

Article

Group Size, Coordination, and the Effectiveness of Punishment in the Voluntary Contributions Mechanism: An Experimental Investigation

Bin Xu ^{1,†}, C. Bram Cadsby ^{2,*}, Liangcong Fan ³ and Fei Song ⁴

¹ Experimental Social Science Laboratory, Zhejiang University and Public Administration College, Zhejiang Gongshang University, Hangzhou, China; E-Mail: xubin211@163.com

² Department of Economics and Finance, University of Guelph, Guelph, ON, N1G 2W1, Canada; E-Mail: bcadsby@uoguelph.ca

³ Guanghua Law School, Zhejiang University, Hangzhou, China; E-Mail: flc1984@163.com

⁴ Ted Rogers School of Management, Ryerson University, Toronto, ON, M5G 2C5, Canada; E-Mail: fsong@ryerson.ca

† Bin Xu is the holder of the grant from the Social Science Experimental Center of Zhejiang University that funded this project. All authors contributed equally to the study.

* Author to whom correspondence should be addressed; E-Mail: bcadsby@uoguelph.ca; Tel.: +1-519-824 4120 ext.53320; Fax: +1-519-763 8497.

Received: 16 November 2012; in revised form: 29 January 2013 / Accepted: 5 February 2013 /

Published: 19 February 2013

Abstract: We examine the effectiveness of the individual-punishment mechanism in larger groups, comparing groups of four to groups of 40 participants. We find that the individual punishment mechanism is remarkably robust when the marginal per capita return (MPCR), *i.e.* the return to each participant from each dollar that is contributed, is held constant. Moreover, the efficiency gains from the punishment mechanism are significantly higher in the 40-participant than in the four-participant treatment. This is true despite the coordination problems inherent in an institution relying on decentralized individual punishment decisions in the context of a larger group. It reflects increased per capita expenditures on punishment that offset the greater coordination difficulties in the larger group. However, if the marginal group return (MGR), *i.e.* the return to the entire group of participants, stays constant, resulting in an MPCR that shrinks with group size, no such offset occurs and punishment loses much but not all of its effectiveness at encouraging

voluntary contributions to a public good. Efficiency is not significantly different from the small-group treatment.

Keywords: public goods; marginal per capita return; MPCR; punishment mechanism; large groups

1. Introduction

The voluntary contribution mechanism (VCM) has been an important topic of research in experimental economics. Among the many issues addressed by laboratory experiments is the relationship between group size and the level of contributions. Isaac, Walker and Williams [1] examined group sizes from four to 100, while simultaneously manipulating the marginal per capita return (MPCR), *i.e.* the return to each participant from each dollar that is contributed, between 0.03 and 0.75. Their main results show that with the MPCR held constant at 0.3, groups of 40 and 100 provide the public good at higher levels of efficiency than groups of four and 10 respectively. However, for an MPCR of 0.75, group size had no significant effect on public good provision. More recently, Weimann *et al.* [2] examined group sizes of 60 and 100 with MPCRs of 0.02 and 0.04. They found an MPCR effect, but little evidence of a group-size effect on contributions.

In a separate line of research, Fehr and Gächter [3,4] demonstrated that informing individual contributors of the contributions made by their peers, and then permitting those contributors to purchase punishments directed at individuals they specify is a remarkably effective means of motivating high contributions among groups of four participants. This is true under both partner and stranger designs. This result is especially noteworthy because the availability of these punishment opportunities does not alter the fact that complete free riding in contributions is still the unique stage-game Nash Equilibrium for the VCM with or without punishment opportunities.

A number of studies have examined the robustness of Fehr and Gächter's results with respect to punishment effectiveness and cost (Egas and Riedl [5] Nikiforakis and Normann [6]; Gardner and West [7]), communication (Bochet, Page, and Putterman [8]), self-selection of punishment *versus* non-punishment institution (Güerker, Irlenbusch, and Rockenbach [9]), monetary *versus* non-monetary punishment (Masclet *et al.* [10]), length of the game (Gächter, Renner, and Sefton [11]), alternative punishment institutions (Casari and Luini [12]), and country (Herrmann, Thöni, and Gächter [13]).¹ Carpenter [16] compares groups consisting of five *versus* ten participants. He also controls for the extent to which subjects can monitor each other. His results show that the availability of punishment promotes contributions for both groups of five and groups of ten, but that restrictions on monitoring can adversely affect contributions.

The effectiveness of the individual punishment mechanism in laboratory groups of four, five, or ten provides a persuasive explanation of how free-riding behavior can be mitigated in relatively small groups that need to mobilize contributions of money or effort towards a common public good. However, it is uncertain whether such a mechanism would continue to be effective in the much larger

¹ Related literatures examine rewards versus punishments (e.g., Rand *et al.* [14]) and the evolutionary emergence of punishment (e.g., Boyd, Gintis and Bowles [15]).

groups that must often cooperate together in the real world for the common good. Carpenter [16] finds that in ten-person groups there is some evidence that individuals punish less because of a bystander effect, *i.e.* second-order free riding in bearing the cost of punishment. He finds however that this is largely offset by the presence of more potential punishers. Casari [17] notes that Carpenter's design employs a punishment mechanism with a fine-to-fee ratio that increases with group size. As Casari points out, a higher fine-to-fee ratio has been associated with increased expenditures on punishment (Anderson and Putterman [18]; Andreoni, Harbaugh, and Vesterlund [19]; Egas and Riedl [5]; Nikiforakis and Normann [6]; Gardner and West [7]; Ostrom, Walker, and Gardner [20]). This could have motivated more punishment expenditures in Carpenter's ten-person than in his five-person groups, mitigating the potential coordination problem in the ten-person groups.

As group size increases, two potential problems arise with the individual punishment mechanism. First, it may become more difficult to identify free riders. For example, if four people share an office and are together obliged to keep the shared facilities clean, it may not be too difficult to identify and punish the one person who neglects to clean the microwave. However, if 40 people share an office, it may be more difficult to identify all ten people who fail to do their share. Since identification of the responsible individuals is necessary in order to punish them, such a problem could detract from the effectiveness of the punishment mechanism.² Second, even if free-riders can all be identified, potential coordination problems in the individual punishment mechanism multiply if each subject trying to decide whether or not to punish a low contributor is unable to observe which of those low contributors may be simultaneously receiving punishments from others. To continue the example, in a four-person office, three of the free-rider's co-workers may find it worthwhile to punish the free-rider by registering disapproval. However, in the 40-person office, it may seem too onerous for all 30 co-workers of the ten free-riders to take the time to punish all of them. If each co-worker instead punishes only one of the free-riders, there will be an average of three punishments per free-rider just as in the four-person office. The difference is that some free-riders may receive more than three punishments, while others receive fewer, and perhaps none at all.

The primary objective of our study is to focus on the latter problem. In particular, we examine the robustness of the individual-punishment mechanism at a constant fine-to-fee ratio in the context of the potential punishment coordination problems that may occur in larger groups even when all free-riders can be identified. Following Isaac, Walker and Williams [1], we compare groups of four *versus* 40 participants. In our four-person groups, the MPCR was set at 0.4. This implies a marginal group return (MGR) of $0.4 \cdot 4 = 1.6$, *i.e.* each contribution of one token results in 1.6 tokens divided equally among the four-person group. In half of our 40-person groups, we held the MPCR constant at 0.4, resulting in a MGR of $0.4 \cdot 40 = 16$, *i.e.* each contribution of one token creates 16 tokens divided equally among the 40-person group. In the other half of our 40-person groups, we held the MGR constant at 1.6, resulting in a reduced MPCR of just 0.04. Of course, we would expect the higher-MPCR group to contribute more to the public good than the lower-MPCR group as occurred in Isaac, Walker and Thomas [21], Isaac and Walker [22], Isaac, Walker and Williams [1] and Weimann *et al.* [2]. We also hypothesize that punishment will be more effective at raising contributions in the high- than in the low-MPCR group. This is because there is more motivation to punish low contributors when their increased

² We thank an anonymous referee for suggesting this example, which we have slightly modified and expanded upon.

contributions would have a greater effect on one's earnings.

It is less clear how an increase in group size, with a constant MPCR, would influence the effectiveness of the individual punishment mechanism. On the one hand, the increase in MGR might be expected to encourage the punishment of low contributors by those who care about the larger potential social surplus. On the other hand, the coordination problem described above may cause free-riding to take hold if some low contributors are not initially punished.

2. Experimental Design

Our specific experimental design adopted key elements from Fehr and Gächter's two important studies [3,4]. Like them, we employed a within-person design of punishment (P) *versus* non-punishment (N) conditions. In particular, each subject played ten rounds of N and ten rounds of P in a session. The order of P and N was reversed for half of the sessions. Henceforth, we call the former the NP order, while the latter is the PN order. Following Fehr and Gächter [3,4], we initially told the participants that they would be playing ten rounds in either the P or N condition. Afterwards, they were informed that they would be playing ten more rounds in a new experiment, and that the session would finish after this second set of ten rounds was played. We used a partner protocol both because of the practical difficulties of using a stranger design with 40-person groups and in order to focus on large groups that may have repeated opportunities for cooperation. We employed scrambled IDs from round to round so that no reputation could be built over time. The fine-to-fee ratio was set at 3:1 as in Fehr and Gächter [4]. Thus, spending one token to punish another person resulted in a three-token loss for that person. This ratio did not vary with either group size or MPCR. A participant could purchase a maximum of ten punishment points directed at each of the other participants.

Each subject was endowed with 20 tokens for each round. As in Fehr and Gächter [3,4], a subject who did not punish others could not lose money. Punishment points received could not reduce income from the contribution stage of the game to less than zero. However, spending money on punishing others created the possibility of losing money. For example, if one received enough punishment points to reduce one's earnings from the contribution stage to zero, any punishment points previously purchased would result in a loss. Following Fehr and Gächter [3,4], we gave each subject an extra sum of tokens at the beginning of the P rounds to reduce the possibility of somebody leaving the session owing the experimenter money. These extra tokens could not be used either to make contributions or to punish others. They were made available only to offset potential losses. We used 25 tokens for the four-person groups as in Fehr and Gächter [3,4]³. The 40-person groups posed a bigger problem in this regard. Within such groups, there was a much greater chance of receiving enough punishment points to reduce a subject's contribution-stage earnings to a very low number or even to zero since each subject could receive punishment points from up to 39 other participants. Thus, each participant was in greater danger of being in a position where the purchase of punishment points could result in owing the experimenter money. Moreover, it was possible to lose a much greater sum of money than in the four-person case since one could potentially purchase punishment points for up to 39 other

³ See the experimental instructions associated with each of Fehr and Gächter [3,4] for details.

participants. Thus, we used 500 tokens to mitigate this possibility for the 40-person groups.⁴ No such losses occurred in the experiment.

The exchange rate was set at 21 Tokens = 1 RMB for group size = 4 and 39.23 (150) Tokens = 1 RMB for group size = 40 with MPCR = 0.04 (0.4). These exchange rates were calculated by holding the mean of the free-riding payoff and the full-contribution payoff plus the 25 (500) tokens for the four- (40-) person P condition equal in RMB between these treatments. Lastly, each subject was also given a 10 RMB show-up fee.

In summary, there are three independent variables: group size (small/large, namely four *versus* 40), MPCR (low/high, namely 0.04 *versus* 0.4), and decision order (NP *versus* PN). MGR is the product of group size and MPCR (low/high, namely 1.6 *versus* 16). Since the MPCR of 0.04 can only be used for 40-person groups, there were six treatments in total:

1. Small group (4), PN, High MPCR (0.4), Low MGR (1.6), ten groups
2. Large group (40), PN, High MPCR (0.4), High MGR (16), three groups
3. Large group (40), PN, Low MPCR (0.04), Low MGR (1.6), three groups
4. Small group (4), NP, High MPCR (0.4), Low MGR (1.6), ten groups
5. Large group (40), NP, High MPCR (0.4), High MGR (16), three groups
6. Large group (40), NP, Low MPCR (0.04), Low MGR (1.6), three groups

The two small-group treatments always have a high MPCR and a low MGR. In what follows, we will refer to them simply as small-group treatments. In contrast, it is necessary to distinguish between the large group treatments with a high MPCR (and high MGR) and those with a low MPCR (and low MGR). The six treatments are displayed in Table 1.

Subjects were randomly recruited via online advertisements at Zhejiang University in Hangzhou, China. All subjects were full-time undergraduate students in diverse majors across the Sciences, Social Sciences, and Humanities. A total of 560 subjects participated in the study. All sessions were run at the Zhejiang University Experimental Social Science Laboratory.

All sessions were computerized.⁵ Upon arrival, each subject was seated at a private computer carrel. Each session lasted about 100 minutes. The average earnings for each subject were approximately 39.6 RMB including a 10 RMB show-up fee. At the time of the experiment, 39.6 RMB was equal to about \$5.82 US. For comparison purposes, the wage rate for Zhejiang University undergraduates who had part-time jobs with the university administration was 12 RMB per hour.

3. Results

3.1. Contributions in the Punishment versus Non-Punishment Condition

Table 1 presents a data summary by treatment of the sum of contributions per capita in the punishment rounds, in the non-punishment rounds and the difference between them. In all cases, the differences between contributions in the P condition and contributions in the N condition are positive.

⁴ University ethics board requirements made it essential to ensure that no participant left the experiment with less money than when s/he arrived.

⁵ Zhijian Wang and Bin Xu jointly designed, tested and implemented the computer program used in this experiment.

Table 1. Data summary. (Note that NP Order means that the ten non-punishment rounds preceded the ten rounds with punishment while PN Order means that the ten rounds with punishment preceded the non-punishment rounds.)

Treatment	Sample Size	P_N difference per capita	P contribution per capita	N contribution per capita	Punishments per capita
1: Small (n=4), High MPCR=0.4, NP Order	10	69.83	124.40	54.58	10.6
2: Large (n=40), High MPCR=0.4, NP Order	3	65.15	175.83	110.68	32.89
3: Large (n=40), Low MPCR=0.04, NP Order	3	32.61	63.60	30.99	8.47
4: Small (n=4), High MPCR=0.4, PN Order	10	38.08	125.18	87.10	6.08
5: Large (n=40), High MPCR=0.4, PN Order	3	43.77	126.54	82.78	26.30
6: Large (n=40), Low MPCR=0.04, PN Order	3	36.41	54.10	17.70	9.68

Table 2 presents regression results and related hypothesis tests using individual data. The dependent variable is the difference between contributions over all ten rounds of the P condition and contributions over all ten rounds of the N condition for each individual participant. Thus, there is one observation for each individual participant, 560 in all. Since the individuals were organized into 32 groups of either four or 40 participants, the individual observations for participants in the same group are not independent. We cannot use group-specific fixed effects to correct for this problem because it is impossible to disentangle such fixed effects from the between-group treatment effects that are the focus of our analysis.⁶ Thus, we use random-effects for each group.⁷ The independent variables are all dummy variables representing the different treatments. Large_High is one for the two 40-participant, high MPCR treatments and zero otherwise. Large_Low is one for the two 40-participant, low MPCR treatments and zero otherwise. PN_Order is one for the PN order and zero for the NP order. There are two interaction variables: Large_High $\times$ PN_Order and Large_Low $\times$ PN_Order.

In Table 2, to ease interpretation, the treatment numbers in square brackets to the right of each coefficient and hypothesis test correspond to the treatment numbers from Table 1. For example, Trmt.1 to the right of β_0 indicates that the constant term represents the value of the dependent variable for the

⁶ Such a regression is completely collinear and thus cannot be run.

⁷ As a robustness check, we also employed two alternative estimation techniques: the robust standard error clustering of errors by group and the combination of a random effect for each group plus robust standard error clustering by group. These different estimation techniques yield identical coefficients, but slightly different standard errors. There are no qualitative differences in inferences regarding treatment effects within either the NP or PN orders. To save space, these results are not reported here, but are available from the authors upon request.

small (*i.e.*, 4-participant) treatment conducted in the NP order, treatment 1 in Table 1. Similarly, the Trmt. 2–3 to the right of the treatment effect $\beta_1 - \beta_2$ indicates that this expression represents the difference between treatment 2 and treatment 3 as defined in Table 1.

Table 2. Regression results on ten-round per capita differences in contributions between the punishment and no-punishment conditions (p-values in parentheses) [Treatment numbers in square brackets as defined in Table 1].

Estimation: DV= $\beta_0 + \beta_1(\text{Large_High}) + \beta_2(\text{Large_Low}) + \beta_3(\text{Order}) + \beta_4(\text{Large_High} \times \text{Order}) + \beta_5(\text{Large_Low} \times \text{Order})$	
Observations: 560 Number of Groups: 32 Overall R squared: 0.109	
Coefficients	Ten-round per capita differences in each treatment
$\beta_0 = 69.83$ (0.000) [Trmt. 1]	$\beta_0 + \beta_1 = 65.15$ (0.000) [Trmt. 2]
$\beta_1 = -4.68$ (0.741) [Trmt. 2–1]	$\beta_0 + \beta_2 = 32.61$ (0.000) [Trmt. 3]
$\beta_2 = -37.22$ (0.009) [Trmt. 3–1]	$\beta_0 + \beta_3 = 38.08$ (0.000) [Trmt. 4]
$\beta_3 = -31.75$ (0.007) [Trmt. 4–1]	$\beta_0 + \beta_1 + \beta_3 + \beta_4 = 43.77$ (0.000) [Trmt. 5]
$\beta_4 = 10.37$ (0.604) [Trmt. (5–2) – (4–1)]	$\beta_0 + \beta_2 + \beta_3 + \beta_5 = 36.41$ (0.000) [Trmt. 6]
$\beta_5 = 35.55$ (0.076) [Trmt. (6–3) – (4–1)]	
Treatment effects	Order effects
$\beta_1 - \beta_2 = 32.54$ (0.044) [Trmt. 2–3]	$\beta_3 + \beta_4 = -21.38$ (0.185) [Trmt. 5–2]
$\beta_1 + \beta_4 = 5.69$ (0.688) [Trmt. 5–4]	$\beta_3 + \beta_5 = 3.80$ (0.814) [Trmt. 6–3]
$\beta_2 + \beta_5 = -1.67$ (0.906) [Trmt. 6–4]	$\beta_5 - \beta_4 = 25.18$ (0.269) [Trmt. (6–3) – (5–2)]
$(\beta_1 + \beta_4) - (\beta_2 + \beta_5) = 7.36$ (0.648) [Trmt. 5–6]	

The first thing to notice is that, for all six treatments, the difference in ten-round per capita contributions between punishment and non-punishment rounds is significant with a p -value of 0.00. Thus, punishment made a significant difference to contributions in all six treatments. Second, in the NP order, the effectiveness of punishment at increasing contributions is significantly higher in the high-MPCR than in the low-MPCR large-group treatment ($p = 0.044$). Third, in the NP order, the effectiveness of punishment at increasing contributions is also significantly higher in the small-group treatment than in the low-MPCR large-group treatment ($p = 0.009$). Fourth, there is no significant difference in the effectiveness of punishment related to group size for a constant high MPCR in the NP order. Fifth, there is a significant order effect in the small-group treatment with punishment being less effective in the PN order ($p = 0.007$). Sixth, there are no significant treatment effects in the PN order.

It may take time for participants to adjust to the change of condition. Thus, it is interesting to examine the analogous results for the last round under each condition.⁸ Table 3 reports these results.

⁸ See footnote 7.

Table 3. Regression results on last-round per capita differences in contributions between the punishment and no-punishment conditions (p-values in parentheses) [Treatment numbers in square brackets as defined in Table 1].

Estimation: $DV = \beta_0 + \beta_1(\text{Large_High}) + \beta_2(\text{Large_Low}) + \beta_3(\text{Order}) + \beta_4(\text{Large_High} \times \text{Order}) + \beta_5(\text{Large_Low} \times \text{Order})$	
Observations: 560 Number of Groups: 32 Overall R squared: 0.164	
Coefficients	Last-round per capita differences in each treatment
$\beta_0 = 12.28$ (0.000) [Trmt. 1]	$\beta_0 + \beta_1 = 10.10$ (0.000) [Trmt. 2]
$\beta_1 = -2.18$ (0.490) [Trmt. 2–1]	$\beta_0 + \beta_2 = 3.44$ (0.186) [Trmt. 3]
$\beta_2 = -8.83$ (0.005) [Trmt. 3–1]	$\beta_0 + \beta_3 = 9.000$ (0.000) [Trmt. 4]
$\beta_3 = -3.28$ (0.191) [Trmt. 4–1]	$\beta_0 + \beta_1 + \beta_3 + \beta_4 = 8.38$ (0.001) [Trmt. 5]
$\beta_4 = 1.55$ (0.728) [Trmt. (5–2) – (4–1)]	$\beta_0 + \beta_2 + \beta_3 + \beta_5 = 2.89$ (0.266) [Trmt. 6]
$\beta_5 = 2.73$ (0.076) [Trmt. (6–3) – (4–1)]	
Treatment effects	Order effects
$\beta_1 - \beta_2 = 6.66$ (0.070) [Trmt. 2–3]	$\beta_3 + \beta_4 = -1.73$ (0.639) [Trmt. 5–2]
$\beta_1 + \beta_4 = -0.63$ (0.843) [Trmt. 5–4]	$\beta_3 + \beta_5 = -0.55$ (0.881) [Trmt. 6–3]
$\beta_2 + \beta_5 = -6.11$ (0.052) [Trmt. 6–4]	$\beta_5 - \beta_4 = 1.18$ (0.821) [Trmt. (6–3) – (5–2)]
$(\beta_1 + \beta_4) - (\beta_2 + \beta_5) = 5.48$ (0.136) [Trmt. 5–6]	

The difference in last-round per capita contributions between punishment and non-punishment rounds is significant for both the small-group ($p = 0.000$ for both NP and PN orders) and the high-MPCR large-group ($p = 0.000$ for NP order and $p = 0.001$ for PN order) treatments, indicating that punishment makes a significant difference in these cases. However, in contrast to the ten-round average data, these differences are not significant for the low-MPCR large-group treatments. Thus, we cannot reject the null hypothesis that punishment makes no difference to the level of contributions when the MPCR is low. In the NP order, the effectiveness of punishment at increasing contributions is significantly lower in the low-MPCR large-group treatment than in the small-group treatment ($p = 0.005$) and lower but with just marginal significance in comparison with the high-MPCR large-group treatment ($p = 0.070$). In the PN order, there is a significant difference in the effect of punishment only between the low-MPCR large-group and small-group treatments ($p = 0.052$). The effectiveness of punishment is not significantly influenced by group size for a constant high MPCR in either the NP or PN order.

In contrast to the ten-round average data, none of the order effects or interactions involving order effects is individually significant for the last-round data. Moreover, a joint test that the coefficients on the main order effect together with those on its interactions with the two other treatment dummies all equal zero yields a Chi-Square statistic of 1.95 with three degrees of freedom ($p = 0.583$). This suggests that the observed differences between the effectiveness of punishment in the NP *versus* the PN order have to do with the transition from N to P relative to the transition from P to N, and vanish by the tenth repetition within the N or P condition. Dropping the order effects, we can aggregate the NP and PN data and re-estimate the regressions using the aggregated data. The results are reported in Table 4.⁹

⁹ See footnote 7.

Table 4. Regression results on last-round per capita differences in contributions between the punishment and no-punishment conditions dropping insignificant order effects (p-values in parentheses) [Treatment numbers in square brackets as defined in Table 1 where 1, 4 indicates treatments 1 and 4 combined].

Estimation: $DV = \beta_0 + \beta_1(\text{Large_High}) + \beta_2(\text{Large_Low})$	
Observations: 560 Number of Groups: 32 Overall R squared: 0.154	
Coefficients	Last-round per capita differences in each treatment
$\beta_0 = 10.64 (0.000)$ [Trmt. 1, 4]	$\beta_0 + \beta_1 = 9.24 (0.000)$ [Trmt. (2, 5)]
$\beta_1 = -1.40 (0.522)$ [Trmt. (2, 5) – (1, 4)]	$\beta_0 + \beta_2 = 3.17 (0.079)$ [Trmt. (3, 6)]
$\beta_2 = -7.47 (0.001)$ [Trmt. (3, 6) – (1, 4)]	
Treatment effects	
$\beta_1 - \beta_2 = 6.07 (0.017)$ [Trmt. (2, 5) – (3, 6)]	

The difference in last-round per capita contributions between punishment and non-punishment rounds continues to be significant for the small-group and high-MPCR large-group cases ($p = 0.000$ in both cases). For the low-MPCR large-group treatment, it now attains marginal significance ($p = 0.079$), yielding some weak evidence that punishment has an effect on contributions even in this case. However, the effectiveness of the punishment condition at increasing contributions is significantly lower in the low-MPCR large-group treatment than in either the small-group ($p = 0.001$) or the high-MPCR large-group ($p = 0.017$) treatments by the last round of each condition. Once again, group size has no significant effect for a constant high MPCR.

3.2. Expenditure on Punishment

Is the punishment condition less effective in the low-MPCR large-group treatment simply because fewer punishments are purchased when the potential gains from further contributions are relatively small? The last column of Table 1 presents per capita expenditures on punishment for each treatment. In both orders, such expenditures appear to be substantially higher in the high-MPCR large-group treatment than in the other two treatments. The high-MPCR large group has a high MGR of 16, while the other two groups have a much lower MGR of just 1.6. It would appear that the higher MGR elicits greater per capita expenditures on punishment.

To investigate this issue further, we regress per capita expenditures on punishment for each group aggregated over all ten punishment rounds on the same dummy variables representing the different treatments as used above. There are 32 observations, one for each group. The estimated coefficients and related hypothesis tests are presented in Table 5. None of the order effects or their interactions with the treatment dummy variables is significant. While per capita punishment expenditures in the high-MPCR large group treatment are significantly higher than in both the small group treatment ($p = 0.000$ and $p = 0.001$ for the NP and PN orders respectively) and the low-MPCR large group treatment ($p = 0.001$ and $p = 0.023$ for the NP and PN orders respectively), there is no significant difference in per capita punishment expenditures between the small-group and the low-MPCR large group treatments for either order. A joint test that the coefficients on the main order effect together with those on its interactions with the two other treatment dummies all equal zero yields, an $F(3, 26)$

statistic of 0.80 ($p = 0.506$). Dropping these order effects leads to qualitatively identical inferences.¹⁰

Table 5. Regression results on ten-round per capita expenditures on punishment (p-values in parentheses) [Treatment numbers in square brackets as defined in Table 1].

Estimation: $DV = \beta_0 + \beta_1(\text{Large_High}) + \beta_2(\text{Large_Low}) + \beta_3(\text{Order}) + \beta_4(\text{Large_High} \times \text{Order}) + \beta_5(\text{Large_Low} \times \text{Order})$	
Observations: 32 Adjusted R squared: 0.475	
Coefficients	Ten-round per capita expenditures on punishment in each treatment
$\beta_0 = 10.60$ (0.000) [Trmt. 1]	$\beta_0 + \beta_1 = 32.89$ (0.000) [Trmt. 2]
$\beta_1 = 22.29$ (0.000) [Trmt. (2–1)]	$\beta_0 + \beta_2 = 8.47$ (0.093) [Trmt. 3]
$\beta_2 = -2.13$ (0.703) [Trmt. (3–1)]	$\beta_0 + \beta_3 = 6.08$ (0.031) [Trmt. 4]
$\beta_3 = -4.53$ (0.240) [Trmt. (4–1)]	$\beta_0 + \beta_1 + \beta_3 + \beta_4 = 26.30$ (0.000) [Trmt. 5]
$\beta_4 = -2.07$ (0.794) [Trmt. (5–2) – (4–1)]	$\beta_0 + \beta_2 + \beta_3 + \beta_5 = 9.68$ (0.057) [Trmt. 6]
$\beta_5 = 5.73$ (0.471) [Trmt. (6–3) – (4–1)]	
Treatment effects	Order effects
$\beta_1 - \beta_2 = 24.43$ (0.001) [Trmt. (5–2)]	$\beta_3 + \beta_4 = -6.59$ (0.346) [Trmt.(5–2)]
$\beta_1 + \beta_4 = 20.23$ (0.001) [Trmt. (5–2)]	$\beta_3 + \beta_5 = 1.21$ (0.862) [Trmt. (6–3)]
$\beta_2 + \beta_5 = 3.60$ (0.521) [Trmt. (5–2)]	$\beta_5 - \beta_4 = 7.80$ (0.429) [Trmt. (6–3) – (5–2)]
$(\beta_1 + \beta_4) - (\beta_2 + \beta_5) = 16.63$ (0.023) [Trmt. (5–6)]	

3.3. Coordination Problem with the Punishment Mechanism in Large Groups

While per capita expenditures on punishment are significantly higher in the high-MPCR large-group treatment, the only high-MGR treatment, than in the other two low-MGR treatments, the effectiveness of the punishment condition at increasing contributions is significantly higher in both the high-MPCR large-group treatment and the small-group treatment than in the low-MPCR large-group treatment. Thus, statistically indistinguishable levels of per capita spending on punishment are significantly more effective at increasing contributions in the small-group treatment than in the low-MPCR large group treatment. Moreover, significantly higher levels of per capita spending on punishment in the high-MPCR large-group treatment relative to the small-group treatment produce increases in contributions that are statistically indistinguishable from each other. We hypothesize that this reflects a coordination problem that afflicts the decentralized punishment mechanism in large groups, making per capita expenditures on punishment less effective at increasing contributions in such groups.

Suppose for example that 25% of participants are low contributors. In a group of four, this implies that there is just one low contributor and three higher contributors who might decide to punish him or her. Suppose that each high contributor purchases one punishment point. The low contributor will receive three punishment points, perhaps an inducement to contribute more in the next round. In an analogous group of 40, there would be ten low contributors and thirty higher contributors who might decide to punish one or more of the ten low contributors. If each high contributor purchases one punishment point, the ten low contributors will together receive thirty punishment points, an average of three per person. It is possible that these thirty punishment points will be divided equally among the

¹⁰ To conserve space, these results are not reported in detail here. They are available from the authors upon request.

ten low contributors. In that case, each low contributor will receive three punishment points just as in the small four-person group. However, there is no mechanism to coordinate the distribution of punishment points among the low contributors. Therefore, it is unlikely that they will be distributed equally. Instead it is probable that some low contributors will receive more punishment points than necessary to motivate higher contributions, while others will receive fewer or none at all.

Table 6. Proportion of low contributors punished averaged across sessions by treatment for two definitions of low contributor.

Treatment	Sample Size	25 th Percentile or Lower and not Highest in Round	Ten or Lower and not Highest in Round
1: Small (n=4), High MPCR=0.4, NP Order	10	0.609	0.606
2: Large (n=40), High MPCR=0.4, NP Order	3	0.761	0.741
3: Large (n=40), Low MPCR=0.04, NP Order	3	0.347	0.188
4: Small (n=4), High MPCR=0.4, PN Order	10	0.674	0.538
5: Large (n=40), High MPCR=0.4, PN Order	3	0.689	0.486
6: Large (n=40), Low MPCR=0.04, PN Order	3	0.476	0.238

Table 6 presents summary data on the proportion of “low” contributors that received at least one punishment point for each treatment. We use two definitions of a low contributor. The first is a relative definition. It defines a contributor to be low if his/her contribution is at or below the 25th percentile in a round and s/he is not one of the highest contributors in that round. The second is primarily an absolute definition. It defines those contributing ten or fewer tokens as low contributors as long as they are not among the highest contributors in the round. According to both definitions, the proportion of low contributors receiving at least one punishment point was substantially lower in the low-MPCR large group treatment than in either of the other two treatments in both the NP and PN orders.

To determine whether there is a significant difference in the likelihood of a low contributor being punished in the low-MPCR large group treatment than in the other two treatments, we employed a negative binomial regression for each definition of a low contributor. For each group of participants, we have one count of the number of times a low contributor received at least one punishment aggregated across all rounds. This is the dependent variable. In addition, we calculate the number of times a low contribution occurred aggregated across all rounds, the log of which is used as the

exposure variable.¹¹ To facilitate interpretation, we report coefficients and the related hypothesis tests as well as the corresponding incidence rate ratios (IRRs). Since IRRs provide a more intuitive interpretation, we focus on them in the discussion that follows.

Table 7 presents the results for the relative definition. Consider the reported IRR for β_2 , which is 0.593. This means that the estimated rate at which low contributors received at least one punishment in the low-MPCR large-group treatment was 59.3% as high as the analogous rate in the small-group treatment for the NP order. Since the p -value is 0.007, this is a significant difference. Similarly, a hypothesis test indicates that the rate at which low contributors received at least one punishment in the high-MPCR large group treatment was 214.9% as high as the analogous rate in the low-MPCR large-group treatment ($p = 0.000$). For the PN order, the incidence rate for the low-MPCR large-group treatment was 70.6% of the rate for the small-group treatment with marginal significance ($p = 0.066$), while the rate for the high-MPCR large-group treatment was 144.1% of the rate for the low-MPCR large-group treatment ($p = 0.053$). There is no significant difference between the incidence rates for the small-group *versus* the high-MPCR large-group treatment in either the NP or PN order. Moreover, there are no significant order effects.

Table 7. Negative binomial regression results for the proportion of times people in the lowest contribution quartile who were not among the highest contributors in a round were punished (p-values in parentheses) [Treatment numbers in square brackets as defined in Table 1].

Estimation: $DV = \beta_0 + \beta_1(\text{Large_High}) + \beta_2(\text{Large_Low}) + \beta_3(\text{Order}) + \beta_4(\text{Large_High} \times \text{Order}) + \beta_5(\text{Large_Low} \times \text{Order})$	
Observations: 32 Pseudo R squared: 0.061	
Coefficients	IRRs
$\beta_0 = -0.54$ (0.000)	
$\beta_1 = 0.24$ (0.202) [Trmt. (2–1)]	1.276
$\beta_2 = -0.52$ (0.007) [Trmt. (3–1)]	0.593
$\beta_3 = 0.14$ (0.462) [Trmt. (4–1)]	1.151
$\beta_4 = -0.23$ (0.397) [Trmt. (5–2) – (4–1)]	0.797
$\beta_5 = 0.17$ (0.522) [Trmt. (6–3) – (4–1)]	1.189
Treatment effects	
$\beta_1 - \beta_2 = 0.77$ (0.000) [Trmt. (2–3)]	2.149
$\beta_1 + \beta_4 = 0.02$ (0.929) [Trmt. (5–4)]	1.017
$\beta_2 + \beta_5 = -0.35$ (0.066) [Trmt. (6–4)]	0.706
$(\beta_1 + \beta_4) - (\beta_2 + \beta_5) = 0.37$ (0.053) [Trmt. (5–6)]	1.441
Order effects	
$\beta_3 + \beta_4 = -0.09$ (0.646) [Trmt. (5–2)]	0.917
$\beta_3 + \beta_5 = 0.31$ (0.103) [Trmt. (6–3)]	1.369
$\beta_5 - \beta_4 = 0.40$ (0.137) [Trmt. (6–3) – (5–2)]	1.492

¹¹ The exposure variable adjusts for the differing numbers of low contributions in each group. The proportions for each treatment presented in Table 6 are averages across such proportions, calculated for each group in a treatment. The numerator of each such group proportion is the count of the number of times a low contributor received at least one punishment, while the denominator is the number of times a low contribution occurred aggregated across all rounds.

A joint test of the null hypothesis that the order effect and its interactions with the treatment variables all equal zero yields a chi-square statistic of 3.41 with three degrees of freedom ($p = 0.332$). Thus, the null hypothesis of no order effects or interactions involving order effects cannot be rejected. Dropping these order effects and re-estimating this negative binomial regression leads to the likelihood of low contributors receiving at least one punishment being significantly lower in the low-MPCR large-group treatment than in either the small-group ($p = 0.003$) or the high-MPCR large-group ($p = 0.000$) treatment. As before, there is no significant difference between the incidence rates for the small-group *versus* the high-MPCR large-group treatment ($p = 0.370$).¹²

Table 8 presents the results for the primarily absolute definition of low contributor. There are marginally significant order effects for the high-MPCR large group treatments ($p = 0.081$) and a significant interaction between the effect of MPCR and order ($p = 0.048$). However, the treatment effects are robust to the altered definition of low contributor. The incidence rates for the low-MPCR large group treatments are significantly lower than for the small-group treatments ($p = 0.000$ for both the NP and PN orders). Moreover, the incidence rate for the high-MPCR large group treatments are significantly higher than for the low-MPCR treatments ($p = 0.000$ for the NP and $p = 0.003$ for the PN order). There is no significant difference between the incidence rates for the small *versus* the high-MPCR large group treatment in either order.

Table 8. Negative binomial regression results for the proportion of times people who contributed ten or less who were not among the highest contributors in a round were punished (p-values in parentheses) [Treatment numbers in square brackets as defined in Table 1].

Estimation: $DV = \beta_0 + \beta_1(\text{Large_High}) + \beta_2(\text{Large_Low}) + \beta_3(\text{Order}) + \beta_4(\text{Large_High} \times \text{Order}) + \beta_5(\text{Large_Low} \times \text{Order})$	
Observations: 32 Pseudo R squared: 0.104	
Coefficients	IRRs
$\beta_0 = -0.61$ (0.000) [Trmt. 1]	
$\beta_1 = 0.29$ (0.228) [Trmt. (2–1)]	1.335
$\beta_2 = -1.06$ (0.000) [Trmt. (3–1)]	0.347
$\beta_3 = 0.08$ (0.720) [Trmt. (4–1)]	1.078
$\beta_4 = -0.51$ (0.119) [Trmt. (5–2) – (4–1)]	0.603
$\beta_5 = 0.16$ (0.608) [Trmt. (6–3) – (4–1)]	1.171
Treatment effects	
$\beta_1 - \beta_2 = 1.35$ (0.000) [Trmt. (2–3)]	3.844
$\beta_1 + \beta_4 = -0.22$ (0.319) [Trmt. (5–4)]	0.805
$\beta_2 + \beta_5 = -0.90$ (0.000) [Trmt. (6–4)]	0.407
$(\beta_1 + \beta_4) - (\beta_2 + \beta_5) = 0.68$ (0.003) [Trmt. (5–6)]	1.980
Order effects	
$\beta_3 + \beta_4 = -0.43$ (0.081) [Trmt. (5–2)]	0.650
$\beta_3 + \beta_5 = 0.23$ (0.304) [Trmt. (6–3)]	1.263
$\beta_5 - \beta_4 = 0.66$ (0.048) [Trmt. (6–3) – (5–2)]	1.942

¹² The detailed results are not reported in order to conserve space. They are available from the authors upon request.

These results together corroborate the coordination hypothesis, supporting the idea that a given per capita expenditure on decentralized individual punishments is more effective at increasing contributions for smaller than for larger groups. In small groups, for a given level of per capita expenditure, a higher proportion of low contributors receive at least one punishment than in large groups. This is the reason that statistically indistinguishable amounts of expenditure on punishment are significantly more effective in the small-group treatment than in the low-MPCR large-group treatment at increasing contributions. It is also the reason that the significantly higher expenditures on punishment observed in the high-MPCR large-group treatment relative to the small-group treatment are necessary to produce similar increases in contributions that are statistically indistinguishable from each other.

3.4. Efficiency of the Punishment Mechanism

Finally, it is interesting to examine whether the punishment mechanism is more or less efficient than the stand-alone VCM in each treatment when both the benefits of contributions and the costs of punishment are taken into account. For each treatment, Table 9 presents the per capita income difference in experimental tokens between the VCM round and the corresponding punishment round as well as the aggregate difference for all ten rounds. In general, efficiency increases from round one to round 10 in all treatments. This is because contributions tend to fall over the VCM rounds, while they tend to rise or at least fall at a slower rate over the punishment rounds. Moreover, expenditures on punishment become less necessary as contributions rise. The greatest gains from the punishment mechanism occur in the high-MPCR large-group treatment. This is primarily because a one-token increase in contributions has a much higher MGR in the high-MPCR large-group treatment (16) than in the other two treatments (1.6), resulting in greater efficiency gains despite the increased expenditures on punishments. It is only in the high-MPCR large-group treatment that earnings in the ten punishment rounds together exceed earnings in the ten VCM rounds. In the other two treatments, efficiency gains start to occur only in rounds seven or eight, and punishment round earnings are lower in aggregate than earnings over the VCM rounds. A regression analysis analogous to the one for expenditures on punishment presented in Table 5 confirms that aggregate efficiency gains in the punishment rounds are significantly higher for the high-MPCR large-group treatment than for either of the other two treatments regardless of order with $p = 0.000$ in all four cases. There are no significant differences in aggregate efficiency gains from punishment between the small-group and low-MPCR large-group treatments. For aggregate efficiency, it is the high MGR emanating from a constant-MPCR pure public good being spread over ten times as many people that causes the availability of punishment to make a significant difference.¹³

¹³ These results are not reported here to save space, but are available from the authors upon request.

Table 9. Per capita income difference in experimental tokens.

	NP Order			PN Order		
	Small High MPCR	Large High MPCR	Large Low MPCR	Small High MPCR	Large High MPCR	Large Low MPCR
Round 1	−5.88	27.00	−5.10	−4.20	−22.50	−5.10
Round 2	−4.83	21.00	−2.35	−3.36	10.50	−2.75
Round 3	−2.52	27.00	−1.57	−1.05	22.50	−1.18
Round 4	−1.26	57.00	−1.18	−0.63	57.00	−1.57
Round 5	−0.21	93.00	−0.78	−0.42	46.50	−0.39
Round 6	1.26	124.50	−0.39	−0.42	51.00	−1.96
Round 7	−0.21	123.00	0.00	1.47	79.50	0.00
Round 8	5.88	111.00	0.00	2.10	88.50	0.78
Round 9	5.46	123.00	0.00	2.73	105.00	0.39
Round 10	2.31	141.00	0.78	2.10	115.50	−0.78
Sum of 10 rounds	−0.04	846.15	−10.36	−1.45	551.70	−12.48

4. Conclusion

The effectiveness of the individual punishment mechanism at increasing contributions to a public good depends critically on what happens to the MPCR of a public good as the potential community of contributors grows. For a pure public good with non-rivalry in consumption, MPCR stays constant and MGR increases proportionally with the size of the community. In this paper, we have demonstrated that the higher MGR produces a significant increase in per capita expenditures on punishment in 40-person relative to four-person groups. At the same time, the larger group creates a coordination problem for the decentralized punishment mechanism, making each dollar spent on punishment less effective at increasing contributions. This occurs because some punishment dollars are inevitably wasted on low contributors who are simultaneously punished sufficiently to increase their contributions by other purchasers of punishment points, while other low contributors escape punishment. In this experimental study, the increase in punishment expenditures was sufficient to offset the reduction in the effectiveness of each punishment dollar. Thus, for a constant MPCR, the individual punishment mechanism proved remarkably robust despite the coordination problems inherent in an institution relying on decentralized individual punishment decisions in the context of a larger group. In fact, despite the rise in cost resulting from the increase in punishment expenditures, the higher MGR on each contribution in the high-MPCR large-group treatment made the punishment mechanism significantly more efficient than in the other two treatments. In fact, the high-MPCR large-group treatment was the only treatment in which aggregate earnings over all ten punishment rounds exceeded aggregate earnings over all ten non-punishment rounds.

However, if the MGR stays constant, resulting in an MPCR that shrinks with group size, per capita expenditures on punishment do not increase. In this case, the coordination problem associated with the 40-person group is not offset by increases in punishment expenditures. This results in the individual punishment mechanism being significantly less effective at increasing contributions for a 40-person than for a four-person community with the same MGR. Examining institutional modifications to

mitigate the coordination problem associated with the decentralized individual punishment mechanism is an important issue deserving further study.

Acknowledgements

We thank Ananish Chaudhuri and Jeffrey Carpenter for helpful comments on an earlier draft of this manuscript. We also acknowledge Qiqi Cheng, Lu Liu, Chao Wang, Tongyu Wu, and Xinchao Zhang for their excellent research assistance, and Zhiwei Fang and Yanmin Qian for their support.

References and Notes

1. Isaac, R.M.; Walker, J.M.; Williams, A.W. Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups. *J. Public Econ.* **1994**, *54*, 1-36.
2. Weimann, J.; Brosig-Koch, J.; Hennig-Schmidt, H.; Keser, C.; Stahr, C. Public-good experiments with large groups. Working Paper, 2012.
3. Fehr, E.; Gächter, S. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **2000**, *90*, 980-994.
4. Fehr, E.; Gächter, S. Altruistic punishment in humans. *Nature* **2002**, *415*, 137-140.
5. Egas, M.; Riedl, A. The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B: Biol. Sci.* **2008**, *275*, 871-878.
6. Nikiforakis, N.; Normann, H. A comparative analysis of punishment in public-good experiments. *Exp. Econ.* **2008**, *11*, 358-369.
7. Gardner, A.; West, S.A. Cooperation and punishment, especially in humans. *Am. Nat.* **2004**, *164*, 753-764.
8. Bochet, O.; Page, T.; Putterman, L. Communication and punishment in voluntary contribution experiments. *J. Econ. Behav. Organ.* **2006**, *60*, 11-26.
9. Gürerk, O.; Irlenbusch, B.; Rockenbach, B. The competitive advantage of sanctioning institutions. *Science* **2006**, *312*, 108-111.
10. Masclet, D.; Noussair, C.; Tucker, S.; Villeval, M. Monetary and non-monetary punishment in the voluntary contributions mechanism. *Am. Econ. Rev.* **2003**, *93*, 366-380.
11. Gächter, S.; Renner, E.; Sefton, M. The long-run benefits of punishment. *Science* **2008**, *322*, 1510.
12. Casari, M.; Luini, L. Cooperation under alternative punishment institutions: An experiment. *J. Econ. Behav. Organ.* **2009**, *71*, 273-282.
13. Herrmann, B.; Thöni, C.; Gächter, S. Antisocial punishment across societies. *Science* **2008**, *319*, 1362-1367.
14. Rand, D.; Dreber, A.; Ellingsen, T.; Fudenburg, D.; Nowak, M. Positive interactions promote public cooperation. *Science* **2009**, *325*, 1272-1275.
15. Boyd, R.; Gintis, H.; Bowles, S. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science* **2010**, *328*, 617-620.
16. Carpenter, J.P. Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Game. Econ. Behav.* **2007**, *60*, 31-51.
17. Casari, M. On the design of peer punishment experiments. *Exp. Econ.* **2005**, *8*, 107-115.

18. Anderson, C.; Putterman, L. Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *J. Econ. Behav. Organ.* **2006**, *54*, 1-24.
19. Andreoni, J.; Harbaugh, W.; Vesterlund, L. The carrot or the stick: Rewards, punishments, and cooperation. *Am. Econ. Rev.* **2003**, *93*, 893-903.
20. Ostrom, E.; Walker, J.; Gardner, R. Covenants with and without a sword: Self-governance is possible. *Am. Polit. Sci. Rev.* **1992**, *86*, 404-417.
21. Isaac, R.M.; Walker, J.M.; Thomas, S.H. Divergent evidence on free riding: An experimental examination of possible explanations. *Public Choice* **1984**, *43*, 113-149.
22. Isaac, R.M.; Walker, J.M. Group size effects in public goods provision: The voluntary contributions mechanism. *Q. J. Econ.* **1988**, *103*, 179-199.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).