroup conflict enhances cooperation via increased sanctioning, no investigation to date has examined the effect of real-world violent conflict on group members’ tendencies to punish non-cooperators and reward cooperators. We therefore conducted experimental economic games before, during and after the 2006 Israel–Hezbollah war.

2. METHODS

The Israel–Hezbollah war, in which thousands were killed or injured, broke out suddenly following a rapid escalation of hostilities. Originally intending to explore other issues, we had run two economic games in Israel before the war began. To measure the impact of the war, we revisited the same population during and after the conflict, running the same economic games each time. All games were played for real stakes. No deception or pretence was employed, and all participants were informed in advance of the complete set of game parameters regardless of the roles assigned to them.

Costly punishments. We use the Ultimatum Game [33] to measure costly punishment. In this game, Player A is endowed with a certain amount of money, and is instructed to propose a division of the endowment between herself and Player B. Player B can either accept or reject this proposal. If Player B accepts, the money is divided as proposed by Player A; if she rejects, both players receive nothing. The game ends following Player B’s decision. Rejection by Player B constitutes costly punishment of Player A’s (selfish) behaviour, as Player B forfeits her respective portion of the endowment.

Costly rewards. We use the Trust Game [34] to measure costly rewarding. Player A is given an endowment, and is instructed to decide what portion of her endowment to transfer to Player B. The amount transferred is tripled in transit. Player B then decides how much of this tripled sum to return to Player A. The game ends following Player B’s decision. Player B thus has the opportunity to reward Player A’s cooperative behaviour by sending money to Player A at her expense. Note that although Player B can also punish uncooperative behaviour by withholding money from Player A, this is less revealing of willingness to punish than are responses in the Ultimatum Game because, rather than being costly, punishment in the Trust Game benefits Player B.

Although the current paper focuses on the behaviours of Player Bs in the Ultimatum and Trust Games, for purposes of completeness, we include a discussion of the behaviours of Player As in the electronic supplementary material, along with complete data for all participants.

We ran two experiment sessions nine months before the war began, two more sessions during the second week of the 34-day war, and one additional session 1 year after the war ended. In all sessions, participants were competent senior citizens residing in the same housing facility in Tel Aviv. This sample was ideal as, unlike most young people, our participants were not called into military service during the war. Each participant played both the Trust Game and the Ultimatum Game, in that order, changing partners between games. Participants were randomly assigned to perform either the Player A role or the Player B role; once assigned, they played that role for both games. We informed participants that they would play each game with a different (anonymous) partner. Neutral language was used in the instructions; for example, the first mover’s decision option was described as ‘propose’ in the Ultimatum Game and ‘transfer’ in the Trust Game, etc. Each individual participated in only one session; all comparisons are thus between aggregates of participants at different periods.

The experiment was administered using pen and paper. Participants were paired at random and interacted anonymously, with 10 pairs (i.e. 20 individuals; 10 per role) in each session.
Participants received a NIS 50 show-up fee in addition to any earnings from the games; at the time of the experiment the exchange rate was approximately US $1 = NIS 4.5. The endowment in the Ultimatum Game was NIS 100; proposals were in NIS 10 increments. We used the ‘strategy method’: for each of the possible divisions of the endowment, Player B is asked to decide in advance whether she wishes to accept or reject the given proposal. Each Player B participant is then matched with a randomly chosen Player A participant, and Player A’s proposal is either accepted or rejected as per Player B’s prior instructions [35]. The endowment in the Trust Game was NIS 50; transfers were in NIS 5 increments. We similarly collected Player B’s return decisions prospectively for all possible transfers.

3. RESULTS

In both games, participants’ decisions did not differ between the pre- and post-war periods (t-test, p > 0.1), hence these were combined for our analyses, producing two categories—wartime and peacetime. The results reported below consist of 20 independent observations (i.e. 20 Player As and 20 Player Bs) for the wartime group and 30 independent observations (i.e. 30 Player As and 30 Player Bs) for the peacetime group.

**Costly punishments.** As illustrated in figure 1, when playing the Ultimatum Game, wartime Player B participants were more likely to reject proposals of less than 40 per cent of the endowment. These rejection rates were significantly different (Mann–Whitney U-test, p < 0.01) from the rejection rates observed among peacetime Player Bs, indicating that overall rejection rates during wartime are higher than rejection rates during peacetime; inspection of figure 1 reveals that this difference is driven by responses to offers below 40 per cent of the endowment. In other words, wartime Player Bs were more willing to incur a cost in order to punish unfair proposals.

**Costly rewards.** In the Trust Game, when the amount transferred by Player A was substantial (approx. three-quarters of the endowment or more), wartime Player Bs returned significantly more money to Player As than did...
their peacetime counterparts (figure 2). Regressing Player B’s standardized returns on the amount sent by Player A reveals an overarching pattern wherein wartime Player Bs rewarded high levels of cooperation, and punished low levels of cooperation, to a greater extent than did peacetime Player Bs ($\beta = 0.024$, $\beta = 0.05$, respectively; $\chi^2(1) = 22.43$, $p < 0.05$). In other words, wartime Player Bs were more willing to incur a cost in order to reward cooperative behaviour; conversely, when the amount transferred was below the required threshold, wartime Player Bs returned less money to Player As than did peacetime Player Bs, revealing an increased propensity to punish uncooperative behaviour.

4. DISCUSSION

Intergroup conflict plays a key role in three prominent theories of the evolution of human cooperation. Via differing pathways, each of these theories generates the prediction that violent intergroup conflict should lead individuals to increase their willingness to inflict costs on non-cooperators and bestow benefits on cooperators. Our investigation of the effects of the 2006 Israel–Hezbollah war on Israeli civilians reveals precisely this pattern. Results from prior models and experiments indicate that, iterated across interactions and contexts, the changes that we documented would serve to increase levels of cooperation within the group during periods of violent intergroup conflict. The findings presented in this paper are thus consonant with the contention that intergroup conflict was an important factor in the evolution of human cooperation.

Ours is the first study to explore the effects of real violent intergroup conflict on actual punishments and rewards in controlled cooperative dilemmas. An important question left unanswered by our findings, however, is the specificity of the effect of conflict on sanctioning behaviours. On the one hand, because punishing non-cooperation and rewarding cooperation each have a direct impact on both subsequent levels of cooperation and individual strategic positioning, we might expect that intergroup conflict primarily affects propensities to react specifically to uncooperative or cooperative behaviour. However, because successful cooperation depends on coordination, and coordination requires that group members conform to many norms, we could also expect that, via the same pathways, intergroup conflict would affect propensities to punish a wide range of non-conformist behaviours and reward a wide range of conformist behaviours. Our study examined second movers’ responses to uncooperative and cooperative behaviour. While such methods usefully shed light on punishment and reward within the circumscribed context of cooperation [13], they are less diagnostic with regard to generalized norm enforcement. Because many norm violations do not involve a second mover upon whom the first mover’s counter-normative behaviour inflicts costs, investigations of the effects of violent intergroup conflict on individuals’ propensity to punish non-conformity and reward conformity will need to employ experimental games that allow disinterested third-party observers to sanction first movers’ behaviour at a cost to themselves [36,37].

Consonant with the possibility that intergroup conflict elevates the enforcement of conformity, history provides troubling indications that war may imperil cultural plurality and democracy as much from within as from without. With this in mind, we conclude by calling for further investigations of the effects of violent intergroup conflict on the propensities to punish and reward members of one’s own group.

We thank Uri Gneezy, Rob Boyd, Karthik Panchanathan, Deborah Small, Joe Henrich, Moshe Hoffmann and two anonymous reviewers for comments. D.F. received support from the U.S. Air Force Office of Scientific Research under award no. FA9550-10-1-0511.

REFERENCES


