UNIVERSITY OF SIENA
DEPARTMENT OF ECONOMICS

LabSi
EXPERIMENTAL ECONOMICS LABORATORY

Marco Casari e Luigi Luini

GROUP COOPERATION UNDER ALTERNATIVE PEER PUNISHMENT TECHNOLOGIES: AN EXPERIMENT

2/2005

WORKING PAPER

Abstract - We study experimentally group cooperation under alternative peer punishment institutions. We find that (1) information about how much others have punished are irrelevant for punishment decisions; (2) when a coalition of at least two subjects in the group must agree before somebody is punished, group performance dramatically increases; we show that this institution reduces perverse punishment and encourages pro-social behavior. Furthermore, the results suggest that the strong emotional component that drives agents to punish alters their ability to behave strategically

JEL Classification: C91, C92, H41, D23

Keywords: public goods, peer punishment, social norms, team production, experiments.

Marco Casari, Krannert School of Management, Purdue University, USA ; e-mail: casari@purdue.edu
Luigi Luini, Department of Economics, Uiversity of Siena, Italy; e-mail: luini@unisi.it

While peer punishment has been shown to increase group cooperation, there is open debate on how cooperative norms can emerge and on what motives drive individuals to punish. In this paper we address both topics by experimentally comparing alternative institutions of peer punishment.

When he welcomes the new employees, the CEO of Diesel introduces them to the company culture with the following words, “If you do not show competence in what you do, they all will walk over you. It will not be me to throw you out, but your very own peers.” [1] We ignore the ways in which this threat is carried out. However, this strong work ethic has contributed to make Diesel an international success case among clothing companies. Forms of peer punishment are reported in many other field settings, including fishing communities where people doing excessive harvesting are the target of vandalisms of boats and fishing nets (Ostrom, 1996). Mining communities also use peer punishment such as social ostracism to discourage others from breaking strikes over working conditions (Francis, 1985).

While peer punishment in field contexts takes many forms, the experimental economic literature has focused on the very specific punishment technology that we call “Baseline,” where group members *simultaneously* choose if and how much to punish each member of the group without ever knowing the punishment choices of others and with wide discretionality on punishment choices (Ostrom et al., 1992, Fehr and Gaechter, 2000, 2002, Sefton et al., 2002, Bochet et al., 2002, Masclet et al., 2003). Important implications have been drawn from this punishment structure, but there are several reasons to examine alternative peer punishment technologies. First, variations in the technology of peer punishment may affect group

[1] From an interview with the CEO of Diesel Renzo Rosso: “Look, here all the doors are open to you. You can climb the ladder or change tasks as much as you wish. But watch out: if you do not show competence in what you do, they all will walk over you. It will not be me to throw you out, but your very own peers,” Stella (1996), pp.24-26, (our translation) Civil servants in some public administrations have equally strict norms but of the opposite sign, where shopping during office hours is allowed and hard working habits are discouraged.

performance.[2] We explore, for given social norms, how changing the punishment technology may lead to more or less group cooperation. If a manager can shape the way team members interact, it may be possible for the manager to increase total production without manipulating the workers' social norms. Second, alternative peer punishment technologies may shed light on the motives to punish. Third, it provides a check for robustness for existing studies. So far the only systematic exploration of the above peer punishment structure concerns changes in the relative cost of punishment (Carpenter, 2002, Andreoni et al., 2003, Putterman and Anderson, 2003).[3] Our paper alters the structure of the technology along multiple dimensions.

In our experiment, subjects can voluntary contribute to a public good. There are treatments without punishment or with one of the following punishment technologies, "Baseline," "Sequential," or "Consensual." When agents are motivated purely by personal monetary earnings, there should be no punishment and complete free riding in all treatments. We know from previous studies that this is not an accurate description of experimental results, but scholars have not settled on a single explanation. The Baseline technology implements a design similar to what is common in the literature and constitutes the benchmark to measure results from other technologies. Studying alternative punishment technologies, reveals, among other things, how subjects enforce their norms of cooperation.

Each person may bring her norms into the lab, i.e. standards of behavior about how individual group members ought to behave in a given situation. Hence, a group may face multiple norms, possibly conflicting ones. In the Baseline technology all requests to punish are carried out and the group outcome results from the mixed impact of all individual norms. Instead, in the

[2] We do not touch the questions concerning the endogenous selection of a punishment institutions or how individuals self-select into a particular punishment institution. A fashinating study in this regard is Gürerk et al. (2004).
[3] For an interesting exception see Decker et al. (2003).

Consensual technology requests to punish are carried out only when there is a minimum coalition of two agents that share a norm. More precisely, at least two members in a group of five must agree on the individual(s) to target for punishment, although they can still disagree on the severity of the punishment. Similarly to the Baseline technology, individual norms are not imposed by the experimenter but are those of the participants. To some extent, the Consensual technology resembles to a legal system where laws are chosen or enforced only through social agreement. Group performance substantially improves under the Consensual peer punishment institution hence generating a “consensual dividend”. Both group cooperation and group net earnings are considerably higher under the Consensual than any other punishment rule studied.

In addition, we vary the information that subjects have about how much others have punished. In the Baseline treatments, group members decide simultaneously how much to punish a specific agent; in the Sequential treatment those decisions are taken one group member at a time. The group members that decide later on in the sequence learn about the cumulative amount of punishment already inflicted on that specific agent. They can decide how much additional punishment to assign. One may expect that this richer information set could have reshaped punishment choices, for example by improving coordination in punishing free riders. Instead, choices are surprisingly similar to the Baseline treatment. We call this the “coordination puzzle.” Possible interpretations of this result are put forward.

This paper is structured into four Sections. In Section 1 we describe the experimental design and the predictions. Aggregate results are presented in Section 2, while individual punishment decision results are presented in Section 3. Conclusions follow in Section 4.

1. The Experimental Design

1.1. Basic Design

Our design consists of a public good experiment with three treatments of differing punishment technologies.[4] There are N=20 participants in each session. In every period the participants are randomly partitioned into four groups of n=5 individuals. In all treatments, subjects participate for twenty periods in a finitely repeated public good game with and without punishment opportunities. In the first ten periods there is no punishment opportunity while in the last ten periods there is. Punishment opportunities are structured in three different ways: Baseline, Consensual, and Sequential (Table 1).

In the Baseline treatment, once group members are informed about each member's contribution to the public good, all punishment requests are simultaneously submitted. At a private cost of one token per punishment point, an agent can decrease the earnings of any other individual in her group by three tokens. In the case an agent receives punishment points from two or more agents, her earnings reduction is the cumulative effect of all requests. Punishment on a targeted agent is carried out irrespectively of the number of requests. This is a common protocol in the experimental literature, adopted, for instance, by Fehr and Gaechter (2000).

In the Consensual treatment, participants simultaneously place their punishment requests. If a subject is the target of the punishment request of just one other subject, that punishment request is ignored. When at least two group members request to punish a subject, their punishment requests are carried out. Subjects are informed of the outcome of their requests.

Finally, in the Sequential treatment, agents decide, in random order, one after another, on their punishment requests. As before, all requests are carried out, but every participant considers

[4] The Instructions for the Consensual treatment can be found in Appendix. The whole experiment was framed in neutral terms.

punishing each one of the other (n – 1) group members in (n – 1) separate steps. In step one, each subject places a punishment request. In step two, a subject knows that one other group member has already had the opportunity to punish the same individual and knows the cumulative amount of punishment already inflicted. The process continues for the remaining steps. Punishment points can be added but never subtracted.

The treatment without punishment opportunity serves as a control for the treatment with punishment opportunity. The same n subjects interact ten periods without punishment opportunities then ten periods with the opportunity to punish.[5]

1.2. Payoffs

In the treatments without punishment, in each period each of the n subjects in a group receives an endowment of y tokens. A subject can either keep these tokens for herself or invest g_i tokens ($0 \le g_i \le y$) into a project. The decisions about g_i are made simultaneously. The period monetary payoff for each subject i in the group is given by

$$\pi_i^1 = y - g_i + a \sum_{j=1}^{n} g_j \qquad (1)$$

where a is the marginal per capita return from a contribution to the public good, $1/n < a < 1$. The total payoff from the no–punishment condition is the sum of the period–payoffs, as given in (1), over all ten periods. Note that (1) implies that full free–riding ($g_i = 0$) is a dominant strategy in the stage game. This follows from $\partial \pi_i^1 / \partial g_i = -1 + a < 0$. However, the group payoff $\sum_{i=1}^{n} \pi_i^1$ is maximized if each group member fully cooperates ($g_i = y$) because $\partial \sum_{i=1}^{n} \pi_i^1 / \partial g_i = -1 + na > 0$.

The major difference between the no–punishment and the punishment conditions is the addition of a second decision stage after the Simultaneous contribution decision in each period. At the

[5] We ran three additional sessions where the ten periods with punishment opportunities were placed before the ten periods without the punishment opportunity. These results are not reported here.

second stage, subjects are given the opportunity to punish each other after they are informed about the individual contribution of the other group members. Group member *j* can punish group member *i* by assigning so–called punishment points p^i_j to *i*. There are three different treatments for the part with punishment opportunities.

In the Baseline (Simultaneous) treatment for each punishment point assigned to *i* the first–stage payoff of *i*, π_i^1, is always reduced by three tokens. Agent *i* takes all punishment decisions at once { $p^1{}_i,\ldots, p^{i-1}{}_i, p^{i+1}{}_i,\ldots, p^n{}_i$}, where $p^k{}_i \in\{0,1,\ldots,7\}$, and simultaneously with the other agents. For received punishment points, agent *i*'s payoff is reduced by $\sum_{k\neq i}^{n} e(p_k^i)$, where $e(p^i_j)= 3\, p^i_j$ is the effectiveness of punishment function. For punishment points given to others, agent *i*'s payoff is reduced by $\sum_{k\neq i}^{n} c(p_i^k)$, where $c(p^j{}_i)= p^j{}_i$ is the cost of punishment. This design has the important feature of holding the fine-to-fee ratio $e(p^i{}_k)/\, c(p^i{}_k)$ constant – and equal to 3 – in order not to alter the "price" of punishment.[6] The pecuniary payoff of subject *i* from both stages, π_i , can therefore be written as:

$$\pi_i = \pi_i^1 - \sum_{k\neq i} e\left(p_k^i\right) - \sum_{k\neq i} c\left(p_i^k\right) \qquad (2)$$

The total payoff from the punishment condition is the sum of the period–payoffs, as given in (2), over all ten periods.

In the Consensual treatment, punishment is Simultaneous and employs the same cost and effectiveness functions as before. However, an agent is punished only if at least two agents requested it. Hence, the payoff function for both stages is:

[6] The reason is to avoid confounding effects in the interpretation of results due to differential "pricing" of punishment. For the same reason we allowed period earnings of a subject to be negative. Not doing so would have increased the fine-to-fee ratio of the marginal punisher. In the experiment, negative period earnings were infrequent. When ignoring the punishment given to others, it amounts to 3.3% of the observations with punishment opportunities. Cumulative earnings were always positive.

$$\pi_i = \pi_i^1 - K(i)\sum_{k\neq i} e\left(p_k^i\right) - \sum_{k\neq i} K(k)c\left(p_i^k\right) \qquad (3)$$

where $K(i) = 1$ if $\left(\sum_k I_{\{i,k\}}\right) \geq 2$ and $K(i) = 0$ otherwise. The function $I(i,k)$ equals one when agent k requests to punish agent i, $p^i_k > 0$, and equals zero otherwise. In practice, only a coalition of 40% of group members or larger is allowed to punish a member. Isolated requests to punish agent i have no effect and no cost is charged for that request. If the punishment request is not carried out then the requesting subject is informed about it and the targeted subject will not know of such request.

In the Sequential treatment the payoffs are given by (2) like in the Baseline treatment but the timing of decisions is different. Instead of being taken all at once, punishment decisions $\{ p^1_i,\ldots, p^{i-1}_i, p^{i+1}_i,\ldots, p^n_i\}$ are broken down into $(n\text{–}1)$ distinct steps where at step k agent i makes a single decision $p^{j(k)}_i$. The order of punishment decisions $j(k)$ is random and such that within the period agent i has an opportunity to target all other agents in the group, $\{j(1),\ldots, j(n\text{-}1)\}=\{1,\ldots, i\text{-}1, i+1,\ldots,n\}$. After each step, there is an update on the cumulative punishment received by each agent in the group.

B. Parameters and Information Conditions

The experiment is conducted in a computerized laboratory where subjects anonymously interact with each other.[7] No subject is ever informed about the identity of the other group members. No communication among subjects is allowed. In all treatment conditions the endowment is given by $y = 20$, groups are of size $n = 5$, the marginal payoff of the public good is fixed at $a = 0.4$, and the number of participants in a session is $N = 20$. In each period subject i can assign up to seven punishment points p_i^j to each group member j, with $j \neq i$ irrespective of

[7] For conducting the experiments we used the experimental software "z-Tree" developed by Urs Fischbacher (1998).

their first stage earnings. In all treatment conditions subjects are publicly informed that the condition lasts *exactly* for ten periods. When subjects play the no–punishment opportunity condition they know that a session consists of two conditions but do not know the rules for the second condition. After period ten of the first condition in a session, they are informed that there will be a "new experiment" and that this experiment will again last exactly for ten periods. They are also informed that the experiment will then be definitely finished.[8]

In the no–punishment conditions the payoff function (1) and the parameter values of y, n, N, and a are common knowledge. At the end of each period subjects in each group are informed about the total contribution Σg_j to the project in their group.

In the punishment conditions the payoff function (2) or (3), in addition to y, n, N, a, and the protocol of the punishment requests are common knowledge. Furthermore, after the contribution stage subjects are also informed about the whole vector of individual contributions in their group. To prevent the possibility of individual reputation formation across periods each subject's own contribution is always listed in the first column of his or her computer screen and the remaining four subjects' contributions are listed without subject ID in the other four columns. Thus, subject i does not have the information to construct a link between individual contributions of subject j across periods. Therefore, subject j cannot develop a reputation for a particular individual contribution behavior. This design feature also rules out that i punishes j in period t for contribution decisions taken in period $t' < t$. Subjects know their own punishment activities, the aggregate punishments imposed on them by the other group members, and the *aggregate* punishment imposed on *other* group members.[9]

[8] Each condition was preceded by a trial period to familiarize the subjects with the software.
[9] This provision can make a difference when subjects do not know the preferences of others. When a subject can only observe the punishment points she gave or received (Fehr and Gaechter, 2000), learning about these

In the Sequential punishment treatment subjects know at what step they are and the cumulative *aggregate* punishment imposed on each *other* group members up to the previous step. Hence, they receive more detailed information about punishment than in the Baseline treatment because they can see both the end-of-period sum and some disaggregated statistics about the individual components of this sum. However, they are not informed about the amount of punishment they have personally received until the end of the period. In all treatments, subjects are also not informed about the *individual* punishment requests of the other group members. Both provisions are meant to prevent, as much as possible, a subject from using punishment to payback others for their requested punishments.

1.3. Predictions

The canonical predictions for the experimental conditions just outlined are well known. If subjects apply the backward induction logic, the equilibrium prediction in all three treatments is that all subjects will contribute nothing to the public good and will punish nothing. In fact, choosing $p^j_i>0$ is a monetary cost that does not generate any monetary benefit in a one-shot interaction. In the experiment the probability that an agent was re-matched with the same four people was less than 2 percent.

2. Aggregate Results

A total of 240 subjects were recruited among the general undergraduate student population of the University of *(omissis)* via ads posted around campus asking to email or call. No subject had participated in this type of experiment before, and each subject participated in only one of the experimental sessions. Twelve sessions were conducted between March and October 2003. Each

---

preferences may be slower than here. In our setting, a subject can see if a social norm was enforced with respect to any other subject in her group.

session lasted between 1 hour and 50 minutes and 2 hours and 30 minutes including instructions. Payment was done privately in cash at the end of each session and totaled 12.40 euros per subject on average.[10]

The results are grouped into two sections, this one referring to aggregate cooperation and net payoff (Results 1-3) and the next one concerning individual decisions to punish (Results 4-5).

When subjects have the opportunity to punish, contributions to the public good increase (Result 1) and stay high over time (Result 2). In addition to replicating these well-known results, we also find sharp differences regarding group cooperation levels and group net earnings according to the punishment technology used. In particular, group cooperation (contribution) and group net earnings are highest in the Consensual treatment (Result 3).

RESULT 1: *The existence of punishment opportunities causes a rise in the average contribution level from 17% to 29% of the endowment. In particular, while the average contribution rises in all treatments, the rise is largest in the Consensual treatment.*

RESULT 2: *In the no-punishment condition average contributions converge over time close to full free riding. In contrast, in the punishment condition average contributions are stable or increasing over time. In particular there is a steady growth in contribution levels in the Consensual treatment.*

Support for Results 1 and 2 comes from Table 2 and Figure 1. Without a punishment opportunity the average individual contribution across all treatments is 3.31 tokens. When the opportunity to punish is introduced, the average individual contribution across all treatments is 5.77. A nonparametric Wilcoxon signed ranks test shows that this difference in contributions is significant at the one percent level (p=0.0061). These average values hide a declining trend when

[10] At the October 2003 rate of $1.17 per euro, it is equivalent to $14.50. This amount includes the show up fee that was 3 euros for the four sessions conducted before October and 5 euros afterwards. The amounts in the instruction were quoted in "Tokens." A token was converted into 0.02 euros.

there are no opportunities to punish - from 5.92 tokens in period one to 1.82 in period ten. With punishment opportunities there is a "jump" in period eleven when there is an average contribution of 5.21 tokens and an ascending trend to 6.50 in period twenty. This jump in contribution between the last period without punishment and the first period with punishment is significant at a one percent level according to a Wilcoxon signed ranks test (p=0.0002).

Besides these common patterns, each punishment technology shows remarkable peculiarities. Overall contributions under a Consensual technology are substantially higher than in the other two (8.46 vs. 4.46 Baseline and 4.38 Sequential).[11] Moreover, while the time trend is increasing for the Consensual technology (period one-ten, 6.94-9.76), it is roughly stationary for the other two (4.01-5.65 Baseline, 4.62-4.10 Sequential). Such differences are summarized by the analysis of relative payoff gains with and without punishment in Result 3.

RESULT 3: *In all treatments punishment opportunities initially cause a relative payoff loss. The Consensual treatment is the only treatment in which relative payoff gains are found, more precisely in the final four periods. In the final period of the Consensual treatment the relative payoff gain is 13 percent. In the Baseline and Sequential treatments the relative payoff losses remain throughout all periods, although they become smaller over time. In the final period of the Baseline and Sequential treatments the relative payoff loss is roughly 20 percent.*

Support for Result 3 comes from Table 2 and Figure 2. Normalizing the earnings in the final period of the no punishment condition to 100, then earnings in the first period with punishment are equal to 57 in the Baseline treatment, 53 in the Sequential, and 85 in the Consensual. By the end of the session, all of these values have increased. While the Baseline is at 80 and the Sequential is at 78, which are still below the reference value without punishment, the Consensual

[11] A non parametric Wilcoxon signed ranks test shows that the difference in contributions with and without sanction opportunities between the Consensual treatment on one side and the other two treatments on the other side is significant at a ten percent level (p=0.0768).

treatment is above, at 113. A Wilcoxon signed ranks test shows that the differences in group net earnings between the last period with and without sanction opportunities are significantly different in the Consensual treatment compared to the other two treatments at a five percent level (p=0.0364).

Interestingly, the high cooperation level of the Consensual treatment was achieved with the lowest level of punishment among all treatments. This is the key point that accounts for its superiority in terms of group net earnings. Let us define the "punishment rate" as the number of punishment points assigned to a particular contribution action. The average punishment rate is $\sum_{t=1}^{10}\sum_{j}^{n}\sum_{k\neq j}^{n} p_{k,t}^{j} / 10\cdot n$ for Baseline and Sequential and $\sum_{t=1}^{10}\sum_{j}^{n} K(j)\sum_{k\neq j}^{n} p_{k,t}^{j} / 10\cdot n$ for Consensual. The average punishment rate was 1.70 in the Consensual compared to 2.47 in the other two treatments (Figure 3). This difference persists after adjusting for the variations in group contribution across treatments (Table 3). For any given contribution level, lower punishment rates translate into a smaller deadweight loss. One reason for the lower punishment rate is that all punishment requests made by just one agent were ignored. Had those requests not been ignored, the punishment rate in the Consensual treatment would have been 29.5 percent higher than our reported rate (full sample, Figure 3).

What cries for an explanation is how a lower threat of punishment observed in the Consensual treatment could provide not weaker but *stronger* incentives to cooperate that in the other treatments. The reason is that the Consensual technology endogenously filtered out the anti-social norm of a minority that was targeting cooperators, thus enhancing the incentives to cooperate. While less than one out of every ten requests to target full free-riders was censored, more than seven out of ten attempts to punish strong cooperators with contributions (15,20] were blocked. Table 3 provides more details.[12] The Baseline and Sequential technology instead allowed a minority to freely harm strong cooperators and hence lower incentives for cooperation.

To provide additional statistical evidence for this explanation and to facilitate the comparison with Fehr and Gaechter (2000) we also present a regression analysis of punishment behavior. As a complement to the use of *absolute* contribution levels employed in Table 3, this analysis also captures the effect of punishment on subjects' *relative* contributions in respect to the group average. Table 4 contains a model and an ordinary least-squared (OLS) regression where the dependent variable is "received punishment points" of a subject and the independent variables comprise "strong cooperator", "others' average contribution", "positive deviation" and "absolute negative deviation", respectively. The latter variable is the absolute value of the actual deviation of a subject's contribution from the others' average in case that his or her own contribution is below the average. This variable is zero if the subject's own contribution is equal or above the others' average. The variable "positive deviation" is constructed analogously. The variable "strong cooperator" is one if the subject's contribution is above fifteen tokens and zero otherwise. This variable retains the absolute scale of the contribution level and may capture the tendency, mentioned above, to target highly cooperative subjects. The model also includes period and session dummies. In all treatments, the coefficient of the "absolute negative deviation" is positive and significant at the one percent level. This result reinforces the conclusion that free riders can reduce the received punishment by increasing their contributions. Although the positive coefficient is not significant, in the Baseline and Sequential treatment, strong cooperators were targeted for punishment. However, in the Consensual treatment, strong

[12] Results in Table 3 are for a subsample only. The data support this point even more strongly when one considers the full sample.

cooperators were less likely to receive punishment (significant at the 10 percent level). This result holds when controlling for the relative contribution with respect to the group.

What stands out in the analysis of group cooperation levels across treatments is the superiority of the Consensual technology. This technology generated punishment costs 10 percent lower that the Baseline technology and realized a contribution level 90 percent higher (Result 3).

3. Individual punishment decisions

We now turn to a comparison of the patterns of individual punishment decisions across treatments. We begin presenting Result 4 about the multiplicity of punishment requests in the Baseline treatment. Result 5 provides new insights for the comparison between Baseline and Sequential treatments.

RESULT 4: *In the Baseline treatment, approximately half of the times that a subject is punished, two or more subjects have requested the punishment.*

Support for Result 4 can be found in Table 5.[13] To interpret this result, it is important to consider that in about 92% of the instances one group member could carry out single-handedly the whole punishment. A subject could distribute up to seven points of punishment to another subject. A total of eight or more points were distributed to a subject only in 8.2% of the cases. It follows that the multiplicity of requests to punish the same agent is not a response to the need to punish free riders more severely because almost always one agent could have done it alone.

One possible explanation for the multiplicity of punishment is a coordination failure when subjects have the following preference structure. Let us make the assumption that punishment is

[13] Inflicting seven points of punishment has a considerable impact on the earnings of an agent, namely a reduction between 40% and 105%. Seven points of punishment reduces earnings by 21 tokens. If everybody free rides, g_i=0 for all i=1,2,3,4 then π_i^1=20. If one agent free rides, g_1=0, and all others fully cooperate, g_i=20 for i=2,3,4, then π_1^1=52 and π_i^1=32 for i=2,3,4.

a "second order public good" because the utility derives from having an agent punished. Consider a situation where there are agents that are willing to punish free riders if no one else does. [14] Under the above assumption, an agent of the above type will happily free ride on punishment if she knows that somebody else will punish. When punishment decisions are simultaneous, there is a strategic element in the decisions that relies on the knowledge of the preferences of others in the group. Either the agents don't know exactly the preferences of others ore they do. If they don't know, then Result 4 does register a coordination failure. Had they had the information, subjects would have punished differently. If they do know, instead, then the preference structure is probably different from the one described above.

Another possible explanation is that a subject's decision to punish does not depend on how much others punish the same subject. Stated differently, a subject gains utility from his/her personal punishment action and does not care about the total amount of punishment received by the targeted subject. This preference structure could describe a strong emotional drive in the motivations for punishment. In that case no strategic element would enter into the punishment decision and the multiplicity of punishment requests would cease to be a puzzle. It would simply reflect the plurality of subjects in each group with a preference for punishment. Moreover, under such preference structure, the procedural differences between Baseline and Sequential would be irrelevant.

Results 5, which will be now presented, provides more support for this "emotional" interpretation of punishment decisions than for the "second order public good" interpretation.

[14] Neurological evidence suggests that sanctioning a norm violator may be a source of utility for the punisher (de Quervain et al., 2004).

RESULT 5: *In the Sequential treatment there is no improvement in coordination in punishment in comparison with the Baseline treatment. In particular, we observe across treatments similar frequencies of multiple requests to punish the same subject and similar distributions of total punishment received by free riders.*

Result 5 has two components, the multiplicity of requests on the same target and the total punishment received by a subject. The similarity in the multiplicity of requests to punish is detailed in Table 5. If there is uncertainty in the punishment preferences of others in the group, coordination in punishment decisions would be more difficult in the Baseline than in the Sequential treatment. In particular, the later movers in the Sequential treatment have additional information on the punishment already distributed and could coordinate more easily. As a consequence, one may expect in the Sequential data a lower number of group members targeting the same agent than in the Baseline. Empirically, that does not seem to be the case.

The evidence on total punishment received by free riders provides further support in the same direction. When two or more subjects in a group are willing to punish there could be a problem in coordinating punishment. If there is an improvement in coordination, one would expect to see in the Sequential results less variability in the punishment received by free riders, i.e. a reduction in the number of free riders escaping punishment or receiving extemely high punishments. Also this conjecture about an improvement in coordination is not supported in the data. We present data relative to groups with two or more members who contribute positive amounts *and* with at least one complete free rider (zero contribution). These situations are very common as they account for 68.1% of the groups in the Baseline and 70.6% in the Sequential treatment. Typically, the complete free riders receive a heavy punishment, an average of 4.83 points in the Baseline and of 4.32 points in the Sequential. The actual punishment does vary widely in level from 0 to 19 points. Yet, the empirical distributions of the punishment points targeting complete free riders are surprisingly similar between Baseline and Sequential treatment. A Kolmogorov-Smirnov test for equality of distribution functions cannot reject the equality hypothesis (p-value of 0.336, no. obs.=183, 219). We conclude that the additional information provided in the Sequential compared to the Baseline treatment does not significantly change either the level or the dispersion of punishment decisions (the standard deviation in punishment points actually grows slightly from 3.46 in the Baseline to 3.85 in the Sequential). To sum up, there is a puzzle here, as the additional information available in the Sequential treatment seems to be ignored.

4. Conclusions

We study group cooperation in the provision of a public good under three peer punishment institutions, where agents have a costly opportunity to decrease the earnings of others in the absence of any personal material benefit. While this study replicates and confirms the robustness of the qualitative results of other experiments (Ostrom et al., 1992, Fehr and Gaechter, 2002, Sefton et al., 2002), it also points to the significant impact of the specific punishment institution. There are three major conclusions.

First, the consensual form of peer punishment performs remarkably better than the others. We document that when a group member can be targeted for punishment only with the agreement of a coalition of agents, the group contributes and earns more than when each agent has free discretionality on whom to punish. We call this the “consensus dividend.” This study has identified one specific set of rules that promotes a strong effect on group cooperation. It would be interesting to characterized the whole class of rules.

Second, introduction of the opportunity to punish has the potential to lower or to increase group net earnings. We find that in two out of three treatments the ability to punish lowers net earnings. This loss is especially pronounced during the periods of transition directly after the opportunity to punish is provided. One concludes that peer punishment does not automatically benefit the group. It crucially depends on the institution that govern peer punishment. Anthropological studies of societies without a judicial system have pointed to the danger of the spontaneous human tendency to engage in peer punishment (Lowie, 1970, Girard, 1977, p.16-22). Our findings provide indirect support for the role of a legal system in the administration of punishment. Legal systems restrict sanctioning to the violation of shared rules and censor individual attempts to punish socially virtuous actions, hence channeling agents' punishment attitudes toward beneficial ends for society (Kosfeld and Riedl, 2004; Casari and Plott, 2003).[15] More studies are needed to explore the behavioral foundations of punishment through legal systems.

Our third conclusion concerns the motivations that drive agents to punish. We report of a "coordination puzzle." Changes in strategic incentives and information levels from the Baseline to the Sequential treatment are surprisingly irrelevant in peer punishment behavior. This similarity in punishment pattern is considered a puzzle because when agents care about the total punishment that another agent receives, they should have no objections to others doing the "dirty job" of punishing. They should actually prefer it because it saves them the punishment cost. In particular, the sequential protocol should favor a better coordination in punishment choices. Instead, no significant differences in punishment patterns were uncovered. Such evidence

[15] In line with most of the experimental literature on punishment, this study did not reveal the identity of the punisher nor permit counter-punishments, hence preventing revenge. Allowing for it would have further exacerbated group conflict and brought up aspects of peer punishment that are detrimental for group performance (Nikiforakis, 2004; Masclet et al., 2004).

supports an interpretation of peer punishment as an emotional response to a norm violation that involves no strategic considerations. According to this interpretation subjects do not perceive punishment as a second-order public good. When it comes to other-regarding attitudes, emotions seem to alter the ability of people to behave strategically.

References

Andreoni, J., Harbaugh, W., and Vesterlund, L. (2003). "The Carrot or the Stick: Rewards, Punishment and Cooperation," *American Economic Review*, 93, 3, 893-902.

Bochet, O., Page, T., and Putterman, L. (2002). "Communication and Punishment in Voluntary Contribution Experiments," Brown University, Department of Economics, *Working Papers no. 2002-29.*

Carpenter, J. (2002). "The Demand for Punishment," *Working Paper 0243*, Middlebury College, Department of Economics

Casari, M. and Plott, C.R. (2003). "Decentralized Management of Common Property Resources: Experiments with Centuries-Old Institutions," *Journal of Economic Behavior and Organization*, 51, 2, 217-247.

de Quervain, D. J.-F.., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004). "The neural basis of altruistic punishment," *Science*, 305, 1254-1258.

Decker, T., Stiehler, A., and Strobel, M. (2003). "A Comparison of Punishment Rules in Repeated Public Good Games." *Journal of Conflict Resolution*, *47*, 6, 751-772.

Fehr, E. and Gaechter, S. (2000). "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 2000, 90, 4, pp. 980-994.

Fehr, E. and Gaechter, S. (2002): "Altruistic Punishment in Humans", *Nature*, Vol.415, 137-140.

Fischbacher, U. (1998). "z-Tree: Zurich Toolbox for Readymade Economic Experiments. Instructions for Experimenters." Mimeo, University of Zurich.

Francis, H. (1985) "The Law, Oral Tradition and the Mining Community," *Journal of Law and Society*, Winter, 12, 3, 267–71.

Girard, R. (1977). *Violence and the Sacred*, Johns Hopkins University Press.

Gürerk, O., Irlenbusch, B., and Rockenbach, B. (2004). "On the Evolvement of Institution Choice in Social Dilemmas,“ mimeo,University of Erfurt.

Kosfeld, M., and Riedl, A. (2004) The Design of (De)centralized Punishment Institutions for Sustaining Cooperation, *Tinbergen Institute Discussion Paper*, TI 2004-025/1

Lowie, R. (1970). *Primitive Society*, New York, W W Norton & Co.

Masclet, D., Denant-Boemont, Noussair, C. (2004). Public Good Game with Sanctions and Metanorms, *working paper*.

Masclet, D., Noussair, C., Tucker, S., and Villeval, M.-C. (2003). “Monetary and Nonmonetary Punishment in the Voluntary Contribution Mechanism,” *American Economic Review*, 93, 1, 366-380.

Nikiforakis, N. S. (2004). “Punishment and Counter-punishment in Public Goods Games: Can we still govern ourselves?,” March, *Working paper,* University of London, Royal Holloway, Department of Economics.

Ostrom, E. (1996). *Governing the commons : the evolution of institutions for collective action*, Cambridge, Cambridge University Press.

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

Putterman, L. and Anderson, C. M. (2003). “Do Non-strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism,” Brown University, Department of Economics, *Working Papers no. 2003-15*.

Sefton, M., Shupp, R., and Walker, J.. (2002). “The effect of rewards and sanctions in provision of public goods,” *CEDEX Working paper no. 2002-2.*

Stella, G. M. (1996). *“Schei”. Dal boom alla rivolta: il mitico Nordest,* Milano, Baldini & Castoldi.

Table 1 – Treatment Conditions

	Baseline	*Consensual*	*Sequential*
Decisions to contribute are Simultaneous	Yes	Yes	Yes
Decisions to punish are Simultaneous	Yes	Yes	No
Punishment when only one agent requested it	Yes	No	Yes
Number of sessions	4	4	4
Total number of participants	80	80	80
Periods without punishment opportunity	10	10	10
Periods with punishment opportunity	10	10	10

Table 2: Summary of individual contributions across experimental sessions

Treatment	*Baseline*				*Consensual*				*Sequential*			
Session date	3/27	10/15	10/17	10/23	5/22	9/16	10/16	10/21	4/14	10/16	10/20	10/22
avg (sd)												
No	3.54	4.40	2.57	3.34	2.56	5.07	3.91	2.27	2.80	4.00	3.07	2.19
punishment	(5.58)	(6.08)	(4.34)	(5.56)	(4.13)	(6.26)	(5.49)	(4.09)	(4.85)	(5.66)	(4.92)	(3.62)
With	7.74	5.14	2.62	2.36	14.11	11.26	5.89	2.57	4.18	5.53	2.42	5.41
punishment	(5.42)	(5.25)	(3.12)	(2.49)	(6.54)	(6.59)	(4.51)	(2.93)	(5.68)	(4.36)	(4.57)	(4.87)

Table 3: Punishment rates by individual level of contribution (Sub-sample)

Individual contribution level	Baseline	Sequential	Consensual			
	Avg. points *(no. obs.)*	Avg. points *(no. obs.)*	Assigned (1)	Requested (2)	Difference (2) – (1)	% Censored [(2) – (1)] /(2)
0	5.06 *(124)*	4.50 *(177)*	4.22 *(68)*	4.50	0.28	6.2%
(0, 5]	1.66 *(261)*	2.34 *(171)*	1.01 *(159)*	1.43	0.43	29.8%
(5, 10]	0.85 *(83)*	1.46 *(153)*	0.45 *(83)*	1.06	0.61	57.9%
(10, 20]	0.78 *(27)*	1.42 *(59)*	0.20 *(30)*	0.93	0.73	78.6%
Total	2.33 *(495)*	2.69 *(560)*	1.44 *(340)*	1.91	0.47	24.6%

Notes: To partially control for the uneven distribution of contribution levels across treatments, the table includes only the sub-sample of experimental data where *group* contribution was in [10, 40], which constitutes about 58% of the sample (1395/2400 obs.).

Table 4: Determinants of getting punished: regression results

	Baseline	*Sequential*	*Consensual*	*All treatments*
High cooperators	0.7999	0.2274	-0.8794	-0.1897
(contributions >15 tokens)	(0.8858)	(0.5818)	(0.3183)*	(0.3975)
Average contribution of	-0.0945	-0.0172	0.0569	0.0025
others in the group	(0.1612)	(0.0358)	(0.0193)*	(0.0438)
Positive deviation from	-0.0707	0.0018	-0.0006	-0.0223
average	(0.0648)	(0.0324)	(0.0134)	(0.0252)
Absolute negative deviation	0.7871	0.6295	0.5688	0.6329
from average	(0.0768)***	(0.1142)**	(0.0434)***	(0.0538)***
Constant	1.5221	-0.6236	0.1575	1.0468
	(1.6952)	(0.3496)	(0.3700)	(0.4299)**
No. of observations	800	800	800	2400
R-squared	0.39	0.41	0.57	0.45

Notes: OLS estimator clustered by session. It includes session and period dummies, not reported. Robust standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Frequency of punishment

	Baseline	*Sequential*	*Consensual*
Contribution choices not punished	**32.1%**	**27.4%**	**66.3%**
Of which: One request to punish	-	-	26.3%
Contribution choices punished	**67.9%**	**72.6%**	**33.9%**
Of which: One request to punish	32.5%	35.8%	-
Two requests to punish	20.0%	23.1%	18.8%
Three requests to punish	12.0%	11.1%	10.4%
Four requests to punish	3.4%	2.6%	4.6%
Total	100.0%	100.0%	100.0%
(observations)	(800)	(800)	(800)

Figure 1: Gross contribution levels

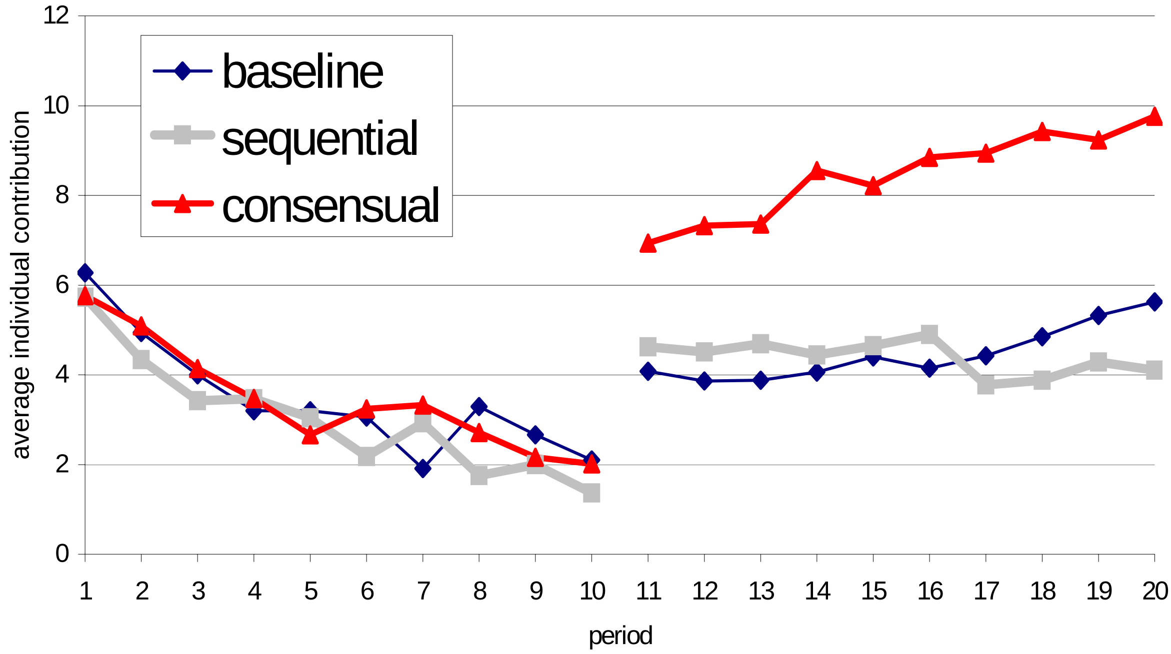
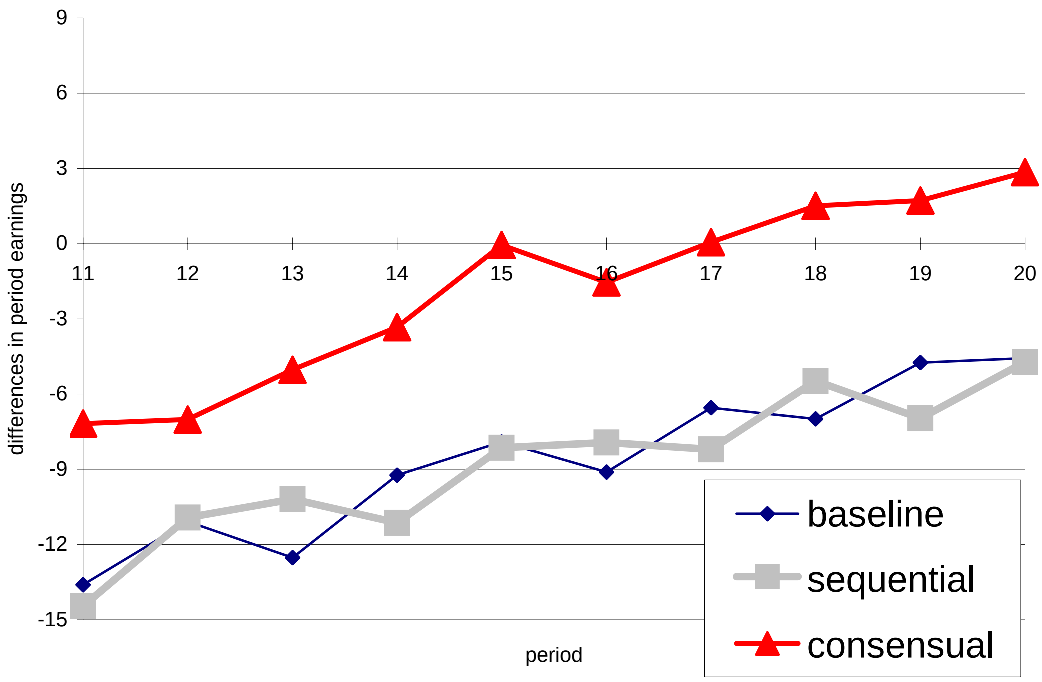
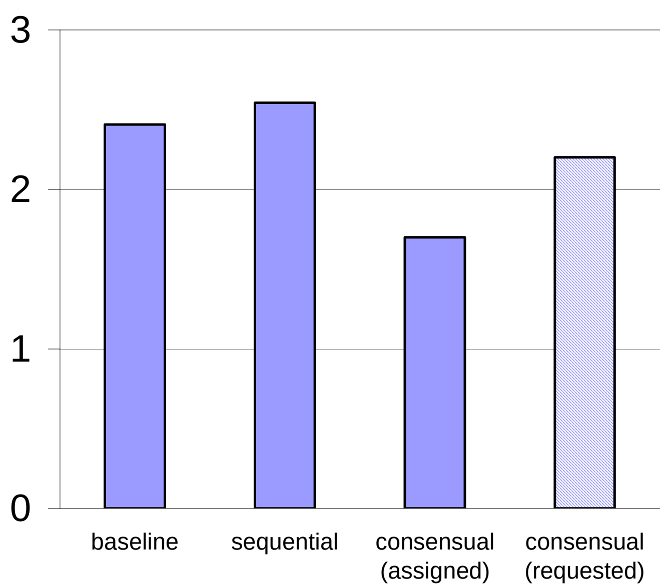


Figure 2: Net aggregate surplus over time



Note: Period by period difference between sequence with sanction and without sanction

Figure 3: Average punishment by treatment



Instructions

The following instructions were originally written in Italian. We document the instructions we used in the first part of the experiment, which were common to all treatments, and present the second part instructions for the Consensual treatment. The instructions for the Baseline and Sequential treatments were adapted accordingly. They are available upon request.

You are now taking part in an economic experiment on decision-making. If you read the following instructions carefully you can, depending on your decision, earn a considerable amount of money.

During the experiment we shall not speak of Euros but rather of Tokens. During the experiment your entire earning will be calculated in Tokens. At the end of the experiment the total amount of tokens you have earned will be converted in Euros at the following rate

1 Token = 2 cent. Euro

At the end of the experiment your earnings will be privately paid in cash. To the amount on your screen you must add 5 Euro as a lump sum participation fee.

During the experiment you will not be asked to reveal your identity and your name will not be associated with the decisions you are going to take. Moreover, you are not allowed to talk or otherwise communicate with the other participants during the experiment.

This experiment is divided into two different parts. The following instructions are related to the first stage. The first stage consists of 10 periods.

Your earnings depend on your decision and on other four participants' decisions. The experiment participants will be randomly re–matched after each period and therefore it is highly likely that in each period you will interact with different people. You do not know the identity of the people with whom you interact.

At the beginning of each period each participant receives 20 points. You task is deciding how you would like to use these tokens. The other participants will face the same scenario. You have to decide how many points of the 20 available you want to contribute to a project. For each point that you keep for yourself you earn an income of one Token. The points you have contributed to the project plus the points that all the other four persons have contributed are converted in a double quantity of Tokens, which will be evenly divided among these five persons. Therefore, after being doubled, you will receive one fifth of the Tokens contributed to the project. To sum up your income consists of two parts:

Your income this period = direct income + income from the project

= (20 – your contribution to the project) + $^1/_5$ x ((sum of yours' and other four people's contribution to the project) x 2)

Each of the four persons will receive from the project the same amount that you will. For example, suppose the sum of the contributions of the five persons is overall 60 points. In this case each person receives from the project 60x2/5=24 Tokens. Instead, if the total contribution to the project is 10 points, each of the five persons receives an income from the project of 10x2/5=4 Tokens. The following table gives you some examples of income from the project:

Sum of the points contributed	0	10	20	30	40	50	60	70	80	90	100
Income from the project for each of the 5 persons	0	4	8	12	16	20	24	28	32	36	40

For each point that you keep for yourself you earn an income of 1 Token. Supposing you contributed this point to the project instead, then the total contribution to the project would rise by one point. Your income from the project would rise by 1x2/5=0,4 Tokens. Your contribution to the project would also raise the incomes of other persons. More precisely, the other four persons will earn an additional 0.4 Tokens each, so that the overall income increase for you and the others would be of 2 Tokens.

After everybody has completed his or her decision, you shall see your period income on the computer screen. Moreover, there will be presented the points contributed to the project by each one of the four persons that could contribute with you as well as their period income. The identity of these other people will change randomly each period.

This procedure will be repeated 10 periods.

Are there any questions? If you have questions during the experiment we kindly ask you to raise your hand and somebody will assist you in private.

These are the instructions for the second and last part of the experiment. As before, the experiment consists of ten periods and in each period you have to make a decision about how many of the 20 tokens available to you.

Different than from before, each period is now composed of two phases, the first phase is identical to the procedure already described, while in the second phase you may choose to reduce the earnings of other people that have profited from the same project.

<u>In the first phase</u> of a period, you have to make the same type of decision as the one in the sequence already described before, and that will here be repeated.

At the beginning of each period each participant receives 20 points. You task is deciding how you would like to use these tokens. The other participants will face the same scenario. You have to decide how many points of the 20 available you want to contribute to a project. For each point that you keep for yourself you earn an income of one Token. The points you have contributed to the project plus the points that all the other four persons have contributed are converted in a double quantity of Tokens, which will be evenly divided among these five persons. Therefore, after being doubled, you will receive one fifth of the Tokens contributed to the project. To sum up your income consists of two parts:

Your income for phase one = direct income + income from the project

= (20 – your contribution + $^1/_5$ x ((sum of yours' and other four people's to the project) contribution to the project) x 2)

After everybody has completed their decision, you shall see on the computer screen the points contributed to the fund by every one of the four persons that could contribute with you as well as their period income. Your decision and result will be shown in the first column. The identity of these other people will change randomly each period.

In the second phase of a period you can reduce or leave equal the income of each of the four persons that have profited from the same project. Conversely, the other persons can lower your earnings as well.

Your decision is about distributing points to the other four persons. There is no way for you to know the identity of the other persons because they have been randomly selected every period among all participants. You have to choose a number of points for each person and you know only his/her contribution decision in the first phase of the period. If you do not want to change the earnings of a person choose 0. If you want to reduce the earnings of a person, you can distribute a number of points from 0 to 7. For each point distributed, the income in that particular person will be reduced by 3 (THREE) tokens. For the person distributing the point, each point costs 1 token. Your overall cost is equal to the sum of the points that you have distributed to each one of the other four persons. Your maximum cost for distributed points is then 28 tokens (7 tokens times 4 persons). Your cost is zero if you do not distribute points to anybody.

As it will now be explained, a request to distribute points is not always carried out. For each person, there are two cases.

When ONLY YOU have requested to distribute points to a given person, your decision has no effect. In particular, there is no reduction in his/her income and no payment on your side for your request. In the opposite case, when BOTH YOU AND OTHERS have requested to distribute points to that same person, then your decision to distribute points is carried out. Requests by others to distribute points to that person will also be carried out. In other words, there have to be at least two requests to distribute one or more points to the same person in order to carry out a reduction of his/her income. It does not matter that the two requests are for distinct amounts.

EXAMPLES. If you distribute 0 points to a person you do not change his/her income. Suppose you request to distribute 6 point to a person. Under some conditions this request does not have any effect, while under other conditions you reduce his/her income by 18 tokens (6x3). More precisely, if all the others distribute 0 points, your request will be ignored. This result will be signaled on the screen at the end of each period by the note “Points distributed? NO” in the column corresponding to the concerned person. On the contrary, if at the same time somebody else has distributed for instance 2 points to the same person, your request is carried out (an 18–token reduction) and you will be charged the fee of 6 tokens. In addition, the request of the other person will be carried out. The cumulated effect of the two requests is an overall income reduction of (6+2) x 3 = 24 tokens. This result is marked on the screen at the end of the period by the note “YES” in the column corresponding to the concern person.

Your total income at the end of the period will be:

Your period income = phase one income – income reduction – cost to distribute points

= phase one income – (sum of received points)x 3 – (total points distributed).

After everybody has completed their decision, you shall see on the computer screen the results for phase two. For each person you will learn the cumulative income reduction due to the points distributed. Individual requests to distribute points will remain confidential in order to preserve the anonymity of decisions.

Are there any questions?

Example of punishment moves in the sequential treatment

Punisher	**Step of punishment (**targeted agent**)**			
	1	**2**	**3**	**4**
A	B	C	D	E
B	A	E	C	D
C	E	D	A	B
D	C	B	E	A
E	D	A	B	C

Notes on how to read the table: In step 1, agent A has an opportunity to punish agent B; in step 2, agent D is informed of the punishment already received by agent A and has an opportunity to punish more. In step 3, agent E is informed of the cumulative punishment already received by agent A and has an opportunity to punish more.