Imperfect public monitoring with costly punishment - An experimental study*

Attila Ambrus† and Ben Greiner‡

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 exactly offset the social costs of punishing. Our findings parallel findings in the empirical literature on gun control in that more severe weapons in criminal acts and in self-defense are used less frequently, as their intimidating factor is often sufficient in preventing offenses.

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

JEL Classification: C72, C92, H41

*We thank Drew Fudenberg, Jeffrey Miron, 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) also 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 contribution, at least not until some later time. In our design such mis-

[1] For the original references in social sciences, see Ostrom et al. (1992) and 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. 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 Bowles (2003).

takes 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 costly punishment option; (ii) a standard punishment technology that is used in Gächter et al. (2008), among other experimental papers (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 imposed and received punishments) monotonically. In the presence of either of the

[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.

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 almost exactly 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. Subjects who were punished "unfairly", in the sense that the punishment followed a contribution by the subject, were less likely to contribute the in next round and more likely to engage in antisocial punishment in subsequent rounds.[6]

[6]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.

Such unfair punishments periodically happen in the imperfect monitoring environment, resulting in relatively low levels of contributions. 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.

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. 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) 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 that in a noisy environment punishment can reduce welfare. 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.[7]

Lastly, our paper provides an experimental counterpart to some empirical findings in the economics literature on gun control. In a survey article,

[7]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).

Cook et al. (2002) report that gun robbers are far less likely to attack and injure their victims than robbers using other weapons, and argues that "the most plausible explanation for this pattern of outcomes is simply that a gun gives the assailant the power to intimidate and gain his victim's compliance without use of force, whereas with less lethal weapons the assailant is more likely to find it necessary to back up the threat with a physical attack." They also point out that the intimidating power of a gun can be beneficial in self-defense, too: according to a study of NCVS data, in burglaries of occupied dwellings only 5 percent of victims who used guns in self-defense were injured, compared to 25 percent of those resisting with other weapons. The counterpart of all these findings is that the type of weapon matters a lot in determining the level of injury if it is used, with guns inflicting much higher fatality rates than other weapons.

Along similar lines, Lott and Mustard (1997) argue that states that liberalized their concealed-carry regulations experienced reductions in violent-crime rates, presumably because would-be assailants were deterred by the increased likelihood that their victims would be armed. Subsequent research, however, has raised concern about the conclusions of Lott and Mustard, pointing to selection issues and data accuracy problems.[8] Dills et al. (2010) conclude that currently available data does not provide obvious support either for the claim that right-to-carry laws increased or decreased violent crime.

Given the above data concerns, which prohibit drawing causal inferences from the empirical papers, our work complements this literature by providing similar findings in a controlled experiment (albeit in a stylized laboratory context). Namely, we show that a severe punishment option, which inflicts heavy damage on the punished, is used less frequently than the regular punishment option, and that its intimidating power is more effective in establishing cooperative behavior.[9]

[8] For a list of critiques of the Lott and Mustard analysis, as well as responses to these critiques, see footnote 21 in Dills et al. (2010).

[9] The repeated public good contribution game with costly punishment option that we examine in our experiments can be viewed as a stylized model of partnership in which partners can retaliate each other for perceived offenses. Presumably, our treatments in-

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* treatments only a "public record" of all group members' choices 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

troducing imperfect monitoring into this game serve as more realistic approximations of reality. As Cook et al. (2002) state: "It is quite possible that most 'self-defense' uses occur in circumstances that are normatively ambiguous: chronic violence within a marriage, gang fights, robberies of drug dealers, encounters of young men who simply appear threatening."

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

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 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, 165 subjects participated in 6 sessions. Upon arrival participants were seated in front of a computer 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$ 29.31, with a standard deviation of AU$ 5.25.

[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.

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	27	5.85		22.93
Regular Punishment	30	12.32	0.53	24.02
Strong Punishment	24	18.53	0.26	27.43
Noise				
No Punishment	30	5.72		22.86
Regular Punishment	30	8.57	1.69	17.61
Strong Punishment	24	16.27	0.75	22.88

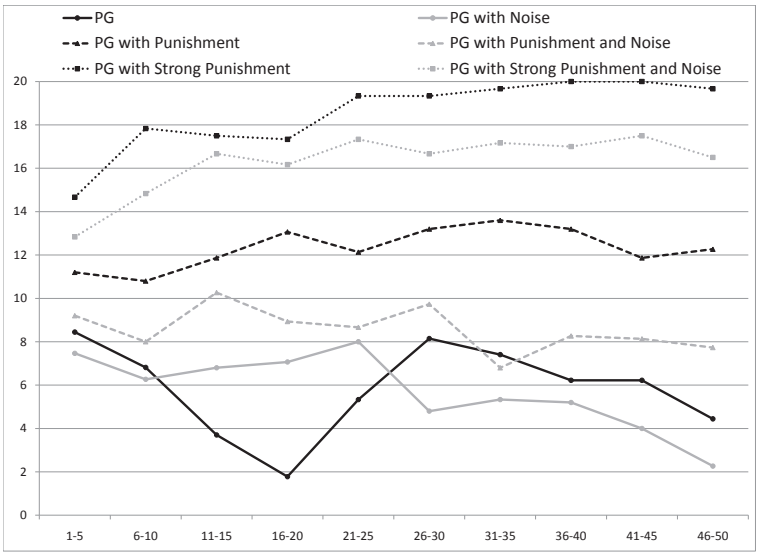
As Table 1 reveals, *noise* leads to lower contributions in all three punishment conditions. This is, however, not significant for *no* and *regular punishment* ($p = 0.775$ and $p = 0.173$, respectively), and only marginally significant for *strong punishment* ($p = 0.092$).

The effects of punishment on contributions are more significant. Without noise both *regular* and *strong punishment* yield significantly increased contributions ($p = 0.045$ and $p = 0.001$, respectively), while there is no difference between those two conditions ($p = 0.211$). With noise, however, *regular punishment* looses its positive effect on contributions compared to *no punishment* ($p = 0.450$), while *strong punishment* still yields sizable ef-

fects ($p = 0.003$ and $p = 0.013$ compared to *no* and *regular punishment*, respectively).

With respect to the average number of assigned punishment points, Table 1 seems to suggest that there are less punishment points assigned when its effect is more severe. This, however, is only significant in the *noise* treatments ($p = 0.029$), while no such effect can be established when there is *no noise* ($p = 0.655$). On the other hand, both *regular* and *strong punishment* are more likely when there is *noise* than if there is *no noise* ($p = 0.007$ and $p = 0.016$, respectively).

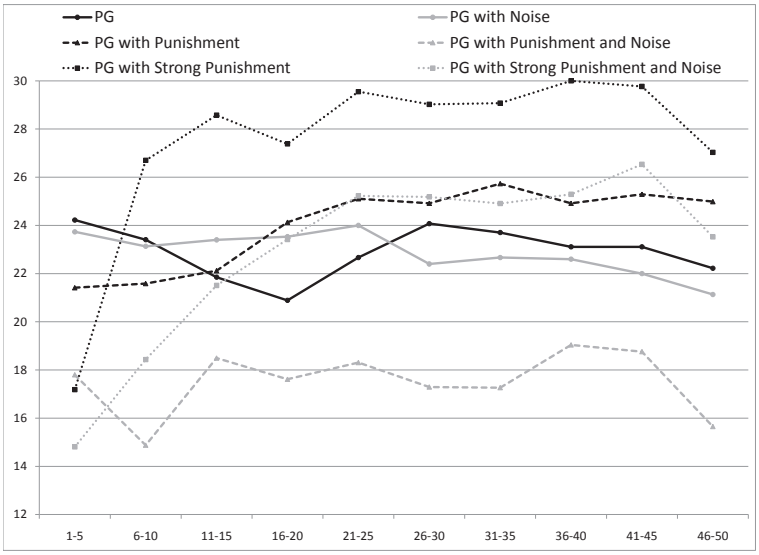
FIGURE 1: AVERAGE CONTRIBUTIONS OVER TIME



Finally, while *noise* does not have a measurable effect on profits when there is no punishment option available($p = 0.775$), it significantly decreases net profits (net of employed and received punishment) when punishment is available ($p = 0.075$ and $p = 0.028$ 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* ($p = 0.008$), while the differences of *regular punishment* to both others are insignificant ($p = 0.902$ and $p = 0.246$ when

compared to *no punishment* and *strong punishment*, respectively). If there is noise then the picture looks different: the *no punishment* condition and the *strong punishment* condition are virtually statistically indistinguishable ($p = 0.902$), while *regular punishment* option yields lower net profits than both the baseline and the *strong punishment* treatment, however, weakly significantly so only when compared to *strong punishment* ($p = 0.083$, vs. $p = 0.190$ when compared to baseline).

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 stabilized in later periods. 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 (all p-values larger than 0.14, with the exception of net profits in the *no-punishment-noise* treatment where a p-value of 0.097 indicates weakly significantly lower net profits in later rounds).

To complement the non-parametric analysis we ran ordinary least-square regressions controlling for interaction effects between our treatments. In

TABLE 2: OLS Regressions of contributions, punishments and net earnings on treatment dummies

Dependent	Public Good Contribution	Assigned Punishment	Net earnings
Intercept	5.75***	0.92***	21.48***
	[1.65]	[0.25]	[0.96]
Period	0.00	-0.02***	0.06**
	[0.02]	[0.00]	[0.02]
Punishment	6.47**		1.1
	[2.99]		[1.74]
Punishment x is strong	6.21**	-0.27	3.41**
	[2.50]	[0.22]	[1.62]
Noise	-0.13		-0.07
	[2.43]		[1.22]
Noise x Punishment	-3.61	1.16***	-6.35**
	[3.92]	[0.42]	[2.90]
Noise x Punishment x is strong	1.48	-0.67	1.87
	[3.74]	[0.46]	[3.04]
N	8250	5400	8250
Adjusted R-squared	0.232	0.089	0.113

Note: Standard errors, clustered at group level, are given in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

particular, we regressed contributions, punishments and net earnings on the dummies *Punishment* (taking the value of 1 if an punishment option was available), *Punishment* x *is strong* (being 1 if additionally that punishment is severe), *Noise* (being 1 in all noise treatments), and interaction effects of *Noise* with the two punishment dummies. As such, the dummies identify the marginal effects between treatments. 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 doubled if punishment is made 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, with no moderation of this effect through different punishment conditions. 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. This leads to a U-shape of net earnings along the severity of punishment dimension under 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: AVERAGE RECEIVED PUNISHMENT POINTS, CONDITIONAL ON CONTRIBUTION AND PUBLIC RECORD

	Punishment	Strong Punishment
No noise		
After contribution decision was		
Contribution	0.237	0.093
Defect	1.012	2.409
Noise		
After public record was		
Contribution	0.460	0.351
Defect	2.472	1.882

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

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 higher than otherwise.[13] However, even for cooperators punishment levels are greater than zero. This might root in anti-social punishment (defectors pun-

[13]This is, however, only strongly significant in the strong punishment treatments, with Wilcoxon matched-pairs signed-rank tests' p=0.016 and p=0.016 under no noise and noise, respectively. The corresponding p-values for regular punishment are p=0.078 and p=0.106. For these and the following tests we use the corresponding averages on the independent group level.

ishing contributors, see also Herrmann et al., 2008), or could be an effect of some subjects also punishing for older offenses. In particular, the higher punishments under noise (both after contribution and defection records)[14] indicate that some subjects exercise retribution for being 'unjustly' punished beforehand. Comparing regular to strong punishment we observe that with noise, the number of attributed punishment points is lower under strong punishment than with the regular punishment technology. However, this decrease does not outweigh the doubled impact of strong punishment, such that the eventually resulting income reduction is larger on average if punishment is more severe.[15] Without noise, on the other hand, 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 (Wilcoxon rank-sum test, p=0.073).

TABLE 4: Expected change in contribution in next round, conditional on innocence and punishment

	No noise			Noise		
	Baseline	Punish-ment	Strong Pnmt	Baseline	Punish-ment	Strong Pnmt
Contributed, PR=contributed						
No punishment	-0.254	-0.047	-0.020	-0.326	-0.166	-0.058
Pnmt > 0		-0.189	-0.091		-0.506	-0.112
Per pnmt point		-0.145	-0.068		-0.214	-0.051
Contributed, PR=not contributed						
No punishment				-0.159	-0.130	0.000
Pnmt > 0					-0.231	-0.040
Per pnmt point					-0.115	-0.013
Not contributed						
No punishment	0.096	0.042	0.156	0.114	0.122	0.100
Pnmt > 0		0.301	0.491		0.177	0.530
Per pnmt point		0.091	0.159		0.058	0.249

Note: Contribution is defined as a binary 0/1 choice variable. Punishment points are not multiplied with factor 3 or 6, yet.

[14] These differences between noise and no noise are only significant for punishment towards contributors, with two-sided Wilcoxon rank-sum test p-values of 0.033 and 0.027 for treatments with regular and strong punishment, respectively.

[15] These observed differences between regular and strong punishment are not significant, neither before nor after multiplication with the punishment severity factor (two-sided Wilcoxon rank-sum tests, all p-values larger than 0.625).

Table 4 calculates changes in contribution decisions of contributors, non-contributors, and falsely labeled non-contributors after receiving punishment or not. Note, however, that 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. Thus, our analysis has to concentrate on the *differences* in changes after punishment compared to no punishment.

When a group member did not contribute (lower part of Table 4), then punishment (compared to no punishment) increased the probability that the person would contribute in the next round. This effect seems to be amplified with a stronger punishment technology. However, under noise, the total effect is somewhat smaller with regular punishment, such that the additional effect of having strong punishment is increased.

Contributors also change their behavior when getting (anti-socially) punished. In all treatments, the likelihood that they contribute again in the next round is lower when they get punished than when they don't get punished. Interestingly, contributors even react averse to punishment when they get punished due to a false public record in the noise treatments. Thus, punishment can also have negative effects on contributions if it affects cooperative people.

Note, however, that due to a low number of observations in each cell when calculating group averages (e.g., some groups would always punish no contribution or never punish contributions, such that for those there are no matching observations in the corresponding cells), none of the battery of applied two-sided Wilcoxon matched-pairs signed-rank tests yielded any statistically significant differences between changes in contributions after punishment compared to no punishment.

As an alternative way of analyzing reactions to punishment, we regress the current round's contribution of a participant 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).

TABLE 5: Regression of current contribution on last round's contribution and received punishment

	Coeff.	StdErr
Intercept	0.097***	[0.026]
$RecPnmt_{LR}$	0.045**	[0.018]
$RecPnmt_{LR}$ x Noise	-0.022	[0.020]
$RecPnmt_{LR}$ x StrPnmt	0.049*	[0.028]
$RecPnmt_{LR}$ x Noise x StrPnmt	0.021	[0.037]
$Contr_{LR}$	0.837***	[0.030]
$Contr_{LR}$ x $RecPnmt_{LR}$	-0.086*	[0.045]
$Contr_{LR}$ x $RecPnmt_{LR}$ x Noise	-0.004	[0.051]
$Contr_{LR}$ x $RecPnmt_{LR}$ x StrPnmt	-0.009	[0.048]
$Contr_{LR}$ x $RecPnmt_{LR}$ x Noise x StrPnmt	-0.002	[0.056]
N	5292	
Adjusted R-squared	0.601	

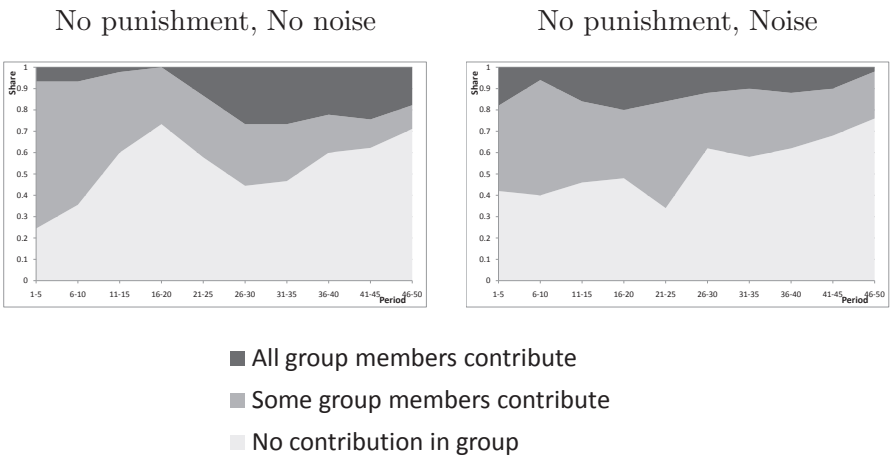
Note: Standard errors, clustered at group level, are given in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

The results of this analysis are presented in Table 5. The Intercept and the coefficient on the $Contr_{LR}$ dummy indicate 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 (weakly) significantly increased when the punishment has a stronger impact. Noise seems not to play a role for these reactions. 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. Whether the punishment is strong or not does not play a role here, and neither does whether noise is present or not.

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, 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. Unfortunately, the number of groups (i.e. independent observations) in each of these category are too low to enable us to corroborate the following observations with statistical tests.

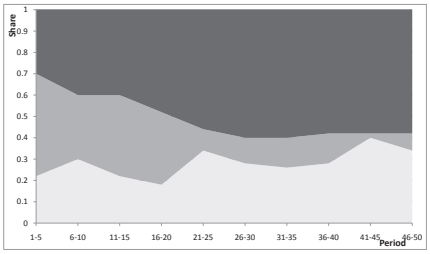
FIGURE 3: No punishment treatments - Group cooperation over time

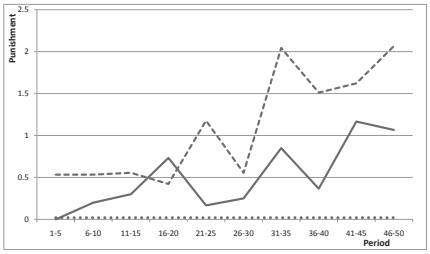


As we observe on the left side of Figure 4, under regular punishment and if there is no noise, most groups polarize such that either all or none of the group members contribute. The majority of groups in this condition become full-contribution groups over time. When we add noise to the information about others contributions, we observe higher dispersion of contributions within groups, such that there is no convergence to polarized groups. Under a severe punishment regime, groups quickly converge to homogenous full-contribution groups. This general tendency stays intact with noise in the

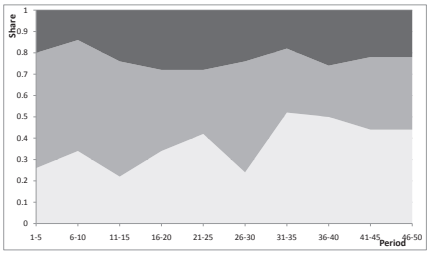
FIGURE 4: Punishment treatments - Group cooperation over time and average punishment in different cooperation classes

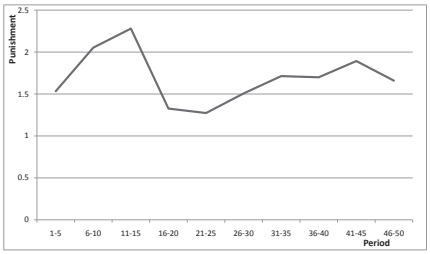
Punishment, No noise



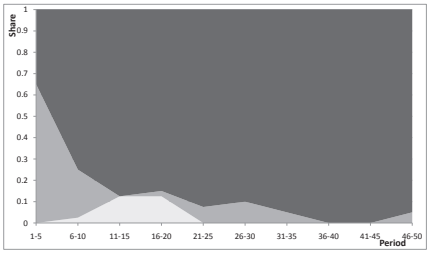


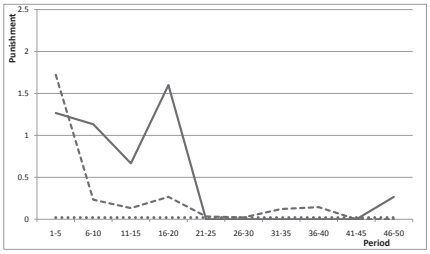
Punishment, Noise



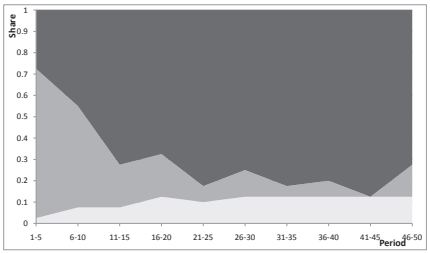


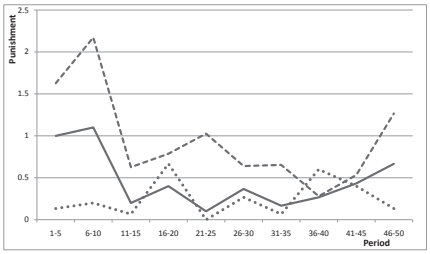
Strong Punishment, No noise





Strong Punishment, Noise





All group members contribute

Some group members contribute

No contribution in group

Group started and finished with full contributions

Group started low,but finished with full contributions

Group did not finish with full contributions

public information, despite a 10% (i.e. one group) which stabilizes on no contributions.

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 disciplination of group members who deviate from the initial path of cooperation).

IV Discussion and Conclusions

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. On the other hand, our nonmonotonicity result suggests that if a society cannot ban the use of all weapons, it might be worse off when banning only serious weapons.

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.

Cook, P., M. Moore and A. Braga (2002): "Gun control," IN Crime: Public policies for crime control, Wilson, J. and J. Petersilia eds., Institute for Contemporary Studies Press, Oakland, CA.

Dills, A., J. Miron and G. Summers (2010): "What do economists know about crime?," IN: The economics of crime: Lessons for and from Latin America; Di Tella, R., S. Edwards and E. Schargodsky (ed.s), University of Chicago Press.

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.

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

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.

Lott, J. and D. Mustard (1997): "Crime, deterrence, and the right-to-carryconcealed handguns," Journal of Legal Studies 26, 1-68.

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.

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.

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