

A TRACE OF ANGER IS ENOUGH: ON THE ENFORCEMENT OF SOCIAL NORMS

Jakub Steiner

CERGE-EI

Charles University
Center for Economic Research and Graduate Education
Academy of Sciences of the Czech Republic
Economics Institute

Working Paper Series
(ISSN 1211-3298)

246

**A Trace of Anger is Enough:
on the Enforcement of Social Norms**

Jakub Steiner

CERGE-EI
Prague, March 2005

ISBN 80-7343-039-8 (Univerzita Karlova v Praze, CERGE)
ISBN 80-7344-028-8 (Národohospodářský ústav AV ČR, Praha)

A Trace of Anger is Enough: On the Enforcement of Social Norms*

Jakub Steiner[†]
CERGE-EI

January 17, 2005

Abstract

It is well documented that the possibility of punishing free-riders increases contributions in one-shot public good games. I demonstrate theoretically that minimal punishment commitments (perhaps provided by anger) may lead to high contribution levels. Thus, almost selfish players may behave as strong reciprocators.

Je dobře zdokumentováno, že možnost trestat příživníky zvyšuje příspěvky v jednorázové hře s obecním majetkem. Ukazují teoreticky, že nepatrný závazek trestat příživníky (snad způsoben hněvem) může vést k vysokým příspěvkům. Téměř sobečtí hráči se tak mohou chovat jako silní reciprokátoři.

JEL classification: D64, H41, Z13.

Keywords: Reciprocity, Emotions, Commitment, Punishment, Public Good.

*I benefited from the comments of James Andreoni, Jeffrey Carpenter, and Simon Gächter. Dirk Engelmann, Andreas Ortmann, and Avner Shaked inspired me in numerous discussions. The work was supported by a CERGE-EI / World Bank fellowship.

[†]Center for Economic Research and Graduate Education, Charles University, and Economics Institute, Academy of Sciences of the Czech Republic (CERGE-EI); Address: Politických Věznu 7, 111 21 Prague, Czech Republic, Tel: +420-605-286-947; E-mail: jakub.steiner@cerge-ei.cz.

1 Introduction

Experimental evidence suggests that people substantially increase their contributions to the public good when the possibility of punishing free-riders is introduced (e.g. Fehr and Gächter, 2000a, 2002; Yamagishi, 1986). Since costly punishment is not subgame perfect, this evidence has been interpreted as another blow to the concept of *homo oeconomicus* and as evidence supporting strong reciprocity theory (e.g. Gintis, 2000). Indeed, some deviation from the *homo oeconomicus* assumptions is needed to explain the (strong) consequences of the possibility to punish. This deviation is typically interpreted as emotions, particularly anger, in the reciprocity literature (e.g. Fehr and Gächter, 2002). Below, I theoretically demonstrate that only traces of anger are needed to explain the well-documented high contribution levels; the deviation from *homo oeconomicus* thus needs to be rather small. Consequently public good game a with punishment option is an efficient institution even if pecuniary stakes are quite high and emotional payoffs are thus relatively small.

The model described below under study reflects the standard experimental design of public good games a with punishment option: N players have the possibility to contribute to a linear public good. All players are thereafter given the opportunity to allocate punishment points to other players. Punishment is costly for both sides.

I assume that the punishment activity is motivated by emotions such as anger. People may punish a bit because of, for example, warm glow (e.g. Andreoni, 1990), altruism (e.g. Andreoni and Miller, 2002), inequality aversion (e.g. Bolton and Ockenfels, 2000; Fehr and

Schmidt, 1999), or reciprocity (e.g. Levine, 1998). The role of emotions as a commitment device was stressed by Frank (1988). There is indeed neuroscientific evidence suggesting that emotions such as anger are triggered by unfair behavior and that people weigh the satisfaction from punishing norm violators with the cost of such action (de Quervain et al., 2004; Sanfey et al., 2003). It is, however, not the purpose of this paper to explain where emotions come from. The focus is on the intensity of emotion that it takes to enable the players to cooperate, where the intensity is measured by the amount of money the players spend on punishment. I choose a black box approach, and I model punishment as an automatic reaction fully governed by a punishment rule, which is a function of the realized contributions.

Below I construct a punishment rule which induces a game with *unique* equilibrium in which all players make high contributions. The rule requires every player to punish only the player with the lowest contribution which in turn motivates her to escape the lowest position. The game with such a punishment rule can be solved by iterated elimination of dominated strategies because the lowest non-eliminated contribution level is always dominated by the level just above it. At the same time, it is possible to implement such a punishment rule at minimal cost because resources from a large group of players are focused on one free-rider. This rule resembles the rule experimentally implemented by Yamagishi (1986). I do not argue that this exact punishment rule is necessarily used in reality; I simply show that a rule inducing high contributions exists.

2 The Model

Each player $i \in \{1, \dots, N\}$ chooses a contribution level c^i from the common strategy set $S = \{\frac{0}{L}\bar{c}, \frac{1}{L}\bar{c}, \dots, \frac{L}{L}\bar{c}\}$, where L represents the number of levels approximating the continuous interval $[0, \bar{c}]$, and L is assumed to be large. The maximal possible contribution is \bar{c} . After the contributions $c = (c^1, \dots, c^N)$ of all players are realized and observed by everyone, the players automatically assign punishment points to each other. I abstract from the individual punishment actions and analyze only the sum of punishments from all players, which is denoted as the aggregate punishment $p^i(c) \geq 0$ of player i . The unspecified punishment rules of individual players may differ in reality (e.g. Fehr and Gächter, 2000b). Then the aggregate punishment rule $p^i(\cdot)$ can be interpreted as the expected punishment of player i given the population the players are drawn from.

Although the game has two time phases, it can be seen conceptually as a one-stage game as the players make decisions only in the contribution phase. The punishment phase can be modeled within the payoff function of the one-stage game.

Definition 2.1 *A game of N players with strategy sets $\{S^i = S\}_{i=1}^N$ and the payoff function¹*

$$U^i(c) = -c^i - p^i(c) \tag{1}$$

is the punishment game.

¹The payoffs from the public good are the same for all players, therefore the public good do not have to be included in the payoff function. The small cost of punishment is also not included because punishment is modeled as an automatic action, not as a result of payoff optimization.

I assume that the average player is able to commit to spend only one penny to punish free-riders:²

$$\mathbf{A1:} \quad \frac{\sum_{i=1}^N p^i}{N} \leq 1.$$

The set of equilibria of the punishment game depends on the particular punishment rule. With the assumption **A1** being valid, one might expect an unattractive equilibrium in which all players free-ride and punishment is ineffectively spread among all the players resulting in no player having an incentive to contribute. However, the punishment rule proposed below induces a game with a unique equilibrium in which all participants make high contributions.

Denote the lowest contribution among players by c_l and the second lowest by c_{sl} where $c_l = c_{sl}$ if there is more than one player with the lowest contribution. Let the punishment rule be

$$p^i(c) = \begin{cases} 0 & \text{if } c^i > c_l, \\ \frac{N}{\bar{c} + \frac{\bar{c}}{L}} (c_{sl} + \frac{\bar{c}}{L} - c^i) & \text{if } c^i = c_l. \end{cases} \quad (2)$$

This punishment rule motivates the player(s) with the lowest contribution to increase her (their) contribution(s) one level above the second lowest contribution. The marginal punishment $\frac{N}{\bar{c} + \frac{\bar{c}}{L}}$ is chosen such that the total punishment expenditures are always at most one unit per player.

Theorem 2.1 1. *The punishment game of N players with punishment rule (2), with the maximal possible contribution satisfying $\bar{c} + \frac{\bar{c}}{L} < N$, and with a large number of contribution levels $L > N$ has a unique*

²Indeed, the average expenditure on punishment was 0.69 points out of 10 available points in the experiment of Fehr and Gächter (2002), (Gächter, email communication, September 12th, 2004).

equilibrium with all players contributing the maximal possible amount

$$\{c^{*i} = \bar{c}\}_{i=1}^N.$$

2. The punishment rule (2) satisfies the low cost assumption **A1**.

PROOF:

1. The player with the lowest contribution always wishes to increase her contribution by at least $\frac{\bar{c}}{L}$ because the marginal punishment she experiences $\frac{N}{\bar{c} + \frac{\bar{c}}{L}} > 1$ is higher than the marginal cost of contribution. Hence, the lowest contribution level, 0, is dominated by $\frac{\bar{c}}{L}$. After the elimination of $\{0, \frac{1}{L}\bar{c}, \dots, \frac{k}{L}\bar{c}\}$, $\frac{k+1}{L}\bar{c}$ is dominated by $\frac{k+2}{L}\bar{c}$ because $\frac{k+1}{L}\bar{c}$ would be the lowest contribution among non-eliminated strategies for $k = 1, \dots, L-2$. Thus, the game can be solved by the iterated elimination of dominated strategies; only \bar{c} survives this process.

2. There is either only one player with the lowest contribution, and then she is the only one being punished. The punishment is largest in this case if $c_{sl} = \bar{c}$ and $c_l = 0$. Then the punishment is $\frac{N}{\bar{c} + \frac{\bar{c}}{L}}(\bar{c} + \frac{\bar{c}}{L} - 0) = N$, so the cost is at most 1 unit per player.

Or there may be many players with the lowest contribution, but then $c_{sl} = c_l$, so each punishment is $\frac{N}{\bar{c} + \frac{\bar{c}}{L}} \frac{\bar{c}}{L} \leq 1$, thus the cost per player is smaller than one unit. Q.E.D.

3 Conclusions

The possibility to punish free-riders acts like a magnifying glass on the players' emotions. Although they give up only a penny voluntarily, they are able to induce each other to contribute N pennies. Thus, the cooperation level increases linearly with the number of players. This

insight is confirmed by Carpenter (2004) who experimentally found a positive group size effect even if one controls for marginal group return of contributions.

The cheapness of the punishment has implications for the evolutionary explanations of strong reciprocity. Clearly the evolutionary forces outweighing the cost of punishment need to be relatively weak. For example, Gintis (2000) assumes that the strong reciprocators must be able to punish all free-riders simultaneously, whereas here it is demonstrated that it is enough to be able to punish only one free-rider. Thus the reciprocator can survive even if the technology of punishment is not very effective.

The punishment rule proposed here has a normative appeal. If a group of subjects wants to overcome the free-riding problem but has only a limited possibility to commit to a punishment threat, they should choose this rule.

Sometimes the public good will be too expensive to be provided through the punishment game discussed above. However, the game can be extended in a straightforward manner to a second punishment stage where those who did not participate enough on punishing free-riders could be punished. Then the players could enforce N^2 contribution levels. Of course, the punishment game requires quite a bit of information: All players have to be able to monitor the actions of all other players. This is realistically possible only in small groups, such as workplaces.

The model proposed here demonstrates that punishment games may be very efficient institutions even for high pecuniary stakes. It

also suggests that the experimental results on the punishment game do not have to be interpreted as a clear refutation of the selfishness assumption. Only minimal deviation from the assumption is needed to explain cooperation in the punishment game. Strong reciprocity is strong in its consequences, not in the emotional requirements of players.

References

- [1] Andreoni J., 1990, Impure Altruism and Donation to Public Goods: A Theory of Warm-Glow Giving?, *The Economic Journal* 100, 464–477.
- [2] Andreoni J. and J.H. Miller, 2002, Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism, *Econometrica* 70, 737–753.
- [3] Bolton G. and A. Ockenfels, 2000, ERC: A Theory of Equity, Reciprocity, and Competition, *American Economic Review* 90, 166–193.
- [4] Carpenter J., 2004, Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods, *Games and Economic Behavior*, forthcoming.
- [5] Fehr E. and S. Gächter, 2000a, Cooperation and Punishment in Public Goods Experiment, *American Economic Review* 90, 980–994.
- [6] Fehr E. and S. Gächter, 2000b, Fairness and Retaliation: The Economics of Reciprocity, *Journal of Economic Perspectives* 14, 159–181.
- [7] Fehr E. and S. Gächter, 2002, Altruistic Punishment in Humans, *Nature* 415, 137–140.
- [8] Fehr E. and K.M. Schmidt, 1999, A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics* 114, 817–868.

- [9] Frank R.H., 1988, *Passions within Reasons: The Strategic Role of the Emotions*, (Norton, New York).
- [10] Gintis H., 2000, Strong Reciprocity and Human Sociality, *Journal of Theoretical Biology* 206, 169–179.
- [11] Levine D.K., 1998, Modeling Altruism and Spitefulness in Experiments, *Review of Economic Dynamics* 1, 593–622.
- [12] 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.
- [13] Sanfey A. G., Rilling J. K., Aronson J. A., Nystrom L. E., Cohen J. D, 2003, The Neural Basis of Economic Decision-Making in the Ultimatum Game, *Science* 300, 1755–1758.
- [14] Yamagishi T., 1986, The Provision of a Sanctioning System as a Public Good, *Journal of Personality and Social Psychology* 51, 110–116.

Individual researchers, as well as the on-line and printed version of the CERGE-EI Working Papers Series (including their dissemination) were supported from the following institutional grants:

- Economic Aspects of EU and EMU Entry [Ekonomické aspekty vstupu do Evropské unie a Evropské měnové unie], No. AVOZ70850503, (2005-2010);
- Economic Impact of European Integration on the Czech Republic [Ekonomické dopady evropské integrace na ČR], No. MSM0021620846, (2005-2011);

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

(c) Jakub Steiner, 2005

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise without the prior permission of the publisher.

Published by

Charles University in Prague, Center for Economic Research and Graduate Education (CERGE)
and

Economics Institute (EI), Academy of Sciences of the Czech Republic

CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.

Printed by CERGE-EI, Prague

Subscription: CERGE-EI homepage: <http://www.cerge-ei.cz>

Editors: Directors of CERGE and EI

Managing editors: Deputy Directors for Research of CERGE and EI

ISSN 1211-3298

ISBN 80-7343-039-8 (Univerzita Karlova v Praze, CERGE)

ISBN 80-7344-028-8 (Národohospodářský ústav AV ČR, Praha)



CERGE-EI
P.O.BOX 882
Politických vězňů 7
111 21 Praha 1
Czech Republic
<http://www.cerge-ei.cz>