



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Games and Economic Behavior 54 (2006) 1–24

**GAMES and
Economic
Behavior**

www.elsevier.com/locate/geb

Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism

Christopher M. Anderson^a, Louis Putterman^{b,*}

^a *Department of Environmental and Natural Resource Economics, University of Rhode Island, RI, USA*

^b *Department of Economics, Box B, Brown University, Providence, RI 02912, USA*

Received 23 June 2003

Available online 10 December 2004

Abstract

The prospect of receiving a monetary sanction for free riding has been shown to increase contributions to public goods. We ask whether the impulse to punish is unresponsive to the cost to the punisher, or whether, like other preferences, it interacts with prices to generate a conventional demand curve. In a series of experiments, we randomly vary the cost of reducing the earnings of other group members following voluntary contribution decisions. In our design, new groups are formed after each interaction and no subject faces any other more than once, so there is no strategic reason to punish. We nonetheless find significant levels of punishment, and we learn that both price and the extent to which the recipient's contribution is below the group mean are significant determinants of the quantity of punishment demanded. Moreover, punishment is mainly directed at free riders even when it costs nothing to the punisher.

© 2004 Elsevier Inc. All rights reserved.

JEL classification: C91; H41; D71

Keywords: Public goods; Collective action; Experiment; Punishment; Demand

* Corresponding author.

E-mail address: Louis_Putterman@brown.edu (L. Putterman).

1. Introduction

The voluntary contribution mechanism is a linear n -person dilemma game that experimentalists use to study, in the laboratory, problems of collective action such as the voluntary provision of a public good. Subjects are put in groups and asked to divide endowments between group and personal accounts. Although the socially efficient outcome is attained when all contribute their full endowment to the group account, individuals maximize their payoff by retaining their endowments, regardless of what others do. In trials, subjects typically contribute an average of over 50% of their endowments in one-shot play or in the initial round of repeated play. In repeated play, contributions then tend to decay toward zero (Ledyard, 1995; Davis and Holt, 1993).

Recently, a number of experiments have been inspired by Fehr and Gächter's (2000a) demonstration that this decay in contributions can be attenuated by allowing subjects to direct costly monetary punishments at other group members after learning of their contributions. The qualitative findings that many subjects engage in costly punishment, that it is aimed mainly at low contributors, and that contributions accordingly stabilize or rise with repetition, has been confirmed by a number of studies including Bochet et al. (in press), Carpenter and Matthews (2002), Falk et al. (2001), Fehr and Gächter (2002), Masclet et al. (2003), Page et al. (in press), and Sefton et al. (2002). Fehr and Gächter (2000b) interpret the evidence as suggesting that substantial numbers of individuals have a propensity to punish free riding, perhaps because it violates a norm of, or a predisposition towards, reciprocity.

Most of these experimental designs make it possible for subjects to interact more than once, so punishment could be motivated by the desire to induce higher contributions from others in later periods. However, there is evidence that at least some costly punishment is nonstrategic. For example, many subjects pay to punish free riders in the last period of finitely repeated interactions. And there is a considerable amount of punishment in the perfect stranger treatment of Fehr and Gächter (2000a) in which no subject encounters another more than once.¹

Such punishment without apparent strategic justification poses a challenge for rational models. When A contributes more to the public good than does B , A earns less than B in the contribution stage of their interaction. Why would A further lower her earnings by incurring a cost to punish B ? One possibility is that punishment could result from anger, or another emotion. Fehr and Gächter (2002) refer to anger when they conclude that "negative emotions towards defectors are the proximate mechanism toward . . . punishment." Although the presence of anger need not rule out systematic behavioral rules—Fehr and Gächter's negative reciprocity or altruistic punishment concept is in fact amenable to such description, as we discuss later—if taken to its limit, the emotion approach might suggest behavior which is simply not amenable to rational analysis.

An alternative explanation for non-strategic punishment is that the punisher may have other-regarding preferences, in the presence of which punishment is utility maximiz-

¹ Falk et al. (2001) analyze experiments designed to determine the share of punishment that is motivated by own payoff maximizing strategic considerations, and conclude that it constitutes a very small part of that observed even in treatments permitting repeat interaction.

ing. One model of other-regarding preferences which has received particular attention in the punishment context is Fehr and Schmidt's (1999) formulation of biased inequality aversion. In their model, an individual's utility may be raised by reducing inequalities, especially those disadvantageous to her, even if her own income falls in the process. The model predicts that high contributors will punish low ones if the former are sufficiently inequality averse and the cost to the person punished exceeds that to the punisher by a sufficient amount.²

A key difference between social preference models and the extreme form of the emotional reaction model is the extent to which punishers will adjust their demand in response to varying punishment prices. In a social preference model, imposing a loss on *B* may give *A* satisfaction, and as with other tastes, the degree of satisfaction anticipated can be reasonably weighed against the cost of acquiring it to determine the amount that should be purchased to maximize self-assessed well-being. Inequality aversion models, such as Fehr–Schmidt, predict a downward sloping demand curve for punishing free riders—as the ratio of costs to punisher and punished falls, punishing becomes utility maximizing over a broader range of inequality aversion levels—just as there is a downward sloping demand curve for consuming boysenberry jam or hot fudge sundaes. The desire to punish free riding can then be thought of as a conventional preference.³ If the irrational anger view is pushed to its extreme, however, economic calculation might not be possible at all, and the observer would have no reason to expect any systematic relationship between the cost of punishment and the amount that is meted out.

We report the results of an experiment to test whether the demand for punishment displays the usual downward slope with respect to price, and thus to test indirectly whether a rational choice-with-social-tastes characterization, or an irrational anger description, is more accurate. To study the demand for punishment, we asked subjects to make contribution and punishment decisions at several prices.⁴ To rule out a strategic incentive to punish, we used a perfect stranger design. Our experiment provides further evidence that non-strategic punishment of free riding is common, and it demonstrates that the demand for punishment does obey the Law of Demand—i.e., the quantity purchased is a decreasing function of the price. This result is robust to controlling for the degree of the target's free riding, and to latent heterogeneity in the propensity to punish free riders. Our price-varying design also sheds light on some related issues. Based on our results, we extend to the case of zero-cost punishment the finding that individuals use punishment mainly to reduce the earnings of free riders. Although our results run counter to a depiction of punishment as irrational, we demonstrate partial violations of the Fehr and Schmidt (1999) inequality-aversion explanation of punishment by showing that punishment occurs even at

² Other models which incorporate other-regarding preferences include Bolton and Ockenfels (2000), Charness and Rabin (2002), Cox and Friedman (2002), Falk and Fischbacher (1998); for a critical survey, see Fehr and Schmidt, 2002. We focus on Fehr–Schmidt because they have received the most attention in the punishment literature.

³ This would parallel the conclusion of Andreoni and Miller (2002) that altruistic behavior is consistent with rationality.

⁴ Preferences for fairness and altruism have also been studied by varying prices in other kinds of experiments; see for instance Eckel and Grossman (1996) and Andreoni and Miller (2002).

higher cost to the punisher than to the target, and also that a surprisingly large share of punishment events involve low contributors punishing high ones.

Concurrent with our research, similar research was independently undertaken by Carpenter (2002). Carpenter also finds that the demand for punishment is decreasing in its price. His experimental design differs from ours in various ways, including

- (a) use of a stranger, but not a perfect stranger, protocol,
- (b) all subjects in a treatment faced the same set of prices in the same preannounced order, which could thus affect contribution levels,
- (c) price-penalty ratios which range from 0.25 to four, with much less emphasis on ratios less than one, and
- (d) no zero price.

Carpenter's analysis complements ours by isolating income and wealth effects at the set of higher prices designed to generate income effects.

2. Experimental design

To test the extent to which monetary factors are considered in determining the quantity of punishment demanded, our experimental design places subjects in a series of five one-shot voluntary contribution games with two stages. In the first stage, subjects are anonymously grouped and choose their level of contribution to the public good. In the second stage, each subject learns the contribution levels of other members of her group and her price for each dollar she chooses to reduce the earnings of other group members. Based on this information, the subject chooses the amount she wishes to reduce each other group member's earnings. Her own earnings are reduced by the cost of the reductions she purchases, and by the total amount of reductions purchased for her by the other group members. After each learns her payoff, the subject is assigned to a completely new group, containing no subject she faced in any previous round.

After extra subjects were dismissed (and paid their \$5 appearance fee), 18 subjects were seated in the laboratory, each at a private computer. The experiment moderator read the instructions (in Appendix A) out loud as subjects read along. The instructions were divided into two parts, one for each stage of the game. After each part, subjects calculated the payoffs given hypothetical practice problems to ensure they understood how decisions mapped to payoffs.

The first stage of the game is a standard voluntary contribution mechanism. Subjects were endowed ten experimental dollars at the beginning of each period. In the first stage, each makes an allocation, x_i , to her group's account, and $(10 - x_i)$ to her own personal account. For each member of the three member group, the group account returns $0.6 \sum_{i=1}^3 x_i$, or 0.6 times the total assigned to the group account. Subjects earned this fraction of the group account allocation regardless of how much they themselves contributed to the group account. Therefore, each subject's earnings from the first stage of the game is $\pi_i = (10 - x_i) + 0.6 \sum_{i=1}^3 x_i$. Given this payoff function, it is dominant strategy to set $x_i = 0$ since the private return on a dollar in the personal account is one dollar, but only

sixty cents in the group account. Subjects indicated their decisions by typing an integer allocation to the group account, x_i into an input box in the experimental software interface. Once all group members made their first stage allocation, the second stage of the game began with each subject being shown how much each other member of the group allocated to the group account. Each was also shown, her (but not other members') private per-unit punishment cost, p_i . In each period, p_i could take on one of five possible values. The five values were 0, 30, 60, 90 and 120 (experimental) cents in treatment 1; 0, 5, 10, 20 and 30 cents in treatment 2; and 30, 40, 50, 60 and 70 cents in treatment 3. Each subject faced each of the five values in a random order throughout the five periods, and subjects' orderings were determined independently, so subjects could infer nothing about the prices of others in their group based on p_i . Subjects were told only that their p_i was chosen randomly from among the five values, and that other subjects' prices were equally likely to be each of the five values in any given period. We assigned punishment costs in this way so that those costs could affect the choice of punishment only directly, and not also by influencing the choice of contribution in a given period.⁵ Three different sets of punishment costs were used to see whether the same or different qualitative results would hold over different ranges of values.

With the contribution and price information presented, subjects made two decisions, one level of punishment (possibly zero) for each of the other two members of their group. Each subject i indicated her choice by typing an integer number of units of punishment, r_{ij} , into a box for each other subject, j , in the group. Subjects could choose any number of punishment units so long as

- (a) the punishment given to any one subject was not greater than that subject's earnings for the period, and
- (b) the total cost of the units purchased did not exceed π_i , the amount earned by the punisher in the first stage of the game in the same period.⁶

Once all group members made their second stage punishment decisions, each subject learned the aggregate amount he or she was punished by other members of the group, and his or her final earnings for the period. Subject i 's earnings for the period are

$$\text{Earnings}_i = (10 - x_i) + 0.6 \sum_{j=1}^3 x_j - p_i \sum_{j \neq i} r_{ij} - \sum_{j \neq i} r_{ji}.$$

The first two terms represent i 's earnings in the voluntary contribution mechanism stage of the game. The third term is the amount i spent punishing other members of her group. This is where the price of punishment, in experimental dollars cost to punisher per experimental

⁵ By having group members' costs be uncorrelated with each other, we eliminated incentives to try to coordinate on punishment, for example a high-cost subject could not assume that the group would have a low cost punisher who would be better placed to carry the burden of punishing a free rider. Subjects were also not told that they were certain to face each cost some time during the experiment; hence they could not be sure what cost they would face next, even in the last period.

⁶ We imposed these budget conditions in each period, rather than over the experiment as a whole, to keep the periods as independent of one another, and thus as much one-shot in character, as possible.

dollar cost to person punished, appears: p_i . We say that the Law of Demand holds for punishment if r_{ij} is a negative function of p_i , all else being equal. The fourth term is the amount i was punished by the other members of her group. Money that is “punished away” is returned to the research fund.⁷ If this equation yields a negative number, i ’s earnings for the period are set to zero.

Subjects participated in this two-stage game five times. After each game, subjects were reassigned to completely new groups, such that they never faced the same subject twice. This matching protocol ensures there is no strategic benefit to punishment: altering others’ behavior will not benefit a subject, for she will never encounter that opponent again, with certainty.⁸ Thus, as in Fehr and Gächter (2000a), there is a straightforward prediction from standard economic theory assuming payoff-maximizing individuals: no one will pay to punish another subject, and for this reason the standard prediction that no one will contribute to the group account still holds for the first stage.⁹

After the five rounds, subjects answered a debriefing questionnaire with basic demographic information and a narrative description of their strategy in the game. Subjects’ earnings were converted from experimental dollars to US dollars at a rate of 1 experimental dollar = \$0.25, \$5 was added as a flat fee for participation, and subjects were paid privately. Earnings averaged \$20.61, with a standard deviation of \$3.58, for sessions which lasted about 75 minutes.

Three sessions of each treatment were run in the Policy Simulation Laboratory at the University of Rhode Island, and one session of each treatment was run in a computer classroom at Brown University. Subjects were recruited from e-mail lists of students who had expressed interest in participating in experiments, from courses, and from on-campus advertisements. Subjects did not necessarily have any training in economics, and, with the exception of a few subjects at Brown, none had experience with similar experiments.¹⁰

⁷ The structure of the contribution and punishment stages is identical to that used by Bochet et al. (in press) and Page et al. (in press) except that in their experiments, groups consist of four subjects, the MPCR is set to 0.4 rather than 0.6, and the punishment cost is uniformly 0.25. Having a fixed cost of taking one monetary unit from the person targeted differs from Fehr and Gächter’s (2000a), who use a rising marginal cost of deducting fixed fractions of the targeted person’s earnings. A fixed cost per monetary unit deducted is used by Fehr and Gächter (2002) and Sefton et al. (2002).

⁸ During periods 1 through 4, one cannot rule out altruistic motivation to help other participants in later periods. However, total punishment in round 5 is not apparently different from that in rounds 1 through 4, and the coefficient on a round 5 indicator variable in a random effects tobit regression, like that in Table 1, is insignificant, indicating that last-round punishment is indistinct from the first four rounds, controlling for punishment price and target contributions.

⁹ The matter is only slightly more complicated for those treatments that include a zero price of punishment. As will be discussed in Result 5 (Section 3.3), at a zero price theory predicts no correlation between the amount of punishment one would receive and the amount one has contributed to the group account. For this reason, the possibility of receiving punishment should also not influence decisions in the contribution stage, and the standard prediction of no contributions continues to hold.

¹⁰ The few exceptions, at Brown, had participated in 10- or 20-period partner-group experiments with a single fixed cost of punishment.

3. Results

Our discussion of experimental results is organized as follows. We begin by discussing two findings that replicate earlier results in the VCM-with-punishment literature. We then turn to our core findings on the sensitivity of punishment to its price. Third, we present two results that extend our knowledge of why people punish and of who punishes whom. Finally, we look at the effects of the price of punishment on the levels of efficiency (earnings) achieved.

3.1. Contributions and punishment: replication of previous results

Each of our 216 subjects made five contribution decisions, for a total of 1080 observations. The average contribution in each round is shown in Fig. 1, broken down by treatment. Initial contributions average about 70% in all treatments, and, consistent with previous experiments of VCM with punishment, decrease only slightly with repetition, in qualitative contrast to voluntary contribution mechanisms without punishment.¹¹

Result 1. Subjects made use of the punishment opportunity, even when it was costly.

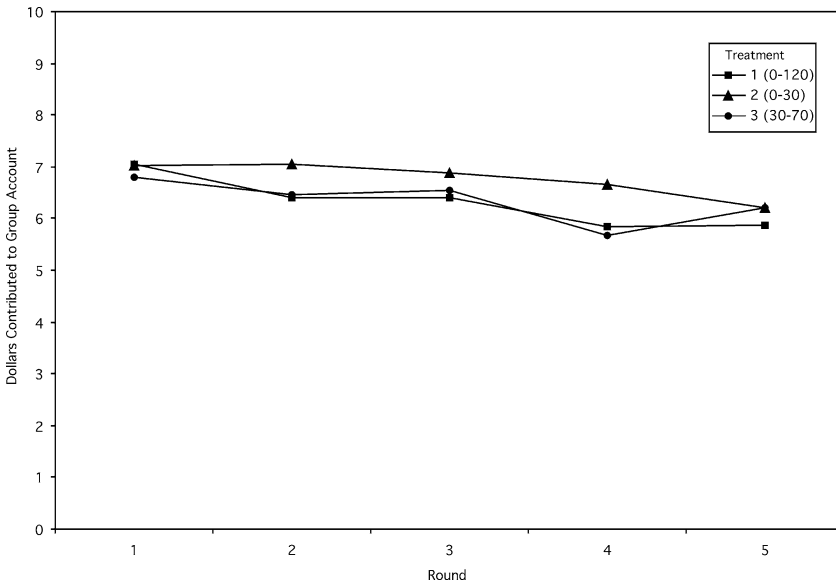


Fig. 1. Average contributions to the group account in each round, by treatment.

¹¹ To study the slopes of the contribution trends over time, we estimated the model $C_i = \sum_{j=1}^3 \{a_j T_{ji} + b_j T_{ji} \text{Period}_i\}$, where C_i is i 's contribution, T_{ji} is an indicator variable which equals one if observation i is from treatment j and zero otherwise, Period_i is the period number from which observation i was observed, and a_j and b_j are model parameters. A random effects regression indicates that all a_j and b_j are significantly different from zero. This is consistent with patterns observed in the VCM with punishment by Page et al. (in press), but different from those observed by Fehr and Gächter (2000a,2002) who find contributions to rise with repetition.

In each period, each subject made two punishment purchase decisions, one for each other member of the group, for a total of 2160 observations. Of these observations, 671 (31.1%) represented a positive amount of punishment by 161 different subjects, 74.5% of the subject pool as a whole. In the 1872 observations in which punishment was costly, 147 different subjects (68.1% of the pool) paid from their own earnings to reduce others' earnings 510 times (27.2% of overall opportunities). Average punishment per period showed no tendency to change over time. Therefore, despite a design which ensures that there is no strategic benefit to punishment, positive levels of punishment are still observed at positive prices, replicating previous experimental results.

Result 2. The demand for punishment is increasing in the recipient's level of free riding.

Figure 2 shows the average level of punishment associated with each level of negative and positive deviation of own contribution from the average by others in one's group. The figure suggests that punishment increases as contribution falls further below the average contributed by others in one's group.

The tobit result in Table 1 provides statistical support for this result.¹² The dependent variable is the amount of punishment subject i purchased for other group members j , and absolute negative mean deviation measures the difference between j 's contribution and the

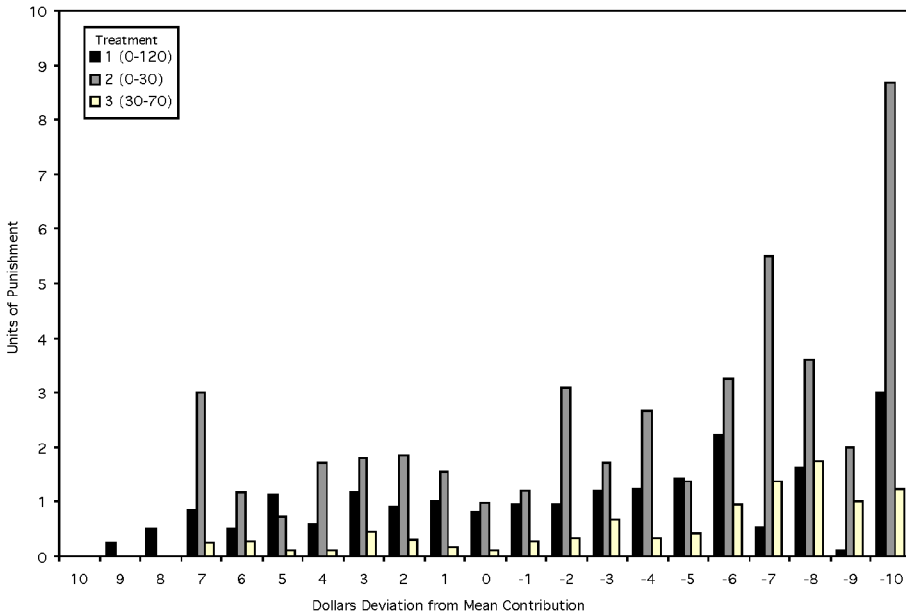


Fig. 2. Average units of punishment given by deviations from the mean of other group members and treatment.

¹² The tobit specification is used because the dependent variable, the amount of punishment, is censored at zero. Random effects are included in the model because our data includes ten observations on each of 216 subjects; within subject, these observations are not independent.

Table 1
Determinants of punishment

Independent variables	Units of punishment Random effects tobit
Average constant ^a	–3.242 NA
Absolute negative mean deviation	0.532 ^{***} (0.059)
Absolute positive mean deviation	0.078 (0.072)
Price of punishment	–0.027 ^{***} (0.005)
Price = 0 dummy	2.547 ^{***} (0.414)
Price = 0 × Abs. neg. mean dev.	0.435 ^{***} (0.125)
σ_ε	3.749
LL	–2456.79
Wald (DOF)	422.00 (16) ^{***}

Random effects tobit model of the determinants of purchasing punishment. Standard errors are in parentheses.

^a A separate constant is estimated for each session. The reported average constant is the average of these constants; all are negative.

* Significance levels noted at $p < 0.10$.

** Idem., $p < 0.05$.

*** Idem., $p < 0.01$.

average contribution of j 's fellow team members (and is set to zero if j 's contribution is above that average), and conversely for absolute positive mean deviation.¹³ As in Fehr and Gächter (2000a), the coefficient on absolute negative deviation is highly positive, indicating that lower than average contributors receive more punishment. An increase in free riding from an absolute negative deviation of two to four dollars increases expected punishment from each other group member from 0.45 units to 0.73 units at a price of 30 cents. At a price of 10 cents, one unit of punishment is expected at an absolute negative deviation of 5.50 dollars, but this decreases to half a unit of punishment at an undercontribution of 2.70 dollars. Carpenter (2002), Falk et al. (2001), Fehr and Gächter (2002), Sefton et al. (2002), and Bochet et al. (in press) also report punishment significantly increasing with the deviation of contribution below the average. As in Fehr and Gächter (2000a), the coefficient on absolute positive deviation is insignificant, indicating that once one's contribution is above the average, further increases in contribution do not significantly reduce expected punishment.

¹³ Proper estimation of the relationship between contribution level and punishment requires simultaneously controlling for other variables, especially price. We delay discussion of additional findings to Results 3 and 5.

3.2. The effect of price on the demand for punishment

Results 1 and 2 establish that our experimental design replicates key results of previous experiments. The next two results establish the previously uncharacterized nature of the relationship between punishment and its price.

Result 3. The level of punishment is decreasing in price, even after controlling for the significant effect of the level of free riding.

The effect of punishment price and treatment on the level of punishment can be seen in Fig. 3, in which each bar indicates the average amount of punishment each person purchased for each other member of the group in each price treatment condition. Across treatments, there is a trend toward smaller quantities of punishment at higher prices. At a price of zero, an average of over three units of punishment are purchased for each other group member. For low, positive prices this demand is just under two units (1.95 at five cents, 1.62 at ten cents). As the price rises to 60% of the value of the punishment purchased, demand falls to 0.49 units per period. At the highest price, \$1.20 or 120% of cost to the subject targeted, demand is only 0.38 units for each other group member.

The decrease in punishment across the full range of prices is also apparent on a treatment-by-treatment basis. Treatment 1, represented by black bars at prices 0, 30, 60, 90 and 120 cents shows a strong decreasing trend, with the quantity of punishment demanded at each price lower than at the next highest price. Treatment 2, represented by gray bars at prices 0, 5, 10, 20 and 30 cents, shows a similar pattern of decreasing quantity de-

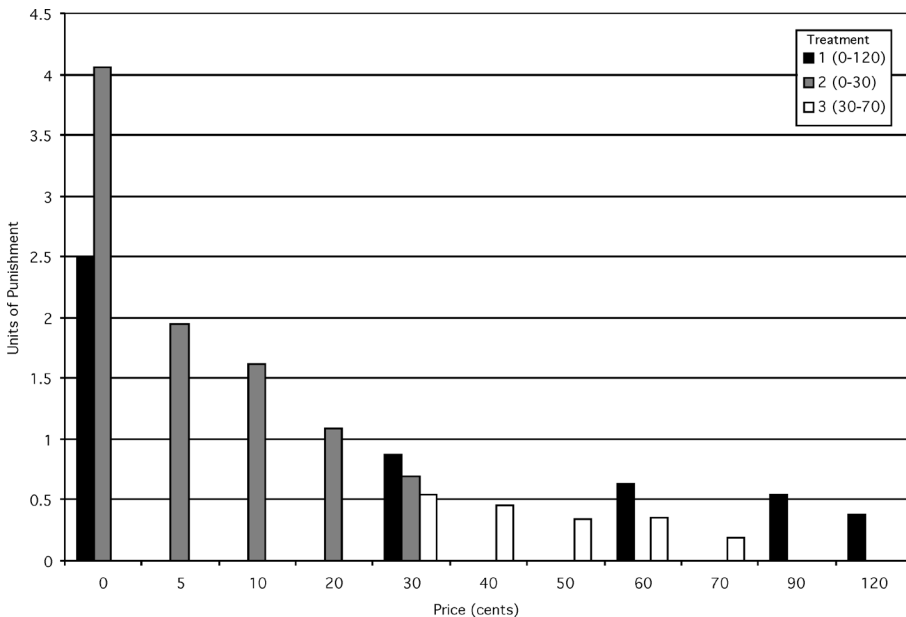


Fig. 3. Average number of units of punishment demanded at each punishment price, by treatment.

manded at each price. Treatment 3, represented by white bars at prices 30, 40, 50, 60 and 70 cents, does not exhibit the same strong monotonicity as there is no decrease between prices of 50 and 60 cents, but there is still an overall trend of reduction.

The random effects tobit regression in Table 1 also establishes the statistical significance of the relationship between punishment and price. The price to punisher i of reducing group member j 's earnings by one experimental dollar has a highly significantly negative coefficient, indicating that r_{ij} is a negative function of p_i , consistent with the Law of Demand. However, the coefficient on the price variable is not only statistically significant, but also its magnitude indicates that the impact of price on punishment is economically significant. For example, an increase in price from forty cents to ninety cents will decrease the punishment that can be expected from each other group member by a target who has undercontributed by two dollars from 0.40 units to 0.20 units. At an undercontribution of \$5.20, one unit of punishment is expected at a price of 5 cents, but the expected punishment drops to half that at a price of 60 cents. The regression also includes an indicator variable for observations at a price of zero. The estimate shows a highly significant positive coefficient on this term, consistent with the general inverse relationship between punishment and price but suggesting that punishment increases between the lowest positive price and the zero price by more than would be suggested by a linear extrapolation. When the price falls to zero, the expected punishment increases by about 0.78 units when the recipient's contribution is the same as the average; increases are larger at higher levels of free riding.

Even though there is less punishment at higher prices, Fig. 3 shows that a positive amount of punishment is observed even when it costs more to the punisher than to the recipient of the punishment. In treatment 1, each subject had one period with a price of \$1.20 to punish another group member by subtracting one dollar from his or her earnings. The average amount of punishment in this condition is 0.38 units, representing a positive amount of punishment in 23 of 144 decisions by 14 of 72 subjects. The demand for punishment even when it costs the punisher more than the person punished suggests that the desire to punish can be quite strong, although no obvious pattern is discernible in the punishments at this price.¹⁴

The significant interdependence of price and free riding shown in Results 3 and 4 is summarized graphically in Fig. 4. The experimental data, represented in Fig. 4a, shows that the most punishment, indicated by the tallest bars, occurs at low prices and high levels of free riding. The expected level of punishment predicted by the tobit model in the second column of Table 1, as a function of price and relative contribution, is shown in Fig. 4b. In general, punishment increases as free riding increases at each price, and punishment decreases as prices increase at all levels of free riding.

¹⁴ In particular, there is no clear tendency to single out free riders: in 7 of the 23 cases, the person punished contributed less than the punisher, in another 10, the same amount, and in the remaining 6, more. Sorting the results by comparison to the average of others' contributions, instead, we find that in 6 of the 23 cases, the person punished had contributed less than the others' average; in another 7, the same as the others' average. A tobit regression with the same specification as in the second column of Table 1 but using only the 144 $p = \$1.20$ cases and dropping the price terms generates significant coefficients on neither the positive nor the negative deviation terms.

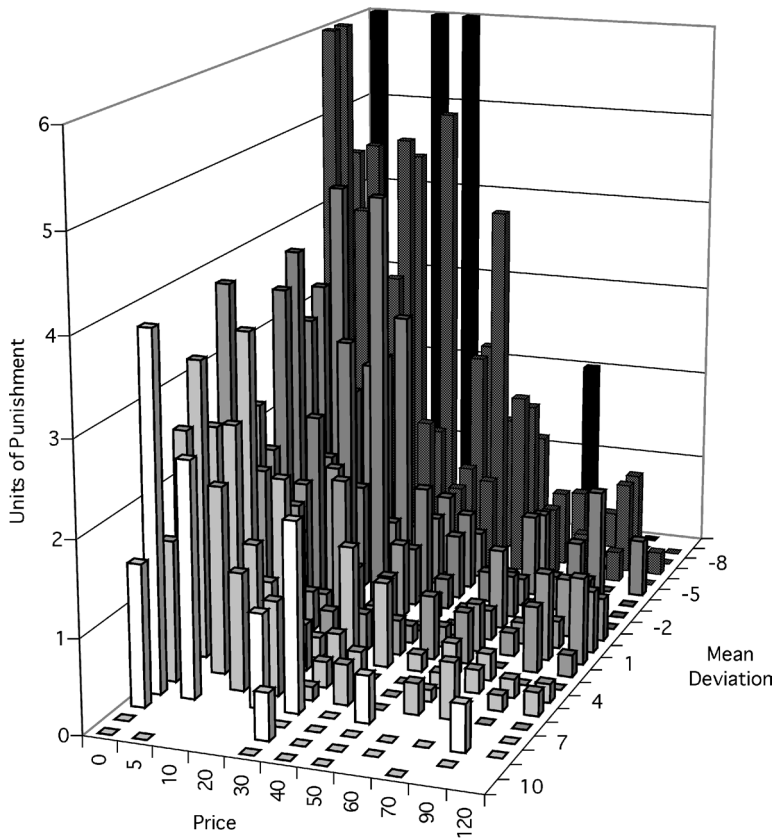


Fig. 4a. Average amount of punishment demanded as a function of the target's deviation from the mean contribution of the other two group members and punishment price.

While the population as a whole is responsive to price and negative deviation of contribution, post-experiment debriefing questionnaires suggest that there was considerable heterogeneity in subjects' punishment strategies. Articulated strategies ranged from people who punished whenever someone contributed below a certain amount or who punished only when the price was zero to people who provided a straightforward argument that punishment did not benefit them, and therefore was irrational. Given this systematic variation, inference based on aggregate data may not fully capture the effect of price variation on punishment.

To investigate subject heterogeneity with regard to punishment, we apply El-Gamal and Grether's (1995, 2000) estimation-classification algorithm to estimate model parameters for different groups of subjects. The estimation-classification algorithm assumes the choices of each person in the sample are described by a function $F(\theta)$, where θ is a vector of unknown model parameters. Heterogeneity is introduced by allowing that the population contains K segments, or "types" of person, with each type described by one of K different θ s. The θ_k s which describe each type and which subjects are which type are es-

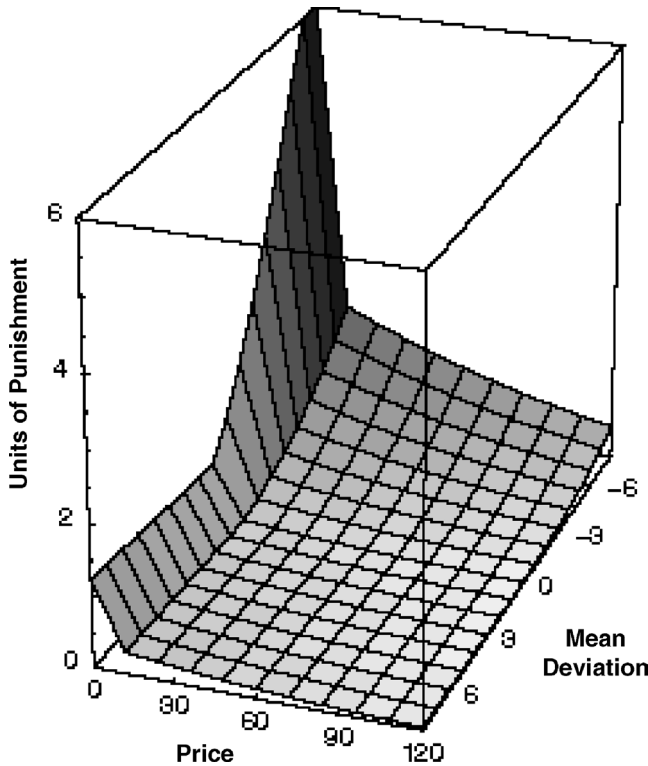


Fig. 4b. Predicted punishment as a function of absolute deviation from the mean contribution of the other two group members and punishment price based on the model of Table 1.

timated simultaneously. In estimation, person i 's contribution to the likelihood function, given $\Theta = (\theta_1, \dots, \theta_K)$, is the maximum of the joint likelihood of all i 's observations across the K types.¹⁵ Conventional maximization algorithms can be used to identify the Θ which maximizes the likelihood of the observed data, with care taken to ensure the global maximum is identified in a likelihood function which often has many local maxima.

Result 4. Sensitivity to punishment price and free riding varies considerably within our sample, but price is significantly negative for all identified groups.

¹⁵ The log-likelihood function is

$$\ln L(Y; X | \Theta, K) = \sum_i \arg \max_k \left\{ \sum_t \sum_j \ln L(Y_{ijt}; X_{ijt} | \theta_k) \right\}$$

where Y are the observed punishment choices, X the independent variables, and $\Theta = (\theta_1, \dots, \theta_K)$, the parameter vectors for each of the K types. The log-likelihood function is the sum over all individuals of the best log-likelihood of the joint probability of their observed choices, levels of punishment for each other group member j and period t , across the k possible types.

Support comes from Table 2, which reports the two- and three-segment estimation-classification tobit models (single-segment results are in Table 1).¹⁶ Each column of the table reports the estimated coefficients, the proportion of the sample best described by those coefficients, the proportion of punishment decisions which resulted in punishment, and the average amount of punishment that those in each segment assign to each other member of their group. The session-specific constants shared among all segments are suppressed.

The two-segment model divides the population into high punishers and low punishers. Both groups respond significantly to the price of punishment, but their demand curves have different slopes. Low punishers, 69.0% of the sample, mete out 0.46 units of punishment at a price of 30 cents in response to a deviation of 2 dollars; high punishers purchase 1.59 units under the same circumstances.

Table 2
Heterogeneous tobit estimates of determinants of punishment

	Two-segment		Three-segment		
	I	II	I	II	III
Proportion of subjects	69.0%	31.0%	30.6%	47.2%	22.2%
Punishment frequency	18%	60%	4%	32%	67%
Average punishment	0.76	1.79	0.24	1.19	2.00
Average contribution	6.62	6.16	5.92	7.06	5.97
Average constant	-6.237	1.226	-9.764	-2.955	1.904
	NA	NA	NA	NA	NA
Abs. neg. deviation	0.983*** (0.114)	0.118* (0.062)	0.565 (0.578)	0.859*** (0.083)	0.012 (0.066)
Abs. pos. deviation	0.063 (0.146)	-0.058 (0.070)	-1.659** (0.812)	0.261** (0.109)	-0.046 (0.068)
Price of punishment	-0.076*** (0.011)	-0.016*** (0.005)	-0.483*** (0.158)	-0.054*** (0.008)	-0.016*** (0.004)
Price = 0 indicator	3.587** (0.662)	1.490** (0.424)	6.110* (3.707)	2.652** (0.578)	1.042** (0.333)
Price = 0 × Abs. neg. dev.	-0.108 (0.193)	0.523*** (0.112)	-188.433 NA	0.677** (0.185)	0.588*** (0.140)
σ_ε	6.166	2.560	11.134	4.349	2.396
LL		-2310.12		-2186.63	

Two- and three-segment tobit models of the determinants of purchasing punishment when the price is zero. Standard errors are in parentheses. Session-specific constants common to all segments are adjusted by segment-specific constants and reported as average constants for each segment to reflect differences in the propensity to punish among segments.

* Significance levels noted at $p < 0.10$.

** Idem., $p < 0.05$.

*** Idem., $p < 0.01$.

¹⁶ Maximization was carried out using Gauss. To maximize the chance we identified the global maximum among many local maxima, we repeated the maximization from one thousand starting points randomly drawn from a large neighborhood around a starting vector based on the single-segment parameters, modified to reflect patterns which emerged from debriefing questionnaires.

Adding the third segment to the model allows a clearer division between subjects who need significant provocation to punish and those who virtually never punish. Price remains significantly negative, with $p < 0.01$, in all three segments, though different segments are differently responsive to price. Segment I, 30.6% of the sample, represents only 26 decisions to punish, by 11 of the 66 subjects in that segment. Six of the 11 only punish at prices of zero or 5 cents, and only three choose to punish in more than one round.¹⁷

With non-punishers and near-nonpunishers estimated to be in their own segment, the punishers are divided into two segments. Segment II characterizes 47.2% of subjects who are responsive both to free riding and to the price of punishing, meting out 1.23 units of punishment for a two dollar undercontribution when the price is 30 cents. Segment III, the remaining 22.2% of subjects, are relatively aggressive and, from a normative standpoint, indiscriminant punishers. A deviation of two units is met with 2.83 units of punishment when the price is 30 cents, but given the small magnitude, and statistical insignificance, of the coefficient on negative deviation, the level of punishment varies little with the target's contribution behavior.

Taken together, these three segments indicate that a significant portion of the sample does not routinely punish, and almost never does so at positive prices. However, the remaining 70% of the sample do punish. About two-thirds of these subjects are primarily responsive to free riding at all prices, and the others punish aggressively at all levels of free riding. Importantly, both punishing groups are responsive to price.

3.3. *Why subjects punish one another*

In Results 3 and 4, we established that the demand for punishment obeys the Law of Demand. This conclusion is paradoxical, from the standpoint of the strongly rational model, because it implies there is a fundamentally rational element, price responsiveness, to a seemingly non-rational decision, that to incur cost without strategic benefit. Nonetheless, Result 2 confirms earlier findings that punishment is increasing in the targeted subjects' negative deviation from other group members' average contribution. This subsection extends Result 2 by shedding further light on the nature of the underlying preference or taste for punishment.

While standard economic theory predicts that no one should punish at positive prices, it makes no prediction when punishment is free. If individuals care only about their own earnings—that is, they have neither altruistic nor envious or spiteful preferences toward others, nor are they reciprocators or adherents to norms such as “each doing his fair share”—then they would be strictly indifferent about punishing at zero price. Since no extra effort is required to punish (all subjects being required in any case to enter some number, possibly 0, in the relevant space), we might expect to see at zero price a smear of data covering all possible punishment values, zero included, with no particular pattern. The next result shows that consumption of free punishment is not random, but rather related to the level of free riding relative to the group mean.

¹⁷ All punishment observations from Segment I occur when the zero price–absolute negative interaction is zero, leading to the enormous magnitude and undefined standard error on the zero price–absolute negative deviation interaction.

Result 5. Cost-free punishment is also an increasing function of the recipient's free riding.

This result is initially demonstrated by the highly significant positive coefficient on the interaction between the zero-price indicator and the absolute negative deviation in the random effects tobit regression of Table 1, which is also reflected in Figs. 4a and 4b. It is further illustrated by the regression of Table 3, which considers only the 288 observations at zero price. As in the whole sample, the coefficient on absolute negative deviation is positive and significant, indicating a preference for punishing free-riders. In this model, at zero price, an increase in free riding from two units to four units leads to an increase in expected punishment from 3.26 units to 4.58 units. Taken together, these results suggest that even when punishment is free, people have specific motives for choosing to punish: they dislike free riding, or have a sense of fairness, or similar other-regarding or normative preferences, and they will generally not punish if the target has not 'earned' the punishment, even though it costs them nothing.

While free riding is a significant motivation for punishment, the next result shows that it is not the only one, as a significant portion of observed punishment is of high contributors from low contributors.

Result 6. High contributors are more likely to punish low contributors than other high contributors. However, there are also many cases of low contributors punishing high contributors.

Figure 5 is a scatterplot of the combinations of punisher and target deviation from the group mean that led to positive punishment demands.¹⁸ In the lower right quadrant are

Table 3

Determinants of punishment when price = 0

Independent variables	Units of punishment Tobit
Average constant	−0.809 NA
Absolute negative deviation	1.033*** (0.223)
Absolute positive deviation	0.245 (0.241)
σ_ε	6.478 LL = −620.74 Wald(9) = 35.59***

Tobit model of the determinants of purchasing punishment when the price is zero. Standard errors are in parentheses. Session-specific constants are omitted.

* Significance levels noted at $p < 0.10$.

** Idem., $p < 0.05$.

*** Idem., $p < 0.01$.

¹⁸ In this plot, the group mean is the average contribution of all three group members, without excluding the target or punisher. This is done so the differences represented on both axes are with respect to the same number. Excluding the punisher from the punisher's mean deviation and the target from the target's mean deviation leads to qualitatively similar results.

240 cases of normal punishment, where someone contributing above the mean punishes someone contributing below the mean (there were 45 additional instances of people above the mean punishing people contributing at the mean). This represents a 37.2% punishment rate of low contributors by high contributors. In contrast, the upper right quadrant shows 76 cases of punishment of high contributors by other high contributors, a rate of 21.4%. A χ^2 test rejects the hypothesis that high contributors were equally likely to punish high and low contributors with $p < 10^{-7}$.

While this pattern of high contributors punishing free riders was expected and is consistent with other published results, we also observed low contributors punishing, and often punishing high contributors.¹⁹ In the lower left quadrant of Fig. 5 are 96 instances

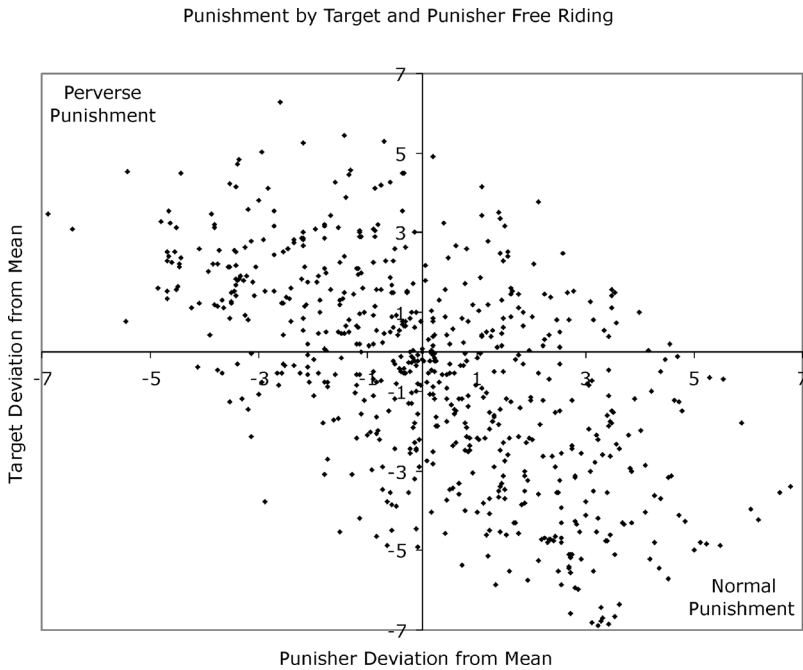


Fig. 5. Noisy scatterplot of combinations of the punisher’s and target’s deviation from the group mean contribution which resulted in positive amounts of punishment. A “noisy” scatterplot adds random noise to the actual values observed. Because there are a limited number of possible deviations from average in a three person VCM, multiple cases are observed at many combinations. The noise, in this case a random number on $[-0.5, 0.5]$, serves to spread out these cases so the distribution of cases is visually apparent.

¹⁹ Any effect of price and punishment quantities on these results can be controlled by adding, to the model in the second column of Table 1, (a) an interaction of the absolute negative deviation variable with an indicator for whether the punisher contributed more than the target, and (b) an interaction of the absolute positive deviation term with an indicator for whether the punisher contributed less than the target. A random effects tobit estimate results in coefficients on both of these interaction terms which are significantly positive ($p = 0.019$ for absolute negative deviation and $p = 0.063$ for absolute positive deviation), suggesting differential rates of punishment of high and low contributors depending on whether the punisher is a high or low contributor herself, consistent with the impression conveyed by the scatterplot. Even with these terms, model results are qualitatively similar to those in Table 1.

of low contributors punishing other low contributors, for a punishment rate of 36.9%, a rate statistically indistinguishable from the rate at which high contributors punish low contributors ($p = 0.991$). However, the upper left quadrant shows 194 cases of perverse punishment (a punishment rate of 30.0%), where someone contributing below the group mean punishes someone contributing above the group mean.²⁰ This rate is significantly higher ($p = 0.004$) than the rate at which high contributors punish other high contributors. This regularity in our data poses a significant challenge to models representing the motivation for punishing, as we observe people consistently paying to increase the inequality, albeit favorable, in earnings.

That high and low contributors punish low contributors at the same rate suggests that subjects who free ride in the contribution stage of the game do not also systematically free ride in the punishment stage, leaving others to incur the cost of punishing free riders. Although the average contributions in Table 2 for the two-segment model suggest the low punishers are also more likely to be high contributors, the three-segment model reveals there is little systematic relationship: subjects in Segment I, who rarely punish, and those in Segment III, who punish the most, have almost identical average contribution levels, while somewhat higher contributions are shown for those in between, in Segment II.

3.4. Punishment and efficiency

Although the availability of punishment has been found to stave off the sharp decline in contributions that typically occurs with repetition in VCM experiments, the impact on efficiency has been found to be more ambiguous, since higher contributions are induced at the expense of costs both to punishers and targets of punishment (Bochet et al., in press). It is thus of interest to see how differences in the costs of punishment affect efficiency in our experiment.

Result 7. Contributions tend to be higher but earnings lower in treatments in which the average available cost of punishment is lower, reflecting the trade-off between the incentive effect of the threat of punishment, and the toll on efficiency from punishments carried out.

Consider again Fig. 1, which shows the time path of average contributions in each of the three price treatments. Recall that subjects in a given treatment knew only the distribution of punishment costs, not the particular costs faced by other group members in given periods, so price realizations could not affect current contributions. With the set of possible costs varying among the three treatments, however, the price set could influence the average contribution in a treatment as a whole. Our results suggest that expected and actual

²⁰ Bochet et al. (in press) describe the punishment of high contributors as “perverse” because it tends to discourage contributions, thus working against efficiency. Calculations by Cinyabuguma et al. (2004) find that perverse punishment accounts for about 20% of all instances of punishment in the partner-group treatments of Fehr and Gächter (2000a), Bochet et al. (in press) and Page et al. (in press), but is less in Fehr and Gächter’s stranger and perfect stranger treatments. The higher incidence of perverse punishment in our own data than in the Fehr–Gächter’s perfect stranger treatment may be partly attributed to the fact that groups have only three members in the present experiment versus four in theirs and the others cited.

punishments, which vary based on price levels within treatment, had small but in some cases statistically significant effects on contributions and earnings.

Table 4 summarizes the components of contributions and earnings in each treatment. The highest average contributions were in the low punishment price treatment, and the lowest were in the high variance punishment price treatment, with the mid-price punishment treatment in the middle. Contributions in the VCM stage of the game are significantly higher in the 0–30 treatment than in the other two treatments, which are not statistically distinct from each other. Pairwise Wilcoxon signed-rank tests comparing the level of contributions in each treatment show the 0–30 treatment has weakly higher contributions than either the 0–120 or 30–70 treatment (with $p = 0.048$ and $p = 0.047$, respectively). The distribution of contributions in the 0–120 treatment is statistically indistinct from that in the 30–70 treatment, with Wilcoxon test $p = 0.98$.

Unlike with contributions, efficiency and earnings differ considerably across treatments. A Wilcoxon signed-rank test of earnings in each period rejects the hypothesis that period earnings in the 30–70 treatment have the same distribution as those in the 0–120 and 0–30 treatments ($p < 10^{-7}$ and $p < 10^{-11}$, respectively); the 0–120 treatment has slightly higher earnings than the 0–30 ($p = 0.04$). The treatment with the highest contributions also has the lowest total earnings because gains from higher contributions are punished away, illustrating the dilemma of punishment. Punishment is most frequent in the 0–30 treatment, where it is cheapest. It is least frequent in the 30–70 treatment, where there is no low-cost punishment available. In the 0–120 treatment, low cost punishment is available some, but not all, of the time. Even controlling for price, Fig. 2 shows there is less punishment in the 30–70 treatment than at comparable prices in the other treatments. However, a Wilcoxon signed-rank test of the data from the 0–120 and 30–70 treatments at the shared prices 30 and 60 does not reject the hypothesis of identical punishment distributions ($p = 0.28$).

4. Discussion and conclusion

This experiment gave subjects in small groups the opportunity to reduce one another's earnings after learning of one another's contributions to a public good. The game was played five times, each time in a completely new group and with a different price encountered in a randomly selected and not preannounced order. Consistent with other

Table 4
Average earnings, punishment and efficiency

	Treatment		
	0–120	0–30	30–70
Average contribution	6.31	6.76	6.34
Contribution stage earnings	15.06	15.41	15.07
Average punishment	0.99	1.88	0.38
Average punishment expenditure	0.32	0.14	0.17
Average total earnings ^a	12.44	11.36	13.97
Efficiency	69%	63%	78%

^a Although subjects' period earnings were given a lower bound of zero, the reported average earning statistics assume negative earnings were allowed. This fully captures the effect of punishment and punishment expenditure.

experiments, subjects engaged in costly punishment even though no strategic gain was possible. Also consistent with those experiments, contributions to the public good remained at relatively high levels despite repetition. The amount of punishment demanded was found to be significantly negatively related to its cost to the punisher, and significantly positively related to the amount of free riding by the person targeted. Punishment was purchased even when it cost the punisher more than the person punished, with the amount being less than at lower prices but showing no special break with the ordinary downward trend. Punishment was mainly used to punish free riding even when it cost nothing to the punisher.

Although we found that most punishment was aimed at low contributors by high ones, we noted that a surprising amount of punishment was also aimed at high contributors, and that this “perverse” punishment was mainly demanded by low ones. Prior to punishment, a higher contributor always earns more than a lower one, and, so long as punishment’s price to the punisher is less than its cost to the person punished, it always reduces (increases) inequality disadvantageous to the punisher when a higher (lower) contributor punishes a lower (higher) one. 51% of the punishment observations in our data at prices less than 100 were “inequality-redressing” cases of a higher contributor punishing a lower one, but 40% were “inequality-exacerbating” cases of a lower contributor punishing a higher one, and another 9% involved an individual punishing someone who had contributed and thus earned (prior to punishment) the same amount. The Fehr–Schmidt theory of inequality aversion, which much of the VCM-with-punishment literature has focused on as a possible explanation for the data, is consistent with the fact that our subjects punish less at higher prices and with a slight majority of our punishment observations, but there is clearly much in our data that it does not explain.²¹

Apart from reconfirming that subjects will punish free riders at cost to themselves without strategic gain and that the anticipation and experience of such punishment can help to sustain contributions to a public good, our experiment sheds light on the nature of the preferences that underlie such punishment. Most importantly, despite the fact that emotions may be the proximate cause of punishment, as argued by Fehr and Gächter and suggested by punishments at price:cost ratios of 1:1 or more,²² the responsiveness of punishment to price suggests that the phenomenon still behaves like a taste for a good being traded off against its money cost, which represents other foregone satisfactions. The most extreme

²¹ Notice that punishment of low contributors by high ones at a price of 120 is also not predicted by the Fehr–Schmidt theory, since it increases the inequality between the punisher and the person punished, although it is consistent with Fehr and Gächter’s idea of punishment as negative reciprocity. With regard to the punishment of high contributors by low ones, a referee suggested that punishment of higher contributors by lower ones might be consistent with inequality aversion if the low contributors were themselves anticipating being punished by the higher ones. Although this possibility cannot entirely be ruled out, punishment of low contributors by high ones left the low contributors with less earnings than the high in only 26% of such punishment cases in our data, so the belief that (simultaneous) retaliation would be needed to redress a disadvantageous inequality would usually have been mistaken. Of course, some subjects anticipating being punished may have wanted to punish as an anticipatory retaliation even knowing that punishing would not have been inequality reducing. Also, low contributors might have punished high ones to show disdain for “goody goodies,” to increase their *relative* earnings, or simply to amuse themselves.

²² Falk et al. (2001) find a fairly high incidence of cases in which a subject punishes free riding when the cost is equal to the punisher and the person punished.

emotive view, in which punishment is in no sense a rational decision, is decisively rejected by our data. Our subjects appear to obtain definite amounts of satisfaction from punishment, similar to that from other goods, rendering the utility gains amenable to rational comparison with cost to reach decisions on the number of units to purchase. In the standard terminology of economics, the desire to punish free riders is a taste, not rational in itself, but an input to the rational decision calculus that differs from a taste for goods primarily in its more obviously social character. While we have not established definitively why punishment brings utility, we confirmed that it is directed mainly, though by no means entirely, at free riders. In addition, like tastes for other goods, the taste for punishment varies considerably in the population.

Acknowledgments

The research reported here was funded by National Science Foundation Grant SES-0001769. We thank the Rhode Island Agricultural Experiment Station for support of the Policy Simulation Laboratory (AES contribution number 4052). We also thank Toby Page for suggestions regarding the design of the experiment, and Dmitri Lemmerman for adapting the computer software used to conduct it.

Appendix A. Experimental instructions

The instructions were read aloud by an experimenter as the students followed along on their computer screens.

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. Each experimental dollar is worth a real \$0.25 (twenty-five cents). The amount you earn will depend on your and others’ decisions. The maximum possible earning is \$32.50 (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.

Groups

There are eighteen participants in this experiment, and the experiment consists of five decision periods. During each decision period, you will be placed in a group with two other participants, hence a group of three. After every period, you will be put in a completely new group. You will not know the identities of the other two members of your group in any given decision period, nor will you be told their identities after the experiment is over.

Earnings

Each of the five periods in the experiment is structured as follows:

At the beginning of the period each member of 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 directly into your earnings.

An amount equal to 0.6 times the group's total assignment to the group account goes into your earnings:

$$\text{Earnings} = (\text{amount in personal account}) + (0.6)(\text{total in group account}).$$

Every group member earns 0.6 times the total in the group account, regardless of how much they personally put into that account.

The next four screens are designed to help you test your understanding of the experiment so far. Corresponding to each screen, there is a sheet of paper on your desk labeled Practice 1, Practice 2, etc. We will fill in the first paper worksheet together, then the first screen. Please note that these are practice exercises only, and do not affect your actual earnings in the experiment.

At this point, an experimenter walked subjects through computing their voluntary contribution game payoffs when all three members of the group contribute \$10 to the group account. The exercise is first completed on paper, and then using the experimental software. Following that, subjects computed, on their own using first paper and then the experimental software, their payoffs for the case where every member assigns \$0 to the group account, and for the case where the subject assigns \$5 to the group account, a second member assigns \$5 to the group account, and a third member assigns \$10 to the group account. Once these exercises were reviewed by an experimenter, the instructions resumed.

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 \$0.60, for a net reduction of \$0.40.

But each of the other people in your group would increase their earnings by \$0.60, for a total increase of \$1.20 for the others in your group.

Reductions

There is another decision that affects your earnings. Once 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, that you decide to reduce *B*'s earnings by \$3, and *C*'s earnings by \$5. The total amount of reductions you make on others' earnings is therefore \$8.

In most periods, it costs something to reduce another group member's earnings. The cost will vary randomly from one period to another during the experiment. There are five possible costs per dollar of reductions: [0, 5 cents, 10 cents, 20 cents, and 30 cents; varies with treatment]. Each group member is equally likely to be faced with any of these five costs in a given period. Each period, you will be told your own cost for imposing reductions that period, but you will not know the reduction costs facing the other two members of your group, which are likely to be different.

As an example, suppose that in this period it costs you 30 cents for each dollar by which you reduce others' earnings. Then in this example, it costs you $(0.30)(\$8) = \2.40 to impose the total of \$8 of reductions on *B* and *C*.

Just as you can reduce others' earnings, others can reduce yours. Suppose *B* reduces your earnings by \$1 and *C* reduces your earnings by \$3. The total reduction of your earnings by others is then

$(\$1 + \$3) = \$4$. You will learn that your earnings have been reduced by a total of \$4 but you will not learn who has reduced your earnings by what amount, how many group members have participated in reducing your earnings, or what it cost them to make these reductions.

Similarly none of the others will learn whether you as an individual have reduced their earnings, by how much, or at what cost, only the total reductions.

Practice sheet 4 takes you through an example of the reduction process. Let's fill out this sheet together, and then enter the information in the corresponding screen.

The experimenter then led subjects through computing their payoffs when they had chosen to punish others, and others had chosen to punish them. The exercise was carried out first on paper, then using the experimental software.

Your net earnings

Your net earnings for a period will be:

Amount in personal account + $(0.6)(\text{Total in group account}) -$

$(\text{Your cost of reducing others' earnings}) * (\text{Total of your reductions of others}) -$

Total of reductions of your earnings made by others.

In this formula, your cost of reducing another person's earnings is the one that is displayed on your screen for that period. If your earnings in a given period are ever negative, they will be reset to zero. Also, the maximum that you can spend on reductions in a given period is the amount of earnings you have for that period before the reduction process begins. Note that each period is independent of the others in that you begin each one with \$10 to allocate, and you can earn as little as zero and as much as \$22 (experimental dollars) in that period, depending on your decisions and those of the others in your group.

Remember that you will be told what your reduction cost is at the beginning of each period, but you will not learn what the reduction costs that face the others in your group are. (That is, knowing your own cost tells you nothing about their costs, since for each of them, any one of the five possible costs is equally likely in any given period.)

At the end of the experiment, the net earnings for the five periods will be totaled and converted from experimental dollars to real dollars. Then \$5 will be added for your participation.

Conclusion

It is important that you understand the procedures and formulas that will be used in the experiment. Once the experiment begins, there will be no further discussion, and no communication of any kind among the participants is permitted, apart from the transmission of the decisions by computer. So please review the experiment in your mind now and raise your hand if you have any questions. Once all questions have been answered, the experiment will begin.

References

Andreoni, J., Miller, J., 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* 70, 737–753.

- Bochet, O., Page, T., Putterman, L., in press. Communication and punishment in voluntary contribution experiments. *J. Econ. Behav. Organ.*
- Bolton, G., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity and competition. *Amer. Econ. Rev.* 90, 166–193.
- Carpenter, J., 2002. The demand for punishment. Working paper 2002-43. Department of Economics, Middlebury College.
- Carpenter, J., Matthews, P., 2002. Social reciprocity. Working paper 2002-29. Department of Economics, Middlebury College.
- Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. *Quart. J. Econ.* 117, 817–869.
- Cinyabuguma, M., Page, T., Putterman, L., 2004. On perverse and second-order punishment in public goods experiments with decentralized sanctioning. Working paper No. 2004-12. Department of Economics, Brown University.
- Cox, J., Friedman, D., 2002. A tractable model of reciprocity and fairness. Discussion paper. University of Arizona and University of California Santa Cruz.
- Davis, D., Holt, C., 1993. *Experimental Economics*. Princeton Univ. Press, Princeton.
- Eckel, C., Grossman, P., 1996. The relative price of fairness: Gender differences in a punishment game. *J. Econ. Behav. Organ.* 30, 143–158.
- El-Gamal, M., Grether, D., 1995. Are people Bayesian? Uncovering behavioral strategies. *J. Amer. Statist. Assoc.* 90, 1137–1145.
- El-Gamal, M., Grether, D., 2000. Changing decision rules: Uncovering behavioral strategies using estimation classification (EC). In: Machina, M., et al. (Eds.), *Preferences, Beliefs, and Attributes in Decision Making*. Kluwer, New York.
- Falk, A., Fischbacher, U., 1998. A theory of reciprocity. Working paper No. 6. Institute for Empirical Research in Economics, University of Zurich.
- Falk, A., Fehr, E., Fischbacher, U., 2001. Driving forces of informal sanctions. Working paper No. 59. Institute for Empirical Research in Economics, University of Zurich.
- Fehr, E., Gächter, S., 2000a. Cooperation and punishment. *Amer. Econ. Rev.* 90, 980–994.
- Fehr, E., Gächter, S., 2000b. Fairness and retaliation: The economics of reciprocity. *J. Econ. Perspect.* 14, 159–181.
- Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. *Nature* 415, 137–140.
- Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. *Quart. J. Econ.* 114, 817–868.
- Fehr, E., Schmidt, K., 2002. Theories of fairness and reciprocity—Evidence and economic applications. In: Dewatripont, M., Hansen, L., Turnovsky, S. (Eds.), *Economics and Econometrics*. In: 8th World Congress, Econometric Society Monographs. Cambridge Univ. Press, Cambridge.
- Ledyard, J., 1995. Public goods: A survey of experimental research. In: Kagel, J., Roth, A. (Eds.), *Handbook of Experimental Economics*. Princeton Univ. Press, Princeton.
- Masclot, D., Noussair, C., Tucker, S., Villeval, M.-C., 2003. Monetary and non-monetary punishment in the voluntary contributions mechanism. *Amer. Econ. Rev.* 93, 366–380.
- Page, T., Putterman, L., Unel, B., in press. Voluntary association in public goods experiments: Reciprocity, mimicry, and efficiency. Working paper No. 2002-19. Department of Economics, Brown University.
- Sefton, M., Shupp, R., Walker, J., 2002. The effect of rewards and sanctions in provision of public goods. Working paper. University of Nottingham and Indiana University.