Legitimate Punishment, Immunity, and the Enforcement of Cooperation

Marco Faillo [a] Daniela Grieco[b] Luca Zarri[c]

(Very preliminary and incomplete version. Please do not quote or circulate)

Abstract

In the framework of a public goods game with costly punishment options, we introduce a novel restrictive setup where only virtuous behavior (that is, being a high contributor) allows one to gain access to punishment options ('legitimacy'). Furthermore, acting virtuously guarantees that it will not be possible to be punished by others ('immunity'). These restrictions rule out punishment driven by targeted or blind revenge (such as the forms of counter-punishment investigated in Nikiforakis, 2008 and in Denant-Boemont, 2007) and censor individual attempts to punish virtuous actions (that is perverse and anti-social punishment; Cyniabuguma et al., 2006; Herrmann et al., 2008). In our experiment, pro-social behaviours are defined in two possible ways: a subject is classified as virtuous if her contribution is higher than the average contribution in her group (first treatment or Average-dependent Legitimacy treatment), or if her contribution is higher than her peers' one (second treatment or Peer-dependent Legitimacy treatment). Our preliminary results show that groups' behavioral patterns exhibit a strong path dependence, as the group's average contribution level in the first period plays a crucial role in the determination of the final outcome.

JEL Classification: C72; C91; C92; D23 ; D72.

Keywords: Experimental Economics; Public Good Games; Costly Punishment; Cooperation; Legitimacy; Immunity.

[a] marco.faillo@unitn.it, Department of Economics, University of Trento.

[b] daniela.grieco@univr.it, Department of Economics, University of Verona.

[c] luca.zarri@univr.it, Department of Economics, University of Verona.

1. Introduction

In a public goods game (*PGG*) or voluntary contribution mechanism (*VCM*) framework, there is a group of subjects who, as the game starts, receive an individual monetary endowment, from which they may contribute any amount to a public good that returns a payoff to each of them. The structure of monetary payoffs in the *PGG* makes it a classical 'social dilemma', as each agent has a dominant strategy to free ride, while, in contrast, at the social optimum each individual contributes his entire endowment. Therefore, the straightforward, standard prediction based on the canonical model of *Homo Oeconomicus* is that everyone should free ride, both in the one-shot and in the repeated *PGG*. However, in the repeated version of the game, the following pattern typically occurs: initially, average contributions are relatively high, whereas, as the game unfolds, they gradually decline and cooperation converges to a near-negligible level (Ledyard, 1995).

In the last years, an increasing number of *VCM* experiments have been investigating the role that *institutions* can play in the enforcement of cooperation. While a strand of experimental research deals with *endogenously* formed institutions (see e.g. Gurerk et al., 2006 and Kosfeld et al., 2009), a second strand encompasses *exogenously* imposed institutions. Within the latter research area, some studies focused on *centralized* mechanisms (see Chen and Plott, 1996; Falkinger et al., 2000; Andreoni, 1993; and Chan et al., 2002), while others explored *decentralized* institutions (Ostrom et al., 1992; Fehr and Gächter, 2000; 2002). Fehr and Gächter (2000; 2002) demonstrate that while in non-punishment treatments (*PGG* without punishment opportunities) cooperation rates indeed tend to fall over time (round after round), this 'decay phenomenon' does not occur insofar as players, by having access to so called 'costly' or 'altruistic' punishment, are allowed to incur a cost to decrease others' monetary payoffs (*PGG* with punishment opportunities)[1]. The presence of punishment opportunities turned out to make the difference and made cooperation sustainable over time[2]. The emergence of voluntary, non-strategic[3] and costly punishment in the laboratory is a puzzle for standard economic theory, as it is in contrast with the so called 'selfishness axiom', that is the idea that subjects act selfishly in order to systematically maximize their monetary gains. In other words, in a finitely repeated *PGG* with punishment options, standard theory predicts that subjects *will not* use such options, due to the net monetary costs associated with their usage[4]. By contrast, despite the

[1] This interesting result (confirmed also by the recent experiment run by Anderson and Putterman, 2006) is in sharp contrast with Binmore's and other non-behavioral scholars' learning argument (known also as the 'discovered preference' hypothesis), according to which players diminish their rate of pro-sociality over time as they gradually *learn* better and better how the game works, so that at the end of the game they are more rational and act selfishly.

[2] Analogously, the introduction of explicit punishment and/or rewarding opportunities significantly affects subjects' behavioral choices in the experimental games studied by Fehr et al. (1997) and Fehr and Rockenbach (2003).

[3] Punishment is 'non-strategic' when interactions are one-shot and involve strangers, so that reputation building motives can be ruled out.

[4] Sethi and Somanathan (1996) observe, on the basis of the case studies cited in their work, that punishments such as social disapproval and physical damage are costly not only for the punishee, but also for the punisher.

seemingly irrational nature of this behavioral attitude, peer punishment of free riders turned out to be a widespread phenomenon both in the field and in the lab. Experimentally, it has been shown to represent a powerful decentralized enforcement device, through which it is possible to induce and successfully sustain cooperation in social dilemmas.

Like these studies, in this paper we focus on a decentralized mechanism and wonder whether the presence in a *VCM* framework of suitably restricted punishment opportunities leads to an increase in the level of aggregate cooperation in a repeated game setting. The peculiarity of our contribution derives from the fact that we set up an experimental design where several restrictions are exogenously imposed with regard to both *who* is allowed to punish and *whom* can be punished[5].

Legitimacy

The first distinctive feature of the design is legitimacy and it holds at the punishment stage (stage 2): at stage 2, subjects are allowed to punish only if their contribution at stage 1 satisfies a condition which varies across treatments. In other words, only 'legitimate punishment' is allowed: at stage 1, subjects need offer a contribution which satisfy a particular condition in order to gain access to the punishment option in the following stage. In the first treatment (*Average-dependent Legitimacy treatment*) they need to contribute above the average of their group, in the second treatment (*Peer-dependent Legitimacy treatment*) their contribution must be higher than the contribution of at least one of the other members of their group. This implies that 'antisocial punishment' (that is, low contributors punishing high contributors; see on this Herrmann et al., 2008) is ruled out. We claim that it is plausible that, within advanced societies, in real-life institutions peer punishment is allowed only insofar as potential punishers are entitled to do this due to their previously virtuous behavior. Therefore, one of the main goals of our restrictive design is to see whether the distinction between 'legitimate' vs 'non-legitimate' peer punishment matters: this feature differentiates us from classic (Fehr and Gächter, 2000; 2002) as well as recent *VCM* studies (Denant-Boemont et al., 2007), where only unrestricted punishment is considered. A recent experiment with a somewhat similar structural feature is Kosfeld et al. (2009), where subjects who do not contribute the efficient amount to the public good are sanctioned. However, unlike our paper, they investigate the linear public good game within a process of endogenous institution formation, by focusing, both theoretically and experimentally, on its impact on cooperation rates and group welfare.

---

[5] Therefore, our design also differs from recent experimental *VCM* protocols where norms prescribing who can punish and/or whom can be punished emerge endogenously within a group (see e.g. Casari and Luini, 2009; Kosfeld et al., 2009).

Immunity

Another important feature of our design is *immunity*: subjects whose contribution satisfy a specific condition cannot be sanctioned. In particular, in the *Average-dependent Legitimacy treatment* immunity is guaranteed to subjects who contributes above the average, in the *Peer-dependent Legitimacy treatment* only subjects whose contribution is not smaller than any other subject's contribution are immune. As a consequence, virtuous behavior (i.e. satisfying the specific condition) automatically allows one to be immune from punishment actions by those who adopt a non-virtuous behaviour. This implies that our design rules out 'perverse first-order punishment' (that is punishment of high contributors, Cyniabuguma et al., 2006). The lack of any restriction on the target of punishment is a typical feature of experiments focusing on first-order punishment (such as, e.g., Fehr and Gächter, 2000; 2002; Masclet et al., 2003; Carpenter, 2005)[6].

2. Experimental setup

In this study, we focus on punishment in a *VCM* framework by means of a restrictive experimental setup having the following two key features: (i) only subjects who act virtuously at stage 1 (the contribution stage) are entitled to get access to punishment options in the subsequent stage; further, (ii) only low contributors are potential punishees at stage 2 (the punishment stage). As we observed in the previous section, these restrictions rule out punishment behavior driven by targeted or blind revenge (that is the forms of counter-punishment investigated in Nikiforakis, 2008 and in Denant-Boemont, 2007) and censor individual attempts to punish socially virtuous actions (Cyniabuguma et al., 2006; Herrmann et al., 2008).

Procedure

The experiment consists of 4 sessions of 20 subjects each, for a total of 80 participants. The first two sessions were conducted at the CEEL Lab of the University of Trento (Italy) in December

[6] In these contributions, it is the case that, as Denant-Boemont et al. (2007) point out, "Because there is only one opportunity to sanction in each period, and there is no means to track the identity of others from period to period, no player can identify individual punishment behavior in a manner that allows him to target an individual for reciprocation" (pp. 146-147). However, it is important to make clear that, in those studies, while immunity of sanctioners *from reprisals* exists, punishers are *not* immune from sanctions *altogether*, as nothing prevents other subjects from sanctioning them. In particular, we might say that in those studies while punishers, unlike in Nikiforakis (2008) are immune from 'targeted revenge' on the part of the punishees, they are not immune from so called 'blind revenge'.

2009, the other sessions are planned for April 2010. The experiment was programmed by using the z-tree platform (Fischbacher, 1999).
The subjects were undergraduate students mainly from Economics. No individual participated in more than one session.

In each session, there are 20 periods of interaction that proceed under identical rules. The participants in a session were assigned to groups of size four with fixed membership, but they did not know the identities of the other members of their group. There were 10 groups, and thus 10 independent observations, for each treatment. At the end of each period, individuals remained in the same group: however, individuals' labels were reassigned on a random basis in each period. For example, the same player could be designated as player *45* in period *t*, as player *6* in period $t + 1$, and as player *38* in period $t + 1$. Like many other studies (e.g. Cinyabuguma et al., 2006; Demant-Boemont et al., 2007), we use a *partner protocol* that keeps the composition of each group constant, but is also characterized by anonymity of the components of the group and change of participants' labels across rounds[7]. The design and the parametric structure of the experiment are based on those of Fehr and Gächter (2000).

Treatments and conditions

The experiment consists of 2 treatments, both made of two stages: the Average-dependent Legitimacy (AL) treatment and the Peer-dependent Legitimacy (PL) treatment. For each treatment, in each session there are $I=20$ subjects divided in $M=5$ groups of $N=4$ (as in standard *VCM* experiments) subjects, who play a finitely repeated public goods game for $T=20$ periods. Participants are aware of the number of rounds they play and of the number of stages: information on the following stages allows to evaluate the effect of the threat of being punished in Stage 2 and on contribution decisions in Stage 1.

I. Average-dependent Legitimacy treatment

This treatment is based on the *VCM* with punishment options (Fehr and Gächter, 2000). In stage 1, at the beginning of each period each participant receives a fixed amount $e = 20$ of tokens. Each participant *i* has to decide whether she wants to invest into a public project or not an amount $g_i \leq e$. Decisions are made simultaneously and with no information about peers' choices. At the end of stage 1, each participant is informed about her current earnings, which consist of two elements:

[7] Although a stranger protocol with random re-matching allows ruling out strategic punishment and reputation motives altogether, a partner protocol has been demonstrated to work as well as a stranger protocol. Nikiforakis (2008), based on Botelho (2004), addresses this issue by comparing results from a stranger protocol to the ones of a partner protocol and finds that differences in punishment decisions are not significant (whereas differences in punishment levels are).

a. The amount of her initial 20 tokens that she has kept for herself (i.e. 20 tokens – Her Contribution to the project);

b. Her income from the project. The income to her is equal to 40% of the total of the four individual contributions to the project.

Therefore, her earnings at the end of stage 1 are calculated by the computer in the following way:

Each participant's earnings after stage 1 = (20 – her contribution to the project)

+ 40% * (total group contribution to the project)

According to the value of g_i, compared to the average level of contributions in the group $\bar{g}$, participants are classified as Contributors (C, if $g_i \geq \bar{g}$) or Deviators (D, if $g_i < \bar{g}$). Participants know that they can go on with stage 2 in the experiment only if they behave as Contributors, that is, as we explained in the previous section, only if they have the legitimacy to do so[8].

In stage 2, Contributors (C) *only* are given the opportunity to simultaneously punish Deviators (D) by assigning a certain amount of points. We restrict the set of possible punishers at C-types, and the set of possible punishees at D-types. This choice allows us to rule out forms of perverse and antisocial punishment and to concentrate on virtuous pro-social behavior only[9]. Therefore, our types at stage 2 will be either Contributors who punish (C, P) or Contributors who do abstain from punishing (C, N). All non-contributors will be (D, N) players. In this case, only Contributors receive information about peers' behaviours, i.e. about the average level of contribution in the group and the specific contribution of each peer, classified as either an above or a below the average player. Contributors might decide to assign up to 10 points to each defector: points assignment is costly and costs are charged according to a standard cost function as in Fehr and Gächter's (2002) table:

[8] It is important to make clear that we never used loaded terms such as 'contributors', 'deviators', 'legitimacy' and 'immunity' during the experiment (see the Instructions, in the Appendix).

[9] On a similar vein, Falkinger et al. (2000) focus on a mechanism that rewards and sanctions players who contribute more and less, respectively, than the average to the public good.

Points	0	1	2	3	4	5	6	7	8	9	10
Cost	0	1	2	4	6	9	12	16	20	25	30

Each participant's earnings at the end of stage 2 are calculated by the computer in the following way:

Each participant's earnings after stage 2 = earnings at the end of stage 1

– cost of points she assigned at stage 2

II. Peer-dependent Legitimacy treatment

This treatment differs from the previous one in the condition that are imposed on the subjects' contributions in order to access legitimacy and immunity. With regard to the right to punish, subject *i* can be a punisher only if her contribution $g_i > g_j$ at least for one $j \neq i$. In other words, participants know that they can go on with stage 2 in the experiment only if their contribution is not the smallest of their group. With regard to immunity, only subjects whose $g_i, \geq g_j$ for every $j \neq i$ cannot be punished. In other words, a subject is immune only if there are not any higher contributions in her group.

3. Preliminary Results (AL treatment only)

Figure 1 illustrates the time path of individual contributions by period, averaged across groups, in the Average-dependent Legitimacy treatment.

[Figure 1]

The period number is shown on the horizontal axis and the average contribution on the vertical axis. The maximum possible individual contribution, corresponding to the group optimum, is 20. The minimum possible contribution is 0. Figure 1 shows that an intermediate and fairly stable average contribution level prevails over the 20 periods. However, by separately looking at each group's contribution behavior, we immediately see that this figure masks an interesting form of heterogeneity across groups: as Figure 2 makes clear, the AL treatment is characterized by a strongly path-dependent divergence in behavioral patterns across groups.

[Figure 2]

Groups can be classified in two categories, according to the path of average contribution and the level of contribution they make in the last periods. Groups 3, 4, 6, 9 and 10 implement virtuous behavior, generally do not free-ride and converge roughly to higher than average contributions (we call them ‘virtuous’). Groups 1, 2, 5, 7 and 8 exhibit an opposite behavior and collapse to lower levels of average contribution, from 3 to 0 (we call them ‘anti-social’). The overall average of 10 tokens represents the ‘critical threshold’ that separates the two major patterns observed in this treatment: with the only exception of group 7, groups that start below (resp., above) the threshold will end up with a low (resp., high) level of aggregate contributions. Figure 1 shows these two key patterns by averaging contribution levels within the two identified ‘macro-groups’, that is the virtuous and the anti-social macro-group.

Average-dependent legitimacy makes initial conditions crucial, leading to a strongly path-dependent pattern of group contributions: Figure 2 makes clear that two different groups may well end up with dramatically divergent average contribution levels even though their initial differences were small[10]. What ultimately matters in this treatment is the combination of such a formal feature of the setup (average-dependent legitimacy) and a substantial component such as groups’ first period average contribution levels: when the initial levels of group contributions are low (resp., high), free riding (resp., cooperation) prevails at the end of the game. Therefore, this form of legitimacy turns out to be a *double-edge sword* for the success of cooperation. Insofar as the groups initially reach high average contribution levels, average-based legitimacy is an institutional mechanism that strongly favors the emergence and sustainability of cooperation. Conversely, whenever a group’s average contribution levels are low from the outset, the institutional feature at work acts as a brake preventing cooperation from ‘taking off’. In other words, average-based legitimacy can be seen as a *morality-dependent institutional arrangemen*t: such a mechanism is conducive to enhanced cooperation in the long run only insofar as a group is autonomously able to reach relatively high contribution levels in the first place.

Let us further observe that we obtain the two key findings typical of the *PGG* (without and with punishment options) within a single treatment. Even though the setup does not change, we observe a strong polarization and, in particular, that half of the groups (the antisocial ones) behave

[10] See e.g. the dynamics of groups 2 and 9: even though their difference is initially small, average contribution levels are dramatically different in the last period. Such a path-dependent divergence is typically observed in evolutionary models. For example, in a dynamic regime characterized by a bistable dynamics, initial conditions play a crucial role: depending on whether a population is located below or above a certain critical threshold, completely different equilibria will emerge (see e.g. Sethi and Somanathan, 1996, where individuals will end up in either a cooperative or an individualistic society in the long run, depending on initial conditions).

as if punishment options were *not available*. They start from low contributions levels and starting from low contributions level in a *PGG* where punishment options are conditional on average-based legitimacy leads to the same qualitative outcome which typically occurs in a *PGG* where punishment options are not available. The occurrence of low initial contribution levels and the presence of average-based legitimacy generate the well-known 'decay phenomenon'. In contrast, the 'virtuous macro-group' – by displaying relatively high average contribution levels from the outset – dynamically behaves as if *unrestricted* punishment options were available (like in Fehr and Gächter, 2000; 2002).

The average contribution for each group in the AL treatment is shown in Table 1, with the standard deviations given in parentheses (to be used later on to compare Treatments).

[Table 1]

There is a significant difference in terms of average contributions across periods between the two macro-groups (virtuous vs. anti-social) (Median test: pearson χ^2 (1)=147; pr=0.000).

The role of path dependence in explaining the contribution levels is confirmed by the regression in table 2 where the variable "high_contr" is defined as a dummy that assumes value equal to 1 when a group has shown an average contribution higher than 10 in the first period (and 0 elsewhere). As in Fehr and Gächter (2000), the contribution level of each subject in each period can be investigated by estimating the following equation:

$$c_i^t = \beta_0 + \beta_1 high_{contr} + \beta_2 pun_{-i}^{t-1} + \beta_3 max\{0, c_i^t - \bar{c}_{-i}^t\} + \beta_4 max\{0, c_{-i}^t - \bar{c}_i^t\} + \beta_5 t + \beta_6\, gend + \beta_7 age + \beta_8 trust + \beta_8 happ$$

where pun_{-i}^{t-1} is the lagged variable representing sanctions received by the subject in the previous period, $c_{-i}^t - c_i^t$ is the difference between the average contribution of peers and own contribution (and viceversa).

[Table 2]

Table 3 shows the average quantity of sanctions assigned by individuals to the rest of the group in each period.

[Table 3]

As an extension of Denant-Boemont *et al*. (2007)'s equation, the number of sanctioning points of each subject in each period can be investigated by estimating the following equation:

$$pun\_given_i^t = \beta_0 + \beta_1 contribution + \beta_2 max\{0, \bar{c}_{-i}^t - c_i^t\} + \beta_3 max\{0, c_i^t - \bar{c}_{-i}^t\} + \beta_4 num_{def} + \beta_5 pun_{-i}^{t-1} + \beta_6 session + \beta_7 t + \beta_8 happ + \beta_9 trust + \beta_{10} gender + \beta_{11} age + \beta_{12} rel + \beta_{13} rel_{att} + \beta_{14} pol + \beta_{15} vol1$$

Results are reported in Table 4.

References

Anderson, C., Putterman, L. (2006). Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism, *Games and Economic Behavior*, 54, 1-24.

Andreoni, J. (1993). An experimental test of the public-goods crowding-out hypothesis, *American Economic Review*, 83, 1317-1327.

Andreoni, J., Harbaugh, W., Vesterlund, L. (2003). The carrot or the stick: rewards, punishments, and cooperation, *American Economic Review*, 93 (3), 893-902.

Botelho 2004

Bowles, S., Gintis, H. (2002). Behavioural science: Homo reciprocans. *Nature*, 415, 125-128.

Carpenter J., 2005, The demand for punishment. *Journal of Economic Behavior and Organization*.

Casari, M., Luini, L. (2007). Peer punishment in teams: expressive or instrumental choice? *Working Paper*.

Casari, M., Luini, L., (). Cooperation under alternative punishment institutions: an experiment, *Journal of Economic Behavior and Organization*, forthcoming.

Chan, K.S., Godby, R., Mestelman, S., Muller, R.A. (2002). Crwoding-out voluntary contributions to public goods. *Journal of Economic Behavior and Organization*, 48, 305-317.

Chen, Y., Plott, C.R. (1996). The Groves-Ledyard mechanism: An experimental study of institutional design. *Journal of Public Economics*, 59, 335-364.

Cinyabuguma, M., Page, T., Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, 9 (3), 265-279.

Denant-Boemont, L., Masclet, D., Noussair, C.N. (2007). Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Economic Theory*, 33, 145-167.

Dreber, A., Rand, D.G., Fudenberg, D., Nowak, M.A. (2008). Winners don't punish, *Nature*, 452, 348-351.

Falkinger, J. (1996). Efficient private provision of public goods by rewarding deviations from average. *Journal of Public Economics*, 62, 413-422.

Falkinger, J. (2004). Noncooperative support of public norm enforcement in large societies, CESifo Working Paper N. 1368.

Falkinger, J., Fehr, E., Gächter, S., Winter-Ebmer, R. (2000). A simple mechanism fort he efficient provision of public goods: Experimental evidence, *American Economic Review*, 90, 247-264.

Fehr, E., Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90 (4), 980-994.

Fehr, E., Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137-140**.**

Fehr, E., Gächter, S., Kirchsteiger, G. (1997). Reciprocity as a Contract Enforcement Device: Experimental Evidence, *Econometrica*, 65 (4), 833-60.

Fehr, E., Rockenbach, B. (2003). Detrimental Effects of Sanctions on Human Altruism, *Nature*, 422, 137-140.

Fischbacher, U. (1999). Z-Tree: a toolbox for readymade economic experiments, Working Paper, Zurich, University of Zurich, Institute of Empirical Research in Economics.

Gurerk, O., Irlenbusch, B., Rockenbach, B. (2006). The competitive advantage of sanctioning institutions. *Science*, 312, 108-111.

Herrmann, B., Thoeni, C., Gächter, S. (2008). Antisocial punishment across societies. *Science* 319, 1362-1367.

Isaac, R.M., Walker, J.M. (1988). Communication and free-riding behavior: the voluntary contributions mechanism. *Economic Inquiry*, 26, 585-608.

Kosfeld, M., Okada, A., Riedl, A. (2009). Institution formation in public goods games. *American Economic Review*, forthcoming.

Ledyard, J. (1995). Public goods: a survey of experimental research, in Kagel, J., Roth, A. (eds.), Handbook of Experimental Economics, Princeton, Princeton University Press.

Masclet, D., Noussair, C., Tucker, S., Villeval, M.C. (2003). Monetary and non-monetary punishment in the voluntary contributions mechanism. *American Economic Review*, 93 (1), 366-380.

Nikiforakis, N. (2008). Punishment and counter-punishment in public geode games: Can we really govern ourselves? *Journal of Public Economics*, 1(2), 91-112.

Oliver, P. (1980). Rewards and punishments as selective incentives for collective action: theoretical investigations. *American Journal of Sociology*, 85, 1356-1375.

Ostrom, E., Walker, J., Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86, 404-417.

Page, T., Putterman, L. (2000). Cheap talk and punishment in voluntary contribution experiments, *Brown University Department of Economics Working Paper*

Panchanathan, K., Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432, 499-502.

Sefton, M.R., Shupp, R., Walker, J. (2007). The effects of rewards and sanctions in provision of public goods. *Economic Inquiry.*

Sethi, R., Somanathan, E. (1996). The Evolution of Social Norms in Common Property Resource Use, *American Economic Review*, 86 (4), 766-88.

Smith, E.A., Bliege Bird, R. (2000). The benefits of costly signaling: Meriam turtle hunters, *Behavioral Ecology,* 14 (1), 116-126.

Wiessner, P. (2003). The pleasures and perils of reward and punishment: strong reciprocity among the Ju/'hoansi Bushmen. Department of Anthropology, University of Utah.

Figures and tables

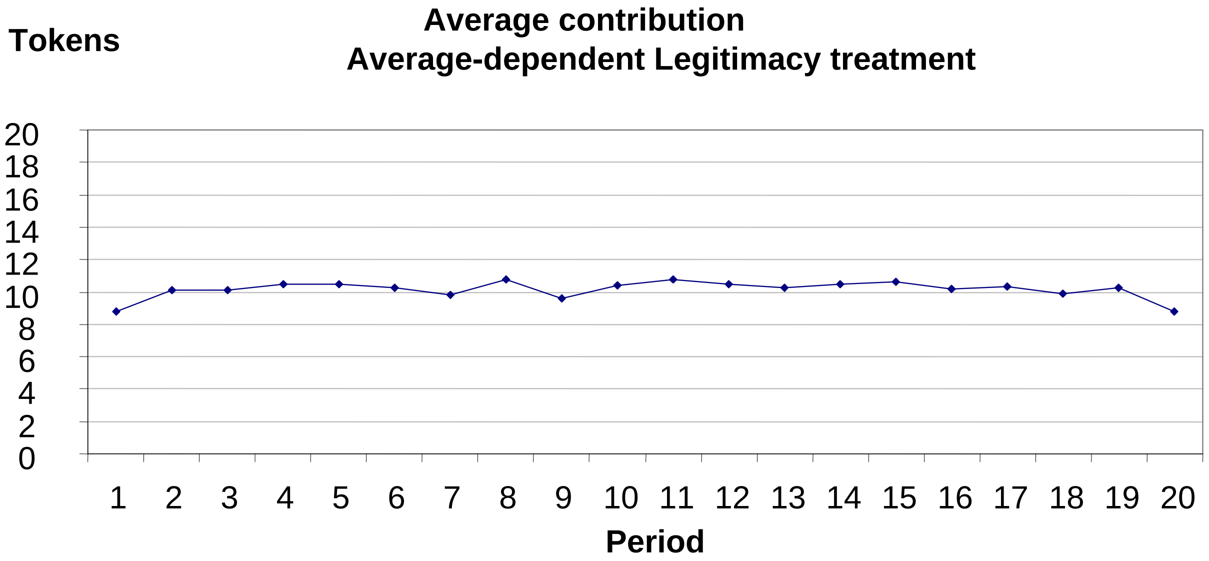

Figure 1 – Average contribution level in the AL treatment

Table 1: Average contributions

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
Average	0.862	2.225	17.437	19.15	1.95	19.45	5.71	3.36	13.87	
Std. dev.	0.813	3.274	3.277	1.933	0,894	1.438	2.151	3.228	3.434	3.687

Table 2:Determinants of the level of contribution

(contribution)	**OLS**
high_contr	7.589**
	(0.544)
pun_rec_lag	0.647**
	(0.138)
neg_dev	-0.829**
	(0.088)
pos_dev	0.956**
	(0.109)
gender	-0.786
	(0.493)
age	-0.679**
	(0.121)
trust	1.381**
	(0.117)
happ	-0.137
	(0.107)
constant	14.562
	(2.943)

Number of obs =760

F(26, 733)= 40.71 Prob > F=0.000

R-squared=0.50

Legend: the dependent variable takes values from 0 to 20. *High_contr*: dummy which takes value 1 when a group has shown an average contribution higher than 10 in the first period. *pun_rec_lag:* points received in the previous period. *neg_dev*: difference between the average contribution of the peers and own contribution. *pos_dev:* difference between own contribution and the average contribution of the peers. *gender:* 1=female; 0=male. *trust:* agreement (from 1 to 10) on the following statement: "Generally speaking, people can be trusted"-
happ: answer (from 1 to 10) on the following question: "How happy are you in this moment".

To control for time the regression model also contains period dummies (coefficients are not significantly different from 0).

** significant at 1%; * significant at 5%; Standard errors in brackets.

Table 3 : Average number of points assigned

	Groups									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mean	1.30	0.95	2.90	2.25	2.75	0.85	1.90	1.45	4.75	6.05
Std.dev	1.15	2.69	3.01	3.32	1.18	1.83	1.74	2.23	3.13	4.79

Number of obs: 80.

Table 4: determinants of punishment points assignment

(pun_given)	**OLS**
contribution	0.023**
	(0.007)
neg_dev	-0.031
	(0.018)
pos_dev	0.027**
	(0.021)
n_defectors	-0.152**
	(0.055)
pun_rec_lag	0.011
	(0.029)
happ	0.012
	(0.023)
trust	0.010
	(0.024)
gender	-0.461**
	(0.095)
age	-0.05*
	(0.027)
rel	0.053**
	(0.168)
rel_att	0.035
	(0.063)
pol	-0.048**
	(0.019)
vol	-0.166
	(0.127)
constant	1.270
	(0.684)

Number of obs =760

F(32, 727)= 14.16; Prob > F =0.000

R-squared=0.38

Adj- R-squared=0.36

Legend: the dependent variable takes values from 1 to 10. *contribution:* contribution to the public good (min=0; max=20). *neg_dev*: difference between the average contribution of the peers and own contribution; *pos_dev:* difference between own contribution and the average contribution of the peers *pun_rec_lag:* points received in the previous period.
To control for time the regression model also contains period dummies (coefficients are not significantly different from 0). *gender:* 1=female; 0=male. *trust:* agreement (from 1 to 10) on the following statement: "Generally speaking, people can be trusted"-
happ: answer (from 1 to 10) on the following question: "How happy are you in this moment?". *Rel:* religion (1=Catholic; 2=Protestant; 3: Muslim; 4=Buddhist; 5=Jewish; 6=Atheist; 7=Agnostic; 8=Other;); *rel_att*= attendance to religious services (1=every day; 2=a few times a week; 3=once a week; 4=a few times a month (less than 4); 5=never); pol:= Coalition voted in the last political election: 1=blank ballot; 2=abstained; 3=right; 4=center-right; 5=center-left; 6=left); *vol*: volunteer association activities (0=no; 1=yes).
** significant at 1%; * significant at 5%; Standard errors in brackets.