ISSN 0819-2642
ISBN 978 0 7340 4022 0

THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 1058

December 2008

Feuds in the Laboratory? A Social Dilemma Experiment

by

Nikos Nikiforakis & Dirk Engelmann

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

Feuds in the Laboratory? A Social Dilemma Experiment*

Nikos Nikiforakis† Dirk Engelmann‡

December 9, 2008

Abstract

Punishing free riders might promote cooperation, but it can also lead to feuds. We use a public good game with punishment opportunities to investigate whether the threat of feuds is efficiency enhancing. Treatments differ with respect to whether a punishment can trigger a feud. In the main treatment (*Feud*) the number of punishment stages is endogenously determined and avoiding revenge is impossible. Participants are generally found to employ strategies that avoid the break out of feuds. While the possibility of a feud affects punishment and contribution patterns, total earnings are not significantly different from that in treatments where punishment opportunities do not exist or where punishment opportunities exist, but there is no possibility of escalation.

JEL Classification: C92, D70, H41

Keywords: feuds, public goods, punishment, revenge, experimental economics

*Financial support from ESRC (project RES-000-22-0948) is gratefully acknowledged. Dirk Engelmann acknowledges financial support from the institutional research grant AV0Z70850503 of the Economics Institute of the Academy of Sciences of the Czech Republic, v.v.i. The experimental instructions and software are available at www.economics.unimelb.edu.au/nnikiforakis/research.htm

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

‡Department of Economics, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, United Kingdom, tel: +44 1784 44 3968, fax: +44 1784 439534, email: *dirk.engelmann@rhul.ac.uk.* Dirk Engelmann is also an affiliate of the Centre for Experimental Economics at the University of Copenhagen and a Senior Researcher at the Economics Institute of the Academy of Sciences of the Czech Republic, v.v.i.

"When there is no central authority to do the enforcement the players must rely on themselves to give each other the necessary incentives to elicit cooperation ... The problem with TIT FOR TAT is that once a feud gets started, it can continue indefinitely"

Axelrod (1984, p.138)

1 Introduction

One of the most enduring problems in economics is finding ways to foster cooperation when free riding incentives exist. A solution that has recently received considerable attention by economists, biologists, political scientists and anthropologists alike is altruistic punishment (see Gintis, Bowles, Boyd, and Fehr, 2005). Altruistic punishers are willing to discipline free riders at their own expense even when they do not anticipate future benefits (Fehr and Gächter, 2002). These punishers are 'altruistic' as their willingness to punish benefits the whole group by discouraging free riding.

This paper uses a social dilemma experiment to study behavior when altruistic punishment can lead to a feud, that is, a long standing argument between parties. Many scholars like Axelrod (1984) and Seabright (2004) have argued that while punishment can help enforce cooperation it can lead to a cycle of revenge. Therefore, one must weigh the benefits of peer punishment against the costs before trying to correct a (perceived) injustice.

There is an abundance of field evidence that avenging a perceived injustice can lead to feuds. One prominent example is workplace feuds which is a major theme in organizational theory (e.g. Bjorkqvist, Osterman, and Hjelt-Backet,1994; Rugala and Isaacs, 2004). Anderson and Pearson (1999) argue that "workplace incivility can spiral, beginning with one party's perception of an incivility and reciprocation with a counterincivility, which can potentially escalate to an exchange of coercive actions" (p.466). The spiral of retribution in the workplace often leads even to homicide (Bensimon, 1994; Dunkel, 1994).

The case of workplace feuds illustrates an important point: Feuds can erupt not only when central enforcement is impossible (e.g. because labor contracts are incomplete or a central authority does not exist), but also when a central authority exists (e.g. the boss) and central enforcement is possible. This can happen if the cost of calling on the central authority exceeds (at the outset at least) the benefit. Therefore, there is a wide array of situations in which altruistic punishment can trigger feuds. Understanding how the risk of starting a feud affects the prevalence and effectiveness of altruistic punishment is therefore an important question.

We use a public-good game with multiple punishment stages to study how the likelihood of triggering a feud affects the willingness to enforce cooperation, cooperation levels and efficiency. The game is as follows. In the first stage, each individual in a group decides how much of an endowment to contribute to a public account. The higher the contributions to the public account, the higher the group earnings. However, every individual has also an incentive to free-ride and not contribute. In the second stage, individuals are informed of each group member's contribution to the public account and can reduce each other's earnings by assigning costly punishment points. In the third stage, individuals can observe the punishment points assigned to each group member and can again assign punishment points to other individuals. Therefore, our experiment permits individuals to counter-punish or punish individuals who have not punished free riders. As long as individuals have funds to punish and punishment is meted out, additional stages are entered. Due to punishment choices being reduced to integers, there is a theoretical upper bound of 44 punishment stages.

One of the novel elements of our experiment is that the number of stages is endogenously determined. With the exception of Denant-Boemont, Masclet and Noussair (2007) and Nikiforakis (2008), all previous studies on altruistic punishment have used the first two stages of the game described above (e.g. 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). That is, altruistic punishers are protected from retaliation. The threat of peer punishment in these experiments has been shown to have a positive effect on cooperation levels which are higher when groups interact repeatedly with each other.

The efficacy of peer punishment is significantly limited when individuals are given an opportunity to retaliate punishment, as shown by Denant-Boemont et al. (2007) and Nikiforakis (2008) who allow for a second punishment stage. Some individuals are willing to respond to altruistic punishment by counter-punishing even if they free ride and, therefore, are justifiably punished, and even if they cannot anticipate monetary benefits from their action (Nikiforakis, 2008). Many individuals are also willing to use counter-punishment strategically to deter future sanctions. Consequently, both studies find that the threat of counter-punishment has a negative impact on the number of individuals who are willing to punish free riders, and, as a result, cooperation unravels and efficiency declines.[1]

Our experimental design allows us to address two questions that have not been investigated previously. The first question concerns how people behave when faced with the threat of a feud: Does the possibility of triggering a feud discourage altruistic punishers from trying to enforce cooperation? Does punishment lead to feuds? To answer these questions we compare punishment behavior in a treatment with two stages where retaliation is impossible, to behavior in treatments with multiple punishment stages. This question is interesting for two reasons. First, the benefits from punishing a free rider are higher when groups interact repeatedly as (former) free riders will cooperate with the others in the group, not only once, but many times in the future. That is, altruistic punishment is more common under fixed matching (Fehr and Gächter, 2000; Nikiforakis, 2008). However, the likelihood of punishment triggering a lengthy feud is also greater when groups interact repeatedly. It is plausible, therefore, that the risk of a feud will outweigh the individual benefits from altruistic punishment. Therefore, we might observe fewer attempts to enforce cooperation exactly in the situation in which the benefits from altruistic punishment are the highest.

Second, examining whether the risk of a feud encourages or discourages altruistic punishers will provide insights into the motives behind altruistic punishment. Elster (1990) argues that revenge is primarily fuelled by a need to protect one's self-esteem. Defying the risk of triggering a feud boosts an individual's self-esteem. Therefore, some individuals might be more willing to enforce cooperation when feuds are likely to ensue despite the fact that punishment is more costly in expected terms. This implies that the more likely

[1] Denant-Boemont et al. (2007) also study behavior in a treatment with 5 punishment stages. Apart from the fact that the number of punishment stages is known ex ante thus prohibiting the break out of feuds, the convex cost of punishment in Denant-Boemont et al. (2007) provides incentives for people to spread their punishment across stages. In other words, observing punishments over multiple stages is not necessarily due to counter-punishment and counter-counter-punishment, but could be due to savings in the cost of punishment.

a feud becomes the greater the punishment activity will be. The evidence from laboratory experiments so far does not lend support to this argument: While many individuals punish anti-social behavior, an increase in the cost of punishment causes a reduction in the severity of punishment (Anderson and Putterman, 2006; Carpenter, 2007a; Nikiforakis and Normann, 2008). However, subjects in these experiments are given only one opportunity to punish other individuals and counter-punishing is impossible. Therefore, the psychological benefits from defying the risk of a feud were absent in previous studies. By observing whether punishment activity is intensified or diminished by the risk of a feud we can infer whether the psychological benefits from punishing outweigh the increase in the expected (monetary) cost of punishment.

The second question this paper wishes to investigate is whether permitting individuals to punish can be efficiency enhancing when this can lead to a feud. The answer is not straightforward, as Axelrod's quote suggests. On the one hand, the fact that feuds usually erupt when law enforcement is absent or too costly implies that the threat of revenge is often the only means of preventing individuals from "socially disruptive behaviors" that will otherwise take place (Boehm, 1984, p.88). Indeed, a well known fact from laboratory experiments is that cooperation in the public good game unravels when punishment is not permitted (Ledyard, 1995). On the other hand, revenge carries significant economic costs. In naturally occurring situations, the parties involved in feuds engage in unproductive and often destructive activities that lower peers' morale, reduce productivity, undermine cooperation, increase health care costs and lead to suboptimal decisions. Inside the laboratory, feuds would lower the earnings of group members and perhaps undermine the willingness to cooperate with each other in the future.

These two conflicting effects make revenge a "deeply ambivalent cultural and social institution" (Waldmann, 2001, p.436). Sociologists and anthropologists studying feuds tend to agree that "[t]he question remains whether feuds created more disruption than they controlled" (Boehm, 1984, p.183). Similarly, Elster (1990), who discusses blood feuds (an extreme form of feud that involves killings), writes that: "To decide whether the feud served the function of reducing the overall level of violence, some *unanswerable questions* would have to be answered" (p.876, emphasis added). The controlled environment of the laboratory allows us to address questions that might indeed be unanswerable in the field.

In particular, laboratory experiments allow us to create counterfactuals and compare individual behavior in the presence and absence of a feud threat, and identify conditions that amplify the likelihood of feuds.

Our results show that subjects employ strategies that avoid the break out of feuds. In our main treatment (*Feuds*) where retaliation is unavoidable as individuals interact repeatedly in fixed groups using fixed identification numbers, we find that individuals largely abstain from punishing to avoid feuds that could span across many periods. Cooperators in this treatment are punished as much as free riders. As a result cooperation unravels in most groups. Surprisingly, a couple of groups manage to fully cooperate without using punishment. Overall, the group savings from the low punishment activity are just enough to compensate for the low levels of cooperation. In another treatment in which individuals interact in fixed groups but are assigned new identification numbers in every period (*Escalation*), thus preventing counter-punishment across periods, we find that participants wishing to punish avoid feuds by punishing their targets severely and depriving them of funds which would allow them to counter-punish.

The rest of the paper is organized as follows. In the next session we present the experimental design, followed by our hypotheses in Section 3. In Section 4 we analyze behavior in the experiment, and Section 5 concludes.

2 Experimental Design

The experiment consists of three treatments. The design of the treatments follows that used in previous public-good experiments with decentralized punishments. The game is divided into a number of stages. The first stage is a standard public good game. Four players simultaneously decide how much of their initial endowment of 20 Experimental Currency Units (ECU) to contribute to a public account. Contributions are multiplied by 1.6 and then distributed equally among the four players. Hence, the marginal per capita return is $1.6/4 = 0.4$. If the players' contributions are $c_i, i = 1, ..., 4$, player i's earnings after the contribution stage are given by

$$\pi_i^0 = 20 - c_i + 0.4 \sum_{h=1}^{4} c_h.$$

The contribution stage is followed by a number of punishment stages. The exact number of the stages, as we discuss below, depends on the treatment. In each punishment stage s, player i can assign *punishment points* to group member j, p^s_{ij}, as long as $\pi^{s-1}_i > 0$ and $\pi^{s-1}_j \geq 0$. Assigning a punishment point costs the punishing player 1 ECU and the punished player 2 ECU.[2] Only integer numbers of points can be assigned and the maximum total number of points a player can assign is equal to his current payoff. That is, $\sum_{j=1,j\neq i}^{4} p^s_{ij} \leq \pi^{s-1}_i$. Player i's earnings at the end of punishment stage s are, therefore, given by

$$\pi^s_i = \pi^{s-1}_i - 2 * \sum_{j=1,j\neq i}^{4} p^s_{ji} - \sum_{j=1,j\neq i}^{4} p^s_{ij.}$$

Table 1 summarizes the experimental treatments. In the (B)aseline treatment, there is one punishment stage similar to most studies so far.[3] In the two main treatments, (E)scalation and (F)euds, the number of punishment stages is ex-ante not restricted, but endogenous. In these treatments, a period ends if no punishment is meted out in a given stage, or if further punishment is not permitted. A punishment in stage s is not permitted if $\pi^{s-1}_i \leq 0$, for all $i \in [1, 4]$, or if at least three players cannot be assigned any further punishment (that is, their earnings in the current period are already negative).[4] The fact that punishment is restricted to integers ensures that a period will end after finitely many punishment stages (namely a maximum of 44). As we will see in Section 4, no period came close to lasting that long.

In all the treatments, in the first punishment stage every player is informed about the contributions to the public account of each player in the group. In further punishment stages, every player is informed about the number of points assigned by each individual

[2] Nikiforakis and Normann (2008) study the effect on cooperation and efficiency of changing the income reduction caused by a single punishment point available at a cost of 1 ECU to the punisher. A factor of 2 was sufficient to stabilize contributions at an intermediate level in a standard design with one punishment stage. Since adding further punishment stages can potentially lead to further increases as well as decreases of the contribution level, using a factor of 2 is appropriate as it permits both improvements beyond and deterioration below the contribution level achieved with the basic design with one punishment stage.

[3] The data for this treatment is taken from Nikiforakis and Normann (2008) that followed exactly the same protocol and drew participants from the same subject pool.

[4] Note that player i with $\pi^{s-1}_i = 0$ can be sanctioned although he cannot sanction. This ensures that a player cannot immunize himself against retaliation by spending all his income on punishing others.

Treatment	No. of punishment stages	Subjects' ID	Number of subjects
(B)aseline	1	Random	24
(E)scalation	≥1, Endogenous	Random	28
(F)euds	≥1, Endogenous	Fixed	24

Table 1: Overview of the Experimental Treatments

player to each of the players in the previous punishment stage. Hence, this permits both retaliation as a player can see who punished him, and sanction enforcement (that is, the punishment of non-punishers) as he can observe who is participating in the punishment of free-riders.

In order to track behavior *within* a period across stages, each player is given an identification number that stays the same for all stages of a given period. To prevent retaliation *across* periods, however, in treatments B and E the identification numbers are assigned randomly and independently for each period. In contrast, in treatment F the identification numbers are fixed for the whole course of the experiment.[5] The reason for keeping identification numbers fixed in treatment F is that outside the laboratory, revenge is unavoidable in repeated interactions. Even in the extreme case that one exterminates one's prime "opponent", a friend, colleague or family member of the victim can (or often must) take revenge. The fixed identification numbers ensure that even if an individual is prevented from counter-punishing in a given period by a budget constraint, he can take revenge in the following period when new funds are received.[6] This characteristic makes treatment F the most "realistic" in the sense that it most closely matches conditions found outside the laboratory.

In all treatments, the stage game is repeated 10 times and group composition remains unchanged. The computerized experiments were run using z-Tree (Fischbacher, 2007) in

[5] This is a crucial difference between our and previous designs. With the exception of Ostrom, Walker and Gardner (1992) who study a common-pool resource game with a single punishment stage, all previous studies on peer punishment use player identification numbers that change across periods so that players can avoid counter-punishment in subsequent periods.

[6] While in reality it is often another person from the same "clan" who seeks revenge, the incentives for avoiding a feud are the same in our set up as outside the laboratory when revenge cannot be avoided. This is the simplest way of analyzing the institution of feud and evaluating its impact on punishment behavior and efficiency. Our implicit assumption is that a family member or a friend of the victim would take revenge much in the same way as the victim would if he was able to do so.

the laboratory of Royal Holloway, University of London. We conducted two sessions per treatment. In each session either 12 or 16 subjects participated. Given the fixed matching into groups of four subjects, we have six independent observations in treatments B and F and seven in treatments E. The total number of subjects was 76. Subjects were students at Royal Holloway from various disciplines who were recruited via email. None of the subjects had previously participated in a public good experiment or had experience with game theory.

Written instructions including control questions were handed out at the beginning of the experiment. The experiment started only after all subjects answered the control questions correctly. Participants were paid in cash at the end of the experiment with earnings calculated at an exchange rate of ECU 25 = £1. An initial capital of £5 (that was used to cover possible losses) was paid in addition.[7] The sessions lasted between 60 and 120 minutes, and average earnings (excluding the initial capital) reached £9.56 (B), £8.56 (E), and £9.25 (F).

3 Hypotheses

We first consider subgame-perfect equilibria (with commonly known selfishness of all players). In the stage game, given that punishment is costly, there will be no punishment in the last possible punishment stage for any history of play (remember that even in the treatments with an endogenous number of punishment stages, there is a theoretical upper limit on the number of possible punishment stages). By backward induction, there will be no punishment in any stage. Given that there is no punishment, the subgame-perfect choice in the contribution stage is to contribute zero. Given that the game is repeated a known finite number of times, the same holds for each period of the repeated game and for all treatments. Behavior in previous experiments, however, deviates substantially from this prediction as individuals contribute positive amounts and punish free riders.

In an attempt to rationalize behavior observed in laboratory experiments, a number of behavioral models have been proposed that include preferences over the distribution of

[7] In the baseline treatment (B), the initial capital was £1. The initial capital was higher for the treatments E and F to account for the possibility of costly escalation.

payoffs (such as Fehr and Schmidt, 1999, or Bolton and Ockenfels, 2000), or preferences for reciprocity (such as Dufwenberg and Kirchsteiger, 2004). According to these models individuals might be willing to punish others. Moreover, the cost of punishing is inversely related to the punishment threat.[8] The possibility of counter-punishments and feuds increases the expected cost of punishment. Consequently, punishment strategies that avoid reprisals are less costly in expected terms.

In our baseline treatment there is no risk of counter-punishment and, therefore, the cost of punishment equals the cost of purchasing punishment points. In treatment E the possibility of costly escalation raises the expected cost of punishing in comparison to B. The extent of escalation is even greater in F as individuals can retaliate not only within but also across periods. In addition, individuals in E can avoid counter-punishments by using extremely severe punishment that deprives victims of any funds which would allow them to take revenge. This is not the case in treatment F where one cannot avoid retaliation and where feuds might escalate across periods. Therefore, we derive

Hypothesis 1: *As the expected cost of punishment increases monotonically from treatments B to F, we expect the frequency of first stage punishment to follow the ranking* $B > E > F$.

In addition to the effect on the punishment frequency, the institutions are likely to affect the severity of punishment. In treatments E and F, individuals can use severe punishment strategically to signal to the victims that they will retaliate acts of counter-punishment. Moreover, an individual in treatment E can avoid counter-punishment and entering a feud altogether by punishing his target severely. This is not possible in F. Therefore, punishment severity should be higher in E than in F. Hence we derive

Hypothesis 2: *We expect the severity of first stage punishment to follow the ranking* $E > F > B$.

Hypotheses 1 and 2 suggest that first stage punishment is both more likely and more severe in E than in F (for given levels of contribution to the public account). Since previous experiments suggest that increased punishment threats lead to higher contribution levels

[8] The hypothesis that the cost of punishment matters is supported by the experimental evidence of Anderson and Putterman (2006), Carpenter (2007a) and Nikiforakis and Normann (2008).

(Nikiforakis and Normann, 2008), we expect contributions to be higher in E than in F. Hypothesis 1 states that punishment is most frequent in B, but Hypothesis 2 states that punishment will be the least severe in B. Therefore, it is difficult to predict the net effect of these contrasting forces. Hence we can only derive

Hypothesis 3: *We expect contributions to the public account to be higher in E than in F.*

Hypotheses 1 to 3 imply that the impact of the different institutions on efficiency (defined as the sum of players' final earnings) is difficult to predict. On the one hand, the higher expected cost will have a positive effect on efficiency by reducing the frequency of punishment. The effect on efficiency can be significant in light of the fact that several previous experimental studies have shown that subjects over-use punishment in treatments like B (for a brief survey see Gächter and Herrmann, in press; and Nikiforakis, 2008). On the other hand, the increase in punishment severity might lead to overall higher punishment expenses. Moreover, even if less is spent on punishment in E and F than in B the weakening of the punishment threat is expected to lower overall contributions and, consequently, efficiency. How these effects add up is not obvious. For the same reason, the prediction of the efficiency ranking between E and F is difficult in spite of the clear prediction regarding punishment and contributions. The absence of clear (or intuitive) predictions regarding the difference in efficiency is one of the main motivations for this study.

The three hypotheses are based on the assumption that punishments occur less frequently as they become more expensive. However, as discussed in the introduction, Elster (1990) argues that the frequency of punishment might be positively related to the expected cost of punishment. One reason for this might be that punishment of wrong-doers protects (or even boosts) one's self-esteem as Elster suggests.[9] Another reason is that punishment can be used strategically by forward-looking individuals to send a signal that one is not afraid to punish even under the threat of escalation. This implies that it is far from trivial that Hypotheses 1, 2 and 3 will be supported.

[9] Akerlof and Kranton (2000) suggest a model where individuals take actions to protect their identity. The authors suggest that the need to reciprocate hostile actions might be due to a need to protect one's self-esteem (p.728).

4 Experimental Results

We start the discussion of the results with a study of the punishment behavior and consider the impact on contribution levels and efficiency in the following sections. With *feuds* we refer to cycles of punishment and counter-punishment. Feuds can occur both in treatment E and in treatment F. The difference is that feuds can span more than one period in F, but not in E.

4.1 Punishment Behavior

4.1.1 Overview

The first question to investigate is whether feuds break out in the experiment. Figure 1 shows the distribution of stages in which punishment was last carried out in individual groups. "0" indicates that no punishment was meted out in a given period in a particular group; "1" indicates that punishment was observed in the first punishment stage, but none in further stages; "2" indicates that there was some punishment in the first and second punishment stages, but none in the third punishment stage, and so on.

Let us first consider feuds within periods. As seen in Figure 1 individuals in the experiment do not engage in feuding. Punishment appears to be more frequently observed in B and least frequently in F – in line with Hypothesis 1.[10] Indeed, in more than half of the cases in F, we observe no punishment activity at all. In treatments E and F, individuals punish infrequently in the second punishment stage (35 and 14 punishment cases, respectively) and very rarely in a higher punishment stage.[11]

A feature of treatment F is that feuds can also occur across periods. A clear case of an

[10] This finding is in contrast to the results of Denant-Boemont et al. (2007) who observe punishment in each of the five punishment stages. The difference could be due to two factors: first, the threat of indefinitely long sequences of retaliation in our experimental design and, second, the convex cost function for punishment used by Denant-Boemont et al., which creates incentives to spread out punishment of an individual across stages.

[11] There is only one case of fourth-stage punishment in each treatment and no cases of punishment in higher stages. In most instances, a period ended because participants decided not to punish further even though they could afford to do so. There is only one case in which a period stopped because punishment was infeasible. This happened in F. However, on a number of occasions (in 15 out of 70 periods in E and in 6 out of 60 periods in F) at least one participant could not afford to retaliate due to negative earnings.

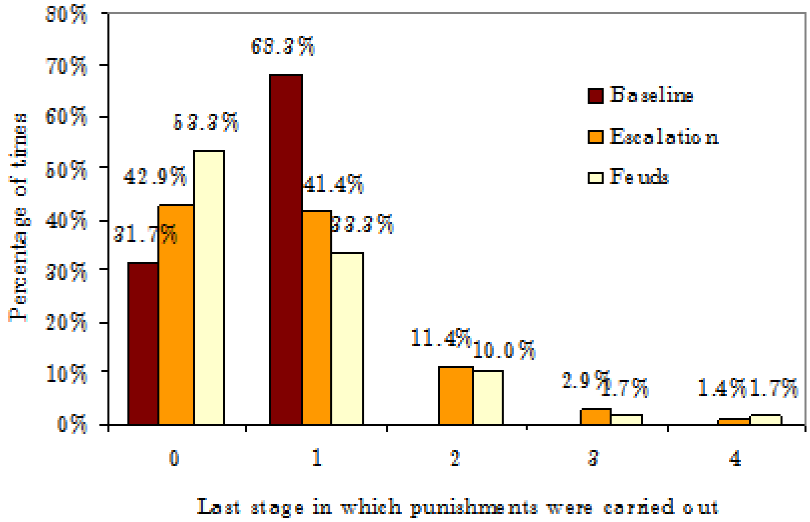

Figure 1: Distribution of stages in which punishment was last carried out, with 0 indicating that no punishment occured in a given period in a particular group.

(interperiod) feud would constitute a pattern where player A punishes player B in period t, B counter-punishes A in period $t+1$, A counter-counter-punishes B in period $t+2$, and so on.[12] In the six groups of treatment F, we find effectively no such activity at all, even if we consider more general patterns. Indeed, the virtual absence of feuds is striking as there appear to be fruitful grounds for the start of a feud, that is, repeated acts of punishment of one player by another.[13]

There is only one case that at a first glance looks like a feud. Two individuals repeatedly punish each other in the first four periods. However, since these subjects already started punishing each other in period 1, it is not clear whether they are counter-punishing or not. In particular, one player punishes all other players in three of the first four periods (and

[12] In general, since the point of starting (or continuing) a feud is to make clear that one does not tolerate "insults", retaliation should typically occur in the same period, or if this is not feasible, in a period following immediately or only shortly after the previous punishment.

[13] As an example of avoiding a feud, one subject was repeatedly punished by the same group member in the first three periods without responding. That is, until the final period where the individual retaliated heavily for the previous punishment. The number of punishment points assigned was exactly the minimum level required to bankrupt the other player so that counter-punishment was impossible.

two players in the fourth) and he is also the only player to contribute fully while others contribute between 0 and 10. Thus it appears that this player's punishment is driven by a desire to punish free-riders rather than continuing a feud. Moreover, if the motivation had been to counter-punish, we would expect this to occur also within the same period - if feasible. This is not the case. We summarize:

Result 1: *Feuds are not observed in the experiment. Individuals rarely use the third and fourth punishment stage and never a punishment stage of higher order. There is also no evidence of feuds across periods in F.*

Figure 1 shows that punishment rarely occurs after the first punishment stage. For this reason, and because our main interest is how the possibility of a feud affects altruistic punishment and contributions to the public good account, our analysis of punishment will focus on first-stage punishment. A detailed discussion of second and higher stage punishment can be found in a companion paper (Engelmann and Nikiforakis, 2009).

4.1.2 First-stage Punishment

Given our hypotheses, we consider the decision to punish (*punishment decision*) separately from the decision of how much to punish (*punishment severity*). Figure 2 shows the likelihood of an individual being punished in the first punishment stage as a function of his deviation from the average contribution of his peers. For simplicity, we refer to individuals contributing less than their peers on average in a given period as *low contributors* and individuals contributing more than their peers as *high contributors.* Two aspects in Figure 2 are noteworthy. First, both low and high contributors are generally less likely to attract punishments in treatment E than in treatment B. Punishment appears to be even less likely in F. This behavior clearly supports Hypothesis 1. Second, while there seems to be a positive correlation between negative deviation (i.e. the absolute difference between own and average contribution of peers for low contributors) and punishment likelihood in treatments B and E (similar to previous experiments), such a relationship does not appear to exist in F. A possible explanation for this behavior is that subjects believe there is a positive relation between anti-social behavior in the contribution stage and anti-social behavior in the punishment stage. Therefore, they might fear that those who completely free-ride are also

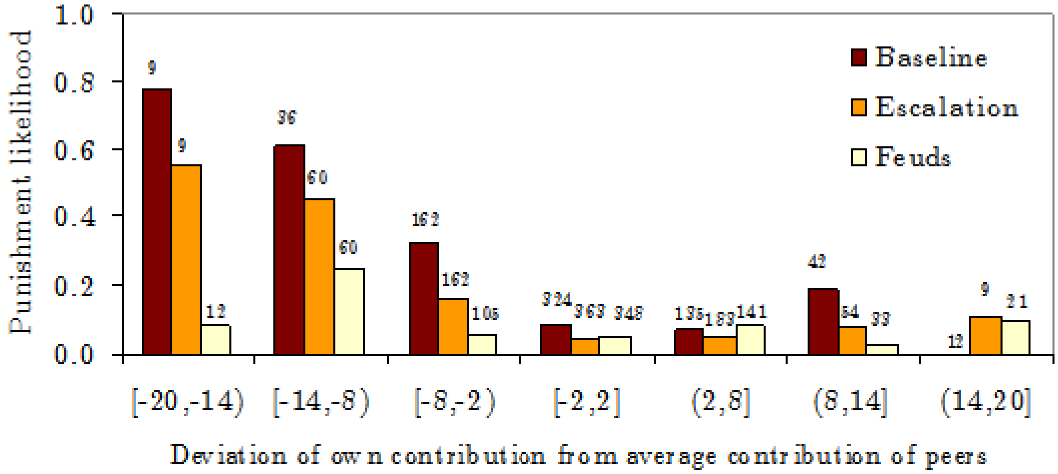

Figure 2: Likelihood of being punished depending on deviation from the average contribution of the other three group members in a given period. Numbers on top of bars indicate the number of observations in the category.

more likely to counter-punish. Note that in the lowest bracket [-20,-14) the other three group members must contribute a significant fraction of their endowment. Therefore, it is not clear who among them should punish the low contributor. Given the high expected cost of punishment (due to retaliation) the second-order public good property of punishing comes to full bearing.

To provide statistical support for these observations we present the results from a regression analysis in Table 2. 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 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 independent variables included in the regression are: E, a dummy variable taking the value of one if the observation comes from treatment E and zero otherwise; F, a dummy variable for treatment F; (*Absolute*) *Negative Deviation* defined as

$negdev_{j,t} = \max\{0, (\sum_{h \neq j} c_{h,t})/3 - c_{j,t}\}$, where $c_{i,t}$ is the contribution of individual i in period t; *Positive Deviation,* defined as $posdev_{j,t} = \max\{0, c_{j,t} - (\sum_{h \neq j} c_{h,t})/3\}$; *Period*, a variable to control for time effects; and the interaction of these variables with the treatment dummies. The repeated nature of the experiment requires that we also control for random effects at the group level.[14]

The (marginally) significant negative coefficients of E and F in the punishment-decision regression indicates that, in line with Hypothesis 1, subjects in the respective treatments are less likely to punish than subjects in B. The significant positive coefficient of *(Absolute) Negative Deviation* reveals that the greater the extent of free riding, the greater is the likelihood of punishment in the baseline treatment. The insignificance of $E*$*(Absolute) Negative Deviation* shows that a similar relation between free riding and punishment exists in E. However, this relation is significantly weaker in treatment F as observed above.[15] Thus, if there is a risk of feuds breaking out, the reciprocal reaction (in the sense of stronger free-riding being more likely to attract punishment) is weakened. This would be consistent with contributors assuming that those who free-ride more are more likely to take revenge.[16] Finally, the likelihood of punishment occurring decreases significantly over time, which is consistent with punishment serving a strategic purpose. We summarize the main finding regarding the frequency of punishment in:

Result 2: *The frequency of first-stage punishment is consistent with Hypothesis 1 and follows the ranking $B > E > F$.*

Result 2 is in line with the assumption that the increased expected cost of punishment in

[14]We also ran all the regressions reported in this paper with group fixed effects, as well as with nested clusters. None of the main results is affected.

[15]If we run the same regression separately for treatment F, the coefficient of *(Absolute) Negative Deviation* is significant (coeff: .039; $p = .04$), even though the slope is substantially flatter. Also, substituting variable E and the respective interaction terms with variable B (dummy for the respective treatment) we find that $F*$*(Absolute) Negative Deviation* is significant (coeff: $-.117$; $p < .01$). This implies that the difference between E and F is also significant.

[16]Indeed, we find some evidence that such a belief might be justified. While we have very few observations that can be attributed as counter-punishment within a period, namely four, all originate with free-riders, with three counter-punishers having contributed 0 and the other 5, when the average of the other three group members was 8.33. Identifying counter-punishment across periods is confounded with other possible reasons of punishment.

	Punishment Decision		Punishment Severity		Income Loss	
	coefficient (std. error)	p-value (z)	coefficient (std. error)	p-value (z)	coefficient (std. error)	p-value (z)
constant	-0.5486 (0.1986)	0.006 (-2.76)	1.8792 (0.9505)	0.048 (1.98)	0.8187 (0.3522)	0.020 (2.32)
E(scalation)	-0.5763 (0.2950)	0.051 (-1.95)	-0.3885 (1.4667)	0.791 (-0.26)	-0.5975 (0.5070)	0.239 (-1.18)
F(euds)	-0.5296 (0.3057)	0.083 (-1.73)	1.7581 (1.7816)	0.324 (0.99)	0.0679 (0.5196)	0.896 (0.13)
$negdev_{j,t}$	0.1309 (0.0186)	<0.001 (7.05)	0.1451 (0.0686)	0.034 (2.11)	0.2911 (0.0383)	<0.001 (7.61)
$E\times\ negdev_{j,t}$	0.0245 (0.0272)	0.369 (0.90)	-0.1041 (0.1165)	0.372 (-0.89)	0.1226 (0.0537)	0.022 (2.28)
$F\times\ negdev_{j,t}$	-0.0921 (0.0268)	0.001 (-3.43)	-0.2562 (0.1461)	0.079 (-1.75)	-0.2539 (0.0530)	<0.001 (-4.79)
$posdev_{j,t}$	-0.0345 (0.0245)	0.159 (-1.41)	0.0462 (0.1256)	0.713 (0.37)	0.0308 (0.0370)	0.406 (0.83)
$E\times\ posdev_{j,t}$	0.0591 (0.0347)	0.089 (1.70)	-0.0115 (0.1924)	0.952 (-0.06)	0.0469 (0.0545)	0.390 (0.86)
$F\times\ posdev_{j,t}$	0.0162 (0.0335)	0.629 (0.48)	-0.1203 (0.2005)	0.548 (-0.60)	-0.0448 (0.0514)	0.383 (-0.87)
period	-0.1532 (0.0244)	<0.001 (-6.28)	-0.0687 (0.1189)	0.563 (-0.58)	-0.1322 (0.0415)	0.001 (-3.19)
$E\times$ period	-0.0009 (0.0370)	0.980 (-0.02)	0.9613 (0.2020)	<0.001 (4.76)	0.0944 (0.0581)	0.104 (1.63)
$F\times$ period	0.0584 (0.0376)	0.120 (1.55)	-0.0564 (0.2322)	0.808 (-0.24)	0.0447 (0.0606)	0.461 (0.74)

Table 2: Hurdle model for the decision whether to punish (first two columns) and the punishment severity conditional on punishment occurring (next two columns). The last two columns capture the resulting expected income loss from being punished.

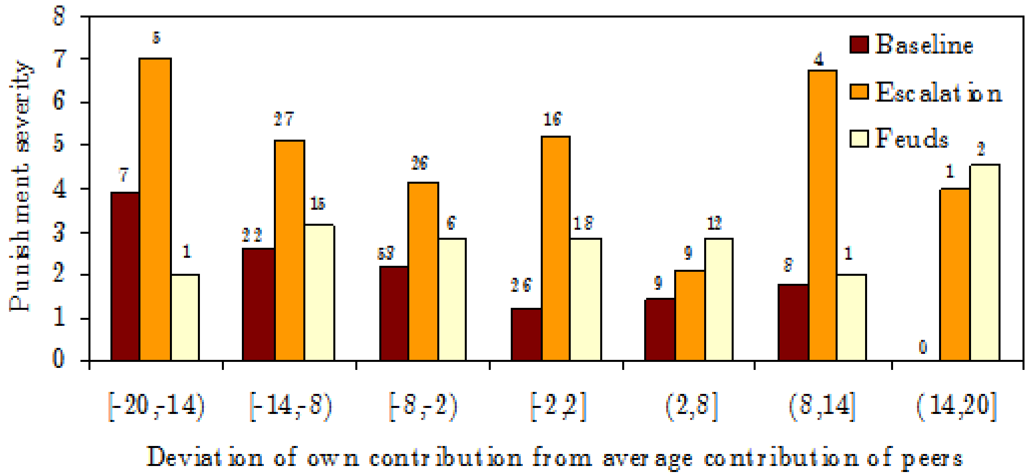

Figure 3: Severity of punishment received conditional on being punished, depending on deviation from the average contribution of the other three group members in a given period. Numbers on top of bars indicate the number of observations in the category (i.e. the number of players deviating by the given amount and being punished).

the form of threats of escalation and feuds reduces the inclination to punish. The alternative hypothesis that the higher threat of feuds might increase the tendency to punish due to reasons related to self-esteem does not find support.[17]

Figure 3 presents the severity of punishment (i.e. the number of punishment points assigned, conditional on punishment occurring) as a function of an individual's deviation from the average contribution of his peers. With the exception of treatment B, there does not appear to be a systematic relation between the extent of positive or negative deviation and the severity of punishment. In line with Hypothesis 2, punishment severity appears to be highest in E followed by F and then B. It is worth noting that the average punishment severity in E is twice that in B (4.2 versus 2.1 punishment points).

The punishment-severity regression in Table 2 reveals some further interesting aspects of subjects' behavior when punishments can lead to feuds. While low contributors in treatment

[17] This does not preclude the possibility that the higher cost increases *some* players' utility from punishing, but only that these concerns are dominated by the direct effects of the cost increase.

B are more heavily punished as the extent of free riding increases, in treatments E and F no such relation is observed.[18] The significant positive relation between $E * Period$ and punishment severity is probably indicative of the fact that subjects learned over time that an effective way of avoiding counter-punishments is to increase the severity of their punishment. The following result summarizes the main findings concerning the severity of punishment.

Result 3: *Punishment severity does not differ among treatments initially, but becomes significantly more severe in E over time, providing partial support for Hypothesis 2.*

Result 3 supports the argument that players might punish more severely in treatment E as this limits the victim's ability to counter-punish. The (weakly) significant negative coefficient of $F * negdev_{j,t}$ suggests that some individuals might interpret the intensity of free-riding as a signal that punishment will be retaliated. As noted above, such a belief might be justified.

Before we turn our attention to the question of how the different institutions affect contributions to the public account, we need to examine the threat that punishment poses to subjects. Figure 4 shows how the differences in the likelihood of being punished and in punishment severity translate into expected income losses in each treatment. Even though punishment is less likely to occur, the punishment threat appears to be generally higher in E than in B due to the higher punishment severity in the former. The punishment threat for free-riders is substantially lower in F than in the other treatments. Note that in all treatments free-riding pays in expected monetary terms as punishment activity is relatively low; total expenses on punishment are below 2 ECU per player per period compared to the 20 ECU endowment at the beginning of each period or the 8 ECU which is the minimum that a player would have available in the first punishment stage (if he contributes all his endowment while none of the other players contributes anything). However, individuals have been shown to react to punishment even if punishment does not reduce their income (Masclet et al., 2003). This suggests that there is a non-monetary component in the "damage" inflicted by punishment. Hence in spite of being substantially smaller than required

[18] Using E instead of B as the omitted category in the regression we find that *(Absolute) Negative Deviation* is far from being significant (coeff: .0411; $p > .6$) and so is $F*$*(Absolute) Negative Deviation* (coeff: $-.152$; $p > .3$).

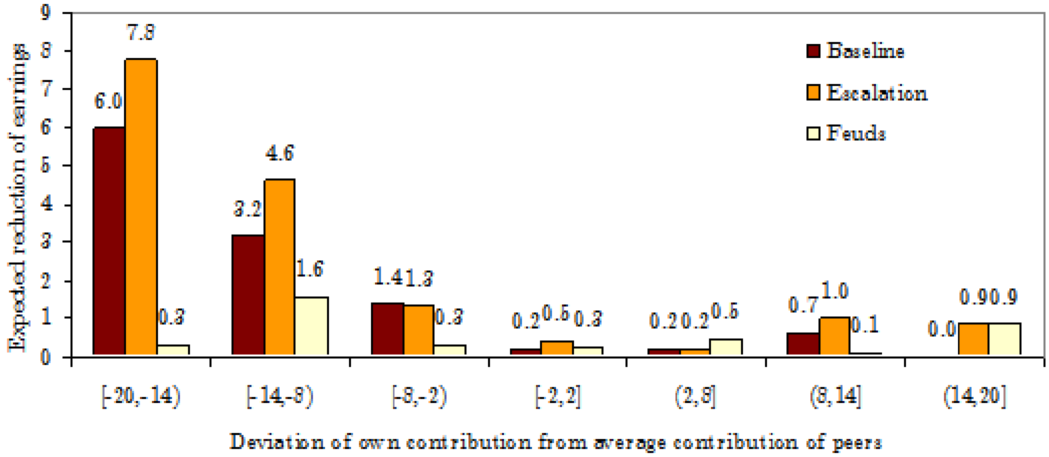

Figure 4: Expected reduction of income through received punishment depending on deviation from the average contribution of the other three group members in a given period. Numbers on top of bars state the expected reduction in ECU.

to deter free-riding in monetary terms, punishment might nevertheless have an effect in enforcing contributions.

The third regression in Table 2 provides support for the observations made in Figure 4. The greater the extent of free riding by the low contributors, the greater is the expected reduction in their earnings due to punishment in treatment B. The expected reduction is even greater in treatment E as the effect of punishment severity outweighs that of punishment frequency. In contrast, the expected income reduction is less for free riders in treatment F. In fact, there is no relation between free riding and income reduction in treatment F.[19] This is striking and suggests that the existence of punishment opportunities is unlikely to support cooperation in treatment F. It is also worth noting that the coefficient of $E*Period$ misses only narrowly the 10-percent level of significance. That is, while the punishment threat decreases over time in B, this trend is substantially weaker in E. We summarize:

Result 4: *There is no relation between the extent of free riding and the income reduction*

[19] Using E instead of B as the omitted category in the regression we find that $F*$*(Absolute) Negative Deviation* is highly significant (coeff: $-.367$; $p < .01$). If instead we use F as the omitted category the coefficient of *(Absolute) Negative Deviation* is not significant (coeff: .037; $p > .3$).

from punishment in treatment F. In contrast, there is a strong positive relation in B which is even greater in E. That is, the treatments' ranking according to the threat low contributors face is $E > B > F$.

4.2 Contributions to the public account

Figure 5 shows the development of the average contribution across treatments over time. For comparison, Figure 5 presents also the evolution of the average contribution in a treatment without punishment opportunities (VCM).[20]

The average contribution starts at 11.13 ECU in treatment B and remains relatively stable throughout the experiment. In treatment F, the average contribution starts at 11.08 ECU, but then declines over time. It appears, however, to be stabilized at a level that is substantially higher than in VCM, which is surprising given the virtually non-existent punishment threat (Figure 4 and Result 4). The average contribution in treatment E follows a reverse pattern to that in F, starting lower (9.61 ECU) but increasing between periods 2 and 5 to coincide almost perfectly with B. Table 5 (in the appendix) presents information at the group-level.

To provide statistical support for the above observations, Table 3 presents the results from a regression of individual contributions on treatment dummies (with B being the omitted category) and the interaction of the treatment dummies with *period*. The regression controls for group-level random effects. While all treatment dummies are far from significant, there are significant differences in the time trends. Compared to the baseline, where contributions do not follow a time trend, contributions increase significantly over time in E and decrease significantly in F.[21] The upward trend in E might actually be an indirect

[20] The data for the VCM treatment consists of six groups of four individuals taken from Nikiforakis and Normann (2008). The VCM sessions followed the same experimental protocol. They were conducted in the same laboratory, drawing subjects from the same pool. Thus, they are equally comparable to treatments B, E, and F.

[21] Regressing contributions to the public account on *period* in the individual treatments, we find that the observed increase over time is significant in E (coeff: .268; $p < .01$) and so is the decrease in F (coeff: $-.607$; $p < .01$). If we add the data from the VCM sessions and use VCM as the omitted category, we find that none of B, E and F differs initially from VCM but that the time trend in all of them is significantly higher than in VCM. If we run the regression without controlling for time trends we find that contributions are higher in B and E than they are in VCM, but the difference between F and VCM is not significant.

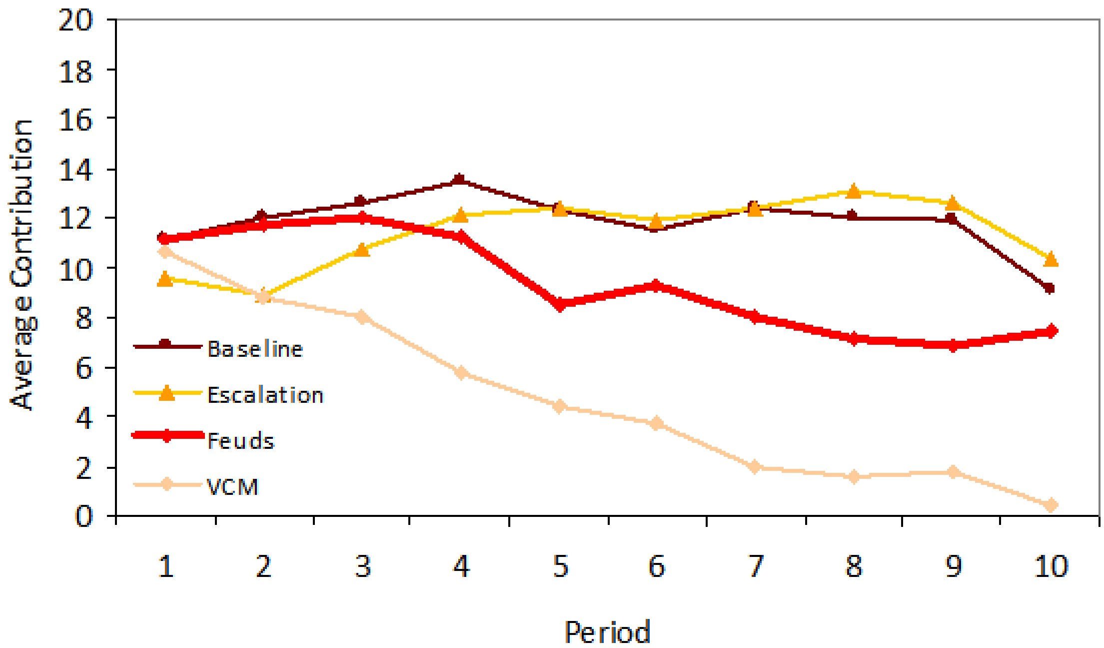

Figure 5: Development of average contribution across periods by treatment. For comparison, VCM shows data from a public good experiment without punishment opportunities. The data are taken from Nikiforakis and Normann (2008) who follow identical procedures. VCM stands for Voluntary Contribution Mechanism.

effect of the strong punishment which we believe to be primarily driven by a desire to avoid counter-punishment. In this specific case, the risk of counter-punishment appears to have a positive effect on contribution since it leads to more severe punishment. We summarize:

Result 5: *Contributions start at a similar level in all treatments. Over time, the average contribution increases significantly in E, remains stable in B, and decreases in F. Therefore, while the initial levels of contribution do not support Hypothesis 3, the time trends are consistent with it.*

The fact that cooperation does not unravel completely in F despite the weak punishment threat seems startling. A closer look at behavior at the group level allows for a better understanding of the evolution of cooperation in F. Table 5 reveals that there is substantial

	coeff	std. error	z	p-value
constant	12.7056	2.5695	4.94	< 0.001
E(scalation)	−2.7913	3.5016	−0.80	0.425
F(euds)	−0.0417	3.6338	−0.01	0.991
period	−0.1586	0.1127	−1.41	0.159
$E\times$ period	0.4268	0.1536	2.78	0.005
$F\times$ period	−0.4485	0.1594	−2.81	0.005

Table 3: Linear regression for the individual contribution with group-level random effects. Treament B is the omitted category, so period covers the time trend in B and the interaction terms the difference between the time trends in the respective treatment and B.

heterogeneity across the six groups in F with respect to cooperation. The average contribution is below 3.80 ECU in four out of six groups in the final period, while in the remaining two groups it is greater than 19.50 ECU. The infrequent use of punishment can at least partly explain the reason why cooperation unravels in the first four groups in F.[22] The high contribution levels in the sixth group in F are most likely due to the fact that all group members appear to be "like-minded" cooperators (Gächter and Thoeni, 2005) in the sense that the average contribution in the first period was unusually high (17.50 ECU). The average contribution in the fifth group is at an intermediate level in period 1 (12.50 ECU). However, even though there is no punishment activity in period 1, the average contribution in period 2 "jumps" to 17.25 ECU. This might be due to two group members contributing the whole of their endowment in periods 1 and 2, thus signalling their intention to continue cooperating.

In contrast to the picture in F, treatment E shows less pronounced patterns: The average contribution is between 8 and 13.75 ECU in the final period in five out of seven groups. Contributions reach the maximum level in only one group (group 4) and collapse completely in another group (group 7).[23]

[22] In two groups, a player attempts to promote cooperation by contributing to the public account and punishing free-riders, but is either met with indifference or is counter-punished.

[23] While there is some decrease from period 9 to 10, this is substantial only in one group, falling from 19.5 to 11.25.

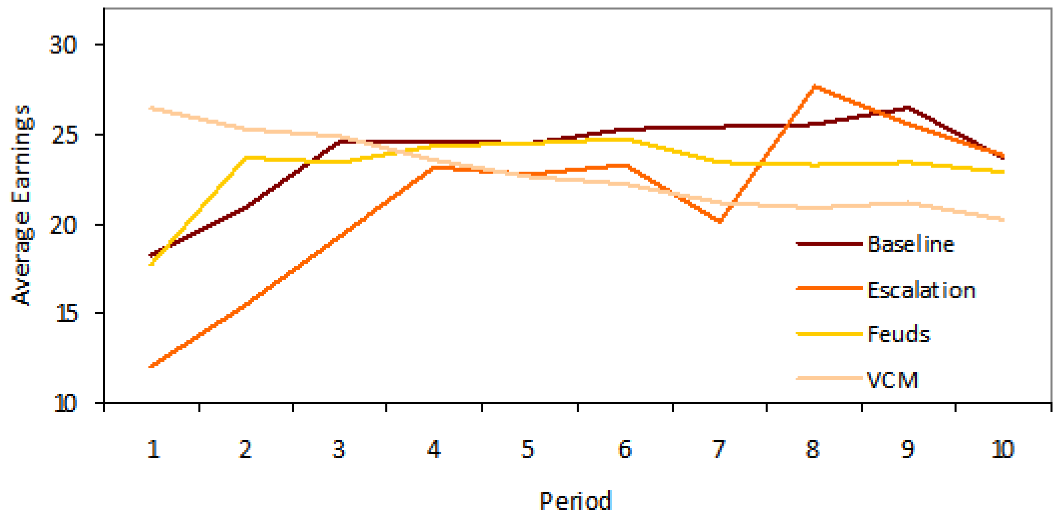

Figure 6: Average total earnings across treatments by period. VCM again shows the data from a public good experiment without punishment for comparison.

4.3 Earnings

The impact of the different institutions on earnings (as a measure of efficiency) is difficult to predict. Cooperation is the lowest in F, but so are the expenditures on punishments. How do these effects add up?

As can be seen from Table 5, the different effects seem to cancel out as earnings do not differ substantially across treatments. Figure 6 shows the development over time of average earnings. We see that earnings in F are at a level similar to those in B for the first six periods. As a result of the severe punishments, which are particularly common in the first three periods, earnings are initially substantially lower in E, but eventually exceed earnings in F and match those in B.

Table 4 shows the results of a linear regression with group-level random effects for individual earnings per period. B is again the omitted category. Confirming the picture in Figure 6, earnings in E are initially significantly lower than they are in B, but earnings increase significantly more in E than in B, even though they also increase significantly in B. Thus while initially harsher punishments lead to lower earnings in E, this is compensated over time by the relative increase in contributions. This effect is, however, not strong enough to compensate for the initial losses as total earnings are lower in E than in B. Even though the time trend in F is below that in B, this difference fails to be significant.[24] These results stand in contrast to the findings of Denant-Boemont et al. (2007) who find earnings to be the lowest in a treatment with five punishment stages. As the two experiments differ in more than one dimension (e.g. endogenous number of stages, punishment technology, subject pool) it is difficult to identify the reason for this difference.

If we add the data from the VCM treatment to the regression and use VCM as the omitted category, we confirm the observations made by looking at Figure 6; earnings in B, E, and F are initially lower than in VCM ($p < 0.001$ for E, $p < 0.05$ for B, $p = 0.1$ for F) and the time trends are significantly greater that in VCM ($p < 0.001$). These effects roughly cancel out across the whole experiment. If we exclude the interaction terms, none of the treatment dummies is significant ($p > 0.5$). The following result provides an answer to one of the two questions posed in the introduction.

Result 6: *Earnings are not different across treatments. Specifically, this means that the opposing effects of the feud institution on contributions and on punishment behavior cancel out. As a result, efficiency is not significantly different from the treatment without punishment opportunities or from a treatment with a single stage of punishment.*[25]

5 Discussion and conclusion

Altruistic punishment can help sustain cooperation, but has the disadvantage that it can lead to feuds. Indeed, feuds are a common phenomenon of every day life with important economic implications. The complex nature of feuds means that evaluating the impact they

[24] If we run the same regression excluding *period* and its interaction with the treatment dummies, we find that there are no significant differences across treatments (p>.3).

[25] We note that while the differences in the initial level of earnings and the time trend between VCM and the other treatments roughly cancel out over the course of the experiment, the picture might look different if the experiment had continued over a substantially larger number of periods. This is not the case, however, for the comparison of B and F, as earnings in these treatments differ significantly neither initially nor over time.

	coeff	std. error	z	p
constant	20.715	2.331	8.89	<0.001
E(scalation)	-6.395	3.1766	-2.01	0.044
F(euds)	0.9833	3.2965	0.30	0.765
period	0.5791	0.1756	3.30	0.001
$E\times$ period	0.687	0.2393	2.87	0.004
$F\times$ period	-0.3206	0.2484	-1.29	0.194

Table 4: Linear regression for the individual earnings with group-level random effects. Treament B is the omitted category.

have on efficiency is extremely difficult to do using field data, if not impossible (Elster, 1990). To address this question we conducted a laboratory experiment based on a social dilemma. The controlled environment permits us to create counterfactuals and make inferences which are otherwise impossible.

The main experimental results are as follows. First, we find that the possibility of triggering a feud makes individuals almost completely unwilling to punish free riders when a feud can span multiple periods. The upshot is that cooperation rates quickly decline towards zero in most groups (although a couple of groups succeed at cooperating fully). Second, we find that individuals react strategically to institutional details. In the treatment in which individuals are given new identification numbers in each period, feuds cannot span multiple periods. As a result subjects can restrict (or completely avoid) retaliation by severely punishing their victims as this deprives them of the funds necessary to retaliate. This is exactly what they are found to do. As a consequence of these two results, we generally observe feuds neither within nor across periods. In our view, this is a positive result: it suggests that individuals aiming to enforce cooperation are likely to choose means that avoid escalation and feuds whenever possible such as ostracism or referring to a central authority. Finally, we find that despite the relatively low levels of cooperation in the "Feud" treatment, efficiency (as measured by earnings) is not any lower than in the other two treatments were peer punishment is available but extensive feuds across several periods are impossible. Earnings in the Feud treatment are also not significantly different to those in a treatment where punishments are not permitted.

One should be careful when drawing inferences from laboratory experiments. Forces that have been neutralized in an experiment might play a significant role outside the laboratory. Therefore, feuds might have a negative or a positive impact on efficiency in different environments. Our experiment should be considered as a first step towards understanding the efficacy of altruistic punishment in the face of risk of punishment escalation and feuds.

One reason efficiency is not lower in the Feuds treatment despite the low levels of cooperation is that we do not actually observe feuds. However, feuds do occur in everyday life. Why then do we not observe feuds in the experiment? One explanation might be that the symmetric nature of the game and the complete information make self-serving interpretations of the situation less likely. Asymmetries between players have been shown to be a source of tension and disagreement (Knez and Camerer, 1995).

A second explanation for the absence of feuds is that there is a common interest for all individuals to cooperate in our experiment. In certain cases where feuds are observed, parties have mutually exclusive interests. One common example of this is colleagues who compete for promotion. This can lead to feuds in the workplace, that have been frequently documented, as discussed in the introduction. Another topical example is terrorism, or more generally political violence. Generally two sides are involved and since one side (the terrorists) are making demands that are considered unacceptable by the other side (the government) such as secession of major parts of territory or a complete change of the fundamentals of government, attacks from one side cannot bring about cooperation, but only counter-attacks. There are numerous crucial differences between such a situation and the one studied in our experiment. The situation is asymmetric and there are no obvious gains from cooperation, so that a mutually agreeable situation is harder to achieve. Furthermore, being ideologically charged, the conflict more easily fuels itself. Moreover, acts of physical violence are likely to produce stronger emotional reactions than the pure monetary (and small-scale) exploitation in our experiment.

Another reason for not observing feuds might be the finite horizon employed in the experiment which limits the scope for strategic punishment.[26] Many interactions in naturally-occurring environments resemble infinitely-repeated games. In these situations, individuals might behave more aggressively and engage in feuding when the benefit of deterring future

[26] A finite horizon was chosen for two reasons. First, we wanted to keep our results comparable to similar public good experiments with punishment. Second, given that we already had an endogenous number of stages in each period, we did not want to further complicate matters and possibly confuse participants with a random stopping rule for determining the number of periods.

misbehavior exceeds the expected cost of a feud. Finally, the likelihood of a feud could well be affected by personal characteristics. For example, our student subjects may have been conditioned on solving conflicts in a calm, rational manner, whereas other people might be more impulsive and more likely to be guided by heuristics than rational calculation. Which of the discussed reasons prevents feuds from breaking out in our laboratory experiment should be the topic of a future study.

References

[1] Akerlof, G., Kranton, R.E., 2000. Economics and Identity. *Quarterly Journal of Economics* 140 (3), 715-753.

[2] 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.

[3] Andersson, L.M., Pearson, C.M., 1999. Tit for Tat? The Spiraling Effect of Incivility in the Workplace. *The Academy of Management Review* 24, 452-471.

[4] Axelrod, R.M., 1984. *The Evolution of Cooperation*. New York: Basic Books.

[5] Bensimon, H., 1994. Violence in the Workplace. *Training and Development* 48 (1), 27-32.

[6] Bjorkqvist, K., Osterman, K., Hjelt-Back, M., 1994. Aggresion Among University Employees. *Aggressive Behavior* 20, 173-184.

[7] Boehm, P., 1984. *Blood Revenge*. University of Pennsylvania Press.

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

[9] Bolton, G.E., Ockenfels, A., 2000. ERC - A Theory of Equity, Reciprocity and Competition. *American Economic Review* 90 (1), 166-193.

[10] Carpenter, J., 2007a. The Demand for Punishment. *Journal of Economic Behavior and Organization* 62 (4), 522-542.

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

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

[13] Dufwenberg, M., Kirchsteiger, G., 2004. A Theory of Sequential Reciprocity. *Games and Economics Behavior* 47, 268–298.

[14] Dunkel, T., 1994. Greatest Danger Zone: Your Office. *Working Woman* 70, 39-41.

[15] Elster, J., 1990. Norms of Revenge, *Ethics* 100 (4), 862-885.

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

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

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

[19] Fischbacher, U., 2007. Z-Tree: Zurich Toolbox for Readymade Economic Experiments. *Experimental Economics* 10, 171-178.

[20] Gächter, S., Herrmann, B., in press. Reciprocity, Culture, and Human Cooperation: Previous Insights and a New Cross-cultural Experiment. *Philosophical Transactions of the Royal Society B – Biological Sciences.*

[21] Gächter, S., Thoeni, C., 2005. Social Learning and Voluntary Cooperation among Like-minded People. *Journal of the European Economic Association* 3, 303-314.

[22] Gintis, H., Bowles, S., Boyd, R., Fehr, E., 2005. *Moral Sentiments and Material Interests.* Cambridge, MA: MIT Press

[23] Knez, M.J., Camerer, C.F., 1995. Outside Options and Social Comparison in Three-player Ultimatum Game Experiments. *Games and Economic Behavior* 10, 65-94.

[24] Ledyard, J., 1995. Public goods: A Survey of Experimental Research. In: Kagel, J., Roth, A. (eds.), *Handbook of Experimental Economics.* Princeton, NJ: Princeton University Press.

[25] 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.

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

[27] Nikiforakis, N., 2008. Punishment and Counter-punishment in Public Good Games: Can we Really Govern Ourselves? *Journal of Public Economics* 92, 91-112.

[28] Nikiforakis, N., Normann, H.T., 2008. A Comparative Statics Analysis of Punishment in Public-good Experiments. *Experimental Economics* 11, 358-369.

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

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

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

[32] Rabin, Matthew, 1993. Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83, 1281–1302.

[33] Rugala, E.A., Isaacs, A. 2004. *Workplace Violence: Issues in Response,* U.S. Department of Justice.

[34] Seabright, P., 2004. *The Company of Strangers.* Princeton University Press.

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

[36] Waldmann P., 2001. Revenge without Rules: On the Renaissance of an Archaic Motif of Violence. *Studies in Conflict & Terrorism* 24, 435-450.

Appendix: Group Data

Treatment	Group	Mean Contribution		Mean Earnings	
		All periods	Periods 6-10	All periods	Periods 6-10
B	1	18.08	19.00	28.15	29.60
B	2	10.53	11.45	23.99	25.97
B	3	10.53	9.15	19.12	20.84
B	4	7.23	4.90	21.34	21.14
B	5	19.33	20.00	29.05	32.00
B	6	5.33	3.75	21.77	21.95
B	average	11.83	11.38	23.90	25.25
E	1	10.88	13.10	13.78	14.81
E	2	10.18	8.75	24.46	24.80
E	3	13.03	13.30	20.47	27.98
E	4	17.65	18.50	21.59	23.75
E	5	11.53	13.60	24.52	28.16
E	6	15.68	17.05	27.83	28.88
E	7	0.80	0.00	16.36	20
E	average	11.39	12.04	21.28	24.05
F	1	3.38	2.15	20.23	19.49
F	2	3.43	0.60	13.66	20.06
F	3	4.63	1.95	20.53	19.52
F	4	7.80	3.90	23.93	21.89
F	5	17.20	17.85	28.90	28.31
F	6	19.53	20.00	31.50	32
F	average	9.33	7.74	23.12	23.56

Table 5: Data for individual groups.