

# Leadership in Altruistic Punishment and Reward

Simon Columbus

Amsterdam University College

University of Amsterdam, Graduate School of Psychology

simon@simoncolumbus.com -- <http://simoncolumbus.com>

#110

Amsterdam University College  
Excellence and Diversity in a Global City



UNIVERSITEIT VAN AMSTERDAM

VU  
VRIJE  
UNIVERSITEIT  
AMSTERDAM

## Abstract

Cooperation in economic games breaks down in the absence of enforcement mechanisms. We show that in the Public Goods Game, cooperation can be sustained by altruistic punishment, but not reward. Voluntary leadership occurs frequently in both conditions, but does not affect the dynamics of contributions and enforcement. While pro-socially punishing leaders are perceived as fairer than anti-social leaders, they have a worse reputation than pro-socially rewarding leaders. Thus punishment appears to carry a reputational penalty even when it is pro-social. Contradicting predictions from indirect reciprocity theory, this provides some support for altruistic punishment as strong reciprocity.

## Introduction

The human species has an extraordinary ability to cooperate. Almost everything around us is the product of the interacting efforts of a myriad of individuals. Moreover, in contrast to other primates—which are confined to small-scale, highly related groups—humans live in large civilisations. Anonymous interactions are common in social and economic life. However, cooperation with strangers depends on the enforcement of specific social norms. Where defectors go unconstrained and free-riders can exploit public goods, cooperation breaks down.

Voluntary punishment and reward can sustain cooperation even when they are individually costly to the enforcer, under the condition that they are directed at free-riders. Such *altruistic punishment* and reward occurs across a wide variety of cultures. Altruistic punishment and reward are thus largely established as effective enforcement mechanisms that can sustain cooperation.

Leadership is a second type of mechanism that can enforce cooperation. Two broad theoretical frameworks have been proposed for the evolution of leadership. The Byproduct Dominance hypothesis argues that leadership and followership are “byproducts of adaptations for dominance and submission” (van Vugt, 2006). In contrast, the Social Coordination hypothesis proposes that “leadership evolved specifically for the purpose of solving coordination problems” (ibid.). In this perspective, leadership, like punishment, would be an altruistic trait.

We test the effects of leadership in a Public Goods Game (PGG) with sequential punishment or reward. These games allow for individuals to take the charge in enforcing norms of cooperation by punishing free-riders or rewarding contributors. Policing of free-riders can be understood as a second-order collective action problem; however, leaders’ options are much more ambiguous in this setting than in coordination games, where group and individual interests align. Furthermore, we consider differences in the effectiveness of punishment and reward as enforcement mechanisms. Finally, using hypothetical scenarios we test the effects of punishment and reward on the reputation of leaders.

## Materials and Methods

96 subjects (62 female) participated in a PGG with real monetary stakes and one of two treatment conditions, Punishment or Reward. In both conditions, groups of  $n = 4$  members played six rounds of the PGG. Identifiers for the players were randomised at the beginning of each round such that reputation building over multiple rounds was not possible. At the beginning of each round, each of the  $n$  players in a group received an endowment of  $y = 20$  tokens. During the Contribution stage, each player could decide whether to make an investment  $g_i$  of part or all of their endowment in a common project ( $0 \leq g_i \leq 20$ ). These contributions were made simultaneously. The total investment in the common project was then multiplied by a factor  $na = 1.6$  and paid out in equal parts to all players.

At the end of the Contribution stage, the full matrix of contributions was revealed to all players. During the following Enforcement round, each player could expend part of their endowment to punish or reward another player. The ratio was set such that player  $i$  had to spend  $p_i = 1$  token to impose a punishment of  $P^j = 3$  tokens on any player  $j$ , and similarly for reward.

The Enforcement round was implemented sequentially, following the design for sequential games from Gillet et al. (Gillet, Cartwright, & van Vugt, 2011). At the start of the stage, there was a three-minute window during which the first player to move and make their punishment allocation revealed their choices to the other players. This window closed as soon as the first player made their move, and all other players then made their decisions simultaneously. If no player chose to make their allocation during the sequential punishment window, the stage would move on to the simultaneous phase. After all players had made their allocation, each player was informed about the punishment they received and their total pay-off from that round.

The total pay-off  $\Pi$  for player  $i$  from the PGG was their average pay-off per round  $t$  [i.e.  $\Pi_i = (\sum_{t=1}^6 \pi_i^t)/6$ ]. Tokens were exchanged for Euros at the end of the experiment at a rate of 3.2 : 1, such that in the Punishment condition the earnings per player in a round with full cooperation and no punishment equal €10. For the Reward condition, tokens were exchanged for Euros at a rate of 12.2 : 1. At this exchange rate, the earnings per player in a round with full cooperation and full reward to the player equal €10. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

## Leader’s Reputation

Since reputation-building was not possible in the Public Goods Game due to the implementation of anonymity, the effects of punishment and reward on leaders’ reputation were assessed via four hypothetical scenarios. The participants were confronted with prompts describing Pro-social (rewarding high contributors resp. punishing low contributors) and Anti-social (rewarding low contributors resp. punishing high contributors) leaders (Figure 1). For each hypothetical scenario, the participants were asked to rate the leader on four dimensions, ‘Dominance’, ‘Altruism’, ‘Being a Good Person’, and ‘Fairness’. Each prompt was to be answered on a 5-option Likert scale (1 = completely agree to 5 = completely disagree).

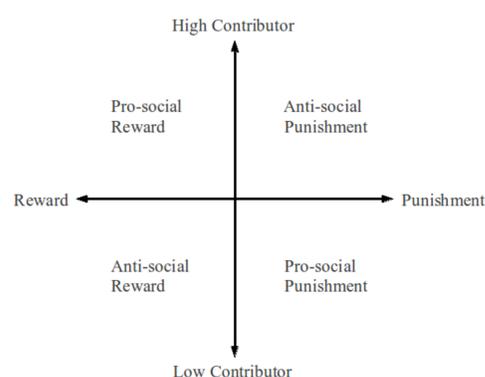


Figure 1: Leadership styles in Punishment and Reward on the pro-/anti-social dimension.

## Results

Punishment and Reward conditions differed in total contributions and earnings, but not in the amount of points spent on enforcement—either by leaders or by all group members—nor in the distribution of leadership ( $\alpha = .05$ ). Further analysis shows that the conditions initially did not differ significantly in contribution levels (independent-samples Mann-Whitney test,  $U = 88.5, p = .65$ , median reward = 11, median punishment = 14 at group level; and  $U = 1,150, p = .145$ , median reward = 12, median punishment = 14 at individual level). Figure 2 illustrates that average contributions to the common good increased in the Punishment condition, but fell off in the Reward condition; i.e., the trends of both conditions diverged over multiple iterations.

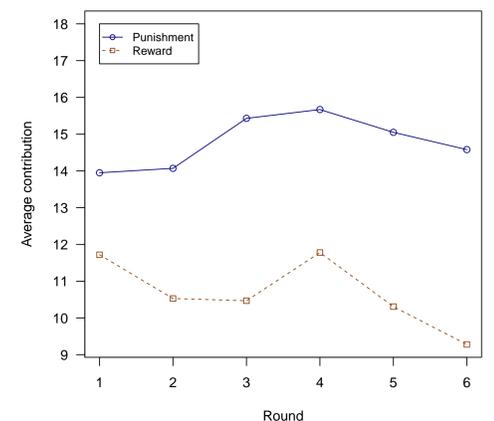


Figure 2: Changes over rounds in the average contribution to the public good.

First-moving occurred during each round of the experiment, and always after only a fraction of the waiting period. Most players moved first at least once (Punishment = 78%; Reward = 78%). The median number of rounds any player chose to move first was one, and monopolisation of leadership by any one player was rare. Thus in global terms, leadership did not differ across conditions. Moreover, expenditure on enforcement did not differ significantly between leaders and non-leaders (Wilcoxon signed-rank test,  $\alpha = .05$ ).

## Reputational Consequences of Pro-social and Anti-social Leadership

Friedman tests for reputation differences between leadership styles were significant for all four variables (all  $p < .001$ ). Pro-social Punishers and Rewarders exhibited no difference on Perceived Fairness ( $p = 1.000$ ), and scored significantly higher than anti-social leaders ( $p < .001$ ). Among the latter, Punishers and Rewarders did not differ significantly ( $p = 1.000$ ).

Pro-social Rewarders were considered more altruistic than their punishing counterparts (median (both) = 3,  $\bar{x} = 2.46$  vs.  $3.07, p = .020$ ). The latter were perceived as significantly more altruistic than Anti-social Punishers (median 3 vs. 4,  $p = .005$ ), but not Anti-social Rewarders (median (both) = 3,  $p = 1.000$ ). A similar pattern can be observed on the perception of being a good person. Pro-social Rewarders scored higher than Punishers (median (both) = 3,  $\bar{x} = 2.34$  vs.  $2.73, p = .048$ ), and the latter were again perceived as more of a good person than Anti-social Punishers (median = 3 vs. 4,  $p < .001$ ), but not Rewarders (median (both) = 3,  $p = .122$ ). Moreover, anti-social Punishers were considered less of a good person than their Rewarding counterparts (median = 4 vs. 3,  $p = .006$ ).

## Conclusions

- Altruistic punishment, but not reward, can sustain cooperation among an anonymous group;
- Leaders do not differ significantly from followers both in terms of personality and of behaviour and leadership does not appear to impact group dynamics;
- Who becomes a leader cannot be predicted by measures of altruism, dominance, Machiavellianism, and sensation-seeking;
- When leaders punish in the interest of the group, they are perceived as more fair than anti-social leaders, but this does not translate into reputational gains in terms of being considered altruistic or being seen as a good person.

## Forthcoming Research

A multi-level analysis of the dynamics involving contributions, leadership, and punishment is currently in progress.

## References

- Fischbacher, U. (2007, February). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–178. Retrieved from <http://www.springerlink.com/index/10.1007/s10683-006-9159-4> doi: 10.1007/s10683-006-9159-4
- Gillet, J., Cartwright, E., & van Vugt, M. (2011, August). Selfish or servant leadership? Evolutionary predictions on leadership personalities in coordination games. *Personality and Individual Differences*, 51(3), 231–236. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0191886910002916> doi: 10.1016/j.paid.2010.06.003
- van Vugt, M. (2006). Evolutionary Origins of Leadership and Followership. *Personality and Social Psychology Review*, 10(4), 354–71. doi: 10.1207/s15327957pspr1004

## Acknowledgements

I thank Mark van Vugt and Thomas Pollet for their support and guidance, as well as Sennay Ghebreab for providing me with access to the B&TA Lab. Further thanks go to Fayette Klaassen, Luc Draisma, Martine van Dusschoten, Lukas Snoek, Jenny Rutten, Bente de Vries, Henk Nieweg, Alina Berendsen, Agnese Logina, Francesca Grandolfo, and Joris Gillet for their help at various stages of this study.

This study was sponsored by the Beta Beurs. The Beta Beurs is devised by the Brain & Technology Amsterdam (B&TA) Lab, co-funded by the Center for Creation, Content and Technology (CCCT), and subsidised by Platform Beta Techniek (PBT).