Version of 1-11-08

Getting Punishment Right: Do Costly Monitoring or Redistributive Punishment Help?

Talbot Page, Louis Putterman and Bruno Garcia

We introduce new treatments of a voluntary contribution mechanism with opportunities to punish, to see how contributions and punishments change when (a) each dollar lost in punishment must be awarded to another team member and/or when (b) obtaining information on individuals' contributions is a costly choice. Conjectures that tying punishments to rewards might reduce punishment of high contributors (perverse punishment) or increase overall punishing are not completely born out, but innovation (a) nonetheless succeeds in making the *net* punishment of high contributors much less common because they receive enough rewards to offset punishment. A surprise finding is that innovation (b) also decreases the incidence of misdirected punishment, since high contributors do more monitoring than low ones while low contributors do most of the perverse punishing. Both innovations raise both contributions and earnings relative to the familiar VCM-with-punishment treatment.

Keywords Public goods, collective action, experiment, punishment

JEL Classification C91, H41, D30

Getting Punishment Right: Do Costly Monitoring or Redistributive Punishment Help?[*]

Talbot Page, Louis Putterman and Bruno Garcia
Brown University

1. Introduction: Voluntary Contributions and Punishment

One of the most active areas of experimental economics research in recent years has been the exploration of subjects' propensities to engage in costly punishment, and the impact of punishment on a variety of interactions, including public goods dilemmas. Punishment is of interest both for the practical reason that it may play a pivotal role in solving these and other dilemma problems by creating self-interested incentives to contribute (Fehr and Gächter, 2000; Gürerk, Irlenbusch and Rockenbach, 2006), and because its robust manifestation is a strong indication of a social preference or psychological predisposition that has broad implications for both theory and policy (Field, 2001; Fehr and Schmidt, 2003).

Although the demand for punishment and its usually pronounced effect on voluntary contributions are remarkable, authors including Cinyabuguma, Page and Putterman (2004), Botelho, Harrison, Costa-Pinto and Rutström (2005), and Gächter and Herrmann (2005) have pointed out that allowing punishment in public goods experiments can have a far less salutary effect on earnings, hence efficiency. Cinyabuguma, Page and Putterman (2006), Ertan, Page and Putterman (2006), and Gächter, Herrmann and Thöni (2005) provide evidence that a major reason for the sometimes inefficient impact of decentralized punishment is that a substantial fraction of it is misdirected at high contributors—a phenomenon we call "perverse punishment."[1]

[*] We are grateful to the Department of Economics at Brown University for funding most of the experiments reported here; one earlier-conducted treatment used funds provided by the MacArthur Foundation network on norms and preferences and National Science Foundation Grant SES-0001769. We are grateful to Omar Ahmed for modifying the programs used to conduct the experiments. Garcia's work was funded by the Chase Manhattan research assistantship.

[1] We call punishment of high contributors "perverse" because the expectation of it tends to discourage contributions, thus reducing social efficiency. Gächter *et al*. use the term "anti-social punishment," with a small difference in application: whereas we use it to refer to punishment received by the contributor of more than a group's average amount (and sometimes in a stronger sense, the contributor of the maximum observed amount in the group), they use it for any instance in which a subject *i* punishes a subject *j* who has contributed more than *i* has.

Recent papers have studied what happens when the cost of punishing varies (Anderson and Putterman, 2006, Carpenter, 2007, Nikiforakis and Norman, 2005), when subjects can determine for themselves whether punishment is permitted (Botelho *et al.* 2005, Ertan, Page and Putterman, 2006, Gürerk, Irlenbusch and Rockenbach, 2005, 2006, Sutter, Haigner and Kocher, 2005), and when punishers themselves can be punished (Nikiforakis, 2005, Cinyabuguma, Page and Putterman, 2006). The efficacy of punishment has also been contrasted with the efficacy of rewards, which have been found to be less effective and less sustained in their impact (Sefton, Shupp and Walker, 2002, Sutter, Haigner and Kocher, 2005), yet less objectionable to some subjects (Sutter *et al.*).

One problem that has not received much attention is that of how observable actions are and of what would happen if observing them entails costs that self-interested agents would not be expected to pay. If a group is to engage in collective action with the help of punishment of free riders by other group members, what is needed is often both mutual monitoring and punishment or peer pressure. Like costly punishment, costly monitoring can be viewed as a second-order public good that can help to solve the first-order public goods problem by making it incentive-compatible to contribute, but that is itself subject to free riding.[2]

Another issue that remains largely unaddressed is the effect of pairing punishment and reward by making punishments "redistributive." Transferring rather than throwing away punished funds is more efficient in its immediate effects, but we wondered about its incentive effects. An experiment that likewise reduces the efficiency cost of punishment is Casari and Plott (2003), in which earnings lost to the person targeted for punishment are transferred to the punisher, so that there is self-interest in punishing. In Sausgruber and Tyran (2007), funds deducted by punishment are distributed to another team member, but punishment is costless to the punisher.

In this paper, we study a set of experimental treatments that permit us to study both costly monitoring and redistributive punishment. In two treatments, subjects learn the contributions of the individual members of their group only if they pay for the information, which then entitles them to impose a certain amount of punishment (or

[2] Grosse, Putterman and Rockenbach (2007) study team members' decisions to invest in costly monitoring when the impact of such investments is to eliminate free riding incentives by making team production incentive-compatible, as envisaged by Alchian and Demsetz (1972).

punishment and rewards) should they wish to. In two treatments, the earnings a punisher deducts from a targeted team member must be assigned by the punisher to some other member of the team, which renders punishment redistributive in nature and makes each dollar of punishment also a dollar of reward.

The inclusion of costly information treatments allow us to compare willingness to incur costs to punish with willingness to incur costs for information on which punishment might be based. It seemed to us that punishment could be either more common or less common when there is a monitoring cost: more common because curiosity might spur payments for information after which subjects punish (in our treatments this entails no additional cost); less common because not having the information displayed might mean that instead of feeling anger at free riders subjects might simply accept the "inevitability" of low contributions and choose not to "throw good money after bad."

We find, surprisingly, that making contribution information costly doesn't significantly alter the overall frequency of punishing, but that it does reduce the proportion of punishment perversely aimed at high contributors. The reason is that it is the higher contributors who disproportionately purchase the contributions information, and since they rarely engage in perverse punishment, having the information be a costly pre-condition to punishment tends to screen out many perverse punishers.

The treatments in which punishment is costly to the punisher but redistributive among the other subjects let us study the incentive effects of such an arrangement. We anticipated that these effects might include (a) providing incentives to contribute more, because higher contributors might be rewarded, (b) eliciting more punishment, if concern about aggregate efficiency discourages some punishing in the heretofore standard treatments, and (c) motivating "pro-social" more than "anti-social" punishers, since the former (those who tend to punish low contributors) might value the rewarding of high contributors as well as the punishing of low ones. We find no evidence of effects (b) or (c), but strong evidence of effect (a).

In addition to incentive effects, having punishment losses be redistributed to other subjects has a resource effect that calls for care in its evaluation. A basic idea in cost-benefit analysis is that a dollar is a dollar, no matter in whose pocket it lands. This dictum allows, as a matter of practicality, cost-benefit practitioners to set aside distributional issues while estimating total net benefits and efficiency gains for a proposed policy. This principle of traditional cost-benefit analysis allows the practitioner to focus, for example, on potential Pareto improvements rather than actual Pareto improvements. A basic idea in the mechanism design literature, in contrast, is that differing distributions have differing incentive effects on efficiency, and thus differing distributions should not be set aside. When it comes to evaluating our experimental results, our ability as experimenters to shift opportunities from, e.g., punishment that "burns" the losses of those targeted to punishment that preserves those dollars and gives them to others, puts our redistributive treatments at an efficiency advantage that is in some respects artificial. It is important, therefore, to disentangle the resource and incentive effects of the institutional changes we induce.

Our experimental design permits us to do just this. Estimating the effects of the alternative policies both inclusive of and exclusive of resource costs, as we will do, illustrates incentive and resource separability in a way that is rarely practicable in policy analysis. We calculate, first, overall efficiencies, a straightforward procedure in the experimental setting. Then we net out the differences in the direct resource costs from "throwing away" or redistributing punishment costs and from making information or punishment either free or costly to the punisher. We find that having one subject's earnings reductions (penalties) be distributed to others can have not only the direct resource benefit inherent in this mechanism but also an incentive benefit that raises efficiency by encouraging more contributions to the public good. And we find that switching from a costly punishment to a costly information mechanism is not deleterious to incentive efficiency, although its small positive effect is not significant.

The rest of the paper proceeds as follows. In Section 2, we give details of the experimental design. In section 3, we discuss relevant theoretical issues and predictions. Section 4 describes and analyses the results of the experiment. Section 5 provides further discussion and conclusions.

2. Design

Our starting point is the well-known design of a voluntary contributions mechanism with punishment, in which subjects are randomly assigned to groups that remain fixed (a "partners" design) for a finite and known number of periods. In this experiment there are 4 members to each group and 20 periods in an experimental session. In the first stage of each period each subject in a group is provided with an initial endowment that he or she is asked to divide between a private account and a group account, before observing the choices of fellow group members. Once the choices are made the subjects learn the amounts others in the group contributed to the group account. The total amount of funds placed in the group account is scaled up by the experimenter and divided equally among the subjects in the group without regard to individual contribution. In the second stage, each subject can reduce the earnings of others in his or her group. A reduction ("punishment") is costly to both the punished and the punisher.

In this baseline treatment with punishment, the payoff function for subject *i* for a period is

(1) Baseline treatment **P** $$y_i = (10 - C_i) + (0.4)\sum_{j=1}^{4} C_j - (0.25)\sum_{j \neq i} P_{ij} - \sum_{j \neq i} P_{ji} \, ,$$

where subject *i* is endowed with 10 experimental dollars (hereafter *E*$) each period, and C_i is *i*'s contribution to the group account; the scaling factor is 1.6 (when divided four way, 0.4) and the summation is taken over the 4 members of *i*'s group; P_{ij} is the amount of punishment *i* imposes on *j*, 0.25 is the cost of punishing per dollar of punishment; and P_{ji} is the amount of punishment *i* receives from each other subject *j*.[3] General constraints on punishment in all treatments were: (*i*) a subject could not spend more than her/his pre-punishment earnings for the period on reducing the earnings of other subjects, (*ii*) a subject's post-punishment earnings for a period would be set to zero if earnings y_i in equation (1) were negative, and (*iii*) a subject *i* could not spend more on reducing the earnings of a subject *j* in any period than would single-handedly reduce *j*'s earnings according in (1) to less than zero.[4]

[3] Experimental dollars converted to U.S. dollars at the rate of one experimental dollar = $0.07 at the end of the session. In the rest of our discussion, the word "dollar" should be understood to mean experimental dollar. Subject earnings over twenty periods plus a $5 show-up fee averaged a little under $25 in U.S. currency.

[4] The payment functions and restrictions are identical to those in Bochet *et al.* (2006) and Page *et al.* (2005). The purpose of (*i*) and (*ii*) was to keep all decisions financially independent of each other while maintaining a guaranteed minimum payment for recruiting reasons. The purpose of (*iii*) was to help subjects to avoid pointless spending on punishment in view of constraint (*ii*). Note, however, that it

This design generates the familiar result (for self-interested subjects) that it is socially optimal for each subject to contribute everything to the group account without any punishment, but privately optimal to contribute nothing and not punish. In replicated experiments, however, subjects make some contributions and impose some punishment. Experimentalists have found that the opportunity to punish creates an incentive to increase contributions but often has little effect on efficiency, partly because punishment is costly to both punisher and the punished, and partly because some punishment is directed at high contributors, thus tending to undermine the incentive to contribute.[5]

To further study incentive and resource effects on punishment, contributions, and efficiency we consider three other treatments in comparison with our baseline treatment. Our first comparison treatment is redistributive punishment (**RP)**. In this treatment each dollar of punishment *i* imposes on another group member *i* must redistribute to one or both of the remaining two members of the group (in other words reductions in earnings act like fines to be redistributed rather than behaving like physical punishments to be endured). As in the baseline treatment it still costs (*E*$0.25) to reduce another's earnings by *E*$1. Subject *i*'s payoff function is:

(2) **RP** $$y_i = (10 - C_i) + (0.4)\sum_{j=1}^{4} C_j - (0.25)\sum_{j \neq i} P_{ij} - \sum_{j \neq i} P_{ji} + \sum_{k \neq i} R_{ki} \ ,$$

where R_{ki} is the number of dollars that others have redistributed to *i*. At the end of the punishment/reward stage, subject *i* is shown the net change in her earnings due to rewards and punishments, i.e. $\sum_{k \neq i} R_{ki} - \sum_{j \neq i} P_{ji}$, but not its addition and reduction

remained possible for subjects to "overspend" on punishing in the sense that both subject *i* and, say, subject *k* might each spend enough to reduce *j*'s earnings for the period to zero, although only one subject's punishment would actually be effective in that case, given (*ii*). This could happen because subjects did not learn of punishment not carried out by or aimed at them, and the design (as in Fehr and Gächter, 2000a) keeps such information private so as not to encourage free riding on punishment.

[5] Thus, Cinyabuguma, Page and Putterman (2004) and Botelho *et a*l. (2006) report earnings to be lower in treatments with than in those without punishment, although Nikiforakis and Norman (2005) find that whether this is so is sensitive to parameter choice and some experiments, such as Gürerk *et al*. (2006), find earnings higher in groups with punishment. Ertan, Page and Putterman (2006) find earnings lower with than without punishment in groups in which who can be punished is unrestricted, but higher with than without punishment in groups that have voted to permit only low contributors to be punished.

components separately. The Appendix shows the screen design for entering an individual's contribution, punishment and reward decisions.

The second comparison treatment is like the baseline treatment (1), except that in each period information on others' contributions is costly, while imposing punishment is free up to a limit, if the subject has paid for the information. The cost of information is *E*\$1 for a period, and the limit on a subject's punishing for that period is *E*\$10. The payoff function for subject *i* for a period for this treatment with punishment and costly information (**CIP**) is

(3) **CIP**
$$y_i = (10 - C_i) + (0.4)\sum_{j=1}^{4} C_j - \sum_{j \neq i} P_{ji} - I_i,$$

where I_i = *E*\$1 if *i* pays for information in the period in question, 0 otherwise; and if *i* pays for information *i* can impose any amount of punishment from *E*\$0 up to the limit of *E*\$10 in total on other group members.

The third comparison treatment is like the redistributive punishment treatment (2) in that money "punished away" from one subject must be assigned to another, and like the costly information treatment (3) in that in each period information on others' contributions is costly, while imposing punishment is free up to a limit, if the subject has paid for the information. The payoff function for subject *i* for a period for this treatment with free redistributive punishment with costly information (**CIRP**) is

(4) **CIRP**
$$y_i = (10 - C_i) + (0.4)\sum_{j=1}^{4} C_j - \sum_{j \neq i} P_{ji} + \sum_{k \neq i} R_{ki} - I_i.$$

Thus, in treatments **P** and **RP** group members learned one another's contributions immediately and could punish or punish and redistribute within limits for 25 experimental cents to the experimental dollar, while in treatments **CIP** and **CIRP** group members chose whether to spend one experimental dollar after the contribution decisions of each period to be informed of one another's individual contributions, and could then punish or punish and redistribute up to a total of *E*\$10.

The payoff functions for the four treatments can be summarized by a single payoff function with three parameters:

(5) $$y_i = (10 - C_i) + (0.4)\sum_{j=1}^{4} C_j - \alpha_1 \sum_{j \neq i} P_{ij} - \sum_{j \neq i} P_{ji} + \alpha_2 \sum_{k \neq i} R_{ki} - \alpha_3 I_i ,$$

where the four payoff functions differ by the values of the α_i's as follows:

Baseline **P**	Redistributive punishment **RP**	Costly information and punishment **CIP**	Costly information & redistributive punishment **CIRP**
$\alpha_1 = 0.25$ $\alpha_2 = 0$ $\alpha_3 = 0$	$\alpha_1 = 0.25$ $\alpha_2 = 1$ $\alpha_3 = 0$	$\alpha_1 = 0$ $\alpha_2 = 0$ $\alpha_3 = 1$	$\alpha_1 = 0$ $\alpha_2 = 1$ $\alpha_3 = 1$

Table 1. Differences in the payoff for the four treatments, as summarized by parameters α_1, α_2, α_3.

where with the different parameter values there will be differing resource and incentive effects.[6] By netting out the resource effects we identify incentive effects on efficiency. Table 2 summarizes the natures of the four treatments in a 2 X 2 schema and shows subject, session, and group counts.

3. Questions and Predictions

The underlying finitely repeated VCM or public goods game at the heart of our treatments has been conducted many times with similar results. Theoretically, the dominant strategy for a payoff maximizing individual interacting with other payoff maximizing individuals is to contribute nothing to the group account, thus earning *E*$10 each period. In the lab, however, contributions typically begin at an average of 50% or more of endowments and decline with repetition. Explanations that have been offered include confusion and learning, altruism, "warm glow," reciprocity, as well as combinations of such factors and heterogeneity of preferences among individuals.

[6] The fact that values of R_{ki} and/or of I_i are not explicitly elicited from subjects in all treatments can be ignored since α_2 and/or $\alpha_3 = 0$ when that is the case.

		Costs to the punisher/monitor	
		Cost to punish, not to monitor	Cost to monitor, not to punish
Redistributive Punishment	No	**P** Cost to punish, not to monitor $(\alpha_1 = 0.25,\ \alpha_3 = 0)$ No redistributive punishment $(\alpha_2 = 0)$ 4 sessions, 16 groups, 64 subjects	**CIP** Cost to monitor, not to punish $(\alpha_1 = 0,\ \alpha_3 = 1)$ No redistributive punishment $(\alpha_2 = 0)$ 4 sessions, 16 groups, 64 subjects
	Yes	**RP** Cost to punish, not to monitor $(\alpha_1 = 0.25,\ \alpha_3 = 0)$ Redistributive punishment $(\alpha_2 = 1)$ 4 sessions, 16 groups, 64 subjects	**CIRP** Cost to monitor, not to punish $(\alpha_1 = 0,\ \alpha_3 = 1)$ Redistributive punishment $(\alpha_2 = 1)$ 4 sessions, 16 groups, 64 subjects

Table 2. 2 x 2 Schema of Treatments with Session and Subject Information

Treatment **P** resembles the original VCM-with-punishment treatment in Fehr and Gächter (2000), Sefton, Shupp and Walker (2002), Carpenter and Matthews (2002), and other experiments, and corresponds exactly to the "punishment only" treatment in Page, Putterman and Unel (2005).[7] The opportunity to impose costly punishment should not be made use of by payoff maximizers knowing they are playing with others of the same type, but many subjects are observed to punish, even in the known last period. Most punishment is aimed at lower contributors, but a minority of punishment is directed at high contributors. The existence of a punishment opportunity typically increases contributions at the outset, implying that some subjects anticipate that they will be punished if they contribute too little. Contributions then either rise, remain roughly the

[7] Differences from Fehr and Gächter (2000) are relatively minor and include use of a constant rather than increasing punishment cost, endowment of 10 rather than 20, lack of a within-subject comparison condition without punishment, and use of the partner matching protocol only.

same, or decline more slowly and significantly less overall than in treatments without punishment.[8]

Having each unit of punishment be balanced by a unit of reward in treatment **RP** may affect the amount of punishment and who it is directed at in at least two different ways. First, if subjects care about total earnings, as proposed by Charness and Rabin (2002), then they will be less reluctant to punish in the **RP** and **CIRP** than in the **P** and **CIP** treatments, so the former treatments would see more punishing overall. Second, any punishment that in ordinary punishment treatments is motivated by the desire to increase the punisher's earnings relative to the average earnings of others in the group should no longer be undertaken, which would tend to reduce the amount of punishment purchased. While the impact on total punishment is thus unclear, the impact on the proportion of punishment given to high contributors appears likely to be negative.[9] As for the effect on contributions, they can be expected to be higher, both because there's likely to be less perverse punishment and (perhaps) more punishment overall, and because of the extra incentive provided by rewards going to high contributors. Notice that a high contributor who is punished (perversely) by one group member but rewarded by others to a greater degree receives no signal of perverse punishment, so any perverse punishing that takes place may be undetectable by its targets.

The effect of making information costly is difficult to predict. Like punishment itself, the information should never be requested by payoff maximizing subjects who assume that others are also payoff maximizing. The fact that many subjects do typically punish suggests that many might also request the information as a first step toward punishing. However, if punishment is an emotional response to evidence of free riding, it seems possible that the absence of the immediate stimulus of information about free riding might cool the anger that can motivate punishing, leading to less punishment. We made up to ten units of punishment costless in the costly information treatments, once the

[8] Nikiforakis and Normann (forthcoming) find the contribution trend to be systematically related to the cost to the punished individual per unit cost to the punisher: the higher the cost of punishment to the person targeted, the more sustained and increasing are contributions.

[9] Notice that a subject who cares about inequalities relative to individuals rather than the group as a whole, disliking inequalities that are disadvantageous to him more than he dislikes those that favor him (Fehr and Schmidt, 1999), may be motivated to punish high earners and give to low earners. But the efficiency-reducing act of punishing indiscriminately to bring all others' earnings down is frustrated by the requirement to redistribute.

information cost had been paid, in order to focus on the costly information decision and not on both costly information and additionally costly punishment. (Ours is the first VCM treatments with costly information, to our knowledge.) Insofar as we compare the amount of punishment given in the costly information treatments to that given in the free information treatments, we keep in mind that the average cost of punishment is much lower in the costly information treatments. Thus, if reluctance to pay for information were not a factor, punishment might be expected to be more common in the costly information treatments, since it costs less to the punisher (compare Carpenter, 2007, and Anderson and Putterman, 2006).

4. Experiment and Results

As Table 2 shows, we conducted four sessions of each treatment, in each of which sixteen inexperienced subjects drawn from the general undergraduate student body (about 5700 students) at Brown University were randomly assigned to groups of four. Thus, 64 subjects per treatment and 256 subjects in all took part. Subjects, recruited by flyer or job posting in an on-line campus magazine, sat at desks in a computer classroom, read the instructions on-screen as the experimenter read aloud, answered practice questions on paper and using a practice version of the experiment's computer interface, had their procedural questions if any answered, then made their series of binding decisions, without communication. **P** treatment sessions were conducted in 2000 and 2001; those of the **CIRP** treatment in the fall of 2003; and those of the **CIP** and **RP** treatments in late winter and early spring of 2006.[10] Instructions for all treatments are available on request.

Contributions

Figure 1a shows average contribution by period in the **P** and **RP** treatments, while Figure 1b shows average contribution by period in the **CIP** and **CIRP** treatments, and the first row of Table 3 indicates contributions in each treatment averaged over all 20 periods. Considering the treatments paired in the figures, we find average contributions

[10] The **P** treatment data are those referred to as the "punishment only" treatment in Page, Putterman and Unel (2005), while the **RP**, **CIP** and **CIRP** treatments are new to this paper. Instructions, screen lay-outs, and experimental protocols were uniform across the four treatments except as required by specific design elements.

higher in both cases when punishment is redistributive, that is higher in **RP** than in **P** and higher in **CIRP** than in **CIP**.[11] But the stronger contribution performance of the **CIP** than of the **P** treatment makes for a smaller contributions gap between **CIRP** and **CIP** (Fig. 1b). In other words, all three of our experiment's innovations (costly monitoring, redistributive punishment, and their combination) boost contributions relative to the now-familiar **P** treatment. The differences vary in significance, however. Mann-Whitney U tests, treating the contribution in each group of 4 subjects averaged over the 20 periods as an observation, find that contributions are higher in **RP** than in **P**, significant at the 5% level in a two-tailed test, that those in **CIRP** exceed those in **P** at the 10% level in a one-tailed test only, and that contributions don't differ significantly among any other pair among the four treatments, including the comparison between **CIP** and **CIRP**.[12] A preliminary conclusion is that the treatments with redistributive punishment especially are quite successful at raising contributions. Two potential explanations—(1) that there are added incentives to contribute more due to rewards, and (2) that there is less perverse punishment of high contributors—will be explored later. The **CIP** treatment also raises contributions, albeit insignificantly, allaying fears that subjects might not engage in monitoring.

Result 1. Both making punishment redistributive and making monitoring rather than punishment costly lead to higher contributions to the public good, with the difference attributable to the first factor being statistically significant.

[11] We graph the treatments in pairs to increase visibility, pairing the treatments on the left side of Table 2 and those on the right side of that table because, being distinguished by the value of α_2 only (versus differentiation of both α_1 and α_3 for left/right comparisons in the table), the members of these pairings are more readily comparable.

[12] The difference in contributions between the **CIRP** and the **P** treatment is significant at the 11% level in a two-tailed test, hence at close to the 5.5% level in a one-tailed test. The difference between the **CIP** and the **CIRP** treatments is insignificant, because despite the averaged trends shown in Figure 1b, several groups in the **CIP** treatment achieved higher average contributions than several groups in the **CIRP** treatment.

	P	**RP**	**CIP**	**CIRP**
Average per period contribution	7.09	9.03	8.33	8.88
Average per period earning	12.88	15.18	13.20	15.09
Adjusted per period earning	12.88	14.20	13.00	13.40

Table 3. Average contributions and earnings, by treatment. The definition of adjusted earnings is given in the text.

Earnings

Figure 2a shows average earnings by period in the **P** and **RP** treatments, while Figure 2b shows average earnings by period in the **CIP** and **CIRP** treatments, and the second row of Table 3 indicates earnings in each treatment averaged over all 20 periods. Not surprisingly, earnings in the treatments with redistributive punishment exceed those in the treatments with only reductions, and this is true in every period. In this case, Mann-Whitney tests show earnings to be higher in **RP** than in **P** and higher in **CIRP** than in **CIP**, in both cases significant at the 1% level in two-tailed tests. Earnings are also significantly higher in the **RP** treatment than in the **CIP** treatment and in the **CIRP** treatment than in the **P** treatment (in other words, earnings are higher in each treatment with redistributive punishment than in each treatment with only reductions). Earnings in the two treatments with only reductions (**CIP** and **P**) and those in the two treatments with redistributive punishment (**RP** and **CIRP**) do not significantly differ from each other.

We cannot be too impressed, however, by the superiority of the **RP** over the **P** and the **CIRP** over the **CIP** treatment in terms of total earnings, since such an outcome was to some extent engineered by us through having the punished funds be "burned" in the first of each pair of treatments but remain in the group (and simply change hands) in the second. We want to check, then, whether there is any efficiency difference if we control for these obvious resource differences. Similarly, we should check whether differences between the **CIRP** and **CIP** treatments, on the one hand, and the **RP** and **P** treatments, on the other, are due to the fact that the cost of information in the former is low compared to the potential cost of punishing in the latter.[13]

We accomplish both ends by adjusting earnings in the **RP**, **CIP** and **CIRP** treatments to put them on a "resource equivalent" footing with respect to each other and to the **P** treatment. Two adjustments are made in this exercise. First, the redistributive and the ordinary (non-redistributive) punishment treatments are put on equal footings by not counting earnings from transfers (rewards) when calculating adjusted earnings in the **RP** and **CIRP** treatments. Second, the costly monitoring treatments are put on the same footing as the costly punishment treatments by subtracting the monitoring costs and adding an *E*$0.25 cost to the punisher for each dollar of punishment imposed. The results, shown in the third row of Table 3, are lower adjusted than actual earnings in the **RP**, **CIP** and **CIRP** treatments, with the change largest in **RP** and **CIRP**. Even so, average adjusted earnings are higher in treatments with redistributive punishment or costly monitoring, but the earnings advantage over the **P** treatment looks large only in **RP**. Mann-Whitney tests confirm that this is the case: **RP** adjusted earnings are higher than those in each of the other treatments, significant at just short of the 5% level in 1-tailed tests, while no other adjusted earnings differences are significant. Thus, having punishments be redistributed has an incentive as well as a resource effect on efficiency in one pair of treatments (**RP** vs. **P**) but not in the other (**CIRP** vs. **CIP**). There is a small efficiency gain from making monitoring rather than punishing costly, but it is not consistent enough across groups to be significant in non-parametric tests, once the resource effect is controlled for.[14]

Result 2: Both making punishment redistributive and making monitoring rather than punishing costly lead to higher earnings, but only the first change leads to a significant (pure) incentive effect on earnings, and only in the costly punishment case.

[13] Although the *E*$1 cost of information was sufficiently high to dissuade the average subject from requesting it in over 75% of periods (see below), once a subject had the information up to *E*$10 of punishment could be given at no further cost, making the effective cost to *i* of giving *E*$1 of punishment to *j* as low as *E*$0.10 in the **CIP** and **CIRP** treatments versus E$0.25 in the **P** and **RP** treatments.

[14] We say "consistent across groups" because the Mann-Whitney test tells us not whether 13.00 is far enough above 12.88, say, for the difference in averages to be significant, but whether enough groups in **CIP** treatment earn more than enough groups in **P** treatment for the two to be likely to come from different distributions, something that can hold regardless of how large or small the numerical differences are.

Punishing and Rewarding Frequencies and Amounts

The sustained high contributions observed in all treatments suggests that many subjects anticipated receiving or actually received punishment when not contributing or when contributing little to their group account.[15] How often did punishment actually occur? In the **P**, **RP**, **CIP** and **CIRP** treatments a subject punished at least one other subject in his/her group in an average 28.8%, 23.9%, 19.1% and 23.2% of periods, respectively, and a subject was punished by one or more other subjects on average in 30.2%, 20.6%, 20.3% and 25.0% of periods, respectively. 78.1%, 76.6%, 81.3% and 82.8% of subjects punished at least one other subject at least once during their experiment session, and 81.3%, 71.9%, 90.6% and 96.9% were punished at least once.[16] The average number of dollars by which a subject *i* reduced the earnings of a subject *j* in a single instance of punishment was *E*$2.59, *E*$3.18, *E*$6.11 and *E*$5.06 in the respective treatments. Because *j* was often punished by more than one individual at a time, for instance during a period in which *j*'s contribution was much lower than the group's average, and because in the **RP** and **CIRP** treatments a subject *j* might receive both reward and punishment (from different team members) in the same period, the total number of dollars by which *j*'s earnings were reduced in a typical instance of being punished differs from the typical amount of punishment given by any one subject; the average (net) punishment received is *E*$3.69, *E*$5.10, *E*$8.39 and *E*$6.39, respectively. In the **RP** and **CIRP** treatments, each dollar taken from an individual *j* was given to another individual, *k*. On average, subjects had their earnings added to (on balance) in 25.2% of periods in **RP** and in 26.3% of periods in **CIRP**, with the total (net) amount added in a given instance (summing together multiple transfers and subtracting off any punishments, when applicable) being *E*$3.49 in **RP** and *E*$5.33 in **CIRP.**

[15] Note that fear of punishment can have not only the direct effect of dissuading those inclined to free-ride from doing so, but also the indirect effect of inducing conditional cooperators to contribute more (or to not reduce their contributions), since the direct effect leads to greater expected contributions by others.

[16] Recalling that a subject in **RP** or **CIRP** treatment could be targeted for punishment by one or even two group members, yet receive no indication of having been punished because others rewarded them still more in the same period, we recalculated the relevant percentages for those treatments to reflect the proportion who received *net* punishment, of which they could be aware, on at least one occasion. The figure for **RP** falls slightly from 71.9% to 70.3%, but the figure for the **CIRP** treatment turns out to be unaffected.

We ask two questions about the effects of treatment on the total amount of punishing:

1. Did subjects do more punishing when punishment was redistributive and was accordingly less directly costly to efficiency, as would be predicted by theories in which agents value aggregate earnings (e.g., Charness and Rabin, 2002)?

The answer is: No. Total dollars of punishment given are no different in the **RP** treatment than in the **P** treatment, and no different in the **CIRP** treatment than in the **CIP** treatment, according to Mann-Whitney tests. In the context of our public goods game, where subjects' attentions seem to be focused on how much to contribute, who is free riding, and who to punish, aggregate earnings don't seem to influence the punishment decision.

Result 3a. Punishers punish no less when the social cost of punishing is higher (i.e. the money taken is "burned").

2. Did subjects do more punishing when the average and marginal cost of punishing is lower?

The answer is: Yes. Although the total number of times that a subject punishes another subject is not significantly different among the treatments, the total number of *dollars* by which subjects are punished is significantly greater in the **CIP** and **CIRP** treatments (where one can give up to 10 dollars of punishment with no additional charge after paying *E*\$1 for contribution information) than in the **P** and **RP** treatments (in which each punishment dollar costs the punisher *E*\$0.25).[17] The finding that more punishment is given when it is less expensive is consistent with Anderson and Putterman (2005), Carpenter (2007), and Nikiforakis and Normann (forthcoming).

Result 3b. Punishers punish less when the private cost of punishing is higher.

[17] According to Mann-Whitney tests, total dollars of reductions by punishment are greater in **CIP** than in **P**, although significant only at the 10% level and only in a one-tailed test. Dollars of punishment in **CIRP** are greater than those in **RP** significant at the 10% level in a two-tailed test and at the 5% level in a one-tailed test. Punishment dollars in **CIRP** also exceed those in **P** at the 10% level in a one-tailed test, and those in **CIP** exceed those in RP at the 10% level in a two-tailed test and at the 5% level in a one-tailed test.

Monitoring

To our knowledge, the **CIP** and **CIRP** treatments are the first in the VCM literature to introduce costly monitoring as a distinct choice. We find that subjects paid for information about others' contributions (and for the right to punish or redistribute earnings) in an average of 20.7% of periods in the **CIP** treatment and 24.1% of periods in the **CIRP** treatment. The total number of requests for information does not significantly differ between these two treatments in a Mann-Whitney test using group-level observations. A subject proceeded to punish at least one other subject or to punish and reward at least one pair of subjects 89.0% of the times that he or she purchased the contribution information in the **CIP** treatment and 94.5% of such times in the **CIRP** treatment. Comparing treatments in which monitoring was costly to their counterpart treatments in which it was not, there are no significant differences among the treatments in the number of events in which some *i* punished some *j*. This suggests that having to pay for information is not in itself a deterrent to engaging in punishment.

Result 4. Making monitoring rather than punishing costly leads to no change in frequency of punishment.

Who was Punished and Rewarded?

Following Fehr and Gächter (2000), we estimate regressions to investigate whether subjects were singled out for punishment as a function of having contributed less or more than others in their group. Define subject *i*'s absolute negative and positive deviations from the average of others' contributions as:

$$\begin{array}{l}\text{Absolute}\\\text{Negative}\\\text{Deviation}\end{array} = \begin{cases} |C_i - \overline{C}_{-i}| & \text{if } C_i < \overline{C}_{-i} \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad \begin{array}{l}\text{Positive}\\\text{Deviation}\end{array} = \begin{cases} |C_i - \overline{C}_{-i}| & \text{if } C_i > \overline{C}_{-i} \\ 0 & \text{otherwise} \end{cases}$$

where $\overline{C}_{-i} = \dfrac{\sum_{j \neq i} C_j}{3}$ is the average of others' constributions.

In the regression equations, the dependent variable is the number of dollars of punishment received by subject *i* in period *t*, and the explanatory variables are the Absolute Negative Deviation, Positive Deviation, the average contributed by others in *i*'s group, and group and period fixed effects. Note that the rewards that may also have been given to *i* in the **RP** and **CIRP** treatments are not considered here, so subjects who were actually net reward recipients may still have positive values of the dependent variable in this regression. We estimate the regressions both by OLS and by Tobit, which accounts for the fact that the amount of punishment cannot be less than 0 and that there are many observations of 0 punishment. All regressions include period and group fixed effects.[18] Results are shown in Table 4.

Treatment	**P**		**RP**		**CIP**		**CIRP**	
Reg. type	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit
Abs. Neg. Deviation	0.625*** (0.028)	1.166*** (0.069)	0.941*** (0.029)	1.762*** (0.094)	1.803*** (0.053)	3.130*** (0.176)	1.293*** (0.051)	2.406*** (0.150)
Positive Deviation	0.064 (0.040)	-0.080 (0.107)	0.203** (0.084)	-0.059 (0.277)	0.153* (0.091)	-1.158*** (0.427)	0.190* (0.115)	-0.866** (0.426)
Av. Oth. Contr.	-0.099** (0.041)	-0.346*** (0.103)	0.214*** (0.076)	-0.097 (0.253)	-0.017 (0.089)	-1.150*** (0.341)	0.111 (0.116)	-0.517 (0.386)
# Obs.	1280	1280	1280	1280	1280	1280	1280	1280
F stat.	24.49		42.05		45.17		25.77	
χ^2		724.15		675.32		590.71		480.64
Adj. R^2	0.405		0.543		0.561		0.418	
Pseudo R^2		0.215		0.254		0.204		0.149

Table 4. **Determinants of punishment received.** Dependent variable: *E*$s of punishment received by subject *j*. Numbers in parentheses are t-statistics. All regressions include group and period fixed effects, not shown. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Both the OLS and the Tobit estimates for all four treatments show that absolute negative deviations have a highly significant positive effect on punishment. The coefficient on positive deviation is positive in all of the OLS estimates but negative in all

[18] A drawback of the Tobit model is that unconditional fixed effects estimates can be biased. Fixed effects for individuals are not used because the data are organized at the level of the recipient of punishment and those giving the punishment could not identify the recipient as the same individual from one period to the next.

four Tobit estimates, with some statistically significant coefficients associated with both signs. While some weight should also be accorded to the Tobit estimates, the positive coefficients in the OLS estimates are hinting at the presence of perverse punishment, on which we report in greater detail later on.

To see whether punishment was mainly motivated by the hope of increasing others' contributions in subsequent periods, we also estimated variants of all regressions shown in Table 4 adding an interaction term between a period 20 dummy variable and each of the first two explanatory variables. According to the OLS results, subjects are punished more for each dollar of absolute negative deviation in period 20 of the **P** and **CIRP** treatments, and less for each dollar of absolute negative deviation in period 20 of the **RP** and **CIP** treatments, as compared with the other nineteen periods. The corresponding Tobit estimates show no significant period 20 differences. These results (which are not shown, to save space) provide little support for the idea that punishment is mainly motivated by strategic considerations, and are more consistent with punishment being a preference-based or emotional response to free riding, as suggested by Casari and Luini (2005) and Hopfensitz and Reuben (2007), among others.

Result 5. Punishment is aimed disproportionately at groups' lower contributors, and is not lessened by the absence of a strategic motivation in the last period.

*Who was rewarded and what determined net reward/punishment in **RP** and **CIRP**?*

We investigate the determinants of which subjects had their earnings added to using a parallel specification to that of Table 4, with results shown in Table 5. In the first four columns, as in Table 4, the dependent variable doesn't account for possibly simultaneous punishments: it is the "gross reward," not the "reward net of any countervailing punishment" that may have been given to *j*. In the four right-hand columns, however, we show the results of similar regressions, but this time the dependent variable is *net* reward, which accounts for both reward and punishment but takes a value of zero for a subject who received only punishment or more punishment than reward. It is this dependent variable that corresponds to reward according to *j*'s information set. In the first four regressions, we find that rewards seem well targeted for efficiency, because there is a significant positive effect of a subject's positive deviation from others' average contribution (subjects who contribute more tend to be rewarded more) and a significant negative effect of absolute negative deviation (subjects who contribute less are rewarded significantly less). The less are others' contributions, the less is one rewarded, *ceteris paribus*—perhaps reflecting that there was less rewarding in groups with fewer high contributors. The four right-hand regressions provide an even stronger indication of efficient targeting.

Dep. Var. Ind. Var.	**Reward Received**				**Net Reward Received (Reward – Punishment)**			
Treatment	**RP**		**CIRP**		**RP**		**CIRP**	
Reg. type	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit
Abs. Neg. Deviation	-.085*** (0.028)	-.559*** (0.127)	-0.109** (0.051)	-1.286*** (0.281)	-.109*** (.027)	-1.307*** (.300)	-.106** (.0507)	-1.495*** (.344)
Positive Deviation	0.434*** (0.081)	0.675*** (0.231)	0.656*** (0.115)	1.077*** (0.324)	.419*** (.078)	.581** (.287)	.616*** (.116)	1.103*** (.367)
Av. Oth. Contr.	-.249*** (0.073)	-.927*** (0.225)	-.390*** (0.116)	-1.463*** (0.332)	-.204*** (.071)	-.973*** (.278)	-.322*** (.117)	-1.434*** (.377)
# Obs.	1280	1280	1280	1280	1280	1280	1280	1280
F stat.	16.17		11.79		15.47		8.27	
χ^2		584.08		390.35		563.59		300.41
Adj. R^2	0.305		0.238		0.2951		0.1739	
Pseudo R^2		0.188		0.111		0.1940		0.0931

Table 5. Determinants of rewards and of net rewards. Dependent variables: left, *E*$s of additions received by subject *j*; right, same minus *E*$s of punishment received by *j*. All regressions include group and period fixed effects, not shown. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

As with punishment, we also checked whether the pattern and amount of rewards was substantially changed in the last period, by re-estimating the regressions of Table 5 adding as an explanatory variable a period 20 dummy variable multiplied by absolute negative deviation and another one multiplied by positive deviation. The resulting coefficients on the positive deviation interaction terms indicate that there was a strong last period effect in the **RP** treatment, fully negating the earlier observed tendency to reward higher contributions, but there was no effect of the last period on rewarding higher contributions in the **CIRP** treatment. As for the negative deviation interactions, the negative impact of *lower* contributions on rewards is significantly *enhanced* in period 20 of **RP** but seems to disappear in period 20 of **CIRP**. Differential rewards do, therefore, persist into the last period in both treatments, although in different ways. (These results are not shown to save space.)

Result 6. Rewards are given disproportionately to groups' higher contributors, and the rewarding of high more than low contributors continues in the last period.

When did Subjects Monitor, and Who Monitored?

Table 6 shows the results of two probit regressions, one for each of the treatments in which a subject learned individuals' contributions only if she chose to pay *E*\$1. The dependent variable takes the value 1 if the individual paid for the information on contributions in the period in question and 0 otherwise. Individual and period fixed effects are included but not shown. The specification is designed to let us investigate two issues: (a) were a group's higher contributors or its lower contributors more likely to request the information? (b) was a subject more likely to request the information when the total contribution of others was smaller (suggesting that there might be free riders to punish)? The specification also closely parallels those of the previous two tables. The results indicate that (a) in both the **CIP** and the **CIRP** treatment, the further above the average contribution of others in her group was a given subject's contribution, the more likely was she to monitor (pay for information), as shown by the significant positive coefficient on positive deviation, and (b) the lower was the average contribution of others, the more likely was a subject to monitor, which is consistent with a desire to punish free riders (i.e., monitoring becomes unnecessary in groups achieving high levels of cooperation).

Treatment	**CIP**	**CIRP**
Abs. Neg. Deviation	.088*** (.033)	.008 (.037)
Positive Deviation	.486*** (.065)	.530*** (.088)
Av. Oth. Contr.	-.216*** (.055)	-.125 (.081)
# Obs.	1100	1080
χ^2	461.35	467.69
Pseudo R^2	0.3798	0.3622

Table 6. **Determinants of monitoring.** Probit regressions. Dependent variable: subject requested contribution information. All regressions include individual and period fixed effects, not shown. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Result 7. Higher contributors monitored more than others, and there was more monitoring when average contributions were lower.

How did punishment affect contributions?

One obvious way in which inclusion of a punishment (or punishment and reward) stage can lead to higher contributions is by causing those targeted to change their behaviors. We investigate this by estimating regressions in which subject *i*'s change of contribution from period *t* to period *t*+1 is the dependent variable, and the explanatory variables measure the amount of punishment given to *i* in period *t*, differentiated to account for how *i*'s contribution compared to that of others in the group in that period. Specifically, we interact the total (in **P** and **CIP**) or net (in **RP** and **CIRP**) punishment dollars received (where net punishment can be negative if *i* was on balance rewarded) by subject *i* in period *t* with dummy variables for the recipient's contribution ranking. We define dummy variable h as 1 if *i*'s contribution was the highest in the group in period *t*, 0 otherwise; dummy variable l as 1 if *i* made the lowest contribution in the group, 0 otherwise; and similarly for dummy variables "2nd"and "3rd," for second highest and for third highest contributor.[19] We also include the dummy variables h, 3rd, and l as non-interacted or "stand alone" terms, because changes in contributions could occur independently of receiving punishment[20] (for example, a highest contributor might tend to reduce her contribution, *ceteris paribus*), and we include individual and period fixed effects. We estimate the regression for each treatment both using OLS and Tobit.[21] Results are shown in Table 7.

The regression results suggest that punishment of lowest contributors tended to be followed by increases in their contributions (all coefficients on $l*R_{ji}$ are positive and significant) while punishment of highest contributors led to reductions in their contributions (significant negative coefficients on $h*R_{ji}$ in the regressions for the **P** and **CIRP** treatments, insignificant coefficients for the **RP** and **CIP** treatments).[22] Punishment of 2nd and 3rd highest contributors seems to have also led to those individuals increasing their contributions in the **P** treatment, but shows no significant effect in the other three treatments.

Result 8. A group's lower (highest) contributors tend to increase (reduce) their contributions after being punished.

[19] If the group has only three contribution levels in the period, we code the highest and lowest contributions as h and l and we code the middle contribution as 2nd if it is above or equal to the average and as 3rd if below the average. If there are only two contribution levels, we code them as h and l. If there is only one contribution level, we code it as h if it is 10, l if 0, and otherwise we drop the observations for that group and period.

[20] The uninteracted dummy for 2nd highest contributor is the omitted category.

[21] The Tobit regressions are estimated using the cnreg command in Stata, treating the dependent variable—$C_{i,t+1}$ - $C_{i,t}$—as possibly left-censored if $C_{t+1} = 0$ and as possibly right-censored if $C_{t+1} = 10$.

[22] These are further confirmations of the negative and hence efficiency-reducing effect of punishment of the highest contributor that led Cinyabuguma *et al*. (2006) to call it perverse punishment. Note that because R_{ji} stands for net punishment (which can take a negative value if the subject received a net reward), the negative impact of punishment on highest contributors' contributions and the positive impact on lowest contributors' contributions incorporates instances in the **RP** and **CIRP** treatments in which rewarded high contributors increased their contributions.

Treatment	**P**		**RP**		**CIP**		**CIRP**	
Reg. type	OLS	Tobit	OLS	Tobit	OLS	Tobit	OLS	Tobit
H	-.962** (.374)	-.108 (.806)	-1.35*** (.443)	2.522 (2.164)	-.667** (.302)	.652 (.718)	-1.179*** (.260)	-.814 (.542)
3^{rd}	2.302*** (.478)	2.561** (1.018)	2.199*** (.640)	1.260 (2.997)	1.387*** (.409)	1.855* (.947)	.547* (.324)	.345 (.651)
L	3.254*** (.416)	2.838*** (.918)	2.096*** (.509)	2.263 (2.527)	1.541*** (.356)	.994 (.849)	1.329*** (.294)	1.686*** (.606)
$h*R_{ji}$	-.184** (.093)	-.379* (.219)	.028 (.030)	.243 (.232)	-.007 (.043)	-.152 (.176)	-.059*** (.016)	-.294*** (.060)
$2^{nd}*R_{ji}$	.360** (.152)	.549* (.318)	-.031 (.108)	.133 (.474)	-.056 (.130)	.022 (.320)	-.035 (.068)	.011 (.138)
$3^{rd}*R_{ji}$	.301*** (.110)	.466* (.247)	-.008 (.127)	.461 (.612)	.076 (.103)	-.093 (.247)	-.011 (.056)	-.008 (.114)
$l*R_{ji}$	.404*** (.053)	.604*** (.136)	.283*** (.037)	.574** (.244)	.191*** (.022)	.236*** (.055)	.186*** (.023)	.239*** (.051)
# Obs.	1212	1212	1216	1216	1216	1216	1216	1216
F stat.	7.54		6.46		4.91		6.12	
χ^2		614.92		369.31		488.03		453.03
Adj. R^2	0.3221		0.2833		0.2206		0.2705	
Pseudo R^2		0.1652		0.1970		0.1722		0.1562

Table 7. Determinants of change in contribution. All regressions include individual and period fixed effects, not shown. Dependent variable: $C_{i,t+1}$ - $C_{i,t}$. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

How common are perverse punishments and rewards?

A major conjecture motivating our experiment was that if earnings taken from punished group members had to be distributed to other members, this might discourage perverse punishment while encouraging pro-social punishment of low contributors, thus increasing efficiency. What is the incidence of perverse punishment in the **RP** and **CIRP** treatments and how does it compare to that in the **P** and **CIP** treatments?

Table 8 lists the percentage of punishment events in which a group's highest contributor was punished, the percentage of punishment dollars given to a group's highest contributor, the percentage of punishment events in which a contributor of more than the group average was punished, and the percentage of punishment dollars given to contributors of more than the group average.[23] The left and right halves of the table differ in two respects. First, entries in the left half of the table differ from those in the right half for the **RP** and **CIRP** treatments because on the left we count punishment without subtracting off countervailing rewards to the same individual, while on the right we net out the rewards, thus causing any case in which the targeted individual received more reward than punishment not to appear as a punishment event. Second, entries on the left treat each time a subject *j* was punished by some subject *i* as a separate event, while entries on the right treat receipt of punishment by *j* in a given period as one event regardless of whether the punishment came from one other subject only, or from two or three other subjects. The counts on the left thus better reflect acts of punishing as seen

Method:	"Gross," with *i,j* interaction as event (Punisher's view)				"Net," with receipt by *j* as event (Recipient's view)			
Percentage of punishment	**P**	**RP**	**CIP**	**CIRP**	**P**	**RP**	**CIP**	**CIRP**
1) events, to highest Contributor	17.2	14.7	9.2	15.1	22.7	11.8	12.3	14.2
2) dollars, to highest contributor	17.2	7.9	7.1	14.3	17.2	3.5	7.1	9.9
3) events, to > average contributors	24.6	16.3	20.2	18.1	32.0	13.6	23.8	16.0
4) dollars, to > average contributors	24.2	9.2	11.0	15.8	24.2	4.8	11.0	10.9

Table 8. Frequency of "perverse" punishment, by treatment and measure.

[23] For purposes of the table, we drop from the counts of both the total instances of punishment and the instances of perverse punishment cases in which all group members contributed 10. Such cases accounted for less than 1% of punishment events in the **P** treatment, but for 13 to 18% of punishment events in the remaining three treatments (where periods with uniformly full contributions were more common). We include the less common punishments in the event of the other occasional tie, that in which all subjects contributed 0, treating it as equivalent to punishing a lowest contributor, and hence as never perverse.

from the punishers' points of view, whereas the counts on the right reflect instances of being punished as seen from the recipients' standpoints.[24]

Judged in terms of punisher behavior (the left side of the table), our conjecture that requiring punished dollars to be given to other group members would reduce the proportion of punishment targeted perversely at high contributors receives only limited support. In particular, for the pair of treatments without monitoring cost—the treatments resembling Fehr and Gächter (2000) and its replications—the share of punishment that is perversely targeted is smaller with redistribution (i.e., in **RP**) than without (in **P**), by all measures. Our conjecture is thus supported in this more familiar setting. But for the pair of treatments with monitoring cost, more perverse punishment is given when punishment is redistributive (in **CIRP**) than when it is not (in **CIP**), by all but one measure.

On the other hand, the right side of Table 8, which views the matter from the operationally more relevant standpoint of the recipient of punishment, shows that making punishment redistributive was in fact successful at generating better incentives to contribute. The proportion of punishment events aimed at highest contributors falls from 17.2% in the **P** treatment to 3.5% in the **RP** and 9.9% in the **CIRP** treatment. The proportions of punishment events and dollars going to above-average contributors are lower in **RP** than in **P** and lower in **CIRP** than in **CIP**, although the dollar difference is negligible for the latter comparison. Only the dollars to highest contributor measure fails to support the idea that making punishment redistributive reduces perverse punishment, and only (again) in the **CIRP** versus **CIP** comparison.

Even this partial failure of making punishment redistributive to reduce the relative shares of perverse punishment in the treatments with costly information can be looked at in a very different light, however. That is, the proportions of punishment which are perverse are strikingly low in *both* costly information treatments, as compared to the **P** treatment and the other experimental treatments cited earlier. It seems that making information costly has itself reduced punishment of high contributors because (a) it is low

[24] Recall that the information conditions were such that the recipient learned only the net outcome of punishment or reward, not the combination of acts by his/her three fellow group members which brought it about. Thus, for instance, a high contributor who was punished by one group member but rewarded a greater amount by others received no indication that a perverse punishment event had occurred, making the right-hand accounting the more appropriate one for gauging the incentive implications of punishment for contribution decisions.

contributors who tend to punish high ones, but (b) low contributors did less monitoring in treatments with costly information. Evidence for (b) is seen in Table 6. For (a), consider Table 9, which shows what proportion of punishment given to highest contributors came from lowest and 2nd-to-lowest ones. In the **P** treatment, with neither redistributive punishment nor costly monitoring, almost two-thirds of punishment given to highest contributors comes from lowest ones, with 85% of such perverse punishment coming from either lowest or 2nd-to-lowest contributors.

Treatmet	Total # of events in which punishment is given to a highest contributor (1)	# of column (1) events in which the punisher was the lowest contributor (2)	Proportion of col. (1) events in which the punisher was the lowest contributor (3)	# of events in which 2nd-to-lowest contributor punished highest contributor (4)	Proportion of column (1) events in which punisher was 2nd-to-lowest contributor[25] (5)
P	95	59	62.11%	22	23.16% 85
RP	57	32	56.14%	8	14.04% 70
CI	33	26	78.78%	5	15.15% 95
CIRP	60	43	71.66%	6	10.00% 81

Table 9. Share of punishment of highest contributors given by lowest and 2nd-to-lowest contributors.

Result 9a. Making punishment redistributive has no clear effect on the targeting of punishment at high contributors by individual punishers, but high contributors receive significantly less* net *punishment in redistributive treatments, thus reducing their disincentive to contribute more.

Result 9b. Making contribution information costly reduces perverse punishment by reducing the representation of low contributors among those punishing.

[25] In this exercise, a subject is called "2nd-to-lowest contributor" if her contribution was third of four different contribution levels or if it was the middle one of three different contribution levels.

Was it harder to free ride in the new designs?

Was free riding thwarted more successfully in the new designs. If by "free riding" we mean contributing little or nothing, then the contribution differences summarized earlier and in Figures 1a and 1b suffice to answer in the affirmative. A more rigorous definition of free riding, however, is that an individual *j* who contributes less than another individual *i successfully* free rides only when *j* ends up earning more than *i* taking into account punishments, rewards, and their costs (including costs of monitoring). We counted all cases in our four treatments in which an individual *j* contributed less than another individual *i* in the same group and period, and we calculated the proportion of those cases in which *j* ended up earning more than *i* in that period. The results are striking. Table 10 shows that whereas only a little more than 20% of potential free riding was eliminated by punishment in the **P** treatment, about half of potential free riding was eliminatd by punishment, or by punishment and rewards, in the **RP**, **CIP**, and **CIRP** treatments.

Treatment	Of events in which $C_j < C_i$, in what proportion was $y_j > y_i$?
P	806/1016 = 0.79
RP	271/566 = 0.48
CI	427/803 = 0.53
CIRP	410/855 = 0.48

Table 10. Proportion of potential free-riding events that *succeeded*, by treatment.

Result 10. Free-riding is more effectively thwarted by punishment in the treatments with costly monitoring and/or redistributive punishment.

5. Discussion and Conclusions

We introduced two new design elements into prototypical public goods or VCM-with-punishment experiments. The first element is a requirement that any money subtracted from a group member's earnings be awarded to some other group member or members. The second makes accessing information about what each individual assigned to the group account a costly choice. Both innovations can be motivated by a concern with the targeting of punishment. The introduction of costly monitoring is further motivated by the fact that it is a relevant element of many real-world collective action problems, but thus far not accounted for in public goods experiments.

We found that both redistributive punishment and costly monitoring reduced the incidence of misdirected or perverse (net) punishment and thus increased contributions to the public good. Redistributive punishment also increased earnings, with or without costly monitoring, although earnings are significantly higher after differences in resource costs are netted out only in our redistributive punishment treatment with free information, **RP**. Making punishment redistributive appears to have reduced the impulse to punish perversely in the treatments without monitoring cost, and it reduced the net outcome of perverse punishment in both treatments because what perverse punishing did occur was often out-weighed by the efficiency-enhancing pattern of rewards. Making monitoring costly appears to have reduced perverse punishment because it disproportionately removed low contributors from the set of potential punishers, and because it is mainly low contributors who punish perversely.

In the real world, punishment of free riders and rewarding of cooperators are often observed in tandem, including in cases where punishment and rewards take the forms of social disapproval and approval. Fines levied on norm violators can be used to finance goods that cooperators value, although it's difficult to think of cases in which pecuniary or material punishment is both decentralized (that is, left to individuals) and redistributive. Centralized redistribution from low to high contributors is automatically effected by the Falkinger mechanism studied by Falkinger, Fehr, Gächter, and Winter-Ebmer (2000), and that pattern is emulated in the recent experiment with cost-free decentralized redistribution by Sausgruber and Tyran (2007).

Studying monitoring in both lab and field is an important area for future research. Our exploratory initial treatments made punishing itself free once information was paid for, but it seems more realistic to make both choices costly. More complex information structures, for instance ones in which the information obtained can be imperfect or in which the accuracy of that information increases with additional expenditure, are also worth investigating.

At a general level, our study provides more evidence, consistent with Ostrom, Gardner and Walker (1992), Fehr and Gächter (2000), Page, Putterman and Unel (2005), Gürerk, Ő., B. Irlenbusch and B. Rockenbach (2006) and others, that many individuals' choices display positive and negative reciprocity (conditionally cooperating, and incurring cost to punish free riders), and that the existence of such individuals can ameliorate free rider problems by rendering it best for selfish individuals to cooperate and by supporting, in this way, conditional cooperators' inclinations to contribute. However, we also find a subset of "contrary" individuals who perversely punish cooperators, reducing the incentives of the latter to voluntarily contribute to a public good. Our experiment shows that reducing the presence of such individuals in the pool of potential punishers by making monitoring costly, and reducing their net effect on those targeted by making punishing redistributive, are both effective ways of reducing their negative impact.

References

Alchian, Armen and Harold Demsetz, 1972, "Production, Information Costs, and Economic Organization," *American Economic Review,* 62**,** 777-795.

Anderson, Christopher M. and Louis Putterman, 2006, "Do Non-strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism," with Christopher M. Anderson, *Games and Economic Behavior* 54 (1): 1-24.

Bochet, Olivier, Talbot Page and Louis Putterman, 2006, "Communication and Punishment in Voluntary Contribution Experiments," *Journal of Economic Behavior and Organization* 60: 11-26.

Botelho, Anabela, Glenn Harrison, Ligia M. Costa Pinto and Elisabet E. Rutström, 2005, "Social Norms and Social Choice," unpublished paper, Dept. of Economics, University of Central Florida.

Carpenter, Jeffrey, 2007, "The Demand for Punishment," *Journal of Economic Behavior and Organization* 62 (4): 522-42.

Carpenter, Jeffrey and Peter Matthews, 2002, "Social reciprocity," Middlebury College Department of Economics Working Paper #29.

Casari, Marco and Luigi Luini, 2005, "Group Cooperation Under Alternative Punishment Institutions: An Experiment" Working Paper, Department of Economics, University of Siena.

Casari, Marco and Charles R. Plott, 2003, "Decentralized Management of Common Property Resources: Experiments with a Centuries-Old Institution," *Journal of Economic Behavior and Organization* 52: 217-47.

Charness, Gary and Matthew Rabin, 2002, "Understanding Social Preferences with Simple Tests," Quarterly Journal of Economics 117 (3): 817-69.

Cinyabuguma, Matthias, Talbot Page and Louis Putterman, 2004, "On Perverse and Second-Order Punishment in Public Goods Experiments with Decentralized Sanctioning," Brown University Department of Economics Working Paper 2004-12.

Cinyabugama, Matthias, Talbot Page and Louis Putterman, 2006, "Can Second-Order Punishment Deter Perverse Punishment?" *Experimental Economics* 9: 265-79.

Ertan, Arhan, Talbot Page and Louis Putterman, 2006, "Can Endogenously Chosen Institutions Mitigate the Free-Rider Problem and Reduce Perverse Punishment?" Brown University Department of Economics Working Paper 2005-13, revised.

Falkinger, Josef, Ernst Fehr, Simon Gächter, and Rudolf Winter-Ebmer, 2000, “A Simple Mechanism for the Efficient Provision of Public Goods: Experimental Evidence,” *American Economic Review* 90: 247-264.

Fehr, Ernst and Simon Gächter, 2000, “Cooperation and Punishment,” *American Economic Review* 90: 980-94.

Fehr, Ernst and Klaus Schmidt, 1999, “A Theory of Fairness, Competition and Cooperation,” *Quarterly Journal of Economics* 114 (3): 817-68.

Fehr, Ernst and Klaus Schmidt, 2003, “Theories of Fairness and Reciprocity – Evidence and Economic Applications,” in M. Dewatripont, et al., eds., *Advances in Economics and Econometrics, 8th World Congress of the Econometric Society v. 1*. Cambridge: Cambridge University Press.

Field, Alexander, 2001, *Altruistically Inclined? The Behavioral Sciences, Evolutionary Theory, and the Origins of Reciprocity*. Ann Arbor: University of Michigan Press.

Gächter, Simon and Benedikt Herrmann, 2005, “Norms of Cooperation among Urban and Rural Dwellers: Experimental Evidence from Russia,” unpublished paper, University of Nottingham.

Gächter, Simon, Benedikt Herrmann, and Christian Thöni, 2005, “Cross-cultural Differences in Norm Enforcement,” *Behavioral and Brain Sciences* 28: 822-3.

Grosse, Stefan, Louis Putterman and Bettina Rockenbach, 2007, “Monitoring in Teams: A Model and Experiment on the Central Monitor Hypothesis,” Brown University Department of Economics Working Paper 2007-04.

Gürerk, Őzgür, Bernd Irlenbusch and Bettina Rockenbach, 2005, “On the evolvement of institution choice in social dilemmas,” University of Erfurt, Working Paper.

Gürerk, Őzgür, Bernd Irlenbusch and Bettina Rockenbach, 2006, “The Competitive Advantage of Sanctioning Institutions,” *Science* 312 pp. 108-110, April 7 2006.

Hopfensitz, Astrid and Ernesto Reuben, 2007, “The Importance of Emotions for the Effectiveness of Social Punishment,” Discussion Paper 05-075, Tinbergen Institute.

Nikiforakis, Nikos, 2004, “Punishment and Counter-punishment in Public Goods Games: Can we Still Govern Ourselves?” unpublished paper, Royal Holloway University of London.

Nikiforakis, Nikos and Hans-Theo Normann, fortchoming, “A Comparative Statics Analysis of Punishment in Public-Good Experiments,” *Experimental Economics*.

Ostrom, Elinor, James Walker and Roy Gardner, 1992, "Covenants with and without a Sword: Self Governance is Possible." *American Political Science Review.* 86 (2): 404-416.

Page, Talbot, Louis Putterman and Bulent Unel, 2005, "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency," *Economic Journal* 115: 1032-53.

Sausgruber, Rupert and Jean-Robert Tyran, 2007, "Pure Redistribution and the Provision of Public Goods," *Economics Letters* 95 (3): 334-8.

Sefton, Martin, Robert Shupp and James Walker, 2002, "The Effect of Rewards and Sanctions in Provision of Public Goods," Working Paper, University of Nottingham and Indiana University.

Sutter, Matthias, Stefan Haigner, and Martin Kocher, 2005, "Choosing the stick or the carrot? – Endogenous institutional choice in social dilemma situations" unpublished paper, University of Cologne, University of Innsbruck and University of Amsterdam.

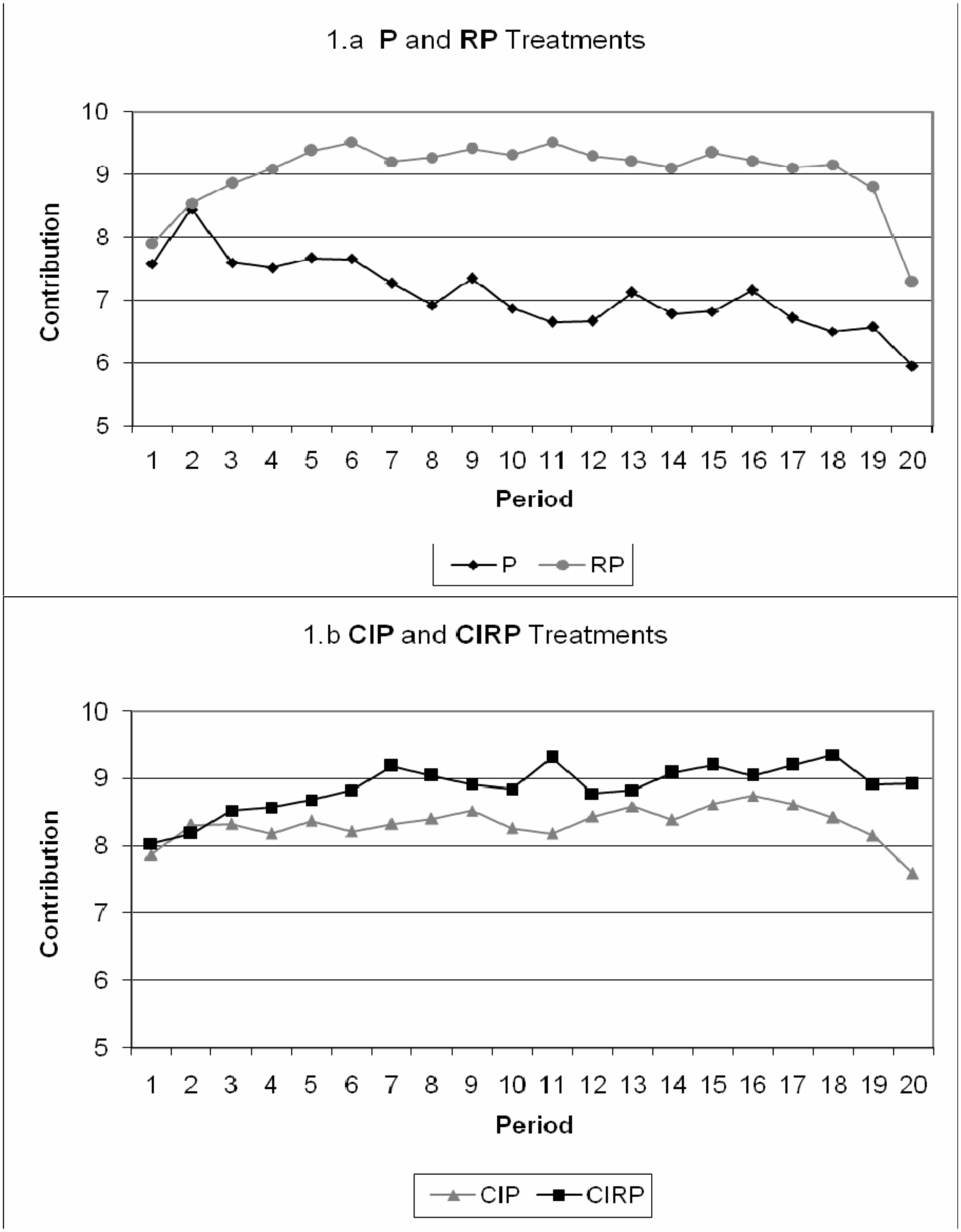


Figure 1. Average contribution by period. Possible contributions range from 0 to 10. The upper figure compares the **P** and **RP** treatments, the lower one the **CIP** and **CIRP** treatments.

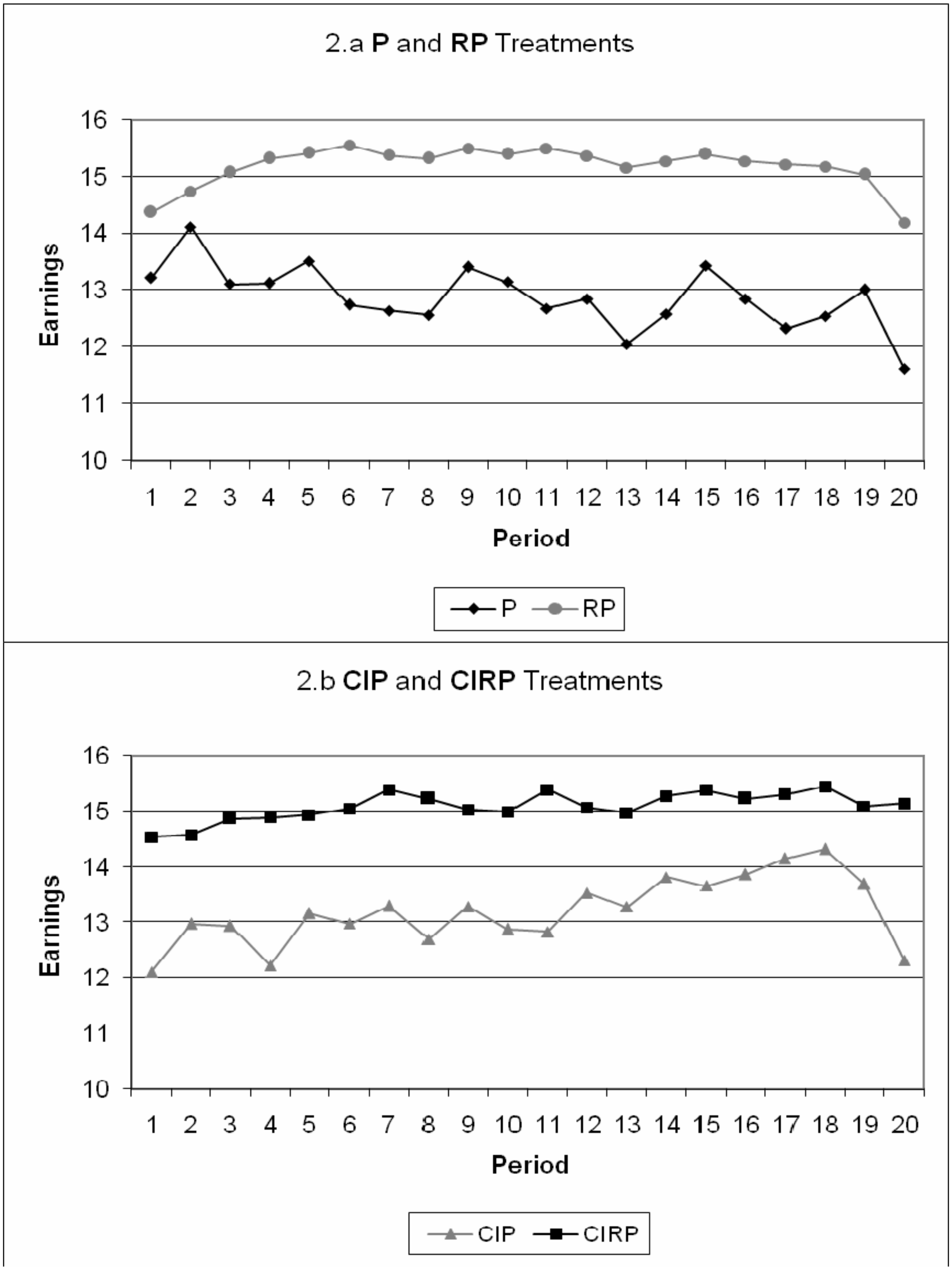
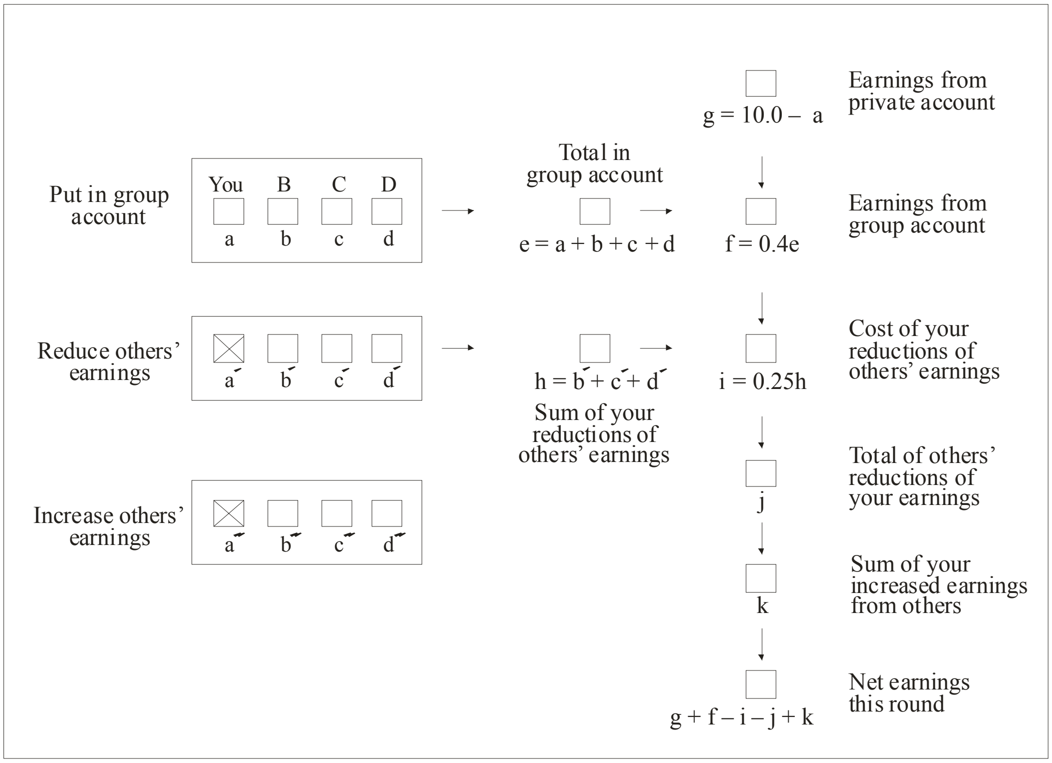


Figure 2. Average earnings by period. Possible earnings range from the Nash equilibrium level of 10 (or less due to punishment and monitoring costs) to social optimum earnings of 16. The upper figure compares the **P** and **RP** treatments, the lower one the **CIP** and **CIRP** treatments.

Appendix


Earnings from private account
g = 10.0 – a
Total in group account
Put in group account
You
B
C
D
a
b
c
d
e = a + b + c + d
f = 0.4e
Earnings from group account
Reduce others' earnings
a′
b′
c′
d′
h = b′ + c′ + d′
Sum of your reductions of others' earnings
i = 0.25h
Cost of your reductions of others' earnings
Total of others' reductions of your earnings
j
Increase others' earnings
a″
b″
c″
d″
Sum of your increased earnings from others
k
Net earnings this round
g + f – i – j + k

Screen design for entering contribution, punishment, and reward decisions, receiving information, and calculating net earnings