Punishing Free-Riders: how group size affects mutual monitoring and the provision of public goods

by

Jeffrey Carpenter

April, 2002

MIDDLEBURY COLLEGE ECONOMICS DISCUSSION PAPER NO. 02-06

DEPARTMENT OF ECONOMICS
MIDDLEBURY COLLEGE
MIDDLEBURY, VERMONT 05753

http://www.middlebury.edu/~econ

Punishing Free-Riders: how group size affects mutual monitoring and the provision of public goods*

April 12, 2002

Jeffrey P. Carpenter
Department of Economics
Middlebury College
Middlebury, VT 05753
(jpc@middlebury.edu)

Standard game theoretic models predict, based on subgame perfection, that public goods will not be provided even if agents are allowed to monitor free riders at some cost. Further, because punishment is not credible in these environments, this prediction is invariant to the size of groups. However, there is now substantial evidence that people are reciprocally motivated and will punish free riders, regardless of the material costs of doing so. To examine the implications of reciprocally-minded agents, we simulate an environment populated with the behavioral strategies often seen in the experimental lab and use the simulation to develop hypotheses that are more specific about why group size should matter when sanctions are allowed. We then test these hypotheses experimentally using the voluntary contribution mechanism. We examine whether the effect of group size is purely due to the number of group members or if information about other group members is what is important. We find large groups provide public goods at levels no less than small groups because punishment does not fall in large groups. However, hindrances to monitoring do reduce the provision of the public good.

Keywords: public goods, mutual monitoring, group size, experiment, simulation

JEL classifications: D63, D64, H41, C63, C91, C92

* Earlier drafts of this paper benefitted from the helpful comments of Steve Burks, Rachel Croson, Armin Falk, Simon Gaechter, Herb Gintis, Peter Matthews, and Eric

Verhoogen. This research is funded by the MacArthur Foundation Norms and Preferences network and the National Science Foundation (SES-0092953).

> *Two neighbors may agree to drain a meadow, which they possess in common: because it is easy for them to know each other's mind; and each must perceive, that the immediate consequence of his failing in his part, is the abandoning of the whole project. But it is very difficult, and indeed impossible, that a thousand persons should agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expense, and would lay the whole burden on the others.*
>
> *David Hume, A Treatise of Human Nature (London: J.M. Dent, 1952, II, 239)*

1. Introduction

Despite standard game-theoretic predictions concerning behavior in social dilemma situations such as the provision of a public good, experimental evidence shows that free riding is not the first impulse of economic agents. In fact, average contributions of between forty and sixty percent of an agent's endowment (despite the dominant strategy to contribute nothing) are common starting points with only a small minority of participants completely free riding in the first period of the standard, finitely repeated game (Davis and Holt [1993], Ledyard [1995]). These results have begun to have an impact on the study of economics. Theoretically, they have been an impetus behind new theories of more realistic, but nonstandard, preferences for reciprocity, fairness and altruism (Andreoni [1990], Rabin [1993], Fehr and Schmidt [1999], Falk and Fischbacher [1998], Bolton and Ockenfels [1999]). And, on a more practical level, this research has provided behavioral foundations for empirical results demonstrating that team production, among other forms of public goods provision, are effective without exogenous intervention (Prendergast [1999]).

Although contributions are positive to start, they do fall off significantly by the end of most experiments which, to some degree, vindicates theories based on subgame perfection. However, there have been a number of recent experiments showing that, if participants can *mutually monitor*, i.e. inspect the decisions made by other participants and sanction behavior deemed antisocial, the incentive to free ride is largely attenuated (Fehr and Gaechter [2000a], Bowles et al. [2001], Page and Putterman [2000], Sefton et al. [2000]). This result also runs contrary to standard theory because punishment, in these experiments, is costly and therefore participants can do better by free riding on the punishment done by others.[1]

Despite subgame perfection preventing punishment from being a component of the standard theory of public goods, equilibria do exist in which punishment is used to elicit positive contributions and, depending on the initial population distribution of strategies, evolutionary models have demonstrated that punishing strategies often survive (Sethi [1996], Carpenter et al. [2001]).[2] These models are much more in line with what one sees in the experimental lab.

Given people punish and evolutionary game theory provides microfoundations for this behavior, in this paper we are interested in the effects of group size on the ability of punishing strategies to survive and enforce contributions to a public good. If we assume people are willing to spend a certain amount of their resources to monitor other group members, there are (at least) two possible, but at this point ambiguous, effects of larger groups. First, larger groups force monitors to spread their resources thinner which might lead to more free riding; but notice there are also more people monitoring each free rider so it is not obvious whether the total amount of punishment each free rider receives will increase or decrease.

A second possible effect has little to do with the fact that there are more people in large groups. Instead, we might consider a logistical hypothesis which states that large groups are less able to elicit contributions because monitoring

[1] Further, being costly implies punishment is not credible.

[2] It is important to note that these models do not assume a preference for, or non-material benefits from, punishing free riders. Instead the models show that strategies with a punishing component survive when selection is based of the material payoff of strategies alone.

becomes more difficult. As groups grow it becomes harder for each individual to monitor everyone else and, as a result, free riding might become easier to hide and punishment becomes less of a deterrent. But, notice that if the "monitoring groups" overlap such that everyone is seen by some minimum fraction of the rest of the group, then punishment might still be sufficient to elicit contributions.

Because the effects of size in terms of numbers of agents and the information agents have about each other is unclear, in section 3 of this paper we present agent-based simulations which were run to derive clearer hypotheses concerning when group size might matter. Then, in sections 4 and 5 we discuss an experiment that was conducted to test our hypotheses with real agents. We begin, in section 2, by discussing other relevant research.

2. Previous and Related Experiments

The impact of group size on behavior in public goods games has already been studied experimentally. So far however, experiments have only been conducted that (1) use game environments that do not allow free riders to be sanctioned and (2) do not consider the logistical problems of large groups. The main result of this literature is that contributions do not fall as groups become larger and, if anything, they tend to increase. Isaac and Walker [1988] and Isaac et al. [1994] in a comprehensive series of experiments examine groups of size four, ten, forty, and one hundred participants. Considering the relatively smaller groups (four and ten persons), they find that size only matters when the return on the public good is low, in which case, contributions actually increase in large groups. When larger groups (forty and one hundred persons) are examined, the authors find that contributions increase relative to smaller groups and the effect is independent of the return on the public good.[3]

---

[3] For a plausible explanation of why contributions do not fall off as quickly in large groups see the discussion of the *minimum profitable coalition* in Davis and Holt [1993].

Interest is growing in the question of whether monitoring can increase contributions in social dilemma experiments. The first public goods experiment incorporating mutual monitoring was Fehr and Gaechter [2000a] who confirm the reciprocity conjecture generated by Andreoni [1988].[4] Andreoni showed that contributions decayed as would be expected by an equilibrium learning hypothesis, but contrary to learning also showed that when the game was restarted contributions returned to significantly high levels. One explanation of this result is that reciprocating participants withhold contributions to punish free riders, and are willing to wipe the slate clean when the experiment is restarted. More directly, Fehr & Gaechter show that when participants have some way, other than withholding contributions, to punish free riders, they do so and contributions increase.

The work of Fehr and Gaechter piqued the interest of other researchers who have confirmed their main result and extended the analysis in other interesting directions. Bowles et al. [2001] develop a team production model based on reciprocity which predicts punishment in equilibrium and test the model experimentally. The experiments substantiate the major hypothesis generated by the model - transferring residual claimancy to a team increases reciprocators' propensity to punish shirkers and this, in turn, increases the productive efficiency of team production. Page and Putterman [2000] also confirm that punishment is used to maintain or increase contributions. In addition, they examine the role of face-to-face communication which, from their preliminary analysis, seems to not effect contribution decisions in the presence of punishment. Another noteworthy study was conducted by Sefton et al. [2000]. Their contribution is to examine the relevance of rewards. In one treatment they allow both rewards and sanctions,

---

[4] However, considering the broader category of social dilemma experiments, Ostrom et al. [1992] were the first to examine mutual monitoring. Their experiment used the common pool resource game in which players cooperate by refraining from extracting a commonly held resource. This work has been extended in Ostrom et al. [1994] and Moir [1998].

and show that initially, rewards are used, but by the end of the experiment, rewards give way to sanctions.

The experiments discussed above demonstrate two behavioral facts. First, the number of people in a group, per se, does not matter. Larger groups appear to be at least as good at providing public goods. Second, punishment is used to elicit contributions in social dilemma situations. While these facts are important in isolation, so far there has been no research linking the roles of punishment and group size in social dilemma situations. In the experiments presented below, we show that when monitoring is possible group size will only result in lower contributions to the extent that larger groups disrupt the amount of information agents have about each other.

3. Using Simulations to Derive the Effects of Group Size

To generate specific hypotheses about how group size will affect the provision of a public good in an environment where some agents punish free riders, we chose to populate a computer with finitely complex automata of the different types we usually see in the experimental lab and have them play a public goods game with some selection pressure.[5] Surveying the experimental literature on punishment in public goods experiments, one generally finds four behavioral types: unconditional free riders, unconditional cooperators, reciprocally-minded tit-for-taters, and cooperators who punish free riders.[6]

We consider the five behavioral types represented in figure 1. Each type is finitely complex (a la Sethi and Somanathan [2001]) because it operates in a finite number of states. Our machines have either one or two states. Call the initial state that each machine starts in the *passive state* and the state that can

[5] This methodology partially follows Miller et al. [2002] and Sethi and Somanathan [2001].

[6] The surveyed papers include those discussed in section 2 plus Fehr and Gaechter [2000b] and Keser and van Winden [2000].

be triggered by the behavior of the rest of the group, the *alert state.*. There are two one-state machines who are permanently in the passive state: *Free Riders* who free ride regardless of what other agents do, and *Contributors* who contribute unconditionally.

The two-state machines are slightly more complex. Each of the these machines starts in a cooperatively passive state and will remain cooperative as long as there are not too many free riders in the group. Define f as the fraction of all agents in the group that are currently free riding. For n agents indexed $i=1...n$, each of machines 3 through 5 are endowed with a trigger, $\alpha_i \in \{0.1,0.2,0.3,...1\}$, which determines whether the machine transits from the passive state to the alert state or back. Specifically, if $f > \alpha_i$ then there are enough free riders to alert the machine. Likewise if $f < \alpha_i$ then the machine remains in, or transits back to, its cooperative passive state. We call machine 3 *Generalized Tit-for-Tat* because it starts in the cooperative state, but will be alerted by a sufficiently high frequency of free riders, in which case, it free rides until the frequency of free-riding falls below α_i.

There are two punishing strategies, Nice Punishers and Mean Punishers and each punisher will spend up to its accumulated material earnings to monitor all the free riders it sees. Each punisher looks at the other group members individually and will continue to punish until it runs out of resources or punishes all the free riders - which ever happens first. To keep things simple, punishers incur a cost, c, to punish each free rider they stumble across and each sanction inflicts $2c$ harm on the target. While each punisher spends at most c to monitor any individual, each free rider accumulates as many sanctions as there are punishers who catch it and still have available resources. *Nice Punishers* contribute in the passive state and once f rises above α_i they punish free riders while continuing to cooperate. *Mean Punishers* differ because they are somewhat hypocritical - in the alert state they punish even though they free ride themselves.

The n group members of different types are arrayed randomly on a circle. There are t periods. At the beginning of each period all the agents decide whether to contribute their unit endowment or free ride. The marginal and average group benefit of a contribution is q and each q is shared equally. Where $q/n<1<q$ holds the group faces a social dilemma. After every agent makes its contribution decision, f is calculated and each agent compares its trigger to f. Those tit-for-taters who have been alerted will free-ride next period (or continue to free ride) and the alerted punishers will search for currently free riding agents. Those agents for whom f is below their trigger transit back to the passive state and cooperate the following period. Punishing strategies look at their immediate neighbors and then flip from left to right until everyone has been monitored or they run out of resources. The payoff function for each agent is:

$$\pi_i = (1-x_i) + \frac{q\sum_j x_j}{n} - c\sum_j p_{ij} - 2c\sum_j p_{ji}$$

where $x_i \in \{0,1\}$ is an agent's contribution, p_{ij} is the sanction agent i assigns to agent j, and p_{ji} is the sanction agent i receives if caught free riding by agent j.

The group evolves in the following sense. At the end of each period a fraction of the population is replaced randomly and the probability that any type takes over an empty position is based on the current population distribution of types. In simpler analytical models, Borgers and Sarin [1997] have shown that this process leads to the standard replicator dynamic (Taylor and Jonker [1978]). Specifically, the lowest performing five percent of the population, in terms of accumulated resources, are culled every period and replaced by an agent who chooses a type by randomly selecting one of the existing group members and imitating it.

We also allow for mutation. Mutation acts only on the trigger of population sensitive agents (i.e. machines 3 through 5). At the very end of each period each agent faces a 1 in 100 chance of having its trigger randomly changed. Lastly, we used $c=1$ and consider two values of q/n, 0.30 and 0.75, which correspond to our experiments with real agents. For what follows *group size* refers to the number of agents in a group and *monitoring fraction* refers to the fraction of the group each punishing type can monitor. We wish to generate two hypotheses: (1) what happens to the contribution rate when groups size increases and (2) what happens to the contribution rate when the monitoring fraction falls. We started each simulation from a balanced population (i.e. 20% of each type). Initial runs of the simulation indicated that most realizations achieved an equilibrium within 100 periods.

Individual realizations of the simulation have the same characteristics as individual groups in the experimental lab. In many of the realizations the group was quickly taken over by free riders and contributions declined to zero. However, it was not uncommon (as seen in Sethi [1996] and Carpenter et al. [2001]) for the group to eliminate free riding and settle on a distribution of machines dominated by unconditional cooperators, but with substantial numbers of nice punishers and tit-for-taters (mean punishers do very poorly because they punish and are punished; however they are rarely driven to extinction after the free riders have disappeared because the selective pressures fall to random drift when everyone cooperates).[7] To account for the likelihood of different equilibria being achieved, for each hypothesis we ran 1000 realizations of the model and report the average contribution rate over these 1000 realizations for 100 periods.

Figures 2 & 3 here

To generate the group size hypothesis we start with a group of 50 agents and then compare what happens to the average contribution rate as we increase the population to 100 and 200 agents. Figures 1 and 2 show the contribution rates for each population size. One can see that independently of the

[7] Changing the parameters of interests (e.g. the cost of punishment, the number of agents culled each period, and the rate of mutation on the trigger) affects only the rate of decline in the average contribution rate because they partially determine the likelihood that punishing strategies will take over the population. The treatment differences we discuss below all persist when the parameters are changed.

productivity of the public good there is no negative effect of group size on contributions when agents punish. In fact, contributions tend to rise as in the experiments of Isaac and Walker [1988] and Isaac et al. [1994]. Apparently, the effect of having more eyes watching dominates the effect of running out of resources with which to punish. We get:

The Group Size Hypothesis: When agents monitor each other, contributions will not decrease as the number of group members increases.

To generate the monitoring fraction hypothesis we maintain the group size of 50 agents but vary the fraction of the group each punisher sees. In the *Full* treatment, punishers can monitor every other group member. In the *Half* treatment, punishers only get to see half the other group members (starting from their immediate neighbors) and in the *Tenth* treatment punishers see only one-tenth of the other group members (again, starting from their immediate neighbors). Returning to figures 2 and 3 we see a substantial effect of the monitoring fraction. The relationship is monotonic - the fewer free riders each punisher can monitor, the lower is the contribution rate. We get:

The Monitoring Fraction Hypothesis: When agents monitor each other, contributions will decrease as the fraction of the group each punisher can monitor decreases.

4. Experimental Design

The experiment closely resembles Fehr and Gaechter [2000a]. The instructions used by Fehr & Gaechter frame the experiment as a group project. Their framing attempts to coordinate the way that subjects perceive the experiment (i.e. as a social dilemma). Although this is a valid approach, the current experiment uses the more standard language of the voluntary contribution mechanism to provide mutual monitoring with as difficult a test as possible and also to relate more directly to the standard literature.[8] Further, Fehr & Gaechter examine mutual monitoring in groups of partners (subjects stay in the same group for the entire experiment) and in groups of strangers (subjects are randomly reassigned to a new group after each period). To avoid the potential confound of repeated game effects, and because this type of strategic behavior is un-modeled in the simulations, all the sessions discussed below used the strangers protocol.

The standard voluntary contribution game was augmented as in Fehr and Gaechter [2000a] to allow participants to monitor each other. Let each participant be endowed with w_i experimental francs and let p_{ij} be the number of *points* assigned to j by i, then each subject's payoff, π_i, can be described by the following payoff function

$$\pi_i = [w_i - x_i + MPCR\sum_j x_j](1 - \min[1, p_i]) - \sum_j c(p_{ij})$$

where $p_i = (\sum_{j \neq i} p_{ji})/10$ is the number points assigned to i by all the other group members divided by 10 so that 1-min[1, p_i] is the fraction of i's payoff she keeps after being punished. $c(p_{ij})$ is the cost of imposing p_{ij} points on player j and x_i is i's public contribution. Define the *marginal per capita return* (MPCR) as the marginal individual benefit from contributing normalized by its opportunity cost - the benefit of not contributing. For free riding to be subgame perfect we simply make MPCR<1. In this experiment subjects participated in either the *low MPCR* treatment (MPCR=.30) or the *high MPCR* treatment (MPCR=.75).

As mentioned above, to monitor free riders participants assigned what the instructions neutrally referred to as points. It was costly to assign points and each point assigned to another player reduced her gross payoff for the period by ten percent, up to a maximum of one hundred percent. Hence, other group

[8] The instructions for a typical session appear in the appendix.

members could reduce a player's gross earnings to zero, but no further (unless the target also punished other players).[9] The punishment cost function used in the experiment, $c(p_{ij})$, was

Points Given Player j, p_{ij} :	0	1	2	3	4	5	6	7	8	9	10
Cost of Points, $c(p_{ij})$:	0	1	2	4	6	9	12	16	20	25	30

Along with changes in the MPCR, group size was manipulated in two ways. To test the group size hypothesis we simply changed the size of groups. In one treatment participants were divided into teams of five persons and in the other we increased the team size to ten persons. To investigate the monitoring fraction hypothesis, we maintained five person groups but varied the number of other subjects each player could monitor. In the *Full* treatment, group members saw the decisions made by everyone else in the group and could monitor them all. In the *Half* treatment, group members could only monitor half of the other group members, but the monitoring groups overlapped so that everyone was seen. In the *Single* treatment, subjects saw only one other group member, but again, everyone was seen by someone else.[10] Hence, the monitoring fraction treatment variable is a proxy for the informational effect of increasing the group size and, because monitoring groups overlap, we conform as closely as possible to the simulated hypothesis.

---

[9] The payoff function for the experiment differs from that used in the simulations because the automata's contribution choice is binary and therefore automata can not condition their punishment choices on how much another group member free rides. This complication might be interesting to examine in a future project where one was trying to mimic experimental results exactly rather than qualitatively.

[10] The instructions made explicit the fact that everyone was seen by at least one other group member.

A total of 20 sessions were conducted with 350 participants.[11] Subjects were recruited by email from the general student population and none had ever participated in an economics experiment before. Each subject was given a five dollar show-up fee when he or she arrived and was then seated at a computer terminal which was partially isolated (there were blinds on each side) so that decisions were made privately. Each session lasted about 45 minutes from sign in to payments and subjects earned $17.44 on average, including the show-up fee.

A typical session was conducted as follows. There were ten periods and each period was split into three stages, the contribution stage, the monitoring stage, and a summary of the results for the period. The contribution decision was to allocate twenty experimental francs between a private account which only benefitted the individual, and a public account which benefitted everyone in the group according to the payoff function. Once all the contributions were recorded, participants learned the group total contribution and their individual gross earnings.

Next, the experiment moved to the second stage. Here participants saw the individual contribution decisions made by other members of their current group. Based on this information subjects could assign up to ten points to each of the other team members. They assigned points by paying out of their accumulated earnings. The total cost of punishing others was calculated according to the cost of punishment function, $c(p_{ij})$. When everyone had finished the second stage, the experiment moved to the third stage where subjects saw a summary of their net payoffs (after subtractions for punishing others and for being punished) for the current period.

[11] Included in these numbers are four control sessions we ran as replications of the standard voluntary contribution game to demonstrate that our protocol and participants were consistent with standard results. Because contributions decayed and the rate of decay depended on the MPCR, we concluded that our treatments and not our language or participants were responsible for the results presented in section 5.

Notice that if we assume standard preferences for participants (i.e. they only care about their own payoffs) then punishing is not a credible threat because it is costly and any potential returns (i.e. getting a free rider to contribute) are divided equally among all the other team members - alternatively, the fact that the game is finitely repeated precludes punishment as part of any subgame perfect equilibrium. This implies that no one should punish, and given this, no one should fear being punished from which it follows that free riding is still a dominant strategy.

5. Results of the Experiment

To begin our analysis of the group size hypothesis we note that contributions to the public good do not diminish as the number of people in a group increases. Figures 4 and 5 plot the average fraction of the endowment contributed to the public good in each period for each Full monitoring treatment. These figures clearly demonstrate that the effect of mutual monitoring does not diminish when we double the size of groups.[12] In five person groups, mutual monitoring largely maintains the initial level of contributions with only a slight increase in the high MPCR treatment.[13] By contrast, in ten person groups mutual monitoring increases contributions to nearly 100 percent by the end of the experiment for both MPCR treatments.[14] Hence, when group members can monitor each other, large groups are actually better at eliciting contributions than small groups. Note

[12] Kolmogorov-Smirnov tests for distributional differences and Wilcoxon tests of differences in contributions pooled across the ten periods support this claim for both MPCR treatments. For the low MPCR we have ks=0.81, $p<0.01$; $z=-7.76$, $p<0.01$ and for the high MPCR we have ks=0.48, $p<0.01$; $z=-5.29$, $p<0.01$.

[13] Slight but significant. The Wilcoxon rank sum statistic for the difference in the first and last periods is $z=-2.63$, $p=.01$.

[14] Despite a slight 'endgame' effect, the averages are not significantly below 100 percent contribution. In the low MPCR case the Wilcoxon signed ranks test yields $z=-0.75$, $p=0.45$, and in the high MPCR treatment we have $z=-1.20$, $p=0.23$.

the similarities between our simulated results and the experiment. In both cases punishment leads to higher contributions when groups grow in size and this effect is independent of the MPCR. Comparatively, the experiment conforms exactly to the simulated group size hypothesis.

Figures 4 & 5 here

Figures 4 and 5 document another unexpected result. To the knowledge of the author, this is the first voluntary contribution experiment in which full or even nearly full contributions have been sustained without face-to-face communication or draconian external regulation.[15] This result is particularly important because contributions approach the social optima independently of the MPCR.

Although our automata are programed to either contribute or not and to punish or not and this programming is invariant to the size of the group, our experimental participants, being more complex, may base their contribution and punishment decisions on the size of the group. Further, it may be the case that our experimental contribution results match the simulations despite the complexity of human subjects simply because the net effect of group size is zero or positive. For these reasons we dig deeper into our punishment data.

An interesting fact from social psychology is that people are typically less likely to act prosocially in large groups. First identified in Latane and Darley [1970], psychologist have shown that, for example, bystanders are less likely to help a stranger when many people have the opportunity to help.[16] In essence a coordination problem arises - people agree help should be provided, but it is

---

[15] Except possibly for Ameden et al. [1998] who show that endogenously sorting players by contribution achieves high levels of contributions, but it is unclear if this result is independent of the MPCR.

[16] Note, free riding can not explain the reduced willingness to help in large groups because free riders would not help in either case - i.e. they would not be sensitive to group size.

unclear in a large group exactly who should help. In our experiment, punishing free riders is a prosocial act like helping strangers. We consider a regression analysis to determine whether group size affects the punishment choices of our participants.

Table 1 here

Table 1 reports the results of an analysis of the individual-level punishment data. We present both a probit analysis of the probability that one agent monitors someone else and a least squares analysis of agents' expenditures on punishment. For the probits, the dependent variable is one if the participant punishes at least one other member of the group (zero otherwise) and the left-hand side variable for the least squares analysis is a player's total expenditure on punishment. The first four columns pool the data from both MPCR treatments and the last four columns separate the data by MPCR to check the robustness of the overall analysis. Lastly, because our experiment generates a panel of data, regressions 1(a) and 1(b) use robust standard errors to partially account for possible interdependencies among our observations while the rest of the regressions employ a different approach, random effects, to control for cross-sectional differences.

The first thing to notice is that our real agents behave similarly to our programed agents in that their decisions to punish or not seem invariant to the number of people in a group. After controlling for differences in the amount of free riding in groups using the Avg Private and Avg Private^2 variables which are the average private investment of the decision maker's group and its square, and the public contribution of the person punishing, Punisher's Pub, we see that in most circumstances the group size regressor, N, is not a significant determinant of whether players punish or not (probits).[17] Further, we see that contributors do most of the punishing (i.e. the Punisher's Pub term is positive and highly significant). This also matches the simulations where, on average, mean punishers do considerably worse than nice punishers. Notice, this result suggests that our participants did not suffer from the kind of coordination problem surrounding who should engage in prosocial acts in big groups that has been found in previous psychological studies.

When we examine the amount a player spends on monitoring we get similar results. The groups size regressor is mostly insignificant, except for when we look at the low MPCR treatment in isolation, in which case, players spend significantly less on punishment in large groups, controlling for how cooperative the group was. The overall insignificance of the group size regressor in this case implies that players are willing to allocate a set fraction of their earnings to punishment and do not adjust this budgeting decision based on the number of people they monitor. In this respect, our players differ from the automata who were programmed to spend all their income on monitoring if there were sufficiently many free riders. The last thing to notice about table 1 is that, while contributors are more likely to monitor other players, they spend no more on monitoring than other players. This fact indicates that on the off chance a free rider monitors in this treatment, she does so with vigor.

The second hypothesis states that the monitoring fraction should matter because punishment will be less of a deterrent when each monitor sees fewer other group members. The data support this hypothesis. Specifically, we find more

---

[17] In other regression not reported in table 1 we corrected for two possible confounds. First, players tend to make higher gross earnings in larger groups and with higher MPCRs. This might lead to a wealth effect - players have more earnings to spend on punishment. Second, the implicit price of punishment in the experiment is not fixed. On average, across treatments, one point reduced the earnings of the targeted player by 2.41 francs which implies the price of reducing a player's earnings by 1 franc was 1/2.41 or 0.41 francs. Rerunning the regressions in table 2 with controls for these potential confounds (i.e. adding player gross earnings and the implied price of punishment) does not change the results and neither effect is significant.

free riding as the monitoring fraction falls. Returning to figures 4 and 5, we see plots of average contributions for the three monitoring fractions. Compare again the similarities between the time paths from our simulations (figures 2 and 3) and the data from the experiment (figures 4 and 5). As one can see, in the low MPCR condition (figure 4) all three treatments start at contributions of forty percent, but spread over time. Similarly, in figure 5 the three treatments when MPCR is high start at approximately fifty-five percent and then trifurcate. Just as the second hypothesis predicts, the single monitoring treatment is the least effective at maintaining contributions, the half treatment is a little more effective, and the full treatment is the most effective.[18, 19] Information appears to be an important component of effective mutual monitoring. In particular, mutual monitoring does not work as well when the number of other group members that one can see falls, even though it was made clear in the instructions that monitoring groups overlap so that every free rider would be seen by someone else.

Table 2 summarizes a regression analysis similar to table 1 for the punishment decisions in the monitoring fraction treatments. Again, we report probits and least squares regressions using robust standard errors and random effects. As when testing the group size hypothesis, we find punishment increases in the severity of free riding, but the effect diminishes slightly in extreme cases (i.e. Avg Private is positive and highly significant and its square is negative and significant) and contributors are significantly more likely to punish than free riders are. With respect to who punishes, compared to the group size data, the probit and least squares results are much more comparable in table 2.

[18] In the low MPCR condition the single and full treatments are clearly different (ks=0.31, p<0.01; z=3.49, p<0.01) as are the single and half treatments (ks=0.26, p=0.01; z=2.88, p<0.01). But, while the trend is for contributions to be maintained in the full treatment and decline in the half treatment, the distributions of group contributions are not different (ks=0.12, p=.0.56; z=0.87, p=0.39).

[19] For the high MPCR condition the test statistics, based, as in note 16, on group contributions, are as follows: full vs. half (ks=0.20, p=0.10; z=2.14, p=0.03), full vs. single (ks=0.44, p<0.01; z=5.47, p<0.01), and single vs. half (ks=0.31, p<0.01; z=4.33, p<0.01).

Specifically, in the monitoring fraction treatments contributors monitor more often and spend more on monitoring.

Overall, there does not seem to be a robust effect of monitoring fraction on participants' likelihood of monitoring. Using robust standard errors, players in the Half treatment seem significantly less likely to monitor compared to players in the Full treatment, but this result is not robust to the inclusion of random effects (compare equations 1a and 2a). Further, participants in the Single treatment are neither more nor less likely to monitor, regardless of the method we use.

Moving to an examination of the effects of our treatments on players' monitoring expenditures, it seems obvious that expenditures, controlling for how cooperative groups were, should fall when players see fewer other group members - there are simply fewer opportunities to monitor. This is more or less what we see in the data. Equations 1b and 2b show that regardless of our method, expenditures on monitoring fall when players can not see all the other group members. However, notice that expenditures fall more in the Half treatment than in the Single treatment indicating that when players can see only one other group member they assume more responsibility for making sure that this person behaves.

Underlying the overall result that players are equally likely to punish in each monitoring fraction treatment, but spend less when they see less than the entire group, are differences in responses by MPCR treatment. The Half treatment appears to have a strong effect in the low MPCR treatment while the Single treatment matters more for players in the High MPCR treatment.

To summarize our results with respect to the effects of monitoring fraction, overall players do not appear systematically less likely to monitor individual other group members when they see fewer of them, but they do spend less money on monitoring when they see fewer group mates. Notice that these results mimic the programming of our automata who punish equally all the free riders they are allowed to see. As in our simulated games, these two results (monitor no less, but allocate fewer total resources) imply free riders are punished less severely when the monitoring fraction falls because they are seen by fewer eyes. As a result, free riding increases.

So far we have established that contributions fall with a reduction in the monitoring fraction because free riders are punished less severely in smaller monitoring groups. We can also investigate another, un-simulated, reason why large groups might suffer more free riding. Again, we are allowing for the fact that real agents may behave more complexly than our automata. An auxiliary hypothesis we can entertain is that the monitoring fraction will influence the relationship between being punished and increasing one's contribution in the future. It might be reasonable to expect that free riders will respond less to punishment when the monitoring fraction falls because they perceive that it is easier to get away with not contributing if there are fewer eyes watching.

We test this extension of Hypothesis 2 using individual data. To do so we regress free rider's responses on the punishment they receive and other treatment variables to see whether free riders respond less to punishment in groups where the number of monitors is restricted. Table 3 summarizes the evidence supporting this proposition. The dependent variable is the difference in a player's contribution between periods t-1 and t. Lag Points is the total number of punishment points assigned to the player in period t-1. There are three dummy variables to identify each treatment. MPCR dummy takes the value one when the MPCR is high, Half dummy is one to designate the Half treatment, and Single dummy is one to indicate the Single treatment. We also examine interactions between the points assigned to the free rider (i.e. FR equals one when a player contributes less than the group average in period t-1) and the treatments to examine our hypothesis about free riders responses. Finally, because the dependent variable is bound between -20 and 20 we use random effects Tobit methods in addition to our benchmark estimates with robust errors.

Equation 1 is a first pass at the data. It confirms that players, in general, respond as predicted to punishment, but the result is only marginally significant suggesting there is more to the story. Equations 2 and 3are much better fits of the data because we allow for the differential impact of punishment on free riders. These two regressions illustrate that punishment directed at free riders significantly changes behavior, but also of interest is the fact that the sign on Lag Points now indicates that contributors reduce contributions if punished. Hence, punishment in our experiment causes some regression to the mean contribution.

Consider equation four which includes dummy variables to account for treatment differences. Note, the reference individual in the following analysis is a contributor in the low MPCR, full monitoring condition. To begin, the MPCR×FR variable has a positive and highly significant coefficient which makes sense. We would expect that as the opportunity cost of contributing falls free riders become more willing to contribute in the future. We also find that the signs on the Half×FR and Single×FR variables are positive indicating that, after controlling for differences in the number of points assigned, free riders increase contributions more than contributors in our limited monitoring treatments which, again, makes sense.

Table 3 here

We can now sharpen the analysis by further interacting terms. The triple interaction of Lag Points by treatment by FR, allows us to ask if punishment has more or less of an effect on free-riding in the different treatments. As one can see in equation 5, in both monitoring fraction treatments, free rider responses to punishment are still significantly negative suggesting that participants in these treatments respond less, perhaps because they fear punishment less. Further, the coefficients suggest that participants become increasingly obstinate when punished. That is, the more punishment free riders receive in the Half and Single treatments compared to the Full treatment, the more they stubbornly free ride. Notice also the relative size of the coefficients on the Half and Single triple interactions. The large negative coefficient on the Single interaction compared to the Half interaction indicates that free riders react less to punishment as fewer monitors observe what they do. These results are interesting because they imply that free riders react differently to collective shunning than to when only one individual confronts them. Generalizing, free riders seem to be more likely to dig in their heels to spite one goodie-goodie who criticizes them, but, at the same time, they bend to follow a collectively established norm.

6. Concluding Remarks

Traditional game theoretic formulations of public goods environments identify free riding as a dominant strategy and therefore suggest that the addition of costly, and therefore incredible, monitoring and increases in group size should not matter. However, this research shows that when we complicate this formulation by taking seriously the behavioral heterogeneity seen in the experimental lab which has been given microfoundations by evolutionary games, group size does matter.

By being more specific about why group size might matter our simulations and experiments have demonstrated that large groups may be equally adept at controlling free riders because members tend to sanction transgressors without considering the fact that other monitors might also punish. That is, everyone in the group wants to show dissatisfaction with free riders, and as a result, free riders are punished much more severely in large groups. Obviously, this behavior acts as a strong disincentive to continue free riding.

At the same time, however, our simulations and experiments suggest that the logistics of large groups may hinder the ability of mutual monitoring to discipline free riders. It is clear that one logistical side effect of growing groups is that it becomes difficult for each group member to keep an eye on all the other members. Our second major result illustrates why this aspect of large groups might matter. When we vary the size of monitoring groups from including everyone to being composed of just one other member we find a significant reduction in the amount that transgressors are punished and a weakening of free riders reactions to punishment which, combined, translate into more free riding.

Our results are interesting because they help explain the dynamics of work teams. More specifically, these results might partially explain why we see large productive teams (Hansen [1997], Knez and Simester [1998]). In large teams, shirkers should fear both a greater likelihood of being caught and increased sanctions because most people monitor, but the extent to which this will act as a deterrent depends on the transparency of the production process. When effort contributions are obvious, monitoring groups will be large and the threat of heavy sanctions will keep shirkers in line. However, when it is easy to hide your free riding or mimic hard work, monitoring group size will fall and shirking will become more prevalent.

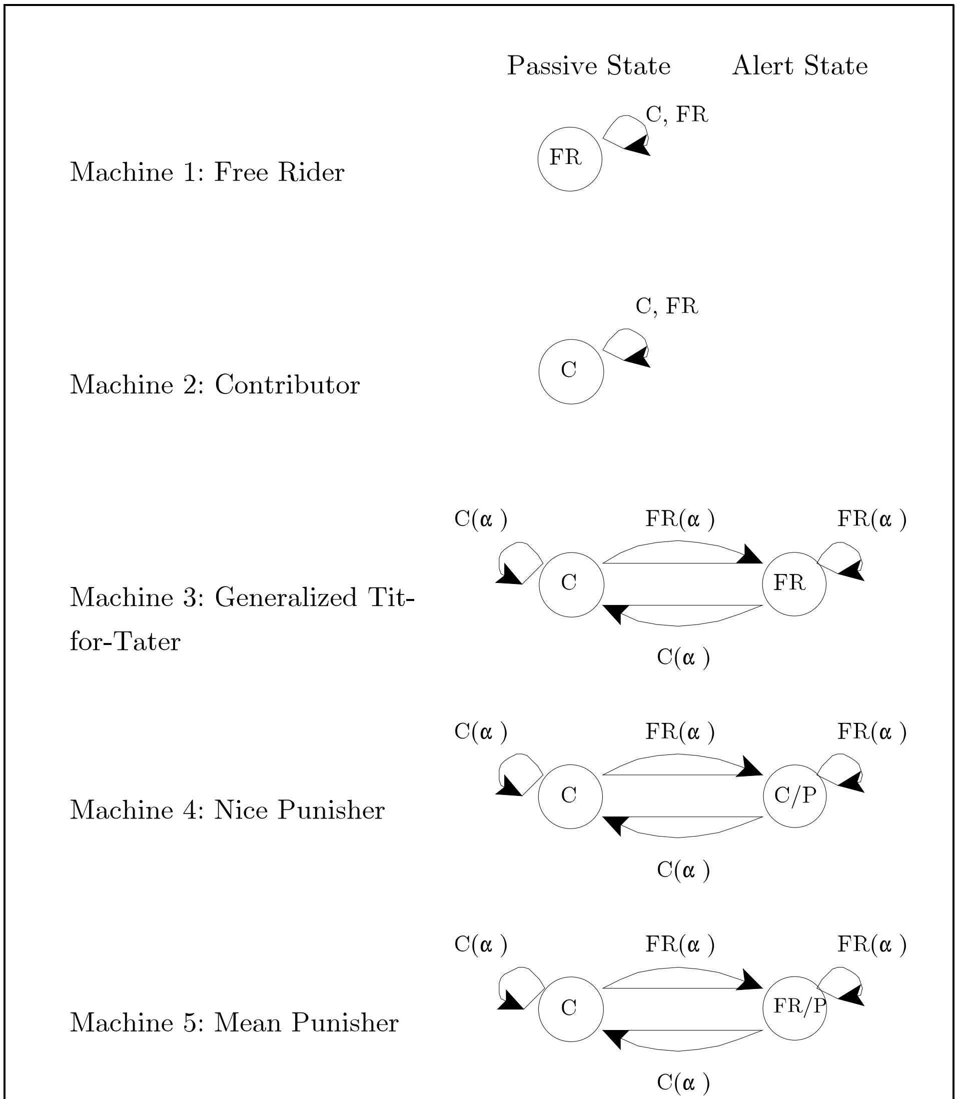

Figure 1 - The Simulated Two-State Automata. C indicates Contribute, FR indicates Free Ride, P indicates punish free riders, and α is the trigger frequency of free riding or contributing that determines behavior.

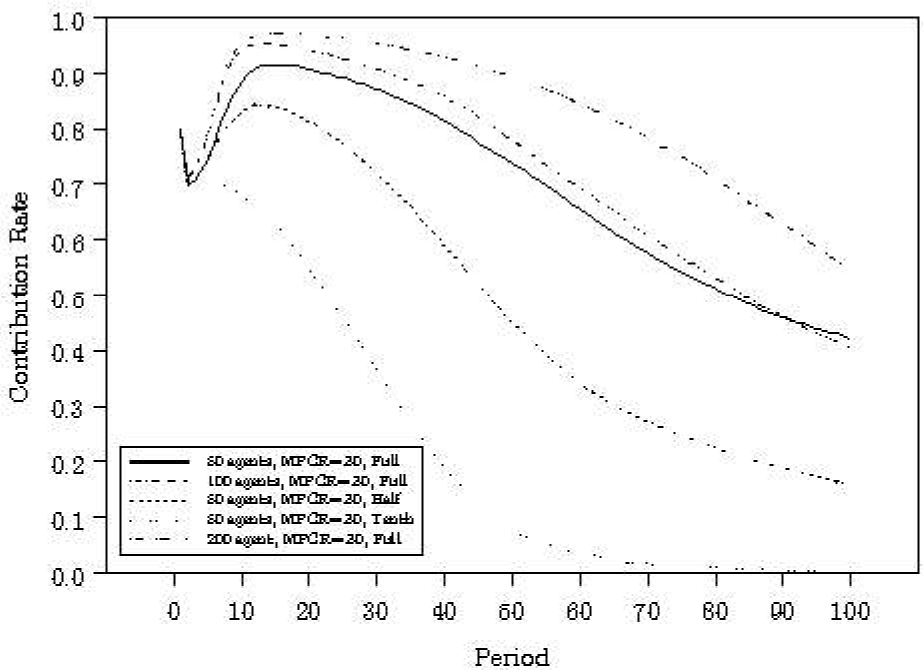

Figure 2 - The Simulated Effect of Increased Group Size and Reduced Agent Information in the MPCR = 0.30 Treatment. (Full implies all group members are monitored by each agent, Half means half are monitored, and Tenth means one-tenth are monitored).

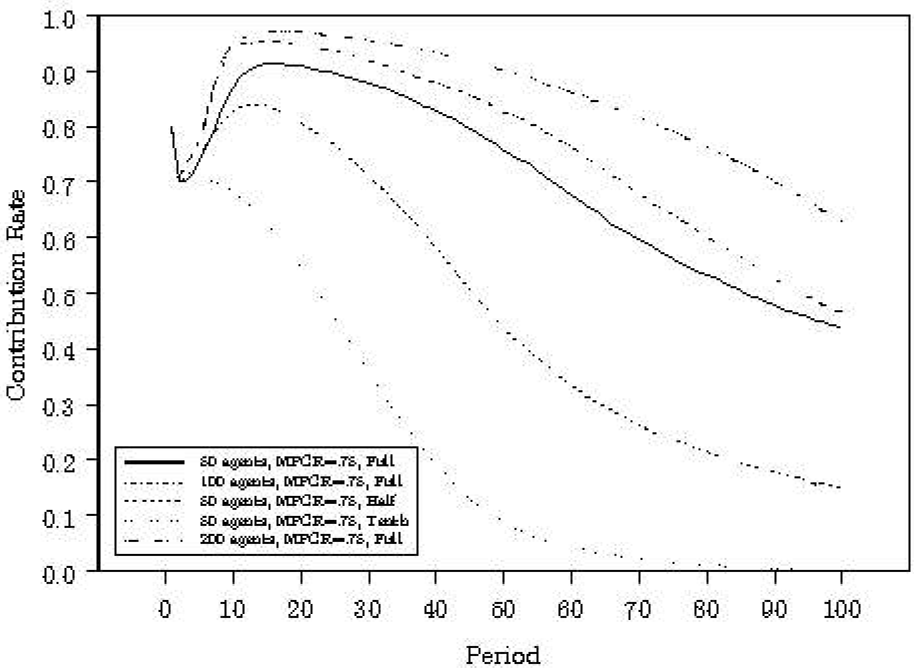

Figure 3 - The Simulated Effect of Increased Group Size and Reduced Agent Information in the MPCR=0.75 Treatment. (Full implies all group members are monitored by each agent, Half means half are monitored, and Tenth means one-tenth are monitored).

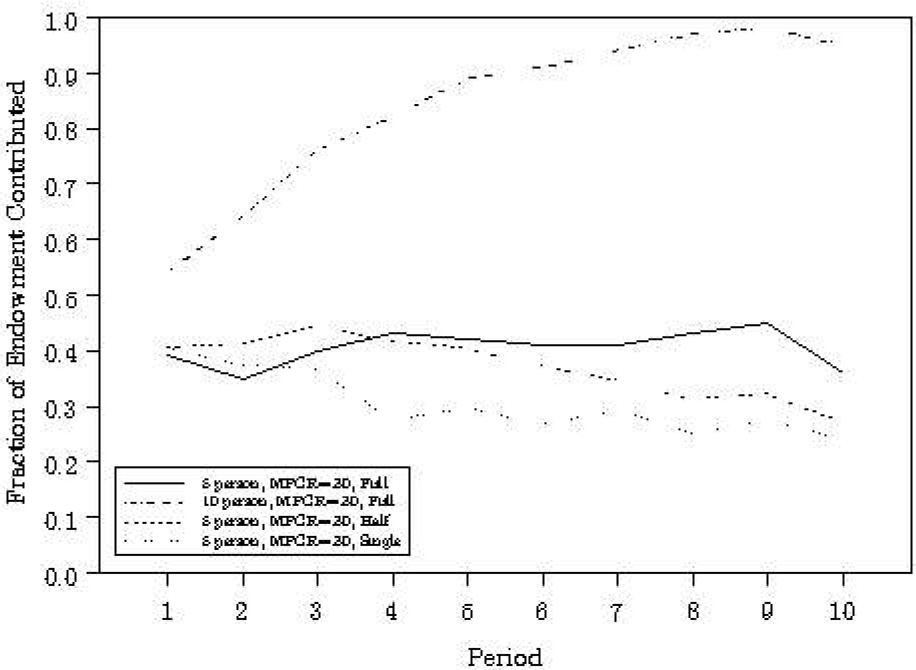

Figure 4 - The Elicited Effect of Increased Group Size and Reduced Participant Information in the MPCR=0.30 Treatment. (Full implies all group members are monitored by each agent, Half means half are monitored, and Single means only one is monitored).

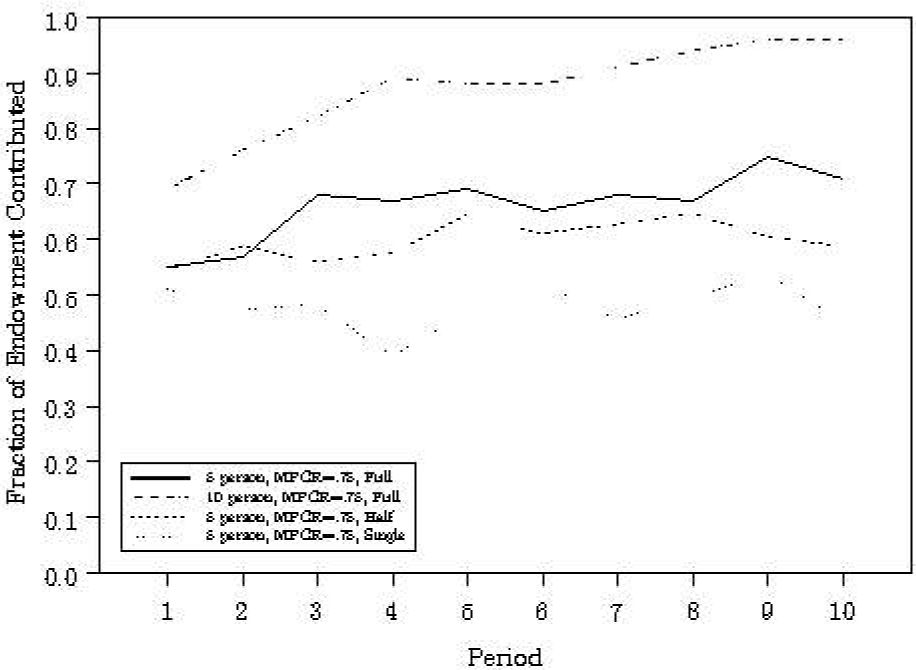

Figure 5 - The Elicited Effect of Increased Group Size and Reduced Participant Information in the MPCR=0.75 Treatment. (Full implies all group members are monitored by each agent, Half means half are monitored, and Single means only one is monitored).

Dependant Variable = Punish or Not (Probit), Expenditure on Punishment (OLS & GLS))								
	All Data		All Data		Low MPCR		High MPCR	
	Robust Standard Errors		Random Effects					
	Probit	OLS	Probit	GLS	Probit	GLS	Probit	GLS
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
N	0.17* (0.09)	-0.14 (0.63)	0.13 (0.22)	-0.29 (0.70)	0.09 (0.30)	-1.32*** (0.48)	0.22 (0.32)	1.11 (1.19)
Avg Private	0.22*** (0.03)	0.58*** (0.14)	0.31*** (0.03)	0.61*** (0.18)	0.32*** (0.04)	0.49*** (0.11)	0.34*** (0.06)	0.94** (0.40)
Avg Private^2	-0.01*** (0.001)	-0.03*** (0.008)	-0.01*** (0.002)	-0.03*** (0.01)	-0.01*** (0.002)	-0.02*** (0.005)	-0.02*** (0.004)	-0.04 (0.63)
Punisher's Pub	0.02*** (0.007)	0.02 (0.04)	0.04*** (0.01)	0.04 (0.05)	0.05*** (0.01)	0.10*** (0.03)	0.04*** (0.01)	0.12 (0.10)
Intercept	-1.07*** (0.17)	1.84* (1.01)	-1.65*** (0.27)	1.47 (1.35)	-1.88*** (0.37)	0.12 (0.84)	-1.66*** (0.43)	-0.25 (2.52)
Pseudo R^2 R^2	0.07 -	- 0.01	- -	- 0.01	- -	- 0.10	- -	- 0.01
Wald χ^2 F-stat	139.14 -	- 6.28	118.28 -	- 18.90	74.34 -	- 97.50	48.11 -	- 10.14

Notes: standard errors in parentheses.
*** significant at the 0.01 level.
** significant at the 0.05 level.
* significant at the 0.10 level.

Table 1 - The Effect of Group Size on Monitoring Decisions

Dependant Variable = Punish or Not (Probit), Expenditure on Punishment (OLS & GLS)

	All Data		All Data		Low MPCR		High MPCR	
	Robust Standard Errors		Random Effects					
	Probit	OLS	Probit	GLS	Probit	GLS	Probit	GLS
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
Avg Private	0.22*** (0.03)	0.57*** (0.16)	0.30*** (0.05)	0.69*** (0.14)	0.40*** (0.09)	1.10*** (0.20)	0.27*** (0.07)	0.69*** (0.24)
Avg Private^2	-0.01*** (0.002)	-0.03*** (0.007)	-0.01*** (0.002)	-0.03*** (0.006)	-0.01*** (0.004)	-0.04*** (0.008)	-0.01*** (0.004)	-0.03** (0.01)
Punisher's Pub	0.02*** (0.006)	0.08*** (0.03)	0.04*** (0.009)	0.12*** (0.03)	0.05*** (0.01)	0.09*** (0.03)	0.04*** (0.01)	0.15*** (0.04)
Half	-0.25*** (0.07)	-2.07*** (0.35)	-0.21 (0.23)	-2.15*** (0.66)	-0.52* (0.29)	-2.48*** (0.62)	0.10 (0.26)	-1.88 (1.22)
Single	0.08 (0.07)	-1.50*** (0.35)	0.42 (0.27)	-1.73*** (0.68)	0.30 (0.30)	-0.97 (0.64)	0.11 (0.23)	-2.71** (1.24)
Intercept	-1.01*** (0.19)	1.30 (1.07)	-1.73*** (0.33)	-0.32 (0.92)	-2.28*** (0.60)	-3.40*** (1.32)	-1.77*** (0.37)	0.24 (1.48)
Pseudo R^2 R^2	0.03 -	- 0.03	- -	- 0.03	- -	- 0.07	- -	- 0.03
Wald χ^2 F-stat	85.91 -	- 12.38	69.75 -	- 52.99	47.09 -	- 54.37	34.01 -	- 22.76

Notes: standard errors in parentheses.
*** significant at the 0.01 level.
** significant at the 0.05 level.
* significant at the 0.10 level.

Table 2 - The Effect of the Monitoring Fraction on Punishment Decisions

Dependant Variable = $x_{it} - x_{it-1}$ (difference in public contribution)

	Robust		Random Effects Tobit		
	(1)	(2)	(3)	(4)	(5)
Lag Points	0.20* (0.11)	-0.23** (0.12)	-0.23** (0.11)	-0.15 (0.11)	0.03 (0.12)
Lag Points^2	0.008 (0.01)	-0.02* (0.01)	-0.02* (0.01)	-0.007 (0.01)	-0.02* (0.01)
Lag Points × FR	-	0.92*** (0.08)	0.92*** (0.08)	0.64*** (0.09)	0.87*** (0.09)
MPCR dummy × FR	-	-	-	1.03*** (0.34)	1.52*** (0.49)
Half dummy × FR	-	-	-	1.42*** (0.36)	1.90*** (0.50)
Single dummy × FR	-	-	-	1.40*** (0.36)	2.88*** (0.52)
Lag Points × MPCR × FR	-	-	-	-	-0.26** (0.12)
Lag Points × Half × FR	-	-	-	-	-0.31** (0.14)
Lag Points × Single × FR	-	-	-	-	-0.81*** (0.17)
Intercept	-0.68*** (0.19)	-0.68*** (0.19)	-0.69*** (0.20)	-1.33*** (0.22)	-1.77*** (0.23)
Adjusted R^2	0.02	0.09	-	-	-
F-stat	15.13	57.46	-	-	-
Wald χ^2	-	-	179.92	228.62	269.22

Notes: standard errors in parentheses.
*** significant at the 0.01 level.
** significant at the 0.05 level.
* significant at the 0.10 level.

Table 3 - Do Free Riders Respond to Punishment?

7. Appendix - Participant Instructions

You have been asked to participate in an economics experiment. For participating today and being on time you have been paid $5. You may earn an additional amount of money depending on your decisions in the experiment. This money will be paid to you, in cash, at the end of the experiment. By clicking the BEGIN button you will be asked for some personal information. After everyone enters this information we will start the instructions for the experiment.

Please be patient while others finish entering their personal information. The instructions will begin shortly.

During the experiment we will speak in terms of Experimental Francs instead of Dollars. Your payoffs will be calculated in terms of Francs and then translated at the end of the experiment into dollars at the following rate: 30 Francs = 1 Dollar

Each participant receives a lump sum payment of 15 Francs at the beginning of the experiment (on top of the $5.00 show-up payment). This one-time payment may be used to offset any losses that are incurred during the experiment. However, it should be noted that you can ALWAYS avoid losses through your own decisions.

The experiment is divided into 10 different periods. In each period participants are divided into groups of 5. You will therefore be in a group with 4 other participants. The composition of the groups will change randomly at the beginning of each period. Therefore, in each period your group will consist of different participants.

Each period of the experiment consists of two stages. In the first stage you will decide how many francs you want to invest in each of two investment accounts. One account is a Private Account, which only you benefit from. The second account is a Public Account, the benefits of which are shared equally by all members of your group. In the second stage of the period you will be shown the investment behavior of the other members of your group. You can then decide whether you want to distribute points to members of your group. If you distribute points to other members of your group, their earnings will be reduced.

Now we will explain the two stages in more depth.

Stage One

At the beginning of every period each participant receives and endowment of 20 francs. You have to decide how much of this endowment you want to invest in each of the two accounts mentioned above. You are asked to invest in whole franc amounts (i.e. an investment of 5 francs is alright, but 3.75 should be rounded up to 4).

To record your investment decision, you will type the amount of francs you want to invest in the Public and/or the Private account by typing in the appropriate text-input box which will be yellow. Once you have made your decision, there will be a green Submit button that will record your investment decision.

After all the members of your group have made their decisions, each of you will be informed of your Gross Earnings for the period.

Your Gross Earnings will consist of two parts:

1) Your return on your Private Account. Your Private Account returns 1 franc for each franc invested. That is, for each franc invested in the Private Account you get 1 franc back.

2) Your return from the Public Account. Your earnings (and everyone else's in your group) is equal to 0.3 times the total investment by all members of the group to the Public Account.

Your Earnings can be summarized as follows:

$1\times$(Investment in Private Account) + $0.3\times$(Group Total Investment in Public Account)

The income of each group member from the Public Account is calculated the same way. This means that each group member receives the same amount from the total investment in the Public Account. For example, consider the case of groups with 5 members, if the total investment in the Public Account is 75 francs (e.g. first group member invests 15 francs, the second 20, the third 10 and the fourth and fifth 15 each) then each group member will receive $0.3\times75 = 22.5$ francs. If the total investment was 30 francs then each group member would receive $0.3\times30 = 9$ francs.

For each franc you invest in the Private Account you get 1 franc back. Suppose however you invested this franc in the Public Account instead. Your income from the Public Account would increase by $0.3 \times 1 = 0.3$ francs. At the same time the earnings of the other members of your group would also increase by 0.3 francs, so the total increase in the group's earnings would be 1.5 francs. Your investment in the Public Account therefore increases the earnings of the other group members. On the other hand your earnings increase for every franc that the other members of your group invest in the Public Account. For each franc invested by another group member you earn $0.3 \times 1 = 0.3$ francs.

Stage Two

In stage two you will be shown the investment decisions made by other members of your group and they will see your decision. Also at this stage you can reduce the earnings made by other member of your group, if you want to. You will be shown how much EACH member of your group invested in both the Public and Private Accounts. Your investment decision will also appear on the screen and will be labeled as 'YOU'. Please remember that the composition of your group will change at the beginning of each period and therefore you will not be looking at the same people all the time.

You must now decide how many points (if any) you wish to give to each of the other member of your group. You distribute points by typing them into the input-text box that appears below the investment decision of each of the other group members.

You will have to pay a cost to distribute points to other group members. This cost increases as you distribute more points to another participant. You can distribute between 0 and 10 points to each other member of your group. Your total cost of distributing points is the sum of all the costs you incur for distributing points to each of the other group members. The following table illustrates the relationship between the points distributed to each group member and the costs of doing so in francs.

Points:	0	1	2	3	4	5	6	7	8	9	10
Cost of Points (in francs):	0	1	2	4	6	9	12	16	20	25	30

Consider the case where there are 5 people per group. Suppose you assign 2 points to a group member. This costs you 2 francs. If you assign 9 points to another group member, it will cost you 25 francs and if you assign 0 points to the rest of the members of your group, you do not incur any cost. In this case your Total Cost of assigning points is (2+25+0+0) or 27 francs. At any time you will be able to calculate your total cost of distributing points by clicking the orange Calculate Cost button that will appear on the screen. When you have finished distributing points you will click the blue Done button.

If you assign 0 points to a particular group member you do not change his or her earnings. However, for each point you assign to a group member, you reduce his or her Gross Earnings in the current period by 10 percent. Hence, if you assign one group member 2 points, his or her Gross Earnings for the period will be reduced by 20%. Assigning 4 points reduces Gross Earnings by 40% etc.

How much a participant's earnings from the first stage are reduced is determined by the Total amount of points he or she receives from all the other group members. If a participant receives a total of 3 points (from all the other group members in the current period) then his or her Gross Earnings would be reduced by 30 percent. If someone is assigned 4 points in total his or her Gross Earnings would be reduced by 40 percent. If anybody is assigned 10 or more points their Gross Earnings will be reduced by 100 percent. In this case the Gross Earnings of this person would be 0 francs for the current period.

For example, if a participant had Gross Earnings of 30 francs from the first stage and was assigned 3 points in the second stage, then his or her earnings would be reduced to 30 - (0.3×30) = 30 - 9 = 21 francs.

In general, your earnings after the second stage will be calculated as follows:

Total Earnings at the end of the Second Stage:

1) If you received fewer than 10 points then Total Earnings equal

(Gross Earnings from Stage One)-[Gross Earnings×(.1×received points)]-(the cost of points you distributed)

2) If you receive 10 or more points then Total Earnings equal

- (the cost of the points you distributed)

Please note that your earnings at the end of the second stage can be negative, if the cost of the points you distribute exceeds your (possibly reduced) earnings from stage one. However, you can avoid such losses by the decisions you make.

After all participants have made their decisions in the second stage, your final earnings for the period will be displayed in a manner similar to what follows:

Earnings Screen at the end of the Period

Your Gross Profits in the Current Period:
The Total Cost of the Points You Assigned to Others:
Number of Points Assigned to You by Others:
Reduction of Gross Profit due to Points Assigned to You: %
Current Period Payoff after Subtractions:
Your Accumulated Earnings Including this Period:

When you have finished reviewing your earnings for the current period you will click the orange Proceed to Next Period button and wait for others to finish. When everyone is done, the experiment will proceed to the next period starting with stage one.

If you have any questions please raise your hand. Otherwise, click the red Finished button when you are done reading.

This is the end of the instructions. Be patient while everyone finishes reading.

8. References

Ameden, H., A. Gunnthorsdottir, D. Houser and K. McCabe (1998). A minimal property right system for the sustainable provision of public goods, University of Arizona.

Andreoni, J. (1988). Why free ride? Strategies and learning in public good experiments. *Journal of Public Economics* 37: 291-304.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm glow giving. *Economic Journal* 50: 464-477.

Bolton, G. and A. Ockenfels (1999). ERC: A theory of equity, reciprocity and competition. *American Economic Review* 90(March): 166-193.

Borgers, T. and R. Sarin (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory* 77(1): 1-14.

Bowles, S., J. Carpenter and H. Gintis (2001). Mutual monitoring in teams: The effects of residual claimancy and reciprocity. mimeo.

Carpenter, J., P. Matthews and O. Ongonga (2001). Social reciprocity. mimeo.

Davis, D. and C. Holt (1993). *Experimental economics*. Princeton, Princeton University Press.

Falk, A. and U. Fischbacher (1998). A theory of reciprocity. mimeo.

Fehr, E. and S. Gaechter (2000a). Cooperation and punishment in public goods experiments. *American Economic Review* 90(4): 980-994.

Fehr, E. and S. Gaechter (2000b). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14(3).

Fehr, E. and K. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114(3): 769-816.

Hansen, D. (1997). Worker performance and group incentives: A case study. *Industrial and Labour Relations Review* 51(1): 37-49.

Isaac, M. and J. Walker (1988). Group size effects in public goods provision: The voluntary contribution mechanism. *Quarterly Journal of Economics* 103(Feb.): 179-199.

Isaac, M., J. Walker and A. Williams (1994). Group size and the voluntary provision of public goods. *Journal of Public Economics* 54(1): 1-36.

Keser, C. and F. van Winden (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics* 102(1): 23-29.

Knez, M. and D. Simester (1998). Firm-wide incentives and mutual monitoring as continental airlines. *Graduate School of Business, University of Chicago*.

Latane, B. and J. Darley (1970). *The unresponsive bystander: Why doesn't he help?* New York, Appleton-Century-Crofts.

Ledyard, J. (1995). Public goods: A survey of experimental research. *The handbook of experimental economics*. J. Kagel and A. Roth Eds. Princeton, Princeton University Press: 111-194.

Miller, J., C. Butts and D. Rode (2002). Communication and cooperation. *Journal of Economic Behavior and Organization* 47: 179-195.

Moir, R. (1998). Spies and swords: Costly monitoring and sanctioning in a common-pool resource environment. mimeo.

Ostrom, E., R. Gardner and J. Walker (1994). *Rules, games and common-pool resources*. Ann Arbor, University of Michigan Press.

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

Page, T. and L. Putterman (2000). Cheap talk and punishment in voluntary contribution experiments. mimeo.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature* 37(1): 7-63.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review* 83(5 December): 1281-1302.

Sefton, M., R. Shupp and J. Walker (2000). The effect of rewards and sanctions in provision of public goods. mimeo.

Sethi, R. (1996). Evolutionary stability and social norms. *Journal of Economic Behavior and Organization* 29: 113-140.

Sethi, R. and E. Somanathan (2001). Understanding reciprocity. *Journal of Economic Behavior and Organization* forthcoming.

Taylor, P. and L. Jonker (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences* 40: 145-156.