ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Carpenter, Jeffrey P.

Working Paper

Punishing Free-Riders : How Group Size Affects Mutual Monitoring and the Provision of Public Goods

IZA Discussion Papers, No. 1337

Provided in Cooperation with:
IZA Network @ LISER, Luxembourg Institute of Socio-Economic Research (LISER)

Suggested Citation: Carpenter, Jeffrey P. (2004) : Punishing Free-Riders : How Group Size Affects Mutual Monitoring and the Provision of Public Goods, IZA Discussion Papers, No. 1337, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/20606

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 1337

Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods

Jeffrey P. Carpenter

October 2004

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods

Jeffrey P. Carpenter
Middlebury College and IZA Bonn

Discussion Paper No. 1337
October 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 1337
October 2004

ABSTRACT

Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods*

Because costly punishment is not credible, subgame perfection suggests that punishment will not deter free riding, regardless of the size or structure of groups. However, experiments show that people will punish free riders, even at considerable cost. To examine the implications of agents who punish, we simulate an environment populated with behavioral strategies seen in the lab and use the simulation to develop hypotheses about why group size should matter when punishment is allowed. We test these hypotheses experimentally and 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 that large groups contribute at rates no lower than small groups because punishment does not fall appreciably in large groups. However, hindrances to monitoring do reduce the provision of the public good.

JEL Classification: C72, C92, H41

Keywords: public goods, punishment, group size, experiment, simulation

Jeffrey P. Carpenter
Department of Economics
Middlebury College
Middlebury, VT 05753
USA
Email: jpc@middlebury.edu

* I thank Corinna Noelke, Peter Matthews and a tough but fair referee for thoughtful comments. This research is funded by the MacArthur Foundation Norms and Preferences network and the National Science Foundation (SES-CAREER 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

Recent experiments have illustrated that public goods can be provided at high levels if participants are allowed to monitor the decisions made by other participants and punish behavior deemed antisocial (Fehr and Gaechter, 2000a; Bowles et al., 2001; Bochet et al., 2003; Sefton et al., 2000). This result is contrary to theories of collective action based on subgame perfection because punishment, in these experiments, is really just a second order public good and therefore participants could always do better by free riding on the punishment done by others.

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., 2004; Carpenter and

Matthews, 2002).[1] 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 extending earlier results in three new directions. On the methodological frontier, we not only abandon the "straw man" of standard economic theory as a starting point from which to generate hypotheses about behavior, we also depart from pen and paper theory altogether and use simulated automata to generate hypotheses about real behavior. One benefit, that we take advantage of, in utilizing simulations is that our model can take the heterogeneity of behavior seen in many experiments much more seriously. At the same time however, one might object that simulating automata requires us to sacrifice on the generality of our predictions. To the contrary, we show that the patterns generated by our agents are robust enough to changes in the structural parameters of the simulation to not only provide interesting predictions for behavior in the lab, but to also be of interest to more accomplished theorists.

On a more practical level, we are interested in exploring the effectiveness of punishment to deter free riding when groups grow in size and information about the contribution choices of all the other group members is not available. If we assume that 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

[1] 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 on material payoffs alone.

people in large groups. Instead, we might consider a logistical hypothesis which states that large groups are less able to elicit contributions because monitoring 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, in this case if the "monitoring groups" overlap such that everyone is seen by some minimum fraction of the rest of the group, then punishment might still deter free riding.

To better understand the effects of group size and the information agents have about each other on the effectiveness of monitoring and punishment, in section 3 we present our simulations that are constructed with agents that mimic the strategies that are commonly described in previous experiments. The simulations allow our agents to interact in the standard public goods experimental environment to generate hypotheses that we expect will be externally valid to human interactions. 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 use game environments that do not allow free riders to be sanctioned or do not consider the logistical problems of large groups. One of the main results 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 (4 and 10 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 (40 and 100 persons) are examined, the authors find that contributions increase relative to smaller groups and that the effect is independent of the return on the public good.[2]

Interest is growing in the question of whether punishing 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).[3] Andreoni showed that contributions decayed as would be expected by an equilibrium learning hypothesis, but contrary to learning, he 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 model of team production 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. Bochet et al. (2003) also confirm that punishment

[2] 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).

[3] 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).

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, 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 Size Hypotheses

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 that mimic the different behavioral types we typically see in the experimental lab and have them play a public goods game with some selection pressure.[4] Surveying the experimental literature on punishment in public goods experiments, one generally finds four behavioral types: unconditional free riders, unconditional cooperators,

[4]This methodology follows Miller et al. (2002) and Sethi and Somanathan (2001) to one degree or another.

reciprocally-minded tit-for-taters, and cooperators who punish free riders.[5]

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 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.

Figure 1 here

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 $f \geq \alpha_i$ 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 in a given

[5] The surveyed papers include those discussed in section 2 plus Fehr and Gaechter (2000b), Keser and van Winden (2000), and Fischbacher et al. (2001).

period 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 but as close to the experimental environment as possible, punishers incur a cost, c, to punish each free rider they stumble across and each sanction inflicts $6.35c$ harm on the target. The ratio of harm to cost is based on pilot experiments (see Carpenter, 1999) in which participants spent an average of 3.54 experimental francs to reduce the earnings of their targets by an average of 22.49 francs.

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 resources available. Nice Punishers contribute in the passive state and once f rises to the level of 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 which is shared equally. Because $q/n < 1 < q$, 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:

$$\pi_i = (1 - x_i) + \frac{q \sum_j x_j}{n} - c \sum_j p_{ij} - 6.35c \sum_j p_{ji}$$

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.

In the standard public goods experiment the distribution of contributions changes as individuals adopt different strategies over time. This social learning process is qualitatively similar to many recent innovations in learning dynamics that are based on the logic of evolution (e.g., Miller and Andreoni, 1991). We allow our simulated population to evolve in the following sense. At the end of each period a fraction of the population is replaced 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) which many interpret as a model of social learning in which individuals compare their current payoffs to the population average payoff and switch strategies with positive probability if they are doing worse than the average. In our version of this process, the lowest performing 8% of the agents in the population, in terms of accumulated earnings, are culled every period and replaced by agents who choose types by randomly selecting one of the existing group members and imitating it.[6] 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 every agent faces a 1 in 100 chance of having its trigger randomly changed.

The simulations were conducted in an institutional environment that was

[6] 8% seems like an odd culling rate, but it assures that a round number of agents is replaced each period.

nearly identical to the experimental protocol described in the next section. Each simulation lasted for 10 periods and at the beginning of each period the agents were randomly reshuffled into new groups (i.e., the strangers protocol). In each session there were 5 groups participating at the same time. The small groups had 5 members and the large groups had 10 members. The agents faced two values of q/n (a.k.a. the marginal per capita return, $MPCR$), 0.375 or 0.75. Combined, these parameters generate marginal group returns, $q = (q/n)n$, that overlap off the diagonal. Specifically, in the small group, low productivity treatment $q = (5 \times 0.375) = 1.875$ and in the large group, high productivity treatment $q = 7.5$, however, the two off-diagonal treatments are where $q = 3.75$. Therefore, these two treatments can be directly compared while controlling for the social return of contributions.

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: what happens to the contribution rate when group size increases and what happens to the contribution rate when the monitoring fraction falls. To do so, we started each simulation from a balanced population (i.e., 20% of each type) and accounted for the likelihood that, depending on the initial distribution of types across groups, different time paths would result by running 1000 realizations of the model for each treatment. We report the average contribution rate over these 1000 realizations.

Although the simulations discussed below were used to generate hypotheses about the effect of monitoring technology and group size, other (unreported) simulations were conducted to explore the equilibrium properties of the system. These simulations indicated that most realizations achieved a stable distribution of types within 50 periods. Individual realizations of the simulation have the same characteristics as individual groups in the experimental lab. In many

of the realizations the session was taken over by free riders and contributions declined to zero. However, it was not uncommon for the sessions 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.

Examining the effect of group size is simple to simulate. We compared the rate of contribution in 5 agent groups to that in 10 agent groups. However, simulating the effects of reduced monitoring fraction is slightly more complicated. In the *No monitoring* condition, agents saw whether the other group members contributed or not, but could not punish them. In this case, the population collapses to three types: unconditional cooperators, unconditional defectors, and tit-for-taters. In the *Full monitoring* condition, agents could punish all the other group members provided they had enough resources to do so. In the *Half monitoring* condition, agents could punish the half the other group members who were located closest to them on the circle. Using this structure, every agent was monitored by the same number of teammates. Finally, in the *Single monitoring* condition agents were allowed to punish only one other group member – the agent to their immediate right.

Figure 2 here

Figure 2 shows the results of the simulations. For each combination of group size and $MPCR$, Figure 2 reports the average agent contribution rate for all four monitoring conditions. Two patterns are particularly conspicuous. Independently of group size and the $MPCR$, large groups contribute at rates no

lower than small groups and there is clear bifurcation with respect to the monitoring fraction.

Apparently, the effect of having more eyes watching dominates the effect of running out of resources with which to punish. Comparing the contribution rates from the first row of Figure 2 to the second row forms the basis for our first hypothesis.

The Group Size Hypothesis: *For each monitoring condition, contributions do not decrease as the number of group members increases.*

Considering the effect of monitoring fraction, Figure 2 illustrates that there appears to be an ordering of contribution rates that is, again, independent of group size and the $MPCR$. Specifically, Full $>$ Half $>$ Single. However, the more striking result is that Full and Half monitoring seem to attenuate the gradual decline of contributions (i.e., prevent free riders from taking over the population) while Single monitoring does no better than the control in which there is no punishment. These patterns inform our second hypothesis.

The Monitoring Fraction Hypothesis: *When agents monitor each other, contributions will decrease as the fraction of the group each punisher can monitor decreases. Further, the difference in contributions will be most noticeable between the Half and Single conditions.*

Before we proceed to the experiment, it is important to establish that these simulation results are robust to changes in the structural parameters. In general, changing the parameters of interest (the ratio of punishment cost to the harm inflicted, the number of groups interacting at the same time, the mutation rate for agent triggers and, the rate at which individuals are culled) may change slightly the levels in Figure 2 but not the patterns of, or the separation between, monitoring conditions. The same patterns emerge for punishment ratios greater

than or less than 6.35 for 1. For example, increasing the ratio to 8 or 10 for 1 results in levels of contributions that are almost identical except that there is slightly more of a difference between the No monitoring condition and the Single monitoring condition in the low $MPCR$ treatment. Increasing the number of groups participating in the same "session" from 5 to 10 or 20 has no discernable effect on the controls without punishment although there is a slight effect on the level of the Single monitoring conditions (i.e., the change in the contribution rate is $\pm 5\%$). This change also has a small effect on the Full and Half monitoring time paths: they both approach 100% contributions a little faster; however, the slight difference between the two conditions persists. Changing the trigger mutation rate has little effect on the contribution time paths but increasing the culling rate initially shifts contributions slightly upward (moving from 8% to 16%) in the three monitoring conditions and then reduces them (moving from 16% to 24%). Again, however, the differences between conditions persist in each cell of the design.

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 punishment with as difficult a test as possible and also to relate more directly to the standard literature. 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 new groups 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 to allow participants to monitor and punish each other. Let each participant be endowed with w_i experimental francs and recall from the previous section that p_{ij} is the number of points (sanctions) 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. As in the simulation, the experimental subjects participated in either the low $MPCR$ treatment ($MPCR = 0.375$) or the high $MPCR$ treatment ($MPCR = 0.75$).

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 members could reduce a player's gross earnings to zero, but no further (unless the target also punished other players).[7] The punishment cost function used in the experiment, $c(p_{ij})$, was

Points Given to 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

[7] 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. However, as mentioned above, the ratio of punishment cost to the harm punishment has on the target in the simulation was calibrated by pilot experiments using the current protocol.

Along with changes in the $MPCR$, group size was manipulated in the same two ways as in the simulations. To test the group size hypothesis, we ran sessions with both 5 person and 10 person groups. To investigate the monitoring fraction hypothesis, we varied the number of other subjects that each player could monitor and punish and, again, we replicated the same 4 conditions as in the simulations. 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.

Including all possible combinations of the parameters results in 16 different treatments (2 $MPCRs$ $\times$ 2 group sizes $\times$ 4 monitoring conditions). A total of 39 sessions were conducted with 735 participants. Subjects were recruited by email from the general student population. 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 \$19.90 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 and punishment 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 benefited the individual, and a public account which benefited 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 that they were allowed to monitor. 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.

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. Figure 3 plots the time path of the average fraction of the endowment contributed to the public good for each of the 16 treatments. To visually examine the effect of group size one compares the first row (group size = 5) to the second row (group size = 10). While the effect is more pronounced in the low $MPCR$ condition, it is clear that contributions do not decline when we double the size of the groups. In five person groups, Full and Half monitoring largely maintains the initial level of contributions while in 10 person groups contributions increase to nearly 100 percent by the end of the experiment. Additionally, for the other two monitoring conditions contributions are not lower in 10 person groups.

Figure 3 here

The degree to which the data conforms to the monitoring fraction hypothesis is also striking. With the exception of the 5 person, $MPCR$=0.375 cell of the design, there is clear separation between the Half and Single monitoring

treatments. Contributions in the Full and Half monitoring conditions, which follow each other closely, are noticeably higher than in the Single and No monitoring conditions which also tend to overlap. By the end of the experiment, the predicted separation has also developed in the 5 person, $MPCR$=0.375 case. Comparing Figures 2 and 3, it is clear that the experiment conforms almost entirely to the simulated hypotheses.

To bolster our visual examination, we can formally test our hypotheses by regressing individual contributions on the structural parameters of the game. Because our contribution data is truncated from above and below, we use the double-censored Tobit regressor, and because our experiment generates a panel of data, we include individual random effects to account for cross-sectional differences. Table 1 reports the results. To test the effect of group size (and the $MPCR$), we pool the data from all four cells in column (1) and find that larger groups do not elicit contributions that are lower than small groups. In fact, controlling for the different monitoring conditions, contributions are significantly higher in large groups ($p<0.01$). In accordance with previous experiments (see Ledyard, 1995), we see that contributions are also significantly higher in the higher $MPCR$ treatments ($p<0.01$). Based on column (1) we conclude that the experimental data support the group size hypothesis.

Table 1 here

We can also use column (1) of Table 1 to test the monitoring fraction hypothesis. The omitted treatment in the regression is the No monitoring condition. Therefore, relative to not punishing at all, both the Full and Half monitoring conditions demonstrate significantly more contributions ($p<0.01$ in both cases) and the Single monitoring condition elicits significantly few contributions

($p<0.01$). These coefficients alone conform with the assertion that there is separation between the Full and Half conditions on one hand and the No and Single conditions on the other. However, we can be more specific by directly comparing the point estimates. Contributions in the Full condition are marginally significantly higher than in the Half condition ($p=0.046$), and contributions in the Half condition are significantly higher than in the Single condition ($p<0.01$). Combining all these statistics, we conclude that both the more specific claim of the monitoring fraction hypothesis – that contributions can be ordered Full $>$ Half $>$ Single – and the less restrictive assertion – that there is separation between the Half and Single conditions - are supported by the experiment.

Columns (2) through (5) of Table 1 disaggregate the data into the four main cells of the experiment (the cells are identified in the row labelled, "Treatment") as a robustness check on our monitoring fraction analysis. As was apparent in Figure 2, the part of the monitoring fraction hypothesis that predicts that the most significant separation will be between the Half monitoring condition and the Single monitoring condition is fully supported in three of the four cells, but finds more limited support in the 5 person, $MPCR$=0.375 case (column (2)). However, the support from regressions is still strong in each case. From columns (3) through (5) we see that there are no significant differences in contributions between the Single monitoring condition and the No monitoring control. Further, there are no significant differences between Full and Half point estimates in any of the four cells. Hence, while the results of the individual cells do not conform as neatly to the Full $>$ Half $>$ Single ordering of contributions, in each case, there is significant bifurcation between the Half and Single conditions.

From both the simulations and the experiment it is clear that, independent of changes in the structural parameters of the game, when agents punish free riders contributions can be sustained at high levels in large groups as long as punishers

are able to monitor sufficiently many of their teammates. Because we know how the simulated agents are "wired," we know what must be driving these results in the simulations. In terms of group size, the potentially dampening effect of being responsible for more agents to monitor in large groups is overwhelmed by the fact that there are more punishers and, at the same time, we seen that punishing agents must be able to monitor at least half of their teammates for punishment to deter free riding. Although our automata are programmed to either contribute or not and to punish or not and this programming is invariant to the size of the group, the same motivation need not be driving our, more complex, experimental participants. That is, despite the striking similarities in the data, the reasons that contributions from human participants follow the paths illustrated in Figure 3 may be different from the reasons underlying the simulated paths in Figure 2. For this reason 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), psychologists have shown that, for example, bystanders are less likely to help a stranger when many people have the opportunity to help. In essence a coordination problem arises - people agree help should be provided, but it is unclear in a large group exactly who should help. In our experiment, punishing free riders is a prosocial act like helping strangers. While the automata are programmed to punish without regard to the number of other potential punishers, we now examine whether the size of the monitoring pool affects the punishment choices of our human participants.

Table 2 here

In Table 2 we analyze whether our participants conditioned their punishment expenditures on the number of other potential punishers. Notice that simply regressing punishment expenditures on group size is not sufficient because the bystander effect (i.e., that expenditures depend on the number of potential punishers) is confounded by the fact that there are also different numbers of free riders when we change the group size or the monitoring group size. To control for the amount of free riding in the group we include indicators for the monitoring fraction which controls for the total number of free riders a punisher can be exposed to in addition to the average contribution and the variance in contributions in the punisher's group. We can then be fairly certain that group size differences above and beyond the differences associated with the amount of free riding one is exposed to can be attributed to a possible bystander effect.[8]

As is now the pattern with our analysis, in column (1) we pool the data and in columns 2 through 5 we analyze the data by cell of the design (again the row "Treatment" identifies each cell of the experiment) to test the robustness of any monitoring fraction results. If our human participants behave as our automata are programmed, we should see no group size effect after controlling for the amount of free riding in the groups. Column (1) of Table 2 suggests that our human participants are indeed more complex than the automata and act like the bystanders examined in Latane and Darley (1970). People in 10 person groups spend approximately 2.5 experimental francs less than people in 5 person groups to punish other group members ($p<0.01$). This amount is significant considering the pooled average expenditure in the experiment is 2.81 francs.

The regression reported in Column (1) of Table 2 highlights other interesting results. Punishment expenditures are higher when the $MPCR$ is higher. One

[8] We employ the same estimation strategy as in Table 1 with one exception: punishment expenditures are not bound from above. For that matter, each regression discussed in the paper uses essentially the same strategy – random effects Tobit.

might expect this to be a wealth effect because for a given level of free riding individuals will have higher gross earnings when the $MPCR$ is higher. However, in an unreported regression we added the gross earnings of the individuals and while it reduced the coefficient on the $MPCR$, the effect persists and is still statistically significant (p=0.03). A possible alternative explanation is that players punish more when the stakes are higher. We also see that participants punish less in groups that have higher contributions (p<0.01) and in groups where contributions vary less (p<0.01) although the effect of variance is small. Further, as one might expect contributors spend more money on punishment than free riders do: a standard deviation increase in one's contribution (7 experimental francs) increases punishment expenditures by 1.54 francs.

Column (1) of Table 2 also shows that punishment expenditures are affected by the monitoring fraction. Expenditures in both the Half monitoring and Single monitoring conditions are significantly lower than in the Full monitoring condition (p<0.01 in both cases). However, the expenditures in the Half and Single conditions are not significantly different (p=0.58) in the pooled data. As columns (2) through (5) suggest this overall pattern is replicated in two of the four cells. When the $MPCR$ is low (columns (2) and (4)) there is significant separation between the expenditures in the Half and Single conditions but when the $MPCR$ is large (columns (3) and (5)) this difference vanishes. Combining the fact (from Table 1) that contributions are vastly different between the Half and Single conditions with the current results that expenditures on punishment are not lower in the Single condition than in the Half condition suggests that the transparency of contributions matters a lot. Specifically, punishment must have more of an effect on free riders in the Half treatment. We analyze this possibility below, but before doing so, we examine punishment from the perspective of receiving it rather than meting it out.

Table 3 here

Table 3 illustrates that the results are similar when we examine the amount of punishment received. In column (1) we see that people receive fewer points in large groups after controlling for how badly they free rode and the number of players monitoring them. In other words, free riders do benefit from the bystander effect. In addition, free riders are punished more severely in the high $MPCR$ treatment and in the pooled data they are punished more severely when they are monitored by more teammates. In the latter case, both the Half and Single indicators are significantly different from the points received in the Full condition, they are also significantly different from each other ($p<0.01$). When we break the analysis down by treatment in columns (2) through (5), however, only in the two cases which share the same marginal social product (columns (3) and (4)), do the number of points received conform to our pooled results. In the small groups with low $MPCR$s (column(2)), free riders actually receive marginally significantly fewer points in the Half condition than in the Single condition ($p=0.06$) and in large groups with high $MPCR$s (column (5)), the number of points received in the Half and Single conditions are statistically equal ($p=0.94$).

Taking stock, with respect to meting out punishment and being punished, our human participants behave similarly to the programming of our automata, but slight differences do exist. Restricting ourselves to the pooled analyses, we do find that punishment expenditures are increasing in the amount of free riding in groups and in the number of other group members that one can monitor. However, unlike our automata, human participants seem to be influenced by the bystander phenomenon and the productivity of the public good.

In addition to looking for differences in the heuristics used to assign punishment, we can also investigate how human participants react to being punished. Although our automata are not explicitly programmed to react to punishment (e.g., they are not more likely to contribute after being punished), they are affected by punishment because being punished affects the likelihood of being culled from the population. If free riding agents are punished severely enough, their payoff falls to the point where they are replaced by a new agent who imitates another surviving member of the population. As the average payoff to free riding falls, new agents are more likely to imitate contributing strategies which are less likely to be culled. In this sense, the population reacts to the level of free rider punishment. Our human participants, on the other hand, will react to punishment at the individual level.

In Table 4 we examine individual reactions to being punished. We separate the data between free riders, who contribute less than the group average, and contributors who give more than the group average because it should make intuitive sense that these two groups will react differently to being punished. To model responses to punishment we regress the difference in a player's contributions between periods t and $t-1$ on the structural parameters of the game, lags of the group average contribution and the number of points assigned by punishers, and interactions between the monitoring conditions and the amount of punishment received.

Table 4 here

Beginning with free riders in column (1) and contributors in column (3), the first thing to notice is that increases in group size and the $MPCR$ are both associated with larger increases in contributions and the magnitudes of

the effects are similar for free riders and contributors. The effect of the $MPCR$ may again have to do with the stakes involved and, given the number of controls, the group size effect may indicate a larger role for conformity when there are more group members to imitate (see Carpenter, 2004).

It is interesting that after accounting for the effects of the group size, the $MPCR$, inertia (i.e., the lagged group average contribution), and punishment, there are still significant direct effects of the monitoring conditions. Free riders increase their contributions less in the Half and Single monitoring conditions ($p=0.07$ and $p=0.05$, respectively) and contributors increase them less in the Single monitoring condition ($p<0.01$) compared to the Full monitoring condition. In addition, free riders and contributors are both affected by inertia in group level contributions. In both cases, the higher were group contributions last period, the less they respond this period. Of course this is partially due to the fact that contributions are capped, but it is also interesting that free riders react more negatively than contributors to inertia. If the coefficient on the lagged average group contribution was entirely a proxy for the capping of contributions, we would expect contributors responses to be larger than free riders. We examine these two phenomenon in more detail below.

In comparison to the more mysterious effects of inertia and the monitoring condition direct effects, the overall role of punishment is more straightforward. We see that punishment causes free riders to increase their responses and punishment causes contributors to reduce their responses. There are two plausible explanations for the response of contributors to being punished: either they are spiteful or they do not have a good grasp of the incentives of the game and believe that reducing their contributions will lead to higher payoffs. In any case, punishment leads contributions to regress to the mean over time.

We can now sharpen the analysis by interacting terms. The interaction of

the points received last period and the Half and Single indicators allows us to ask if punishment has more or less of an effect on behavior in the different monitoring conditions. These interactions help us understand the dynamics of Figure 3. The significantly positive coefficient on the Half $\times$ Points Received$_{t-1}$ interaction in column (1) suggests that free riders respond almost twice as much to punishment in the Half condition than in the baseline. Hence, while free riders receive less punishment in the Half condition (Tables 2 and 3), they react more and this causes the effect of punishment to be similar and the contribution time paths of the Full and Half conditions to converge.

By contrast, free riders respond significantly less to punishment in the Single condition, perhaps because they fear punishment less or resent it more. That is, the more punishment free riders receive in the Single condition compared to the Full condition, the more they stubbornly free ride. This result is interesting because it implies 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.

Contributors, on the other hand, do not respond differently to punishment in the Half condition which also partially explains why the Full and Half contribution time paths in Figure 3 are similar. However, contrary to the responses of free riders, in the Single condition contributors appear to be significantly more embarrassed when just one other group member punishes them than when they are seen by more eyes.

To conclude our analysis of player reactions to punishment we try to offer a partial explanation for why the monitoring conditions have direct effects in addition to any differential effects they have on the amount of punishment players receive. One possibility is that in addition to being correlated with the

amount of punishment players receive, the monitoring conditions might also be correlated with the amount of inertia in group contributions. To explore this possibility, we add additional interactions between the monitoring conditions and the lag of the group average contribution in columns (2) and (4). As one can see by comparing columns (1) and (2) for free riders, controlling for differential inertial effects causes the direct effect of the Single condition to no longer be significant and reduces the significance of the Half monitoring indicator from p=0.07 to p=0.10. As for the interactions themselves, we find no difference in the effect of inertia in the Half monitoring condition, but find that free riders respond even less to increases in the group average in the Single monitoring condition. This fact also helps to explain why contributions in the Single condition mimic those in the No monitoring control. Lastly, notice that adding the interactions to the contributor regression has the same effect. In column (4) of Table 4 we see that neither monitoring condition indicator is significant and contributors respond like free riders to inertia in the Single condition.

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, punishment 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 and think harder about the implications of growing groups for the deterrence value of punishment, group size does matter in the sense that it may hinder the ability of norm enforcing agents to catch all the free riders before their contagion affects the entire group.

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 too much consideration for the fact that other monitors might also punish. While our experimental data matches our simulation data rather well, our human agents do not behave identically to our automata. However the behavioral differences to not appear to be large enough to affect the outcome appreciably. Even though real players seem to be somewhat afflicted by bystander psychology (i.e., they punish less in large groups), the effect is not large enough to diminish the stronger effect of more monitors punishing each free rider.

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

free riding or mimic hard work, monitoring group size will fall and shirking will become more prevalent.

7 References

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

Bochet, O., Page, T., Putterman, L., 2003. Communication and punishment in voluntary contribution experiments. Journal of Economic Behavior & Organization. forthcoming.

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

Bowles, S., Carpenter, J., Gintis, H., 2001. Mutual monitoring in teams: Theory and evidence on the importance of residual claimancy and reciprocity. mimeo.

Carpenter, J., 1999. Mutual monitoring in teams: Theory and experiments. mimeo.

Carpenter, J., 2004. When in Rome: Conformity and the provision of public goods. Journal of Socio-Economics. forthcoming.

Carpenter, J., Matthews, P., 2002. Social reciprocity. Middlebury College Department of Economics Working Paper 2002-29.

Carpenter, J., Matthews, P., Ong'ong'a, O., 2004. Why punish? Social reciprocity and the enforcement of prosocial norms. Journal of Evolutionary Economics. forthcoming.

Davis, D., Holt, C., 1993. Experimental economics. Princeton University Press, Princeton.

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

Fehr, E., Gaechter, S., 2000b. Fairness and retaliation: The economics of reciprocity. Journal of Economic Perspectives. 14, 159-181.

Fischbacher, U., Gaechter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Economic Letters. 71, 397-404.

Hansen, D., 1997. Worker performance and group incentives: A case study. Industrial and Labor Relations Review. 51, 37-49.

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

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

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

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

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

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

Miller, J., Andreoni, J., 1991. Can evolutionary dynamics explain free riding in experiments? Economic Letters. 36, 9-15.

Miller, J., Butts, C., Rode, D., 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., Gardner, R., Walker, J., 1994. Rules, games and common-pool resources. University of Michigan Press, Ann Arbor.

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

Sefton, M., Shupp, R., Walker, J., 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., Somanathan, E., 2001. Understanding reciprocity. Journal of Economic Behavior and Organization. 50, 1-27.

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

8 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 0.3×75 = 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 $0.3\times1 = 0.3$ 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:	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.

9 Tables and Figures

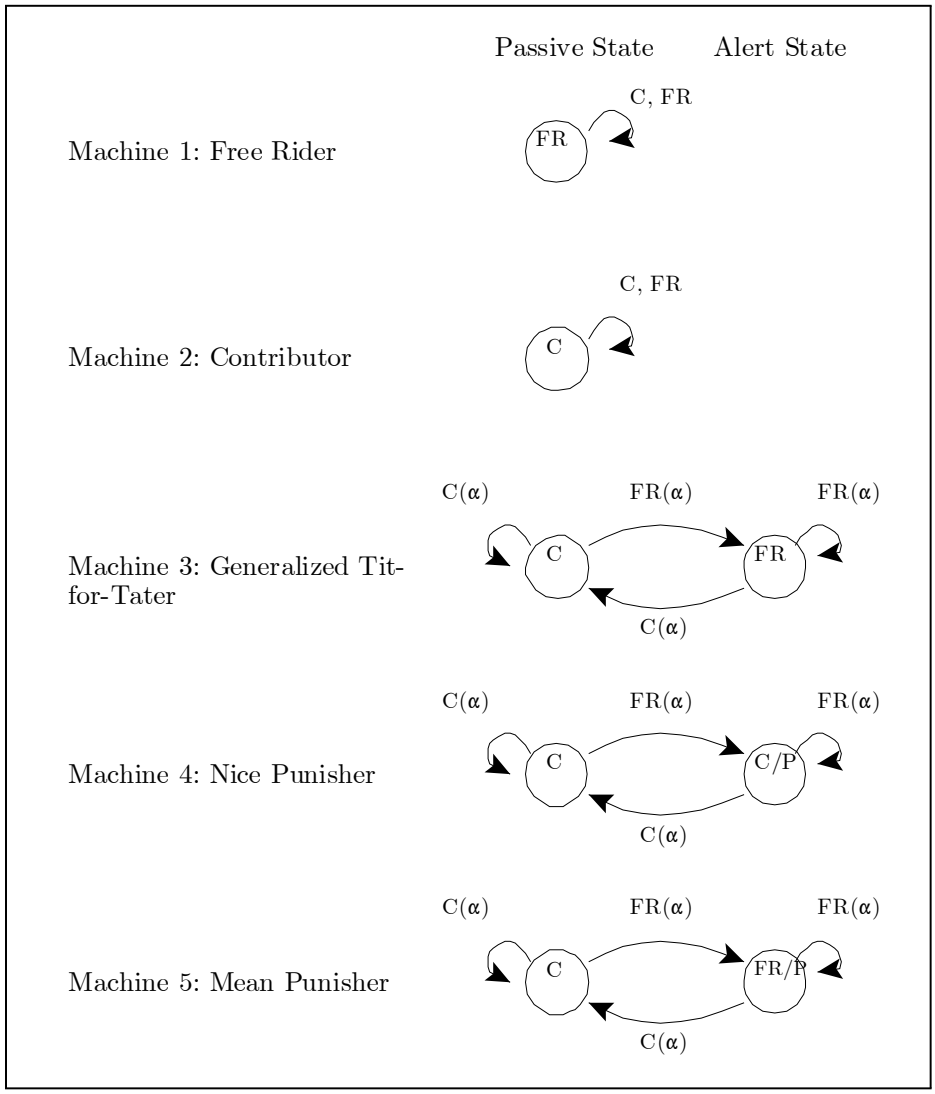

Figure 1: The Simulated Two-State Automata. C indicates Contribute, FR indicates Free Ride, P indicates punish free riders, and α is the trigger frequency of free riding or contributing that determines behavior.

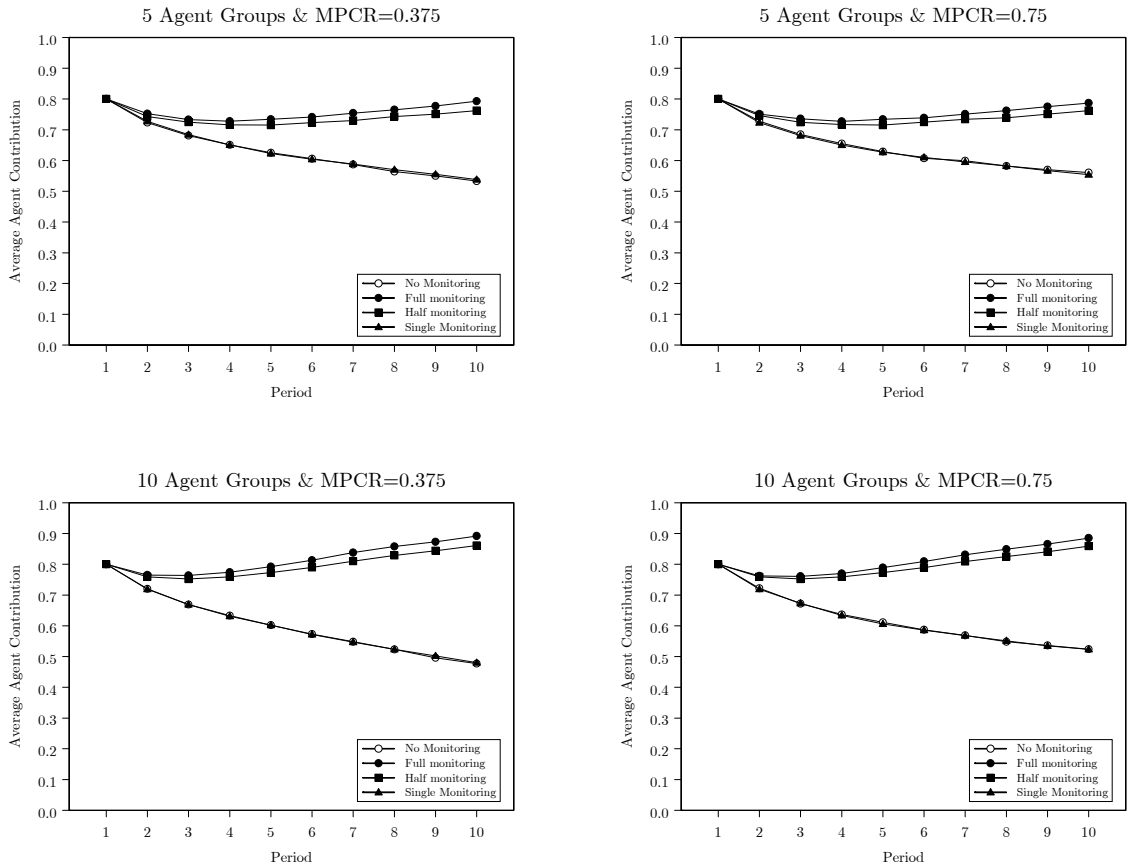

Figure 2: Automata Contribution Rates

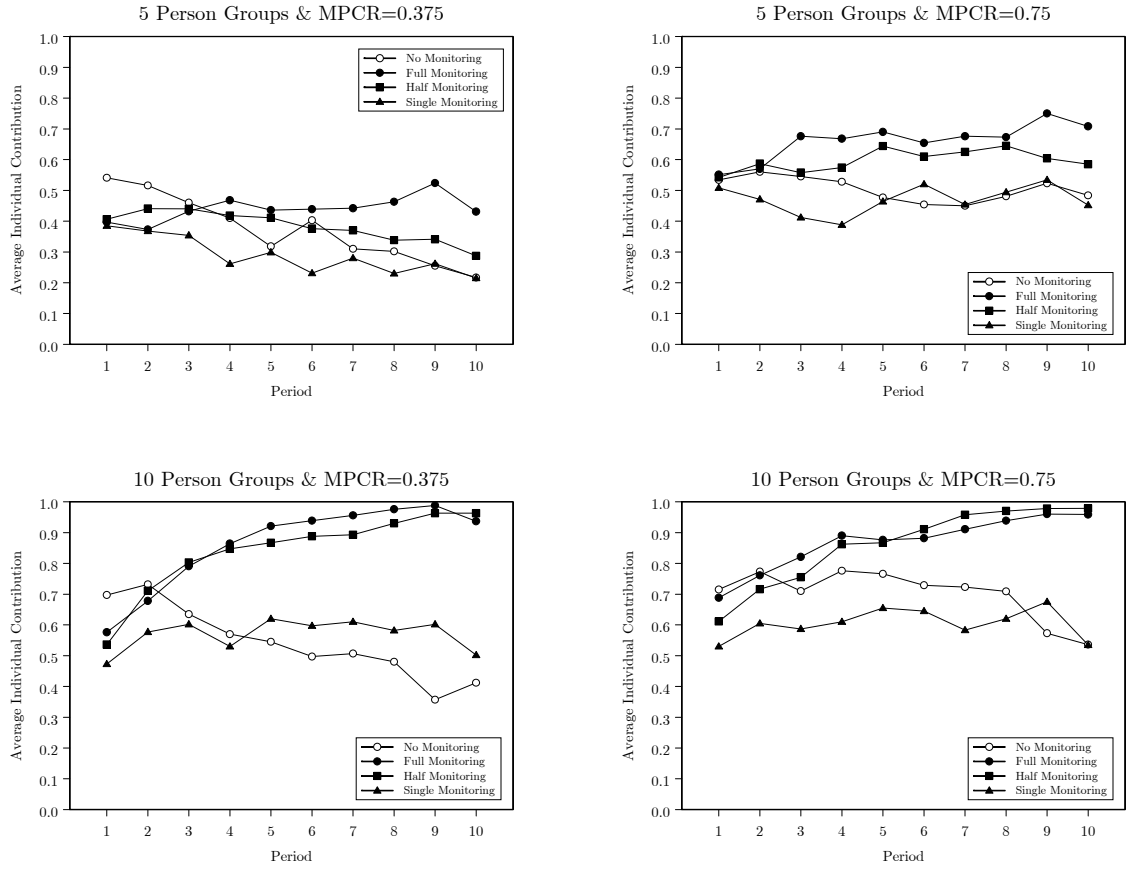

Figure 3: Contribution Rates in the Experiment

Do the Treatments Affect Contributions? (Dependent variable is individual i's contribution in period t, $x_{i,t}$)					
	(1)	(2)	(3)	(4)	(5)
Group Size (5 or 10)	1.59***				
	(0.10)				
MPCR (0.375 or 0.75)	9.16***				
	(1.31)				
Full Monitoring	4.15***	1.69*	2.82***	11.31***	3.39***
	(0.69)	(0.96)	(0.72)	(2.10)	(0.85)
Half Monitoring	2.68***	0.56	3.16***	9.89***	3.79***
	(0.69)	(0.72)	(0.85)	(2.15)	(0.95)
Single Monitoring	-1.99***	-1.47**	-0.55	0.11	-1.33
	(0.65)	(0.72)	(1.45)	(2.05)	(0.94)
Intercept	-4.83***	6.96***	10.08***	11.66***	13.90***
	(1.12)	(0.55)	(0.40)	(1.86)	(0.76)
Treatment:					
(Group Size, MPCR)	Pooled	(5, 0.375)	(5, 0.75)	(10, 0.375)	(10, 0.75)
Wald χ^2	535***	15***	24***	89***	64***
Obs., individuals	7350, 735	2500, 250	1650, 165	1600, 160	1600, 160

Notes: (i) The regression is a Tobit with both upper and lower censoring and individual random effects, (ii) standard errors in parentheses, (iii) *** indicates significance at the 99% level, ** at the 95% level and * at the 90% level.

Table 1:

Do the Treatments Affect Expenditures on Punishment? (Dependent variable is individual i's expenditure on punishment in period t, $\sum c_{i,t}$)					
	(1)	(2)	(3)	(4)	(5)
Group Size (5 or 10)	-0.50***				
	(0.13)				
MPCR (0.375 or 0.75)	8.32***				
	(1.53)				
Half Monitoring	-5.33***	-2.71***	-3.27***	-2.36***	-9.11***
	(0.71)	(0.73)	(1.23)	(0.86)	(2.91)
Single Monitoring	-5.76***	-0.92	-3.80***	-4.76***	-9.33***
	(0.68)	(0.75)	(1.07)	(1.09)	(2.47)
Average Contribution	-0.52***	-0.25***	-0.43***	-0.46***	-0.61***
	(0.05)	(0.06)	(0.10)	(0.09)	(0.23)
Contribution Variance	0.02***	0.004	0.01	0.06***	0.05*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
Punisher's Contribution	0.22***	0.17***	0.24***	0.26***	0.20*
	(0.03)	(0.03)	(0.06)	(0.04)	(0.11)
Intercept	4.71***	3.39***	5.61***	1.04	7.01
	(1.17)	(0.73)	(1.52)	(1.63)	(4.48)
Treatment:					
(Group Size, MPCR)	Pooled	(5, 0.375)	(5, 0.75)	(10, 0.375)	(10, 0.75)
Wald χ^2	253***	48***	41***	122***	33***
Obs., individuals	5250, 525	1650, 165	1000, 100	1400, 140	1200, 120

Notes: (i) The regression is a Tobit with both upper and lower censoring and individual random effects, (ii) standard errors in parentheses, (iii) *** indicates significance at the 99% level and * at the 90% level.

Table 2:

Do the Treatments Affect the Amount of Punishment One Receives? (Dependent variable is the amount of points individual i receives in period t, $\sum p_{i,t}$)					
	(1)	(2)	(3)	(4)	(5)
Group Size (5 or 10)	-0.19***				
	(0.02)				
MPCR (0.375 or 0.75)	4.10***				
	(0.31)				
Target's Contribution	-0.26***	-0.14***	-0.18***	-0.53***	-0.30***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)
Half Monitoring	-2.10***	-2.10***	-1.14***	-1.12***	-4.07***
	(0.14)	(0.22)	(0.37)	(0.42)	(0.36)
Single Monitoring	-2.58***	-1.68***	-2.35***	-4.22***	-4.05***
	(0.14)	(0.23)	(0.38)	(0.47)	(0.38)
Intercept	4.59***	4.28***	5.29***	7.60***	7.27***
	(0.24)	(0.20)	(0.38)	(0.49)	(0.49)
Treatment:					
(Group Size, MPCR)	Pooled	(5, 0.375)	(5, 0.75)	(10, 0.375)	(10, 0.75)
Wald χ^2	1306***	180***	95***	378***	258***
Obs., individuals	5250, 525	1650, 165	1000, 100	1400, 140	1200, 120

Notes: (i) The regression is a Tobit with both upper and lower censoring and individual random effects, (ii) standard errors in parentheses, (iii) *** indicates significance at the 99% level.

Table 3:

Do Responses to Punishment Differ by Treatment? (Dependent variable is the difference in contributions between periods *t* and *t-1*)				
	(1)	(2)	(3)	(4)
Group Size (5 or 10)	0.47***	0.48***	0.38***	0.40***
	(0.06)	(0.06)	(0.06)	(0.06)
MPCR (0.375 or 0.75)	1.74**	1.82***	1.57***	1.79***
	(0.73)	(0.73)	(0.59)	(0.59)
Half Monitoring	-0.77*	-1.47*	-0.25	0.30
	(0.43)	(0.88)	(0.28)	(0.80)
Single Monitoring	-0.86**	0.65	-1.83***	0.48
	(0.43)	(0.88)	(0.34)	(0.81)
Average Contribution $_{t-1}$	-0.25***	-0.24***	-0.09***	-0.05
	(0.04)	(0.05)	(0.03)	(0.04)
Points Received $_{t-1}$	0.32***	0.32***	-0.28***	-0.25***
	(0.05)	(0.05)	(0.06)	(0.07)
Half × Points Received $_{t-1}$	0.28***	0.29***	0.18	0.17
	(0.09)	(0.09)	(0.13)	(0.13)
Single × Points Received $_{t-1}$	-0.19**	-0.20**	0.44***	0.38***
	(0.09)	(0.09)	(0.10)	(0.10)
Half × Average Contribution $_{t-1}$		0.06		-0.03
		(0.06)		(0.05)
Single × Average Contribution $_{t-1}$		-0.16**		-0.20***
		(0.07)		(0.06)
Intercept	0.53	0.26	-3.29***	-4.23***
	(0.57)	(0.75)	(0.49)	(0.66)
Free Riders or Contributors:	Free Riders		Contributors	
Wald χ^2	170***	181***	92***	102***
Obs., individuals	2088, 441	2088, 441	2637, 486	2637, 486

Notes: (i) Free Riders contribute less than the group average and Contributors give more then the average, (ii) The regression is a Tobit with both upper and lower censoring and individual random effects, (iii) standard errors in parentheses, (iv) *** indicates significance at the 99% level, ** at the 95% level and * at the 90% level.

Table 4: