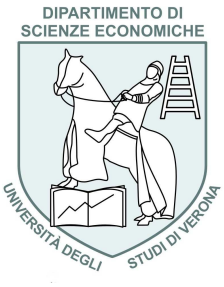

Working Paper Series
Department of Economics
University of Verona

Top Contributors as Punishers

Daniela Grieco, Marco Faillo, Luca Zarri

WP Number: 24 December 2013

ISSN: 2036-2919 (paper), 2036-4679 (online)

Top Contributors as Punishers

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

Abstract

The puzzle of human cooperation among strangers is still one of the fundamental open questions in contemporary social sciences. We experimentally investigate a finitely repeated public goods game where in each round access to sanctioning power is exclusively awarded to the group's *top contributor*. We show that this novel 'Top Contributors as Punishers' mechanism is extremely effective in raising cooperation and welfare, compared to other peer-to-peer sanctioning institutions. Despite the potential (first and second-order) free riding problem, the lure of the top contributor role induces many subjects to significantly contribute and incur relevant costs to sanction others. Our findings yield efficiency implications for the design of mechanisms intended to foster cooperation in social dilemma environments, from teamwork to voluntarily maintained web-based projects.

JEL Classification: C73; C91; D02 ; D63.

Keywords: Public Goods Games; Cooperation; Top Contributors; Behavioral Mechanism Design.

[a]daniela.grieco@unibocconi.it, Department of Economics, Bocconi University.

[b]marco.faillo@unitn.it, Department of Economics and Management, University of Trento.

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

1. Introduction

Explaining the emergence and sustainability of human cooperation in social dilemma environments, where a strong temptation to free ride on others exists, has long been a core problem for social scientists. In the last years, an increasing number of economic experiments have been contributing to shed light on the issue by investigating the role that *institutions* can play in enhancing cooperation (see e.g. Yamagishi, 1986, and Casari and Plott, 2003). Since in social dilemmas the maximization of social welfare conflicts with individual payoff maximization, the role of sanctioning institutions aimed at castigating deviant behavior has been extensively explored (Ostrom et al., 1992). On the whole, so far, laboratory studies have concentrated on two broad classes of punitive mechanisms, tackling the problem from two different angles: *decentralized* and *centralized* punishment.

Under voluntary, decentralized punishment, players are usually free to sanction each other arbitrarily and this institutional arrangement turned out to be extremely successful in stabilizing cooperation rates over time, due to many participants' willingness to engage in (costly) punishment of inappropriate behavior (see Fehr and Gächter's (2000; 2002) pathbreaking studies). However, the experimental literature has recently identified a 'dark side' of unrestricted peer punishment, by shedding light on some serious drawbacks of this peer-based sanctioning mechanism. First, in many cases it may undermine the scope for self-governance, as sanctioning may take the form of misdirected, 'antisocial' punishment – that is, low contributors inefficiently meting out sanctions on high contributors (Herrmann et al., 2008; Gächter and Herrmann, 2011)[1]. Relatedly, and even more importantly, recent work documents that the success of 'vigilante justice' in enforcing cooperation

[1] A further problem with discretionary sanctioning is that when multiple stages of punishment are allowed, so that immunity of sanctioners from reprisals is removed, counterpunishment and feuds are likely to be triggered, limiting, once again, successful self-governance and leading, eventually, to a demise of cooperation (Denant-Boemont et al., 2007, Nikiforakis, 2008 and Nikiforakis and Engelmann, 2011).

comes at a substantial cost: unless we consider a significantly longer time horizon (Gächter et al., 2008), average earnings are *lower* than in the absence of sanctioning options (Denant-Boemont et al., 2007; Dreber et al., 2008). This is a major shortcoming of unrestricted punishment, as it risks turning into a wasteful activity for those communities or organizations that adopt it (Nosenzo and Sefton, 2014).

Thus, it is natural to think that an alternative, viable solution could be to delegate the sanctioning power to an *external* enforcer, i.e. a Hobbesian 'Leviathan' entitled to monitor individuals' behavior and wield a 'sword' against free riders. However, even a purely centralized solution appears to be largely unsatisfactory under some important respects. A first reason has to do with the *informational* dimension (see on this also Baldassarri and Grossman, 2011). In many jobs, due to lack of physical proximity, employers cannot observe the exact contribution provided by each worker to the production of total output. The underlying argument is that in many socio-economic contexts the relevant knowledge is dispersed and a decentralized system is better able to detect it and fulfil its potential, compared to a centralized one. Next, even apart from informational problems, monitoring individuals can be extremely *costly*[2]. In this regard, recent studies also highlight the importance of potentially significant 'hidden costs of control' (Falk and Kosfeld, 2006; Schnedler and Vadovic, 2011).

Therefore, in light of such serious drawbacks characterizing 'extreme' (i.e. purely decentralized and purely centralized) punishment-based incentive schemes, in this paper we decided to experimentally test a novel sanctioning mechanism that occupies a somewhat intermediate position: we call it the 'Top Contributors as Punishers' mechanism. This peer-based institution prescribes that the sanctioning power is concentrated in the hands of a single player in each round: we exogenously impose that in each round *only the top*

[2] It is also worth noting that, in many areas of the world, serious obstacles in establishing a central authority endowed with the power to sanction wrongdoing arise due to lack or weakness of the rule of law.

contributor – that is, the player who acts more virtuously at the contribution stage – is granted the power to sanction others at the punishment stage. In principle, we view this as a hybrid institutional arrangement that combines the key advantages of purely centralized and purely decentralized sanctioning schemes: under this scheme, (i) the sanctioning activity is *exogenously restricted* (like in purely centralized systems, such as e.g. police and courts), so that antisocial punishment is ruled out, but at the same time (ii) it is *directly administered* by group members themselves (like in purely decentralized systems), without relying on external authorities.

Real-life examples of groups appointing only one peer at a time as a potential punisher and changing her over time include the choice of a scientific journal's editor, some horizontal relationships in the workplace and peers' interaction in web communities based on social networks. Insofar as it takes place after a careful scrutiny of her publication record, the appointment of a journal's editor is based on merit: she is usually a 'top contributor', in the sense that she is one of the most distinguished scholars in her field, and is appointed for a relatively short time span[3]. In labor environments, the highest ability worker in a team often has the authority (and sometimes even the formal legitimacy – e.g. when he becomes the 'team leader' or the 'team captain', in team sports) to sanction the coworkers who exert low effort and jeopardize the achievement of the team's goals[4]. A similar logic underlies the functioning of Wikipedia, the most world famous web-based encyclopedia. While a very large community of users voluntarily develops and maintains the pages, only the top contributors (that is the volunteers who are active and regular Wikipedia contributors for at least several months, have considerable experience and, therefore, are expected to have the trust and confidence of the

[3] Principal investigators in grant application processes provide a further example along these lines.

[4] Many online initiatives provide further examples along these lines: in web communities like Stack-exchange.com, participants accumulate credits when helping each-other – typically by providing suggestions on how to fix problems – and are awarded the title of moderators. This turns into the power of sanctioning peers who behave impolitely or write posts that are off-topic.

community) are likely to become administrators, i.e. editors who hold sanctioning power as they have been granted the ability to block user accounts and IP addresses from editing pages and other actions[5].

On the whole, our experimental analysis in this paper includes three peer-based punishment treatments. While in one treatment punishment is potentially *diffuse*, in the sense that more than one player in each group may sanction others in each round, in the remaining two treatments sanctioning power is *concentrated* in the hands of a single player. More specifically, we aim at comparing the performance, in terms of both cooperation and efficiency, of the 'Top Contributors as Punishers' mechanism to the following two treatments: a baseline based on a classic, unrestricted peer punishment mechanism *à la* Fehr and Gächter (2000) and an institution based on a solitary but randomly selected punisher (O'Gorman et al., 2009). This will allow us to also contribute to the thin but growing stream of literature that has been focusing on the impact that restricting access to punishment opportunities to one player at a time can have on the enforcement of cooperation.

Our findings indicate that the 'Top Contributors as Punishers' institutional arrangement effectively fosters cooperation and welfare, thanks to many subjects' tendency to significantly contribute to the public good and to top contributors' high willingness to costly mete out sanctions on their peers. This interestingly occurs despite the lack of monetary incentives to do so and the presence of both a first-order and a second-order free riding problem that, in principle, may have had a demotivating effect, leading subjects to perceive the new institution at work as an excessively demanding one and possibly to avoid being the highest contributors in order to leave on other players' shoulders the burden of sanctioning activity.

[5] The Wikipedia website clearly states that administrators are not external subjects but members of the same community of users: "Administrators were not intended to develop into a special subgroup. Rather, administrators should be a part of the community like other editors".

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 illustrates the experimental design. Section 4 reports our core findings and Section 5 concludes the paper.

2. Related literature

As to experiments based on exogenous restrictions on the number of peer punishers, the most closely work is the following. Three recent papers (O'Gorman et al., 2009; Carpenter et al., 2012, and Nosenzo and Sefton, 2014), in line with our study, investigate the finitely repeated public goods game framework and depart from Fehr and Gächter's (2000) seminal article by exogenously imposing that *only one punisher per group* can emerge in each round. The presence of such restriction significantly differentiates these papers from most previous studies on peer punishment[6].

A potential advantage of a solitary punisher mechanism is clearer accountability, as, by concentrating the sanctioning power in the hands of a single player, "it lacks the second-order free rider problem – which has been the central focus of much theoretical effort – and it avoids the problem of uncoordinated over punishment" (O'Gorman et al., 2009)[7].

O'Gorman et al. (2009) explore a solution in which responsibility for punishment is borne by one specific, designated and randomly selected individual. They find that under solitary punishment cooperation can be successfully sustained and reaches levels comparable with that maintained by diffuse punishment. Further, group earnings levels are higher: a sole punisher solution reduces inefficient losses, as sanctioning efforts are not unnecessarily duplicated. Carpenter et al. (2012) also have a treatment where the monitoring and sanctioning power is concentrated in the hands of one randomly selected

[6] Other experimental work based on restricted sanctioning (though not on the appointment of single punishers) includes Carpenter (2007) and Eriksson et al. (2013). For a recent study focusing on a stochastic two-person prisoner's dilemma environment where only cooperators can punish non-cooperators, see Xiao and Kunreuther (2014).
[7] See on this point also Nosenzo and Sefton (2014).

group member. However, unlike O'Gorman et al. (2009), they show that under this form of restricted punishment both contributions and average earnings are *lower* than under peer-to-peer punishment. The third closely related study (Nosenzo and Sefton, 2014), by means of a five-treatment experimental design, analyzes both reward and punishment schemes, under both centralized and decentralized systems. In one of their central monitoring treatments, all punishment power is concentrated in the hands of one, randomly assigned group member and this was the same group member in all ten periods. The authors show that concentrating punishment power *reduces* the effectiveness of punishment, compared to (unrestricted) peer punishment. Hence, their findings are similar to Carpenter et al.'s results: in both papers, unlike in O'Gorman et al., concentrating the power in the hands of a solitary punisher is not an effective means to sustain cooperation. These three papers on the whole present mixed evidence on the effectiveness of solitary punishers in raising cooperation and earnings levels.

Since in our 'Top Contributors as Punishers' treatment each player in each round has the chance to "earn the right" to be the sole punisher – as this possibility crucially depends on her behavior at the contribution stage –, we claim that our design innovation allows us to analyze a form of 'endogenous leadership': top contributors may be viewed as 'team leaders' who, far from being exogenously appointed through a random selection mechanism, succeed in earning the leader role *throughout the game*, i.e. thanks to their contribution decisions in the first stage of the game. In this regard, our experiment also relates to the body of research in economics and social psychology focusing on sequential versions of social dilemmas and dealing with the role of 'leading-by-example' as a cooperation enforcement device (Van Vugt and De Cremer, 1999; Güth et al., 2007; Arbak and Villeval, 2013)[8]. On the whole,

[8] Compared to other work in this literature, Güth et al.'s (2007) study is closer to ours as it associates leadership with punishment power, within a public goods game framework.

this literature documents an ambiguous effect of leadership on public goods provision.

A further related paper is Andreoni and Gee (2012), who explore an enforcement device based on authority delegation to a third party which they term the 'hired gun' mechanism. Compared to classic unrestricted peer punishment, the hired gun mechanism acts as a low cost device that fosters cooperation and welfare. Unlike their mechanism, in our central treatment we explore a peer-to-peer punishment institution, rather than an external third party[9].

Our work has also some similarities with the large literature on *tournament-based* incentive schemes in organizational settings. In particular, by imposing that only top contributors can have access to punishment power, we are close to the body of research exploring the effectiveness of reward-based rank-order tournaments extensively used in real world settings by managers as motivational tools to encourage employees to compete and work harder than their colleagues (Konrad, 2009). In so called 'reward tournaments', the most productive agent in the team receives the best prize, usually consisting of bonuses (for example, for the 'employee of the month') or promotions (Lazear and Rosen, 1981). Therefore, our central treatment also incorporates a reward tournament component: since by design in each group only one player at a time – and, in particular, the top contributor – is assigned the power to punish her coplayers, we are implicitly providing an incentive to compete for the top.

The key difference with the aforementioned literature is that while in tournament-based incentive schemes employees typically compete for *monetary* benefits, in our setting top performers get a non-monetary reward, i.e. the power to sanction others. It is worth noting that, since sanctioning others is costly, our reward for top contributors is not simply monetarily

[9] A second key difference between our design and Andreoni and Gee's study is that, as we clarified above, punishers, in our experiment, far from being randomly selected, have to *earn* this right by being the highest contributors in their group.

neutral, but even *costly*. In this regard, it is plausible to believe that our key treatment, by granting exclusive access to punishment power, has connections also with the recent literature exploring the incentive effects of *decision rights*, and in particular the non-pecuniary utility associated with authority per se (Fehr et al., 2013). In their authority-delegation game, Fehr et al. (2013) find that the principals show a proclivity for retaining authority in situations in which they could improve their expected income by delegating it. Also Bartling et al. (2014) shed light on the motivating power of authority providing experimental evidence on the *intrinsic* – rather than purely instrumental – value of decision rights. de Quervain et al. (2004) provide neuroscientific evidence that people derive non-pecuniary utility from punishing others. They examine the neural basis for punishment of unfair behavior and find that sanctioning activates reward-related brain regions. In particular, they show that the activation in the dorsal striatum reflects the anticipated satisfaction from punishing individuals who behave unfairly.

Finally, this study has some similarity with recent experimental work focusing on the *legitimacy* of sanctioning institutions. In previous work (Faillo et al., 2013), we investigate a sanctioning institution where, in principle, more than one punisher exists in each group (unlike in the 'Top Contributors as Punishers' mechanism), but where we exogenously impose that, in each round, player *i* can punish player *j* only insofar as *i* contributes more than *j* (in the same round). Our results indicate that such a 'legitimate punishment' scheme yields substantial benefits to cooperation and welfare, compared to a 'vigilante justice' institution. Other recent lab and field experiments have concentrated on punishment mechanisms which either derive their legitimacy from a process of *endogenous* choice (Ertan et al., 2009; Sutter et al., 2010; Eckel et al., 2010; Baldassarri and Grossman, 2011) or fruitfully interact with moral messages to sustain cooperation (Dal Bó and Dal Bó, 2014).

3. Experimental Design

3.1. Procedures

A total of 168 subjects participated voluntarily in the experiment at the CEEL Lab of the University of Trento. Nine sessions were conducted between November 2009 and November 2013. The experiment was programmed by using the z-tree platform (Fischbacher, 2007). Subjects were undergraduate students (46% from Economics, 48% females, 77 % Italians). We employed a between subjects design: no individual participated in more than one session. In each session, the participants were paid a 5 euros show up fee, plus their earnings from the experiment. The average payment per participant was 12.52 euros (including the show-up fee) and the sessions averaged approximately 90 minutes. At the beginning of each session, participants were welcomed and asked to draw lots, so that they were randomly assigned to terminals. Once all of them were seated, the instructions[10] were handed to them in written form before being read aloud by the experimenter. The participants had to answer several control questions and we did not proceed with the actual experiment until all participants had answered all questions correctly.

For each treatment, participants in each session were randomly assigned to groups of size N=4, so that they did not know the identities of the other members of their group. Like other experimental studies (see e.g. Cinyabuguma et al., 2006; Denant-Boemont et al., 2007), we used a partner protocol that kept the composition of each group constant over rounds, so that, at the end of each period, individuals remained in the same group. The reason why we use a partner design is that repeated interaction is a typical feature of many real world settings (e.g., businesses or web-based communities) in which sanctioning often takes place (Xiao and Houser, 2011). However, individuals' labels were randomly reassigned in each period. For example, the same player could be designated as player *32* in period *t*, as player *5* in period *t*

[10] A translation of the instruction sheet is provided in Appendix C. Original instructions were written in Italian. They are available upon request from the authors.

+ 1, and as player *43* in period $t + 2$. Therefore, our partner protocol was also characterized by anonymity of the components of the group and change of participants' labels across rounds. The parametric structure of the experiment is based on Fehr and Gächter (2000).

3.2. Treatments

The structure of monetary payoffs in a laboratory linear public goods game or voluntary contribution mechanism (*VCM*) makes it a classical 'social dilemma', characterized by a tension between individual and group incentives: each player has a dominant strategy to free ride, while at the social optimum each agent allocates his entire endowment to the group account. Therefore, in the finitely repeated version of this game, if we supposed that individuals are driven by selfish concerns over their own material payoffs only (and that there is common knowledge of this), we should expect all of them to avoid contributions and systematically free ride from the outset. By contrast, a robust result from the experimental literature is that initially cooperation rates are relatively high ('overcontribution'), while they gradually decline over time, leading to the so called 'decay' phenomenon, with complete free riding often prevailing towards the end of the game (Ledyard, 1995; for a recent selective survey of this literature, see Chaudhuri, 2011).

In our experiment, participants played, in all treatments, a finitely repeated public goods game with punishment options for 20 periods. In every period, the experimental game consisted of two decision stages: at stage 1 (contribution stage), players choose how much to contribute to the public good and at stage 2 (punishment stage) they have access to punishment options. In each treatment, participants were informed about these features of the game to be played. The following three treatments were implemented: (1) a baseline, unrestricted punishment (Baseline) treatment, (2) a restricted, random solitary punishment (SR) treatment and (3) a restricted, solitary punishment (SP)

treatment where only top contributors can act as punishers[11]. There were 3 sessions for the Baseline, 3 sessions for the SR and 3 sessions for the SP: in all cases, we had two sessions with 20 subjects and one with 16 subjects.

3.2.1. Baseline

In the Baseline treatment, punishment is unrestricted and subjects are provided with full information, that is there is feedback about *all* their group co-players' individual contributions. This is a replication of the standard linear *VCM* with punishment and partner protocol (Fehr and Gächter, 2000), where everyone can freely punish everyone else in the group. In stage 1, each participant receives a fixed amount e =20 of tokens and has to decide whether she wants to invest or not an amount $g_i \leq e$ into a public project. 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 are calculated by the computer in the following way:

$$\pi_i = (20 - g_i) + 0.4 \sum_{j=1}^{4} g_j$$

In stage 2 subjects are informed about the contribution by the other members of their group and can decide to assign between 0 and 10 punishment points to any of them. Points assignment is costly and costs are charged according to a convex cost function as in Fehr and Gächter's study (Table 1).

[TABLE 1]

Each point that a subject receives reduces her earnings at stage 1 by 10%, with 100% as the maximum total reduction. Punishment is anonymous: subjects who get punished do not know the identity of the punisher. Each participant's net earnings at the end of stage 2 are given by her earnings at the end of stage 1 minus the costs of assigned and received punishment points: they are

[11] Part of these data (namely, the ones from the Baseline) were also analyzed in Faillo et al. (2013).

calculated by the computer and each participant sees her cumulative net earnings on the screen at the end of each round.

3.2.2. Randomly selected solitary punisher

The SR treatment differs from the Baseline as now punishment power is concentrated in the hands of *one subject* only in each round. As before, subjects are provided with full information on the contribution levels of their peers in the group. However, in the punishment stage only one *randomly selected* participant receives the right to punish their peers. The random draw is repeated in each period, potentially assigning the sanctioning power to different subjects over time (like in O'Gorman et al., 2009). Like in the Baseline, the punisher can decide to assign between 0 and 10 punishment points to any peer and costs are charged according to the same cost function.

3.2.3. Restricted solitary punishment: the 'Top Contributors as Punishers' mechanism

The key feature of the central treatment that we analyze in this study, that is the SP, is that in each round only the 'Top Contributor' is allowed to sanction her group coplayers[12]. A participant is classified as Top Contributor when her contribution is the highest in the group ($g_i > g_{-i}$). In this treatment, in stage 2, the Top Contributor *only* is given the opportunity to punish other subjects in the group by assigning a certain amount of points.

In case the highest amount is contributed identically by two or more subjects, only one of them gets randomly selected to make a punishment choice that will be actually implemented[13]. Like in the other two treatments, all subjects receive information about their peers' contribution behavior.

[12] It is important to make clear that, in order to minimize so called 'experimenter demand effects' (Zizzo, 2010), we never used loaded terms such as 'punishment', 'top contributor' and 'free riding' during the experiment.

[13] The other highest contributors are asked to answer a hypothetical question asking what they would do in terms of punishment, knowing that their decision won't produce real consequences in the game.

However, only the Top Contributor might decide to actually assign up to 10 points to each co-player, with punishment costs being charged according to the previously illustrated cost function. Each participant knows that she can go on with stage 2 in the experiment only if she is the Top Contributor at stage 1, that is only if she earns the right to sanction others.

It is worth pointing out that, since punishing others is monetarily costly, in SP, like in a standard, finitely repeated *VCM* with unrestricted punishment options, insofar as all the subjects are supposed to be driven by material self-interest only and this information is common knowledge, the unique subgame perfect equilibrium is for all agents to *never punish* and *never contribute.*

By running two treatments where only one punisher at a time can arise (i.e. the SR and the SP), we also contribute to the growing stream of experimental literature on peer punishment dealing with its *concentrated* vs. *diffuse* nature and described in Section 2. The key question addressed by these studies is: does concentrating the punishment power in the hands of some subjects only make a difference? In principle, concentrating the punishment power solves the coordination problem and reduces inefficiencies in sanctioning. But the flip side of the coin is that punishment itself is a *second-order public good* (Yamagishi, 1986), so that asking only some subjects to carry the burden of providing it can be subjectively perceived as excessively demanding by the players themselves. This holds true especially in SP, where they are aware that they need to be the highest contributors to gain access to punishment. Therefore, introducing this form of solitary punishment may ultimately lead to *less* punishment and cooperation. We claim that this trade-off makes it worth addressing the above research question in the lab.

4. Results

4.1. Contribution levels

Figure 1 displays the pattern of average contributions by period in the three treatments.

[FIGURE 1]

In neither treatment average contributions decline over time. By comparing the levels of cooperation across treatments, we find that mean contributions (Table 2) in SP are *significantly higher* than those of both Baseline and SR (Wilcoxon Rank-sum Test with group averages as independent observations: SP vs. Baseline: $z = 2.38$, p-value: 0.017; SP vs. SR: $z = 2.27$, p-value=0.02). Therefore, the introduction of the 'Top Contributors as Punishers' mechanism has a positive effect on the level of contributions. Further, there is no significant difference between the level of contributions of Baseline and SR (Wilcoxon Rank-sum Test: Baseline vs. SR: $z = 0.55$, p-value=0.58). These results are supported also by a random effect Tobit estimation (Table 3).

Result 1. The introduction of the 'Top Contributors as Punishers' mechanism significantly increases the level of cooperation.

[TABLE 3]

4.2. Punishment behavior

As Result 1 shows, the introduction of the 'Top Contributors as Punishers' mechanism determines higher contribution levels, compared to the two treatments characterized by unrestricted punishment (Baseline and SR). With regard to the distribution of punishment points, in all the treatments we observe the typical decreasing pattern (Figure 2). In SR, punishment is significantly lower than in the other treatments (Wilcoxon Rank-sum: Baseline

vs. SP: z=1.35, p-value=0.17; Baseline vs. SR: z=2.50, p-value=0.01; SP vs. SR: z=3.31, p-value=0.00).

[FIGURE 2]

Result 2. On average, when a single punisher is randomly selected, the number of punishment points assigned is significantly lower than in the other treatments.

It is worth noting that in SP solitary punishers turn out to be willing to incur, on average, a very high personal cost for punishing others. Therefore, SP successfully fosters cooperation by relying on the key role played by some individuals with regard to both the first-order and the second-order public good to be voluntarily provided: far from free riding on others' efforts, these players are both willing to be the top contributors, at the contribution stage, *and* ready to incur relevant costs at the sanctioning stage.

In case of equally ranked top contributions, the Top Contributors who are extracted and get the concrete chance of punishing do assign an average of 2.61 punishment points, whereas the average number of hypothetical punishment points virtually assigned by non-extracted Top Contributors is 1.99. This indicates that Top Contributors' propensity to punish is higher when they know that their assigned punishment points have a real, rather than virtual, nature.

When punishment activity is not restricted, a non-negligible share of punishment points can be classified as 'antisocial', since they are directed to subjects who contribute *more than the punisher* (see on this also Faillo et al., 2013).

Antisocial punishment turns out to be a widespread, quantitatively relevant phenomenon: the percentage of antisocial points is 19,5% in the Baseline and reaches 30% in SR (See the details in Appendix A). Thus, we interestingly

show that in SR subjects are less prone to punish and about one third of the punishment points assigned by them are not used to sustain cooperation. This finding suggests that an important 'enemy' of cooperation such as antisocial punishment manifests itself as a quantitatively relevant phenomenon not only when discretionary punishment is potentially diffuse (Baseline), but also when one (randomly selected) participant at a time is free to sanction others (SR). More specifically, we conjecture that the higher percentage of antisocial points observed in SR compared to the Baseline depends on the noisy signals sent to individuals by a sanctioning institution where potential punishers are *randomly* selected. This is in stark contrast with mechanisms where players need to *earn* the punishment power, e.g. passing through voting (Baldassarri and Grossman, 2011), by being the top contributors in the group, as it is the case in SP, in this study, or by means of (comparatively) virtuous behavior at the contribution stage (Faillo et al., 2013) or in an unrelated trivia quiz (Eckel et al., 2010).

4.3. Average net earnings

Taking group average earnings as independent observations, we observe that average earnings in SP are significantly higher than in the Baseline (Wilcoxon Rank-sum: Baseline vs. SP: $z=2.11$, p-value=0.03). Average earnings in SR are slightly greater than those in the Baseline ($z=1.65$, p-value =0.09). Earnings in SP are not significantly different from those in SR($z=0.82$, p-value =0.41)

[FIGURE 3]

Result 3. Average earnings are significantly lower in the Baseline than in SP. The SR treatment performs slightly better than the Baseline, due to the lower usage of punishment.

4.4. Determinants of changes in individual contributions

The key design feature of SP is that in this treatment only one subject at a time – namely, the Top Contributor – is awarded the right to punish others. Therefore, the role of Top Contributor (and her contribution level) is made *salient* by the very structure of the game. However, it is important to note that, in principle, this saliency per se should *not* make that role *attractive*, due to the high personal monetary costs associated to being the Top Contributor. In other words, SP appears to be a rather motivationally demanding punishment institution, since its structure implicitly 'asks' single players not only to compete with each other and try to be the most virtuous player at the contribution stage (first-order public good), but also to carry the whole burden of punishment activity in the same round (second-order public good). Therefore, this feature of SP may lead to lower contributions (in order to avoid to be the one in charge of punishment activity) and, therefore, to a *reduction* in the effectiveness of this mechanism as a cooperation enforcement device, compared to other sanctioning institutions. Alternatively, in SP we might observe many individuals who contribute a lot but only to gain *immunity* from others' sanctioning: as a consequence, these players, after earning the top contributor role, would strategically act as second-order free riders, failing to incur costs to sanction others.

In fact, if we look at the entitled and extracted members of each group in SP across periods (see Appendix B for details), we observe that the number of times players are entitled *far exceeds* the number of times they are actually extracted: this reveals that widespread attempts to be the highest contributor and sanction others take place – despite the fact that both contributing and punishing are costly. Therefore, the fact that many Top Contributors are also willing to significantly sanction others allows us to rule out that many players' willingness to contribute a lot is exclusively driven by an instrumental motive such as the search for immunity.

More specifically, we find that, i) on average there is *more than one* entitled member per period in each group and that ii) the role of potential punisher tends to be assigned to *different members in different periods*. It follows that, in the presence of institutional constraints imposing that only one member at a time (namely, the top contributors) can have access to punishment, a form of longitudinal distribution of punishment activity among group members arises endogenously. This rotation in the top contributor and punisher role naturally arises and, due to the high personal (monetary) costs associated to being a Top Contributor who also significantly punishes others, it is likely to play a key complementary role – together with many individuals' willingness to significantly contribute and sanction – for the sustainability of cooperation and, therefore, for the success of our novel peer-based sanctioning mechanism[14].

The presence vs. lack of rotation in the sole punisher role seems to make a difference also in the closely related experimental studies recalled in Section 2: as pointed out by Nosenzo and Sefton (2014), the reason why Carpenter et al. (2012) and O'Gorman et al. (2009) reach different conclusions is likely to be related to the fact that while in the former roles remained *constant* throughout the experiment, in the latter the role of central monitor was randomly assigned to a new subject *in each new round of play*, so that this may have attenuated the negative impact of the appointment of a 'bad monitor' (that is, a monitor who never sanctions free riding). We add that rotation makes contributing and punishing *less demanding* for each single player, compared to a treatment where the same person acts as a monitor throughout the whole game[15].

[14] Concerning the demographic features characterizing the identity of the Top Contributor, the probability to become a Top Contributor increases when the subject is a male and his major is Economics, and decreases in the number of experiments in which the subject took part. Once there is a tie, and a random draw among the entitled Top Contributors determines who is the one who can actually punish. Neither males nor students in Economics are more represented in the extracted pool.

[15] We view an institution where in each group the punisher may change from round to round as applying some of the key design principles that, according to Ostrom (2000), drive the

In order to shed light on the forces driving the success of SP as a sanctioning institution, we investigate the determinants of the increase in subjects' contribution level from *t* to *t+1* in all the three treatments (Table 4).

[TABLE 4]

Our analysis interestingly reveals that, given the distance from one's own contribution and the highest contribution in the group, in all the treatments *but SP*, punishment is effective: the higher the number of punishment points a subject receives, the higher the increase in contribution in the next period. Furthermore, in all the treatments *but SP* the distance from Top Contributor's contribution in period *t-1* - measured as the absolute difference between subject's contribution and the highest contribution in the group - *positively* affects subjects' contribution in period *t*: the larger the distance, the larger the increase in contribution.

Therefore, in SP, the increase of contributions is not driven by received punishment points, that turn out to be ineffective. The only aspect that seems to play a role is the Top Contributor's level of contributions, that drives significantly, but now *negatively,* subjects' contributions. Players appear to use the feedback on the highest contribution (corresponding to the choice of the only peer who can punish) as a reference level to tailor their own choice: the closer they are to the top the bigger the increase of their contribution – i.e. their effort to reach the top.

Hence, introducing restrictions in the usage of punishment by means of an experimental design that makes the Top Contributor salient seems to effectively favor the enforcement of cooperation but through the operation of

---

functioning of self-organized resource regimes sustained by local users (such as e.g. in-shore fisheries): "By creating official positions for local monitors, a resource regime does not have to rely only on willing punishers to impose personal costs on those who break a rule. *The community legitimates a position*. In some systems, *users rotate into this position* so everyone has a chance to be a participant as well as a monitor" (italics added).

different forces, compared to the other forms of peer-based sanctioning under study.

5. Conclusions

A recent lesson from work in behavioral mechanism design is that the way in which formal rules and incentive schemes interplay with individual preferences is often counterintuitive (Gneezy et al., 2011; Bowles and Polania-Reyes, 2012). Many subjects are to some extent driven by social concerns and their perception of the institutional environment they are embedded in is likely to significantly shape their decision-making process. Therefore, it is far from obvious that exogenous institutional changes that leave monetary incentives unchanged will be neutral, as, in principle, both crowding-in and crowding-out outcomes are possible (Galbiati and Vertova, 2008; Dale and Morgan, 2010). We argue that this methodological caveat is important also for incentive schemes based on punishment. In this regard, our findings reinforce the view, advanced by previous experimental studies, that individual behavior crucially depends on the nature of the sanctioning system at work (see e.g. Herrmann et al., 2008, Ambrus and Greiner, 2012, Xiao, 2013 and Faillo et al., 2013).

In particular, our central treatment reveals that a mechanism assigning punishment power to top contributors only successfully raises cooperation and welfare, compared to sanctioning institutions where antisocial punishment is permitted – despite the relevant monetary costs incurred by solitary punishers. Previous experimental work indicates that, within the public goods game framework, a design that makes top contributions salient to the players is not sufficient *per se* to foster cooperation. Dale and Morgan (2010) document a perverse *negative* effect of asking individuals to contribute the socially optimal amount, as this tool turns out to be, at best, ineffective. Similarly, Samek and Sheremeta (2014) show that recognizing only the highest contributors, by displaying their identities, does not increase contributions. By

contrast, when the top contributor role is made salient due to its link with the sanctioning institution, as it is the case in our 'Top Contributors as Punishers' mechanism, cooperation and welfare significantly increase.

We believe that the most plausible interpretation of our result is that, even though being a top contributor who punishes others is monetarily costly, many players are attracted by a non-pecuniary motive such as the 'lure of authority' (Fehr et al., 2013) associated with exclusive access to punishment power: the search for exclusive punishment seems to be sufficient to induce several players to engage in a competition within the group, countervailing the demotivating effect that knowing to be the only provider of the second-order public good may produce. This interpretation is compatible with de Quervain et al.'s (2004) neuroeconomic finding that punishment of unfair behavior activates reward-related brain regions: this supports the view that individuals derive *satisfaction* from a materially costly activity such as sanctioning others.

More generally, we view our results as consistent with Ostrom et al.'s (1992) claim that policymakers responsible for the governance and management of common pool resources should not presume that individuals facing social dilemma situations are caught in inexorable tragedies from which there is no escape. Our findings yield efficiency implications for the design of mechanisms intended to foster cooperation – from teamwork to voluntarily maintained web-based projects – and challenge the Hobbesian view that an external enforcer is always necessary to grant cooperation: sustainable self-governance is possible as individuals can successfully engage in collective action even without calling for an external enforcer, provided that they are able to rely on 'motivationally appropriate' institutional arrangements.

Acknowledgements: We thank Antonio Fasano, Enrique Fatas, Arnout van de Rijt, George Loewenstein, seminar participants at LUISS University in Rome and participants in the 2014 IMEBESS Conference in Oxford for useful suggestions and comments. The usual caveats apply. We gratefully acknowledge the University of Verona (2011 Joint Projects on "Punishment and Decision-making: Neuroeconomic Foundations, Behavioural Experiments and Implications for Law and Economics"),

the Department of Economics and Management of the University of Trento and the Department of Economics of the University of Verona for financial support.

References

Ambrus, A., Greiner, B. (2012). Imperfect public monitoring with costly punishment: an experimental study. American Economic Review, 102 (7), 3317-3332.

Andreoni, J., Gee, L.K. (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. Journal of Public Economics, 96, 1036-1046.

Arbak, E., Villeval, M. (2013). Voluntary leadership: motivation and influence. Social Choice and Welfare, 40 (3), 635-662.

Baldassarri, D., Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. Proceedings of the National Academy of Science, 108, 11023-11027.

Bartling, B., Fehr, E., Herz, H. (2014). The intrinsic value of decision rights. Econometrica, forthcoming.

Bowles, S., Polania-Reyes, S. (2011). Economic incentives and social preferences: substitutes or complements? Journal of Economic Literature, 50 (2), 368-425.

Carpenter, J. (2007). Punishing free-riders: how group size affects mutual monitoring and the provision of public goods. Games and Economic Behavior, 60, 31-51.

Carpenter, J., Kariv, S., Schotter, A. (2012). Network architecture, cooperation and punishment in public good experiments. Review of Economic Design, 16 (2), 93-118.

Casari, M., Plott, C. (2003). Decentralized management of common property resources: experiments with a centuries-old institution. Journal of Economic Behavior and Organization, 51, 217-247.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Experimental Economics, 14 (1), 47-83.

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

Dale, D.J., Morgan, J. (2010). Silence is golden. Suggested donations in voluntary contribution games. Mimeo.

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.

Dal Bó, E., Dal Bó, P. (2014). "Do the right thing:" The effects of moral suasion on cooperation. Journal of Public Economics, forthcoming.

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.

Eckel, C.C., Fatas, E., Wilson, R. (2010). Cooperation and status in organizations. Journal of Public Economic Theory, 12 (4), 737-762.

Eriksson K., Strimling P., Ehn M. (2013). Ubiquity and efficiency of restrictions on informal punishment rights. Journal of Evolutionary Psychology, 11, 17-34.

Ertan, A., Page, T., Putterman, L. (2009). Who to punish? Individual decisions and majority rule in mitigating the free rider problem. European Economic Review, 53 (5), 495-511.

Faillo, M., Grieco, D., Zarri, L. (2013). Legitimate punishment, feedback, and the enforcement of cooperation. Games and Economic Behavior, 77, 271-283.

Falk, A., Kosfeld, M. (2006). The hidden costs of control. American

Economic Review, 96 (5), 1611-1630.

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., Herz, H., Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. American Economic Review, 103 (4), 1325-1359.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10, 171-178.

Gächter, S., Herrmann, B. (2011). The limits of self-governance when cooperators get punished: Experimental evidence from urban and rural Russia. European Economic Review, 55 (2), 193-210.

Gächter, S., Renner, E., Sefton, M. (2008). The long-run benefits of punishment. Science, 322, 5907, 1510.

Galbiati, R., Vertova, P. (2008). Obligations and cooperative behaviour in public goods games. Games and Economic Behavior, 64 (1), 146-170.

Gneezy, U., Meier, S., Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. Journal of Economic Perspectives, 25 (4), 1-21.

Güth, W., Levati, M.V., Sutter, S., van der Heijden, E. (2007). Leading by example with and without exclusion power. Journal of Public Economics, 91, 1023-1042.

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

Konrad, K.A. (2009). Strategy and dynamics in contests, Oxford, Oxford University Press.

Lazear, E.P., Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. Journal of Political Economy, 89 (5), 841-864.

Ledyard, J. (1995). Public goods: a survey of experimental research, in Kagel,

J., Roth, A. (eds.), Handbook of Experimental Economics, Princeton, Princeton University Press.

Nikiforakis, N. (2008). Punishment and counter-punishment in public goods games: Can we really govern ourselves? Journal of Public Economics, 92, 91-112.

Nikiforakis, N., Engelmann, D. (2011). Altruistic punishment and the threat of feuds. Journal of Economic Behavior and Organization, 78 (3), 319-332.

Nosenzo, D., Sefton, M. (2014). Promoting cooperation: the distribution of reward and punishment power, in Van Lange, P.A.M., Rockenbach, B., Yamagishi, T. (eds.), Social dilemmas: New perspectives on reward and punishment, Oxford, Oxford University Press.

O'Gorman, R., Henrich, J., Van Vugt, M. (2009). Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. Proceedings of the Royal Society B, 276, 323-329.

Ostrom, E. (2000). Collective action and the evolution of social norms. Journal of Economic Perspectives, 14 (3), 137-158.

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

Samek, A.S., Sheremeta, R. (2014). Recognizing contributors: an experiment on public goods. Experimental Economics, forthcoming.

Schnedler, W., Vadovic, R. (2011). Legitimacy of control. Journal of Economics and Management Strategy, 20 (4), 985-1009.

Sutter, M., Haigner, S., Kocker, M.G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. Review of Economic Studies, 77 (4), 1540-1566.

Van Vugt, M., De Cremer, D. (1999). Leadership in social dilemmas: the effects of group identification on collective actions to provide public goods. Journal of Personality and Social Psychology, 76 (4), 587-599.

Xiao, E. (2013). Profit seeking punishment corrupts norm obedience. Games and Economic Behavior, 77, 321-344.

Xiao, E., Houser, D. (2011). Punish in public. Journal of Public Economics, 95, 1006-1017.

Xiao, E., Kunreuther, H. (2014). Punishment and cooperation in stochastic prisoner's dilemma game. Journal of Conflict Resolution, forthcoming.

Yamagishi, T. (1986). The provision of a sanctioning system as a public good. Journal of Personality and Social Psychology, 51, 110-116.

Zizzo, D.J. (2010). Experimenter demand effects in economic experiments. Experimental Economics, 13 (1), 75-98.

Figures and tables

Figure 1. Average contributions

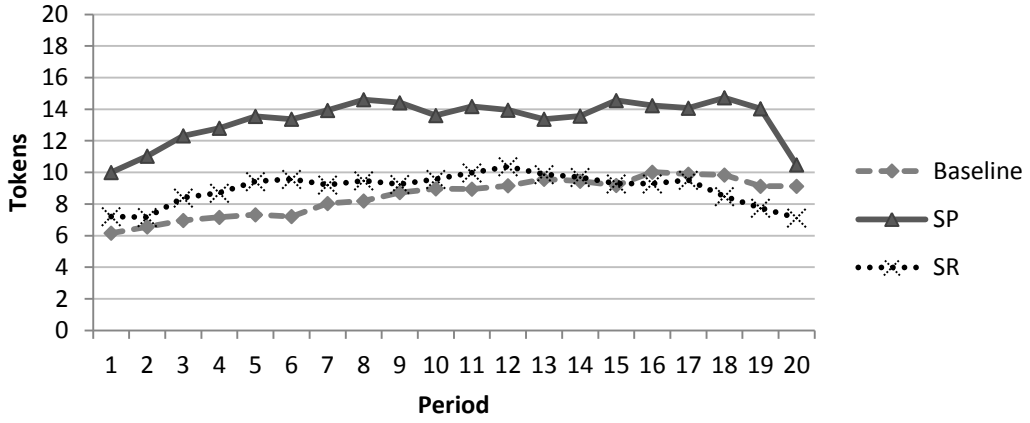

Figure 2. Average quantity of points given

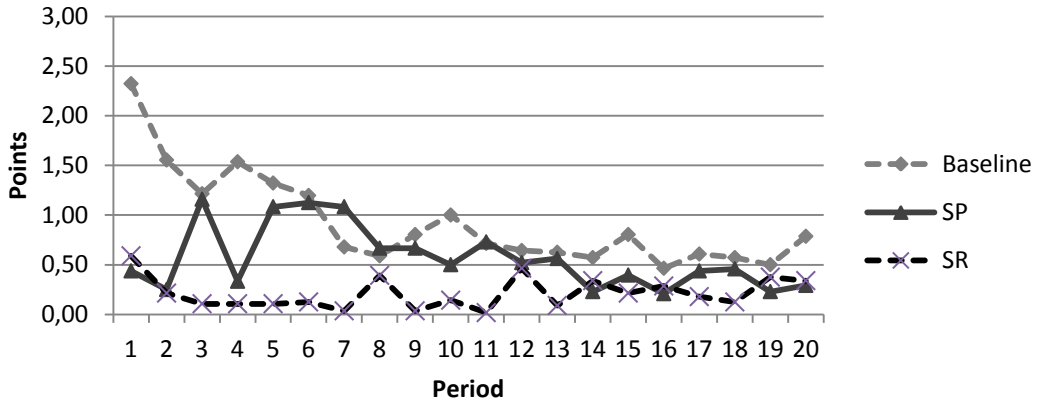

Figure 3. Average profits

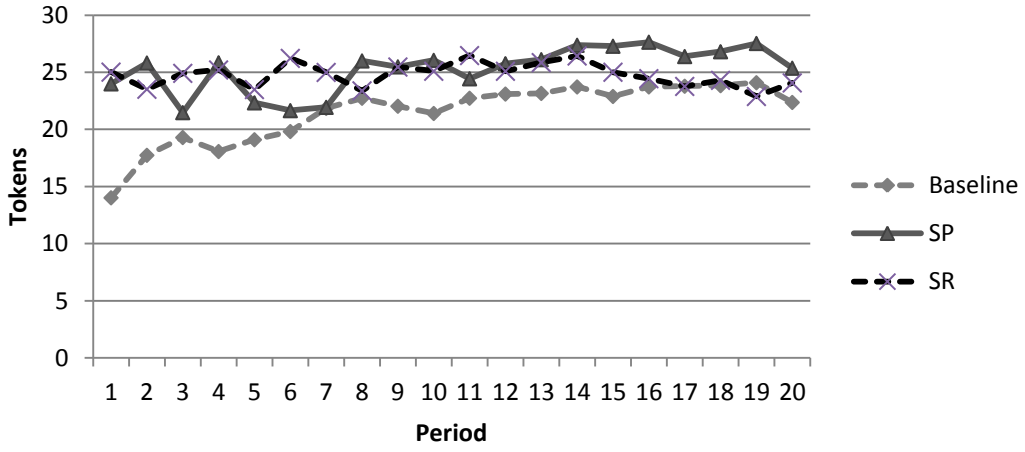

Table 1. Cost function

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

Table 2. Mean contributions

Group	Baseline	SP	SR
1	13.76	19.06	16.61
	(7.34)	(2.46)	(3.77)
2	18.40	11.86	11.55
	(3.31)	(3.90)	(2.70)
3	4.94	14.19	4.55
	(1.32)	(1.45)	(1.53)
4	11.30	11.29	5.48
	(4.03)	(1.58)	(3.35)
5	12.85	9.33	15.96
	(4.04)	(4.22)	(4.90)
6	4.58	17.95	2.45
	(2.95)	(2.71)	(1.86)
7	6.46	11.70	6.96
	(0.85)	(3.94)	(0.75)
8	2.18	16.25	4.84
	(0.72)	(2.05)	(0.68)
9	4.39	14.89	3.04
	(2.37)	(4.45)	(3.16)
10	1.64	17.85	6.51
	(0.94)	(2.93)	(1.83)
11	2.84	5.39	4.90
	(2.11)	(2.22)	(2.70)
12	15.13	17.31	16.44
	(4.77)	(3.61)	(3.10)
13	7.05	6.96	11.64
	(1.63)	(2.45)	(5.32)
14	11.11	12.75	14.63
	(2.95)	(2.43)	(3.35)
Mean	8.33	13.34	8.96

Table 3. Determinants of contributions

Contribution at t	Random Effect Tobit
Legitimacy	6.011***
	(1.511)
Solitary punisher	.910
	(1.471)
Constant	10.662*
	(5.516)
Log-likelihood	-8474.2768
Wald Chi(2)	34.34
N. Of obs.	3360

Legend: the dependent variable takes values from 0 to 20.

Legitimacy" is a dummy variable that takes value 1 when punishment is restricted, and 0 elsewhere. "Solitary Punisher" is a dummy variable that takes value 1 when only one subject is allowed to punish, and 0 elsewhere.
Controls: gender, age, faculty, nationality, and previous experiments in which the subject took part.
**** significant at 1%; ** significant at 5%; * significant at 10%.*

Table 4. Determinants of the increase in contribution levels

Contribution at t - contribution at t-1	Baseline	Solitary random (SR)	Solitary restricted (SP)
Points received at t-1	.270***	.519***	-.076
	(.076)	(.169)	(.149)
Distance from highest at t-1	-.196***	-.199***	.285***
	(.035)	(.030)	(.060)
Constant	1.211	1.391	-1.075
	(1.809)	(1.906)	(4.964)
Log-likelihood	-1569.7654	-1671.4109	-1884.5978
Wald Chi(2)	68.59	60.60	30.01
N. Of obs.	620	625	590

Legend: the dependent variable is defined takes values from -20 to 20.

The variable Distance from highest at t-1 *is defined as the difference between subject's contribution in t-1 and the highest contribution in the group in t-1, excluding subjects whose contribution was the highest in t-1.*

Controls: gender, age, faculty, nationality, and previous experiments in which the subject took part.

**** significant at 1%.*

Appendix A: Antisocial punishment

BASELINE

Group	(1) Points given by *i* to *j*	(2) Antisocial Contri<Contrj	(3) % of antisocial points
1	29	5	17.2%
2	42	0	0.0%
3	2	0	0.0%
4	45	12	26.7%
5	59	10	16.9%
6	72	29	40.3%
7	25	0	0.0%
8	45	2	4.4%
9	133	5	3.8%
10	181	48	26.5%
11	132	18	13.6%
12	15	0	0.0%
13	130	66	50.8%
14	123	7	5.7%
Total	1033	202 (19.5%)	
Mean	73.79	14.43	

SR

Group	(4) Points given by *i* to *j*	(5) Antisocial Contri<Contrj	(6) % of antisocial points
1	1	1	100.0%
2	14	13	92.9%
3	31	8	25.8%
4	40	15	37.5%
5	2	0	0.0%
6	25	2	8.0%
7	16	1	6.3%
8	17	4	23.5%
9	5	5	100.0%
10	37	7	18.9%
11	26	0	0.0%
12	36	19	52.8%
13	14	2	14.3%
14	12	6	50.0%
Total	276	83 (30%)	
Mean	19.71	5.93	

Appendix B. Members entitled and extracted. Treatment SP

Group	Name	Number of times entitled (1)	Number of times Extracted (2)	Quantity of points given (3)	Number of punishment episodes (4)	Average number of points given: 3/2 (5)
1	1	19	3	5	3	1.67
1	2	16	0	0	0	
1	3	19	1	0	0	0.00
1	4	20	0	0	0	
2	1	10	8	24	8	3.00
2	2	3	2	6	2	3.00
2	3	8	5	10	5	2.00
2	4	7	5	25	3	5.00
3	1	19	19	46	15	2.42
3	2	0	0	0	0	
3	3	1	1	0	0	0.00
3	4	0	0	0	0	
		20				
4	1	8	5	3	3	0.60
4	2	2	2	7	2	3.50
4	3	4	4	3	2	0.75
4	4	10	9	9	5	1.00
5	1	9	8	4	4	0.50
5	2	6	4	4	2	1.75
5	3	4	1	2	1	2.00
5	4	9	7	0	0	0.00
6	1	19	13	11	6	0.85
6	2	9	2	0	0	0.00
6	3	4	1	1	1	1.00
6	4	10	3	2	1	0.67
7	1	9	8	50	8	6.25
7	2	5	4	0	0	0.00
7	3	0	0	0	0	
7	4	8	8	161	7	20.13
8	1	19	10	0	0	0.00
8	2	11	6	25	6	4.17
8	3	0	0	0	0	
8	4	13	4	6	3	1.50
9	1	14	10	38	10	3.80
9	2	4	3	0	0	0.00
9	3	1	1	0	0	0.00
9	4	12	5	15	3	3.00
10	1	8	3	8	3	2.67
10	2	13	6	8	6	1.33
10	3	3	2	2	2	1.00
10	4	15	6	15	6	2.50
11	1	14	13	2	1	0.15
11	2	3	3	13	1	4.33
11	3	1	1	0	0	0.00
11	4	2	2	6	1	3.00
12	1	8	5	11	5	2.20
12	2	2	1	1	1	1.00
12	3	9	4	0	0	0.00
12	4	11	4	11	4	2.75
13	1	4	4	7	4	1.75
13	2	4	4	10	3	2.50
13	3	10	9	34	9	3.78
13	4	3	3	1	1	0.33
14	1	9	6	25	6	4.17
14	2	8	5	7	2	1.40
14	3	7	6	22	6	3.67
14	4	5	3	12	1	4.00

Appendix C: Instructions

Good morning, thank you for participating in this experiment. You are taking part into a study on economic decisions. During the experiment, you can, depending on your decisions and on other participants' decisions, earn a considerable amount of money in addition to the 5 euros you will receive anyway. The answers you give and the choices you make will be totally anonymous. The experimenters will not be able to associate your choices and your answers to your name.
During the experiment you cannot communicate with other participants (otherwise you would be excluded from the experiment) and you should be very careful in reading the instruction that will appear on your screen and will be read out by one of the experimenters. If you have any questions, please ask the experimenters.
Your earnings will be calculated in tokens; each token will be converted in euros at the following ratio: 1 token = 0,02 euros.
At the end of the experiment, you will be asked to fill a short questionnaire; afterwards, we will proceed with the payment, that will occur in cash.

THE PARTICIPANTS
In this experiment there are in total 20 participants, which are divided into 5 groups with 4 members each (*in the sessions with 16 participants, we have 4 groups with 4 members each*). The group composition will be the same for the whole experiment. Therefore you will always interact with the same three people, but you do not know their identity, and they do not know your own identity.

The experiment is composed by 20 rounds. In each round, the other 3 persons in your group will be randomly identified by means of numbers ("labels"). Note that the labels will be changing across rounds, and you will not be able to associate the choices made by a specific participant to a specific label. For example: in the first round the labels will be 1, 2 and 3. In the second round the labels will be 7, 5 e 11, in the third they could be 45, 2, 23 or 22, 32 and 11. But there is no relation among the different labels. The participant that is labeled with 1 in the first round, in the second could be indicated with 32, 54, or 33.

THE STAGES

Each round is composed by 2 stages.

During the first stage, you will decide how many tokens you will contribute to a "project". In the second stage, you will receive information on the number of tokens that the other 3 members of the group have decided to contribute to the project and therefore you will be able to reduce or not the earnings of any member of the group according to your and their levels of contribution at stage 1. This could be done by assigning points using your endowment of tokens. The following paragraphs will describe the experiment in detail.

STAGE 1

At the beginning of each round each participant will receive 20 tokens. We call this amount "endowment". Your task is deciding how to use your endowment. You have to decide how many tokens you want to use to contribute to the project and how many tokens you will keep for yourself. The number of the round appears in the left-top corner of the screen, whereas in the right-top corner you will see the amount of tokens you earned in the round and your total earnings up to that moment.

You have to decide how many tokens to contribute to the project by typing a number between 0 and 20 in the specific area (Figure 1). You can access to

that area by clicking with the mouse. After typing the number of tokens you want to contribute to the project, you should press [CONTINUE]. Once you have taken your decision, you cannot modify it.

Figure 1. Example of choice screen.

At the end of stage 1, you will be informed individually, on the screen of your computer, about your earnings, that consist of two elements:

a. The amount of the 20 initial tokens you kept for yourself (i.e. 20 tokens minus your contribution to the project);

b. Your payment deriving from the project, that is equal to the 40% of the sum of all the individual contributions to the project in your group (your contribution is included).

YOUR EARNINGS AT STAGE 1

Therefore, your earnings at the end of Stage 1 are calculated from the computer in the following way:

Your earnings after Stage 1 = (20 tokens – your contribution to the project) + 40% * (total contribution to the project)

Each group member's earnings are calculated in the same way; moreover each individual receives the same payment from the project. Assume, for instance, that in your group Member 1 will contribute 4 tokens, Member 2 will contribute 2 tokens, Member 3 will contribute 3 tokens and you will contribute 1 token. As a consequence, the total amount that the group will contribute is 10 tokens. Therefore, each member of the group will receive an amount equal to the 40% of 10 tokens = 4 tokens. The earnings of the 4 members of the group will be:

- Member 1: 20-4+4=20
- Member 2: 20-2+4=22
- Member 3: 20-3+4=21
- You: 20-1+4=23

As a further example, if Member 3 contributes 20 tokens, while you and Members 5 and 6 contribute 0 tokens, the whole amount of tokens that the group contributes is 20 tokens and each member of the group receives a payment from the project that is equal to the 40% of 20 tokens = 8 tokens. In this case, the earnings of the 4 members of the group will be:

- Member 3: 20-20+8=8
- Member 5: 20-0+8=28
- Member 6: 20-0+8=28
- You: 20-0+8=28

As a final example, if Member 13 contributes 0 tokens and you and Member 11 and 15 contribute 10 tokens each, the whole amount that the group contributes is 30 tokens and each member receives a payment from the project

that equals 40% of 30 tokens = 12 tokens. In this case, the earnings of the 4 members of the group will be:

- Member 11: 20-10+12=22
- Member 13: 20-0+12=32
- Member 15: 20-10+12=22
- You: 20-10+12=22

STAGE 2

At the beginning of Stage 2, you can see how much the other members of the group have contributed to the project and which is the average level of contribution in the group.

Only the participant whose contribution has been **the highest in the group** (*only for the SP Treatment; in the SR Treatment: only one participant selected at random; no restrictions hold in the Baseline Treatment*) will go on with the experiment and take part into Stage 2. The other participants have to wait the next round and will be updated on the amount of their earnings as soon as the other members will have taken their decisions.

In this stage, if your contribution is the highest in the group, you can reduce or let unchanged the earnings of the members of the group (you can assign up to 10 points). Any other member of the group can, if she like, reduce your earnings if her contribution is the highest in the group. Each point reduces the earnings at stage 1 of the participants who receive them by 10%. You have to type the number of points you want to assign to each member of the group whose contribution has been lower than yours. If you choose to assign 0 points to a specific member of the group, you do not modify her earnings. If you choose to assign 1 point to a specific member of the group, you reduce her earnings of 10%. The amount of points you assign to each member of the group determines the amount you reduce their earnings at stage 1. If an

individual receives 4 points, her earnings will be reduced by 40%, and if she receives 10 or more points her earnings will be reduced by 100%.
If you assign points, you face a cost that depends on the number of points you distribute in total. The table below shows the relation between the number of points you assign to a participant and the cost you incur:

# points assigned	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Cost you incur	**0**	**1**	**2**	**4**	**6**	**9**	**12**	**16**	**20**	**25**	**30**

For example, in your group you contributed 18 tokens. The other members (Members 1, 2 e 3) have contributed 10, 15 and 5 tokens respectively. Therefore, your contribution is the highest in the group and you can decide to assign points to all other members incurring the cost indicated in the table above, or keeping their earnings constant by assigning 0 points (and this will cost 0). If you assign, for instance, 0 point to Member 1 and 3 points to Member 3, the earnings of Member 1 at Stage 1 will be constant whereas the earnings of Member 3 will be reduced by 30%. As the number of points you assigned at Member 3 is 3 and you assigned points only to her, the cost you incur in total is 4 tokens.
When you have taken your decision, click on [CONTINUE].

YOUR EARNINGS AT STAGE 2
Your earnings at the end of Stage 2 will be calculated by the computer in the following way:

If your contribution has been the highest of the contributions in the group:
Your earnings at the end of Stage 2 = Your earnings at the end of Stage 1 - cost of the points assigned at Stage 2

If your contribution has not been the highest in the group:

Your earnings at the end of Stage 2 = Your earnings at the end of Stage 1 - points you received *10%* (Your earnings at the end of Stage 1)

Please note that at the end of the second stage your earnings could be negative. This occurs when the cost of the points you decided to assign is higher than your earnings at the first stage. In general, you can avoid this by paying a little attention.