PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Costly punishment prevails in intergroup conflict

Lauri Sääksvuori, Tapio Mappes and Mikael Puurtinen

Proc. R. Soc. B 2011 **278**, 3428-3436 first published online 30 March 2011
doi: 10.1098/rspb.2011.0252

PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2011) **278**, 3428–3436
doi:10.1098/rspb.2011.0252
Published online 30 March 2011

Costly punishment prevails in intergroup conflict

Lauri Sääksvuori[1,2,*], Tapio Mappes[3] and Mikael Puurtinen[3]

[1]*Max Planck Institute of Economics, IMPRS, Kahlaische Strasse 10, 07745 Jena, Germany*
[2]*Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington, IN, USA*
[3]*Centre of Excellence in Evolutionary Research, The Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35, Jyväskylä 40014, Finland*

Understanding how societies resolve conflicts between individual and common interests remains one of the most fundamental issues across disciplines. The observation that humans readily incur costs to sanction uncooperative individuals without tangible individual benefits has attracted considerable attention as a proximate cause as to why cooperative behaviours might evolve. However, the proliferation of individually costly punishment has been difficult to explain. Several studies over the last decade employing experimental designs with isolated groups have found clear evidence that the costs of punishment often nullify the benefits of increased cooperation, rendering the strong human tendency to punish a thorny evolutionary puzzle. Here, we show that group competition enhances the effectiveness of punishment so that when groups are in direct competition, individuals belonging to a group with punishment opportunity prevail over individuals in a group without this opportunity. In addition to competitive superiority in between-group competition, punishment reduces within-group variation in success, creating circumstances that are highly favourable for the evolution of accompanying group-functional behaviours. We find that the individual willingness to engage in costly punishment increases with tightening competitive pressure between groups. Our results suggest the importance of intergroup conflict behind the emergence of costly punishment and human cooperation.

Keywords: cooperation; group conflict; public goods; punishment

1. INTRODUCTION

The ability of humans to uphold cooperative relationships among large numbers of unrelated partners is an evolutionary puzzle. Among numerous proposed solutions to the problem of cooperation [1], punishment of uncooperative individuals has attracted considerable attention as a proximate reason why cooperative behaviours might proliferate [2–5]. While abundant experimental evidence [6,7] and direct neurobiological measurements [8] indicate that human readiness to incur costs to sanction uncooperative individuals is motivated by emotional mechanism, the evolution of individually costly punishment has been difficult to explain. Theoretical research [9–11] suggests that the evolutionary origin of group-beneficial behavioural traits and traditions is embedded in intergroup conflict. Consequently, costly behaviours that increase cooperation in groups may proliferate at the expense of less cooperative groups and individuals through extinction and emulation. This process is often seen as a consequence of military, economic and other forms of intergroup rivalries.

Warlike activity recorded among prehistoric humans [12,13] and quasi-experimental preference measurements in modern conflict areas [14] suggests that intergroup conflict has shaped human behaviour during the evolutionary history of the species. At the same time, anthropological annals [15,16], present-day field observations [17,18] and behavioural experiments among diverse human populations [19,20] all portray a picture of considerable variation in social organization and livelihood between human communities that explicitly differ in their willingness to sanction norm-violating behaviour. While a variety of different punishment mechanisms has been identified, sanctioning in many primitive societies without a judicial system [21] as well as in groups managing common-pool resources [22] is not coordinated by a central authority, but by individual group members or informal coalitions.

Since punishment is costly both for those who punish and typically even more so for those who are punished, punishment is expected to evolve only when the benefits of increased cooperation outweigh its costs. Several recent papers employing experimental set-ups in isolated groups have found clear evidence that while costly punishment increases the level of cooperation, the net effect in terms of material pay-off is often negative, decreasing the success of groups and individuals [23–26]. These results have cast doubt on the idea that costly punishment could evolve as a group-beneficial trait. Concurrently, arguments have been presented that the reduction in group and individual success owing to costs of punishment is likely to be overcome under longer time horizons [27] and coordinated punishment activity [28]. The disparity in the conclusions from experiments with isolated groups underscores the lack of direct evidence on the selective merit of costly punishment in intergroup interactions.

In light of the consistent and strong empirical evidence, there is no question whether humans readily

* Author for correspondence (saaksvuori@econ.mpg.de).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2011.0252 or via http://rspb.royalsocietypublishing.org.

Received 4 February 2011
Accepted 10 March 2011

Table 1. Summary of experimental treatments. Table summarizes the key configurations of our treatments. Hyphen in the treatment title refers to a treatment with group conflict.

treatment	punishment	group configuration	group size	no of subjects
NOPUN	no	no competition	8	48
PUN	yes	no competition	8	48
PUN–NOPUN	partly	asymmetric competition	8 + 8	96
PUN–PUN	yes	symmetric competition	8 + 8	48
NOPUN–NOPUN	no	symmetric competition	8 + 8	48

sacrifice individual resources to discipline uncooperative individuals without immediate tangible benefits. The unresolved question, however, is why humans incur costs to discipline their fellows. In this study, we explicitly address the possibility that costly punishment could evolve as a group-functional trait in intergroup conflict. For the purpose, we conducted a series of public goods experiments where we systematically varied the competitive environment between groups and individual opportunities to punish fellow group members. Our experimental design excludes the effects of direct reciprocation, reputation scores, communication and other conceivable proximate mechanism potentially supporting human cooperation in an effort to consistently focus on the importance of intergroup conflict in determining the selective benefit of costly punishment.

2. MATERIAL AND METHODS

(a) *Experimental design*

To study the effectiveness of costly punishment in direct intergroup conflict, we conducted a series of public goods experiments with and without group competition. Most essentially, we varied the possibility of individuals to engage in costly punishment towards fellow group members. Altogether five different treatments were conducted (table 1).

In two control treatments, there was no group competition. In the no-punishment (NOPUN) treatment, participants played the linear public goods game without punishment opportunity, whereas in the punishment (PUN) treatment an equivalent public goods game was played with an opportunity to punish. These control treatments allow comparison with studies that have explored the effects of costly punishment in public goods games without incorporating between-group competition [5,23,24,27]. More importantly, by comparing the behaviour in group competition treatments with control treatments, we test the importance of group competition for the effectiveness of costly punishment.

In the asymmetric group competition treatment (PUN–NOPUN), participants of one group had the possibility to engage in costly punishment, while members of the rival group did not have this opportunity. Consequently, the comparison of net earnings of individuals in groups with contrasting punishment possibilities gives direct evidence on the benefits of costly punishment in intergroup competition, and shows whether costly punishment could evolve by influencing the success of individuals through direct intergroup competition.

In our symmetric group competition treatments (PUN–PUN and NOPUN–NOPUN), all members of the competing groups either had or did not have the opportunity to punish their fellow group members, respectively. The symmetric competition treatment without punishment (NOPUN–NOPUN) establishes an important benchmark to study if intergroup conflict is sufficient by itself to stabilize cooperation within strategically interdependent groups. The treatment with symmetric punishment opportunities (PUN–PUN) reveals the effects of punishment in a population where all competing groups have equal punishment opportunities.

In all treatments, eight participants played the game within their own group. The composition of groups stayed intact throughout the game. Participants' identities within a group were shuffled between periods. The game lasted for 30 identical periods. In the beginning of each period, participants received 20 monetary units (MUs) and simultaneously allocated these between group and personal accounts. The total amount allocated to group account was doubled by the experimenter and divided equally among all group members. In treatments with punishment, participants could then assign deduction points to their own group members after each period. Punishment was costly. Each deduction point cost the punisher 1 MU and reduced the earnings of the receiver by 3 MUs. We apply the 3 : 1 punishment ratio as well as the identity shuffling to facilitate the compatibility of our results with the pertinent literature [5,24,27].

In all treatments with group competition, the performances of competing groups were compared in each period after the public goods game was played within groups. The group with more MUs invested into group account won twice the difference in total investments. The group with lower investments lost an equivalent amount of MUs. Wins and losses from group competition were divided equally among the group members. Thus, the pay-off consequences of conflict were endogenously determined by the performance of the groups [29]. This model of group competition introduces two important improvements to other existing experimental models of group competition [30,31]. First, as group competition does not involve an external prize, earnings can readily be compared between treatments with and without group competition. Second, the effect of group competition depends linearly on the difference between group performances. In other words, the more unequal the performances of the competing groups are, the more impact group competition has on individual earnings. When the group performances differ only slightly, group competition has only a minor effect on earnings. Finally, when the group performances are tied, group competition has no effect on earnings. In many scenarios relevant to the study of human behaviour, this structure can be seen as a more suitable way of modelling group competition than the probabilistic intergroup conflict used to model 'winner-takes-it-all' situations [32,33].

After the group competition stage, participants were informed about the contributions of their fellow group members and the total contribution made by the competing

group. In groups with punishment opportunity, participants could then assign deduction points to their own group members using the same procedure as in treatments without competition. This means that the winner in the group competition was the group with highest level of cooperation before subtracting the costs of assigned and received punishments from the individual payments. It appears natural to assume that during the course of human evolution the success in group conflicts has been primarily determined by the coordination and cooperation among the group members, whereas net wealth has not been easily transformable to fighting power before the emergence of industrial societies. Likewise, it is unlikely that individuals mete out punishments at the time when informal punishments directly jeopardize individual or group success. This intuition is further supported by our data showing that the actual outcome of the conflict affects the likelihood and severity of assigned punishments (see §3 below). Notably, however, when we make inferences about the selective benefits of costly punishment we always compare net earnings that account for the costs of assigned and received punishments. Overall, the design reflects the importance of cooperation in surviving periodic war and abrupt environmental crises in conditions likely to have been experienced by late Pleistocene and early Holocene humans [34]. For a more detailed discussion pertaining to the group conflict model and punishment, see the electronic supplementary material.

(b) *Experimental procedure*

The experiment was conducted at the laboratory of the Max Planck Institute of Economics in Jena (Germany). In 12 different experimental sessions, a total number of 288 participants took part in the experiment. The vast majority of the 169 female and 119 male participants were undergraduate students studying a range of different disciplines. None of the participants had previous experience with social dilemma experiments. In all treatments, participants were informed about the individual contributions and corresponding earnings in their own group after the contribution stage. In addition, the total amount of contributions in the directly competing group was revealed to participants. Participants were not informed about the individual punishment decisions of other participants. In the asymmetric competition treatment (PUN–NOPUN), where punishing and non-punishing groups were compared, no information about the punishment opportunity was revealed to the group without punishment.

The total earnings of a participant equalled the sum of net pay-offs over all 30 periods in all treatments. One experimental session lasted on average 90 min. Earnings per participant ranged from €9 to €36 with an average of €20. The experiment was programmed and run using z-Tree [35]. A full description of the experimental procedure including sample instructions is available in the electronic supplementary material.

3. RESULTS AND DISCUSSION

Examining the behaviour in isolated PUN and NOPUN groups, we find that the effect of costly punishment on cooperation was, on average, substantial but only marginally significant owing to large variation between groups with punishment opportunity (figure 1*a*, mean contributions in PUN 14.5 MUs and in NOPUN 7.7 MUs; Mann–Whitney $U_{6,6} = 30$, exact $p = 0.065$, two-tailed). Consequentially, punishment did not significantly increase the net material pay-off that accounts for the cost of assigned and received punishment points (figure 1*b*, mean total net pay-offs in PUN 989 MUs and in NOPUN 830 MUs; Mann–Whitney $U_{6,6} = 28$, exact $p = 0.132$, two-tailed). A closer look at the distribution of net pay-offs reveals that the pay-offs vary widely among groups with punishment (figure 2*b*). The inconsistent effect of punishment in the absence of group competition is further illustrated by individual-level data showing that the highest individual pay-off was earned by an individual belonging to a group without punishment (compare figure 2*a*,*b*). These findings are consistent with several previous studies that have not found substantial benefits of punishment in isolated groups [4,23,24,36]. It is noteworthy that the games in our experiments lasted for 30 periods. Thus, the non-significant effect of punishment cannot be explained by short game duration. Our results from isolated groups using a larger group size than typical, at eight participants, suggest that the findings [27] stressing the long-term benefits of punishment perhaps apply to small groups (groups size of three in [27]), but do not readily extrapolate to larger groups.

In the asymmetric group competition treatment (PUN–NOPUN), the possibility of punishment had a dramatic effect on cooperation. In groups with punishment opportunity, contributions to the group account quickly rose and levelled close to maximum investment that significantly exceeds the contributions of non-punishing groups (figure 1*c*, groups with punishment 19.3 MUs, groups without punishment 13.6 MUs, Wilcoxon signed-rank test for six-matched observations, $t = -21$, exact $p = 0.031$, two-tailed). The effect of punishment on net pay-offs was even more pronounced (figure 1*d*, groups with punishment 1485 MUs, groups without punishment 586 MUs, Wilcoxon signed-rank test for six-matched observations $t = -21$, exact $p = 0.031$, two-tailed). Importantly, even the lowest earning individual in groups with punishment opportunity earned more than the highest earning individual in groups without punishment (figure 2*c*). The data unequivocally reveal that in an asymmetric group conflict where groups with and without punishment are in direct competition, punishment opportunity benefits both the group and the individual.

We provide further evidence on the effects of punishment in intergroup conflicts and test if intergroup competition alone suffices to maintain cooperation by examining the behaviour in symmetric conflicts where both of the groups either had or did not have the opportunity to punish. While group competition alone had a weak tendency to increase net pay-offs when compared with a situation without competition (mean total net pay-offs in NOPUN 830 MUs and in NOPUN–NOPUN 998 MUs; Mann–Whitney test assuming independence of observations: $U_{6,6} = 29$, exact $p = 0.093$, two-tailed), it was not sufficient to maintain stable cooperation (figure 1*e*). By contrast, in PUN–PUN, cooperation quickly stabilized near maximum contributions (figure 1*e*). Comparing the symmetric competition treatments reveals that mean contributions were higher in punishing groups than in groups without punishment (PUN–PUN 19.4 MUs and NOPUN–NOPUN 13.3 MUs, Mann–Whitney test assuming

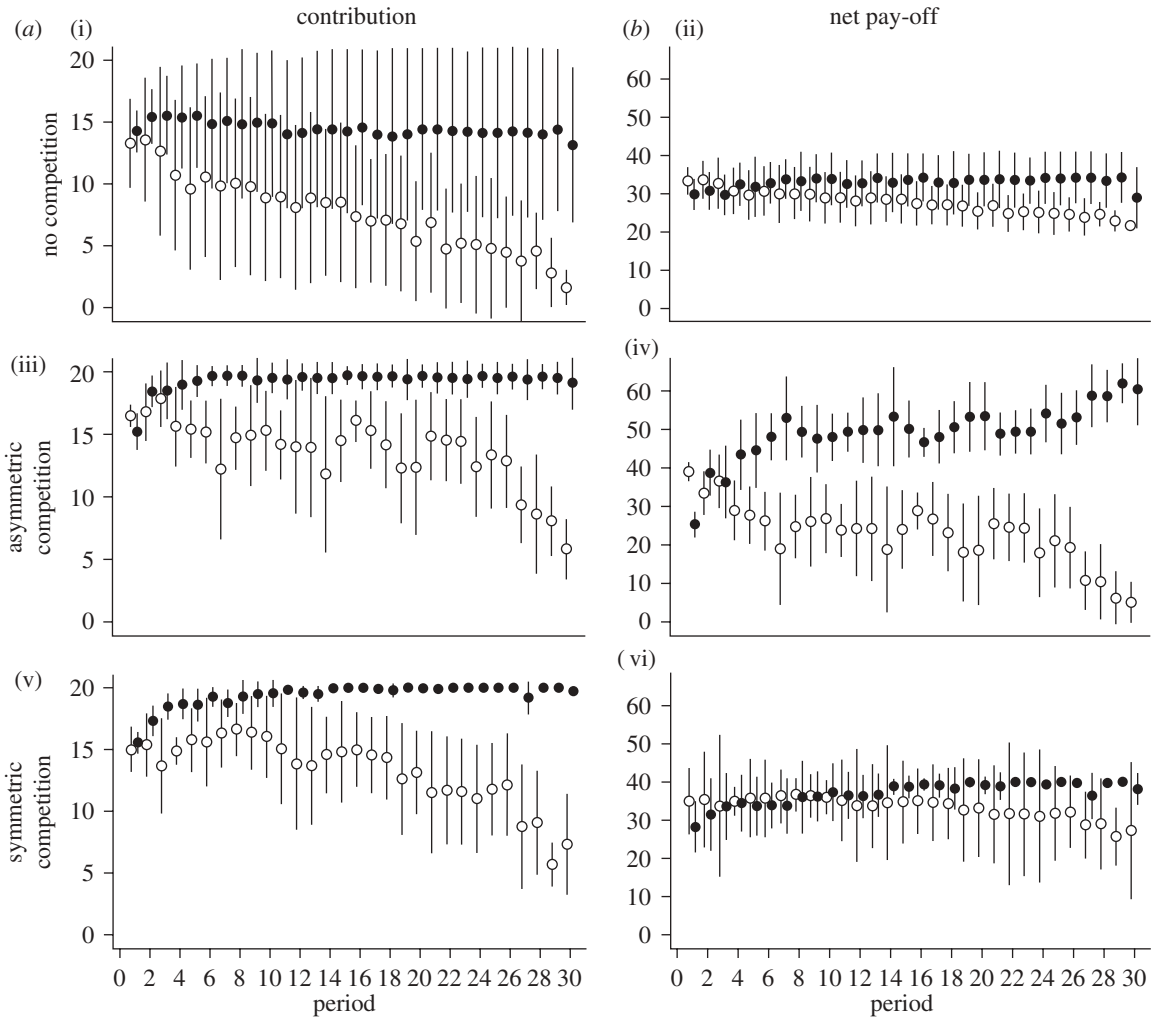

Figure 1. (*a*) Cooperation rates and (*b*) net pay-offs in groups with and without punishment. Solid black dots denote the mean of groups with punishment opportunity, whereas transparent symbols correspond to groups without punishment opportunity. Error bars denote the 95% confidence interval for the mean value. Punishment opportunity led to significantly higher contribution levels in experiments with direct group comparisons (iii,v). This effect was only marginally significant when groups were in isolation (i). (ii,iv,vi) On the right display the mean net pay-offs after accounting for the cost of assigned and received punishment points. The net pay-offs were significantly higher in punishing groups in asymmetric group conflict (iv), marginally higher in punishing groups in symmetric competition treatments (vi), whereas no statistically significant effect was found in the absence of group competition (ii). For statistical test values see §3.

independence of observations: $U_{6,6} = 36$, exact $p = 0.002$, two-tailed). Punishment maintains high level of cooperation even when both parties of the group conflict adopt the culture of peer-punishment. However, the comparison of symmetric competition treatments in terms of net pay-offs reveals that the pay-off superiority of punishing groups becomes effective only with a longer time horizon and remains modest, even though growing, over time (figure 1*f*, mean total net pay-off in PUN–PUN 1113 MUs and in NOPUN–NOPUN 998; Mann–Whitney test assuming independence of observations: $U_{6,6} = 30$, exact $p = 0.065$, two-tailed). The narrow benefit of punishment over non-punishment in symmetric competition treatments may lead one to (erroneously) conclude that the disposition to punish may not need to proliferate in the long run, as tribes composed of several interacting group with a practice to sanction uncooperative individuals do not significantly outperform tribes without punishment. However, rivalrous interactions do not occur only within tribes, but also along the boundaries of their tribal territories. In conflict along the boundary, groups with punishment would prevail and punishment would thus inevitably spread to the entire population. Further, in a population consisting only of punishers, any group or tribe renouncing punishment would perish as evidenced by the major advantage for individuals in punishing groups in the asymmetric competition treatment (figures 1*d* and 2*c*). In sum, punishment appears to be very resistant to invasion in an environment characterized by group conflict.

An intriguing and robust finding is that the within-group variation in individual pay-offs is substantially smaller in groups with punishment than in groups without punishment (see the caption of figure 2). Clearly, the propensity to incur costs in order to sanction not only ensures a higher level of cooperation and net earnings in group conflicts, but also generates substantially greater equality within the group vis-à-vis a group without such opportunities. The result that punishment decreases within-group variation in success may have important ramifications to understanding the evolution of group-functional behaviours in humans. Selection favours group-beneficial but individually costly traits (the cost

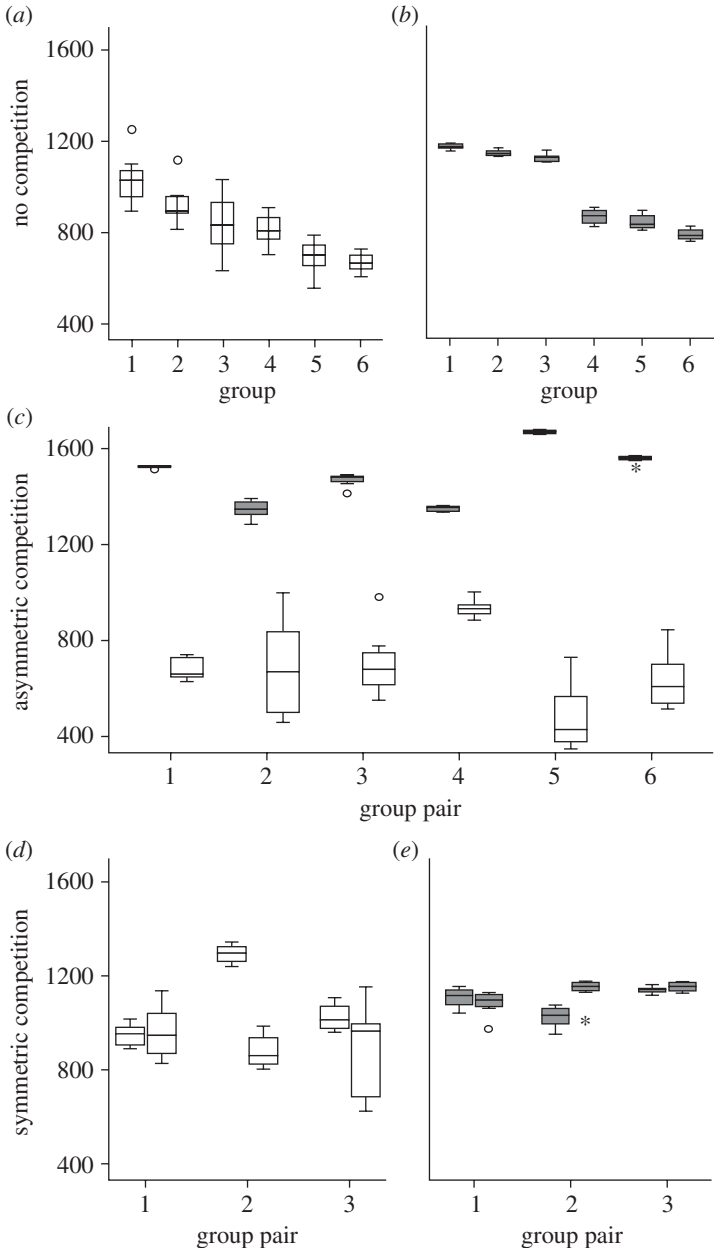

Figure 2. Punishment decreases within-group variation in success. Boxplot graphs depict the median and the degree of dispersion in individual total net earnings in each group of subjects. (*a*) NOPUN treatment, (*b*) PUN treatment, (*c*) PUN–NOPUN treatment, (*d*) NOPUN–NOPUN treatment and (*e*) PUN–PUN treatment. Boxplots with grey shading denote groups with punishment opportunity. The within-group variation in individual pay-offs is smaller in groups with punishment across treatments (non-significant in a comparison of symmetric competition treatments). The average variance of individual net pay-offs within groups in no-competition treatments are PUN 517 MUs and NOPUN 8295 MUs (Mann–Whitney $U_{6,6} = 36$, $p = 0.002$, two-tailed exact); in asymmetric competition PUN–NOPUN 395 MUs and 15307 MUs (Wilcoxon signed-ranks test for six-matched observations, $p = 0.031$, two-tailed exact); and in symmetric competition PUN–PUN 1652 MUs and NOPUN–NOPUN 10 156 MUs (Mann–Whitney test assuming independence of observations $U_{6,6} = 29$, $p = 0.132$, two-tailed exact). In asymmetric group competition, the lowest earning individual in the punishment group earned more than the highest earning individual in the no-punishment group without exception. In isolated groups, highest individual pay-off was earned by an individual belonging to a group without punishment.

being relative to other members of the group) only when variation in success is low within and high among groups [37]. Consequently, by repressing within-group differences in success, punishment attenuates selection operating against individually costly but group beneficial traits. In other words, punishment can function as a form of reproductive levelling that is likely to change the selective environment so that it becomes more favourable to the evolution of behaviours that increase the success of the group relative to other groups [13]. While the idea that the repression of within-group competition shifts competition (and selection) to between-group level is in general well developed in social evolution theory [10], it has thus far been largely overlooked in an effort to understand the origin and effects of costly punishment. Our results suggest a fundamental role for costly punishment in shaping the selective regime in which human social behaviour has evolved.

To explore the factors motivating observed punishment behaviour more closely, we constructed various regression

Table 2. Determinants of received punishment. Multi-level regression coefficient on the determinants of received punishment points in treatments with punishment opportunity (PUN, PUN–NOPUN and PUN–PUN). Models 1, 2 and 3 show the most important motivational factors behind the decision to punish. Model 4 displays the motivational factors behind the punishment in treatments with competition adding the group competition outcome variable and its interaction term with the deviation from average contribution as well as the treatment dummy and its interaction with the slope of punishment. The benchmark treatment for the dummy variable is the symmetric punishment treatment (PUN–PUN). Model 5 includes all punishment data allowing to assess the harshness of punishment behaviour between treatments. The benchmark treatment for the dummy variables (PUN and PUN–NOPUN) is the symmetric punishment treatment (PUN–PUN). Numbers in parentheses indicate standard errors. See electronic supplementary material, tables S2 and S3 for further analyses regarding the determinants of received punishments.

	received punishment points				
independent variables (fixed effects)	no competition	PUN–NOPUN competition	PUN–PUN competition	competition data	all data
	(1)	(2)	(3)	(4)	(5)
deviation from group average	−0.417**	−0.596**	−0.924**	−0.911**	−0.954**
	(0.012)	(0.013)	(0.018)	(0.013)	(0.013)
group average	−0.013	−0.366*	−0.425**	−0.462**	−0.163**
	(0.009)	(0.023)	(0.035)	(0.020)	(0.012)
period	−0.028**	−0.007*	−0.020**	−0.011**	−0.031**
	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)
group competition outcome		−0.008*	−0.140**	−0.007+	
		(0.003)	(0.019)	(0.004)	
group competition outcome × deviation from group average		0.017**	0.014*	0.020**	
		(0.002)	(0.005)	(0.002)	
treatment (PUN–NOPUN)				−0.375+	−0.410
				(0.219)	(0.290)
treatment (PUN–NOPUN) × deviation from group average				0.334**	0.415**
				(0.021)	(0.021)
treatment (PUN)					−1.010**
					(0.296)
treatment (PUN) × deviation from group average					0.532**
					(0.018)
constant	1.288**	7.662**	9.532**	9.989**	4.577**
	(0.385)	(0.578)	(0.711)	(0.443)	(0.018)
controls	yes	yes	yes	yes	yes
random intercepts					
subject–within group	yes	yes	yes	yes	yes
group (std.)	yes	yes	yes	yes	yes
observations	1440	1440	1440	2880	4320
	(48) (6)	(48) (6)	(48) (6)	(96) (12)	(144) (18)
log-likelihood	−2130.47	−1697.79	−2191.60	−4016.57	−6351.51
prob > $\chi 2$	<0.000	<0.000	<0.000	<0.000	<0.000

**Significant at 1%.
*Significant at 5%.
+ Significant at 10%.

models that account for the fact that both individuals and groups undergo repeated measurements and each conflict pair creates a cluster of related groups (table 2). All models control for individual demographic factors referred to as controls (age, gender and cultural background). Given the collected data and our regression-based statistical models, we do not find any consistent demographic differences among our participants with respect to cooperation, punishment or response to punishment (see electronic supplementary material, tables S2 and S3 for more detailed information and estimates). Table 2 indicates that free-riding (negative deviation from the average contribution) was a major factor explaining the number of received punishment points in all treatments. Unlike many previous studies [38–40], we found no evidence for antisocial punishment targeted towards cooperators (electronic supplementary material, figure S1). Group's average contribution proves to be significant only in competing groups (table 2: models 2, 3 and 4) where it is proportional to the cooperative effort in the rival group, indicating that the motivations to punish are qualitatively different between groups with and without competitive pressure. In all treatments, the number of received punishment points decreased as the game proceeded (see also figure 3). In competing groups, the outcome of group competition (amount of MUs transferred to/from each group member) affected participants' eagerness to impose a penalty on their peers so that the severity of defeat increased the harshness of individual punishments (model 4). Overall, the intensity of group conflict had a marked effect on the willingness to punish uncooperative group members (figure 3). In PUN–PUN, where rivalry between groups

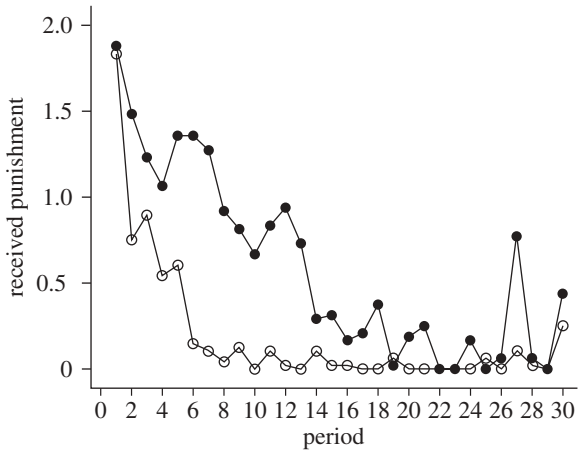

Figure 3. Intensity of group conflict increases punishment. The average number of received punishment points per individual over all 30 periods of play in treatments with punishment and group competition. Symbols with black filling denote the symmetric group competition treatment with punishment (PUN–PUN) and transparent symbols denote the asymmetric group competition treatment (PUN–NOPUN). Free-riders were punished harder in PUN–PUN where competition between groups was intensified owing to the more comparable chance of equal group performance between the competing groups (see also models 4 and 5 in table 1).

was pronounced, the same amount of free riding was punished more severely than in PUN–NOPUN where groups with punishment opportunity dominated in group competition (models 4 and 5 in table 2). Results suggest that the growing external threat to group and individual success increases readiness to sacrifice individual resources to disciplinary action. This result corresponds with the real life observations of heightened sanctions adopted in times of group conflict, like the increased social sanctioning targeted towards surviving deserters [41].

4. CONCLUSIONS

Darwin [42] suggested that competition between bands could select for individual traits such as courage and faithfulness, which benefit the group in conflict situations. The little direct historical evidence available on intergroup variation and patterns of extinctions over the evolution of human social behaviour stresses the prospect that lethal group conflict may have been frequent enough to allow the proliferation of individually costly, but group beneficial traits [11,43]. Likewise, the importance of behavioural traditions for group cohesiveness is attested in many avenues of present-day social life. Intergroup rivalries are present in varying organizational levels including war, competition for foreign direct investment, promotion tournaments in labour markets, and team sports.

The inclination to punish norm violators is a human universal [44] but accounting for its evolution is an evolutionary puzzle. Our experimental results from direct intergroup conflict demonstrate that costly punishment entails undisputable individual benefits in a population of competing groups with repeated intra-and intergroup interactions. Moreover, we find that the competitive pressure between conflicting groups evokes qualitative differences in the use of punishment. Given the robust result that costly punishment generates significantly more equal distributions of material pay-offs, it prepares the ground for the evolution of parallel group-functional traits and traditions. These results support the importance of intergroup competition in the emergence of costly punishment and human cooperation, stressing the prospect that parts of the human behavioural repertoire have evolved as group-functional traits through conflicts between human communities.

This paper has created an illustrative setting to shed light on the ultimate cause behind the widely observed costly punishment. At the same time, it prepares the ground for more comprehensive experimental investigations to identify various parallel evolutionary causes such as reputation, signalling and moral standards that may as well maintain cooperation. In fact, earlier research suggests that various proximate causes may efficiency interact to boost cooperation [45]. Our aim has been to conduct an experiment as parsimonious as possible without neglecting any of the design principles important for the emergence of costly punishment. We demonstrate the principle, not the actual course of human history. The results might have been different, had we, for instance, allowed more direct behavioural means for retaliation [46,47]. Likewise, the detrimental habit of punishing cooperators in some human societies [40] is shown to create a possible caveat to the coevolution of punishment and cooperation [48]. Consequently, in future studies it would be worthwhile to examine the effect of between-group competition on the degree of antisocial punishment in societies where antisocial punishment is common.

The demonstrated success of costly punishment in situations where groups interact should not be understood as something that inevitably leads to behavioural adaptations, helping humans to establish a culture of cooperation. The characteristics of human evolution and socio-ecological trajectories are utterly complex phenomena where cooperative predisposition may concurrently co-evolve with destructive elements leading to ruinous rivalries [32,34]. A deeper understanding of these phenomena will help us to prevent tragedies.

The authors would like to thank Samuel Bowles, Werner Güth, Sebastian Krügel, Kari Nissinen, Elinor Ostrom, James Walker and Johannes Weisser for suggestions and discussions when preparing and revising the manuscript. The paper has benefited from comments made by the participants of the Experimental Reading Group at the Workshop in Political Theory and Policy Analysis during the autumn semester 2009 as well as the seminar participants at the Max Planck Institute of Economics, the ESA North-American Meeting 2009 and the IMEBE 2010 conference. Authors are indebted to Rico Löbel for his research assistance and to Abelheid Baker for her editing help. Financial support from the Max Planck Society and the Center of Excellence in Evolutionary Research, Academy of Finland, is gratefully acknowledged. The authors declare no conflict of interest.

REFERENCES

1 Hardin, G. 1968 The tragedy of the commons. *Science* **162**, 1243–1248. (doi:10.1126/science.162.3859.1243)
2 Yamagishi, T. 1986 The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110–116. (doi:10.1037/0022-3514.51.1.110)

3 Boyd, R. & Richerson, P. J. 1992 Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **13**, 171–195. (doi:10.1016/0162-3095(92)90032-Y)
4 Ostrom, E., Walker, J. & Gardner, R. 1992 Covenants with and without a sword: self-governance is possible. *Am. Political Sci. Rev.* **86**, 404–417. (doi:10.2307/1964229)
5 Fehr, E. & Gächter, S. 2002 Altruistic punisment in humans. *Nature* **415**, 137–140. (doi:10.1038/415137a)
6 Falk, A., Fehr, E. & Fischbacher, U. 2005 Driving forces behind informal sanctions. *Econometrica* **73**, 2017–2030. (doi:10.1111/j.1468-0262.2005.00644.x)
7 Houser, D. & Xioa, E. 2005 Emotion expression in human punishment behavior. *Proc. Natl Acad. Sci. USA* **102**, 7398–7401. (doi:10.1073/pnas.0502399102)
8 de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. & Fehr, E. 2004 The neural basis of altruistic punishment. *Science* **305**, 1254–1258. (doi:10.1126/science.1100735)
9 Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. 2003 The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **00**, 3531–3535. (doi:10.1073/pnas.0630443100)
10 Frank, S. A. 2003 Perspective: repression of competition and the evolution of cooperation. *Evolution* **57**, 693–705.
11 Bowles, S. 2009 Did warfare among ancestral hunter–gatherer affect the evolution of human social behaviors? *Science* **324**, 1293–1298. (doi:10.1126/science.1168112)
12 Keeley, L. H. 1996 *War before civilization: the myth of the peaceful savage*. Oxford, UK: Oxford University Press.
13 Bowles, S. 2006 Group competition, reproductive leveling, and the evolution of human altruism. *Science* **314**, 1569–1572. (doi:10.1126/science.1134829)
14 Voors, M., Nillesen, E., Verwimp, P., Bulte, E., Lensink, R. & van Soest, D. In press. Does conflict affect preferences? Results from field experiments in burundi. *Am. Econ. Rev.*
15 Knauft, B. 1991 Violence and sociality in human evolution. *Curr. Anthropol.* **32**, 391–428. (doi:10.1086/203975)
16 Boehm, C. 1993 Egalitarian behavior and reverse dominance hierarchy. *Curr. Anthropol.* **34**, 227–254. (doi:10.1086/204166)
17 Gibson, C. C., William, J. T. & Ostrom, E. 2005 Local enforcement and better forests. *World Dev.* **33**, 273–284. (doi:10.1016/j.worlddev.2004.07.013)
18 Coleman, E. A. & Steed, B. C. 2009 Monitoring and sanctioning in the commons: an application to forestry. *Ecol. Econ.* **68**, 2106–2113. (doi:10.1016/j.ecolecon.2009.02.006)
19 Marlowe, F. W. *et al.* 2008 More 'altruistic' punishment in larger societies. *Proc. R. Soc. B* **275**, 587–592. (doi:10.1098/rspb.2007.1517)
20 Henrich, J. *et al.* 2010 Markets, religion, community size, and the evolution of fairness and punishment. *Science* **327**, 1480–1484. (doi:10.1126/science.1182238)
21 Boehm, C. 1999 *Hierarchy in the forest: the evolution of egalitarian behavior*. Cambridge, MA: Harvard University Press.
22 Ostrom, E. 1990 *Governing the commons: the evolution of institutions for collective action*. Cambridge, UK: Cambridge University Press.
23 Sefton, M., Shupp, R. & Walker, J. M. 2007 The effect of rewards and sanctions in provision of public goods. *Econ. Inq.* **45**, 671–690. (doi:10.1111/j.1465-7295.2007.00051.x)
24 Egas, M. & Riedl, A. 2008 The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B* **275**, 871–878. (doi:10.1098/rspb.2007.1558)
25 Dreber, A., Rand, D. G., Fudenberg, D. & Nowak, M. A. 2008 Winners don't punish. *Nature* **452**, 348–351. (doi:10.1038/nature06723)
26 Wu, J.-J., Zhang, B.-Y., Zhou, Z.-Z., He, Q.-Q., Zheng, X.-D., Cressman, R. & Tao, Y. 2009 Costly punishment does not always increase cooperation. *Proc. Natl Acad. Sci. USA* **106**, 17 448–17 451. (doi:10.1073/pnas.0905918106)
27 Gächter, S., Renner, E. & Sefton, M. 2008 The long-run benefits of punishment. *Science* **322**, 1510. (doi:10.1126/science.1164744)
28 Boyd, R., Gintis, H. & Bowles, S. 2010 Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* **328**, 617–620. (doi:10.1126/science.1183665)
29 Puurtinen, M. & Mappes, T. 2009 Between-group competition and human cooperation. *Proc. R. Soc. B* **276**, 355–360. (doi:10.1098/rspb.2008.1060)
30 Bornstein, G. & Ben-Yossef, M. 1994 Cooperation in intergroup and single-group social dilemmas. *J. Exp. Soc. Psychol.* **30**, 52–67. (doi:10.1006/jesp.1994.1003)
31 Baron, J. 2001 Confusion of group-interest and self-interest in parochial cooperation on behalf of a group. *J. Confl. Resolution* **45**, 283–296. (doi:10.1177/0022002701045003002)
32 Abbink, K., Brandts, J., Herrmann, B. & Orzen, H. 2010 Intergroup conflict and intragroup punishment in an experimental contest game. *Am. Econ. Rev.* **100**, 420–447. (doi:10.1257/aer.100.1.420)
33 Leibbrandt, A. & Sääksvuori, L. 2010 More than words: communication in intergroup conflicts. Jena Economic Research Papers 2010-065. Friedrich-Schiller-University Jena, Germany. Max-Planck-Institute of Economics.
34 Choi, J.-K. & Bowles, S. 2007 The coevolution of parochial altruism and war. *Science* **318**, 636–640. (doi:10.1126/science.1144237)
35 Fiscbacher, U. 2007 z-Tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* **10**, 171–178. (doi:10.1007/s10683-006-9159-4)
36 Janssen, M., Holahan, R., Lee, A. & Ostrom, E. 2010 Lab experiments for the study of social–ecological systems. *Science* **328**, 613–617. (doi:10.1126/science.1183532)
37 Price, G. R. 1972 Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**, 485–490. (doi:10.1111/j.1469-1809.1957.tb01874.x)
38 Cinyabuguma, M., Page, T. & Putterman, L. 2004 On perverse and second-order punishment in public goods experiments with decentralized sanctioning. Working Papers 2004–12. Providence, RI: Brown University, Department of Economics.
39 Denant-Boemont, L., Masclet, D. & Noussair, C. 2007 Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Econ. Theor.* **33**, 145–167. (doi:10.1007/s00199-007-0212-0)
40 Herrmann, B., Thöni, C. & Gächter, S. 2008 Antisocial punishment across societies. *Science* **319**, 1362–1367. (doi:10.1126/science.1153808)
41 Costa, D. L. & Kahn, M. E. 2007 Deserters, social norms, and migration. *J. Law Econ.* **50**, 323–353. (doi:10.1086/511321)
42 Darwin, C. 1871 *The descent of man and selection in relation to sex*, 2nd edn. London, UK: John Murray.
43 Soltis, J., Boyd, R. & Richerson, P. J. 1995 Can group-functional behaviors evolve by cultural group selection?

An empirical test. *Curr. Anthropol.* **36**, 473–494. (doi:10.1086/204381)
44 Henrich, J. *et al.* 2006 Costly punishment across human societies. *Science* **312**, 1767–1770. (doi:10.1126/science.1127333)
45 Rockenbach, B. & Milinski, M. 2006 The efficient interaction of indirect reciprocity and costly punishment. *Nature* **444**, 718–723. (doi:10.1038/nature05229)
46 Nikiforakis, N. 2008 Punishment and counter-punishment in public good games: can we really govern ourselves? *J. Public Econ.* **92**, 91–112. (doi:10.1016/j.jpubeco.2007.04.008)
47 Janssen, M. A. & Bushman, C. 2008 Evolution of cooperation and altruistic punishment when retaliation is possible. *J. Theor. Biol.* **254**, 541–545. (doi:10.1016/j.jtbi.2008.06.017)
48 Rand, D. G., Armao, J. J., Nakamaru, M. & Ohtsuki, H. 2010 Anti-social punishment can prevent the co-evolution of punishment and cooperation. *J. Theor. Biol.* **265**, 624–632. (doi:10.1016/j.jtbi.2010.06.010)

Supporting Information for Costly Punishment Prevails in Intergroup Conflict

Lauri Sääksvuori,[1*] Tapio Mappes,[2] Mikael Puurtinen[2]

[1]Max Planck Institute of Economics
Kahlaische Strasse 10, 07745 Jena, Germany
[2]Centre of Excellence in Evolutionary Research,
The Department of Biological and Environmental Science,
University of Jyväskylä, PO Box 35, Jyväskylä 40014, Finland

[*]To whom correspondence should be addressed;
E-mail:saaksvuori@econ.mpg.de

1 Experimental Design

Our general design principle has been to develop an experimental design that facilitates compatibility of our results with the pertinent literature. In every treatment, the composition of participants in each group stayed intact throughout the game. At the beginning of each period, subjects were randomly assigned an ID number (from 1 to 8) to distinguish their action from the others within a period. Reassigning the ID numbers after each period ensured that participants could not create a link between the actions of other participants across periods. In all treatments, participants were informed about the individual contributions and corresponding earnings in their own group after the contribution stage in each period. In addition, the total amount of contributions in the directly competing group was revealed to participants. Participants were not informed about the individual punishment decisions of other participants. That is, they knew neither who of their peers punished them nor whether other group members were punished. In an asymmetric competition treatment (PUN-NOPUN), where punishing and non-punishing groups were compared, no information about the punishment opportunity was revealed to the group without punishment. One experimental session lasted on average 90 minutes. The total earnings of a participant equaled the sum of payoffs over all periods. Earnings per participant ranged from €9 to €36 with an average of €20. The experiment was programmed and run using z-Tree [1].

In treatments with punishment opportunities (PUN, PUN-NOPUN and PUN-PUN), participants had in each period a chance to allocate a maximum of five punishment points to each member in their group. Each punishment point cost the punisher 1 MU and reduced the earnings of the receiver by 3 MUs. Participants could refrain from punishing their own group members by entering 0 in the corresponding field on a computer screen. An experimental

rule guaranteed that no participant would incur negative payoff due to received punishment points. The possibility to assign punishment points was guaranteed after all possible outcomes by allowing subjects to produce negative earnings through the costs of punishment.

The applied punishment procedure largely follows the standard practice that has been established in the literature since the introduction of linear punishment technology [2]. The cost-impact ratio of 3:1 can be, however, deemed very effective in comparison to many other ratios whose effectiveness has been experimentally examined [3]. At the same time, related studies have found that the regularly applied punishment technologies without explicit means to coordinate individual punishment may severely hinder the true effectiveness of the punishment activity [4, 5]. We have chosen the 3:1 ratio in view of these results to facilitate the compatibility of our results with the other existing studies.

One could think that an experimental rule guaranteeing that no participant will incur negative payoff due to received punishment points would possibly generate an unfair competitive advantage in a group competition against a group without punishment opportunity. However, it is important to note that besides facilitating the compatibility between studies this rule is of minor empirical relevance in our study. Only two out of 4320 (0.05%) of individual per period payoffs in punishing groups yielded negative nominal payoffs.

A similar concern regarding to a possible unjustified competitive advantage of punishment is thinkable due to an experimental practice that holds group members liable to cover the existing difference between group accounts provided that some of their own group members do not have enough MUs to cover their proportion of losses originating from the group comparison. Also in this case, empirical perusal reveals that the situation in which other group members are held liable to cover the losses of other group members never took place in groups with punishment

opportunity during the course of actual experiment. This situation happened though in non-punishing groups at the rate of 0.12% (7/5760).

2 Group Conflict Model

Let $\Pi_i(x_i, X_A, X_B)$ denote the payoff of a representative player i, where x_i is the contribution made by player i, X_A is the sum of contributions in player i's group and X_B is the total contribution in the competing group. E_i denotes individual endowment. The payoff function for a player i is given by equation (1)

$$\Pi_i = E_i - x_i + \frac{\alpha \sum_{a=1}^{N} x_i}{N} + \beta(\frac{X_A - X_B}{N}) \text{ for all } i \in X_A, \tag{1}$$

where α denotes the intragroup productivity and β signals the intensity of intergroup conflict. Consider equation (1) re-written as:

$$\Pi_i = E_i - x_i + \frac{\alpha \sum_{a=1}^{N} x_i}{N} + \beta\frac{\sum_{a=1}^{N} x_i}{N} - \beta\frac{\sum_{b=1}^{N} x_b}{N} \text{ for all } i \in X_A \tag{2}$$

and take the partial derivative subject to x_i

$$\frac{\partial \Pi_i}{\partial x_i} = -1 + \frac{\alpha}{N} + \frac{\beta}{N}. \tag{3}$$

Let the group size (N) be 8 and $\alpha = \beta = 2$ as in the experiment. As noticed from (3) given parameters indicate that investing one additional MU to group account equals a net benefit of -0.5 MUs. The game is a finitely repeated social dilemma in which dominant strategy for a self-interested opportunist remains to contribute nothing despite the group conflict. The collective welfare is, however, maximized only when players make full contributions.

3 Supporting Analyses

Table S1 shows the complete multilevel regression models including the control variables on the determinants of assigned punishments. We do not find any significant effect based on age, gender or cultural background. We have chosen a mixed effect regression to account for the fact that that both individuals and groups undergo repeated measurements and each conflict pair creates a cluster of related groups. In table S2, we study the robustness of the mixed effect regression coefficients reported in the main article by estimating alternative random effects models where individual are additionally clustered within their group. The random effects regression estimates with clustered group level observations yield similar results to the mixed effect regression for our data.

In table S3, we study if the response to punishment differs according to the competition regimes or demographic factors. Models presented in Table S3 allow us also to examine the nature of punishment in more detail. While the models (1), (2) and (3) estimated separately for each treatment with punishment opportunity suggest that there is a qualitative difference in responses to punishment (the effect received punishment points being highly significant in PUN-(NOPUN) and PUN-PUN, but non-significant in PUN) such that the participants adjust their behaviour more in treatments with competition, the model that directly compares these treatments with the help of treatment dummies does not indicate statistically significant differences in response to received punishment points between treatments. We do not find any effects in response to punishment driven by age, gender or cultural background.

Given the estimated effect size, one assigned punishment point roughly increases contributions to a group account by 0.2 MUs in the subsequent period. The cumulative effect of increased contributions to the group account is highly dependent on the period punishment

takes place (assuming that the effect of received punishment point is lasting unchanged over periods). Based on the estimated effect size, one assigned punishment point leads to an individual benefit of 0.1 MUs. Given the cost of 1 MU for each assigned punishment point, it takes about 10 periods to recover the individual cost of assigned punishment points through increased investments into the group account. Therefore, it can be well deemed that punishments that are assigned before the 20th period can be guided by self-interest. Combining this result with the significant decrease in the number of assigned punishment points over periods further supports the potential importance of self-interested motivations behind the decision to punish. However, not all observed punishment can be self-interested as it becomes evident from the relatively large number of punishment points assigned in the last period of the game. This interesting and quite common pattern of 'end-game punishment' is perhaps best deemed as a truly spiteful behaviour along the lines of Hamilton's [6, 7] original definition.

Figure S1 shows the average number of received punishment points for a given deviation from group's average contribution. Unlike many previous studies [8, 9, 10], we find no evidence for antisocial punishment targeted toward cooperators. Punishment was directed toward free-riders, i.e. below average contributors, in all treatments. Figure S2 displays separately the average contributions, net pay-offs and number of received punishment points in each group in each treatment.

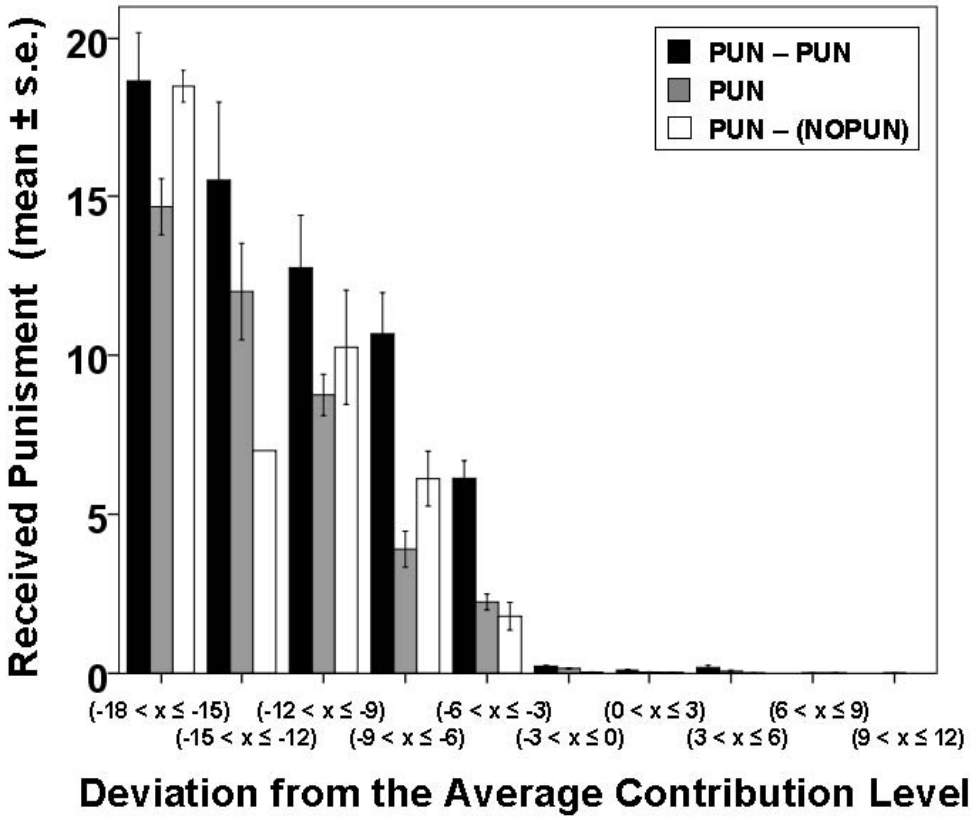

Figure S1:The average number of received punishment points for a given deviation from group's average contribution. Deviations are grouped into intervals of equal size. Punishment was directed towards free-riders, i.e. below average contributors.

Table S1: Multilevel regression coefficients on the determinants of received punishment points in treatments with punishment opportunity (PUN, PUN-NOPUN and PUN-PUN). Models 1, 2 and 3 show the most important motivational factors behind the decision to punish. Model 4 displays the motivational factors behind the punishment in treatments with competition adding the group competition outcome variable and its interaction term with the deviation from average contribution as well as the treatment dummy and its interaction with the slope of punishment. The benchmark treatment for the dummy variable is the symmetric punishment treatment (PUN-PUN). Model 5 includes all punishment data allowing to assess the harshness punishment of punishment behavior between treatments. The benchmark treatment for the dummy variables (PUN and PUN-NOPUN) is the symmetric punishment treatment (PUN-PUN). Variable 'culture' refers to participant's cultural background measured by the location where he/she completed the high-school education [0=The new federal states of Germany; 1=The old federal states of Germany]. **Significant at 1%; *Significant at 5%; +Significant at 10%. Numbers in parentheses indicate standard errors.

	Received punishment points				
Independent variables	PUN No-Competition	PUN-(NOPUN) Competition	PUN-PUN Competition	Competition Data	All Data
(fixed effects)	(1)	(2)	(3)	(4)	(5)
Deviation from group average	-0.417**	-0.596**	-0.924**	-0.911**	-0.954**
	(.012)	(.013)	(.018)	(.013)	(.013)
Group average	-0.013	-0.366**	-0.425**	-0.462**	-0.163**
	(.009)	(.023)	(.035)	(.020)	(.012)
Period	-0.028**	-0.007*	-0.020**	-0.011**	-0.031**
	(.003)	(.003)	(.004)	(.002)	(.002)
Group competition outcome		-0.008*	-0.140**	-0.007+	
		(.003)	(.019)	(.004)	
Group competition outcome x		0.017**	0.014*	0.020**	
Deviation from group average		(.002)	(.005)	(.002)	
Treatment [PUN-NOPUN]				-0.375+	-0.410
				(.219)	(.290)
Treatment [PUN-NOPUN] x				0.334**	0.415**
deviation from group average				(.021)	(.021)
Treatment [PUN]					-1.010**
					(.296)
Treatment [PUN] x					0.532**
deviation from group average					(.018)
Age	-0.012	-0.010	-0.016	-0.012	-0.015
	(.015)	(.016)	(.012)	(.009)	(.009)
Gender [0=Women]	-0.007	0.041	0.016	0.021	0.023
	(.076)	(.107)	(.067)	(.053)	(.053)
Culture [0=Eastern Germany]	0.105	0.007	-0.063	-0.033	0.009
	(.130)	(.170)	(.096)	(.080)	(.082)
Constant	1.288**	7.662**	9.532**	9.989**	4.577**
	(.385)	(.578)	(.711)	(.443)	(.018)
Random Intercepts					
Subject -within group (std.)	0.145	0.283	0.091	0.160	0.214
Group (std.)	0.092	0.349	0.202	0.362	0.492
Observations	1440	1440	1440	2880	4320
	(48) (6)	(48) (6)	(48) (6)	(96) (12)	(144) (18)
Log-likelihood	-2130.47	-1697.79	-2191.60	-4016.57	-6351.51
Prob > χ^2	< 0.000	< 0.000	< 0.000	< 0.000	< 0.000

Table S2: Random effects regression estimates with clustered group level observations on the determinants of received punishment points in treatments with punishment opportunity (PUN, PUN-NOPUN and PUN-PUN). Models 1, 2 and 3 show the most important motivational factors behind the decision to punish. Model 4 displays the motivational factors behind the punishment in treatments with competition adding the group competition outcome variable and its interaction term with the deviation from average contribution as well as the treatment dummy and its interaction with the slope of punishment. The benchmark treatment for the dummy variable is the symmetric punishment treatment (PUN-PUN). Model 5 includes all punishment data allowing to assess the harshness punishment of punishment behavior between treatments. The benchmark treatment for the dummy variables (PUN and PUN-NOPUN) is the symmetric punishment treatment (PUN-PUN). Variable 'culture' refers to participant's cultural background measured by the location where he/she completed the high-school education [0=The new federal states of Germany; 1=The old federal states of Germany]. **Significant at 1%; *Significant at 5%; +Significant at 10%. Numbers in parentheses indicate robust standard errors.

	Received punishment points				
Independent variables	PUN No-Competition	PUN-(NOPUN) Competition	PUN-PUN Competition	Competition Data	All Data
(Random effects)	(1)	(2)	(3)	(4)	(5)
Deviation from group average	-0.415**	-0.576**	-0.924**	-0.912**	-0.954**
	(.068)	(.141)	(.085)	(.070)	(.068)
Group average	-0.016	-0.232**	-0.443**	-0.385**	-0.066**
	(.011)	(.086)	(.116)	(.098)	(.029)
Period	-0.028**	-0.012**	-0.019**	-0.015**	-0.034**
	(.010)	(.003)	(.010)	(.005)	(.006)
Group competition outcome		-0.009+	-0.137**	-0.007	
		(.005)	(.035)	(.007)	
Group competition outcome x		0.022**	0.014*	0.019**	
Deviation from group average		(.007)	(.013)	(.006)	
Treatment [PUN-NOPUN]				-0.373+	-0.408
				(.192)	(.122)
Treatment [PUN-NOPUN] x				0.331*	0.419**
deviation from group average				(.161)	(.180)
Treatment [PUN]					-0.546**
					(.194)
Treatment [PUN] x					0.533**
deviation from group average					(.094)
Age	-0.010	-0.000	-0.004	-0.002	-0.018
	(.006)	(.009)	(.011)	(.007)	(.006)
Gender [0=Women]	-0.011	0.018	0.003	0.023	0.007
	(.069)	(.055)	(.060)	(.047)	(.035)
Culture [0=Eastern Germany]	0.078	0.094	-0.108	-0.011	0.009
	(.140)	(.079)	(.103)	(.084)	(.067)
Constant	1.272**	4.939**	9.587**	8.338**	2.441**
	(.250)	(.578)	(2.294)	(1.882)	(.627)
Observations	1440	1440	1440	2880	4320
R^2 - Overall	0.490	0.5962	0.792	0.731	0.642
R^2 - Within	0.482	0.633	0.788	0.746	0.652
R^2 - Between	0.684	0.549	0.852	0.646	0.627
Prob $> \chi^2$	-	-	-	< 0.000	< 0.000

S3: Random effects regression estimates with clustered group level observations indicating individual response to received punishments in the preceding period in treatments with punishment opportunity (PUN, PUN-NOPUN and PUN-PUN). Responses by individual i is measured as the difference in contributions to group account from period t-1 to t. Models 1, 2 and 3 show the significance of received punishment separately in each treatment with punishment. Model 4 compares responses between treatments. The benchmark variable for the treatment dummies (PUN and PUN-PUN) is the punishment in the asymmetric competition treatment (PUN-NOPUN). Variable 'culture' refers to participant's cultural background measured by the location where he/she completed the high-school education [0=The new federal states of Germany; 1=The old federal states of Germany]. **Significant at 1%; *Significant at 5%; +Significant at 10%. Numbers in parentheses indicate robust standard errors.

Difference in contributions from period t-1 to t				
Independent variables	PUN No-Competition	PUN-(NOPUN) Competition	PUN-PUN Competition	All Punishment Data
(Random effects)	(1)	(2)	(3)	(4)
Number of received	0.210	0.217**	0.107**	0.168
punishment points t-1	(.152)	(.055)	(.041)	(.105)
Contribution t-1	-0.705**	-0.522**	-0.721**	-0.717**
	(.106)	(.039)	(.080)	(.066)
Group sum t-1	0.060**	0.036**	0.023**	0.055**
	(.013)	(.009)	(.004)	(.009)
Period	-0.010	-0.012+	0.024**	-0.006
	(.008)	(.007)	(.006)	(.005)
Gender [0=Women]	1.420	0.209*	0.241	0.701
	(1.013)	(.104)	(.122)	(.362)
Gender [0=Women] x	0.229	0.209	0.063+	
Received punishments t-1	(.235)	(.104)	(.065)	
Treatment [PUN-PUN]				-0.085
				(.133)
Treatment [PUN-PUN] x				0.027
Received punishments t-1				(.095)
Treatment [PUN]				-1.163
				(.974)
Treatment [PUN] x				0.115
Received punishments t-1				(.103)
Age	0.332	-0.344	-0.007	-0.012
	(.078)	(.025)	(.022)	(.022)
Culture [0=Eastern Germany]	0.440	0.328	-0.148	0.025
	(.578)	(.122)	(.269)	(.218)
Constant	2.271+	5.434**	10.170**	5.581**
	(1.354)	(1.568)	(1.742)	(1.218)
Observations	1392	1392	1932	4176
R^2 Overall	0.102	0.406	0.466	0.203
R^2 Within	0.428	0.500	0.490	0.465
R^2 Between	0.007	0.081	0.360	0.010

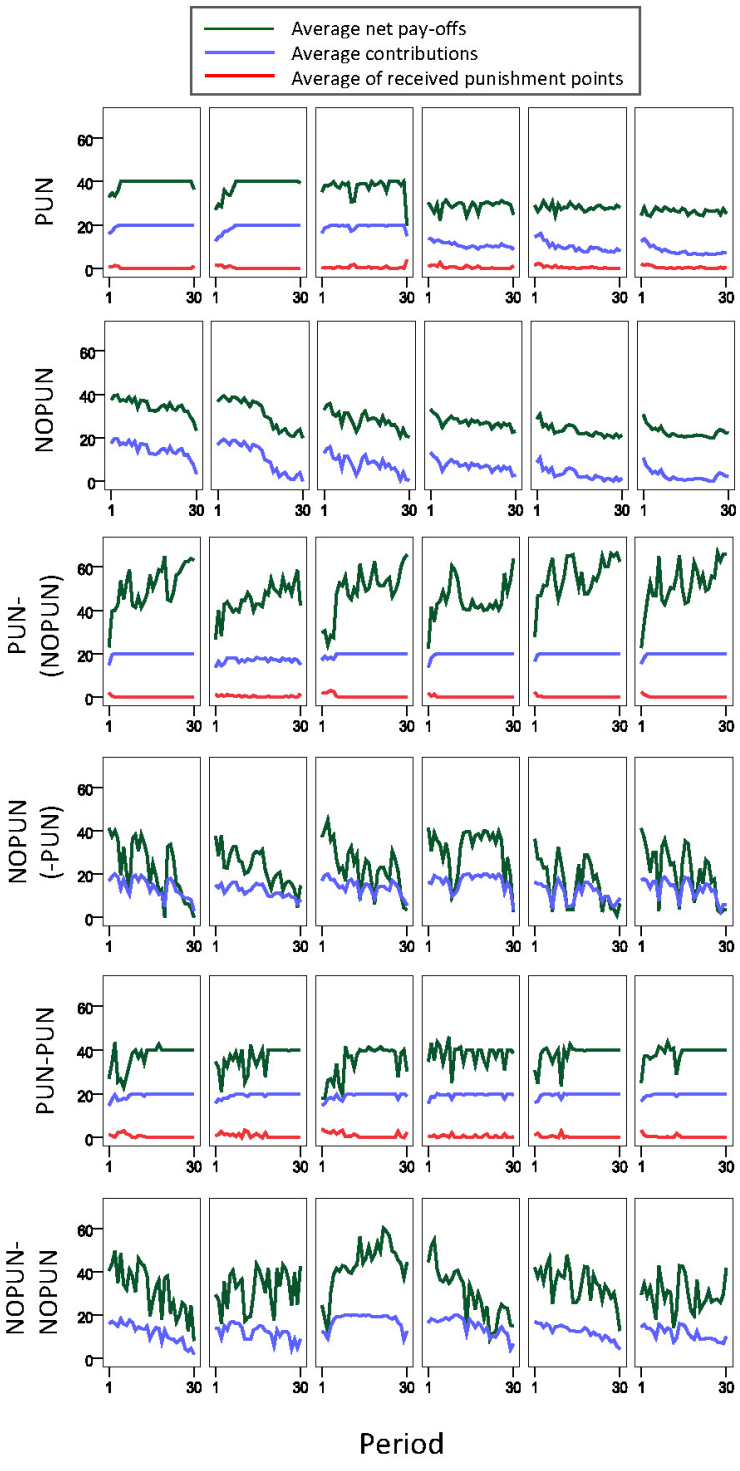

Figure S2: Average contributions, net pay-offs and the number of received punishment points separately for each eight participant group in each treatment.

References

[1] Fiscbacher, U. 2007 z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, **10**, 171–178.

[2] Fehr, E. and Gächter, S. 2002 Altruistic punisment in humans. *Nature*, **415**, 137–140.

[3] Egas, M. and Riedl, A. 2008 The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B-Biological Sciences*, **275**, 871–878.

[4] Janssen, M., Holahan, R., Lee, A., and Ostrom, E. 2010 Lab experiments for the study of social-ecological systems. *Science*, **328**, 613–617.

[5] Boyd, R., Gintis, H., and Bowles, S. 2010 Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, **328**, 617–620.

[6] Hamilton, W. D. 1964 The genetical evolution of social behaviour, I & II. *Journal of Theoretical Biology*, **7**, 1–52.

[7] Hamilton, W. D. 1970 Selfish and spiteful behaviour in an evolutinary model. *Nature*, **288**, 1218–1220.

[8] Cinyabuguma, M., Page, T., and Putterman, L. 2004 On perverse and second-order punishment in public goods experiments with decentralized sanctioning. Working Papers 2004-12, Brown University, Department of Economics.

[9] Denant-Boemont, L., Masclet, D., and Noussair, C. 2007 Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, **33**, 145–167.

[10] Herrmann, B., Thöni, C., and Gächter, S. 2008 Antisocial punishment across societies. *Science*, **319**, 1362–1367.

Instructions

Thank you for coming! You are now about to take part in an experiment on decision making. You have earned 2.50 Euro for showing up on time. Reading carefully the following instructions and taking part in the experiment you can earn a considerable amount of money depending both on your own decisions and on the decisions of others.

These instructions and the decisions to be made are only for your private information. During the experiment you are neither allowed to communicate in the laboratory nor with someone outside the laboratory. Please switch off your mobile phone. Any violation of these rules will lead to exclusion from the experiment and all payments. If you have any questions regarding the rules or the course of this experiment, please raise your hand. An experimenter will assist you privately.

During the experiment all decisions and transfers are made in Experimental Currency Units (ECUs). Your total income will be calculated in ECUs and at the end of the experiment converted to Euros at the following rate:

1 ECU = 0.02 Euro

The experiment consists of thirty (30) consecutive decision periods. Your total earnings will be determined as a sum of your earnings from all these periods. **At the beginning of the experiment, participants will be divided into groups of eight (8) individuals.** During the experiment you will interact with your own group members and one other group of eight participants. The composition of the groups will stay the same in each period. This means that you interact throughout the experiment with the same people both within your own group and in the other group. You will never be informed about the real identity of other participants in this experiment; neither will they know with whom they interact. Your total earnings will be privately paid in cash at the end of the experiment.

1. Stage 1

You will be a member in a group of 8 participants. At the beginning of each round all members are endowed with 20 ECUs. Your task is to allocate them either into your private account or you can invest them fully or partially into a project. Each unit not invested into a project automatically remains in your private account. You earnings compose of your private and project accounts. All participants in your group will simultaneously face the same decision situation.

1.1 Your income from the private account

You will earn one ECU for each unit allocated to your private account. No other member in your group will earn from your private account.

1.2 Your income from the project account

You will earn from the project account based on the sum of investments made by all members in your group. Each member will profit equally from the invested amount. This means that you will earn from your own investment as well as from the investment of others. The income for each member in your group from the project account will be the determined as follows:

The sum of all investments will be doubled and divided equally among all members in your group

Example. The sum of all contributions into a project account is 100 ECUs. The number of contributions will be doubled to (2 * 100 =) 200 ECUs and divided equally among all eight members in your group. Your earnings from the group account will be 200/8 = 25 ECUs.

1.3 Your total income from Stage 1

Your total income consists both from the amount on your private account and the total amount of investments into a project account.

Your income = Income from your private account + Income from the project account

Example. Assume that you have allocated 10 ECUs into your private account and 10 ECUs into a project account. The total amount of investments into the project account in your group is 80 ECUs. Your income will be 10 + 80 * 2 / 8 = 10 + 160/8 = 10 + 20 = 30 ECUs.

2. Stage 2

In the second stage, the sum of tokens contributed into the project account in your group will be compared with another group's project account. This comparison will always be made between the same two groups. Should the sum of contributions in your group's project account exceed the sum of contributions in the other group**, your group wins twice the difference between project accounts.** Correspondingly, should the sum of contributions in your group's project account be below the sum of contributions in the other group, your group loses twice the difference between project accounts.

The wins and losses from the group comparison are divided equally among the group members.[1] Possible losses will be deducted individually from the combined private and project account income.

Example. The sum of contributions into your group's project account is 140 ECUs. The group with whom you are compared to has made a contribution of 100 ECUs. The difference between the project contributions is 40 ECUs. This difference will now be doubled to (2 * 40 =) 80 ECUs and divided equally among the group members. Thus, you earn 10 ECUs from the comparison between groups. Correspondingly, each member in the other group will lose 10 ECUs.

[1] In case some group member(s) does not have enough ECUs to cover her proportion of losses from the group comparison, other members in the corresponding group are held liable to cover the existing difference between group accounts. Should the total loss (twice the difference between group accounts) exceed the amount of ECUs earned by all group members, each member will lose her earnings from the current round. However, the winning group can only win as many ECUs as the losing group has in total. Negative individual earnings from the group comparison are not possible.

3. Stage 3

You will see how much each member in your group contributed into a project account and their corresponding individual earnings after the group comparison stage. **You will now make a decision whether to decrease the earnings of your group members by assigning deduction points to them**. All members in your group have the same opportunity.

All individual contributions into a project account in your group are displayed to group members in random order. For example, the first column on left could present a different group member in different periods. The same holds for all columns.

Your task is to decide how many deduction points you want to assign to each other member in your own group. **You may assign up to 5 points to each group member.** If you do not want to change the earnings of a specific group member, you have to enter 0 into a corresponding input field. **Each deduction point you assign costs you 1 ECU and will decrease the earnings of its target by 3 ECUs.** Similarly, the other members in your group have the possibility to assign deduction points to you. Each received deduction point will decrease your earnings by 3 ECUs.

Example. You assign a total amount of 10 deduction point to four different members in your group. Assume that the corresponding individual allocation of deduction points is 1 point to Member A, 2 points to Member B, 2 points to Member D and 5 points to Member F. Your cost of assigning deduction points will be 10 ECUs. The corresponding payoff deductions will be 3 ECUs from Member A, 6 ECUs from Member B, 6 ECUs from member D and 15 ECUs from member F.

All deductions from the earnings after the group comparison stage will be determined as a sum of assigned and received deductions points from the current round. There is only one exception to this rule. Should the cost of **received** deduction points exceed the individual earnings after the group comparison, earnings will be reduced to zero. Nevertheless, participant has always to incur the costs of all deductions points she **assigns.**

Your total income from each round will be calculated as follows.

Should the earnings after the group comparison be equal or higher than the effect of received deductions points,

Total income =

+ earnings after the group comparison

- sum of received deduction points * 3

- sum of deduction points you have assigned

Should the earnings after the group comparison be less than the effect of received deduction points,

Total income =

0 – sum of deduction points you have assigned.

After all participants have made their decisions, the number of ECUs you earned in the corresponding round will displayed to you and stored in the computer. Your earnings from the earlier rounds cannot be used in the following rounds. You will receive a new endowment of 20 ECUs in the beginning of each round.
