Democratic punishment in public good games with perfect and imperfect observability*

Attila Ambrus† and Ben Greiner‡

May 11, 2018

Abstract

In the context of repeated public good contribution games, we experimentally investigate the impact of democratic punishment, when members of a group decide by majority voting whether to inflict punishment on another member, relative to individual peer-to-peer punishment. Democratic punishment leads to more cooperation and higher average payoffs, both under perfect and imperfect monitoring of contributions. A control treatment with random dictator punishment verifies that this effect primarily works by curbing anti-social punishment and thereby establishing a closer connection between a member's contribution decision and whether subsequently being punished by others.

Keywords: public good contribution experiments, punishment, voting

JEL Classification: C72, C92, H41

*Financial support through an Australian Research Council Discovery Grant (DP150104169) is gratefully acknowledged. We thank Justin Cheong and Johannes Hoelzemann for excellent research assistance, and an anonymous reviewer, the editor, as well as Marco Casari, Nikos Nikiforakis, and Louis Putterman for helpful comments.

†Duke University, Department of Economics, Durham, NC 27708, e-mail: aa231 AT duke.edu

‡Wirtschaftsuniversität Wien, Institute for Markets and Strategy, Welthandelsplatz 1, 1020 Vienna, Austria, e-mail: bgreiner AT wu.ac.at

I Introduction

Several papers in the experimental literature, starting from Fehr and Gächter (2000), demonstrated that the availability of a costly punishment option for individuals can increase cooperation in public good contribution games. Gächter, Renner and Sefton (2008) showed that this increases overall net payoffs in the population, provided that the time horizon for interaction is long enough. However, Grechenig, Nicklisch and Thöni (2010) and Ambrus and Greiner (2012) found that the above results hinge on the assumption that individuals can perfectly monitor each others' actions. If there is a small amount of noise in monitoring, then the availability of costly individual punishment does not help the participants' welfare, and it can even decrease it. The reason is that with imperfect monitoring from time to time a contributor gets punished by fellow team members who received an incorrect negative signal regarding the contribution. This discourages future contributions and can trigger antisocial punishment by the contributor who was "unfairly" punished.[1] Hence even in the long run, contribution levels stay away from the socially efficient levels, and individuals keep on punishing each other, further decreasing each others' payoffs. Moreover, in a recent paper Fischer, Grechenig and Meier (2013) find that if monitoring is imperfect, centralizing punishment, in the form of delegating punishment rights to a particular individual, does not remedy the issues above, and cooperation levels remain low.[2]

In this paper we find that democratic punishment, in the form of group members after each round of the contribution game deciding which members to punish using simple majority rule, outperforms individual punishment, both in terms of

[1] In experiments on social dilemma games with imperfect observability and no direct punishment option available, Aoyagi and Fréchette (2009) and Fudenberg, Rand and Dreber (2012) find that players under noise are more forgiving than without noise. On the prevalence of anti-social punishment in public good contribution games with individual punishment, see Cinyabuguma, Page and Putterman (2006), Herrmann, Thöni and Gächter (2008), and Ertan, Page and Putterman (2009). The latter use the term "perverse punishment", that refers to punishment of above-average contributors regardless of the punisher. Hauser, Nowak and Rand (2014) provide a theoretical analysis in the context of a dynamic learning model, explaining why punishment might not promote cooperation when anti-social punishment is possible. Nikiforakis (2008) shows that the possibility of subsequent (anti-social) counter-punishment neutralizes the positive effects of the existence of the peer-punishment institution. Kamei and Putterman (2015), however, provide evidence that this negative effect is mitigated when there is very detailed information available about individual contribution and punishment choices.

[2] In a recent paper Rand, Fudenberg and Dreber (2015) find evidence that it is the inability to directly observe each other's contribution intentions which leads to deterioration of cooperation in environments with imperfect monitoring.

cooperation levels and average payoffs, and in both perfect and imperfect monitoring environments. A key reason is that democratic punishment mitigates anti-social punishment, and makes the relationship between one's contribution decision and whether she gets subsequently punished clearer: Specifically, it makes it more likely that contributing members do not get punished, and that non-contributing members get punished. In particular it greatly reduces the opportunities of those who get punished by others for not contributing to punish back, either preemptively or subsequently. This is because such retaliatory punishments are almost never supported by a majority of members, and for this reason the institution of majority voting suppresses such attempts.

However, we find evidence that individual punishing intentions also differ between the two types of punishment environments. A group member is more likely to vote to punish a non-contributor in a democratic procedure than deciding to punish in the individual punishment environment. Interestingly, this holds not only for contributing members, but also for non-contributing ones (when deciding whether to punish fellow non-contributors). We also find evidence that individuals react differently, with respect to subsequent contributions to the public good, when they are punished democratically by group members versus when they get punished individually by fellow members. In both cases getting punished after not contributing increases expected contribution in the next round. The difference is that when an individual gets punished even though he contributed, this punishment discourages her to contribute in the next round in the individual punishment treatment, but not in the democratic punishment treatment. These findings are in line with the finding in several papers ((Frey, 1994; Frey, Benz and Stutzer, 2004; Pommerehne and Weck-Hannemann, 1996)), that there is a positive relationship between direct-democratic participation rights and pro-social behavior.[3]

Our experimental design involves groups of five subjects, playing twenty times repeated public good contribution games. In the individual punishment treatment, after each round each group member decides independently which other members to punish. In the democratic punishment treatment, after each round members simultaneously cast votes which members should be punished, and punishment is inflicted on those members who received at least three votes. In order to put the

[3]See also Dal Bó, Foster and Putterman (2010), who show that endogenous democratic adoption of a policy that automatically fines unilateral non-contributors increases cooperation relative to when the same policy is imposed on the group exogenously.

two punishment schemes on an equal footing, we set payoffs in the democratic punishment treatment such that if the group votes to punish a member, the punishment inflicted is the same as if all four other members decided to punish the member in the individual punishment treatment. Similarly, the cost of a group punishment for each of the other members is the same as the cost of punishing in the individual punishment treatment.

From the perspective of an expected-utility decision-maker (who conditions her choice on being pivotal), this design implies two possible channels for differences in behavior between individual and democratic punishment. First, the democratic voting procedure may have a filtering effect on individual choices (only punishments supported by a majority will be implemented), but second a *pivotal* group member's punishment effect is larger while individual costs are kept constant, compared to an individual in the individual punishment treatment. In order to disentangle these effects, we also implemented a control treatment, dictator punishment. In this treatment after each round of contributions, for each possible punishment recipient a random group member gets selected, and this group member's punishment choice applies to every other member of the group. Thus, in this treatment, the individual cost, the cost imposed on others, and the amount of punishment are exactly the same for a group member in the dictator punishment treatment (conditional on being selected) as in the democratic punishment treatment (conditional on being the pivotal group member). We find that at the aggregate, dictator punishment does not lead to different contribution, punishment, or profit levels than individual punishment. Compared to democratic punishment, punishment levels with dictator punishment are significantly higher, and net profits are significantly lower. As opposed to democratic punishment, dictator punishment does not curb anti-social punishment.[4]

In our setting, members of a group cannot commit ex-ante to a particular punishment rule, instead in each round a majority decides ad-hoc on whether to punish someone or not. There are several papers in the literature taking a different approach, in which there is a democratic group decision at the beginning of the game, deciding on whether to adopt a punishment scheme (either the option of individual

[4] The results from the control treatment also indicate that increased punishment of noncontributors in the democratic punishment treatment relative to individual punishment is not because of dispersion of responsibility, as would be posited by the identifiability theory of group shifts (Wallach, Kogan and Bem, 1962, 1964). With dictator punishment the punishing player is clearly identified, still punishment levels remain high.

punishment or an automated punishment rule) and in some cases on features of the punishment scheme (how severe punishment is allowed to be, or who can be punished): see Andreoni and Gee (2012); Dal Bó et al. (2010); Ertan et al. (2009); Kamei, Putterman and Tyran (2015); Markussen, Putterman and Tyran (2014); Sutter, Haigner and Kocher (2010); Tyran and Feld (2006). Other studies allow the punishment to be delegated to a specific subject, who carries them out without commitment: see for example Baldassarri and Grossman (2011); Fehr and Fischbacher (2004); Kamei et al. (2015); Leibbrandt and López-Pérez (2011, 2012).[5] More related to our investigation are Decker et al. (2003), Cinyabuguma et al. (2006), Casari and Luini (2009) and Van Miltenburg, Buskens, Barrera and Raub (2014). Decker et al. (2003) investigate collective punishment schemes in which each of three group members make proposals on how much punishment to impose on a fourth member, and either the minimum, or the median, or the maximum proposal is carried out, as well as an individual punishment scheme. They find no clear differences between the above schemes in terms of average payoffs. Cinyabuguma et al. (2006) study a setting in which after each round group members can vote whether to expel certain members of the group, and show that the threat of expulsion can facilitate more cooperation. Casari and Luini (2009) show that in a repeated public good contribution game with punishment, it increases average payoffs if only coalitions of at least two members can inflict punishment (with the whole cost of punishment being borne by members of the coalition). In contrast to these results, Van Miltenburg et al. (2014), not in a repeated games context but in a setting in which partners are randomly rematched after each round, find that group voting on whether to punish certain group members underperforms individual punishment, in terms of the level of cooperation achieved. All of the above papers only consider settings with perfect monitoring, as opposed to our study, which tests the robustness of democracy effects in an imperfect monitoring environment.

II Experimental Design

We implemented six treatments in a 3×2 factorial design. Our main comparison is between a repeated 5-person public good game that allows for *individual punishment* and a public good game in which a majority of group member votes is required in order to punish another group member (*democratic punishment*). We

[5]For a related theoretical analysis, see Aldeshev and Zanarone (2014).

employ both games in two different environments, one with perfect observation of other group members' contributions, and one in which the signal about other group member's contribution is noisy, such that there is a small chance of 10 percent that a contribution is displayed to others as a defection.[6]

In addition to the two punishment stage designs, we tested a further design (with and without noise) with *dictator punishment*. In this condition, each member states whether she would like to punish other members, and for each punishment receiver one of the other group members is randomly selected and the punishment of this group member is implemented on behalf of this *and* the three remaining members. In the *dictator* treatment punishment is still individual (and not filtered through a democratic majority requirement), but – under the traditional assumption of an expected-utility decision-maker – the individual incentives for the punishing group member are the same in this treatment (conditional on being dictator) as in the *democratic punishment* treatment (conditional on being the pivotal voter).[7] Hence the *dictator punishment* condition allows us to distinguish whether differences between *individual punishment* and *democratic punishment* are due to the effect of majority rule or due to the effect of increased punishment effectivity from the perspective of the decisive member.

At the beginning of the experiment, participants were matched to groups of five, that stayed constant for all 20 rounds. Within each group, participants were assigned IDs from 1 to 5. Group IDs also stayed constant for the course of the experiment, such that group members were identifiable. Each round consisted of 2 stages, a public good contribution stage and a punishment stage. In the public good contribution stage, each group member was endowed with 50 points, and decided whether she wanted to contribute these 50 points to a "project" or not. If the endowment was kept, it increased the participant's payoff by 50 points. If the

[6] The same design of imperfect monitoring was used in Ambrus and Greiner, 2012. Markussen, Putterman and Tyran (2016) find that such Type I errors have a similar effect in undermining cooperation as Type II errors (where a non-contribution may be shown as a contribution), but that when given a choice, experimental subjects dislike Type I errors more than Type II errors.

[7] When deciding what choice to make, an expected utility maximizer only considers the conditioning events when her choices are payoff relevant. In the democratic punishment treatment the relevant conditioning event for an individual is when she is pivotal. In the dictator punishment treatment the relevant conditioning event is when being selected as a dictator. We designed these treatments such that conditional on the relevant conditioning event, the costs and induced punishments are the same between these treatments. Empirically, the (theoretically irrelevant) likelihood to become a dictator punisher was 20%, while the likelihood to be pivotal was 10.3% and 3.5% in the democratic punishment treatment with and without noise, respectively.

endowment was contributed, it benefitted each of the five group members by 0.3 times 50 = 15 points. Thus, if no group member contributed, each would earn 50 points, while the symmetric efficient outcome of 75 points for each could be reached if all contributed their endowment.

Our treatments differ only in the second stage of each round. First, after their simultaneous decisions in Stage 1, participants were informed about the contribution of each group member in their group. In our *No noise* treatments, the actual contribution of the respective participant was displayed. In the *Noise* treatments, the display showed a "public record" of each group member's contribution. Participants were informed that if a group member did not contribute his endowment, then the public record would always indicate "no contribution". If the group member contributed, however, then there was a 10 percent chance that the public record showed "no contribution" rather than "contribution". The same public record of a member was displayed to all other group members.

Second, participants were asked to indicate their willingness to monetarily punish ("reduce the earnings of") each other group member. In our *Individual Punishment* treatments, each group member could directly reduce the earnings of another group member by 15 points, at a cost of 5 points. In the *Democratic punishment* condition, group members simultaneously cast votes for each group member whether to punish that group member or not. Thus, for each group member, votes from all four other group members were collected. If three or more group members voted to punish a participant, then the earnings of that participant were reduced by 60 points, and each of the other four group members (independent of how they voted) incurred a cost of 5 points for this punishment. If no majority was reached (because two or less group members voted for punishment), then no points are reduced and no costs incurred. Thus, the equivalent of a punishment by a group (when majority is reached) in the *Democratic Punishment* treatments is being punished by each other group member in the *Individual Punishment* treatments, and the equivalent of no group punishment (because there was no majority to punish) is not being punished at all in the *Individual Punishment* treatments. In the *Dictator Punishment* treatments, each group member was asked to make a punishment decision as under *Individual Punishment*. After all group members made their choices, for each punishment recipient the computer randomly selected one of the other group members, and that group member's punishment decision was implemented for all remaining group members as well. Thus, the individual monetary incentives (conditional on being selected / being the marginal voter) are the same in the *Democratic Punishment* and the *Dictator Punishment* treatments (a choice of punishment implies a reduction of 60 points at a cost of 5 points), while (in expectation) the relation between individual punishment decisions and implemented punishment decisions are the same across *Individual Punishment* and *Dictator Punishment*, since in the latter treatment implemented choices are a random draw of individual choices.

After all participants simultaneously made their punishment decisions, they were informed about the punishments and votes in their group, and the consequences for their round payoffs. In the *Noise* treatments, any payoff information was provisional based on public records; participants were informed about their true earnings in each round at the end of the experiment.

Twelve experimental sessions, on *Individual Punishment* and *Democratic Punishment*, took place in March and April 2014 at the Business School Experimental Research Laboratory at the University of New South Wales. Six further sessions, on *Dictator Punishment*, were conducted in the same laboratory in June 2016. Experimental subjects were recruited from the university student population using the online recruitment system ORSEE (Greiner, 2015). Overall, 480 subjects participated in 18 sessions, with either 20, 25, or 30 subjects per session. Upon arrival participants were seated in front of a computer at desks which were separated by dividers. Participants received written instructions and could ask questions which were answered privately. The experiment was programmed in zTree (Fischbacher, 2007). Sessions lasted about one hour. At the end of the experiment, participants filled out a short demographic survey. They were then privately paid their cumulated experimental earnings in cash (with a conversion rate of AU$ 0.02 per point) plus a AU$ 5 show-up fee. Participants could incur losses in a particular round, but session losses were capped at the show-up fee. No participant incurred losses over the whole session. The average earning was AU$ 26.59 (including showup-fee), with a standard deviation of AU$ 4.81, a minimum payoff of AU$ 7.70 and a maximum payoff of AU$ 35.30.

III Results

III.A Aggregate results

In our analysis, we will first focus on our main comparison between individual and democratic punishment. Then, in order to highlight the channels through which the described effects work, we will discuss how the results from the dictator punishment treatment relate to the individual and democratic punishment treatments. Tables and figures refer to data from all treatments.

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

	N part.	N groups	Avg. contr.	Avg. punishm.	Avg. net profits
No noise					
Individual punishment	75	15	23.33	5.96	53.72
Democratic punishment	80	16	36.75	2.40	65.18
Dictator punishment	70	14	27.57	10.63	49.61
Noise					
Individual punishment	80	16	18.78	6.36	50.92
Democratic punishment	90	18	27.58	4.37	57.97
Dictator punishment	85	17	21.03	10.94	45.93

TABLE 2: P-values from Non-parametric Wilcoxon Ranksum tests across treatment dimensions

	Contributions	Received Punishment	Net profits
Individual Punishment vs. Democratic punishment			
with No noise	0.012**	0.005***	0.000***
with Noise	0.055*	0.137	0.003***
Individual Punishment vs. Dictator punishment			
with No noise	0.631	0.028**	0.359
with Noise	0.627	0.121	0.177
Dictator Punishment vs. Democratic punishment			
with No noise	0.050*	0.000***	0.001***
with Noise	0.137	0.009***	0.000***
No noise vs. Noise			
with Individual punishment	0.489	0.874	0.385
with Democratic punishment	0.023**	0.030**	0.005***
with Dictator punishment	0.275	0.937	0.634

Table 1 lists the average contributions, punishments, and net profits observed in our six treatments. Figures 1, 2 and 3 display the evolution of public good contributions, punishment, and net profits over time. As groups stay constant over all 20 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 2 reports the results.

In both the perfect monitoring and the noisy environment, we observe higher contributions, less punishment (only significant for the *No Noise* condition), and consequently higher net profits when groups vote over punishment compared to when group members can punish individually. Introducing *Noise* in the observation of other group members' contribution behavior lowers contributions and net profits, and increases observed punishment for both when punishment is individual as well as when punishment is a group decision, but statistically significantly so only for the latter environment.[8,9]

The regressions reported in Table 3 confirm and further detail these results. We estimate the likelihood of contribution (Model 1), the amount of punishment points received in a round (Models 2-5), as well as the net profits in a round (Model 6) using treatment dummies and a *Round* control. The dummy *Noise* equals 1 in the Noise treatments and 0 otherwise; the dummy *Democratic Punishment* is 1 for the treatments with voting over punishment and 0 in the individual punishment treatments; and the interaction effect *Noise*×*Democratic punishment* equals 1 only in the respective treatment with democratic punishment under noise. Similarly, the dummy *Dictator Punishment* equals 1 in dictator punishment treatments and 0 otherwise, and the interaction effect *Noise*×*Dictator punishment* captures the additional effect of noise in these conditions. For each estimation we ran additional post-estimation F-tests in order to determine the total effect of *Noise* under demo-

[8] Ambrus and Greiner (2012) only study an individual punishment environment and find a significant effect of noise on all three observables. However, in Ambrus and Greiner (2012) the game was repeated 50 times (while only 20 times here) and featured smaller (3-person) groups.

[9] Figure 1 suggests different time trends in contributions across our treatments, at least until just before the end. When running Probit models of contribution decision on the only explanatory variable *Round* (excluding Round 20), we reject the Null hypothesis of no time trend for the *Individual Punishment* treatments at the 5% significance level (p-values of average marginal effect of *Round* $p = 0.032$ and $p < 0.001$, respectively), while we cannot reject a zero time trend for *Democratic Punishment* at this level ($p = 0.352$ and $p = 0.100$, respectively). Under *Dictator Punishment* we observe a negative time-trend with noise ($p = 0.003$) but not without noise ($p = 0.242$).

cratic and dictator punishment (Noise + N×DemP, Noise + N×DicP) and the total effect of *Democratic/Dictator punishment* under noise (DemP + N×DemP, DicP + N×DicP).

FIGURE 1: Average contributions over time

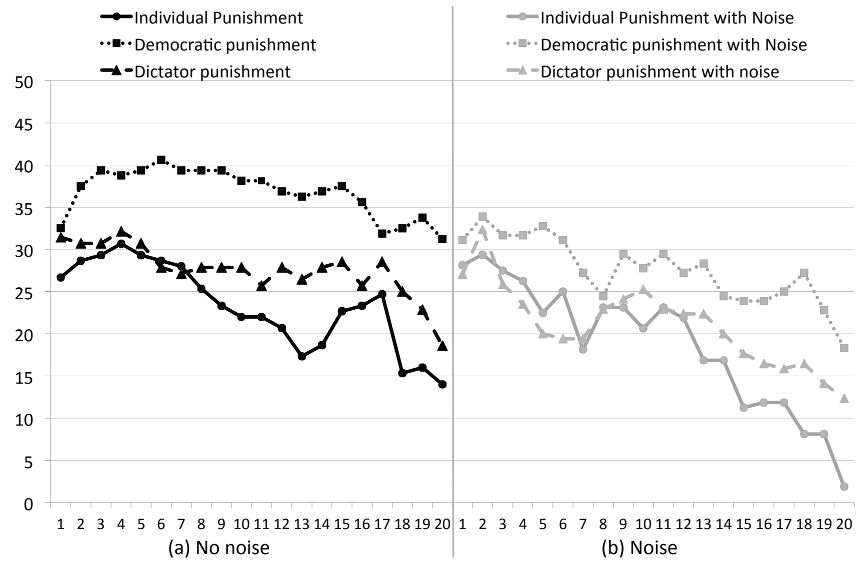


FIGURE 2: Average received punishment over time

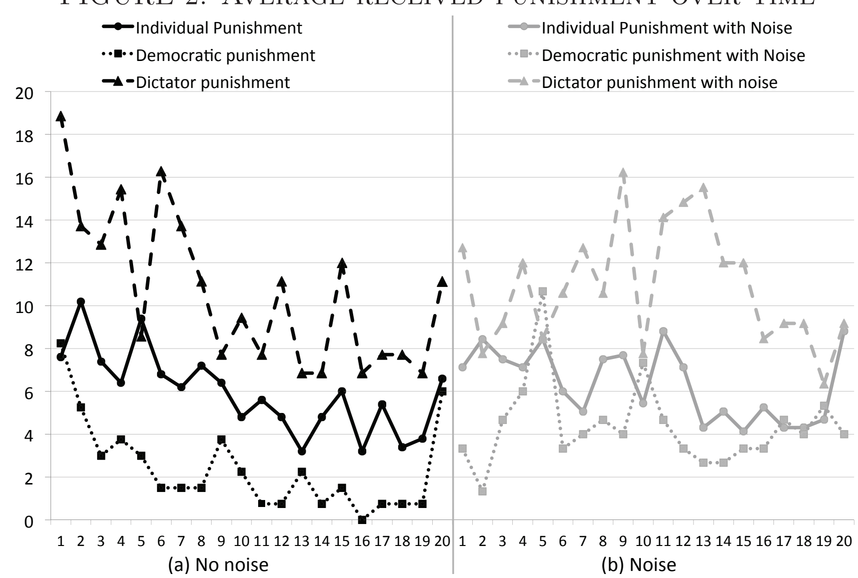
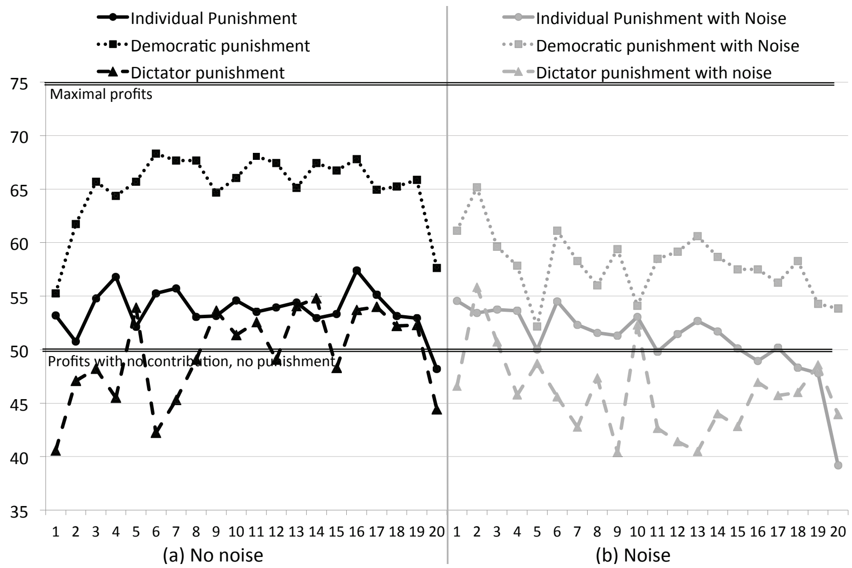


FIGURE 3: Average net profits over time



The results of Model 1 and 6 in Table 3 replicate the non-parametric tests, in that we observe a significant increase in contributions and net profits when the group votes to punish compared to individual punishment (both when there is perfect and imperfect monitoring), and that noise has a statistically significant detrimental effect on contributions and net profits only in the democratic punishment condition.

The Models 2 to 5 in Table 3 explore effects of treatment conditions on punishment behavior. Model 2 predicts all punishments (independent of towards whom they were directed), and shows that introducing democratic punishment significantly reduces overall punishment in both non-noisy and noisy environments. *Noise* increases punishments when groups punish but not when individuals punish, which is related to the observation that democratic voting seems to be less effective in reducing punishments under noise than when there is no noise. Models 3 and 4 regress punishment of defectors (as identified by their public record) and cooperators, respectively. The results show that democratic punishment leads to a significant decrease of punishment of cooperators in both environments, but to a decrease of punishment of defectors only in the noise environment and only weakly significantly. Model 5 serves the purpose of showing that due to the relatively low likelihood of “noise” in public records, the punishment patterns towards “true cooperators” (some of which might have a wrong public record of no cooperation) are very similar to those towards the subset of cooperators who are clearly identified as such by their public record.

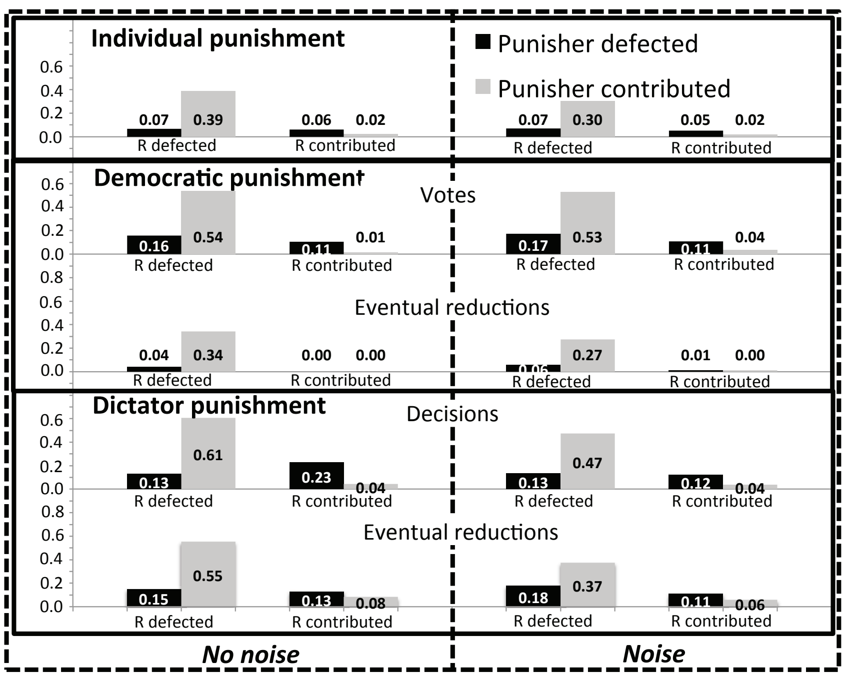
TABLE 3: PROBIT/TOBIT/OLS ESTIMATIONS OF CONTRIBUTIONS, PUNISHMENTS AND NET EARNINGS BASED ON TREATMENT DUMMIES

Dependent	Public Good	Received Punishment				Net
	Contribution	All	PR Defect	PR Coop.	True Coop.	Profits
Model	Probit	Tobit	Tobit	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept		-78.10***	-19.56	2.61***	2.43***	54.77***
		[24.00]	[16.38]	[0.83]	[0.85]	[2.01]
Round	-0.013***	-1.80***	-3.17***	-0.06*	-0.04	-0.10
	[0.002]	[0.58]	[0.74]	[0.03]	[0.04]	[0.08]
Noise	-0.090	10.28	-0.33	-0.02	1.25	-2.80
	[0.085]	[15.24]	[16.43]	[1.03]	[0.99]	[2.47]
Democratic punishment	0.270**	-103.87***	-38.20	-2.03***	-2.05***	11.46***
	[0.106]	[23.01]	[33.35]	[0.69]	[0.69]	[2.71]
Noise ×	-0.098	28.03	-2.92	0.20	0.73	-4.40
Democratic punishment	[0.132]	[23.88]	[38.60]	[1.04]	[1.03]	[3.33]
Dictator punishment	0.081	3.09	14.44	3.63**	3.61**	-4.11
	[0.105]	[18.54]	[24.93]	[1.77]	[1.78]	[3.72]
Noise ×	-0.035	-7.39	-11.66	-0.84	-0.31	-0.88
Dictator punishment	[0.128]	[26.32]	[32.37]	[2.38]	[2.26]	[4.84]
P-values from post-estimation F-tests						
Noise + N×DemP = 0	0.066	0.046	0.926	0.110	0.000	0.002
DemP + N×DemP = 0	0.021	0.000	0.040	0.018	0.082	0.000
Noise + N×DicP = 0	0.193	0.894	0.668	0.690	0.647	0.377
DicP + N×DicP = 0	0.532	0.810	0.894	0.080	0.020	0.110
N	9600	9600	4875	4725	4957	9600
Pseudo R-squared	0.058	0.026	0.013			
N left-censored		7952	3510			
N right-censored		777	649			
Adjusted R-squared				0.047	0.038	0.073

Note: For the Probit estimation on contributions, we report marginal effects at the mean, rather than coefficients. Received punishment points are censored at 0 and 60, but Models 4 and 5 do not converge as Tobit models, so we report results from OLS regressions in these cases. 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.

Dictator punishment has very different effects compared to *Demoratic punishment*. Allowing one individual to punish on behalf of all group members does not significantly affect overall contribution and punishment levels or net profits, with and without noise (see Models 1, 2, and 6 in Table 3). However, when considering only punishment towards contributors in Models 4 and 5 in Table 3, we observe an actual *increase* in punishment under *Dictator punishment* compared to *Individual punishment*, both with and without noise. We will explore these effects more deeply in the next subsection. On the aggregate level, as Tables 1, 2, and 3 indicate, *Dictator punishment* leads to very similar contribution, punishment and net profits levels as *Individual punishment*, and significantly higher punishment levels and significantly lower net profits than *Democratic punishment*.

FIGURE 4: Frequency of (vote for and eventual) punishment, conditional on punisher's own contribution and receivers' public record



III.B Punishment pattern

Figure 4 shows the punishment pattern in our six treatments. It displays the frequency of punishment conditional on whether the punisher contributed or not and whether the punishment receiver contributed or not. For the democratic punishment treatments, the figure distinguishes between votes for punishment and eventual punishment (when votes for punishment reached the required majority). For the dictator punishment treatments, we distinguish between the punishment decisions of all (potential dictator) members and the decisions of the eventually selected dictators (which should be the same in expectation and differ only due to the random selection process).

Table 4 complements Figure 4 by displaying results from non-parametric tests that compare the values shown in the figure along treatment dimensions, punishment source, and target characteristics.[10] In the following, we summarize our observations from Figure 4. Table 4 is organized in the same order, such that it supports these observations statistically point by point.

1. Unsurprisingly, defectors attract more punishment than contributors. For punishers who contributed, this is significant across treatments and conditions (rows No Noise, P contr and Noise, P contr). Interestingly, under democratic punishment defectors are also more likely to punish other defectors than other contributors, both when looking at votes as well as when looking at eventual outcomes (rows No noise, P defect and Noice, P defect). While the latter observation could be caused by majorities of cooperators dragging defectors along to punish another defector, the former result suggests that this is not the case: defectors also intend to punish other defectors more than cooperators.

2. In their individual choices/votes, contributors are much more likely than defectors to punish defectors. This is highly significant except for dictator punishment under *No Noise*. Contributors are also less likely than defectors to

[10]Since we are employing a full battery of tests here, we decided to adjust the p-values required for a particular significance level with a Bonferroni correction. We assume each set of four tests in Table 4 to belong the the same 'family' of hypotheses, and correspondingly divide the required p-value for a particular significance level by 4. As a result, a Null hypothesis is rejected at the 10% level when the p-value is 0.025 or below, and it is rejected at the 5% level (1% level) when the p-value is 0.0125 (0.0025) or below, respectively. Table 4 reports original p-values obtained from the tests, but stars represent the corrected significance level. Group-level averages serve as independent observations.

TABLE 4: P-values from non-parametric tests comparing results reported in Figure 4

1. Receiver defected vs. Receiver contributed

	IndP	DemP votes	DemP evtl	DicP ind	DicP evtl
No noise, P defect	0.022*	0.012**	0.002***	0.041	0.006**
No noise, P contr	0.001***	0.000***	0.001***	0.001***	0.001***
Noise, P defect	0.063	0.002***	0.000***	0.236	0.013*
Noise, P contr	0.001***	0.000***	0.000***	0.001***	0.000***

2. Punisher defected vs. Punisher contributed

	IndP	DemP votes	DemP evtl	DicP ind	DicP evtl
No noise, R defect	0.001***	0.001***	0.071	0.002***	0.293
No noise, R contr	0.728	0.023*	no diff	0.000***	0.000***
Noise, R defect	0.001***	0.000***	0.004**	0.140	0.163
Noise, R contr	0.074	0.045	0.084	0.030	0.163

3. Individual punishment choices/votes across treatments

	IndP vs. DemP		*IndP vs. DicP*		*DemP vs. DicP*	
	No noise	Noise	No noise	Noise	No noise	Noise
P defect, R defect	0.029	0.000***	0.011**	0.059	0.645	0.121
P defect, R contr	0.791	0.171	0.020*	0.031	0.222	0.426
P contr, R defect	0.093	0.000***	0.047	0.017*	0.190	0.166
P contr, R contr	0.511	0.049	0.510	0.005**	0.183	0.201

4. Democratic punishment: votes vs. elected punishment

No noise		*Noise*	
P defect, R defect	0.728	P defect, R defect	0.001***
P defect, R contr	0.003***	P defect, R contr	0.000***
P contr, R defect	0.030	P contr, R defect	0.000***
P contr, R contr	0.005**	P contr, R contr	0.000***

5. Eventual punishments across treatments

	IndP vs. DemP		*IndP vs. DicP*		*DemP vs. DicP*	
	No noise	Noise	No noise	Noise	No noise	Noise
P defect, R defect	0.187	0.730	0.002***	0.026	0.433	0.072
P defect, R contr	0.000***	0.010**	0.215	0.293	0.000***	0.000***
P contr, R defect	0.373	0.796	0.275	0.358	0.647	0.198
P contr, R contr	0.001***	0.051	0.090	0.058	0.000***	0.001***

6. No noise vs. Noise

	IndP	DemP votes	DemP evtl	DicP ind	DicP evtl
P defect, R defect	0.874	0.894	0.401	0.486	0.748
P defect, R contr	0.791	0.648	0.092	0.081	0.028
P contr, R defect	0.206	0.091	0.051	0.226	0.397
P contr, R contr	0.343	0.046	0.176	0.159	0.005**

Note: All tests rely on averages at independent group level. For comparisons across treatments (within a treatment) we employ Wilcoxon Ranksum tests (Wilcoxon Matched Pairs Signed Ranks tests), respectively. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively, after applying an ex-post Bonferroni correction for repeated hypothesis tests, assuming each set of n=4 tests to be a test family (that is, dividing the required p-value for a level by 4).

engage in anti-social punishment and punish contributors. These differences, however are not or only weakly statistically significant (except for dictator choices in the *No Noise* condition).

3. Across treatments, we generally observe a higher likelihood to vote/opt for punishment in democratic and dictator decisions compared to the willingness to individually punish in the same situation. These differences, however, are statistically only significant for the democratic vs. individual punishment under noise and towards defectors. Notably, we do not observe significant differences in individual votes/choices between democratic and dictator punishment.

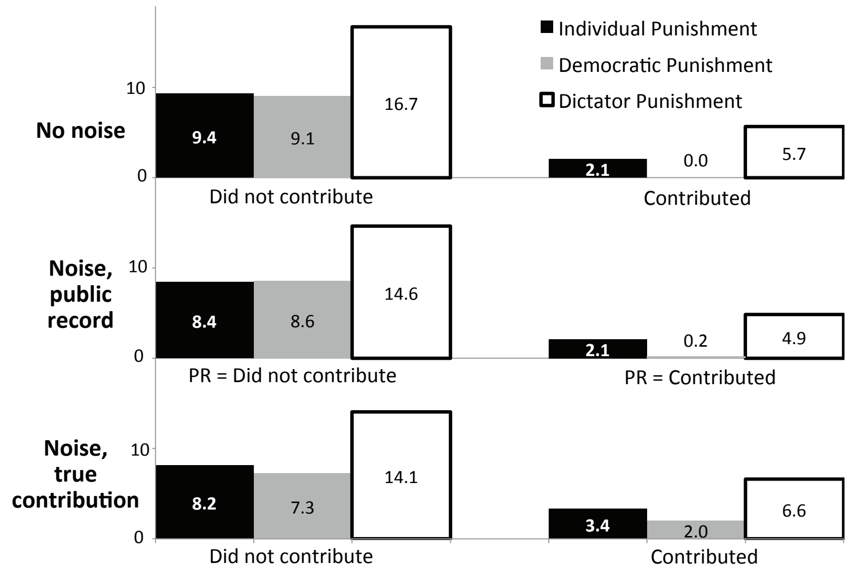
4. With *Democratic Punishment*, we observe a drop in punishment frequency from votes to eventual punishments, indicating that often some group members wanted to punish but did not reach the required majority. This drop is significant across all types of punishment interactions for the *Noise* treatment, but only for punishment towards contributors (where it literally dropped down to zero) in the *No Noise* condition.

5. Albeit Figure 4 displays quite some variation in *eventual* punishments across our three treatments, for the different drilled-down cases these numbers do not differ statistically, with one important exception: (anti-social) punishment towards contributors is significantly lower under *Democratic Punishment* compared to *Individual* and *Dictator Punishment*.

6. We cannot detect systematic significant differences in punishment pattern between our *No Noise* and *Noise* conditions.

In sum, the main effect of introducing *Democratic Punishment* on punishment patterns is that cooperators are effectively not punished anymore. Figure 5 visualizes these consequences. It displays the resulting average number of received punishment points conditional on whether (the public record indicated that) the participant had contributed or not. Participants who did not contribute were deducted an average of 9.4 points (8.4 points) when there was *No Noise* (*Noise*), and this changed only slightly to 9.1 points (8.6 points) when employing *Democratic Punishment*. But for punishment towards contributors, introducing the voting procedure resulted in a drop from an average of 2.1 points (2.1 points) to literally 0 points (0.2 points) when there was *No Noise* (*Noise*). In the *Noise* treatment, this drop did not benefit

FIGURE 5: Average punishment points deducted, conditional on punished subject's (true) contribution and public record



all contributors, since some of them were burdened with a "no contribution" public record, such that the real expected punishment of a contributor decreased from 3.4 points on average with noisy *Individual punishment* to 2.0 points with noisy *Democratic punishment*. With *Dictator punishment*, on the other hand, we observe almost a doubling of received punishment points, on average, both for contributors and non-contributors, and both with and without noise in the observation of contribution decisions.[11]

III.C Reactions to received punishment

In Table 5 we report results from Probit regressions that explore how participants' contribution behavior responds to punishment received in the previous round. In Model 1 and 3 of Table 5, we regress the current contribution of a participant on the number of punishment points that were deducted from his income in the pre-

[11]Statistically, based on Wilcoxon ranksum tests, the received punishment of contributors is significantly lower at the 1%-level in the democratic punishment treatment than in individual or dictator punishment treatments, and all other comparisons, while partly strong in magnitude, do not yield statistical significance at the 10%-level. The only exceptions to this rule are found under Noise, True contribution, Contributed: Individual vs. Democratic punishment $p = 0.239$, Individual vs. Dictator punishment $p = 0.040$.

TABLE 5: Probit estimations of current contribution based on previous round behavior

	Model 1	Model 2	Model 3	Model 4
Noise	-0.066			
	[0.048]			
DemPun	0.156**	0.128***		
	[0.064]	[0.047]		
Noise × DemPun	-0.044			
	[0.071]			
DicPun	0.044	0.045		
	[0.060]	[0.043]		
Noise × DicPun	-0.003			
	[0.069]			
$RecPnmt_{PR}$	0.007***	0.007***	0.007***	0.005***
	[0.002]	[0.001]	[0.001]	[0.002]
$RecPnmt_{PR}$ × Noise	0.002		-0.001	0.001
	[0.002]		[0.002]	[0.002]
$RecPnmt_{PR}$ × DemPun	-0.001	-0.003*	0.001	-0.002
	[0.002]	[0.001]	[0.002]	[0.002]
$RecPnmt_{PR}$ × Noise × DemPun	-0.002		-0.001	
	[0.002]		[0.002]	
$RecPnmt_{PR}$ × DicPun	-0.004*	-0.004***	-0.003*	
	[0.002]	[0.001]	[0.001]	
$RecPnmt_{PR}$ × Noise × DicPun	-0.001		0.000	
	[0.002]		[0.002]	
$Contr_{PR}$	0.533***	0.427***	0.561***	0.440***
	[0.026]	[0.032]	[0.026]	[0.032]
$Contr_{PR}$ × $RecPnmt_{PR}$	-0.016***	-0.011**	-0.017***	-0.013***
	[0.004]	[0.004]	[0.004]	[0.004]
$Contr_{PR}$ × $RecPnmt_{PR}$ × Noise	0.006		0.006	
	[0.005]		[0.005]	
$Contr_{PR}$ × $RecPnmt_{PR}$ × DemPun	0.007**	0.003	0.009***	0.005
	[0.003]	[0.007]	[0.003]	[0.007]
$Contr_{PR}$ × $RecPnmt_{PR}$ × Noise × DemPun	*Omitted due to collinearity*			
$Contr_{PR}$ × $RecPnmt_{PR}$ × DicPun	0.010**	0.005	0.011***	0.007
	[0.004]	[0.004]	[0.004]	[0.004]
$Contr_{PR}$ × $RecPnmt_{PR}$ × Noise × DicPun	-0.006		-0.006	
	[0.005]		[0.005]	
$Contr_{PR}$ × $PRwrong_{PR}$		0.082*		0.077
		[0.049]		[0.050]
$Contr_{PR}$ × $PRwrong_{PR}$ × $RecPnmt_{PR}$		0.008		0.009**
		[0.005]		[0.004]
$Contr_{PR}$ × $PRwrong_{PR}$ × $RecPnmt_{PR}$ x DemPun		-0.001		-0.002
		[0.007]		[0.007]
$Contr_{PR}$ × $PRwrong_{PR}$ × $RecPnmt_{PR}$ x DicPun		-0.003		-0.005
		[0.005]		[0.005]
N	9120	4845	9120	4845
Pseudo R-squared	0.234	0.152	0.218	0.144

Note: We report average marginal effects. Robust standard errors are clustered at the group level and reported in brackets. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

vious round ($RecPnmt_{PR}$). We control for whether the participant contributed in the previous round or not ($Contr_{PR}$), and interact previous punishment and previous contribution with each other as well as with treatment dummies that indicate whether noise was present (*Noise*), whether punishment was determined individual, democratically, or by a dictator(*DemPun*, *DicPun*), or combinations of these (*Noise* × *DemPun*, *Noise* × *DicPun*).[12] Models 2 and 4 in Table 5 only look at choices in the three *Noise* treatments and analyze whether having received a wrong public record in the previous round (dummy $PRwrong_{PR}$, indicating that the public record displayed that participant hasn't contributed even though he did) changes next-round reactions to received punishment. Since all these regressions include the previous round's contribution as an independent, the actual treatment effect on contributions is already contained in last round's contribution, they become treatment-specific time trend indicators.[13] To verify robustness of our estimates, we report both models with (1 and 2) and without (3 and 4) baseline treatment effects included. Our estimates and conclusions remain the same across these variations.

Models 1 and 3 of Table 5 show that participants who did not cooperate increase their next-round contribution for each punishment point received ($RecPnmt_{PR}$ is positive), but (weakly significantly) less so under *Dictator Punishment* than in the other two punishment conditions ($RecPnmt_{PR} \times DicPun$ is negative). When the participant cooperated, then is effect is reversed: The joint effect of $RecPnmt_{PR} + Contr_{PR} \times RecPnmt_{PR}$ in Model 3 equals $0.007 - 0.017 = -0.010$ and is significantly different from zero ($p = 0.009$). That is, when a participant cooperated, then higher punishment leads to *less* cooperation in the next round. This averse reaction of cooperators against punishment, however, is mitigated under democratic and dictator punishment ($Contr_{PR} \times RecPnmt_{PR} \times DemPun$ and $Contr_{PR} \times RecPnmt_{PR} \times DemPun$ are both statistically significantly positive, and the total joint effects of punishment on cooperators in these treatments are not different from zero, $p = 0.748$ and $p = 0.526$, respectively). Models 2 and 4 only consider the noise treatments. Similarly, they show that defectors and cooperators

[12]Since under *Democratic Punishment*, the only punishment of contributors happens when there is *Noise* (and never when there is *No Noise*), we have a problem of perfect collinearity in that condition, and therefore do not include the variable "$Contr_{PR} \times RecPnmt_{PR} \times$ Noise × VotePun" in our estimations.

[13]They now represent how the treatment affects the absolute difference between last round's and this round's contribution, after controlling for a relative time trend (coefficient on last round contribution).

react differently to punishment ($RecPnmt_{PR}$ is positive but $Contr_{PR} \times RecPnmt_{PR}$ is negative). Here, however, the negative effect of punishment on cooperators is reduced if that punishment came after the cooperator's action was wrongly displayed as a defection ($Contr_{PR} \times PRwrong_{PR} \times RecPnmt_{PR}$ is positive).

IV Conclusion

In this paper we observed that democratic punishment, when punishment decisions in a group are decided by majority voting, facilitates more cooperation and higher payoffs than individual punishment. It achieves so by establishing a stronger connection between a member's contribution decision and whether the member gets punished, in particular by decreasing anti-social punishment while keeping the same level of pro-social punishment. We also see some evidence that participation in democratic punishment makes punishment intentions themselves more pro-social. The findings suggest that social norms or institutions that help members of a group to coordinate punishment decisions, and make it contingent on majority approval, can be welfare enhancing, even without the ability to make future commitments for punishment. A direction for future research is investigating what voting rule for punishments is optimal for society's welfare, for different levels of noise in observations, although addressing this question would ideally require larger groups than in our study. Presumably the expected welfare in the group is non-monotonic in the strictness of the voting rule, since if the threshold for punishing is very low, outcomes might be similar to individual punishment, while if they are too high then it might become impossible for the group to agree upon punishing someone, resulting in widespread free riding.

References

Aldeshev, G. and Zanarone, G. (2014), Endogenous enforcement institutions, Technical report, Working Paper, University of Namur.

Ambrus, A. and Greiner, B. (2012), 'Imperfect public monitoring with costly punishment: An experimental study', *American Economic Review* **102**(7), 3317–3332.

Andreoni, J. and Gee, L. (2012), 'Gun for hire: Does delegated enforcement crowd out peer punishment in giving to public goods?', *Journal of Public Economics* **96**, 1036–1046.

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

Baldassarri, D. and Grossman, G. (2011), 'Centralized sanctioning and legitimate authority promote cooperation in humans', *Proceedings of the National Academy of Sciences of the USA* **108**(1-5).

Casari, M. and Luini, L. (2009), 'Cooperation under alternative punishment institutions: An experiment', *Journal of Economic Behavior and Organization* **71**, 273–282.

Cinyabuguma, M., Page, T. and Putterman, L. (2006), 'Can second-order punishment deter perverse punishment?', *Experimental Economics* **9**, 265–279.

Dal Bó, P., Foster, A. and Putterman, L. (2010), 'Institutions and behavior; Experimental evidence on the effects of democracy', *American Economic Review* **100**, 2205–2229.

Ertan, A., Page, T. and Putterman, L. (2009), 'Who to punish? Individual decisions and majority rule in mitigating the free rider problem', *European Economic Review* **53**, 495–511.

Fehr, E. and Fischbacher, U. (2004), 'Third-party punishment and social norms', *Evolution and human behavior* **25**, 63–87.

Fehr, E. and Gächter, S. (2000), 'Cooperation and punishment in public goods experiments', *American Economic Review* **90**, 980–994.

Fischbacher, U. (2007), 'z-Tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics* **10**, 171–178.

Fischer, S., Grechenig, K. and Meier, N. (2013), Cooperation under punishment: Imperfect information destroys it and centralizing punishment does not help, Technical report, Working Paper, University of Bonn.

Frey, B. (1994), 'Direct democracy: Politico-economic lessons from Swiss experience', *American Economic Review Papers & Proceedings* **84**, 338–342.

Frey, B., Benz, M. and Stutzer, A. (2004), 'Introducing procedural utility: Not only what, but also how matters', *Journal of Institutional and Theoretical Economics* **160**, 377–401.

Fudenberg, D., Rand, D. G. and Dreber, A. (2012), 'Slow to anger and fast to forget: Cooperation in an uncertain world', *American Economic Review* **102**, 720–749.

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

Grechenig, C., Nicklisch, A. and Thöni, C. (2010), 'Punishment despite reasonable doubt - A public goods experiment with sanctions under uncertainty', *Journal of Empirical Legal Studies* **7**, 847–867.

Greiner, B. (2015), 'Subject pool recruitment procedures: Organizing experiments with ORSEE', *Journal of the Economic Science Association* **1**(1), 114–125.

Hauser, O., Nowak, M. and Rand, D. (2014), 'Punishment does not promote cooperation under exploration dynamics when anti-social punishment is possible', *Journal of Theoretical Biology* **360**, 163–171.

Herrmann, B., Thöni, C. and Gächter, S. (2008), 'Antisocial punishment across societies', *Science* **319**, 1362–1367.

Kamei, K. and Putterman, L. (2015), 'In broad daylight: Fuller information and higher-order punishment opportunities can promote cooperation', *Journal of Economic Behavior and Organization* **120**, 145—159.

Kamei, K., Putterman, L. and Tyran, J. (2015), 'State or nature? Endogenous formal versus informal sanctions in the voluntary provision of public goods', *Experimental Economics* **18**, 38–65.

Leibbrandt, A. and López-Pérez, R. (2011), 'The dark side of altruistic third-party punishment', *Journal of Conflict Resolution* **55**, 761–784.

Leibbrandt, A. and López-Pérez, R. (2012), 'An exploration of third and second party punishment in ten simple games', *Journal of Economic Behavior and Organization* **84**, 753–766.

Markussen, T., Putterman, L. and Tyran, J. (2014), 'Self-organization for collective action: An experimental study on sanction regimes', *Review of Economic Studies* **81**, 301–324.

Markussen, T., Putterman, L. and Tyran, J. (2016), 'Judicial error and cooperation', *European Economic Review* **89**, 372–388.

Nikiforakis, N. (2008), 'Punishment and counter-punishment in public good games: Can we really govern ourselves?', *Journal of Public Economics* **92**, 91–112.

Pommerehne, W. and Weck-Hannemann, H. (1996), 'Tax rates, tax administration and income tax evasion in switzerland', *Public Choice* **88**, 161–170.

Rand, D., Fudenberg, D. and Dreber, A. (2015), 'It's the role that counts: The role of intentions in reciprocal altruism', *Journal of Economic Behavior and Organization* **116**, 481–499.

Sutter, M., Haigner, S. and Kocher, M. (2010), 'Choosing the carrot or the stick? - endogenous institutional choice in social dilemma situations', *Review of Economic Studies* **77**, 1540–1566.

Tyran, J. and Feld, L. (2006), 'Achieving compliance when legal sanctions are non-deterrent', *Scandinavian Journal of Economics* **108**, 135–156.

Van Miltenburg, N., Buskens, V., Barrera, D. and Raub, W. (2014), 'Implementing punishment and reward in the public goods game: the effect of individual and collective decision rules', *International Journal of the Commons* **8**(1), 47–78.

Wallach, M. A., Kogan, N. and Bem, D. J. (1962), 'Group influence on individual risk taking', *Journal of Abnormal and Social Psychology* **65**, 75–86.

Wallach, M. A., Kogan, N. and Bem, D. J. (1964), 'Diffusion of responsibility and level of risk taking in groups', *Journal of Abnormal and Social Psychology* **68**, 263–274.