AUSTRALIAN
SCHOOL OF BUSINESS™
THE UNIVERSITY OF NEW SOUTH WALES

The University of New South Wales
Australian School of Business

School of Economics Discussion Paper: 2011/10

Imperfect public monitoring with costly punishment – An experimental study

Attila Ambrus and Ben Greiner

School of Economics
Australian School of Business
UNSW Sydney NSW 2052 Australia
http://www.economics.unsw.edu.au

ISSN 1837-1035
ISBN 978-0-7334-3080-0

Imperfect public monitoring with costly punishment - An experimental study*

Attila Ambrus† and Ben Greiner‡

August 5, 2011

Abstract

This paper experimentally investigates the effects of a costly punishment option on cooperation and social welfare in long finitely repeated public good contribution games. In a perfect monitoring environment increasing the severity of the potential punishment monotonically increases both contributions and the average net payoffs of subjects. In a more realistic imperfect monitoring environment, we find a U-shaped relationship between the severity of punishment and average net payoffs. Access to a standard punishment technology in this setting significantly decreases net payoffs, even in the long run. Access to a very severe punishment technology leads to roughly the same payoffs as with no punishment option, as the benefits of increased cooperation offset the social costs of punishing.

Keywords: public good contribution experiments, imperfect monitoring, welfare implications of costly punishment

JEL Classification: C72, C92, H41

*We thank Drew Fudenberg, Jeffrey Miron, Andreas Nicklisch, and Ori Weisel for helpful comments and suggestions. Financial support through an Australian School of Business Research Grant is gratefully acknowledged.

†Department of Economics, Harvard University, Cambridge, MA 02138, e-mail: ambrus AT fas.harvard.edu

‡University of New South Wales, School of Economics, Sydney, NSW 2052, email: bgreiner AT unsw.edu.au

I Introduction

A large and growing experimental literature in economics, starting with Fehr and Gächter (2000), demonstrates that the possibility of costly punishment facilitates increased cooperation in finite-horizon social dilemma situations such as prisoner's dilemma and public good contribution games.[1] A recent paper by Gächter et al. (2008) shows that if the game horizon is long enough, the possibility of punishment also increases average net payoffs in the population.[2] That is, while in early periods of the game (roughly the first ten periods in the 50-period game investigated) the welfare-improving effect of increased cooperation is more than counter-balanced by the welfare-reducing effect of relatively frequent use of the punishment option, in the rest of the game a high level of cooperation is maintained with little explicit use of the punishment option. This result is consistent with group selection models of cooperation and punishment.[3]

In this paper we investigate how the option of costly punishment affects welfare in a more realistic environment, in which subjects observe each others' decisions with a small amount of noise. In particular, we investigate a public good contribution game in which after each contribution decision the public record of a player, that is the information on the subject's contribution announced publicly to all players, might differ from the true contribution of the subject: even if the subject contributed to the public good, with 10% probability the public record indicates no contribution. This design corresponds to partnership situations in which even if a member of the partnership contributes to a joint project, the others do not recognize the

[1] For the original references in social sciences, see Yamagishi (1986), Ostrom et al. (1992), and the theoretical contribution of Boyd and Richerson (1992). For empirical evidence for the relevance of costly punishment outside the lab, see Krueger and Mas (2004) and Mas (2008).

[2] An earlier string of papers (Fehr and Gächter, 2002; Gurerk et al., 2006; Herrmann et al., 2008; Egas and Riedl, 2008; and Dreber et al., 2008) shows that in repeated games with a shorter time horizon, the social costs of punishment tend to outweigh the benefits coming from increased cooperation. Rand et al. (2009) investigates the effect of access to punishment versus reward options in long (50-period) contribution games. For a theoretical investigation of the potential social costs and benefits of punishment, see Hwang and Bowles (2010).

[3] See Boyd et al. (2003), and Chapter 13 in Bowles (2003).

contribution, at least not until some later time. In our design such mistakes in the public record only influence the subjects' information, not their payoffs, which are determined by their true actions.[4]

Our design is in most parts similar to that of Gächter et al. (2008). In particular, we examine 50-period public good contribution games, and we adopt the same mapping between contributions and payoffs.[5] The only different aspect is that in our experiments subjects can only choose between contributing all or none of their endowments in each round. This was implemented in order to simplify the noise structure, with the intent that subjects understand better how their public records depend probabilistically on their decisions. Because of this change, we also ran a control design in which subjects observed each others' contributions perfectly. The other dimension in which we varied the design was the amount and effectiveness of costly punishment subjects could inflict on each other: we employed (i) a no punishment environment; (ii) a standard punishment technology that is used in Gächter et al. (2008), among other experimental papers, in which a subject can inflict a damage of 3 tokens for every token spent on punishment, and there is an upper limit on the amount of damage that could be inflicted; and (iii) a strong punishment technology, in which a subject can inflict a damage of 6 tokens for every token spent on punishment, and there was no upper limit on the amount of punishment. Hence, our experiments facilitated investigating the effects of increasing the severity of punishment in both perfect and imperfect monitoring environments.

We found that in the benchmark perfect monitoring condition increasing the severity of punishment increased both the amount of contributions and the average net payments (that is payments net the costs implied by

[4]The realized payoffs were revealed to subjects at the end of the experiment.

[5]As expressed in Gächter et al., there is an assertion in the experimental literature that play in long finitely repeated games, aside the last few periods, is similar to play in indefinitely repeated games with a large continuation probability. We are not aware of a formal test of this claim. Our results are relevant for infinite-horizon situations to the extent that the above assertion is adopted. In the real world there are both situations which are well approximated by an finite-horizon model (if there is a highlighted point of time after which the probability of continued interaction is very small), and ones which are better approximated by an infinite-horizon model.

imposed and received punishments) monotonically. This reinforces the findings of Nikiforakis and Normann (2008), the first paper in the literature that investigated the effects of varying the severity of punishment.[6] In the presence of either of the punishment options subjects learned to cooperate. In the strong punishment design this learning quickly led to almost full cooperation in the public good game, and virtually no use of the punishment option after a few initial periods.

In the imperfect monitoring environment the observed patterns are very different. The possibility of using the standard punishment option, while increasing contributions by a modest amount, significantly decreased average net earnings. Contribution levels stayed far away from full cooperation, and subjects kept on using the punishment option regularly throughout the whole game. In fact, average per period net earnings stabilized for the second half of the experiment, suggesting that the same qualitative conclusions would hold in even longer time horizons.

In contrast to standard punishment, the strong punishment option does increase average contributions significantly, even in the imperfect monitoring environment. However, the use of the punishment technology remains relatively frequent throughout the game. In our experiment these contrasting effects on the payoffs cancel each other out, and average net earnings with the strong punishment option are about the same as with no punishment option.

To summarize, in a noisy environment, it is not clear whether the costly punishment option is beneficial for society, even in the long run. Moreover, we find a U-shaped relationship between the severity of possible punishment and social welfare: the possibility of an intermediate level of punishment significantly decreases social welfare relative to when no punishment is available, while the possibility of severe punishment results in payoffs has a roughly zero net benefit for society.

A closer look at the data provides hints for why costly punishment is less effective in a noisy environment in establishing cooperation. First, subjects

[6] Nikiforakis and Normann investigated punishment effectiveness ratios 1:1, 1:2, 1:3, and 1:4 in 10-times repeated public good contribution games with perfect monitoring.

who were punished "unfairly", in the sense that the punishment followed a contribution by the subject, were less likely to contribute in the next round.[7] Such unfair punishment happens more often in the imperfect monitoring environment, following a wrong public record. The above effect gets curtailed in the design with strong punishment, but at the cost that when punishment occurs (and it does occur from time to time) then it inflicts heavy damage. Second, in the case of regular punishment, the positive effect of punishing non-contributors on their subsequent contributions is reduced. This suggests either that non-contributors do not believe that others will keep on punishing them for public records of not contributing in a noisy environment, or that they keep on not contributing because of the possibility that they get a wrong public record and get punished anyway even if they contribute.

Our paper complements findings in a number of recent papers. Bereby-Meyer and Roth (2006) show that players' ability to learn to cooperate in a repeated prisoner's dilemma game is substantially diminished when payoffs are noisy, even though in their experiment players could monitor each other's past actions perfectly.[8] In contrast, we find that a small noise in monitoring, albeit decreasing contributions in all conditions, does so significantly only in the strong punishment treatment. Abbink and Sadrieh (2009) find that if contributions are observed perfectly but there is noise in observing punishment then subjects punish each other more, reducing overall efficiency. Bornstein and Weisel (2010) and Patel et al. (2010), using different designs, show that the benefits of costly punishment are diminished when there is uncertainty regarding the realized endowment of subjects (but contributions are perfectly observed). Most closely related to our investigation is Grechenig et al. (2010), who in a work independent from ours also point out that in a noisy environment punishment can reduce welfare.

[7] This is consistent with the findings of Hopfensitz and Reuben (2009) in that punishment facilitates future cooperation, but only when it evokes shame and guilt, not when it evokes anger. The paper uses information on players' emotions captured through a questionnaire during the experiment. Herrmann et al. (2008) also find that (antisocial) punishment of contributors lowers their subsequent contributions.

[8] See also Gong et al. (2009) on repeated prisoner's dilemma games with stochastic payments, in a group versus individual decision-making context.

They do not investigate the effects of increasing the severity of punishment technology, which is the main focus of our paper, and instead examine the effects of varying the level of noise in observations. Furthermore, like all the above papers, Grechenig et al. focus on relatively short repeated games, in which the welfare benefits of costly punishment are ambiguous even without noise (see footnote 2).

We also contribute to the small but growing experimental literature on repeated games with imperfect public monitoring (Miller, 1996; Aoyagi and Fréchette, 2009; Fudenberg et al., 2010) although these papers investigate issues largely unrelated to ours.[9]

II Experimental Design

We implemented six treatments in a 3x2 factorial design. In the punishment dimension we varied between no, regular and strong punishment options, and in the noise dimension we employed either no noise in the information about other group members' contributions, or small noise. In our baseline experimental design, the instructions and procedures follow closely those of Gächter et al. (2008). Namely, experimental subjects participated in a 50-rounds repeated public good game. At the beginning, participants were randomly and anonymously matched to groups of three which stayed constant over all 50 rounds. In each round, each of the three participants in a group was endowed with 20 tokens and asked to either contribute all or none of these tokens to a group account.[10] If the amount was kept it benefitted the participant by 20 points, while if the amount was contributed it benefitted each of the three group members by $0.5 \times 20 = 10$ points.

After all group members made their choice simultaneously, they were informed about the outcome of the game. In the *no noise* conditions participants were informed about the choices in their group, while in the *noise*

[9] Earlier experimental papers that investigate manipulating players' information in repeated games in less standard ways (such as presenting information with delay, or in a cognitively more complex manner) include Kahn and Murnighan (1993), Cason and Khan (1999), Sainty (1999) and Bolton et al. (2005).

[10] This binary choice differs from Gächter et al. (2008), as we aimed to implement a simple noise structure.

treatments only a "public record" of each group member's choice was displayed. If a group member did not contribute, then the public record would always indicate "no contribution". If the group member contributed, there was a 10% chance that the public record showed "no contribution" rather than "contribution". Participants were fully informed about the structure of the noise.

In the *no punishment* conditions the round ended after that information was displayed, and the experiment continued with the next round. In the *punishment* conditions subjects participated in a second stage in each round. Here they were asked whether they would like to assign up to 5 deduction points to the other two members of their group.[11] Assigning deduction points did incur a cost to the punisher of one point per deduction point. In the *regular punishment* treatments each assigned deduction point implied a reduction of 3 points of the punished group member's income. However, the effect of received punishment was capped at the earnings from the public goods game, while a punisher always had to pay for assigned punishment points. Thus, participants could incur losses in a round only in the size of their own punishment to others. This punishment technology mimics the one used in Gächter et al. (2008) and many other public good experiments in the literature. In the *strong punishment* treatments, each assigned reduction point reduced the income of the punished group member by 6 points, and that income reduction was not capped, such that negative round incomes were allowed.[12]

The experimental sessions took place in February and March 2010 and 2011 at the ASB Experimental Research Laboratory at the University of New South Wales. Experimental subjects were recruited from the university student population using the online recruitment system ORSEE (Greiner 2004). Overall, 339 subjects participated in 12 sessions, between 24 and 30 per session. Upon arrival participants were seated in front of a computer

[11] Public records of the other two group members were always displayed anonymously in random ordering. Punishment choices were elicited on that same ordering, such that punishment could be dedicated, but reputation effects across rounds were excluded.

[12] However, the overall experiment income was capped at zero such that participants would go home with no less than their show-up fee of AU$ 5.

at desks which are separated by dividers. Participants received written instructions and could ask questions which were answered privately.The experiment started after participants completed a short comprehension test at the screen. The experiment was computerized and programmed in zTree (Fischbacher 2007). At the end of the experiment, participants filled in a short survey asking for demographics. They were then privately paid out their cumulated experiment earnings in cash (with a conversion rate of AU$ 0.02 per point) plus a AU$ 5 show-up fee and left the laboratory. Average earnings were AU$ 28.94, with a standard deviation of AU$ 5.31.

III Results

III.A Aggregate results

As groups stay constant over all 50 rounds, each group in our experiment constitutes one statistically independent observation. To test for treatment differences non-parametrically we apply 2-sided Wilcoxon rank-sum tests, using group averages as independent observations.

Table 1 lists the average contributions, punishments and net profits observed in our six treatments. Figures 1 and 2 display the evolution of public good contributions and net profits over time.

TABLE 1: Average contributions, punishment and net profits in treatments

	N participants	Avg. contribution	Avg. punishment	Avg. net profits
No noise				
No Punishment	57	5.59		22.80
Regular Punishment	57	12.40	0.64	23.66
Strong Punishment	54	17.61	0.48	25.45
Noise				
No Punishment	57	4.04		22.02
Regular Punishment	60	9.60	1.45	19.10
Strong Punishment	54	16.04	0.65	23.48

As Table 1 reveals, *noise* leads to lower contributions in all three punishment conditions. This is, however, only statistically significant for *strong punishment* ($p = 0.011$) and not significant for *no* and *regular punishment* ($p = 0.511$ and $p = 0.144$, respectively).

The effects of punishment on contributions are more significant. Contributions increase monotonically from *no punishment* over *regular punishment* to *strong punishment* both under no noise (p-values of 0.005, 0.030, and 0.001 for regular punishment vs. no punishment, strong punishment vs. regular punishment, and strong punishment vs. no punishment, respectively) and noise (p-values of 0.004, 0.001, and 0.001, respectively).

With respect to the average number of assigned punishment points, Table 1 seems to suggest that there are less punishment points assigned when their effect is more severe.[13] This, however, is only significant in the *noise* treatments ($p = 0.001$), while statistically no such effect can be established when there is *no noise* ($p = 0.385$). On the other hand, both *regular* and *strong punishment* are more likely when there is *noise* than if there is *no noise* ($p = 0.001$ and $p = 0.068$, respectively).

Finally, while *noise* does not have a measurable effect on profits when there is no punishment option available ($p = 0.511$), it (weakly) significantly decreases net profits (net of employed and received punishment) when punishment is available ($p = 0.024$ and $p = 0.069$ for *regular* and *strong punishment*, respectively). Along the punishment dimension, when there is *no noise*, only *strong punishment* has a significant positive effect on payoffs compared to the baseline with *no punishment* ($p = 0.035$), while the differences of *regular punishment* to both others are insignificant ($p = 0.737$ and $p = 0.352$ when compared to *no punishment* and *strong punishment*, respectively). If there is noise then the picture looks different: the *regular punishment* condition yields lower net profits than both the baseline and the *strong punishment* condition, though this effect is only significant for the latter ($p = 0.319$ and $p = 0.033$, respectively). The robustness of these

[13] This observation is closely related to the endogenously lower number of non-contributions. For a breakdown of punishment by reason see Table 3 and the discussion in Section III.B below.

results is confirmed by further tests applied to data from only the last 30 or last 20 rounds.

FIGURE 1: Average contributions over time

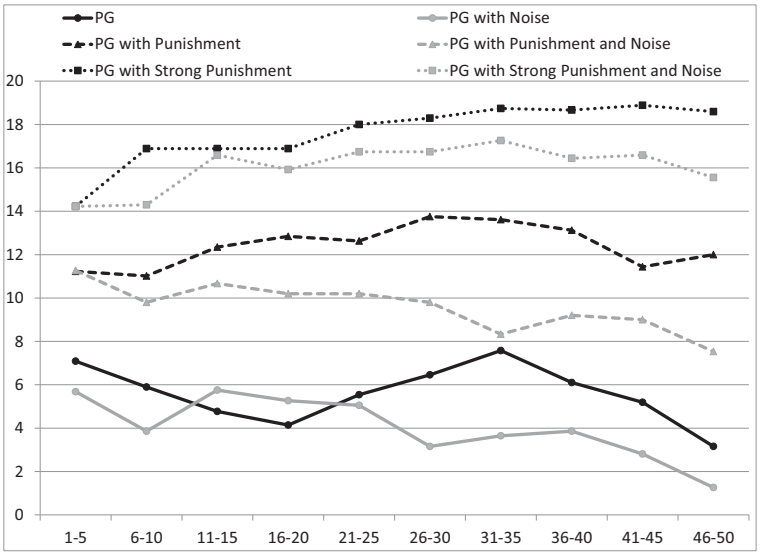
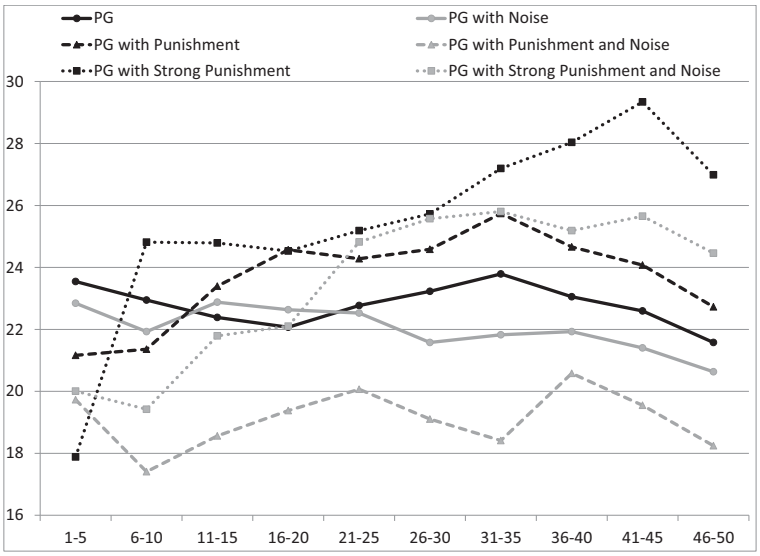


FIGURE 2: Average net profits over time



Figures 1 and 2 suggest that after some initial volatility, contributions and net profits in the different treatments tend to stabilize over time, aside from relatively small endgame effects in the very last periods (analogously to Gächter et al. 2008). This observation is corroborated by a battery of two-sided Wilcoxon matched-pairs signed-rank test comparing the average contributions and net profits in rounds 11 to 30 to rounds 31 to 50, which all yield p-values larger than 0.132, with the following exceptions: contributions increase over time with no noise and strong punishment ($p = 0.052$) and decrease with noise when there is no punishment or it is weak ($p = 0.016$ and $p = 0.011$, respectively), and net profits in the *no-punishment-noise* treatment were lower in later rounds ($p = 0.010$).

To complement the non-parametric analysis we ran Probit and ordinary least-square regressions controlling for interaction effects between our treatments. In particular, we regressed contributions, punishments and net earnings on the treatment dummies *Regular Punishment* and *Strong Punishment*, dummy *Noise* (being 1 in all noise treatments), and interaction effects of *Noise* with the two punishment dummies. All regressions also control for trends over time. As the groups of three participants are our units of statistically independent observations, we cluster standard errors on that level.

Table 2 lists the results from this analysis. We find a strong positive effect of punishment on contributions to the public good, which is even almost doubled if punishment is more severe. Noise, on the other hand, has no significant effect on how much participants contribute. The number of assigned punishment points is not significantly affected when punishment is more severe, but noise increases this number significantly, though less so when is punishment is strong. With respect to net earnings, punishment has a significant general positive effect only when it is strong. When noise is existent in addition to punishment, net payoffs are significantly reduced, but only under the regular punishment technology.[14] This leads to a U-shape of net earnings along the severity of punishment dimension under

[14]Hypothesis F-tests confirm at the 5%-level that the joint effect of *Weak Punishment* and *Noise*x*Weak Punishment* is negative, but cannot reject that the joint effect of *Strong Punishment* and *Noise*x*Strong Punishment* is different from zero (p=0.133).

TABLE 2: PROBIT/OLS ESTIMATIONS OF CONTRIBUTIONS, PUNISHMENTS AND NET EARNINGS BASED ON TREATMENT DUMMIES

Model	Probit	OLS	OLS
Dependent	Public Good	Assigned	Net
	Contribution	Punishment	earnings
Intercept		0.99***	21.54***
		[0.19]	[0.70]
Period	-0.001	-0.01**	0.05***
	[0.001]	[0.00]	[0.02]
Regular Punishment	0.332***		0.86
	[0.096]		[1.36]
Strong Punishment	0.576***	-0.16	2.65**
	[0.072]	[0.23]	[1.33]
Noise	-0.099	0.81***	-0.78
	[0.101]	[0.28]	[0.80]
Noise x Regular Punishment	-0.041		-3.77*
	[0.150]		[2.01]
Noise x Strong Punishment	-0.030	-0.64*	-1.19
	[0.169]	[0.34]	[1.65]
N	16950	11250	16950
Pseudo R-squared	0.195		
Adjusted R-squared		0.047	0.053

Note: For the Probit estimation on contributions, we report marginal effects rather than coefficients. For all estimations, robust standard errors are clustered at group level and given in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

noise: regular punishment has a negative effect on net earnings, but with strong punishment this negative effect is mitigated by the additional positive earnings effect in that condition.

III.B Punishment pattern

Table 3 displays the average number of received punishment points conditional on the published contribution of a subject. Obviously, punishment received following a public record of no contribution is considerably

TABLE 3: Average punishment points spent, conditional on receiver's contribution and public record

	All rounds		Only first round	
	Punish-ment	Strong Pnmt	Punish-ment	Strong Pnmt
No noise				
After contribution decision was				
Contribution	0.212	0.316	0.114	0.771
Defect	1.338	1.681	1.636	3.000
Noise				
After public record was				
Contribution	0.411	0.262	0.742	0.583
Defect	2.236	1.666	1.414	1.444

Note: Punishment points are not multiplied with factor 3 or 6, yet.

higher than otherwise.[15] However, even for cooperators punishment levels are greater than zero. This might root in anti-social punishment (defectors punishing contributors, see also Herrmann et al., 2008), or could be an effect of some subjects also punishing for older offenses. With regular punishment we observe higher punishment levels under noise (but significantly so only for punishment towards contributors, $p = 0.030$), while punishment levels are unaffected by noise or even slightly less when punishment is strong ($p = 0.331$ and $p = 0.033$ for punishment after contribution and defection records, respectively).

Comparing regular to strong punishment we observe that punishment towards contributors is not affected by the punishment technology, neither with nor without noise, and neither in terms of assigned or (multiplied) received punishment points (all p-values larger than 0.266). With respect to defectors, however, the number of received (multiplied) punishment points, the eventually resulting income reduction, is larger if punishment is more severe, both without and with noise ($p = 0.001$ and $p = 0.027$, respectively), while the number of assigned points is only different if there is no noise

[15] This is strongly significant in all four punishment treatments, with all p-values smaller than 0.006. These and the following tests are based on the corresponding averages on the independent group level.

($p = 0.015$, vs. $p = 0.827$ with noise). As a result, a stronger punishment technology leads to a larger discrimination between contributors and defectors: while the former attract (not significantly) less punishment points, the latter are punished even harsher.

All these described effects are already existent when only looking at the very first round of the game (see the right part of Table 3), and statistically significantly so except for the differences between regular and strong punishment. Since in the first round subjects cannot punish for older offenses, this provides clearer evidence that a public record of not contributing in a given round attracts more punishment points in the same round than a public record of contributing. Contributors do receive some punishment even in the first round though, indicating the existence of purely antisocial punishment.

We employ Probit regression analysis to analyze reactions to received punishment and other previous experiences. In Model 1 of Table 4, we estimate the current round's contribution of a participant based on the number of punishment points she received in the last round ($RecPnmt_{LR}$, not yet multiplied with the punishment factor). We control for the last round's contribution of this participant ($Contr_{LR}$), and interact with treatment dummies on whether noise was present (Noise), whether the strong punishment technology was present (StrPnmt), or both (Noise x StrPnmt).

Due to the binary nature of contribution decisions, contributors can only fix or reduce their contribution, while non-contributors' contributions can only stay the same or increase. The large and significant effect of the $Contr_{LR}$ dummy indicates the general differences in trends between participants who contributed before or not. Our main interest, however, lies in the interactions. We find that for non-contributors, the higher the received punishment, the more likely they are to contribute in the next round. This effect is significantly increased when the punishment has a stronger impact. When, on the other hand, contributors get punished, then they are likely to decrease their contribution in the next round, and more so the higher the punishment. The punishment technology effect discussed above now works in the other direction, softening this discouraging effect when punishment is strong. In both cases, noise does not seem to play a role.

TABLE 4: Probit estimations of current contribution based on last round behavior

	Model 1	Model 2	Model 3
$RecPnmt_{LR}$	0.041***	0.022**	0.012
	[0.011]	[0.011]	[0.008]
$RecPnmt_{LR}$ x Noise	-0.009		0.004
	[0.013]		[0.010]
$RecPnmt_{LR}$ x StrPnmt	0.039**	0.0571***	0.023**
	[0.016]	[0.018]	[0.012]
$RecPnmt_{LR}$ x Noise x StrPnmt	0.012		-0.003
	[0.023]		[0.015]
$Contr_{LR}$	0.794***	0.705***	0.535***
	[0.025]	[0.037]	[0.046]
$Contr_{LR}$ x $RecPnmt_{LR}$	-0.144***	-0.124***	-0.073***
	[0.040]	[0.023]	[0.024]
$Contr_{LR}$ x $RecPnmt_{LR}$ x Noise	0.046		0.024
	[0.041]		[0.026]
$Contr_{LR}$ x $RecPnmt_{LR}$ x StrPnmt	0.076*	0.025	0.051*
	[0.043]	[0.026]	[0.027]
$Contr_{LR}$ x $RecPnmt_{LR}$ x Noise x StrPnmt	-0.070		-0.031
	[0.046]		[0.030]
$Contr_{LR}$ x $PRwrong_{LR}$		0.016	
		[0.057]	
$Contr_{LR}$ x $PRwrong_{LR}$ x $RecPnmt_{LR}$		0.085***	
		[0.021]	
$Contr_{LR}$ x $PRwrong_{LR}$ x $RecPnmt_{LR}$ x StrPnmt		-0.021	
		[0.034]	
$OtherContr_{LR}$			0.386***
			[0.051]
$Contr_{LR}$ x $OtherContr_{LR}$			0.121*
			[0.065]
N	11025	5586	11025
Pseudo R-squared	0.454	0.353	0.535

Note: We report marginal effects rather than coefficients. Robust standard errors, clustered at group level, are given in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively. $Contr_{LR}$ and $RecPnmt_{LR}$ refer to contribution and punishment received in the last round, respectively, while $PRwrong_{LR}$ indicates whether the public record of a contributor in the last round was wrong, and $OtherContr_{LR}$ represents the average contribution (scaled [0,1]) of the other two group members in the last round. Noise and StrPnmt are dummies indicating whether noise or the strong punishment technology were present.

The Probit Model 2 reported in Table 4 concentrates on choices under Noise, and explores whether having been a contributor with a *wrong* public record in the last round ($PRwrong_{LR}$) has an effect on how that participants reacts to being punished by her group members. While in the new model any other effects are robust, the lack of significance for $Contr_{LR}$x$PRwrong_{LR}$ suggests that having had a wrong public record does not influence contributions by itself, the significant positive effect on the interaction term with the received punishment indicates that those contributors are less likely to reduce their contribution when being punished, and similar so in both punishment regimes. Nevertheless, the net effect of increasing the punishment of a subject with a wrong public record on the next period contribution of this subject is still negative.

Finally, Model 3 includes the average contribution of the other two group members ($OtherContr_{LR}$, scaled to [0,1]) as a control into the estimation equation of Model 1. We find that current contributions are indeed highly correlated with the other group members' last contributions (more for previous contributors). This might be interpreted as an alternative type of punishment by reducing future payoffs (though such punishment cannot be targeted towards an individual), or as evidence for coordination on and convergence to a group norm. The inclusion of these controls reduces the positive effect of punishment on subsequent contributions of non-contributors, but the effect remains significantly positive in the strong punishment treatment. The negative effects of punishment on contributors subsequent choice are robust against including the controls. These results, however, have to be interpreted with care due to multicollinearity, as the relation between own and others' contributions in the last round ($Contr_{LR}$ and $OtherContr_{LR}$) is highly correlated with the subsequently received punishment ($RecPnmt_{LR}$).

III.C Evolution of cooperation and punishment in groups

In Figures 3 and 4 we classify the groups in the different treatments by whether there was full, partial, or no contribution to the public good in different periods, and study the emergence of such groups over time. Figure 4

additionally includes the pattern of punishment over time for groups which started and ended with full public good contributions, groups which started low but converged to full contributions after some time, and groups which did not manage to reach full contributions.

FIGURE 3: No punishment treatments - Group cooperation over time

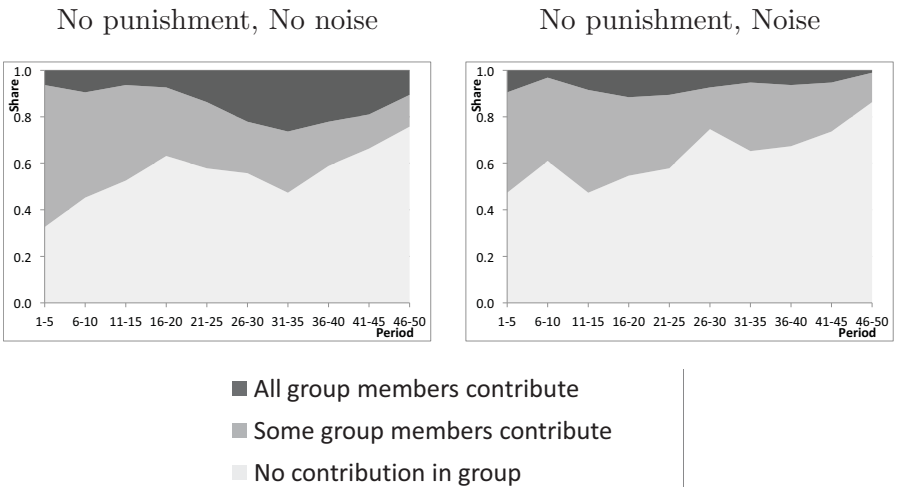
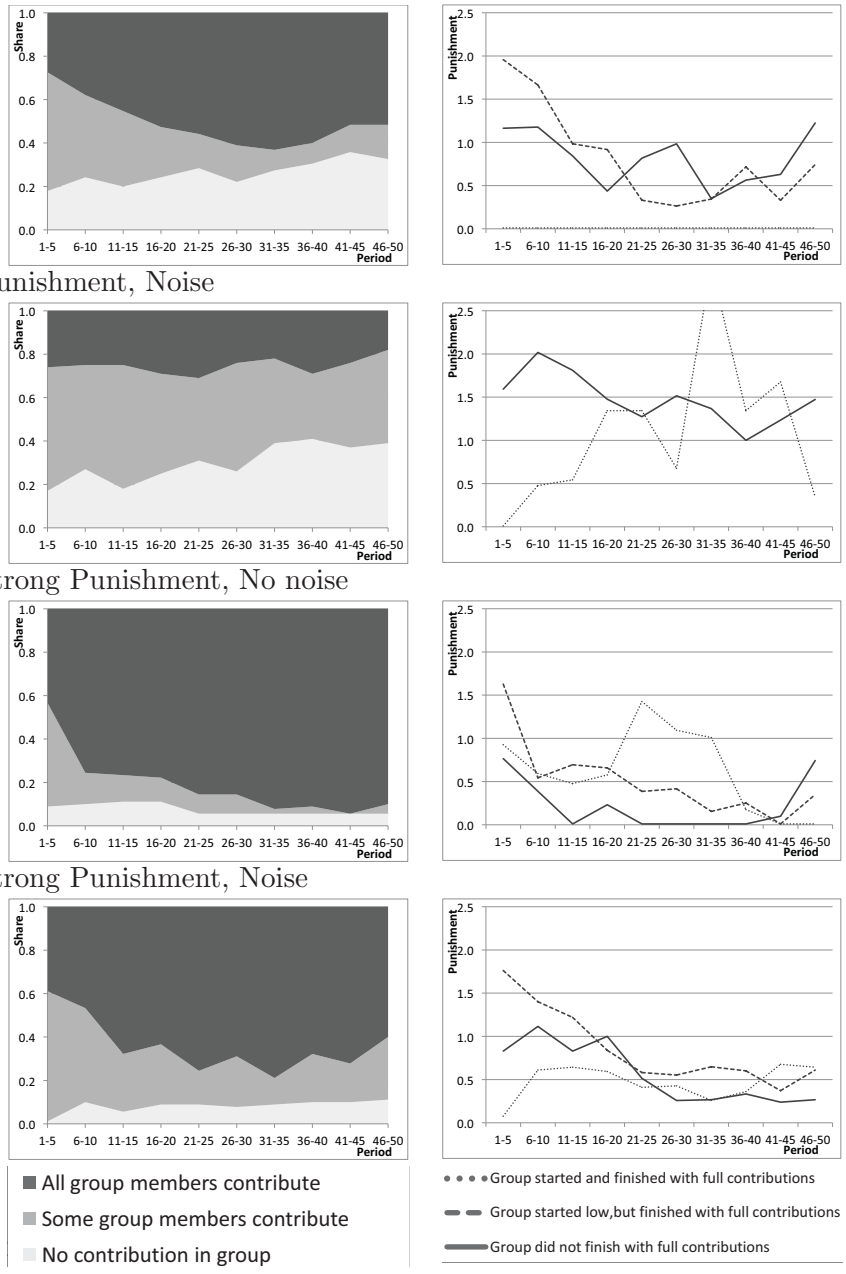


Figure 3 shows that when there is no punishment available, groups who started out with at least some contributions become no-contribution groups over time. As we observe on the left side of Figure 4, under regular punishment and if there is no noise, most groups polarize such that either all or none of the group members contribute. When we add noise to the information about others' contributions, we observe higher dispersion of contributions within groups, such that there is no convergence to polarized groups, but some consistent increase in the number of no-cooperation groups. Under a severe punishment regime, groups quickly converge to homogenous full-contribution groups. This general tendency stays intact with noise in the public information.

FIGURE 4: Punishment treatments - Group cooperation over time and average punishment in different cooperation classes



We statistically confirm these observations with a battery of Fisher Exact tests comparing the shares of different groups in the very first and the very last round of a treatment. We find that when there is no punishment available, then the share of groups who at least partly contribute shrinks and the share of groups with no contributions at all increases over time, both with and without noise (all p-values smaller than 0.01). On the contrary, under the strong punishment regime, the share of partly contributing groups decreases, too, but is accompanied by a significant increase in full contribution groups (all p-values smaller than 0.05), again no matter whether noise is existent or not. With regular punishment, however, we observe a significant decrease in the share of partly contributing groups when there is no noise ($p = 0.045$), and we find a weakly significant increase in the share of groups who do not contribute at all when there is noise ($p = 0.082$), in both cases with no significant effect on the individual shares of other two group types.

When comparing statistically between noise and no noise treatments, we do not find any significant differences in the first round of all punishment regimes (all p-values larger than 0.25), and differences for the last round only for regular punishment: the share of groups with full contributions in the last round is significantly lower ($p = 0.003$) when there is noise than when there is no noise. When comparing between the punishment conditions, we find that treatments do not start out with different distributions of group types, except that under strong punishment there are less no-contribution groups in the first round than without punishment options ($p = 0.020$ and $p = 0.008$ for no noise and noise, respectively). For the last round, however, we find significant differences in the distribution of group types across punishment conditions. When there is no noise, then we have a monotone increase of the last-round share of full-contribution groups and a monotone decrease of the share of no-contribution going from no punishment to regular punishment and strong punishment (all p-values smaller than 0.01). Comparing the end of the treatments with noise, this pattern only holds true for strong punishment (all p-values smaller than 0.05), while regular punishment now features more partial-contribution groups ($p = 0.014$) rather than full-contribution groups ($p = 1.000$), compared to no punishment.

The right side of Figure 4 displays average punishment in different classes of groups. If there is no noise, then groups which start with full contributions and end with full contributions experience no punishment at all during the game. While we do not observe such groups under noise and regular punishment, we observe some but low punishment in such groups under noise and a strong punishment regime (potentially indicating successful but costly coordination on cooperation).

IV Conclusion

This paper finds that while in a perfect monitoring public good contribution environment increasing the severity of a costly punishment option unambiguously increases average net payoffs, in an imperfect monitoring environment the above relationship is nonmonotonic. Moreover, at least for some punishment technologies, the presence of costly punishment can be detrimental for society. This weakens the case that group selection evolutionary procedures lead to emotional responses like anger and revenge, inducing individuals to punish cheaters.

A possible direction for future research is reexamining the questions addressed in this paper using data from real world environments in which dissatisfied participants can punish each other, such as feedback scores in electronic commerce, or grades and teacher evaluations in higher education.

References

Abbink, K. and A. Sadrieh (2009): "The pleasure of being nasty," Economics Letters, 105, 306-308.

Aoyagi, M. and G. Fréchette (2009): "Collusion as public monitoring becomes noisy: Experimental evidence," Journal of Economic Theory, 144, 1135-1165.

Bereby-Meyer, Y. and A. Roth (2006): "The speed of learning in noisy games: Partial reinforcement and the sustainability of cooperation," American Economic Review, 96, 1029-1042.

Bolton, G., E. Katok and A. Ockenfels (2005): "Cooperation among strangers with limited information about reputation," Journal of Public Economics, 89, 1457-1468.

Bornstein, G. and O. Weisel (2010): "Punishment, Cooperation, and Cheater Detection in 'Noisy' Social Exchange," Games, 1(1), 18-33.

Bowles, S. (2003): "Microeconomics: Behavior, institutions, and evolution," Princeton University Press, Princeton NJ.

Boyd, R., H. Gintis, S. Bowles and P. Richerson (2003): "The evolution of altruistic punishment," Proceedings of the National Academy of Sciences (USA), 100, 3531-3535.

Boyd, R. and P. Richerson (1992): "Punishment allows the evolution of cooperation (or anything else) in sizable groups," Ethology and Sociobiology, 13, 171-195.

Cason, T. and F. Khan (1999): "A laboratory study of voluntary public good provision with imperfect monitoring and communication," Journal of Development Economics, 58, 533-552.

Dreber, A., D. Rand, D. Fudenberg and M. Nowak (2008): "Winners don't punish," Nature, 452, 348-351.

Egas, M. and A. Riedl (2008): "The economics of altruistic punishment and the maintenance of cooperation," Proceedings of the Royal Society, 275, 871-878.

Fehr, E. and S. Gächter (2000): "Cooperation and punishment in public goods experiments," American Economic Review, 90, 980-994.

Fehr, E. and S. Gächter (2002): "Altruistic punishment in humans," Nature, 415, 137-140.

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

Fudenberg, D., D. Rand and A. Dreber (2010): "Turning the other cheek: Leniency and forgiveness in an uncertain world," mimeo Harvard University.

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

Gong, M., J. Baron and H. Kunreuther (2009): "Group cooperation under uncertainty," Journal of Risk and Uncertainty, 39, 251-270.

Grechenig, K., A. Nicklisch and C. Thöni (2010): "Punishment Despite Reasonable Doubt Public Goods Experiment with Sanctions under Uncertainty," Journal of Empirical Legal Studies, 7(4), 847-867.

Greiner, B. (2004): "An Online Recruitment System for Economic Experiments," in: Kurt Kremer, Volker Macho (eds.): Forschung und wissenschaftliches Rechnen 2003. GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung, 79-93.

Gürerk, O., B. Irlenbusch and B. Rockenbach (2006): "The competitive advantage of sanctioning institutions," Science, 312, 108.

Herrmann, B., C. Thöni and S. Gächter (2008): "Antisocial punishment across societies," Science, 319, 1362-1367.

Hopfensitz, A. and E. Reuben (2009): "The importance of emotions for the effectiveness of social punishment," Economic Journal, 119, 1534-1559.

Hwang, S. and S. Bowles (2010): "Is altruism bad for cooperation?," working paper Santa Fe Institute.

Kahn, L. and J. Murnighan (1993): "Conjecture, uncertainty, and cooperation in prisoners' dilemma games: Some experimental evidence," Journal of Economic Behavior and Organization, 22, 91-117.

Krueger, A. and A. Mas (2004): "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires," Journal of Political Economy, 112, 253-289.

Mas, A. (2008): "Labor Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market," Review of Economic Studies, 75, 229-258.

Miller, J. (1996): "The evolution of automata in the repeated prisoner's dilemma," Journal of Economic Behavior and Organization, 29, 87-112.

Nikiforakis, N. and H. Normann (2008): "A comparative statics analysis of punishment in public-good experiments," Experimental Economics, 11, 358-369.

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

Patel, A., E. Cartwright and M. van Vugt (2010): "Punishment cannot sustain cooperation in a public good game with free-rider anonimity," mimeo University of Gothenburg.

Rand, D., A. Dreber, T. Ellingsen and M. Nowak (2009): "Positive interactions promote public cooperation," Science, 325, 1272-1275.

Sainty, B. (1999): "Achieving greater cooperation in a noisy prisoner's dilemma: an experimental investigation," Journal of Economic Behavior and Organization, 39, 421-435.

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