Punishment and Counter-punishment in Public Goods Games: Can we still govern ourselves?

Nikos S. Nikiforakis

Royal Holloway University of London

-1st Draft-

March 2004

Punishment and Counter-punishment in Public Goods Games: Can we still govern ourselves*?

Nikos Nikiforakis

Royal Holloway University of London

Abstract

In the public goods literature, there have been recently a number of experiments which demonstrate how the problem of the under-provision of a public good can be solved through mutual monitoring and sanctioning between the members of a group when antisocial behavior is observed. In many circumstances, however, we can not allow for punishment and exclude the possibility of counter-punishment occurring. We design a public goods experiment based on Fehr and Gaechter (2000) where we allow for both punishment and counter-punishment. We find that in both Partner and Stranger treatments average contributions decline steadily over time, at a rate similar to the treatment were no punishment was allowed, and tend towards full free-riding. The reason for this change seems to be that under the threat of counter-punishment people are less willing to punish. An important result is that participants squander their endowment in punishment and counter-punishment actions leading to a relative payoff loss, in comparison to the treatment without punishments.

* The title is a reference to Ostrom, Walker and Gardner (1992): "Covenants with and without a sword: Self governance is possible". I would like to express my gratitude to Hans Normann for his invaluable help at every stage of this paper. I would also like to thank Marco Casari for the helpful comments. Please send further comments to n.nikiforakis@rhul.ac.uk

1. Introduction

Contrary to the predictions of standard economic theory that people will not contribute voluntarily for the production of a public good, a considerable amount of experiments have shown that, initially, people give on average between 40 and 60 percent of their endowment. The contribution level, however, decreases with repetition under the influence of free-riders [Davis and Holt (1993), Ledyard (1995), Fehr and Fischbacher (2003)].

Recent experiments have also shown that people are willing to punish behaviour, which is deemed to be antisocial. Under the punishment threat free-riding is curtailed [Fehr and Gaechter (2000), Bowles, Carpenter and Gintis (2001), Page and Putterman (2000), Sefton, Shupp and Walker (2002), Masclet, Noussair, Tucker, Villeval (2003), Carpenter (2002)][1].

In the seminal paper by Fehr and Gaechter (2000) (hereafter F&G), participants played a two-stage public good game. In the first stage, they were asked to divide their endowment between a public and a private account. The returns from each account were designed so that group earnings were maximized when participants contributed all their money in the public account. However, each individual had an incentive to keep his endowment for himself. The results confirmed previous findings with significant contributions (40-60 percent) in the beginning of the experiment, which declined over time.

In the second stage, participants were allowed to assign punishment points to the other members in their group after they were notified about individual contributions in the public account[2]. Punishment was costly for both the punisher and its receiver. The introduction of punishment

[1] For a brief description of these papers, which deal with punishment in a public goods environment see section A.3 in the appendix. The observed rejection of positive offers in experimental ultimatum games is another indication of people's willingness to punish unfair behaviour (Gueth and Tietz [1990], Camerer and Thaler [1995], Roth [1995], Fehr, Gaechter, and Kirchsteiger [1997], Forsythe, Horowitz, Savin and Sefton [1994], Davis and Holt [1993]. See also Zizzo and Oswald (2001).

[2] The announcements were made in such a way that the formation of individual reputation across periods was not possible.

opportunities led to significantly higher contributions and to an eventual payoff gain relative to a treatment where participants were not allowed to punish. We will return to this paper and its results later on.

This experimental design of F&G has become the standard by which to study the influence of punishment on cooperation in a public goods environment. However, in every day life, one can often observe many cases of free-riding and the unwillingness of the cooperators to punish. At the same time, there exists an abundance of anecdotal evidence that people are willing to engage in costly counter-punishment. An example of the unwillingness to punish can be found in the case of the refusal by some countries to sign the Kyoto Protocol for the reduction of the emissions of greenhouse gasses in the atmosphere[3]. Many countries, which agreed to the conditions of the Protocol, expressed their dissatisfaction with the "free riders" without, however, taking any measures to discipline them.

We conjecture that the reason for the avoidance of punishment, in cases where free-riding is observed, is the fear of counter-punishment. More over, often, we can not allow for punishment and exclude the possibility of counter-punishment occurring. Since counter-punishment is inseparable from punishment, we shall refer to the type of punishment in models as the one by F&G as "*one-sided punishment*", in contrast to the "*two-sided punishment*" where counter-punishment is allowed.

To test the hypothesis that the threat of counter-punishment can be the explanation for the observed free-riding, we designed a public goods experiment with two treatments: one without any form of punishment, the familiar voluntary contribution mechanism (VCM), and one with punishment and counter-punishment (P&CP). To have a solid basis for comparison of our results, we based the experimental design to the one by Fehr and Gaechter (2000) (F&G). The two treatments were run both under the partner and the stranger protocol.

[3] Air is a textbook case of a pure public good.

In the VCM treatment, as we will see, average contribution exhibited a similar behaviour to the one reported so far in other experiments, by starting between 40 and 60 percent of the endowment and decreasing over time. The introduction of counter-punishment opportunities in the P&CP treatment seems to cancel out, to a large extent, punishment's so far observed disciplinary effect and participants behave similarly to the VCM treatment with average contribution declining with repetition. In the words of Girard (1979): "Reciprocal violence now demolishes everything that unanimous violence has erected". We show that an explanation for this is that under the counter-threat, people are less willing to punish and as a result, people are almost free to free ride.

To our knowledge, there is no other paper testing for the effect that the existence of counter-punishment opportunities has on the level of cooperation. Although in our experiment no explicit coordination opportunities exist, in the partner treatment, the fact that the composition of the groups remains the same might lead to the formation of behavioural norms that will alleviate free-riding more effectively[4].

The remaining of the paper is structured as follows: section 2 introduces the experimental design and the procedures of the experiment, while section 3 presents the predictions of standard economic theory for this set up. Section 4 discusses the experimental results and section 5 concludes.

2. The Experiment

2.1 *The experimental design*:

To have a clear picture of the effect that counter-punishment has we based our design on F&G (2000). The experiment consists of two treatments using a related sample design: one without any punishment (VCM), *and* one with *two-sided* punishment i.e. with punishment and counter-punishment (P&CP). We run the treatments both under the partner protocol, where the

[4] Masclet et al. (2003) show that when the same group of people play a finitely repeated public goods game the expression of disapproval towards anti-social behavior can also play a significant role in decreasing free-riding.

composition of each group remains unchanged throughout the experiment and under the stranger protocol, where the participants where randomly re-matched in each period. For each treatment there were 12 subjects who were randomly divided in groups of 4 people and played a finitely repeated public goods game for 10 periods.

All participants were aware that each treatment would last exactly 10 periods. However, they were not aware that a second treatment was to follow[5]. The related sample design has the advantage that additionally to across-subjects comparison we can make within-subjects comparisons of the average level of contribution, which have much more statistical power. To test for sequence effects, in session 1 (stranger) and session 3 (partner) the participants played the P&CP treatment first and the VCM second, whereas in sessions 2 (stranger) and 4 (partner) the order was reversed. All this can be summarised in table 1:

Table 1: Treatment Conditions

	P&CP / VCM	**VCM / P&CP**
Stranger	Session 1: 3 groups of 4 participants	Session 2: 3 groups of 4 participants
Partner	Session 3: 3 groups of 4 participants	Session 4: 3 groups of 4 participants

2.1.1 *The VCM treatment*:

The first treatment is the standard voluntary contribution mechanism as presented first by Isaac, Walker and Thomas (1984). In the beginning of each of the ten periods, every participant received a fixed amount of 20 Experimental Currency Units (ECUs)[6]. The participant had then to decide

[5] Following the example of F&G, to keep the results from the first treatment unaffected by the existence of a second treatment.

[6] The ECU was exchanged at a rate of 1 ECU = 4 p.

how many ECUs to keep for himself and how many to invest into a project. All the participants made their decision simultaneously and without being aware of the others' decisions. The monetary payoff for each subject in each period was given by:

(1) $$\pi_i^{VCM} = 20 - g_i + 0.4 * \sum_{j=1}^{n} g_{j,}$$

where 20 is the endowment in ECUs, g_i is the amount of ECUs subject *i* invests in the project ($0 \leq g_i \leq 20$) and 0.4 is the marginal return per capita (MRPC) from the project. This payoff function implies that each player's income comes from two sources: the money he keeps for himself, as indicated by $20 - g_i$ and a fraction of the total amount that the group invested in the project, $0.4 * \sum_{j=1}^{n} g_j$. The total payoff from the no-punishment condition is equal to the sum of the 10 period payoffs as given by (1) i.e. $\sum_{n=1}^{10} \pi_i^{VCM}$.

Equation (1) also implies that full free-riding (g_i=0) is a dominant strategy in the stage game. This follows from $\partial \pi_i^{VCM} / \partial g_i$=-1+0.4<0, which means that the more an individual contributes to the project the less her income will be in that stage. However, the aggregate payoff, $\sum_{i=1}^{4} \pi_i^{VCM}$ is maximized if each group member fully cooperates (g_i=y), since $\partial \sum_{i=1}^{4} \pi_i^{VCM} / \partial g_i$=-1+4*0.4>0. This inequality shows that the more people contribute, the higher the aggregate payoff; therefore, the total payoff of the group will be at its highest point when the participants contribute their whole income.

In the first treatment, the payoff function (1), the amount of the endowment (20 ECUs), the MPRC (0.4), the number of the subjects and the duration of the treatment were all common knowledge between the players.

2.1.2 *The P&CP treatment*:

In the second treatment, two more stages were added to the simple voluntary contribution mechanism, which now became the first of three stages. In the second stage subjects were given the opportunity to simultaneously punish each other after being informed of the individual contributions[7]. To do so, group member *i* had to assign *punishment points* to group member *j*. This had two different effects in the payoffs of members *i* and *j*: for each point received by player *j* his income from the first stage, π_t^1, was reduced by 10%. Note that the first stage income could never be reduced below zero, so if player *j* received more than 10 punishment points his income was reduced by 100%. Additionally, player *i* also faced a cost for distributing punishment points to player *j*. This cost was given by the following convex cost function, $\underset{i\neq j}{c}(p_i^j)$:

Table 2: Punishment points per player and associated costs for the punishing subject

p_i^j	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$\underset{i\neq j}{c}(p_i^j)$	0	1	2	4	6	7	12	16	20	25	30

Given the above information, the payoff at the end of the second stage for subject ***i*** is equal to:

(2)
$$\pi_t^2 = \pi_t^1 * \left[\frac{\max(0, 10 - \sum_{j\neq i} p_j^i)}{10} \right] - \sum_{i\neq j} c(p_i^j)$$

[7] For the whole experiment we used neutral framing. Punishment was referred to as "assigning points" in order to "reduce" another participant's income. The public good itself was named "project".

Again, the payoff function (2), the cost function ($\underset{i\neq j}{c}(p_i^j)$), the amount of the endowment (20 ECUs), the MPRC (0.4), the number of the subjects and the duration of the treatment were all common knowledge.

Up to the end of the second stage, the experiment is identical to the one by F&G. In the third and final stage, the subjects were informed how many points each of the other members in their groups assigned to them. They then were given a last opportunity to reduce the income of the participants who punished them during the second stage[8]. For the calculation of the end-of-period income the same function for the payoff and the punishing points were used as before. The only difference is that now the second stage income was used as a basis instead of the first stage income, i.e.:

(3)
$$\pi_t^3 = \pi_t^2 * \left[\frac{\max(0,10 - \sum_{j\neq i} p_j^i)}{10} \right] - \sum_{i\neq j} c(cp_i^j)$$

where cp_i^j is the number of counter-points that player *i* assigns to player *j* and the cost of counter-points is equal to the cost of punishment points i.e. $c(cp_i^j) = c(p_i^j)$. Note that the cost for assigning points works accumulatively i.e. if player *i* punished player *j* in his group with 2 points during the second stage and then with 2 further (counter-) points in the third stage, his total cost from points would be equal to 6 i.e. the cost of 4 points.

To prevent the possibility of forming an individual reputation, every player received a number between 1 and 4, in the beginning of each period, which he retained for the duration of the period, but which changed in the next one.

[8] Note that only the subjects who were punished were allowed to punish back. This was done to avoid strategic punishing. By strategic, we mean that a subject, if allowed, could have punished lightly or not at all in the second stage to avoid counter-punishment and then more heavily in the third. Obviously, such a design would not be appropriate to measure the effect of counter-punishment.

2.2. *Procedures*

The experiment took part between December 2003 and January 2004 in the experimental laboratory of Royal Holloway, University of London. It consisted of five sessions (2 partner, 2 stranger and 1 control), which lasted approximately an hour and forty-five minutes[9]. The participants were recruited via e-mail by using the Royal Holloway Economics Experiments Team mailing list. The total number of the subjects was 60. Twenty-four of them took part in the partner treatment (VCM and P&CP), twenty-four in the stranger (VCM and P&CP) and 12 more in the control session (stranger VCM and One-Sided Punishment). The sample consisted of students with different nationalities and backgrounds including Economics[10]. Being in the above-mentioned mailing list implied that the participants might have taken part in economics experiments before, although not in a public goods experiment.

The subjects were gathered outside the laboratory and then entered the lab in a random order one by one. They sat in such a way that it would have been impossible for them to see who the other participants in their group were, in order to avoid the communication effect.

At the beginning of each of the treatments, the participants were given a different set of instructions explaining in detail what was to happen[11]. They were then given as much time as they needed to read the instructions and to fill in a brief control questionnaire. Once the participants were ready, a supervisor approached and privately checked the answers and answered any questions that the participants had.[12] Afterwards, and due to the complexity of the experiment, the supervisor read out a pre-written summary of the key points and asked for any possible questions. Finally, a trial period was used were the participants were introduced to the computer

[9] The control treatment lasted slightly less.
[10] Contrary to other findings (Marwell and Ames [1981]) the economists-to-be were arguably the strongest supporters of cooperation.
[11] The instructions can be found in section A.1 of the appendix.
[12] In general, the only explanations that the supervisors had to give were regarding incorrect answers in the control questionnaire. These cases however were very few.

screens they would have to use to make their decisions. Again, a pre-written text was used for this, to ascertain that all subjects would receive the same explanations regardless of the session they participated. The experiment was programmed and conducted with the software z-Tree (Fischbacher [1999]). Participants earned on average £18.05. No show up fee was given.

3. Predictions

The subgame-perfect Nash equilibrium prediction is in all treatments that participants should contribute nothing to the project, i.e. g_i=0, for every i. In specific, in the treatment without punishment the dominant strategy is to free ride. Using backward induction for the ten periods we find that the dominant strategy is to contribute nothing in the project.

In the P&CP treatment the subgame-perfect Nash equilibrium prediction is that people will never counter-punish in the third stage, since this is costly and yields no material benefits. The same applies for punishment in the second stage. Finally, at the first stage, the participants understand that no one is going to punish them no matter whether they cooperate or not, and therefore they have no reason to contribute to the project, thus choosing to contribute zero. Applying backward induction for the ten periods we arrive at the prediction that g_i=0, p_i^j=0 and cp_i^j=0.

4. Experimental Results

We will begin by analyzing the effect of counter-punishment first under the stranger and then under the partner protocol.

4.1 *The impact of counter-punishment under the stranger protocol*

If the introduction of counter-punishment is of no importance then we should observe no difference in the behaviour of the participants in comparison to other experiments who studied one-sided punishment. This

means that in the P&CP treatment people should increase significantly their level of contribution to the public good in comparison to the VCM treatment and continue to do so under the threat of punishment. However, there is a significant difference between this behaviour and the one observed when counter-punishment was possible.

Result 1: *The existence of punishment and counter-punishment causes a minor aggregate increase in the average contribution level, which is, however, considerably smaller of the one when only punishment was present.*

Table 3: Mean contributions in the stranger-treatment

	mean contribution in all periods		**mean contribution in the final periods**	
Session	VCM	P&CP	VCM	P&CP
1	3.97	6.80	2.17	3.83
	(1.66)	(1.71)	(2.69)	(3.13)
2	3.55	2.47	0.58	0.92
	(3.23)	(1.86)	(1.44)	(1.51)
mean	3.76	4.63	1.38	2.38
	(2.35)	(1.71)	(2.26)	(2.83)
	VCM	Punishment	VCM	Punishment
FG mean	3.7	11.5	1.9	12.
	(5.7)	(5.9)	(4.1)	(5.6)
FG session 3	4.5	10.7	2.0	13.1
	(6.0)	(4.9)	(3.8)	(4.0)
NSN control	6.9	10.4	2.83	9.25
	(2.29)	(1.14)	(4.20)	(5.83)

Note: The numbers in parentheses indicate standard deviations. In session one the treatment with P&CP was played first and then the VCM whereas in session 2 the roles were reversed. NSN refers to the authors initials.

Support for the first result comes from Table 3. On the first part of table 3, comparison of columns 2 and 3 shows that in session 1 we had an increase on the average contribution level, whereas in session 2 (when the VCM was played first) counter-punishment led to a decrease[13]. For completeness we add the aggregate results from F&G[14].

[13] It has been shown that the outcome of a public goods game is largely dependent on the mixture of selfish and altruistic individuals, and the environment in which the game is played (Fehr and Fischbacher [2003]). In session 2, 4 participants could be characterized as

The first thing that one should notice is the striking similarity of the results in the VCM treatment between the two experiments. On average, the contribution level increases from 3.76 to 4.63 that is by 23%, which is significantly different from the 211% increase that the introduction of punishment opportunities caused in F&G. This result supports our hypothesis that the effect of counter-punishment should eliminate to a large extent the positive effect of punishment on cooperation.

To test for differences in behaviour across countries based on cultural characteristics (Burlando and Hey [1997]) we also run a session identical to session 3 of F&G. In the last rows of table 3 we see that the introduction of one-sided punishment in our sample increases average contribution. Again, the average contribution in the one-sided punishment treatment across all periods is very similar between the two experiments, as the participants were able to sustain cooperation. However, whereas in F&G average contribution was higher in the final period, in our case, there was an end-of-treatment effect[15]. The results indicate that the difference in behavior can indeed be attributed to the introduction of counter-punishment opportunities and not in cultural differences. Our next result deals with the evolution of average contribution over time.

Result 2: *In both the VCM and the P&CP treatments, average contributions converge to free-riding over time.*

A first indication for result 2 can be found in table 3 by examining columns and 5: we can see that there is a small difference between the final periods of the two treatments. Looking at columns 3 and 5, we observe the decline in average contributions. By comparison of the means, we witness that there

"perfect free riders" as they contributed zero in all periods. These subjects were able to drag down cooperation very quickly.

[14] Fehr and Gaechter had 3 independent observations each one with 24 subjects. Dufwenberg and Sneezy (2000) have shown that there is no difference in the results when using 12 or 24 subjects.

[15] The evolution of average contribution can be seen in figure13 in the appendix.

is a small increase in average contribution in our experiment, which however is significantly smaller than in F&G and our control. In contrast to these, one can see the increase in cooperation that punishment alone caused when looking columns 3 and 5 on the lower part of table 2.

Result 2 is better understood by looking at figure 1 and 2. Though the two treatments were played in sequence, we place the results over the same time period to underline the similarities in behaviour.

Figure 1: Average contribution over time in the stranger-treatment (session 1)

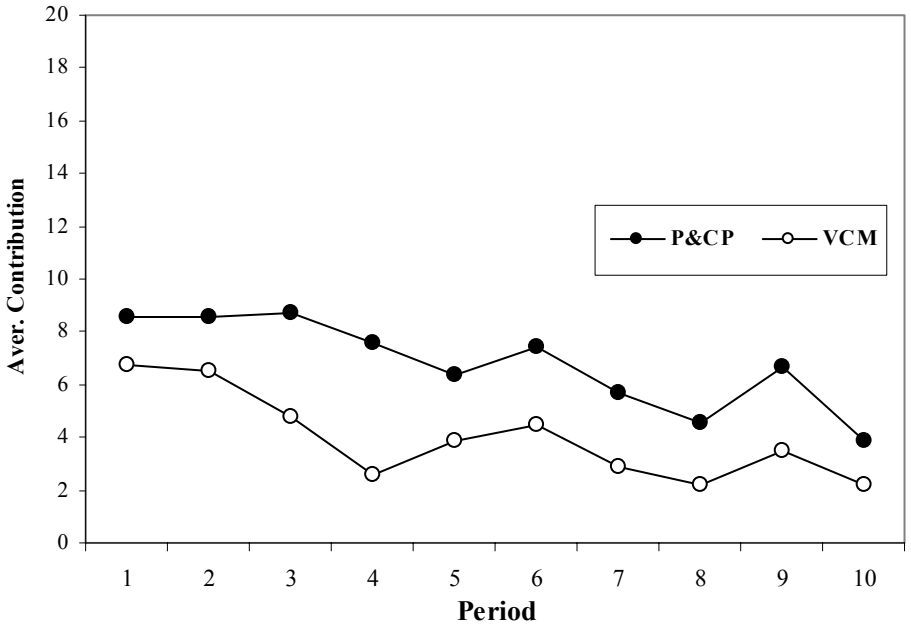
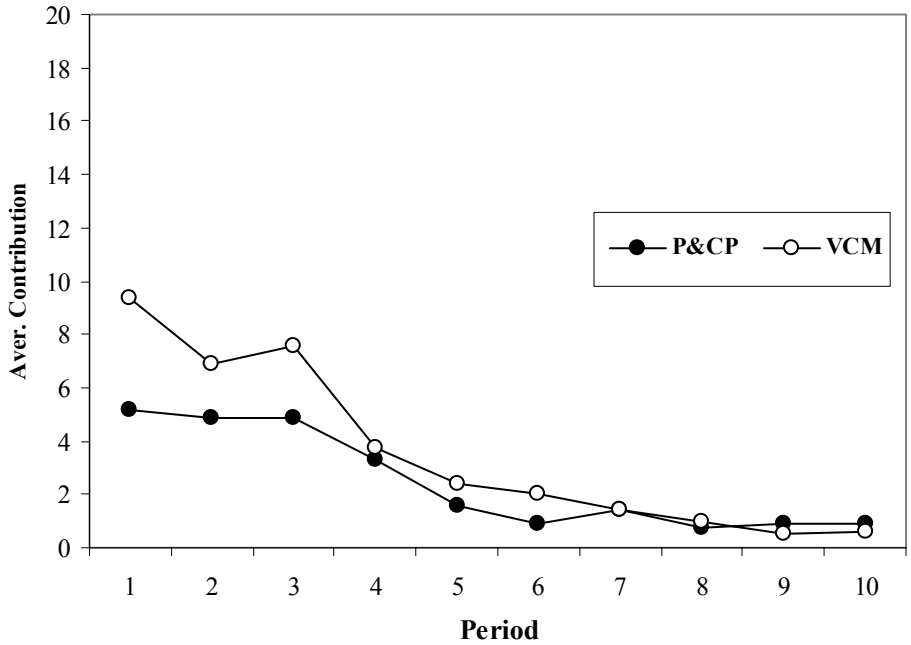


Figure 2: Average contribution over time in the stranger-treatment (session 2)



From both figures it is apparent that counter-punishment has a dramatical effect on cooperation. Whereas in experiments with one-sided punishment average contribution was increasing over time, in our experiment the course of average contribution is reversed and cooperation is deteriorating. In session 1, people achieve a higher level of contribution than in VCM, but it appears to be falling towards the end. In session 2, where the VCM was played first, punishment is no longer able to increase cooperation and it remains most of the time at an even lower level than in the treatment with no punishment reaching ultimately reaching complete free-riding. If we aggregate the results of the two sessions we get figure 3.

Figure 3: Average contribution over time in the stranger-treatment (session 1 & 2)

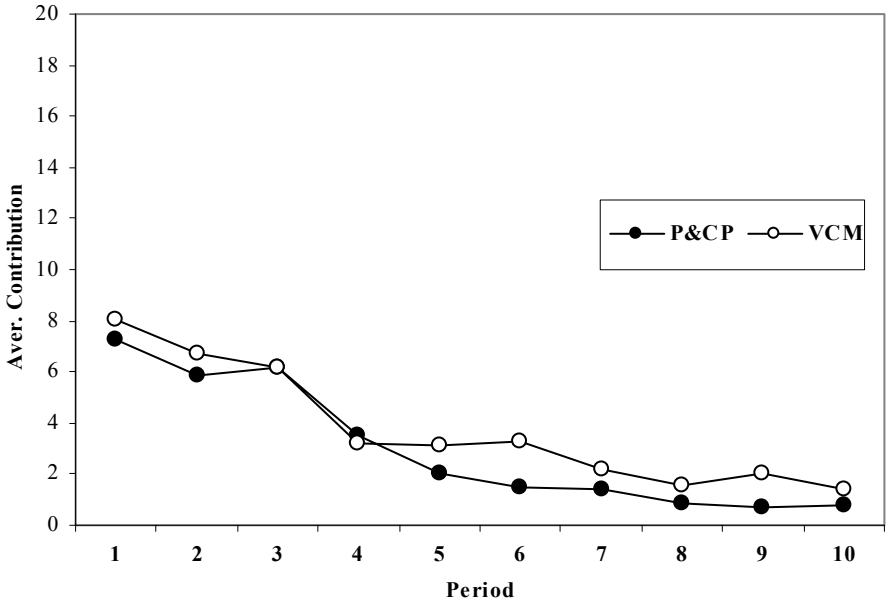


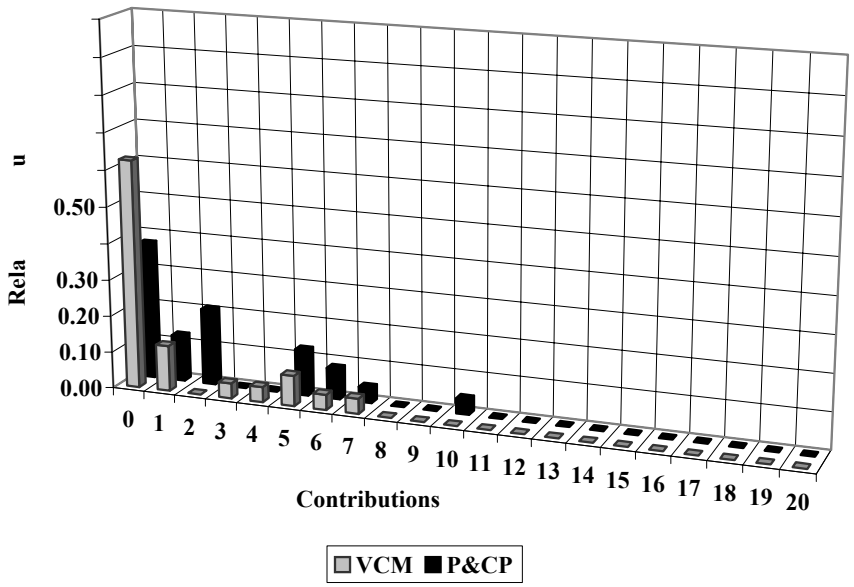
Figure 3 shows how strong the effect of counter-punishment is in the stranger-treatment. The evolution of the average contribution over time in the two treatments is almost identical, which suggests that in the stranger-treatment counter-punishment balances off the effect of punishment completely.

Results 1 and 2 deal only with average contributions. To have a deeper understanding we take a look at the behavioural regularities at the individual level. Result 3 summarizes the findings.

Result 3: *There is very similar behaviour in the final period of both treatments and free riding emerges as the modal action.*

The aforementioned result comes from figure 4. Although the percentage of people who free-ride completely in the P&CP treatment is significantly lower than in the VCM, one can still notice the similarity of the results and the total absence of participants who contributed more than 10 ECUs.

Figure 4: Distribution of contributions in the final period of the stranger-treatment.



4.2 *The impact of counter-punishment in the partner-treatment*

The first result in the partner-treatment deals with the average contribution over all periods.

Result 4: *The introduction of punishment and counter-punishment opportunities causes a rise in the average contribution level.*

Evidence for result 4 can be found in table 4. By comparing column 2 with column 3 we notice that contribution has increased on average in all the groups. According to a Wilcoxon matched pairs test, with group averages as observations, this difference is statistically significant (p=0.028, two-tailed). On average, subjects contribute from 1.7 (group 6) to 4.9 (group 3) times more than in the no-punishment condition. In the P&CP condition, participants contribute on average 42 percent of their endowment. The increase in contribution (136%), in comparison to the VCM treatment, is similar in amount to the one found by F&G, although the aggregate levels in both conditions seem to be half in our case.

If we compare column 2 with column 4 and column 3 with column 5 we find again that in both treatments and for all 6 groups there has been a decline on the average level of contribution. In the final period of the P&CP, participants contribute on average only 2.71 ECUs. This is result 5:

Table 4: Mean contributions in the partner-treatment

	mean contribution in all periods		**mean contribution in the final periods**	
Group	VCM	P&CP	VCM	P&CP
1	4.45	13.03	0	10
	(2.55)	(1.44)	(0)	(3.56)
2	0.73	2.33	0.25	0
	(1.51)	(3.09)	(0.5)	(0)
3	1.58	7.73	0.25	0.5
	(3.20)	(6.30)	(0.5)	(0.58)
4	3.7	7.15	0	3.25
	(3.90)	(2.84)	(0)	(3.95)
5	2.95	7	0	5
	(3.24)	(1.07)	(0)	(5.77)
6	7.85	13	0	0.25
	(5.52)	(5.68)	(0)	(0.5)
Mean	3.54	8.37	0.07	2.71
	(4.1)	(5.32)	(0.28)	(4.61)
	VCM	Punishment	VCM	Punishment
FG mean	7.5	17	3.2	18.2
	(6.8)	(4.5)	(4.4)	(2.3)

Note: The numbers in parentheses indicate standard deviations. In session three (groups 1, 2, 3) the treatment with P&CP was played first and then the VCM whereas in session 4 (groups 4, 5, 6) the roles were reversed.

Result 5: *Both in the VCM and the P&CP conditions of the Partner-treatment average contributions decrease over time and converge to full free riding.*

Additionally to the evidence given in table 4, result 5 can be better summarized in figures 5 and 6, which represent session 3, where the P&CP condition was played first, and session 4, where the order was reversed, respectively, and show the evolution of the average contribution over time.

Figure 5: Average contribution over time in the partner-treatment (session 3)

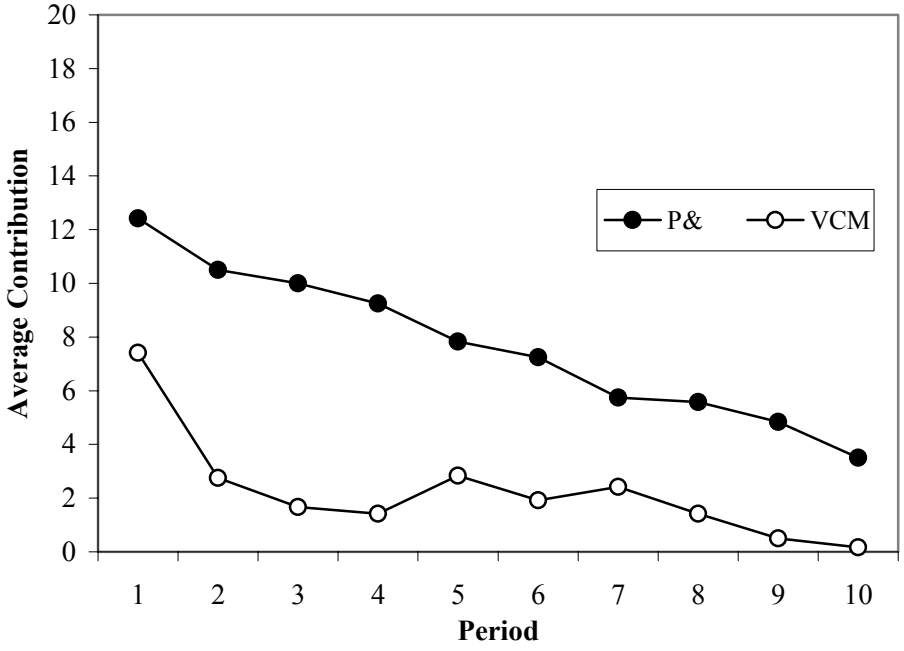
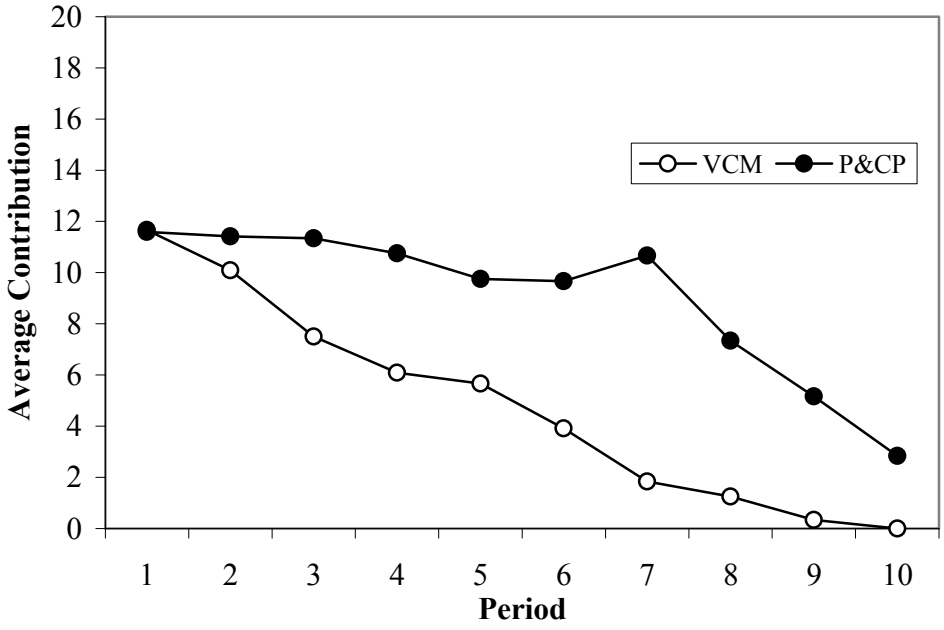


Figure 6: Average contribution over time in the partner-treatment (session 4)

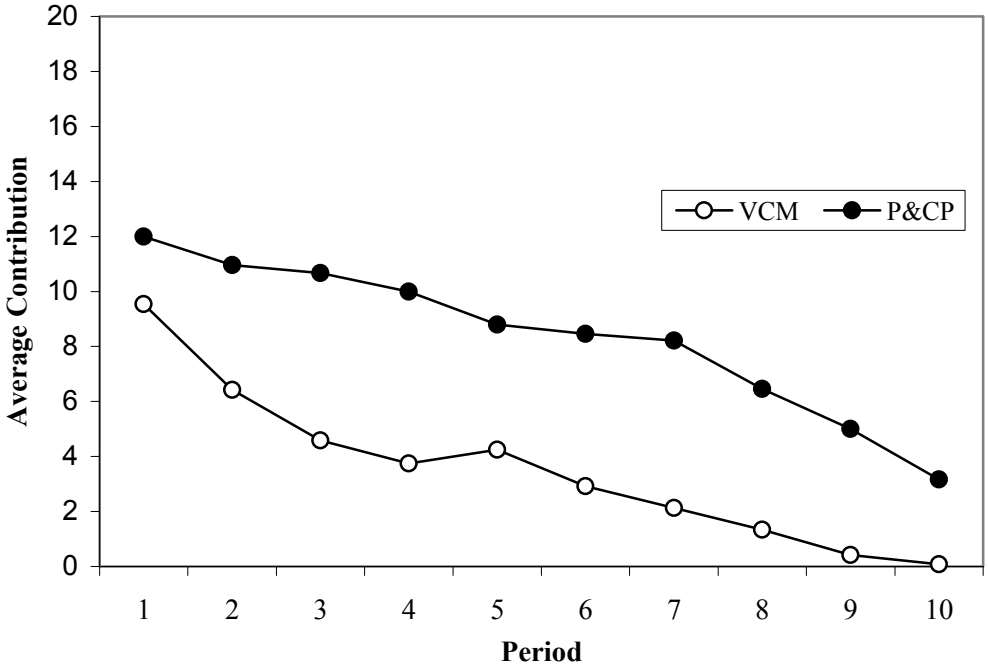


Figures 5 and 6 show again that counter-punishment draws away some of the power that punishment had to discipline free riders. In both sessions, the average contribution to the public good in the P&CP treatment initially is roughly 12 ECUs and then follows a similar negative trend and settles at approximately 3.5 ECUs. The VCM treatment has the same characteristics as in most reported experiments. People are conditionally cooperative and begin by contributing a significant fraction of their endowment which varies between 40 percent (session 1) to 60 percent (session 2). However, soon the free-riders drug the cooperation in both cases down until it reaches almost complete free riding.

In the P&CP treatment there is on average a higher level of contribution, which might reflect the hopes of the subjects that they will be able to control the free riders and also their own fear of being punished. Still in both cases, the subjects, as they become more experienced, start contributing less and average contribution converges towards free-riding.

If we put the results of the two sessions together we get figure 7. Average contribution declines at a similar rate in both treatments. The contribution level difference which was not present in the stranger-treatment might be attributed to the willingness to avoid disapproval (Masclet et al.[2003])or at the repeated interaction between the participants (Fehr and Fischbacher [2003]) .

Figure 7: Average contribution over time in the partner-treatment (session 3 & 4)



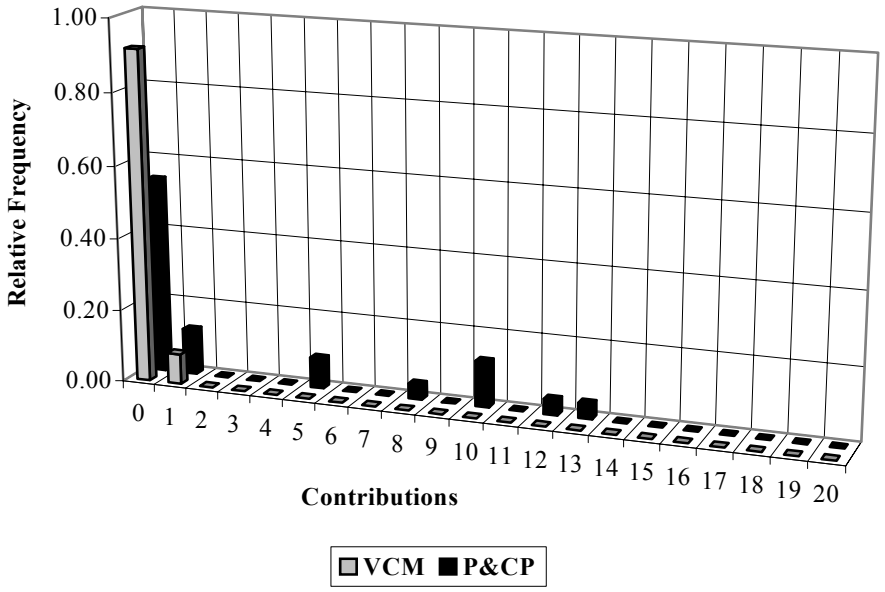
Our last result concerning the partner-treatment has again to do with the behavior at the individual level in the final period.

Result 6: *In both treatments, free-riding emerges as the modal action.*

Evidence for result 6 is drawn from the histogram in figure 8, which shows the relative frequency of contributions in the final period. As we can see, for both treatments zero contribution is the mode. In the P&CP condition, 54 percent of the participants choose to free-ride completely and 13 percent

more to contribute just one ECU[16]. There are some individuals with higher contributions. In the VCM treatment, 92 percent decide to free-ride completely and the remaining 8 percent contribute one ECU.

Figure 8: Distribution of contributions in the final periods of the partner-treatment.



4.3 *Willingness to punish*

So far we have shown that the introduction of counter-punishment opportunities has a drastic effect to the level and the evolution of average contribution. The initial contributions in F&G are very similar to ours, however, as the experiments proceed the results diverge. In F&G, as well as in other experiments using one-sided punishment, average contribution increases with repetition. In our experiment, average contribution decreases in both sessions, at a rate similar to that of the treatment without punishment, and it tends towards full free riding. The question that arises therefore is what triggers this different behaviour?

Punishment is a second order public good since everyone benefits from its existence, but every individual would rather avoid its cost. The

[16] This is a vast departure from the 82.5 percent of participants who chose to cooperate completely in F&G when counter-punishment was absent.

possibility of counter-punishment and the uncertainty of its harshness make punishment more costly and people less willing to punish. If this is the case indeed, we should observe a decline in the number of sanctions, which would then explain the existence of free riding.

To have a basis for comparison we will juxtapose the evolution of the average number of sanctions over time from this experiment and the one of F&G. Our findings are summarized by result 7.

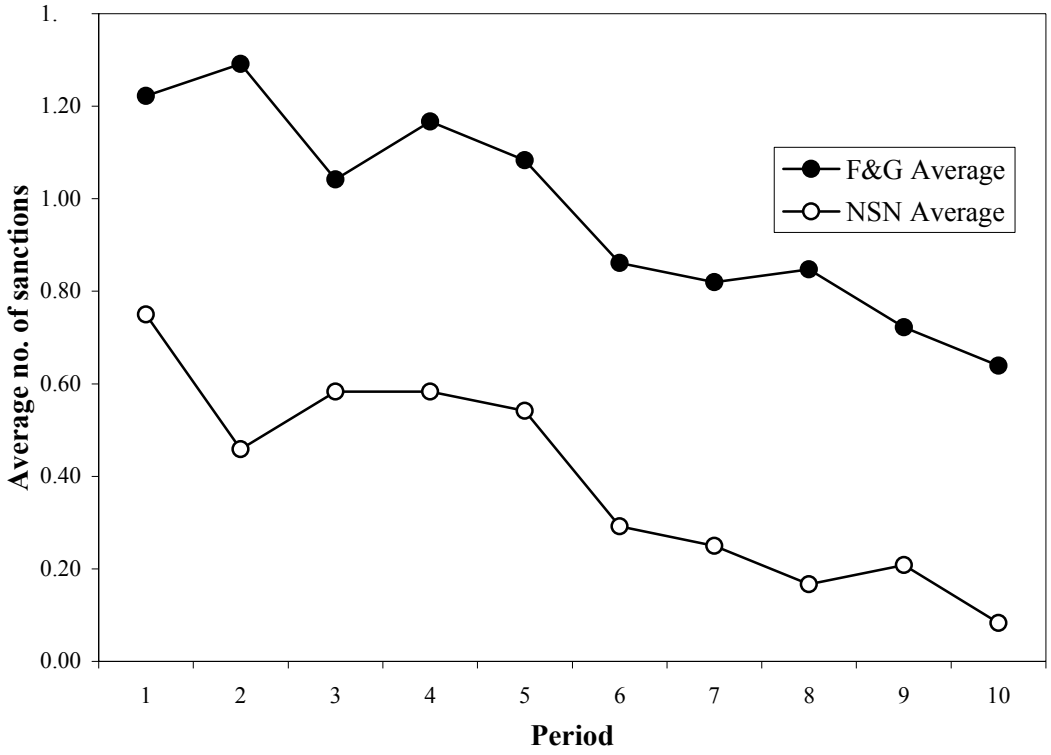
Result 7: W*hen we allow for counter-punishment, the average number of sanctions decreases significantly in both the partner and the stranger-treatment, even though average contribution declines.*

Evidence for Result 7 is drawn from figures 9 and 10, which depict the evolution of the average number of sanctions over time. As we can see in figure 9, in the stranger-treatment of F&G there is a decline in the average number of sanctions over time reflecting mainly the increase on the level of contribution. The average number settles at approximately 0.65[17]. This implies that the participants, having realized the effectiveness of punishment, try to push the last non-cooperators to contribute more until the last moment.

In our experiment, the average number of sanctions starts at a much lower level and pursues a similar course, which could now be attributed to the realization that the threat of punishment can not alleviate free-riding and also that punishment can be punished. In the final period, almost no one punishes[18].

[17] An average of "0.25" implies that on average there was one sanction per group. An average of "1" implies that on average there were 4 sanctions per group i.e. one per player.
[18] The existence of participants who are willing to punish in the last period is an indication of the willingness of people to punish even when they do not anticipate any material benefits.

Figure 9: Evolution of the average number of punishments sanctions in the stranger-treatment

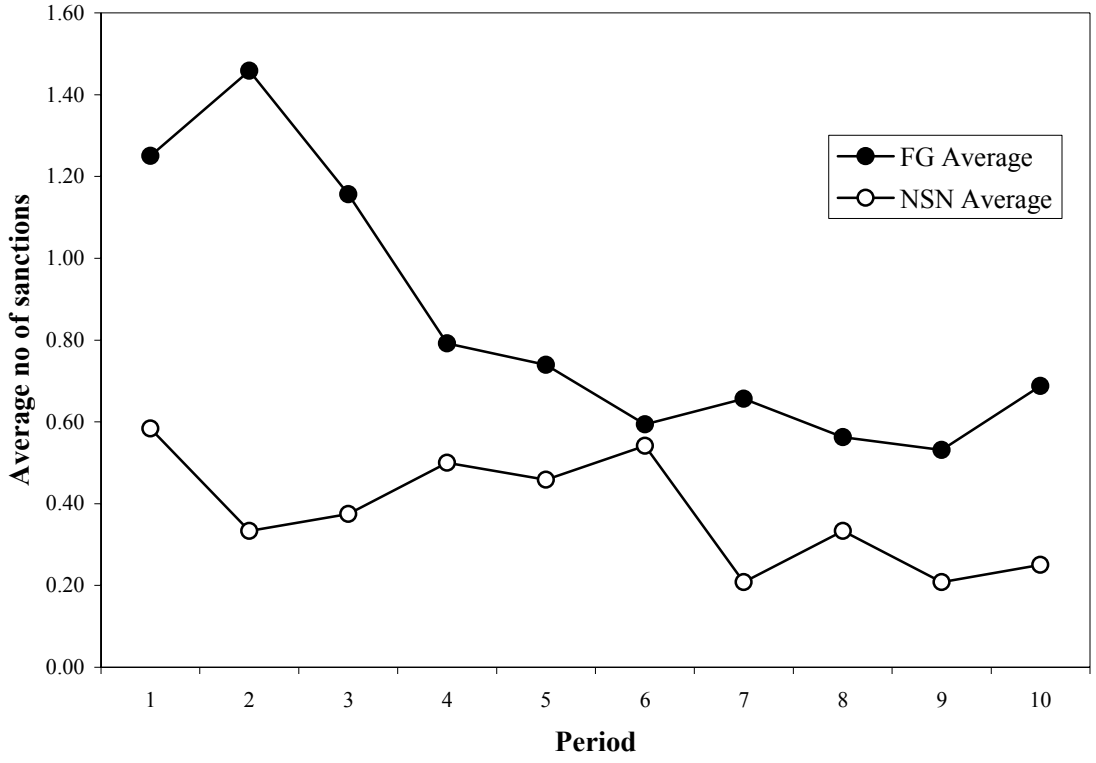


In the partner-treatment, after the second period, the average number of sanctions in F&G falls sharply following the increase in the subjects' cooperation levels and continues to do so with minor increases until it is finally stabilized around 0.6.

In our experiment, the average number of sanctions in the partner-treatment is also downward sloping, but moves at much lower levels. Between periods 3 and 5 there appears to be an effort from the cooperators side to discipline the free riders in order to stop declining contribution. However, after period 5 there is a major decline on the average number of punishments possibly reflecting the "surrender" of the cooperators and the average settles at very low levels (less than one punishment action per group). The difference of the average number of sanctions across treatments

is statistically significant under both protocols according to a Mann-Whitney U test with the average number of sanctions per group as observations[19].

Figure 10: Evolution of the average number of punishments sanctions in the partner-treatment



Result 7 becomes even more remarkable if we take to consideration the fact that in our experiment, where average contribution was at a much lower level, participants had a more serious reason to want to punish. On the other hand, in the experiment by F&G, average contribution was constantly increasing approaching full cooperation eliminating the reasons for punishment. These findings lend support to the hypothesis that counter-punishment makes people less willing to punish

[19] These results are supported by our findings in the control treatment. The difference between the average willingness to punish in the control treatment (NSN) and session 3 of F&G is not significant as it becomes apparent from figure 14 to be found in the appendix.

4.4 *Effectiveness of punishment*

The effect that counter-punishment has on the willingness to punish is not the only one: counter-punishment appears to diminish the effectiveness of punishment.

In F&G, 89 and 78 percent of the participants increased their contribution in the partner and in the stranger-treatment respectively, after they were punished. The average increase was 4.6 ECUs under the partner and 3.8 ECUs under the stranger protocol. In this experiment only 30 percent (partners) and 27 percent (strangers) increased their contribution level by an average of 3.6 ECUs and 4 ECUs respectively, following a punishment. So why are people less responsive to punishment?

First, we have to see whether the actual size of the punishments is now different i.e. do people punish more lightly in order to avoid retribution? In the partner-treatment of F&G, the weighted average size of punishment was 1.71, whereas in this experiment it was equal to 2.20[20]. So, if anything, participants punished even more on average when counter-punishment was present. The answer, therefore, to the previous question can not be found here.

The situation is reversed in the stranger-treatment, where the weighted average size of punishment in F&G was 1.90, in contrast to the 1.47 of our experiment[21]. In this case, therefore, part of the observed lack of reaction to punishment might be attributed to the lower average size of punishment.

This fact is somewhat surprising since even people who did not counter-punish were unwilling to raise their contribution. An explanation to this might be that participants, sensing the modest willingness of cooperators to punish free riders, were aware of the constant decline in

[20] Under the partner protocol, the average size of punishment in session 1 and session 2 of F&G was 1.73 and 1.68 respectively. In this experiment the average size of punishment was 1.86 and 2.79 for session 1 and session 2 respectively.

[21] Under the stranger protocol, the average size of punishment in session 1, session 2 and session 3 of F&G was 1.96, 1.90 and 1.85 respectively. In this experiment the average size of punishment was 1.36 and 1.57 for session 1 and session 2 respectively.

cooperation. As a result, participants who got punished chose not to raise their contribution in subsequent rounds pre-empting the decay of cooperation.

Result 8: *In the presence of counter-punishment, people react less to punishment.*

4.5 *Payoff Consequences of Two-Sided Punishment*

We saw earlier that the Nash equilibrium in the voluntary contribution mechanism with zero contribution (i.e. g_i=0) and an individual payoff of 20 ECUs, is not the Pareto-dominant, welfare-maximizing solution, where g_i=20 and the individual payoff equals to 32 ECUs. It has been shown [Fehr and Gaechter (2000)] that punishment alone can force people to cooperate and though it comes with a cost (i.e. money given to buy punishment points and the income loss that punishment implies) it eventually leads to an improvement for the society as whole.

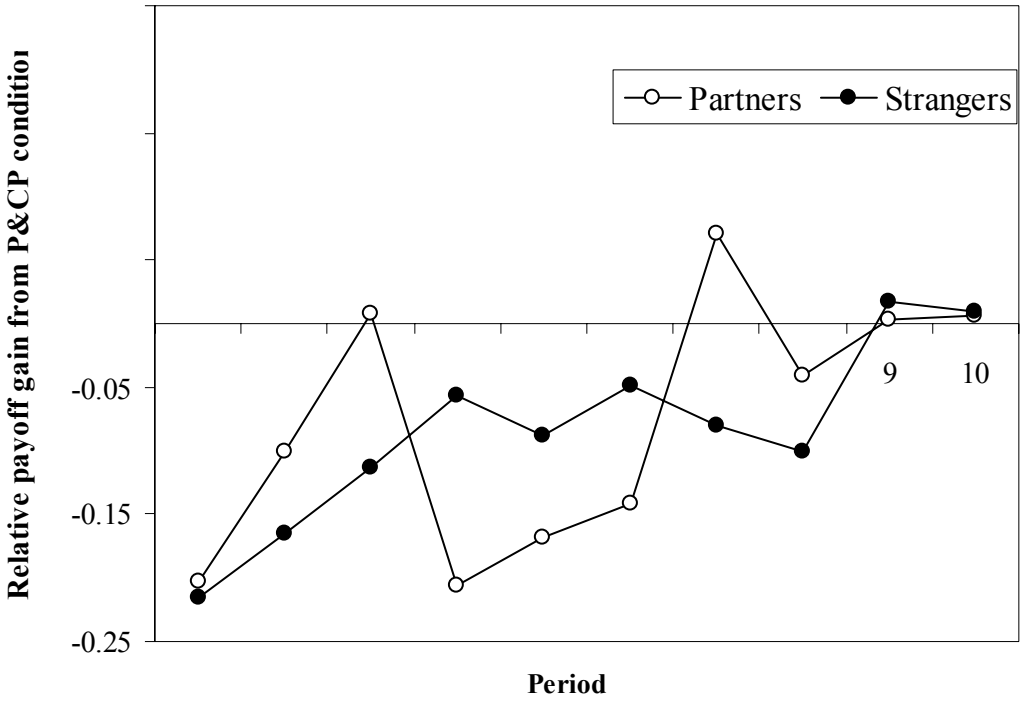
As pointed out earlier, in the real world, punishment is not restricted to one-sided and is often followed by an act of counter-punishment. Counter-punishment on the other hand implies additional costs and is a weapon on the hands of the free riders. One key question, therefore, to be answered is: how does the punishment option together with counter-punishment affect the average group payoff? Is the group better off now than before? Or, more simply, was Mahatma Gandhi right when he predicted: "An eye for an eye will make us all blind"?

To answer this question we calculate the relative payoff gain of the punishment and counter-punishment, which is equal to the difference of the average group payoff of between two treatments normalized by the average group payoff of the no punishment treatment. In mathematical terms:

$$relative.gain = \frac{aver.group.payoff.P\ \&\ CP - aver.group.payoff.VCM}{aver.group.payoff.VCM}$$

Figure 11 depicts the payoff consequences that counter-punishment and punishment have over time in both the stranger and partner treatment. As we can see, the squander of the endowments in sanctions and counter-sanctions leads to a relative payoff loss; in 8 out of 10 periods in the stranger treatment and in 7 out of 10 in the partner.

Figure 11: Payoff consequences of punishment and counter-punishment in the partner and stranger treatment



Under the stranger protocol there is an almost constant convergence towards 0, which is the result of the declining number of punishments, which almost reach zero at the last periods (see figure 9) and of the almost identical contributions between the VCM and the P&CP treatment (figure 3). In the partner-treatment, however, where punishment is more effective in raising contributions and there are implicit opportunities for coordination, relative payoff follows a more turbulent path. In the last two periods, in both

conditions the relative difference approaches zero, which implies that the average payoff in the different treatments is approximately identical. This is the result of similar contributions and the declining number of sanctions.

This finding indicates how harmful mutual monitoring can be to a society. It also demonstrates that in the presence of counter-punishment, where the controlling of the free riders is harder, participants might be better off free riding and avoiding costly punishment. In combination to the previous results it serves as a sign that counter-punishment might lead, eventually, to similar outcomes to the treatment where no punishment was possible i.e. similar contributions, no punishments and similar payoffs.

Result 9: *Under both protocols, punishment with counter-punishment leads to a relative payoff loss for most of the experiment until the participants learn to behave as in the VCM treatment i.e. not contribute and not punish.*

4.6 *What drives punishment?*

The standard economic theory predicts that participants will neither punish nor counter-punish given that this is costly and their actions are expected to yield no material benefits. Our findings contradict this prediction.

Under the partner protocol there were 91 sanctions 30% of which were answered back. Of the latter, 40.7% punished with more points than the ones received. The average size of counter-punishment was 2.29 points opposed to the average size of punishment which was 2.20.

Under the stranger protocol, 20% of the 94 sanctions were answered back. Out of them 73.7% answered back giving as many counter-points as the punishment points they received. Only 15.8% were answered back giving more points than those received. This might be an indication that under the partner protocol people tried to establish a reputation for the group that punishment will not be tolerated. The average size of counter-punishment was 1.60 points, when the average size of punishment was 1.47.

The high propensity towards punishment makes it important to identify the forces behind the participants' punishing activities. To do this we run a Tobit regression on the following model as in Masclet et al. (2003):

$$p_i^j = \alpha_0 + \alpha_1[\max(0, c_i^t - c_j^t)] + \alpha_2[\max(0, c_j^t - c_i^t)] + \alpha_3[\max(0, c_i^t - \overline{c^t})] + \alpha_4[\max(0, \overline{c^t} - c_i^t)]$$

where p_i^j is the punishment points that player i assigns to player j, c_i^t is player i's contribution in period t and $\overline{c^t}$ is the average contribution of in period t of i's group. To model time effects we also include period dummies along with a group dummy for the partner-treatment and a session dummy for the stranger-treatment. We wish to see whether the introduction of counter-punishment has altered the significance of the independent variables. The results from the regression can be found in table 4.

Table 4: Determinants of punishment

	Dependent Variable: punishment points given	
Independent Variables	Strangers	Partners
Negative Deviation from punisher's contribution (α_1)	0.227***	0.252***
	(0.050)	(0.084)
Positive Deviation from punisher's contribution (α_2)	-0.088	0.091
	(0.091)	(0.085)
Negative Deviation from Group's Average contribution (α_3)	0.150*	0.090
	(0.082)	(0.120)
Positive Deviation from Group's Average contribution (α_4)	0.049	-0.231
	(0.136)	(0.167)
Constant (α_0)	-4.321***	-4.307***
	(0.794)	(1.053)
Observations	720	720
Uncensored Observations	94	91

Notes: Numbers in parentheses indicate standard errors
*** Significant at the 1-percent level
** Significant at the 5-percent level
* Significant at the 10-perecent level

As we can see in both treatments the main driving force of punishment seems to be the negative deviation from the contribution of the punisher. The less the punished contributes in comparison to the punisher the higher

the punishment will be. This result is in agreement with the findings of Masclet et al. (2003) and Falk et al. (2000) and might be taken as an indication of spitefulness. In the stranger-treatment, negative deviation from the group's average is significant at the 10-percent level, but the same does not apply for the partner-treatment. This is a surprising result since we would expect the group standard to be more important under the partner protocol as in F&G (2000). Also surprising are the positive sign for the coefficient of α_2 in the partner-treatment and of α_4 in the stranger-treatment[22].

Due to the limited number of counter-punishments, it is impossible to derive a meaningful explanation for its determinants through regression analysis. It appears that a small fraction of the participants dislike punishment so much that they counter-punish at every opportunity given, whereas others seem to take in consideration the initial cause of their punishment i.e. their contribution in relation to the others in their group and punish only when they feel they did not deserve punishment.

4.7 *Selfish vs. Altruistic individuals*

The careful reader might have noticed in table 3 a difference in contributing behaviour between the different groups. This observation makes a deeper look at the individual actions essential.

In general, in contrast to the experiments with one-sided punishment there seems to be a big variation in individual activities that are alleviated towards the end. Under both protocols, the initial contributions vary from 0 to 20 ECUs. Most of the subjects appear to decrease their contribution over time, some keep it relatively constant at either high or low levels of contribution and some appear to be undecided about whether to contribute a

[22] Masclet et al. (2003) run the same regressions, though only for the partner-treatment and found all of the aforementioned variables to be significant at the 1-percent level.

lot or little. Some individuals contribute zero throughout the P&CP treatment[23]. These are summarized in tables 5 and 6[24].

Table 6 is particularly useful since we can observe how the actions of a participant affect the future decisions of the other group members. It appears that it takes only one determined free-rider to bring cooperation down. This cannot be better illustrated than in the case of group 6 (participants 21-24), where 3 participants were strong supporters of cooperation contributing for most of the experiment 20 ECUs. Subject 22, who contributed not more than 13 ECUs at any instance, forced the other three members to drop substantially their contributions from period 7 onwards. Note that none of the cooperators used punishment extensively. The ability of the free riders to obliterate cooperation under this set up can also be seen in the cases of group 3 (subjects 9-12), group 4 (subjects 13-16) and in lesser extent group 2 (subjects 5-8).

Another notable case is group 5: subject 20, a strong reciprocator[25], spent most of his money in the experiment to sanction the other group members[26]. However, his 77 points(!) were not enough to increase cooperation within the group. The consequence was that by the end of the experiment he had also decreased his contribution.

An enlightening exception to this is group 1 (subjects 1-4). All four members were like-minded people whose initial contributions did not vary greatly. As a result, though they could not increase cooperation, they were able to sustain it at the initial levels. All these are summarized in result 10.

Result 10: *The level of cooperation when counter-punishment is allowed depends on whether or not selfish individuals exist: one determined selfish*

[23] It is interesting to observe that most of these participants also spend no money on punishment activities.
[24] See appendix.
[25] A "strong reciprocator" is an individual willing to engage in costly activities, even when this yields no future material benefits for him (Fehr and Fischbacher [2003]).
[26] As a result, subject 20 won, by far, the least money amongst the participants in all sessions.

individual can obliterate cooperation like in the VCM treatment. Cooperation seems possible only between like minded individuals.

5. Conclusion

In the last years there has been a considerable amount of papers indicating the significance of mutual monitoring and decentralised sanctions in the provision of public goods. These papers show that contrary to standard economic theory people are willing to punish and under this threat contribution levels rise significantly. However, often, we can not allow for punishment excluding counter-punishment. Our hypothesis is that punishment elicits negative emotions amongst the punished, which in turn might lead to counter-sanctions.

Our results show that when we introduce counter-punishment, punishment stops being a valid mechanism for the discipline of selfish individuals and the efficient provision of public goods. Under both the stranger and the partner protocol, contributions decrease over time and in some cases approach full defection.

The reason behind this behaviour is the decreased willingness of cooperators to turn into punishment activities in order to alleviate free riding. In this environment, one determined free rider appears to be enough to bring down cooperation.

Mutual monitoring amongst individuals is now a harmful devise since it leads to a large squander of resources without any beneficiary result up to the point where participants realise that they can not control the free-riders and give up cooperating. In our opinion, this serves as a warning that in many cases people are unable to achieve cooperation and a formal independent body is needed to enforce it.

The situation might even be understated. We believe that one of the characteristics of the individuals who chose to free ride in the real world is often their relative "strength" to the cooperators. In that case, people might be even less willing or not willing at all to punish free riders in fear of

a severe counter-punishment. An additional reason which affects the willingness to punish negatively might be the group size; punishment is a second order public good, counter-punishment, however, is not. As a result, we believe that the greater the group size is the weaker the incentive to counter-punishment will be.

Our results are related to that of Carpenter (2002) who shows that when the price of punishment increases the demand for it decreases. This diminishes the threat of punishment and leads to a raise in free-riding. In an indirect way, the threat of counter-punishment increases the price an individual has to pay in order to punish. However, in our view, punishment comes always at an (expected) high cost since counter-punishment amongst agents cannot be separated from sanctions.

References:

Bowles, S., J. Carpenter, and H. Gintis, (2001): "Mutual monitoring in teams: The effects of residual claimancy and reciprocity, mimeo.

Burlando, R. and J.Hey (1997): "Do Anglo-Saxons free ride more?", *Journal of Public Economics,* 64, 41-60.

Camerer, C. and R. Thaler (1995): "Anomalies: Ultimatums, Dictators and Manners", *Journal of Economic Perspectives*, 9:2 (spring), 209-219.

Carpenter, J. (2002): "The Demand for Punishment", mimeo.

Davis, D. and C. Holt (1993): *Experimental Economics,* New Jersey: Princeton University Press

Dufwenberg, M. and U. Sneezy (2000): "Price competition and market concentration: an experimental study", *International Journal of Industrial Organization*, Vol. 18, 7-22.

Falk, A., E. Fehr, and U. Fischbacher, (2000): "Informal Sanctions", Working paper, University of Zurich

Fehr, E. and U. Fischbacher (2003): "The nature of human altruism", *Nature*, Vol.425, 785-791.

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

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

Fehr, E., S. Gaechter, and G. Kirchsteiger (1997): "Reciprocity as a constant enforcement device- Experimental Evidence", *Econometrica,* Vol. 65 (4), 833-860.

Fischbacher, U. (1999): "z-Tree: Zurich Toolbox for Readymade Economic Experiments. Instruktionen für Experimentatoren", mimeo, University of Zurich.

Forsythe, R., J. Horowitz, N. Savin, and M. Sefton (1994): "Replicability, Fairness, and Pay in Experiments with Simple Bargaining Games", *Games and Economic Behavior* 6,3 (May 1994):347–369.

Girard, R. (1979): *Violence and the Sacred*, Johns Hopkins University Press.

Gueth, W. and R. Tietz (1990): “Ultimatum Bargaining Behaviour – A survey and comparison of experimental results”, *Journal of Economic Psychology,* 11:3 (September), 417-449.

Isaac, Walker, and Thomas (1984): “Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations”, *Public Choice,* 43, 113-149.

Kagel, J., and Roth A. (1995): *Handbook of Experimental Economics,* ed. by John Kagel and Alvin Roth, Princeton University Press

Ledyard, J. (1995): “Public Goods: A survey of experimental research” in: Handbook of Experimental Economics ed. by J. Kagel and A. Roth, Princeton University Press

Marwell, G. and R. E. Ames (1981): "Economists Free Ride, Does Anyone Else? Experiments of the Provision of Public Goods," *Journal of Public Economics*, 15, 295-310.

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

Ostrom, E., J. Walker and R. Gardner (1992): “Covenants with and without a sword: Self governance is possible”. *American Political Science Review,* Vol. 86 (2), 404-417

Page, T. and L. Putterman (2000): “Cheap talk and punishment in voluntary contribution experiments”, Working paper

Roth, A. (1995): “Bargaining Experiments”, in *Handbook of Experimental Economics,* ed. by John Kagel and Alvin Roth, Princeton University Press

Sefton, M., R. Shupp, and J. Walker (2002): “The effect of rewards and sanctions in provision of public goods”, Working paper

Zizzo, D.J, and A. Oswald (2001): “Are people willing to pay to reduce others’ incomes?”, Annales *d’Economie et de Statistique,* 63-64, 39-62.

Appendix:

A.1 Instructions

These are the instructions we used in session 3. For maximum comparability we tried to adopt the instructions of F&G (2000). The instructions to the other sessions are available from the author upon request.

You are now taking part in an economic experiment. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions.

The instructions which we have distributed to you are for your private information. **Please do not communicate with the other participants during the experiment.** Should you have any questions please ask us.

During the experiment we shall not speak of Pounds, but of Experimental Currency Units (ECU). Your entire earnings will be calculated in ECUs. At the end of the experiment the total amount of ECUs you have earned will be converted to Pounds at the rate of **1 ECU = 4 p** and will be immediately paid to you in cash.

At the beginning of the experiment the participants will be randomly divided into groups of four. You will therefore be in a group with 3 other participants. **The composition of each group will remain the same throughout the experiment.** The experiment lasts 10 periods and each period is divided into 3 stages.

The first stage:

At the beginning of each of the 10 periods each participant will receive 20 ECUs. In the following, we shall refer to this amount as the "endowment".

In the first stage, your task is to decide how to use your endowment. **You have to decide how many of the 20 ECUs you want to contribute to a project** (from 0 to 20) **and how many of them to keep for yourself**. The consequences of your decision are explained in detail below.

Once all the players have decided their contribution to the project you will be informed about, the group's total contribution, your income from the project and your payoff in this period. Your payoff in each period is calculated using the following simple formula. Again, if you have any difficulties do not hesitate to ask us.

Income from the 1st stage = Endowment of ECUs - Your contribution to the Project + 0.4*Total contribution to the Project

This formula shows that your first stage income consists of two parts:

1) The ECUs which you have kept for yourself (endowment – contribution)
2) The income from the project, which equals to the 40% of the group's total contribution.

The income of each group member from the project is calculated in the same way. This means that each group member receives the same income from the project. Suppose the sum of the contributions of all group members are 60 ECUs. In this case, each member of the group receives an income from the project of: 0.4*60=24 ECUs. If the total contribution to the project is 9 points, then each member of the group receives an income of: 0.4*9=3.6 ECUs from the project.

You always have the option of keeping the ECUs for yourself or contributing them to the project. Each ECU that you keep raises your end of period income by 1 ECU. Supposing you contributed this point to the project instead, then the total contribution to the project would rise by 1 ECUs. Your income from the project would thus rise by 0.4*1=0.4 ECUs.

However, the income of the other group members would also rise by 0.4 ECUs each, so that the total income of the group from the project would be 1.6 points. Your contribution to the project therefore also raises the income of the other group members. On the other hand you also earn an income for each point contributed by the other members to the project. In particular, for each point contributed by any member you earn 0.4 ECUs.

In addition to the 20 ECUs per period, each participant receives a one-off lump sum payment of 25 ECUs at the beginning of this part. This one-off payment can be used to pay for eventual losses during the experiment. **However, you can always evade losses <u>with certainty</u> through your own decisions.** Note that this lump sum payment will not be used to calculate the income from the period. It will only be added to your total income from all the periods at the very end.

<u>The second stage:</u>

At the second stage you will be informed how much each group member contributed individually to the project at the first stage. At this stage you can **reduce or leave equal** the income of **each** member of your group by **distributing points**. The other group members can also reduce your income if they wish to.

If you choose 0 points for a particular group member, you do not change his or her income. However if you give a member 1 point you reduce his or her income by 10 percent. If you give a member 2 points you reduce his or her income by 20 percent, etc. The amount of points you distribute to each member determines, therefore, how much you reduce their income from the first stage. If one player receives in total 4 points his income will be reduced by 40% and if he receives **10 or more** his income from the first stage will be reduced by 100%.

If you distribute points you have costs in ECUs, which depend on the amount of points you distribute. You can distribute between 0 and 10 points to each group member. The more points you give **to any** group member, the higher your costs. Your total costs are equal to the **sum of the costs of distributing points to each of the other three group members**. The following table illustrates the relation between distributed points to **each** group member and the cost of doing so in ECUs.

Points	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Cost of points per person	*0*	*1*	*2*	*4*	*6*	*9*	*12*	*16*	*20*	*25*	*30*

Example: Supposing you give 2 points to player 1 this costs you 2 ECUs; if you also give 8 points to player 3 this costs you a further 20 ECUs; and if you give 0 points to the last group member this has no additional cost for you. In this case, your total costs of distributing points would be 22 ECUs (2+20+0) and not 30 ECUs.

Your total income from the two stages is therefore calculated as follows:

Total income (in ECUs) at the end of the 2nd stage =

= (Income from the 1st stage)*[1-(1/10)*received points] - costs of distributed points , where received points can't be more than 10 points

Please note that your income in ECUs at the end of the *second* and the *third* stage can be negative, if the costs of your points distributed exceeds your (possibly reduced) income from the 1st stage. **You can however evade such losses with certainty through your own decisions.** Should your income become zero or negative at the end of the second stage you will not be able to continue to the third stage. If your income becomes zero or negative at the end of the third stage you can simply use your 25 ECUs that we gave you in the beginning in order to pay this off.

The third stage:

In the third and final stage, after being informed of the points that the other group members assigned to you, you will be given one last opportunity of assigning points back to the other participants, thus reducing their income. We shall call this points "counter-points". **You will only be able to assign counter-points to participants who assigned points to you during the second stage**

The costs of assigning points, as well as the income reduction caused by each point remain the same as before. **Note**: if you distribute 2 points in the second stage to player 1 you have a cost of 2 ECUs If in the third round you decide to distribute 3 more to player 1 then the total cost is 9 ECUs.

Your profit at the end of the period is equal to:

Total income (in ECUs) at the end of the period =
= (Income from the 2nd stage)*[1-(1/10)*received counterpoints] - costs of distributed counterpoints , where received counterpoints can't be more than 10 points

If you have any further questions please raise your hand and one of the supervisors will come to help you.

Control Questionnaire

1. Each group member has an endowment of 20 ECUs. Nobody (including you) contributes any ECUs to the project. What is:

a. Your income at the end of the first stage? ……

b. The income of the other group members?……

2.Each group member has an endowment of 20 ECUs. You contribute 20 ECUs to the project. All other group members contribute 20 ECUs each to the project. What is:

a. Your income at the end of the first stage? ……

b. The income of the other group members?……

3.Each group member has an endowment of 20 ECUs. The other three group members contribute together a total of 30 ECUs to the project. What is:

a. Your income at the end of the first stage if you contribute 0 ECUs to the project? …….

b. Your income at the end of the first stage if you contribute 15 ECUs to the project? …….

4. Each group member has an endowment of 20 ECUs. You contribute 8 ECUs to the project. What is:

a. Your income at the end of the first stage if the other group members together contribute a further total of 7 ECUs to the project?.....

b. Your income at the end of the first stage if the other group members together contribute a further total of 22 ECUs to the project?.....

5.At the second stage you distribute the following points to your three other group members: 9, 5, 0. What are the total costs of your distributed points?….

6.What are your costs if you distribute 0 points? ……

7.By how many percent will your income from the first stage be reduced when you receive from the other group members a total of:

a. 0 points? …

b. 4 points? ...

c. 15 points? …

8. At the second stage you distribute the following points to your three other group members: 2, 2, 0. In the third stage you distribute the following points to your three other group members: 1, 1, 1. What are the total costs of your distributed points?….

A.2 Graphs from the control treatment

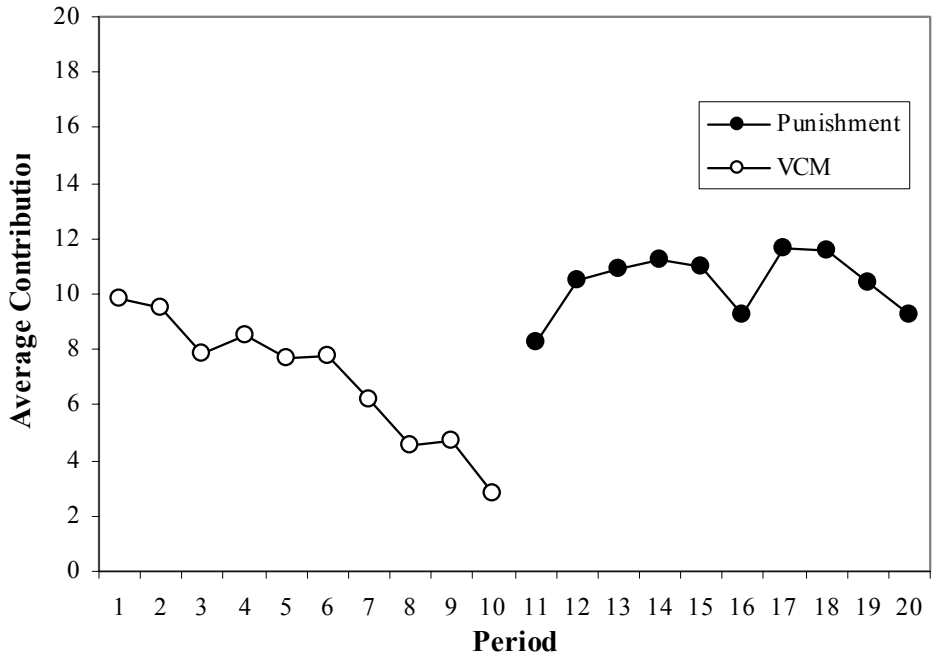
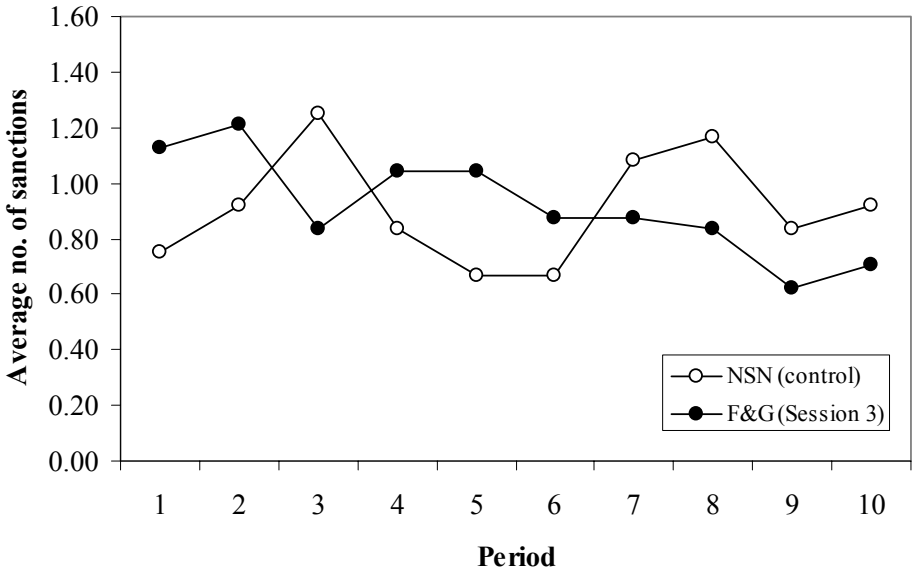


Figure 14: Evolution of the average number of punishments sanctions in the stranger-treatment



A.3 Brief summary of related articles

In the introduction, we mentioned a number of papers were one-sided punishment was used. Here we take a more detailed look in some of them.

Sefton et al. (2002) have a similar treatment to F&G with similar results; moreover they include a treatment with rewards, where participants can transfer a fraction of their endowment to someone else in their group, and a treatment with both sanctions and rewards. They find, amongst others, that subjects use both rewards and sanctions, but in a different way: initially rewards are at a higher level than sanctions, though they appear to decline faster and are not as capable for sustaining cooperation as sanctions. Sanctions, however, come at a higher social cost than rewards due to the income loss they entail. Their results show that the treatment that allows for both sanctions and rewards is the most effective for the production of a public good.

Masclet et al. (2003) study the influence of informal sanctions on the level of cooperation. They show that the results of F&G can partially be

attributed to the existence of informal sanctions. To do this, in addition to the punishment treatment of F&G, they used a treatment where no sanctioning was possible and another where individuals were allowed to express their dissatisfaction to others by assigning costly points to them. These points, in contrast to the F&G treatment, did not entail any income loss to their receiver. By comparing the differences in the average level of contributions between treatments, they show that informal sanctions can indeed play a significant role in the provision of public goods, though this requires repeated interactions and their influence tends to wear off.

Bowles et al. (2001) create a model for team production where the effectiveness of mutual monitoring depends on the willingness of some participants to engage in costly punishment. They then run an experiment identical to F&G where they were able to replicate their results. In addition they tested for the effect of the residual claim and the group size on the willingness to punish; they find the former to be positively correlated with the willingness to punish and the latter to be insignificant.

Carpenter (2002) argues that agents who forego material benefits to punish others can also be viewed as rational. Using a similar design to F&G and by altering the price of punishment points he shows that punishment is an ordinary and normal good.

Table 5- STRANGERS

Subject	Average contribution	Evolution of contribution	Punishments given		Punishments received		Counter-punishments given	
			No of sanctions	Total points	No of sanctions	Total points	No of sanctions	Total points
1	4.4	0,10,0,8,0,5,5,6,5,5	0	0	4	7	0	0
2	9.0	8,5,15,3,0,19,8,7,20,5	0	0	7	9	1	1
3	6.8	10,9,10,8,7,8,5,6,3,2	0	0	2	3	1	1
4	9.6	16,12,11,11,11,9,7,7,6,6	9	14	0	0	0	0
5	7.5	7,9,14,5,5,8,8,5,8,6	6	10	3	4	1	2
6	8.4	15,15,12,10,8,6,5,5,6,2	10	15	2	3	1	2
7	5.1	5,5,6,6,5,7,6,6,4,1	4	4	3	4	0	0
8	3.2	8,8,9,3,4,0,0,0,0,0	1	1	4	7	4	5
9	13.5	20,15,15,15,15,15,10,10,10,10	0	0	2	2	1	1
10	7.8	7,6,7,15,10,5,6,0,15,7	3	5	3	4	3	4
11	0.7	0,1,1,0,1,1,1,1,1,0	0	0	9	13	0	0
12	5.6	7,8,5,7,10,6,7,2,2,2	6	7	3	3	3	4
13	4.0	7,8,7,5,7,2,3,0,3,2	13	13	1	1	1	1
14	0.8	3,2,0,0,0,0,3,0,0,0	2	2	4	4	0	0
15	4.1	10,9,9,8,2,3,0,0,0,0	3	4	0	0	0	0
16	4.4	14,10,10,10,0,0,0,0,0,0	6	8	1	1	0	0
17	0.0	0,0,0,0,0,0,0,0,0,0	0	0	9	20	0	0
18	0.0	0,0,0,0,0,0,0,0,0,0	0	0	6	8	0	0
19	0.0	0,0,0,0,0,0,0,0,0,0	2	10	8	11	1	7
20	3.1	8,10,10,0,0,0,0,2,2,2	0	0	4	5	0	0
21	6.6	10,10,10,6,5,5,5,5,5,5	17	26	0	0	0	0
22	0.0	0,0,0,0,0,0,0,0,0,0	0	0	10	14	0	0
23	3.9	5,7,9,7,2,0,6,1,1,1	0	0	3	4	1	3
24	2.0	5,2,3,4,3,1,0,1,0,1	8	11	5	6	1	1

Subjects 1-12 took part in session 1 and subjects 13-24 in the session 2.
Contributions refer to the P&CP treatment.

Table 6- PARTNERS

Subject	Average contribution	Evolution of contribution	Punishments given		Punishments received		Counter-punishments given	
			No of sanctions	Total points	No of sanctions	Total points	No of sanctions	Total points
1	13.4	11,12,11,15,13,15,15,15,14,13	1	1	4	4	2	3
2	14.1	13,10,15,15,15,15,16,16,14,12	6	9	1	2	0	0
3	10.7	10,10,9,11,11,10,12,12,12,10	6	8	8	14	1	1
4	13.9	14,15,15,15,15,15,15,15,15,5	3	6	3	4	2	3
5	3.3	3,10,0,5,0,15,0,0,0,0	2	2	0	0	0	0
6	4.0	5,10,8,5,2,0,0,0,0,0	1	1	1	1	1	1
7	2.0	20,0,0,0,0,0,0,0,0,0	0	0	1	1	0	0
8	0.0	0,0,0,0,0,0,0,0,0,0	0	0	1	1	1	1
9	7.4	20,20,12,13,1,1,3,4,0,0	2	4	4	12	1	3
10	8.9	15,10,20,20,15,1,5,1,1,1	3	5	7	17	5	12
11	5.1	8,9,10,2,12,5,2,1,1,1	10	27	6	9	6	16
12	9.5	20,20,20,10,10,10,1,3,1,0	2	3	1	2	1	2
13	7.0	6,8,8,8,0,8,8,8,8,8	8	25	7	14	0	0
14	9.3	20,10,12,12,12,12,15,0,0,0	4	20	4	10	2	10
15	9.8	10,11,12,12,12,15,10,11,0,5	8	13	2	4	0	0
16	2.5	5,12,0,0,8,0,0,0,0,0	4	6	11	34	0	0
17	5.6	5,7,10,7,5,0,8,8,6,0	1	2	10	25	2	5
18	0.5	0,0,0,0,1,0,4,0,0,0	1	1	10	46	0	0
19	9.9	8,10,11,10,10,11,9,10,10,10	0	0	7	9	0	0
20	11.9	15,14,13,12,11,11,11,11,11,10	25	77	0	0	0	0
21	14.0	20,20,20,20,10,20,20,10,0,0	0	0	1	1	0	0
22	7.3	10,5,10,8,7,9,13,0,10,1	3	3	1	2	1	4
23	14.7	20,20,20,20,20,10,20,10,7,0	1	2	1	1	1	2
24	17.0	20,20,20,20,20,20,20,20,10,0	0	0	1	1	1	1

Subjects 1-12 took part in session 3 and subjects 13-24 in the session 4. Subjects 1-4 formed group 1, 5-8 group 2 etc.
Contributions refer to the P&CP treatment.