ISSN 0819-2642
ISBN 978 0 7340 4002 2


**THE UNIVERSITY OF MELBOURNE**

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 1036

April 2008

# Feedback, Punishment and Cooperation in Public Good Experiments

by

Nikos Nikiforakis

Department of Economics
The University of Melbourne
Melbourne Victoria 3010
Australia.

# Feedback, Punishment and Cooperation in Public Good Experiments*

Nikos Nikiforakis†

April 18, 2008

**Abstract**

A number of studies have shown that peer punishment can sustain cooperation in public good games. This paper shows that the format used to give subjects feedback is critical for the efficacy of punishment. Providing subjects with information about the earnings of their peers leads to lower contributions and earnings compared to a treatment in which subjects receive information about the contributions of their peers even though the feedback format does not affect incentives. The data suggest that this is because the feedback format acts as a coordination device, which influences the contribution standards that groups establish.

JEL Classification: C92, D70, H41

Keywords: feedback format, peer punishment, public good game, altruistic punishment, cooperation.

*I would like to thank John Creedy, Dirk Engelmann, Lata Gangadharan, Simon Loertscher, Hans-Theo Normann, Steve Tucker and Anne van den Nouweland for their comments on the paper. I would also like to thank Tim Cason, Dan Friedman, Simon Gächter, Jacob Goeree, Charles Noussair, and Charlie Plott for insightful discussions. The paper benefited from comments made at the 2007 International Meeting of the European Economic Association in Rome, and seminars at Maastricht University, the University of Canterbury, the University of East Anglia, the University of Melbourne, and the University of New South Wales. Funding from ESRC (project RES-000-22-0948) and the Faculty of Economics and Commerce at the University of Melbourne is gratefully acknowledged.

†Department of Economics, The University of Melbourne, Victoria 3010, Australia, tel: +61 383449717, fax: +61 38344 5104, email: *n.nikiforakis@unimelb.edu.au*

# 1 Introduction

The question of how to promote cooperation when private and collective interest are at odds is a recurring theme in economics. The interest in the topic stems from the difficulty of finding mechanisms that will eliminate the incentives to free ride when people are assumed to be self regarding. Recently, however, laboratory experiments have provided convincing evidence that many individuals have other-regarding preferences and are not only willing to cooperate with others, but also to punish free riders even when they cannot expect monetary benefits from their actions (e.g. Fehr and Gächter, 2000; 2002; Masclet, Noussair, Tucker, Villeval, 2003). The threat of punishment helps discipline free riders and leads to high levels of cooperation.[1]

The simplicity of the mechanism and its efficacy in promoting cooperation has attracted a lot of attention amongst economists and other social scientists (see Fehr and Schmidt, 2003). However, despite multiple studies on the subject, little is known about whether cooperation established under the threat of peer punishment is robust to institutional changes. The main reason is that all studies carried out since the publication of the seminal work by Fehr and Gächter (2000, 2002) have followed similar experimental protocols and used as a testing ground the public-good game as modified by Fehr and Gächter.

The two-stage game of Fehr and Gächter (2000, 2002) is as follows. In the first stage, each group member is given an endowment that he must divide between a private and a public account. Contributing the whole of the endowment to the private account is a dominant strategy, but contributions to the public account create a positive externality shared equally by all group members. In the second stage, individuals are informed about how much each of their peers contributed to the public account and are given the opportunity to reduce the earnings of any group member at a personal cost.

This paper investigates whether the feedback format used in the experiments is important for the efficacy of peer punishment in promoting cooperation and efficiency. A common characteristic of all public good experiments with peer punishment is that in the second stage participants receive feedback about each group member's *contribution* to the public account. However, experiments in oligopolistic markets have shown that subjects receiving information about the *earnings* of their peers behave more competitively and tend to be less cooperative than subjects who instead receive information about ei-

[1] See Anderson and Putterman (2006), Bochet, Page and Putterman (2006), Carpenter (2007a, 2007b), Fehr and Gächter (2000, 2002), Masclet, Noussair, Tucker, Villeval (2003), Nikiforakis and Normann (2008), Noussair and Tucker (2005), Page, Putterman and Unel (2005), Sefton, Shupp and Walker (2007).

ther the aggregate group production (Huck, Normann and Oechssler, 1999; 2000) or the individual levels of production (Offerman, Potters and Sonnemans, 2002) even though incentives are unaffected. That is, the way in which the same information is presented affects the likelihood of collusion.

One reason institutional details such as feedback format might affect cooperation is that, under certain conditions, peer punishment transforms the public good game from a social dilemma to a coordination game. Indeed, Fehr and Gächter (2000) write that subjects in their experiment "quickly established a *common* group standard that did not change over time" (p.992; emphasis in original). Fehr and Schmidt (1999; Proposition 5) formally show that the two-stage game has multiple Pareto-ranked equilibria if a single (sufficiently) inequity-averse individual exists. The patterns found in the data are consistent with the theoretical predictions of the model. For example, punishments are typically found to increase in severity as the deviation from the group's average contribution increases, while the average group contribution is itself not important in determing the extent of punishment (Anderson and Putterman, 2006; Carpenter, 2007b; Fehr and Gächter, 2000). In other words, punishments are not harsher when the overall level of contributions to the public account is lower.[2]

If peer punishment transforms the public good game to a coordination game as the data suggests, it seems surprising that groups never seem to adopt inefficient contribution standards and instead adopt nearly (or even fully) efficient standards. Van Huyck, Battalio and Beil (1990, 1991) were the first to show that groups repeatedly fail to coordinate at the payoff dominant equilibrium in games with multiple Pareto-ranked equilibria. The fact that the severity of punishment is not affected by the group's total contribution to the public account suggests that, apart from the threat of punishment, there must be other factors that promote cooperation. The evidence from oligopolistic experiments discussed above implies that one of these factors might be the feedback format used.

To see how feedback format can affect cooperation consider the examples in Tables 1a and 1b. Each of four players is given an endowment of \$20 and must decide how much to contribute to a public account. As the public account generates income for all group members the individual with the lowest contribution (Player 4) is the person with the highest income (and the player who will be most severly punished in period $t$). Place yourself in the role of Player 4 and try to imagine how much you would contribute in

[2]The experiment presented in this paper uses a fixed matching protocol (explained in detail in section 2). Hence, the findings discussed in the introduction do not refer to evidence from experiments employing random matching protocols.

period $t+1$ assuming your goal is to contribute as much as Players 1, 2 and 3. Do this separately for Tables 1a and 1b.

Insert Tables 1a and 1b here

The information in Tables 1a and 1b is equivalent (see section 2). However, the feedback format seems to highlight different aspects of the outcome: The feedback about individual contributions (*contribution feedback*) displayed in Table 1a emphasizes who is cooperating, who is not, and, ultimately, the social benefit of making contributions to the public account. In contrast, the feedback about individual earnings (*earnings feedback*) displayed in Table 1b makes salient to all players the private benefit associated with contributing to the private account. Since players choose their contributions simultaneously, they have to form expectations about the contributions of their peers. It is possible, that the different emphasis placed on social and private incentives affects expectations. Feedback format might therefore act as a coordination device helping groups select contribution standards. Consequently, subjects in a position similar to that of Player 4 could be more likely to adjust their contribution upwards when receiving contribution feedback rather than earnings feedback.[3]

The experiment presented in the following section aims to test whether feedback format can affect coordination, cooperation and efficiency in a public good game with peer punishment. The three treatments differ only with respect to the feedback format. In the first treatment, similar to previous experiments, individuals receive contribution feedback, while in the second treatment, participants receive only earnings feedback. The information in both treatments is equivalent as there is a clear one-to-one relation between contributions and earnings. Nonetheless, the different emphasis offered by the feedback format might affect the contribution standards chosen. The third treatment is of particular interest: Individuals receive both *contribution* and *earnings feedback*. Given that the two formats stress conflicting elements of the decision task individuals should find it harder to converge towards a contribution standard.

The results show that feedback format has a significant impact on the efficacy of peer punishment. Contributions are sustained at relatively high levels under contribution feedback. Earnings feedback has a negative impact on contributions. When subjects

[3] I asked several economists to take part in this exercise and tell me how much they would contribute in period $t+1$ in each scenario. Most replied that they would undoubtedly increase their contribution when faced with Table 1a. However, when faced with Table 1b, most of them responded that they would not increase their contribution as others would most likely lower their contribution or that they would increase their contribution slightly.

receive *only* earnings feedback cooperation breaks down. In line with the argument that feedback format affects coordination, subjects are found to have difficulty in establishing contribution standards when they receive feedback in both formats. This leads to progresively harsher punishments and earnings that are even lower than those predicted by the non-cooperative Nash equilibrium.

The rest of the paper is organized as follows. Section 2 introduces the experimental design. Section 3 presents the experimental results. Section 4 discusses alternative explanations for the results. Section 5 concludes.

# 2 The Experiment

The experiment was conducted in the Experimental Economics Laboratory at the University of Melbourne between February and April 2007. The 124 participants were students from the University of Melbourne recruited randomly using ORSEE (Greiner, 2004) from a pool of more than 1000 volunteers. Each subject took part in only one of the three treatments, and none of the subjects had previous experience with economics experiments.

Upon arrival at the laboratory, participants are seated in partitioned computer terminals and read *the same* set of instructions regardless of treatment.[4] Each participant must answer ten control questions before the experiment can begin. The questions aim to help participants understand the incentives in the game and how contributions translate to earnings and vice versa.

At the beginning of the experiment, subjects are randomly divided into groups of four individuals. The same group of individuals plays a finitely repeated public good game for 10 periods. That is, group composition remains unchanged throughout the experiment.[5] The game consists of two decision stages: the contribution and the punishment stage (referred to as stage 1 and stage 2 in the instructions). At the beginning of each period, each participant is given an endowment of E$20 (experimental dollars). In the first stage, players must decide simultaneously and without communication how much of the endowment to contribute to a public account, $c_i$, where $0 \leq c_i \leq 20$. The rest $(20 - c_i)$ remains in the player's own account. In addition to the money that player $i$ keeps, $i$

[4]The only difference in the instructions is one word that refers to the feedback format. Instructions were written in a neutral language.

[5]This is the obvious choice of matching protocol given the focus of the study which depends on the ability of groups to create common contribution standards. Under random matching the variance in contributions does not decrease over time (e.g. Fehr and Gächter, 2000). As a result, feedback format should not have the same (if any) effect when random matching is used.

receives a fixed percentage of the group's total contribution to the public account, 0.4. The earnings of player $i$ at the end of the first stage are given by equation (1) which was also used to generate the numbers in Tables 1a and 1b

$$\pi_i^1 = 20 - c_i + 0.4 \sum_{h=1}^{4} c_h. \tag{1}$$

At the beginning of the second stage each participant receives detailed feedback about stage one. The format used to provide feedback depends on the treatment. In treatment C individuals receive information about each group member's contribution, $c_i$ (*contribution feedback*), in treatment E individuals receive information about each group member's earnings at the end of stage one, $\pi_i^1$ (*earnings feedback*), while in treatment CE individuals receive information about each group member's contribution *and* earnings (see Table 2). Equation (1) shows that the different types of feedback are equivalent: An individual with higher earnings has contributed less to the public account. That is, $\pi_j^1 > \pi_i^1$ if and only if $c_j < c_i$.[6]

Insert Table 2 here

After participants are informed about the contributions/earnings of each group member subjects have to decide whether they wish to punish any of their group members. To do so they must purchase punishment points. Punishment is costly for the punisher as every point costs E$1. At the same time, each punishment point reduces the earnings of its recipient by E$2.[7] Let $p_{ij}$ denote the number of punishment points that player $i$ assigns to $j$ (where $i, j$=1, ..., $n$; $j \neq i$). Player $i$'s earnings at the end of the period are accordingly

$$\pi_i = 20 - c_i + 0.4 \sum_{h=1}^{4} c_h - \sum_{\substack{i=1 \\ i \neq j}}^{4} p_{ij} - 2 \sum_{\substack{j=1 \\ j \neq i}}^{4} p_{ji}. \tag{2}$$

The maximum number of points a participant can distribute to others is equal to his

[6] One could argue that calculating the earnings of each individual using contribution feedback is simpler than calculating the contribution of each group member using earnings feedback. To avoid this confound, at the end of the first stage subjects are also informed about the group's total contribution, $\sum_{h=1}^{4} c_h$, and their earnings up to that point, $\pi_i^1$. This information simplifies the calculation of individual contributions using earnings feedback and makes incentives salient in a similar way across treatments.

[7] Nikiforakis and Normann (2008) show that there is a monotonic relation between the reduction a punishment point causes and contributions to the public account (and earnings). A reduction E$2 per punishment point was sufficient to stabilize contributions at an intermediate level. To permit increases in contributions as well as decreases relative to the basic design a reduction factor of E$2 was used.

earnings from the first stage, that is, $\sum_{j \neq i} p_{ij} \leq \pi_i^1$, while punishment decisions are made simultaneously and without communication. To make punishments possible, at the beginning of each period subjects are randomly assigned an identification number between 1 and 4 to distinguish their actions from those of the others. To prevent the formation of individual reputation that could lead to counter-punishments and blur incentives, these numbers change across periods.

At the end of each period, participants are informed about the punishment points they received in total, the associated income reduction and their earnings as in (2). Payoff functions (1) and (2), the duration of the experiment (10 periods), and the instructions are common knowledge amongst participants in all treatments. Note also that, as is common in public good experiments with punishment opportunities, each subject is given a one-off lump-sum payment of E$25 to compensate for any losses he or she might incur in the duration of the experiment.

The experimental sessions lasted approximately fifty minutes and the average payment was A$17.80 or roughly $16.40. The exchange rate between experimental and Australian Dollar was E$1 = A$0.085. The experiments were conducted using zTree (Fischbacher, 2007).

# 3 Results

## 3.1 Punishment behavior

Punishment behavior should be a good indicator of how feedback format affects behavior. As mentioned earlier, a common finding in public good experiments is that individuals tend to punish those who contribute less than their peers. Therefore, if feedback format serves as a coordination device helping group members establish a contribution standard, punishment behavior should not differ in C and E. However, punishment behavior might be different in CE as subjects might find it harder to coordinate their contributions when receiving feedback in two formats that highlight both the private and the social benefit which are at odds with each other. The following results summarize the effect of feedback format on punishment behavior.

**Result 1:** *The likelihood that punishment will occur is unaffected by the feedback format.*

**Result 2:** *Punishment severity is the same in C and E, but is significantly higher in CE.*

Figures 1, 2, and Table 3 provide support for Results 1 and 2. Figure 1 depicts the likelihood of punishment as a function of how much an individual's contribution deviated from the average contribution of his peers. In all treatments, the likelihood seems to increase with the negative deviation from the contribution of one's peers. There are no apparent differences in the punishment likelihood across treatments. This suggests that the incentives in the game are understood equally well by subjects in all treatments.

Insert Figures 1 and 2 here

Figure 2 shows how deviating from the contributions of one's peers affects the severity of punishment conditional on punishment occurring. While there appear to be no differences between treatments C and E, punishments in CE are substantially harsher (especially in the first two brackets on the left). The harsher punishments in CE might be due to subjects having difficulties in creating common standards for contributions to the public account. This interpretation is supported by the considerably smaller number of observations in the interval [-2, 2] in CE compared to C and E in Figure 1. Further support is given in Figure 3 which depicts the evolution of punishment severity over the course of the experiment (the numbers in Figure 3 indicate the likelihood of being punished in each period). The severity of punishment is similar across treatments in the first period. However, punishment severity appears to progressively increase in CE, while remaining constant in C and E. In the final period of the experiment the punishments are more than twice as harsh in CE than in C or E. The temporal pattern of punishment severity in CE suggests an unresolved tension. The source of this tension will become clear in section 3.3.[8]

Insert Figure 3 here

In support of these observations Table 3 presents the results from a regression analysis. The fact that punishments do not take place in the majority of cases (see Figure 1) requires that one models the decision to punish (*punishment decision*) separately from the decision of how much to punish (*punishment level*). The appropriate specification to capture the two-stage process is a hurdle model. The likelihood function of the hurdle model is given by the product of two separate likelihoods. First, the likelihood

[8]It is worth emphasizing that the likelihood of punishment is the same across treatments despite the apparent ineffectiveness of punishment in raising contributions in some treatments. This is evidence for the fact that punishments are not driven only by strategic considerations, but also by emotions (e.g. Fehr and Gächter, 2000; 2002; de Quervain, Fischbacher, Treyer, Schellhammer, Schnyder, Buck and Fehr, 2004).

that an individual will be punished, which is captured by a standard Probit model, and second, the conditional likelihood that the individual will receive a certain number of punishment points, which is captured using a truncated linear regression. The two parts of the model are estimated separately (McDowell, 2003). The repeated nature of the experiment requires that one also controls for random effects at the group level.

The independent variables included in the regression are: $CE$, a dummy variable taking the value of one if the observation comes from CE and zero otherwise; $E$, a dummy variable for treatment E; (*Absolute*) *Negative Deviation* defined as $\max\{0, (\sum_{h\neq j} c_{h,t})/3 - c_{j,t}\}$, where $c_{j,t}$ is the contribution of individual $j$ in period $t$; *Positive Deviation,* defined as $\max\{0, c_{j,t} - (\sum_{h\neq j} c_{h,t})/3\}$; *Others' Average Contribution,* that is, $(\sum_{h\neq j} c_{h,t})/3$; and *Period*, a variable to control for time effects. For simplicity in exposition, individuals contributing more (less) than their peers on average will be sometimes referred to as *high (low) contributors.*

Insert Table 3 here

The results in Table 3 support the evidence from Figures 1, 2, and 3. Starting with the *punishment decision,* the coefficients of CE and E are insignificantly different from C (the omitted category) at all conventional levels. This implies that the likelihood of being punished does not differ significantly across treatments. Similarly to what has been found in previous studies, we see that the higher negative (positive) deviation is, the higher (lower) the likelihood of punishment is. Punishment severity, on the other hand, is significantly higher in CE than in C or E. The greater the extent of free riding (as captured by *Absolute Negative Deviation*) the harsher the punishment is likely to be. Also, similarly to the results of previous studies that employ a fixed matching protocol, we observe that the average contribution of the other group members affects neither the punishment decision nor the punishment level.[9]

## 3.2 Contributions to the public account

The fact that the punishment threat is not weakened by the provision of earnings feedback creates ideal conditions for testing the effect feedback format has on contribution levels. If contributions are lower in CE and E than they are in C we will have strong evidence supporting the hypothesis that feedback format affects contributions by emphasizing the private benefit of free riding. If this hypothesis is correct we should also observe higher contributions in CE than in E due to the lack of emphasis on the social

[9] This is the case even if one considers the punishment behavior in the early periods.

benefit of contributions in E. The following result summarizes the impact of feedback format on contributions.

**Result 3:** *Earnings feedback has a negative effect on contribution levels, while contribution feedback has a positive effect. The highest contribution levels are observed in C followed by CE and then E.*

Support for Result 3 comes from Figure 4 and Table 4. Figure 4 presents the evolution of average contribution over time in each treatment. Average contribution starts at similar levels across treatments. This is to be expected as subjects receive feedback for the first time after the contribution stage of period 1. Any differences, therefore, should be observed from period 2 onwards. Indeed, average contribution remains stable in CE over the duration of the experiment, but increases slightly in C, and declines in E.

Table 4 provides statistical support for these observations using a linear regression with group random effects. The first model compares average contributions across C, CE, and E. The differences are significant across all treatments (the difference between CE and C is significant at the 10% level). The second regression models separately the different time trends by introducing interaction terms between Period, CE and E. The insignificance of the coefficients of CE and E suggests that there are no differences across treatments at the beginning of the experiment. Somewhat surprisingly, given the upward trend observed in Figure 4, the coefficient of Period is not significant. The reason for this becomes clear in the third regression that excludes observations from the final period. Regression (3) indicates that there is a significant increase in C that was masked in the second regression by a strong end-game effect. The differences between CE and C, and E and C increase in every period as indicated by the negative coefficients of CE*Period and E*Period.

Insert Figure 4 and Table 4 here

Result 3 is consistent with the findings of Casari and Luini (2005) who use a public good game with punishment opportunities in which subjects receive contribution and earnings feedback. They find low and stable levels of contribution to the public account even when peer punishment is permitted. Casari and Luini (2005), however, neither examine the effect of feedback format nor do they provide an explanation for the low contribution levels observed in their experiment.[10]

[10] A recent study by Hermann, Thoeni and Gächter (2008) suggests that another reason for the unusually low contribution levels in Casari and Luini (2005) (lower than in CE) might be the Italian subject pool used.

## 3.3 Convergence of contributions

If feedback format acts as a coordination device and helps groups establish contribution standards, then providing individuals with information in two different formats highlighting the conflict between private and collective interest should slow convergence towards a given contribution level. In other words, the standard deviation of contributions should decline more slowly in CE, than in C or E. Indeed, a Wilcoxon test shows that standard deviation declines significantly in the second half of the experiment in C (by 24%, $p$-value$< .1$) and in E (by 45%, $p$-value$< .01$), but the 16% decline in CE (see Table A1 in the appendix) is not significant ($p$-value$> .3$).[11]

**Result 4:** *Standard deviation of contributions decreases significantly over time in C and E where feedback is given in only one format, but not in CE where feedback is given in two different formats.*

How exactly does feedback format prevent convergence in CE? The fact that contribution levels differ significantly despite similar punishment patterns in all treatments begs the question: How do individuals respond to punishments? This question is important as Figures 1 and 2 suggest that, in all treatments, individuals can reduce the expected loss of income due to punishments by contributing the same amount on average as their peers (or slightly more).[12] Therefore, the way individuals react to punishment should reflect their expectation about what their peers will contribute in the following period.[13]

**Result 5:** *Earnings feedback lowers the likelihood that a low contributor will respond to a punishment by raising his contribution in the following period.*

In treatment C, 74% of the punishment victims react by increasing their contribution in the following period. This is comparable to the reaction in previous experiments. In sharp contrast, only 56% and 51% of the punished subjects raise their contribution in CE

[11] The standard deviation is higher in the second half of the experiment in 4 out of 10 groups in CE, 2 out of 10 in C, and only 1 out of 11 in E.

[12] Masclet, Noussair, Tucker, and Villeval (2003) find that subjects increase their contribution to the public account following a punishment even if the punishment does not reduce their income. This suggests that punishment has an additional non-monetary cost for the recipient that should be taken into consideration when one considers the cost and benefit of choosing a particular contribution level.

[13] One could pay subjects to state their beliefs about what their peers will contribute on average in the following period. However, this procedure is known to affect behavior in public good experiments (Gächter and Renner, 2006). Given that the main interest of this study is to examine whether the feedback format is partly responsible for the high contribution levels seen in previous studies I decided against eliciting subjects' beliefs.

and E, respectively. This suggests that, while most subjects in C expect the contribution of their peers to be higher in the following period than their contribution in the current period, a large number of subjects punished in CE and E expect the average not to be higher than (or close enough to) their current contribution. In other words, earnings feedback seems to induce a considerable fraction of punished subjects to expect a large drop in contributions.

Insert Table 5a and 5b here

The different reactions to punishment might be partly affected by the different trends in group contributions. That is, an individual might not increase her contribution in E as she observes the contribution of her peers to be also decreasing from one period to the next or might choose to increase it in C as group contribution has been increasing. To provide formal statistical support for Result 5, Table 5a presents the results from a Probit regression analysis of how low contributors adjust their contributions between periods. The dependent variable is the likelihood that an individual will increase his contribution in period $t+1$ compared to his contribution in period $t$. In addition to variables explained previously, the independent variables include *Other's Average Contribution Change*$= \sum_{h \neq i} c_{h,t} - \sum_{h \neq i} c_{h,t-1}$; *Punished,* a dummy variable taking the value 1 if a subject was punished and zero otherwise; and the interaction of these variables with treatment dummies. The significance of variable (*Absolute*) *Negative Deviation* $*$ *Punished* indicates that low contributors punished in treatment C (the omitted category) are more likely to increase their contributions in the following period than low contributors not punished. This tendency is significantly weaker in treatments CE and E as indicated by the (significant) negative coefficients of *CE* $*$ (*Absolute*) *Negative Deviation* $*$ *Punished* and *E* $*$ (*Absolute*) *Negative Deviation* $*$ *Punished*, respectively. Increases in the contributions of the other group members raises the likelihood of an individual increasing his contribution by a similar amount in all treatments. To understand why individuals fail to converge in treatment CE one needs to also consider the behavior of high contributors.

**Result 6:** *High contributors in CE are as likely to lower their contribution in the following period as high contributors in C. High contributors in E are more likely to lower their contribution in the following period than high contributors in C.*

Support for Result 6 is presented in Table 5b that shows the adjustment in the contribution of high contributors. The negative and significant coefficient of *Positive*

*Deviation* indicates that the higher an individual's contribution is compared to that of her peers the less likely she is to raise her contribution in the following period (or, equivalently, the more likely she is to lower or leave her contribution unchanged) in treatment C. The most interesting finding in the second regression is the insignificance of the coefficient of $CE * Positive\ Deviation$ and, at the same time, the significance of $E * Positive\ Deviation$. The latter implies that high contributors in treatment E are less likely to increase their contribution in the following period compared to high contributors in treatment C (similar to the case of low contributors). However, the insignificance of the coefficient of $CE * Positive\ Deviation$ implies that high contributors in CE react similarly to high contributors in C. This indicates that high contributors in CE have somewhat different expectations to low contributors in the same treatment: While high contributors adjust their contributions in a similar way to high contributors in C, low contributors are found to be less likely to increase their contribution than low contributors in C.

Results 5 and 6 taken together explain why contributions fail to converge in treatment CE: High contributors seem to expect low contributors to adjust their contribution upwards, while low contributors seem to expect high contributors to adjust their contribution downwards. It, therefore, appears as if high contributors focus their attention on the contribution feedback and low contributors on the earnings feedback.

**Conjecture:** *In treatment CE subjects interpret feedback in a self-serving manner. This tendency hampers convergence of contributions.*

The different reactions of high and low contributors in treatment CE are most likely the reason punishment severity increases dramatically over time in CE (Figure 3). The ineffectiveness of punishment in convincing low contributors to increase their contribution levels presumably leads some high contributors to use the severity of punishment as a signal of their intentions. The continuous increase in punishment severity in CE also suggests that the lack of responsiveness to punishment causes frustration (or anger) to punishers. The latter is supported by the fact that punishments in CE are substantially more severe even in the final period.

## 3.4 Earnings across treatments

The use of punishment is costly for groups: First, subjects must pay to punish group members, and, second, punishment victims lose part of their income. For punishment to be beneficial for the groups, the benefits that accrue from its threat and use must

outweigh the costs. Most previous studies have found that peer punishment does not lead to increases in earnings and often leads to decreases (for a brief review see Nikiforakis, 2008).

If all individuals contribute their endowment to the public account and abstain from punishing average earnings are maximized and equal E$32 (see equation 1). On the other hand, if individuals behave as prescribed by the subgame-perfect Nash equilibrium, then average earnings equal E$20. Feedback format has a strong impact on earnings.

**Result 7:** *Earnings are significantly lower in treatments CE and E in which subjects receive earnings feedback compared to C in which subjects receive only contribution feedback.*

Figure 5 provides the relevant support for Result 7. Earnings feedback has clearly a negative effect on earnings. This is due to investments in peer punishment that are at least as costly and lower contribution levels. Most notably, the severe punishments in treatment CE lead to earnings that are even lower than those predicted by the inefficient equilibrium. The average earnings of E$24.20 in C are significantly higher than the average earnings in E (Mann-Whitney, $p$-value$=$ .016) and CE (Mann-Whitney, $p$-value$<$ .01).

Insert Figure 5 here

At this point, one could argue that groups receiving earnings feedback are not worse off than they would be if peer punishment was totally absent. However, this argument would overlook the fact that individuals voluntarily contribute significant fractions of their endowments even when not threatened by punishment. Nikiforakis and Normann (2008) compare earnings in a treatment similar to C to those in a treatment without peer punishment. Average earnings are E$23.90 in the treatment with punishment (in their paper the treatment is referred to as "2") and E$22.81 in the treatment without peer punishment. This difference was not found to be significant.

# 4 Alternative Explanations

The experimental results support the hypothesis that feedback format serves as a coordination device that influences the level of contributions to the public account by focusing subjects' attention on either collective or private benefit. Nevertheless, it is worth considering alternative explanations for the experimental findings.

The first explanation one must consider is that participants do not fully understand the experiment. As a result, earnings feedback help subjects discover their dominant strategy. This does not seem to be the case. Firstly, before the experiment can begin participants must answer ten questions that help them understand the incentives in the game and how contributions translate to earnings and vice versa. Secondly, the post-experimental questionnaire shows that participants understand the game by frequently referring to the need for cooperation and the incentives to free ride. Thirdly, the fact that participants understand the incentives they face is also evinced by their punishment behavior which is similar across treatments and follows patterns seen in previous experiments.[14] This evidence suggests that the differences are not driven by confused subjects.

A second related explanation is that information about earnings allows individuals to imitate the choice of the most successful of their peers. Vega-Redondo (1997) shows that if individuals are boundedly rational then competition in an oligopolistic market will be more fierce if information about earnings is provided. Huck, Normann and Oechssler (1999, 2000) and Offerman, Potters and Sonnemans (2002) provide evidence in support of Vega-Redondo's model from experimental Cournot oligopolies.[15] The findings in this experiment, however, challenge the notion that the impact of earnings feedback is due to boundedly rational individuals. If individuals are boundedly rational (in the sense that they cannot calculate their optimal decision), then they should be unable to imitate the choices of their most successful peers unless they can observe both the choice and the respective earnings. This justification, therefore, fails to explain the difference in contributions in treatments C and E. Furthermore, imitation should be easier in treatment CE where information is readily available. This would imply that standard deviation should decrease in this treatment as all group members imitate the one with the highest earnings. However, as we saw standard deviation does not decrease significantly over time in CE, in contrast to C and E. Imitation cannot account for the harsher punishments that were observed in CE. Therefore, the results from the present experiment add to the existing evidence questioning the empirical significance of imitation in laboratory experiments (see Boesch-Domenech and Vriend, 2003; Apesteguia, Huck, Oechssler and Weidenholzer, 2007).

A third explanation is that feedback format alters the interpretation of punishments. In treatment C it should be clear that punishments are meted out to the least cooperative

[14] The reader is reminded that punishment severity is found to be very similar in the first period of the experiment.

[15] See also Abbink and Brandts (2007), Apesteguia, Huck and Oechssler (2007) and Selten and Apesteguia (2005).

subjects. This is supported by the fact that most subjects react to punishment by raising their contribution in the following period. In treatments E and CE, however, the person punished is the highest earner - the person who "did best". It is, therefore, possible that punishment is interpreted as a sign of jealousy (or envy) in E and CE, rather than as an "invitation" to cooperate as it is in C. Jealousy would prescribe that a punishment by individual $i$ is followed by a reduction in $i$'s contribution in the following period. As we saw, however, this is not what high contributors (typically the ones meting out punishment points) do in CE. Subjects must have been aware of this from early on. Consider the contributions in period 2 of the subjects that punished other group members in period 1. In C, 68% of the punishers in period 1 (13 out of 19 punishers) did not reduce their contribution to the public account in period 2. Similarly, 64% of the punishers in CE (9 out of 14 punishers) did not reduce their contribution in period 2. The behavior of punishers is similar in treatments C and CE and, therefore, it seems unlikely that the interpretation of punishment changes. (For completeness, only 38% of the punishers - 6 out of 16 punishers - in E did not reduce their contribution from period 1 to period 2.) Therefore, the differences across treatments cannot be attributed to feedback format altering the interpretation of punishments.

A fourth explanation is that there is an experimenter's effect in public good games with peer punishment. The feedback format changes subjects' perception about what the experimenter's "preferred" behavior is. In C subjects might be aware that punishment serves to increase contributions. In contrast, in E and CE, the presence of information about earnings might make subjects perceive the game differently (e.g. as a tournament). It seems difficult, however, to reconcile this argument with the fact that punishment behavior follows similar patterns across treaments.

# 5 Discussion

This paper presented evidence from a public good experiment showing that cooperation sustained under the threat of peer punishment can be sensitive to changes in institutional details. In particular, changes in the format used to provide subjects with feedback about the actions of their peers lead to significantly different levels of cooperation and earnings even though they do not affect incentives. Feedback about the earnings of the other group members impacts negatively both cooperation and earnings.

The data suggest that feedback format affects the coordination process. Unlike in the treatments where a single feedback format is used, in the treatment where feedback is given in two formats that emphasize the conflicting social and private benefits, individual

contributions to the public account do not converge. Contribution feedback emphasizes the social benefit of contributing to the public account, while earnings feedback highlights the private benefit from contributing to the private account. This explanation differs from those previously given for similar framing effects (e.g. Huck, Normann and Oechssler, 1999, 2000; Offerman, Potters and Sonnemans, 2002). The explanation offered here is most closely related to that in Dufwenberg, Gächter, and Hennig-Schmidt (2006). These authors formally show how framing effects can be explained within a rational choice framework. In particular, Dufwenberg, Gächter, and Hennig-Schmidt show that frames can affect beliefs about the actions of others, which in turn affect the actions subjects take.

The results cast doubt on the common belief that peer punishment can always solve the free rider problem (e.g. Ostrom, Walker and Gardner, 1992; Fehr and Gächter, 2002). Different institutional details such as the feedback format might influence subjects' expectations in a way that either enhances or undermines the efficacy of peer punishment. Institutional effects are likely to be more pronounced in larger groups where communication is difficult.

Institutional details might also affect the likelihood of conflict escalation. The self-serving interpretation of information in the treatment in which subjects receive feedback in two formats prevented contributions from converging and consequently led to harsher punishments. In a more general setting where reprisals are permitted, harsher punishments are more likely to trigger counter-punishments (Denant-Boemont, 2007; Nikiforakis, 2008). This, in association with the self-serving interpretation of feedback, might lead to conflict escalation and lower efficiency if multiple rounds of punishment and counter-punishment are allowed (Nikiforakis and Engelmann, 2008).

The negative impact of earnings feedback on cooperation suggests that a reconsideration of some previous experimental findings might be desirable. For example, if coordination in the public good game can be improved by omitting information about earnings, one might wonder whether similar (superficial) changes in institutional details can prevent the well-documented coordination failure in games with multiple Pareto-ranked equilibria (e.g. Van Huyck, Battalio and Beil, 1990;1991). For example, Charness, Fréchette, and Kagel (2004), show that the use of payoff tables leads to less gift-exchange compared to a treatment in which subjects are presented only with payoff functions. While the gift-exchange game is not a coordination game, payoff tables undoubtedly make incentives in the game salient. As subjects in pure-coordination experiments are confronted with payoff tables, coordination failure might be less of a problem when payoff functions instead of tables are used.

The evidence from the present experiment also emphasizes the need for economists to go beyond the standard assumption that agents have self-regarding preferences. Only then will we be in the position to identify situations in which costless changes in institutional details can improve outcomes. Institutional details might help facilitate cooperation in the presence of peer punishment opportunities and free riding incentives - as in the treatment with contribution feedback - or lead to significant efficiency losses - as in the treatment with contribution and earnings feedback.

# References


[1] Abbink, K., Brandts, J. 2007. 24 - Pricing in Bertrand Competition with Increasing Marginal Costs, forthcoming in Games and Economic Behavior.

[2] Apesteguia, J., Huck, S., Oechssler, J., 2007. Imitation - Theory and Experimental Evidence, Journal of Economic Theory 136, 217 – 235.

[3] Apesteguia, J., Huck, S., Oechssler, J., Weidenholzer, S., 2007. Imitation and the Evolution of Walrasian Behavior: Theoretically Fragile but Behaviorally Robust, working paper.

[4] Anderson, C., Putterman, L., 2006. Do Non-Strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism, Games and Economic Behavior 54 (1), 1-24.

[5] Bochet, O., Page, T., Putterman, L., 2006. Communication and Punishment in Voluntary Contribution Experiments, Journal of Economic Behavior and Organization 60 (1), 11-26.

[6] Boesch-Domenech, A., Vriend, N., 2003. Imitation of Successful Behaviour in Cournot Markets, The Economic Journal 113, 495–524.

[7] Carpenter, J., 2007a. Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods, Games and Economic Behavior, 60(1), 31-52.

[8] Carpenter, J., 2007b. The Demand for Punishment, Journal of Economic Behavior & Organization 62 (4), 522-542.

[9] Casari, M., Luini., L., 2005. Group Cooperation Under Alternative Peer Punishment Technologies: An Experiment, mimeo.

[10] Charness, G., Fréchette, G.R., Kagel, J.H., 2004. How Robust is Laboratory Gift-Exchange?, Experimental Economics 7, 189-205.

[11] de Quervain, D. J. F., Fischbacher, U., Treyer V., Schellhammer, M., Schnyder, U., Buck, A., Fehr, E., 2004. The Neural Basis of Altruistic Punishment, Science 305 (5688), 1254-1258.

[12] Denant-Boemont, L., Masclet, D., Noussair, C., 2007. Punishment, Counterpunishment and Sanction Enforcement in a Social Dilemma Experiment,Economic Theory 33(1), 145-167, 2007.

[13] Dufwenberg, M., Gächter, S., Hennig-Schmidt, H., 2006. The Framing of Games and the Psychology of Strategic Choice. CeDEx Discussion Paper No. 2006-20.

[14] Fehr, E., Gächter, S., 2000. Cooperation and Punishment in Public Goods Experiments, American Economic Review 90, 980-994.

[15] Fehr, E., Gächter, S., 2002. Altruistic Punishment in Humans, Nature 415, 137-140.

[16] Fehr, E., Schmidt, K., 1999. A Theory of Fairness, Competition and Co-operation, Quarterly Journal of Economics 114, 817-868.

[17] Fehr, E., Schmidt, K., 2003. Theories of Fairness and Reciprocity: Evidence and Economic Applications, In: Dewatripont, M. et al. (eds), Advances in Economic Theory, Eighth World Congress of the Econometric Society, Vol. I, 208-257, Cambridge: Cambridge University Press.

[18] Fischbacher, U., 2007.z-Tree: Zurich Toolbox for Ready-made Economic Experiments, Experimental Economics 10 (2), 171-178.

[19] Gächter, S., Renner, E., 2006. The effects of (incentivized) belief elicitation in public good experiments, CeDEx Discussion Paper No. 2006-16.

[20] Greiner, B., 2004. The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. University of Cologne, Working Paper Series in Economics 10.

[21] Hermann, B., Thoeni,C., Gächter,S., 2008. Antisocial punishment across societies. Science 319, 1362-1367.

[22] Huck S., Normann, H.T., Oechssler, J., 1999. Learning in Cournot oligopoly: An experiment, Economic Journal 109, C80-C95.

[23] Huck S., Normann, H.T., Oechssler, J., 2000. Does information about competitors' actions increase or decrease competition in experimental oligopoly markets?, International Journal of Industrial Organization 18, 39-57.

[24] Masclet, D., Noussair, C., Tucker, S., Villeval M.C., 2003. Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism, American Economic Review 93, 366-380.

[25] McDowell, A., 2003. From the Help Desk: Hurdle Models, The Stata Journal 3 (2), 178-184.

[26] Nikiforakis, N., 2008. Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?, Journal of Public Economics 92, 91–112.

[27] Nikiforakis, N., Engelmann, D., 2008. Feuds in the Laboratory?, University of Melbourne Working Paper Series.

[28] Nikiforakis, N., Normann, H.T., 2008. A Comparative Statics Analysis of Punishment in Public-Good Experiments, forthcoming in Experimental Economics.

[29] Noussair, C., Tucker, S., 2005. Combining Monetary and Social Sanctions to Promote Cooperation, Economic Inquiry 43(3), 649-660.

[30] Offerman, T., Potters, J., Sonnemans, J., 2002, Imitation and Belief Learning in an Oligopoly Experiment, The Review of Economic Studies 69 (4), 973-997.

[31] Ostrom, E., Walker, J., Gardner, R., 1992. Covenants With and Without a Sword: Self Governance is Possible, American Political Science Review 86, 404-417.

[32] Page, T., Putterman, L., Unel, B., 2005. Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency, The Economic Journal 115 (506), 1032-1053.

[33] Sefton, M., Shupp, R., Walker, J., 2007. The Effect of Rewards and Sanctions in Provision of Public Goods, Economic Inquiry 45 (4), 671-690.

[34] Selten, R., Apesteguia, J., 2005. Experimentally observed imitation and cooperation in price competition on the circle, Games and Economic Behavior 51, 171–192.

[35] Van Huyck, J.B., Battalio, R.C., Beil, R.O., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. The American Economic Review 80, 234-248.

[36] Van Huyck, J.B., Battalio, R.C., Beil, R.O., 1991. Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games. The Quarterly Journal of Economics 106, 885-910.

[37] Vega-Redondo, F., 1997. The Evolution of Walrasian Behavior, Econometrica 65, 375-384.

Table 1 - An example of different feedback formats

Table 1a

| Player | Contribution in period $t$ |
|---|---|
| 1 | 20 |
| 2 | 15 |
| 3 | 10 |
| 4 | 5 |

Table 1b

| Player | Earnings in period $t$ |
|---|---|
| 1 | 20 |
| 2 | 25 |
| 3 | 30 |
| 4 | 35 |

Table 2 - Experimental treatments

| Treatment | Information | Number of participants |
|---|---|---|
| C | Contribution feedback | 40 |
| CE | Earnings feedback | 44 |
| E | Contribution and earnings feedback | 40 |

Table 3 - Determinants of punishment

| | Punishment Decision | Punishment Level |
|---|---|---|
| CE | 0.11 | 1.35*** |
| | (0.19) | (0.36) |
| E | 0.22 | 0.49 |
| | (0.19) | (0.37) |
| (Absolute) Negative Deviation | 0.16*** | 0.25*** |
| | (0.01) | (0.03) |
| Positive Deviation | -0.04*** | 0.04 |
| | (0.01) | (0.05) |
| Others' Average Contribution | -0.01 | 0.00 |
| | (0.01) | (0.03) |
| Period | -0.03*** | 0.13*** |
| | (0.01) | (0.04) |
| Constant | -1.31*** | -0.28 |
| | (0.19) | (0.47) |
| Observations | 3720 | 619 |
| Wald chi2 | 477.55*** | 125.20*** |
| Log likelihood | -1339.66 | |

Punishment decision is a probit with group random effects; Punishment severity is a truncated linear regression with group random effects; Standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4 - Contributions to the public account

| | (1) | (2) | (3) |
|---|---|---|---|
| CE | -2.82* | -1.74 | -1.30 |
| | (1.62) | (1.74) | (1.72) |
| E | -5.45*** | -2.47 | -1.86 |
| | (1.58) | (1.70) | (1.68) |
| CE*Period | | -0.20* | -0.31** |
| | | (0.11) | (0.13) |
| E*Period | | -0.54*** | -0.71*** |
| | | (0.11) | (0.13) |
| Period | | 0.09 | 0.25*** |
| | | (0.08) | (0.09) |
| Constant | 10.61*** | 10.09*** | 9.54*** |
| | (1.15) | (1.23) | (1.22) |
| Observations | 1240 | 1240 | 1116 |
| Wald chi2 | 11.86*** | 49.25*** | 46.32*** |

Linear regression with group random effects; * significant at 10%; ** significant at 5%; *** significant at 1%; (3) excludes observations from final period

Table 5a - Likelihood of increasing contribution in period $t$

| Low Contributors (i.e. Negative Deviation>0) | |
|---|---|
| (Absolute) Negative Deviation | -0.06 |
| | (0.06) |
| (Absolute) Negative Deviation * Punished | 0.24*** |
| | (0.07) |
| Others' Average Contribution Change | 0.12** |
| | (0.05) |
| CE * (Absolute) Negative Deviation | 0.19** |
| | (0.08) |
| CE * (Absolute) Negative Deviation * Punished | -0.28*** |
| | (0.09) |
| CE * Others' Average Contribution Change | -0.10 |
| | (0.07) |
| E * (Absolute) Negative Deviation | 0.05 |
| | (0.08) |
| E * (Absolute) Negative Deviation * Punished | -0.17* |
| | (0.09) |
| E * Others' Average Contribution Change | -0.07 |
| | (0.07) |
| Constant | -0.11 |
| | (0.11) |
| Observations | 471 |
| Wald chi2 | 42.77 |

Probit with group random effects; Standard errors are in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5b - Likelihood of increasing contribution in period $t$

| High Contributors (i.e. Positive Deviation>0) | |
|---|---|
| Positive Deviation | -0.07* |
| | (0.04) |
| Positive Deviation * Punished | -0.06 |
| | (0.13) |
| Others' Average Contribution Change | 0.04 |
| | (0.05) |
| CE * Positive Deviation | -0.05 |
| | (0.05) |
| CE * Positive Deviation * Punished | -0.08 |
| | (0.19) |
| CE * Others' Positive Contribution Change | -0.01 |
| | (0.08) |
| E * Positive Deviation | -0.16** |
| | (0.07) |
| E * Positive Deviation * Punished | 0.18 |
| | (0.16) |
| E * Others' Average Contribution Change | -0.11 |
| | (0.08) |
| Constant | -0.62*** |
| | (0.12) |
| Observations | 459 |
| Wald chi2 | 18.06 |

Probit with group random effects; Standard errors are in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

Table A1 - Group data

| | | All periods | | First period | | Periods 6-10 | |
|---|---|---|---|---|---|---|---|
| Treatment | Group | Contribution | Standard deviation | Contribution | Standard deviation | Contribution | Standard deviation |
| C | 1 | 10.20 | 2.65 | 6.75 | 4.57 | 12.80 | 1.80 |
| C | 2 | 17.33 | 2.04 | 12.25 | 8.34 | 19.60 | 0.48 |
| C | 3 | 8.70 | 4.23 | 7.75 | 3.20 | 8.50 | 4.48 |
| C | 4 | 7.68 | 7.05 | 8.75 | 8.38 | 6.60 | 7.05 |
| C | 5 | 6.58 | 4.53 | 8.75 | 5.38 | 5.25 | 3.24 |
| C | 6 | 3.80 | 2.85 | 6.25 | 2.50 | 2.30 | 2.42 |
| C | 7 | 14.83 | 3.10 | 9.00 | 8.60 | 17.05 | 2.57 |
| C | 8 | 16.00 | 3.82 | 12.25 | 6.13 | 17.30 | 3.25 |
| C | 9 | 12.65 | 3.08 | 12.00 | 5.66 | 13.20 | 3.75 |
| C | 10 | 8.35 | 5.02 | 8.75 | 4.79 | 6.55 | 4.12 |
| **C** | **average** | **10.61** | **3.84** | **9.25** | **5.76** | **10.92** | **3.32** |
| CE | 1 | 6.63 | 4.35 | 6.50 | 4.43 | 7.50 | 4.86 |
| CE | 2 | 11.40 | 5.39 | 8.75 | 5.38 | 13.70 | 4.72 |
| CE | 3 | 8.35 | 4.38 | 8.75 | 2.50 | 6.20 | 5.08 |
| CE | 4 | 7.65 | 5.08 | 9.50 | 9.47 | 7.15 | 2.44 |
| CE | 5 | 2.50 | 3.78 | 6.25 | 8.10 | 1.55 | 2.88 |
| CE | 6 | 3.63 | 2.43 | 6.00 | 3.92 | 2.55 | 1.80 |
| CE | 7 | 13.33 | 9.20 | 10.00 | 8.79 | 13.45 | 9.21 |
| CE | 8 | 12.13 | 5.35 | 9.00 | 7.44 | 12.20 | 6.30 |
| CE | 9 | 7.90 | 3.71 | 9.50 | 7.14 | 7.60 | 2.76 |
| CE | 10 | 4.43 | 2.45 | 5.00 | 3.46 | 3.45 | 1.87 |
| **CE** | **average** | **7.79** | **4.61** | **7.93** | **6.06** | **7.54** | **4.19** |
| E | 1 | 3.15 | 3.28 | 11.25 | 6.29 | 0.40 | 0.80 |
| E | 2 | 6.43 | 4.79 | 8.50 | 8.70 | 5.45 | 4.36 |
| E | 3 | 5.25 | 3.41 | 9.25 | 7.89 | 4.05 | 2.25 |
| E | 4 | 7.15 | 3.90 | 10.25 | 8.66 | 5.60 | 2.07 |
| E | 5 | 3.08 | 2.40 | 8.50 | 7.68 | 1.70 | 1.23 |
| E | 6 | 1.65 | 1.55 | 5.75 | 2.99 | 0.10 | 0.20 |
| E | 7 | 3.35 | 2.93 | 5.00 | 4.40 | 3.30 | 2.17 |
| E | 8 | 4.05 | 2.90 | 6.75 | 3.95 | 3.55 | 2.55 |
| E | 9 | 3.78 | 2.97 | 10.25 | 6.13 | 2.30 | 2.45 |
| E | 10 | 9.95 | 1.23 | 6.25 | 1.50 | 11.60 | 1.25 |
| E | 11 | 8.95 | 6.08 | 12.00 | 7.83 | 8.15 | 5.89 |
| **E** | **average** | **5.36** | **3.22** | **8.25** | **5.97** | **4.20** | **2.29** |

Figure 1 – The likelihood of punishment as a function of deviation from the contribution of one's peers (numbers on top of bars indicate number of observations)

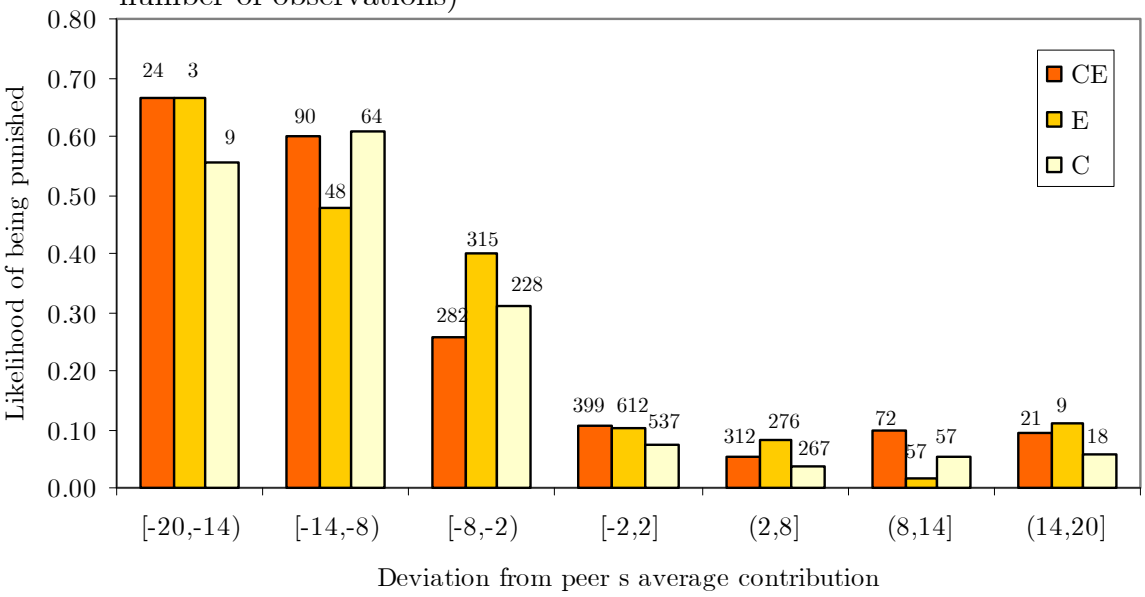


Figure 2 – The severity of punishment as a function of deviation from the contribution of one's peers (conditional on punishment occurring)

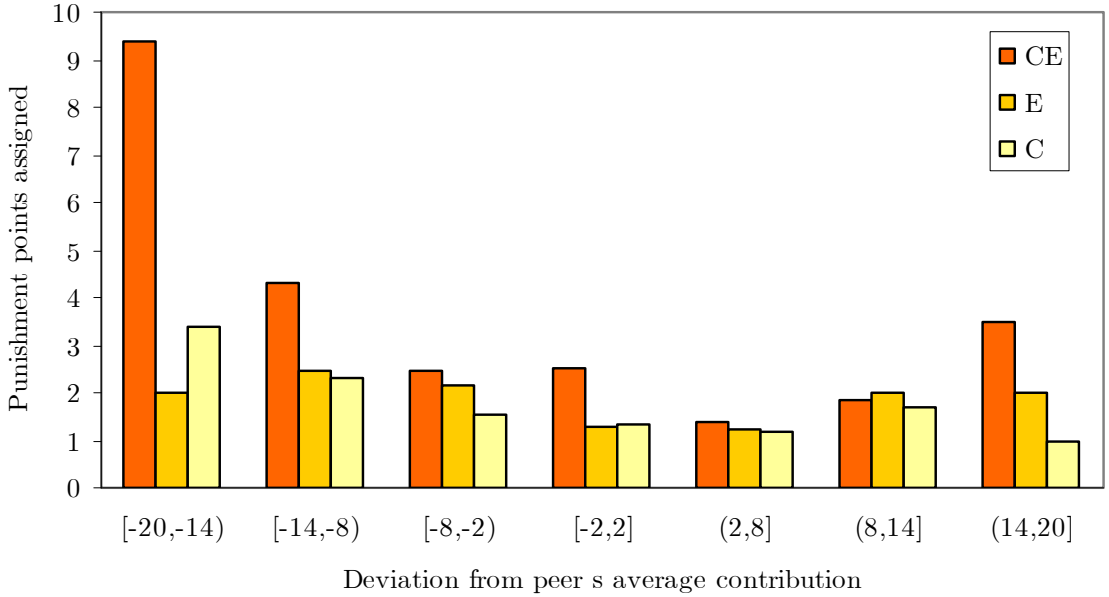

Figure 3 – The evolution of punishment severity (numbers indicate the likelihood of being punished in each period)

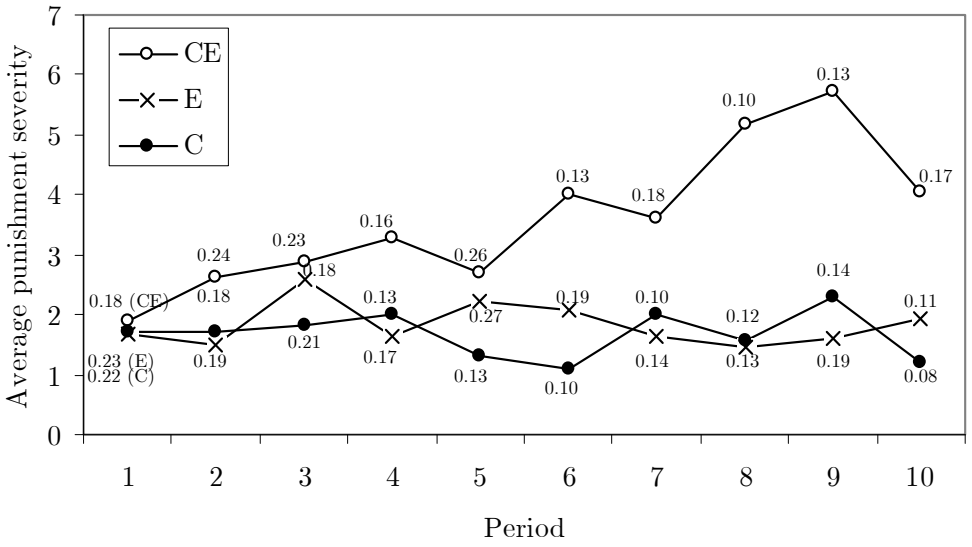


Figure 4 – The evolution of average contribution across treatments

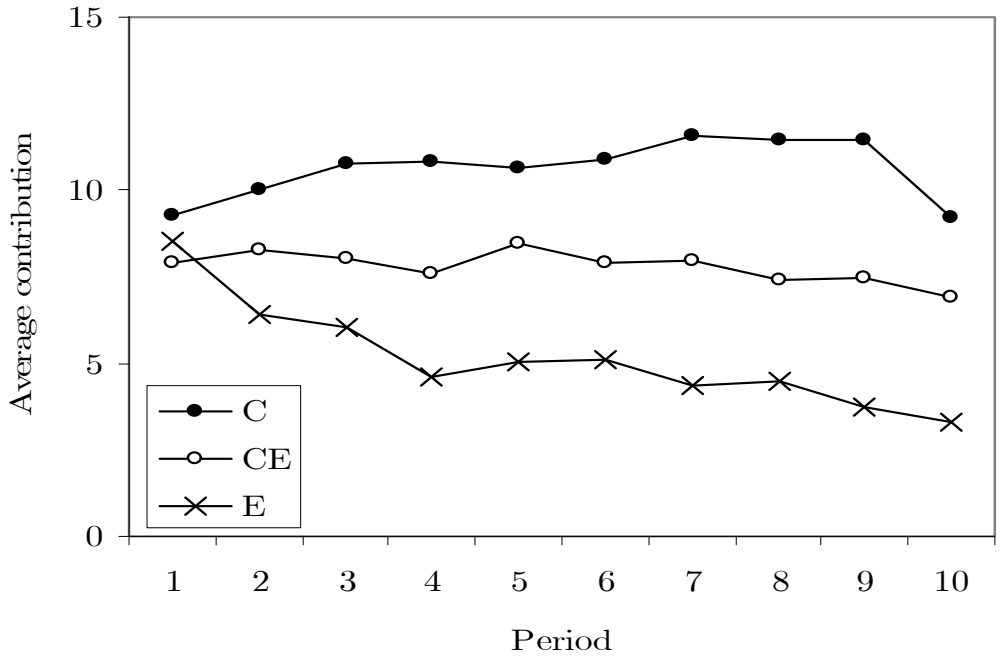

Figure 5 – Average Earnings Across Treatments