Can Endogenously Chosen Institutions Mitigate the Free-Rider Problem and Reduce Perverse Punishment?*

by Arhan Ertan,† Talbot Page‡ and Louis Putterman‡
Brown University

Abstract

Previous experiments on public goods dilemmas have found that the opportunity to punish leads to higher contributions and reduces the free rider problem; however, a substantial amount of punishment is targeted on high contributors. In the experiment reported here, subjects are given the opportunity to vote on rules governing punishment. We found that, from their first opportunity to vote, no group ever allowed punishment of high contributors, most groups eventually voted to allow punishment of low contributors, and a minority of groups never allowed any form of punishment. Groups allowing punishment of low but not high contributors had significantly higher efficiency and contributions than comparison groups with unrestricted punishment.

JEL Classification: C91, C92, D71, H41
Keywords: Public goods, collective action, punishment, voting, institutions.

* The research reported here was supported by National Science Foundation grant SES-0001769. Please send comments to Arhan_Ertan@Brown.Edu, Talbot_Page@Brown.Edu, Louis_Putterman@Brown.Edu.
† Ph.D. candidate, Department of Economics, Brown University.
‡ Professor of Economics, Brown University.

Can Endogenously Chosen Institutions Mitigate the Free-Rider Problem and Reduce Perverse Punishment?*

by Arhan Ertan, Talbot Page and Louis Putterman
Brown University

In an experimental study of a social dilemma Fehr and Gächter found (2000) that experimental subjects were willing to punish free riders. The punishment, mainly targeted at free riders, led to less free-riding and higher contributions. In their experimental design, the (iterated dominant strategy) equilibrium for self-interested players predicts no contributions and no punishment, but the results are consistent with predictions of models with heterogeneous preference types including other-regarding cooperators as well as self-interested payoff maximizers.[1]

While Fehr and Gächter found punishment was largely directed at those making less than average contributions they also found some directed at higher-than-average contributors (pp. 990). In a related experiment on a social dilemma (where there were incentives to overuse a common pool resource) Ostrom, Gardiner, and Walker (1992) found most punishment directed at the over-users, but some directed at the cooperative users. In another social dilemma, Saijo and Nakamura (1995) suggested that some subjects appeared to be motivated by a desire to increase their relative earnings by lowering others' earnings more than their own. Ostrom *et al*. and Fehr and Gächter suggested other possible reasons for punishing cooperators, including mistakes and preferences to harm others suspected of harming oneself in "blind revenge."[2] In reviewing our own and others' experiments on voluntary contribution mechanisms (VCM) with punishment, we found about 20% or more of punishments were targeted at higher-than-average contributors and sometimes at those who contributed their whole endowment.

* The research reported here was supported by National Science Foundation grant SES-0001769. Please send comments to Arhan_Ertan@Brown.Edu, Talbot_Page@Brown.Edu, Louis_Putterman@Brown.Edu.

[1] For example in the 2-person prisoners' dilemma of Kreps, Milgrom, Roberts, and Wilson (1982) and the centipede game of McKelvey and Palfrey (1992). See Camerer and Fehr (2004) for a summary of evidence concerning heterogeneous preference types.

[2] Typically in experiments on social dilemmas with a punishment option, subjects are not informed of who punished whom, so someone seeking revenge for being punished would do so blindly.

To study the effects of punishment of high contributors and to investigate how rules governing punishment might evolve endogenously, we designed an experiment where the subjects could vote on rules governing punishment and then participate in a voluntary contributions mechanism with punishment governed by the chosen rules. In its design the (iterated dominant strategy) equilibrium predicts no contributions and no punishment for self-interested players, and any vote on rules of punishment is a (weakly) dominant strategy.[3] It seemed likely, however, that groups would at least sometimes vote to prohibit punishment of high contributors while allowing punishment of low contributors.

More specifically, we asked: (1) If experimental subjects are given the opportunity to vote on rules governing punishment, how often would they vote to prohibit punishment of high contributors? (2) If some groups allowed punishment of low contributors but not of high contributors, would this lead to increased contributions and efficiency, compared with other rules governing punishment? (3) And if voters voted to allow punishment of low but not high contributors, would high contributors earn more than low contributors, thus reversing incentives to free ride?

The results were stronger than we had anticipated. Out of 160 group votes, even when groups had no prior experience with unrestricted punishment, no group ever voted to allow unrestricted punishment or punishment of high contributors. In contrast, there was a gradual but clear evolution toward allowing punishment of low contributors, while still prohibiting punishment of high contributors. At the individual level, higher contributors were more likely than others to vote to allow punishment of low contributors, and lower contributors were more likely than others to vote to allow punishment of high contributors.

We found that contributions were highest in groups that voted to allow punishment of low contributors while prohibiting punishment of high contributors, next highest in groups under the (exogenous) rule allowing unrestricted punishment, and lowest under the rule prohibiting all punishment. In contrast, the ordering for efficiency was highest in groups that voted to allow punishment of low contributors while prohibiting punishment of high contributors, next highest in groups under the rule prohibiting all punishment, and lowest

[3] Because of the dominant strategy of no punishment, the outcome of the vote does not matter for self-interested subjects. There was no monetary cost of voting.

under the (exogenous) rule allowing unrestricted punishment (this same ordering of rules occurred in the frequency of voting outcomes).

Besides the papers already cited, our experiment is related to others as follows. In Falkinger, Fehr, Gächter, and Winter-Ebmer (2000), the Falkinger mechanism taxes lower-than-average contributors and subsidizes higher-than-average contributors, and thus rules out punishment of higher-than-average contributors while punishing lower-than-average contributors. The Falkinger *et al.* experiment differs from ours in that their tax and subsidy incentives are imposed exogenously and convert the game from a free-rider one to one where the Nash equilibrium for purely self-interested subjects is one with the efficient level of contributions from each player, contrasting with the inefficiently zero contributions of the Nash equilibrium of our experiment for purely self-interested subjects.

Sutter, Haigner, and Kocher (2005) compared the effects of endogenously chosen rules to punish or reward. One set of rules are those of the basic VCM without punishment; a second set is the VCM with punishment (unrestricted as to target); and the third is the VCM with rewards. The rules are chosen by repeated voting until unanimity is achieved. Our experiment differs from theirs by using majority rule to vote on separate ballot items that distinguish the treatment of high and low contributors by the endogenously chosen rules.

Other studies of VCMs with punishment, Carpenter and Matthews (2002), Sefton, Shupp and Walker (2002), Page, Putterman and Unel (2005) and Bochet, Page and Putterman (forthcoming), in addition to Fehr and Gächter (2000), found the opportunity to punish without restriction had a dramatic effect on contributions but little overall effect on efficiency. These studies suggested to us that along with the direct costs of punishment to the punisher and the punished (found in everyday life and included in these experiments) the incentive costs of punishing high contributors might also play an important role.[4]

We estimated incentive costs of punishing high contributors in a voluntary contributions mechanism (VCM) with unrestricted punishment, finding that each dollar of punishment of a group's highest contributor substantially decreased his or her next period contribution (while each dollar of punishment of a group's lowest contributor substantially increased his or her next period contribution). This result supported our conjecture that a rule

[4] Ways to reduce the efficiency costs of punishment are explored by Masclet, Noussair, Tucker and Villeval, 2003, who allow only the nonmonetary punishment of reprimands, and by Casari and Plott, 2003, who make fines transferable to the "bounty-hunters" who identify norm-violators.

allowing punishment of low but not high contributors might significantly increase overall efficiency.[5]

In the larger picture, researchers have studied interactions of self-interested preferences in combination with other preferences (such as ones for conditional cooperation and altruism) through controlled play of ultimatum, dictator, trust, centipede, VCMs, and other games. These studies (for example, McKelvey and Palfrey, 1992, Andreoni and Miller, 1993, Palfrey and Prisbrey, 1997) have provided at least rough estimates of frequencies of heterogeneous preference types, under varying experimental conditions and with replication. In the final section of this paper we briefly discuss how this research on the "demography" of heterogeneous preference types offers a new perspective on voting models, compared with earlier models assuming only self-interested preference types.

The paper is organized as follows. Section 1 explains the experimental design, Section 2 summarizes the predictions, Section 3 presents the analysis and results, and Section 4 provides a concluding discussion.

1. Design

1.1 Basic Design

Our design extends the basic voluntary contributions mechanism (VCM) in which subjects are randomly assigned to groups that remain fixed (a "partners" design) for a finite and known number of periods. Each subject in a group is provided with an initial endowment that he or she is asked to divide between a private account and a group account. Any funds placed in the group account are scaled up by the experimenter and divided equally among the subjects in the group without regard to individual contribution. This design generates the familiar result (for self-interested subjects) that it is socially optimal to contribute everything to the group account, but privately optimal to contribute nothing.

To this basic VCM we added punishment and voting opportunities, using two designs to study how rules restricting or allowing punishment might evolve over a series of votes.

[5] Cinyabuguma, Page and Putterman (2004). To check an alternative hypothesis that high and low contributors adjust their contributions toward their group average, even without punishment, the authors have also estimated regressions including stand-alone dummy variables for the rank of the contribution level. Results for the coefficients of interest are unaffected.

Our "3-Vote" design allowed for a gradual process of learning and familiarization, see Figure 1A. In the first instructions, given at the beginning of the experiment, subjects were read and shown on the computer screen instructions describing the basic VCM mechanism (without mention of punishment or voting to come later). Then they participated in the VCM for three periods without punishment or voting.

At the beginning of the 4th period, the subjects received their second instructions. These explained the opportunity of voluntary punishment, which was unrestricted except for some budgetary constraints (these instructions did not mention the voting to come later). Then the subjects participated in three periods of the VCM with unrestricted punishment.

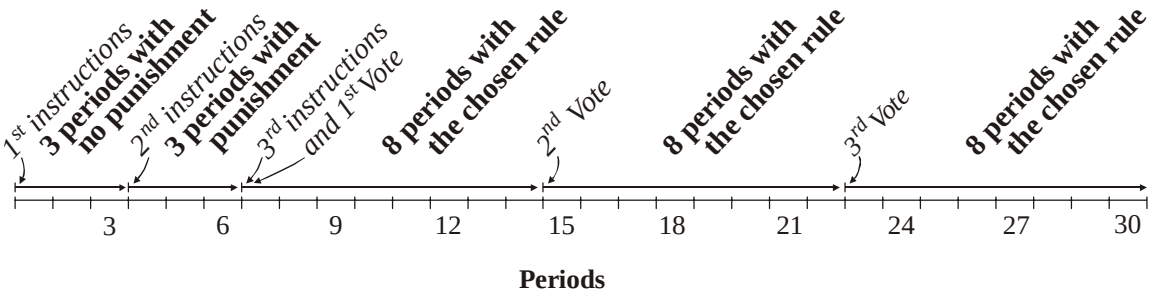


FIGURE 1A. THE 3-VOTE DESIGN

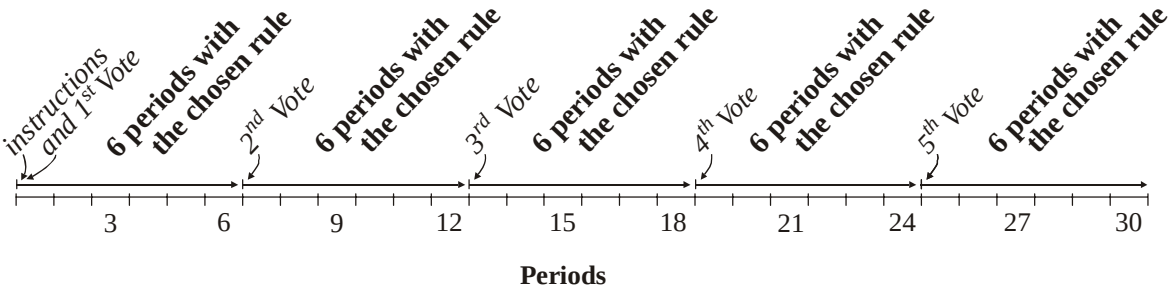


FIGURE 1B. THE 5-VOTE DESIGN

At the beginning of the 7th period, the subjects received their third instructions, which explained the voting process, and took their first vote on the rules governing who, if anyone, could be punished for the next eight periods. At the beginning of the 15th period a second vote was taken and new rules regulating punishment were chosen. Then the subjects participated in eight periods of the VCM with punishment (if any) governed by the second chosen rules. At the beginning of the 23rd period the third and final vote was taken, and the

remaining eight periods were conducted with possible punishment governed by this last vote. We included practice exercises in each of the three sets of instructions.

Surprised to find that out of 60 group votes none allowed punishment of high contributors, we asked if this uniform pattern of voting would hold up if experimental subjects had no prior experience with unrestricted punishment and more opportunities to vote. The question led to an additional, "5-Vote" design, Figure 1B. In this design the task of learning and familiarization was harder. The first and only instructions, given at the beginning of the experiment, explained the basic VCM mechanism without punishment, possible rules governing punishment, and the opportunity to vote on them. Again, the instructions included practice exercises.

After the instructions at the beginning of the 5-Vote design, the subjects voted to allow or restrict punishment. Then they participated for six periods in the VCM, governed by the chosen rules of punishment. At the beginning of the 7^{th} period, the subjects voted again, and then participated in 6 periods of the VCM, governed by the chosen rules of punishment. The same process repeated for three more times, as shown in Figure 1B. The 5-Vote design had the same number of periods (30) as the 3-Vote design and took about the same amount of time.

In both designs, sessions had sixteen subjects assigned randomly to four groups of four subjects who remained together throughout the session. Each subject knew there were 16 subjects in the experiment room but could not tell which among the others in the session belonged to her group. Contribution and punishment choices (if any) were announced to other group members under randomly changing labels B, C, and D, for one's fellow members, so that the behaviors of individuals could not be tracked from period to period, except by conjecture. A subject learned the total amount of punishment she had received, but not which group members punished her or by how much.

1.2 Payoffs

All periods shared the same underlying structure. In each period, each subject had to decide on a division of 10 experimental dollars, in integer amounts, between a private account and a group account, before observing the choices of fellow group members. In a period, subject *i* earned

(1) $$y_i = (10 - C_i) + (0.4)\sum_{j=1}^{4} C_j$$

where C_i is *i*'s contribution to the public account and the summation is taken over all members of *i*'s group, including *i*. After all four made their decisions, each was informed of the choices of the others. When punishment was permitted, it cost a subject 0.25 experimental dollars to reduce the earnings of another person by 1.00 experimental dollar. Subject *i*'s earnings after punishment were thus

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

where R_{ij} is the number of dollars by which *i* reduced *j*'s earnings, and conversely for R_{ji}. General constraints on punishment in all treatments were: (*i*) a subject could not spend more than her/his pre-punishment earnings for the period on reducing the earnings of other subjects, (*ii*) a subject's post-punishment earnings for a period would be set to zero if earnings y_i in equation (2) were negative, and (*iii*) a subject *i* could not spend more on reducing the earnings of a subject *j* in any period than would single-handedly reduce *j*'s earnings according in (2) to less than zero.[6] The Appendix shows the screen design for entering an individual's contribution and punishment decisions.

1.3 Voting

In a voting stage, each subject checked off one of three boxes beside each of three ballot items, on a screen set up as follows:

I vote to allow a person's earnings to be reduced if

(a) that person assigns less than the average amount[7] Yes No No preference

[6] The purpose of (*i*) and (*ii*) was to keep all decisions financially independent of each other while maintaining a guaranteed minimum payment for recruiting reasons. The purpose of (*iii*) was to help subjects to avoid pointless spending on punishment in view of constraint (*ii*). Note, however, that it remained possible for subjects to "overspend" on punishing in the sense that both subject *i* and, say, subject *k* might each spend enough to reduce *j*'s earnings for the period to zero, although only one subject's punishment would actually be effective in that case, given (*ii*). This could happen because subjects did not learn of punishment not carried out by or aimed at them, and the design (as in Fehr and Gächter, 2000) keeps such information private so as not to encourage free riding on punishment.

[7] As explained in the instructions, "average amount" meant the average over the four members of the group in the contribution stage of the period in question. It could vary among groups and within a given group from one period to the next. Note that a vote to allow punishment of those contributing less than the group average of 4 players is the same event as a vote to allow punishment of those contributing less than the average of the 3 others.

to the group account ☐ ☐ ☐

(b) that person assigns the average amount to the group account — Yes ☐ No ☐ No preference ☐

(c) that person assigns more than the average amount to the group account — Yes ☐ No ☐ No preference ☐

In each group of four subjects, of those expressing a preference in ballot item (a), if there was a majority or tie of "No" votes against punishment of low contributors, then punishment of low contributors would be prohibited for the next 8 periods in the 3-Vote design and 6 periods in the 5-Vote design; and if a majority voted "Yes," punishment of low contributors would be allowed; and correspondingly for ballot items (b) and (c).[8] After the vote, each group's members received a message reporting the voting outcome, which was one of $2^3 = 8$ possible punishment rules (*i.e.*, combinations of the three ballot item choices).[9]

When a group voted to restrict punishment, a fixed zero appeared in the punishment box[10] for all individuals to which the restriction applied during the punishment stages that followed each contribution stage. For example, members of a group that had voted to prohibit all punishment saw the standard punishment stage screen with fixed 0's in all the punishment boxes, indicating that no punishment was allowed in this case. Just before the second and later votes, each subject was informed of the punishment rule chosen in the preceding votes of each of the four groups in a session along with each group's average contributions and earnings during the periods the rule governed (the information was new for the most recently taken vote, and was repeated for the earlier votes).

We conducted five sessions of each design using a total of 160 subjects (see Table 1).[11] All of the sessions were conducted by computer in a computer lab at Brown University.

[8] We expected few cases where someone was exactly an average contributor, but for symmetry we treated the average contributor on a separate ballot item.

[9] Only "Yes" and "No" votes were counted in determining majorities; for example, if 2 voted "Yes" and 2 voted "No," the proposal did not pass, but if 2 voted "Yes," 1 "No" and 1 "No preference," the proposal passed. Subjects were informed of whether a ballot item passed or not, but not by how many votes or who voted which way.

[10] See the boxes labeled b′, c′ and d′ on the lower left portion of the diagram in the Appendix showing the screen design.

[11] Subjects were Brown undergraduates, recruited by (a) distribution of flyers in the mailboxes of all undergraduates, (b) distribution of flyers in a large introductory economics course, (c) distribution of table slips at a student dining hall, and (d) advertising under the heading of employment in an on-line campus magazine, the *Brown Daily Jolt*. Analysis of information provided in the post-experiment debriefing shows that the

At the end of each session, cumulative earnings for the thirty periods were totaled and converted to real money at the rate of 25 experimental dollars to one real dollar, and $5 was added as a participation fee. Sessions typically lasted a little less than two hours including instructions, and subjects' overall earnings averaged approximately $25. Instructions for both designs are similar and available in our Working Paper.

TABLE 1. NUMBERS OF GROUPS, SUBJECTS, AND VOTES

Session design	Number of sessions	Number of groups in each session	Number of subjects in each group	Total number of subjects	Total number of group votes on rules
3-Vote	5	4	4	80	60
5-Vote	5	4	4	80	100

2. Predictions

Since punishment of higher-than-average contributors seems against the interest of both selfish payoff maximizers and cooperators, we conjectured that at least some groups would vote against the third ballot item, allowing "punishment of higher-than-average contributors." And because such punishment seemed likely to reduce efficiency, we refer to it in what follows as "perverse punishment" for short.[12] But due to the apparent scattershot pattern of unrestricted punishment, with substantial amounts of punishment targeted at high contributors, and with unknown frequencies of self-interested, cooperator, and other-regarding preference types, we made no predictions on how often perverse punishment would be prohibited, the efficiency effects of voting, or how much mitigation of free-rider incentives might result from allowing an endogenous choice of institutional rules.

3. Results

3.1 The Voting Pattern

subjects majored in a large range of concentrations, with the economics concentration being that of only 15%, about 5% more than the proportion in the overall student body. A little less than half the subjects had taken no economics courses at the college level. 67% of the subjects were female, somewhat higher than the 53% share in the general student body. Brown's undergraduate population numbers about 5500, so students participating in a given session tended not to know one another.

[12] In the instructions and experiment we used neutral language and did not use words like "free riding," "punishment," and "perverse punishment."

In the 3-Vote design there were 720 individual votes (80 subjects each voting 3 times on 3 ballot items), and in the 5-Vote design 1200 individual votes. Table 2 shows the number of individual votes on each ballot item. The table shows a substantial number of individuals voted to allow punishment of higher-than-average contributors, but many more voting to allow punishment of less-than-average contributors.

TABLE 2. NUMBERS OF INDIVIDUAL VOTES TO ALLOW PUNISHMENT OF HIGH, AVERAGE, AND LOW CONTRIBUTORS, BOTH DESIGNS

	Yes	No	No Preference
Allow punishment of less than average contributors	410	211	19
Allow punishment of average contributors	46	577	17
Allow punishment of above average contributors	111	493	36

In the 3-Vote design there were 60 group votes (see Table 1), and in the 5-Vote design there were 100 group votes. In the 160 group votes altogether, only 4 of the 8 possible combinations of rules were ever chosen by majority rule. These were to allow: (*i*) no punishment, 56 group votes; (*ii*) punishment of lower-than-average contributors and no other punishment, 98 votes; (*iii*) punishment of low-or-equal-to-average contributors and no other punishment, 4 votes; and (*iv*) punishment of equal-to-average contributors and no other punishment, 2 votes. Conspicuously absent from this list is that no group ever voted to allow punishment of higher-than-average contributors.

RESULT 1. *No group ever voted to allow punishment of higher-than-average contributors, so perverse punishment was ruled out from the first opportunity to vote*.

In ruling out perverse punishment, every group also ruled out unrestricted punishment from the beginning. The two rules “punishment of lower-than-average contributors and no other punishment” and “punishment of low-or-equal-to-average contributors and no other punishment” are similar and we grouped them together under the heading of allowing punishment of “low-but-not-high” contributors. Figure 2 shows how the group voting

evolved, over the sequence of votes for the 3-Vote design and 5-Vote designs. Result 2 summarizes the voting pattern over time.

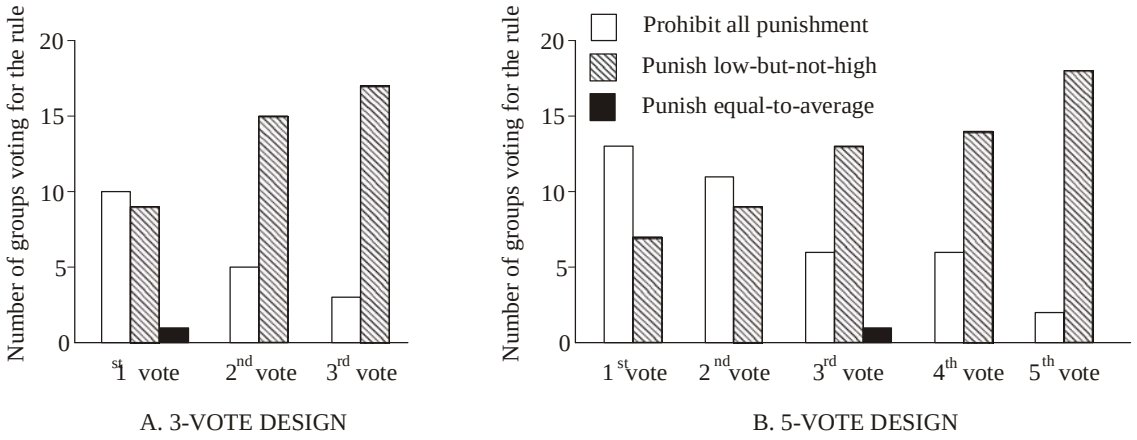


FIGURE 2. EVOLUTION OF THE VOTING RULES

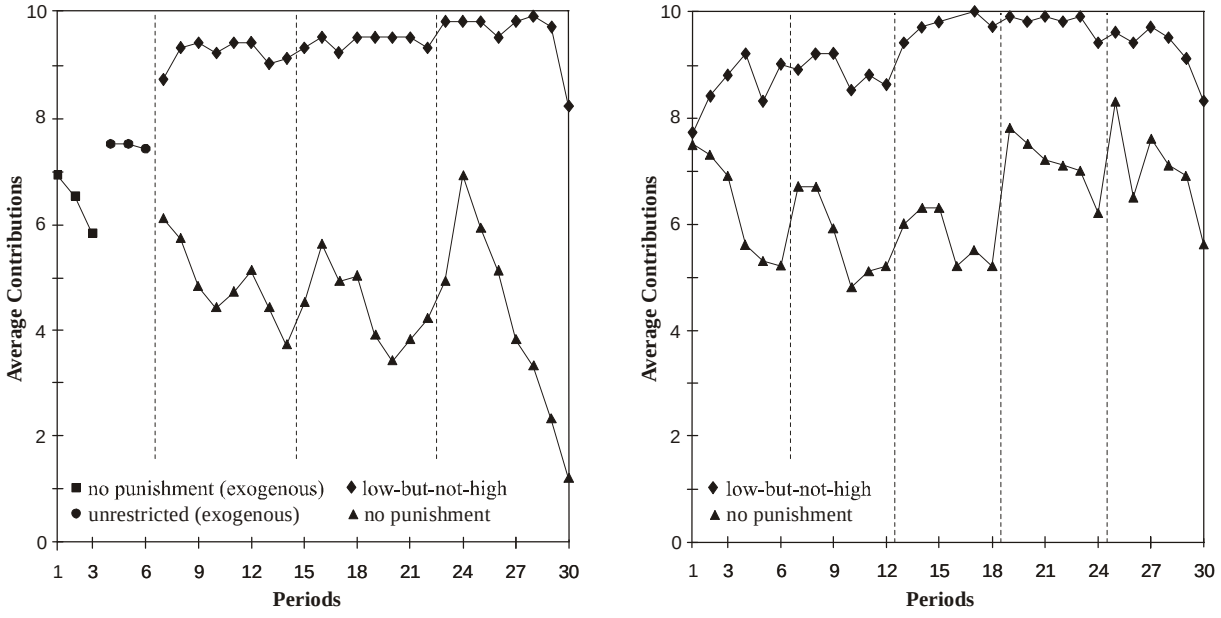
RESULT 2. *In both designs, a plurality of groups voted in their first vote to prohibit all punishment, with a substantial minority of groups voting to allow punishment of low-but-not-high contributors. Over the sequence of votes, this ordering reversed, so that in the final vote, nearly all groups voted to allow punishment of low-but-not-high contributors, with only a few remaining groups voting to prohibit all punishment.*

3.2 Contributions and Efficiency

Figure 3 shows period-by-period contributions of groups for the two composite rules most frequently chosen. In both the 3- and 5-Vote designs, groups that allowed punishment of low-but-not-high contributors achieved substantially higher levels of contributions than did groups that prohibited punishment altogether, significant ($p < 0.001$) in Mann-Whitney tests.[13] In both designs, contributions in groups that permitted punishment of low-but-not-high contributors tended to increase over time until the end-game fall off. In contrast, 3-Vote design

[13] The test compared all groups over the periods governed by voted rules. The observations used are averages of contributions in groups during the 8 periods that a given voted rule governs. For example, in the 3-Vote design if a particular group of 4 subjects voted to allow no punishment during periods 7-14, voted to allow punishment of low contributors in periods 15-22, and again voted to allow punishment of low contributors in periods 23-30, its average contribution of the first 8 periods would constitute one observation in the no-punishment category, those of the second 8 periods one observation in the punish-low-but-not-high category, and those of the last 8 periods another observation in that category.

groups that prohibited all punishment had falling levels of contributions over time, replicating earlier results on basic VCMs without punishment, and in the 5-Vote design contributions had a slightly increasing trend in the middle periods.[14]



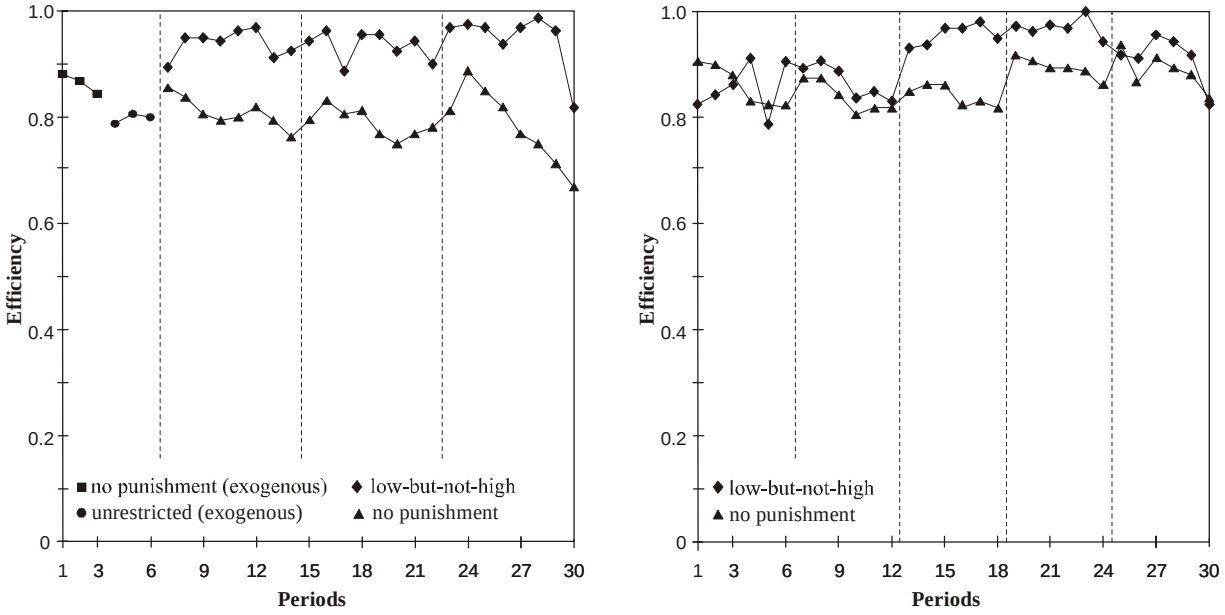
A. 3-VOTE DESIGN B. 5-VOTE DESIGN

FIGURE 3. AVERAGE CONTRIBUTIONS FOR THE TWO DESIGNS, BY PERIOD AND PUNISHMENT RULE

Figure 4 shows period-by-period efficiency[15] of groups that voted to prohibit all punishment and groups that voted to prohibit perverse punishment while allowing punishment of low contributors. In Figure 4A, average period efficiency was always higher under the rules allowing punishment of low contributors, and similarly in Figure 4B, except in six periods. In Mann-Whitney tests comparing all groups over the periods governed by voted rules, the groups which voted to allow punishment of low-but-not-high contributors had significantly higher efficiency than groups which voted for no punishment, at the 0.1% level for the 3-Vote design and 1% for the 5-Vote design.

[14] In Figure 3 contributions under the endogenously chosen rule of no punishment are more sustained and decline more slowly than is typical in a VCM without punishment. But endogenous choice includes its process, including repeated voting and the ability to change rules, possibly leading to commitment effects (see Sutter *et al.*), restart effects (see the dashed vertical lines in Figure 3), and selection effects as groups change rules in response to free-riding behavior.

[15] We follow the usual definition of efficiency in experiments, as the observed sum of earnings divided by the maximum possible sum of earnings, for a specified group or groups and periods.



A. 3-VOTE DESIGN B. 5-VOTE DESIGN

FIGURE 4. EFFICIENCY FOR THE TWO DESIGNS, BY PERIOD AND PUNISHMENT RULE

Table 3 compares contributions and efficiency under the two most voted rules, and the exogenously imposed conditions of unrestricted punishment (periods 4-6 of the 3-Vote design) and no punishment (periods 1-3).[16] The results of the five tests of Table 3 are summarized in Result 3:

RESULT 3. *For each of the Wilcoxon matched pair tests on contributions, contributions are higher under the rule of punish low-but-not high than under the rule of unrestricted punishment, and contributions are higher under the rule of unrestricted punishment than under the rule of no punishment, and this ordering is transitive. Correspondingly, efficiency is higher under punish low-but-not high than under unrestricted punishment, and efficiency is higher under no punishment that under unrestricted punishment, and this ordering is transitive.*

[16] For example, in comparing contributions under the rule of punish low-but-not-high with contributions under the (exogenous) rule of unrestricted punishment in Test 1, we considered the 17 groups of the 3-Vote design that eventually chose the rule allowing punishment of low-but-not-high contributors (see Figure 2A). For each of these groups we calculated the average group contribution over the first three periods that the group was governed by this rule. We matched this average with the same group's average contribution over the three periods of unrestricted punishment (periods 4-6 of the 3-Vote design). In the 17 matched pairs, 14 groups had higher contributions under the rule of punish low-but-not-high, 2 groups had higher contributions under unrestricted punishment, and 1 group was tied. The difference is significant ($p = 0.001$) in a two-tailed Wilcoxon matched pair test.

TABLE 3. EFFECTS OF THE PUNISHMENT RULE ON CONTRIBUTION AND EFFICIENCY

Test	Ranks of Contributions by the Punishment Rule	Test	Ranks of Efficiency by the Punishment Rule
1	punish low > unrestricted****	4	punish low > no punishment***
2	unrestricted > no punishment**	5	punish low > no punishment**
3	unrestricted > no punishment**	2	no punishment > unrestricted*
4	punish low > no punishment***	3	no punishment ~ unrestricted
5	punish low > no punishment****	1	punish low > unrestricted***

Notes: Wilcoxon Matched Pair Tests. Tests 1–4 are for groups in the 3-Vote design. **Test 1** compares the average contributions of the first three periods, if any, in which a group chose "punish low" matched with the average contributions of the same group in periods 4–6 of "unrestricted punishment" (the number of distinct groups matched and compared is $n = 17$); and correspondingly for efficiency. **Test 2** compares contributions (efficiency) for groups in periods 1–3 with contributions for the same groups in periods 4–6, $n = 20$. **Test 3** compares periods 4 – 6 with 7-9, for the groups that chose no punishment in periods 7–9, $n = 10$. **Test 4** compares the same groups before and after a switch from "no punishment" to "punish low," comparing the 8-period averages before and after the switch, $n = 9$. **Test 5** is the same as Test 4, except it is for the 5-Vote design and 6-period averages are compared before and after the switch, $n = 17$. **** indicates significance at the 0.1% level, *** at the 1% level, ** at the 5% level, * at the 10% level, and ~ insignificant, in two-tailed tests. "Punish low" indicates "punish low-but-not-high."

Because of the difference in the orderings for contributions and efficiency, the sequence or tests in Table 3 for efficiency are rearranged to show the transitivity. The difference in the orderings of contributions and efficiency is likely due to the cost of punishment.

3.3 Mitigating the Free-Rider Problem

The symmetric design of this and other VCM experiments suggests a simple definition of free riding: a subject A experiences free riding when someone else in his group, B, contributes less to the public good but earns more than A does.[17] For a specified punishment rule, sequence of periods, and collection of groups, we define the *frequency* of free riding as the number of cases of free riding divided by the number of observations, and an *observation* as a pairing in a group, where one subject in a group has a higher contribution than the other subject of the pair. By the design of a basic VCM without punishment and its payoff equation (1), every time someone contributes more than someone else, there is a case of free riding because the higher contributor always has lower earnings. Thus, in this

[17] Under this definition, if everyone in a group contributed the same low amount, there would be no free riding (it is only defined for unequal contributors).

definition of free riding, the frequency of free riding for the basic VCM is 100% (as shown in the first bar of Figure 5). But the frequency of free riding may decrease when sufficient punishment is directed at low contributors.

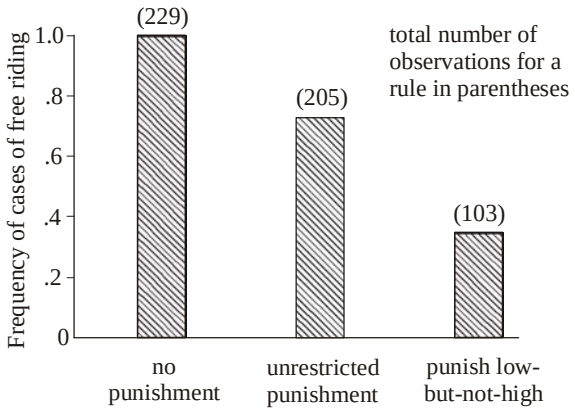


FIGURE 5. FREQUENCY OF CASES OF FREE RIDING, BY PUNISHMENT RULE

For the rule of unrestricted punishment, over all 20 groups in periods 4-6 of the 3-Vote Design, there were 205 observations of pairs of unequal contributions by subjects in a group, and 148 cases of free-riding, for a frequency of 72% (see the middle bar). In comparison, the frequency of free riding in the first three periods after a group voted for the rule of punishing low-but-not-high contributors was 35% of the 103 observed unequal pairs. This is a striking reduction, considering that the rule of punish low-but-not-high does not prevent a higher-than-average contributor from free riding on a still higher contributor. The difference in free riding between unrestricted punishment and punish low-but-not-high contributors is significant ($p < 0.0001$) in a Fisher exact test.[18]

RESULT 4. *In comparing VCMs with rules governing punishment, we find the highest frequency of free-riding in groups operating with no punishment, less free-riding in groups*

[18] We also did a Wilcoxon matched pair test to compare the frequency of free-riding in the first three periods (if any) after a group votes for "punish low-but-not-high" matched with the frequency of free-riding of the same group in periods 4–6 with the rule of "unrestricted punishment". The number of distinct groups matched and compared was $n = 12$ (since 3 groups never voted for "punish low-but-not-high" and in 5 groups all the subjects contributed the maximum possible amount) and the result was significant with $p = 0.005$ in a two-tailed test.

with unrestricted punishment, and least free riding in groups allowing punishment of low-but-not-high contributors.

A regression analysis of incentives to free ride finds the same ordering as in Result 4. In the regressions below, we follow Fehr and Gächter in defining subject *i*'s absolute negative and positive deviations from the average of others' contributions as:

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

where $\bar{C}_{-i} = \frac{\sum_{j \neq i} C_j}{3}$ is the average of others' contributions.

Using Fehr and Gächter's specification (see their Table 5, p. 991), we first consider behavior in the three periods of the exogenously imposed rule of unrestricted punishment (periods 4–6 of the 3-Vote design, see column (1) of Table 3), and compare this with the first three periods of the endogenously chosen rule allowing punishment of low-but-not-high in both the 3 and 5-Vote designs (columns (2) and (3)).[19] Then, secondly, we consider behavior for the punish low-but-not-high rule over all the periods which it governs in the 3- and 5-Vote designs (columns (4) and (5)).

In each regression of Table 4 the dependent variable is each subject *i*'s punishment received in each period (three periods for regressions (1), (2), and (3), and up to 24 and 30 periods in regressions (4) and (5) respectively). The independent variables are the Average Contribution of Others, *i*'s Absolute Negative Deviation, *i*'s Positive Deviation, and period and group dummies (not shown).[20,21]

[19] We include observations for only the first three periods under a rule in columns (2) and (3) to achieve comparability with the regression for periods 4-6 (column (1)), in view of the possibility that learning or other factors might change behaviors with more repetitions.

[20] In both the unrestricted (Column 1) and restricted (Columns 2-5) punishment regressions, only the observations of individuals who could potentially be punished are included. The difference is that under unrestricted punishment, anyone can be punished.

[21] The regressions were also estimated by the Tobit method, treating 0 punishment observations as potentially left-censored. Resulting coefficients are similar and similarly significant except in the case corresponding to Column (1), where they are not significant at conventional levels.

TABLE 4. DETERMINANTS OF PUNISHMENT RECEIVED

Dependent Variable: experimental dollars of punishment

Independent variables	First three periods of the rule			All periods of the rule	
	Unrestricted Punishment 3-Vote Design (1)	Punish Low-But-Not-High 3-Vote Design (2)	Punish Low-But-Not-High 5-Vote Design (3)	Punish Low-But-Not-High 3-Vote Design (4)	Punish Low-But-Not-High 5-Vote Design (5)
Constant	-0.74 (1.067) $p = 0.490$	4.086* (2.353) $p = 0.088$	19.754*** (4.587) $p < 0.001$	0.587 (2.222) $p = 0.792$	11.483*** (4.367) p = 0.010
Average Contribution by Others	0.388** (0.175) $p = 0.028$	-0.230 (0.244) $p = 0.350$	-1.090*** (0.405) $p = 0.009$	0.228 (0.206) $p = 0.269$	-0.654** (0.269) p = 0.016
Positive Deviation	0.377** (0.152) $p = 0.014$	n.a.	n.a.	n.a.	n.a.
Absolute Negative Deviation	0.888*** (0.221) $p < 0.001$	1.217*** (0.148) $p < 0.001$	1.039*** (0.122) $p < 0.001$	1.054*** (0.138) $p < 0.001$	0.967*** (0.095) $p < 0.001$
R-squared	0.54	0.91	0.86	0.75	0.78
Observations	240	82	92	241	176

Notes: Punishment received as a function of deviation from group average in unrestricted and restricted punishment conditions. OLS regressions with period and group fixed effects, not shown. Unrestricted punishment, in Column 1, is observed in periods 4-6, where each observation is for one subject and one period. Columns 2–5 include one observation per subject under the rule allowing punishment of low-but-not-high contributors. In Columns 2 and 3, only the first three periods in which a group adopted the rule for the first time are included, while Columns 4 and 5 include all periods of restricted punishment. Numbers in parentheses are White heteroskedasticity-consistent standard errors; *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

To compare how the rules of punishment affect the estimated incentives toward contributing $1 less, Table 5 focuses on coefficients of Absolute Negative Deviation and Positive Deviation from Table 4. For example, in Column (1) of Table 5 the coefficient for Absolute Negative Deviation is $0.89, the estimated punishment for a $1 reduction in contribution for a less-than-average contributor, under the rule of unrestricted punishment, in the first three periods of the 3-Vote design, and shown as a negative gain of $-0.89 in Column (1) of Table 5. In Column (2) of Table 5 the coefficient for Absolute Negative Deviation is $1.22, the estimated punishment for a $1 reduction in contribution for a less than average contributor, under the rule of punish low-but-not-high contributors, in the first three periods of the 3-Vote design, and shown as a negative gain of $-1.22 in Column (2) of Table 5, etc.

The $+0.60 throughout Table 5 is the $1 gain in the private account from reducing one's contribution by $1, minus the $0.40 loss in the individual's earnings from the group account. In Column (1) of Table 5 the coefficient for Positive Deviation is $0.38, the estimated punishment for each $1 of additional contribution for a higher-than-average contributor, under the rule of unrestricted punishment, in periods 4-6 of the 3-Vote design. The $+0.38 in Column (1) of Table 5 is the positive gain from contributing $1 less and avoiding $0.38 in perverse punishment, for a higher-than-average contributor. The cases labeled n.a. in Table 5 are for the rule of punish low-but-not-high in Columns (2)-(5), in which case punishment of higher-than-average contributors is not allowed.

TABLE 5. INCENTIVES TO CONTRIBUTE $1 LESS

	Less-than-average contributors, subject to punishment				
	(1) unrestricted punishment	(2) punish low-but-not-high	(3) punish low-but-not-high	(4) punish low-but-not-high	(5) punish low-but-not-high
Abs. neg. deviation	$-0.89	-1.22	-1.04	-1.05	-0.97
$1 account shift	+ 0.60	+ 0.60	+ 0.60	+ 0.60	+ 0.60
Net gain	$-0.29	-0.62	-0.44	-0.45	-0.37
Higher-than-average contributors, subject to punishment only in Column (1)					
	(1)	(2)	(3)	(4)	(5)
Positive deviation	+0.38	n.a.	n.a.	n.a.	n.a.
$1 account shift	+ 0.60	+ 0.60	+ 0.60	+ 0.60	+ 0.60
Net gain	+0.98	+0.60	+0.60	+0.60	+0.60

Note: Net gain is the change in earnings from contributing $1 less.

Table 5 shows that for less than average contributors the net gain from contributing $1 less is negative for each of the cases in Columns (1)-(5). The $-0.29 in Column (1) suggests that unrestricted punishment can reverse a subject's incentive to free ride, for a subject contributing less than average, replicating Fehr and Gächter's earlier finding for the case of less-than-average contributors. But the negative gains for less than average contributors is even more negative in Columns (2)-(5), suggesting that the incentive against free-riding is strengthened for less than average contributors under the rule of punish low-but-not-high.

Table 5 suggests that the incentives to contribute $1 less for higher-than-average contributors is *not* reversed under unrestricted punishment or the rule of punish low-but-not-high. In Column (1) under unrestricted punishment, a subject with a higher-than-average contribution makes an estimated net gain of $0.98 from contributing $1 less (a gain of $0.38 from reduced perverse punishment added to the $0.60 gain from shifting away from the group account). In Columns (2)-(5), under the rule of punish low-but-not-high, a higher-than-average contributor bears no punishment, but still gains the $0.60 from a $1 shift from the public account. While neither rule reverses the incentive for a higher-than-average contributor to contribute less, the incentive toward free riding is less under the rule of punish low-but-not-high than under unrestricted punishment.

In other words, when it is allowed, perverse punishment appears to exacerbate the incentive problem for high contributors. In fact there was considerable perverse punishment in periods 4–6 of the 3-Vote design. Of the 129 events of punishment, 28% were punishments aimed at higher-than-average contributors for the period and group in question, 19% at the highest contributor for the period and group in question and 11% at individuals who contributed their full endowment.[22]

3.4 Do Subjects Vote "According to their Type"?

We conjectured that even though some subjects use opportunities to perversely punish (when punishment is unrestricted) and would likely vote to allow perverse punishment in our experiment, punishment of high contributors might nonetheless be ruled out since few groups would have a majority of members of this type. Results at group level are consistent with this conjecture. Is there also evidence at the level of individuals, however, that subjects tended to vote "according to type"? Logit regressions provide some affirmative evidence.

We estimated regressions in which the dependent variable is 1 if subject *i* voted to permit punishment specified by a particular rule and 0 otherwise. We used logit rather than probit regressions because of the need to control for group and vote (period) effects. Our first

[22] These percentages are calculated by counting each event (rather than dollar amount) of someone punishing someone else. They may be atypically high due to the short duration of the unrestricted punishment portion of our experiment. Yet similarly large amounts of perverse punishment are found in some other studies; see for example Cinyabuguma, Page and Putterman (2004), Anderson and Putterman (forthcoming), Gächter and Hermann, and for a regression result similar to column (1), in which the absolute positive deviation term also has a positive significant coefficient, Ones and Putterman (forthcoming), Table 2.

explanatory variable and the main focus of the analysis is Relative Contribution, which is the difference between subject *i*'s contribution and the average contribution of the other three members of *i*'s group, calculated for each of the eight (six) periods prior to the vote and then averaged over those periods. This is a larger (smaller) number if *i* tended to be a high (low) contributor. We conjectured that higher contributors were more likely to include subjects favorably disposed towards cooperation, and that such individuals would therefore also be more likely to vote to allow punishment of low contributors (alternatively, high contributors, feeling resentment at free riders, would be more likely to vote to allow subjects to punish them). Similarly, because it is lower contributors who are found most likely to punish high contributors and resist cooperation, we thought that any votes to allow punishment of high contributors would more likely come from lower contributors.

The other explanatory variables used in the regressions control for the possible effects of experience with punishment and for possible clues to subjects' types observable in their own decisions to punish. Punishment Received is the average number of experimental dollars worth of punishment that *i* received per period in the previous eight (six) periods. It includes only punishment when a below-average contributor, and thus excludes the perverse punishment of periods 4 – 6 of the 3-Vote design. Punishment Given is the average number of experimental dollars worth of punishment that i gave to others per period in the previous eight (six) periods, again excluding any perverse punishment given in periods 4 – 6 of the 3-Vote design. Because past behaviors are used to predict the current vote, the first vote in the 5-Vote design is not included in the regressions for that design. We include the first vote of the 3-Vote design but because no punishment was possible in periods 1 – 3, the explanatory variables are averages over periods 4 – 6 only.

The resulting regressions are shown in Table 6. As conjectured, the coefficients on the subjects' relative contribution were positive in the regressions on voting to allow punishment of low contributors, significant at the 5% level or better for both the 3- and the 5-Vote design, and negative in the regressions on voting to allow punishment of high contributors, significant at the 10% level in the regression for the 3-Vote, but not significant in that for the 5-Vote design. Only one out of eight coefficients on the punishment variables is significant, but they are included because of their potential relevance and due to the improvement of fit when they are included. The one significant coefficient, which is for

Punishment Received in the regression for voting to allow punishment of low contributors in the 5-Vote design, may be explained in part by the desire of those who received punishment to retain the right to punish others, in part by the experience of punishment helping to educate a subject to its use as a tool to benefit the group, but it is best not to speculate given that the result is not very robust.[23]

Table 6. Individual votes on allowing punishment as a function of relative contribution and punishment given and received.

	3-Vote Design		5-Vote Design	
	Punish Low	**Punish High**	**Punish Low**	**Punish High**
Relative Contribution	0.4203 (0.1650) p = 0.011	-0.6791 (0.3925) p = 0.084	1.1479 (0.3823) p = 0.003	-0.1344 (0.9954) p = 0.893
Punishment Received	-0.0227 (0.0487) p = 0.640	0.0418 (0.0749) p = 0.576	0.5651 (0.2593) p = 0.029	-0.2908 (0.6524) p = 0.656
Punishment Given	0.0207 (0.0563) p = 0.714	0.0584 (0.0560) p = 0.297	0.1563 (0.3034) p = 0.606	-1.8468 (2.2798) p = 0.418
Vote-3 Dummy	0.5915 (0.5683) p = 0.298	-1.5401 (1.2067) p = 0.202	0.6816 (0.5017) p = 0.174	0.1067 (1.3963) p = 0.939
Vote-4 Dummy	0.8894 (0.4068) p = 0.029	-1.3007 1.0840 p = 0.230	1.2283 (0.6438) p = 0.056	-0.0471 (1.3003) p = 0.971
Vote-5 Dummy	NA	NA	1.8248 (0.8232) p = 0.027	-0.4075 (1.3230) p = 0.758
Pseudo R^2	0.1748	0.2992	0.1063	0.0820
# of Obs.	160	72	172	76

Note: logit regressions with group fixed effects (not shown) and robust standard errors. In the columns labeled "Punish Low (High)," the dependent variable is 1 when the subject voted to allow punishment of below- (above-) average contributors and zero otherwise.

RESULT 5: *Subjects were more (less) likely to vote to allow punishment of less- (greater-) than-average contributors the higher on average was their contribution above their group's average contribution in the eight (six) periods.*

[23] Apart from the 1st vote in the 3-Vote design, which follows periods in which punishment is made possible exogenously in all groups, only groups that permitted punishment during the previous eight (six) periods can be included in the regressions, given the two punishment variables. Additional observations are lost if all subjects vote identically in a group in every included period. In 12 groups in the 3-Vote design and for 10 groups in the 5-Vote design, every subject voted against allowing punishment of high contributors in every included vote, leading to the smaller number of observations for the regressions about allowing such punishment.

4. Discussion and Conclusion

The main conclusion is that the endogenous choice of institutional rules of punishment can mitigate the free rider problem and reduce perverse punishment. We found that, given the opportunity to vote, no group voted to allow unrestricted punishment and no group allowed punishment of high contributors. The favored choice of a punishment rule, allowing punishment of low-but-not-high contributors, increased both contributions and efficiency, compared with unrestricted punishment and no punishment.

In a parallel way with Fehr and Gächter's and others' VCM experiments, our results are inconsistent with the predictions of the iterated dominance equilibria for purely self-interested players, but consistent with predictions for players with heterogeneous preference types, including both conditional cooperators and active resistors of cooperation (e.g. perverse punishers). Analysis of earlier experiments and of the periods of unrestricted punishment in our 3-Vote design suggests that 20% or more of punishment is targeted at higher-than-average contributors.

In our experiment we find that not only is about 20% of punishment directed at high contributors but also roughly the same percentage of the individual votes are to allow such punishment (out of 640 individual votes on the ballot item to allow punishment on high contributors 111 votes, or 17%, were "Yes" votes). With a "demography" of preference types in which roughly one in five favor such a rule and with considerable punishment of high contributors occurring when it is permitted, fully decentralized decisions about punishment leads to "scattershot" behavior reducing punishment's efficiency. Majority votes on when punishment is allowed suppressed this behavior, because the proposal to allow punishment of cooperators was voted down every time.

We suspect that perverse punishment – or more generally, resistance to those promoting group cooperation – may be a widespread phenomenon, not just limited to public goods experiments with unrestricted punishment. One student, who had severely punished someone who had contributed her whole endowment to the group account, offered in his debriefing a quite general reason for his action, explaining that when someone does something morally superior to you, you feel uncomfortable and want to get back at that person. The inclination to retaliate against those punishing one's free riding is also manifest

in recent experiments by Nikiforakis (2004) and Cinyabuguma, Page and Putterman (2004). Such retaliation might be done informally by poor-mouthing, ridicule, or ostracism.

In our experimental design, inefficient free riding is a dominant strategy for purely self-interested actors, leading to low efficiency. But under majority rule voting self-interested actors combined with cooperators to outvote a minority of perverse punishers, thus freeing cooperative types to punish free-riding without fear of retaliation, thereby ameliorating the free-rider problem and achieving higher efficiency. In the larger context of practical politics, Ordeshook (pp. 213-215) analyzed a model of majority rule voting for political pork (legislators voting for projects benefiting one's own district with the costs borne by the whole country). In his model of purely self-interested legislators it is a dominant strategy for every legislator to externalize the costs of his district's project, for cases when the costs of each project exceed the benefits, leading to low efficiency.

But when there is a combination of self-interested types combined with some conditional cooperators, the cooperators may decline to add a pet project to a pork bill and vote against it. Would ever there be enough cooperators to form a majority and vote down a pork barrel bill? Elsewhere, we found (Page, Putterman, and Unel, 2005) that cooperators could be a majority, under favorable conditions when they were allowed to group themselves together, but this majority did not manifest itself in the baseline condition when the experimental subjects were grouped randomly. In actual legislatures of course, legislators form coalitions and it is common for some legislators to vote against pork bills, occasionally voting down such bills.

As another example, Meltzer and Richard's (1981) model of the level of redistributive taxation uses a median voter solution assuming strictly self-regarding preferences. More accurate explanations of the level of redistribution and its variation over time and place would consider the strength of preferences for greater equality, on the parts of some citizens, and resentment of the "undeserving poor," on the parts of others (see, for instance, Benabou and Tirole, 2005). Such an addition of two almost opposite social preference types alongside self-interested types resembles the specific case studied in this paper, where self-interested subjects co-exist with both cooperation-preferring and cooperation-resisting types, with the associated demographic leading to predictable voting outcomes. Our study of endogenous institutional choice illustrates how the experimental

method can be applied to study majority rule voting when self-interested and other preference types interact.

References

Anderson, Christopher and Louis Putterman, forthcoming, "Do Non-strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism," *Games and Economic Behavior* (in press).

Andreoni, James and John H. Miller, 1993, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *Economic Journal* 103: 570-85.

Benabou, Roland and Jean Tirole, 2005 "Belief in a Just World and Redistributive Politics," C.E.P.R. Discussion Paper 4952.

Bochet, Olivier, Talbot Page and Louis Putterman, forthcoming, "Communication and Punishment in Voluntary Contribution Experiments," *Journal of Economic Behavior and Organization* (in press).

Camerer, Colin and Ernst Fehr, 2004, "Measuring social norms and preferences using experimental games: a guide for social scientists," Chapter 3, pp. 55-95 in *Foundations of human sociality: economic experiments and ethnographic evidence from fifteen small scale societies,* (ed. Joseph Henrich, et al.); Oxford; New York: Oxford University press.

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

Casari, Marco and Charles R. Plott, 2003, "Decentralized Management of Common Property Resources: Experiments with a Centuries-Old Institution," *Journal of Economic Behavior and Organization* 51 (2): 217-47.

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

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

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

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

Kreps, David, Paul Milgrom, John Roberts and Robert Wilson, 1982, "Rational Cooperation in Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory* 27: 245-52.

Masclet, David, Charles Noussair, Steven Tucker and Marie-Claire Villeval, 2003, "Monetary and Non-Monetary Punishment in the VCM," *American Economic Review* 93(1): 366-80.

McKelvey, Richard and Thomas Palfrey, 1992, "An Experimental Study of the Centipede Game," *Econometrica* 60(4): 803-36.

Meltzer, Allan and Scott Richard, 1981, "A Rational Theory of the Size of Government," *Journal of Political Economy* 89 (5): 914-27.

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

Ones, Umut and Louis Putterman, forthcoming, "The Ecology of Collective Action: A Public Goods and Sanctions Experiment with Controlled Group Formation," *Journal of Economic Behavior and Organization* (in press).

Ordeshook, Peter, 1986, *Game Theory and Political Theory: An Introduction*, Cambridge; New York: Cambridge University Press.

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

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

Palfrey, Thomas R. and Jeffrey E. Prisbrey, 1997, "Anomalous Behavior in Public Goods Experiments: How Much and Why?" American Economic Review 87 (5): 829-46.

Saijo, Tatsuyoshi and Hideki Nakamura, 1995, "The 'Spite' Dilemma in Voluntary Contribution Mechanism Experiments," *Journal of Conflict Resolution* 38 (3): 535-60.

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

Sutter, Matthias, Stefan Haigner, and Martin Kocher, 2005, "Choosing the stick or the carrot? – Endogenous institutional choice in social dilemma situations" (draft, April 24, 2005).

Appendix

A.1 Screen Design

Figure A is the screen design for an individual to enter her contribution to the group account (box a), to learning of others' contributions (boxes b, c, and d), to enter her punishment decisions (boxes b′, c′, and d′), and to observe the computer's calculation of net earnings for a round.

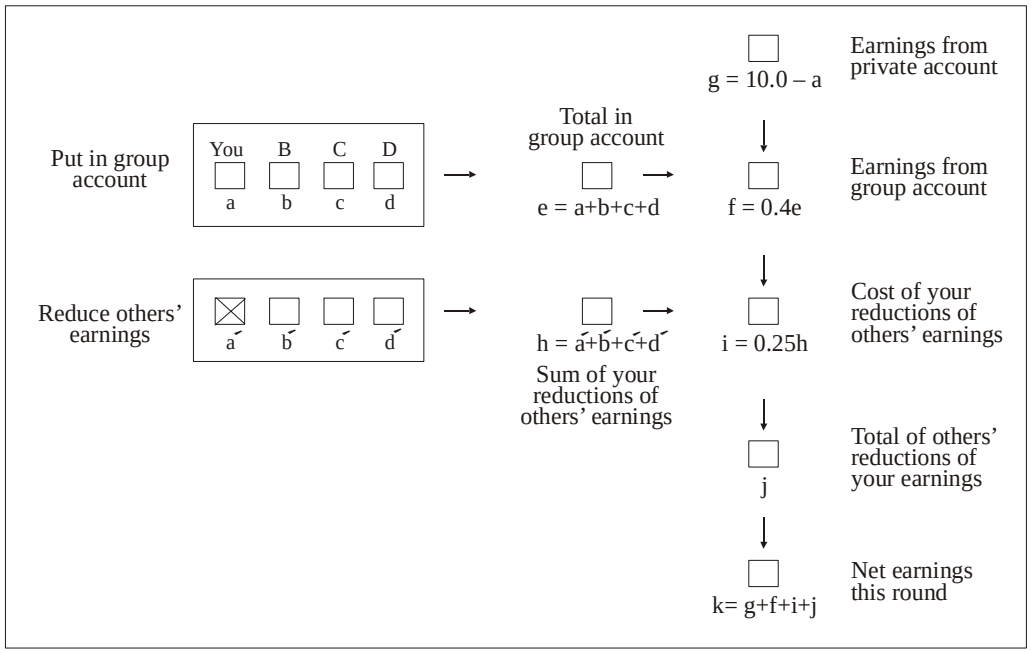


FIGURE A. SCREEN DESIGN FOR ENTERING CONTIBUTION AND PUNISHMENT DECISIONS, RECEIVING INFORMATION, AND CALCULATING NET EARNINGS

A.2 Instructions

A.2.1 Instructions for 3-Vote Experiments:

[Screen 1]

This is an experiment, funded by a research foundation, to study decision-making. You will be earning money in "experimental dollars" during the experiment. At the end of the experiment you will be paid in cash in real dollars (25 experimental dollars converts to 1 real dollar). The amount you will earn will depend on your and others' decisions. The maximum possible earning is $31.40 (real dollars) and the minimum possible is $5. You are likely to earn an amount in between. Please make sure you understand the decision process.

[Screen 2]

Your Group

The experiment consists of thirty distinct periods or rounds of decision-making, which will be played in five groups: two groups of 3 periods followed by three groups of 8 periods. All of these periods share a common structure.

At the beginning of the experiment, you will be randomly matched with three other participants, to form a group of four that will remain together throughout the experiment. The other three people who are in your group will be identified to you as "B," "C," and "D," although the letters will be shuffled from period to period, so that the person identified as "B" in one period is equally likely to be called "C" or "D" in the next one. You will not know the actual identities of the other members of your group either while making your decisions or after the experiment.

[Screen 3]

Your First Decision: Assigning Money to Group and Personal Accounts

At the beginning of every period each person in your group will receive $10 (experimental dollars). Each of you must decide how to divide this amount between a group account and a personal account.

The money you assign to your personal account goes into your earnings.

An amount equal to 0.4 times the group's total assignment to the group account goes into your earnings.

Your earnings = (amount in your personal account) + (0.4)(total in group account)

[Screen 4]

The next four screens illustrate how the experiment works. Fill in the blanks of your worksheet first, then enter the information in the practice decision screen. The numbers you type in the practice screens are for practice only and will net affect your earnings from the experiment.

Practice Questions

Practice 1.

The four members of your group each have $10. Every member of your group has assigned $10 to the group account and $0 to their personal account. Fill in the blanks on the right.

(1) Amount you assigned to group account . . . . . $_____
(2) Amount you assigned to your personal account . . . . . $_____
[= $10 – group account assignment on line (1)]
(3) Total number of dollars assigned to your group account . . . . . $_____
(4) Income from the group account for a member of your group . . . $_____
[0.4 • group account total in line (3)]
(5) Your earnings after the assignment decisions . . . . . $_____
[group account income in line (4) + personal account income in line (2)]

Now, go back to the practice screen. Type in your assignment to the group account, press enter, and check your calculation.

Practice 2.

The four members of your group each have $10. Every member of your group has assigned $0 to the group account and $10 to their personal account. Fill in the blanks on the right.

(1) Amount you assigned to group account . . . . . $_____
(2) Amount you assigned to your personal account . . . . . $_____
[= $10 – group account assignment on line (1)]
(3) Total number of dollars assigned to your group account . . . . . $_____
(4) Income from the group account for a member of your group . . . $_____
[0.4 • group account total in line (3)]
(5) Your earnings after the assignment decisions . . . . . $_____
[group account income in line (4) + personal account income in line (2)]

Type in your assignment to the group account, press enter, and check your calculation.

Practice 3.

Person B assigned $10 to the group account and $0 to his or her personal account,
person C assigned $5 to the group account and $5 to his or her personal account,
person D assigned $0 to the group account and $10 to his or her personal account, and
you assigned $5 to the group account and $5 to your personal account.

Fill in the blanks on the right.

(1) Amount you assigned to group account . . . . . $_____
(2) Amount you assigned to your personal account . . . . . $_____
[= $10 – group account assignment on line (1)]
(3) Total number of dollars assigned to your group account . . . . . $_____
(4) Income from the group account for a member of your group . . . $_____
[0.4 • group account total in line (3)]
(5) Your earnings after the assignment decisions . . . . . $_____
[group account income in line (4) + personal account income
in line (2)]

Type in your contribution, press enter, and check your calculation.

[Screen 5]

Consider what would happen in practice 3 if you increase your assignment to the group account by $1.

Your personal account would go down by $1, reducing your earnings by $1.

Your group account would go up by $1, increasing your earnings by $0.40, for a net reduction of $0.60

But each of the other people in your group would increase their earnings by $0.40, for a total increase of $1.20 for the others in your group.

[Screen 6]

The experiment will begin with three periods of play. Each period you begin with a new $10 to allocate, and each period's earnings are independent of the others.
After these three periods, there will be further instructions. After all thirty periods have been completed, your net earnings will be totaled and converted from experimental dollars to real dollars. Then $5 will be added for your participation. You will receive your earnings in cash before leaving the experiment.

During the experiment, there is to be no communication of any kind among participants, apart from the entering of decision numbers that will be transmitted to the other members of your group. It is important that you fully understand the decision process before we begin. Please raise your hand now if you have any questions.

BREAK FOR FIRST THREE PERIODS.
INSTRUCTIONS FOR SECOND SET OF PERIODS:

[Screen 7]

A New Decision

After this break for instructions, you and the same three members of your group will be interacting for another three periods. As with the three periods just completed, each of these periods begins with a decision on assigning ten dollars to a group account or to a personal account. This time, however, each period also includes a second stage of decision-making.

After you learn the others' assignments to the group account, you have a chance to reduce others' earnings, and others have a chance to reduce your earnings. Suppose, in the last example, you decide to:

reduce B's earnings by $2
reduce C's earnings by $3
reduce D's earnings by $4

The total amount of reductions you make on others' earnings is $9.

It costs you $0.25 for each $1 you reduce others' earnings. So your own earnings are reduced by (0.25)($9) = $2.25 in this example.

Now, suppose

B reduces your earnings by $2
C reduces your earnings by $1
D reduces your earnings by $0

The total reduction of your earnings by others is ($2 + $1 + $0) = $3. Your screen will tell you how much your earnings have been reduced, but not who has reduced your earnings by what amount.

Similarly none of the others will learn by how much you have reduced their earnings. They will only learn their total reductions by others in the group as a whole.

Please fill in the sheet labeled practice 4 and the corresponding practice decision screen.

Practice 4.

You assigned $5 to the group account and $5 to your personal account,
person B assigned $10 to the group account and $0 to his or her personal account,
person C assigned $5 to the group account and $5 to his or her personal account, and

person D assigned $0 to the group account and $10 to his or her personal account.

You reduce person B's earnings by $2,
person C's earnings by $3, and
person D's earnings by $4.

You receive a total of $3 in reductions from other members of your group.

(1) Amount you assigned to group account . . . . . $_____
(2) Amount you assigned to your personal account . . . . . $_____
[$10 – group account assignment on line (1)]
(3) Total number of dollars assigned to your group account . . . . . $_____
(4) Income from the group account for a member of you group . . . $_____
[(0.4) • group account total in line (3)]
(5) Your earnings after the assignment decisions . . . . . .$_____
[group account income in line (4) + personal account
income in line (2)]
(6) You reduced the earnings of others in your group by a total of . .$_____
(7) This cost you . . . . . .$_____
[(0.25) • the sum of your reductions from line (6)]
(8) Other members of your group reduced your earnings by . . . . . $_____
(9) Your total earnings for this period . . . . . $_____
[Your earnings after the assignment decisions on line (5) minus
your reduction cost on line (7) minus the amount by which your
earnings were reduced on line (8)]

Enter your reductions, press enter, and check your calculation.

[Screen 8]

Your Net Earnings

Your net earnings for a period will be:

Amount in personal account + (0.4)(total in group account) - (0.25)(total of your reductions of others) - total of reductions of your earnings made by others.

If this results in a negative number in any period, your earnings for that period will be set to zero.

[Screen 9]

As before, each period you begin with a new $10 and each period's earnings are independent of the others. As before, there is to be no communication while the experiment is in progress, apart from the entering of your decisions. After the next three periods are over, there will be a break for instructions concerning the remaining twenty-four periods of play.

Are there any questions?

BREAK FOR PLAY OF PERIODS 4 – 6.
INSTRUCTIONS FOR VOTING AND REMAINING 24 PERIODS:

[Screen 10]

In the remaining parts of the experiment, you will play for three sets of eight periods each in the same group of four subjects. Before this part begins, each group will decide, by voting, whether to permit subjects to reduce one another's earnings after learning of their assignments to the group account. It will be possible to allow reductions of subjects who assign more than the group's average to its group account, of subjects who assign exactly the group average, and/or of subjects who assign less than the group average. Once the decision has been made by your group, it will be in force for the next eight periods of the experiment.

[Screen 11]

Your voting screen will look like the one below.

I vote to allow a person's earnings to be reduced if

	Yes	No	No preference
(d) that person assigns less than the average amount to the group account	□	□	□
(e) that person assigns the average amount to the group account	□	□	□
(f) that person assigns more than the average amount to the group account	□	□	□

Please answer "Yes," "No," or "No preference" by clicking the box to the right of each of the three choices. If a majority of those expressing a preference vote yes, the reductions in question will be allowed; otherwise they will not.

[Screen 12]

Once the voting is done, you will be told what if any reductions will be allowed for your group. Then the next eight periods of play will begin. Note that the reduction boxes for any individuals whom your group has decided cannot have their earnings reduced will automatically appear with zeros inside. If no reductions are allowed in your group, you may still have to wait while other groups engage in their reduction stages.

Are there any questions?

- - - - -

END OF INSTRUCTIONS, FOLLOWED BY ACTUAL VOTING SCREEN AS ABOVE:

I vote to allow a person's earnings to be reduced if

(g) that person assigns less than the average amount to the group account	Yes ☐	No ☐	No preference ☐
(h) that person assigns the average amount to the group account	Yes ☐	No ☐	No preference ☐
(i) that person assigns more than the average amount to the group account	Yes ☐	No ☐	No preference ☐

Please check your choices before pressing the button to submit. Remember to vote on every item, and remember that you will vote only once. Press the submit button when ready.

Submit

MESSAGE TO THE SUBJECTS PRIOR TO COMMENCEMENT OF THE 7th PERIOD:

Possible messages:

Your group has voted not to allow group members to reduce one another's earnings.

Your group has voted to allow members to reduce the earnings of any other group member.

Your group has voted to allow members to reduce the earnings of any group member who assigns less than the average amount to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns the average amount or less to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns the average amount or more to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns less than the average amount, or more than the average amount, to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns the average amount to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns more than the average amount to the group account.

MESSAGE BEFORE 15th AND 23rd PERIODS:

You will now vote again on whether to allow reductions of the earnings of others in your group. As before, reductions of a given kind will be allowed only if there is a majority of Yes votes for it. After the vote, you will be informed of the outcome under which your group will operate for the next eight periods. Before the vote, you will see a screen listing the decisions made by each of the groups in the experiment today, the average amount the group's members contributed to their group account, and the average amount that they earned, during the periods covered by that vote.

A.2.2 Instructions for 5-Vote Experiments:

[Screen 1]

This is an experiment, funded by a research foundation, to study decision-making. You will be earning money in "experimental dollars" during the experiment. At the end of the experiment you will be paid in cash in real dollars (25 experimental dollars converts to 1 real dollar). The amount you will earn will depend on your and others' decisions. The maximum possible earning is $31.40 (real dollars) and the minimum possible is $5. You are likely to earn an amount in between. Please make sure you understand the decision process.

[Screen 2]

Your Group

The experiment consists of thirty distinct periods or rounds of decision-making. All of these periods share a common structure.

At the beginning of the experiment, you will be randomly matched with three other participants, to form a group of four that will remain together throughout the experiment. The other three people who are in your group will be identified to you as "B," "C," and "D," although the letters will be shuffled from period to period, so that the person identified as "B" in one period is equally likely to be called "C" or "D" in the next one. You will not know the actual identities of the other members of your group either while making your decisions or after the experiment.

[Screen 3]

Your First Decision: Assigning Money to Group and Personal Accounts

At the beginning of every period each person in your group will receive $10 (experimental dollars). Each of you must decide how to divide this amount between a group account and a personal account.

The money you assign to your personal account goes into your earnings.

An amount equal to 0.4 times the group's total assignment to the group account goes into your earnings.

Your earnings = (amount in your personal account) + (0.4)(total in group account)

[Screen 4]

The next four screens illustrate how the experiment works. Fill in the blanks of your worksheet first, then enter the information in the practice decision screen. The numbers you type in the practice screens are for practice only and will net affect your earnings from the experiment.

Practice Questions

Practice 1.

The four members of your group each have $10. Every member of your group has assigned $10 to the group account and $0 to their personal account. Fill in the blanks on the right.

(1) Amount you assigned to group account . . . . . $_____
(2) Amount you assigned to your personal account . . . . . $_____
[= $10 – group account assignment on line (1)]
(3) Total number of dollars assigned to your group account . . . . . $_____
(4) Income from the group account for a member of your group . . . $_____
[0.4 • group account total in line (3)]
(5) Your earnings after the assignment decisions . . . . . $_____
[group account income in line (4) + personal account income in line (2)]

Now, go back to the practice screen. Type in your assignment to the group account, press enter, and check your calculation.

Practice 2.

The four members of your group each have $10. Every member of your group has assigned $0 to the group account and $10 to their personal account. Fill in the blanks on the right.

(1) Amount you assigned to group account . . . . . $_____
(2) Amount you assigned to your personal account . . . . . $_____
[= $10 – group account assignment on line (1)]

(3) Total number of dollars assigned to your group account . . . . . $_____
(4) Income from the group account for a member of your group . . .$_____
[0.4 • group account total in line (3)]
(5) Your earnings after the assignment decisions . . . . . .$_____
[group account income in line (4) + personal account income
in line (2)]

Type in your assignment to the group account, press enter, and check your calculation.

Practice 3.

Person B assigned $10 to the group account and $0 to his or her personal account,
person C assigned $5 to the group account and $5 to his or her personal account,
person D assigned $0 to the group account and $10 to his or her personal account, and
you assigned $5 to the group account and $5 to your personal account.

Fill in the blanks on the right.

(1) Amount you assigned to group account . . . . . $_____
(2) Amount you assigned to your personal account . . . . . $_____
[= $10 – group account assignment on line (1)]
(3) Total number of dollars assigned to your group account . . . . . $_____
(4) Income from the group account for a member of your group . . .$_____
[0.4 • group account total in line (3)]
(5) Your earnings after the assignment decisions . . . . . .$_____
[group account income in line (4) + personal account income
in line (2)]

Type in your contribution, press enter, and check your calculation.

[Screen 5]

Consider what would happen in practice 3 if you increase your assignment to the group account by $1.

Your personal account would go down by $1, reducing your earnings by $1.

Your group account would go up by $1, increasing your earnings by $0.40, for a net reduction of $0.60

But each of the other people in your group would increase their earnings by $0.40, for a total increase of $1.20 for the others in your group.

[Screen 6]

A Second Stage

After you and the others in your group learn of one another's assignments to the group account, it may be possible for you to alter your and their earnings. In particular, if your group decides to allow it, you will be able to reduce the earnings of one or more of the other members of the group at some cost to your own earnings. You would pay $0.25 for each $1 by which you reduced another person's earnings. You would be free to decide not to reduce a person's earnings, and not to incur any reduction cost, by simply typing a zero where possible reductions are entered.

[Screen 7]

To get a sense of how this would work, suppose, in the last example, that you decide to:

reduce B's earnings by $2
reduce C's earnings by $3
reduce D's earnings by $4

The total amount of reductions you make on others' earnings is $9.

It costs you $0.25 for each $1 you reduce others' earnings. So your own earnings are reduced by (0.25)($9) = $2.25 in this example.

Now, suppose

B reduces your earnings by $2
C reduces your earnings by $1
D reduces your earnings by $0

The total reduction of your earnings by others is ($2 + $1 + $0) = $3. Your screen will tell you how much your earnings have been reduced, but not who has reduced your earnings by what amount.

Similarly none of the others will learn by how much you have reduced their earnings. They will only learn their total reductions by others in the group as a whole.

Please fill in the sheet labeled practice 4 and the corresponding practice decision screen.

Practice 4.

You assigned $5 to the group account and $5 to your personal account,
person B assigned $10 to the group account and $0 to his or her personal account,
person C assigned $5 to the group account and $5 to his or her personal account, and
person D assigned $0 to the group account and $10 to his or her personal account.

You reduce person B's earnings by $2,
person C's earnings by $3, and

person D's earnings by $4.

You receive a total of $3 in reductions from other members of your group.

(1) Amount you assigned to group account . . . . . $_____
(2) Amount you assigned to your personal account . . . . . $_____
[$10 – group account assignment on line (1)]
(3) Total number of dollars assigned to your group account . . . . . $_____
(4) Income from the group account for a member of you group . . . $_____
[(0.4) • group account total in line (3)]
(5) Your earnings after the assignment decisions . . . . . .$_____
[group account income in line (4) + personal account income in line (2)]
(6) You reduced the earnings of others in your group by a total of . .$_____
(7) This cost you . . . . . $_____
[(0.25) • the sum of your reductions from line (6)]
(8) Other members of your group reduced your earnings by . . . . . $_____
(9) Your total earnings for this period . . . . . $_____
[Your earnings after the assignment decisions on line (5) minus your reduction cost on line (7) minus the amount by which your earnings were reduced on line (8)]

Enter your reductions, press enter, and check your calculation.

[Screen 8]

Whether or not members of your group have the option of reducing one another's earnings will be decided by voting. At the beginning of the experiment, you will be asked to vote on whether or not to allow reductions. The majority decision of your group will determine the rules under which your group will make the first six periods of decisions. After that, you will vote again, and the (possibly different) rules that your group selects will be in effect for the next six periods. Altogether, you will vote on the rules for your group five times, and each vote will be followed by six rounds of assignment and (possibly) reduction decisions. The experiment as a whole will consist of 30 periods, each having an assignment and possibly a reduction stage.

[Screen 9]

When determining the rules for your group, it will be possible to allow reductions of subjects who assign more than the group's average to its group account, of subjects who assign exactly the group average, and/or of subjects who assign less than the group average.

[Screen 10]

Your voting screen will look like the one below.

I vote to allow a person's earnings to be reduced if

		Yes	No	No preference
(j)	that person assigns less than the average amount to the group account	□	□	□
(k)	that person assigns the average amount to the group account	□	□	□
(l)	that person assigns more than the average amount to the group account	□	□	□

Please answer "Yes," "No," or "No preference" by clicking the box to the right of each of the three choices. If a majority of those expressing a preference vote yes, the reductions in question will be allowed; otherwise they will not.

[Screen 11]

Once the voting is done, you will be told what if any reductions will be allowed in your group for the next six periods. During the reduction stage of each round, the reduction box of any subject whose earnings you are not allowed to reduce, because of the rules decided by your group, will automatically show a zero, which cannot be changed. If your group decides to allow no reductions, you will see zeros in all of the boxes. But you may still have to wait a few moments if there are other groups that have decided to allow reductions.

[Screen 12]

Your Net Earnings

Your net earnings for a period or round will be:

Amount in personal account + (0.4)(total in group account) - (0.25)(total of your reductions of others) - total of reductions of your earnings made by others.

If this results in a negative number in any period, your earnings for that period will be set to zero.

[Screen 13]

Each period you begin with a new $10 and each period's earnings are independent of the others. After all thirty periods have been completed, your net earnings will be totaled and converted from experimental dollars to real dollars. Then $5 will be added for your participation. You will receive your earnings in cash before leaving the experiment.

[Screen 14]

During the experiment, there is to be no communication of any kind among participants, apart from the entering of decision numbers that will be transmitted to the other members of your group. It is important that you fully understand the decision process before we begin. Please raise your hand now if you have any questions.

- - - - -

END OF INSTRUCTIONS, FOLLOWED BY ACTUAL VOTING SCREEN AS ABOVE:

I vote to allow a person's earnings to be reduced if

(m) that person assigns less than the average amount to the group account	Yes □	No □	No preference □
(n) that person assigns the average amount to the group account	Yes □	No □	No preference □
(o) that person assigns more than the average amount to the group account	Yes □	No □	No preference □

- - - -

THIS WILL BE FOLLOWED BY A MESSAGE TO THE SUBJECTS PRIOR TO COMMENCEMENT OF THE 1st, 7th, 13th, 19th, and 25th PERIODS:

Possible messages:

Your group has voted not to allow group members to reduce one another's earnings.

Your group has voted to allow members to reduce the earnings of any other group member.

Your group has voted to allow members to reduce the earnings of any group member who assigns less than the average amount to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns the average amount or less to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns the average amount or more to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns less than the average amount, or more than the average amount, to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns the average amount to the group account.

Your group has voted to allow members to reduce the earnings of any group member who assigns more than the average amount to the group account.

- - - - -

MESSAGE BEFORE 2nd AND LATER VOTES:

You will now vote again on whether to allow reductions of the earnings of others in your group. As before, reductions of a given kind will be allowed only if there is a majority of Yes votes for it. After the vote, you will be informed of the outcome under which your group will operate for the next six periods. Before the vote, you will see a screen listing the decisions made by each of the groups in the experiment today, the average amount the group's members contributed to their group account, and the average amount that they earned, during the periods covered by that vote.