

The Ecology of Collective Action: A Public Goods and Sanctions Experiment with Controlled Group Formation*

Umut Ones, Brown University
Louis Putterman, Brown University

Abstract

Accumulating evidence suggests that the outcomes of laboratory public goods games reflect the presence of individuals having differing preferences and beliefs. We designed a public goods experiment with targeted punishment opportunities to (a) confirm subject heterogeneity, (b) test the stability of subjects' types and (c) test the proposition that differences in group outcomes can be predicted with knowledge of the types of individuals who compose those groups. We find that differences in the inclination to cooperate have persistence, and that significantly greater social efficiency can be achieved by grouping less cooperative subjects with those inclined to punish free riding while excluding those prone to perverse retaliation against cooperators.

JEL #s: D91, D92, H41, D23

Keywords: public goods, voluntary contribution mechanism, heterogeneous preferences, group formation

Corresponding author: Louis Putterman, Department of Economics, Box B, Brown University, Providence, RI 02912. Tel.: 401-863-3837. Fax: 401-863-1970.
Louis_Putterman@Brown.Edu.

* We are indebted to Toby Page, who collaborated with Putterman on a number of related experiments, for his help in designing the experiments on which this paper reports. We thank Dennis Zachary Shubert for his work on the program with which the experiments were run. The research was funded by National Science Foundation grant SES-0001769.

The Ecology of Collective Action: A Public Goods and Sanctions Experiment with Controlled Group Formation

0. Introduction

The prisoners' dilemma game models a dynamic common to many economic interactions. Workers in teams, partners in firms, and communities of individuals who share a common environment or common interest often confront the dilemma that all cooperating toward a shared goal is in their joint interest, yet each is better off if others cooperate while she herself shirks responsibility. In some work teams, partnerships, irrigation associations, village woodlot projects, and other groups, collective action succeeds beyond the expectations of the conventional free rider analysis; but in others, failure is the norm.

To better understand why collective action sometimes succeeds and at other times fails, economists have conducted dozens of experiments with an n -person linear public goods game known as the voluntary contribution mechanism (VCM). These studies exhibit a high degree of concurrence in finding that (a) in one shot public goods games and in the first period of repeated games, subjects contribute an average of 50% or more of their endowments to the public good, and (b) in repeated play, contributions tend to decline with repetition, reaching an average of 10 or 15% in an announced last period (for reviews of the literature, see Davis and Holt, 1993; Ledyard, 1995).

The natural question for economists to ask about these results was whether they made necessary a reconsideration of conventional game theory, or whether with a little effort they could be reconciled with it. The dominant strategy of a rational agent intending to maximize his or her own payoff only, faced with other players of the same type who have common knowledge of their types, is to contribute nothing to the public good. Since subjects in most VCM experiments can't contribute negative amounts, errors due to unfamiliarity with the game would produce a natural upward bias. The decay in contributions might thus be interpreted as evidence of learning. But several kinds of evidence are leading experimentalists to reject a pure learning interpretation. First, contributions regularly rise again, even for experienced subjects, when the game is restarted (Andreoni, 1988). Second, when high contributors are grouped by the experimenter with other high contributors, their contributions are sustained at high levels (Gunnthorsdottir *et al.*, 2002). Third, when subjects have an opportunity to impose costly

monetary punishment on other group members, high contributors tend to continue contributing while punishing free riders, who respond by raising their contributions (Fehr and Gächter, 2000a, hereafter FG).¹ Rather than being due to some “typical” subject learning with experience that it’s best to free ride, argue FG, the usual fall-off of contributions in the standard VCM might better be attributed to the interactions between subjects prone to free riding and others more inclined to cooperate conditional on others’ doing likewise.

FG’s approach is one of several that suggest that the outcomes of public goods experiments can’t be understood without recognizing the presence of subjects having different preferences. In addition to the *actual* presence of subjects whose subjective payoffs don’t coincide with the material payoffs of the game, the existence of *beliefs* that such subjects may be present, and that other subjects may also believe such types to be present, can explain observed behaviors in a Bayesian model along the lines of Kreps, Wilson, Milgrom and Roberts (1982).² Andreoni’s (1995) analysis leads him to conclude that “on average about half of all cooperation comes from subjects who understand free-riding but choose to cooperate out of some form of kindness.” Offerman, Sonnemans and Schram (1996) and Palfrey and Prisbrey (1997) find evidence of “warm glow” giving, in which the donor acts as if obtaining utility from contributing to the public good irrespective of the benefit received by others. FG suggest that conditionally cooperative subjects reciprocate the “kind” contributions of other cooperators and punish the “unkind” free riding of self-interested types. Ahn, Ostrom and Walker (2003) argue that most behavior in public goods experiments can be explained by subjects having varying degrees of inequality aversion, with some subjects simply being payoff maximizers. Fischbacher, Gächter and Fehr (2001) and Kurzban and Houser (2001) identify many subjects in their conditional and circular contribution games as cooperators and conditional cooperators.

If subjects differ in type and if the decay of contributions typical in experiments with randomly formed groups is attributable to the way that conditional cooperators respond to free riders in the absence of punishment or partner selection mechanisms, then the study of group

¹ FG’s basic result has been replicated by Carpenter and Matthews (2002), Sefton, Shupp and Walker (2002), Fehr and Gächter (2002), Masclet, Noussair, Tucker and Villeval (2003), Page, Putterman and Unel (forthcoming, hereafter PPU), and Bochet, Page and Putterman (forthcoming, hereafter BPP).

² For models using this approach to explicitly show the viability of cooperative behaviors, see Guttman (2000), (2003).

behaviors becomes a study of an *ecology of interacting types*.³ Nature may have given rise to heterogeneity among human individuals either because the relevant traits have not had time to reach fixation or because an ongoing interplay of types proved evolutionarily stable;⁴ but humans may be able to design institutions that give more or less beneficial results by manipulating the types of individuals comprising particular groups, and by exposing people to social environments that may help (along with possibly varying in-born predisposition) to determine type.

Gunnthorsdottir *et al.* (2002) show that cooperative players in a VCM experiment can achieve superior outcomes when grouped by the experimenter with other cooperators, without their knowledge; Page *et al.* (forthcoming, hereafter, PPU) let cooperators seek one another out and observe a similar result. In this paper, we carry their approaches further by controlling group formation in a more complex collective action environment in which subjects have not one but two decision variables under their control—decisions on how much to contribute, and decisions on whether and by how much to impose costly punishment on other group members after learning of their contributions, as in FG.

Our paper proceeds as follows. In section 1, we discuss the theoretical framework of our study and its relationship to the existing literature. In section 2, we explain the design of our experiment. In section 3, we describe the results as they illustrate the general character of cooperation and punishment behaviors. Section 4 focuses on the differentiation among groups in our experiment. Section 5 analyzes the persistence of individuals' behaviors, evidence of environmental influences, and the relationship between the tendency to contribute and the tendency to punish. Section 6 concludes with a discussion and summary.

1. Theoretical framework

a. Preferences and Types

³ As in Schelling's famous "ecology of micromotives" (1971), the emergent properties of the social system are distinct from the intentions and not immediately predictable from the actions of the individuals, taken in isolation. In our case, we use the term "ecology" to emphasize the importance of interactions, and the importance, for purposes of predicting outcomes, of knowing which types of agents it is that are interacting. We study mainly the interactions of given types of agents, although the question of how such interactions might change the composition of a population, either by causing given individuals to change their type or by altering what kinds of individuals are present in the population, is a related and interesting one.

⁴ For a survey of evolutionary models of preference formation, including ones in which both reciprocator and payoff maximizing behaviors exist in equilibrium, see Sethi and Somanathan (2003). Heterogeneity of human behavioral inclinations may be due to differences of culture and individual upbringing, as well as of genes. See Boyd and Richerson (1985), Durham (1991) and Ben-Ner and Putterman (1998).

Several kinds of preferences, among them inequality aversion, altruism, and “warm glow,” offer potential explanations for the persistence of positive contributions in public goods experiments. We focus on the preference called *reciprocity* or *conditional cooperation* because it is consistent with the actions of many subjects in voluntary contribution and other experiments,⁵ and because we find merit in the growing literature on the topic by anthropologists, sociologists, socio-biologists, evolutionary psychologists, and economic theorists and experimentalists.⁶ According to Hoffman, McCabe and Smith (1998) and Fehr and Gächter (2000b), reciprocity entails an inclination to confer benefits on those who help one and to impose costs on those who harm one. The first, favor-returning part, can be called “positive reciprocity,” the second, harm-returning part, “negative reciprocity.” In both parts, the reciprocator shows a willingness to incur costs, and accordingly his or her actions are inconsistent with an exclusive preference for maximum material payoff. Although a reciprocating agent may obtain material benefit in the long run if she creates an expectation of like future actions in repeated play with the same partner or when reputation carries into play with others, strong reciprocators reciprocate even in one shot games or end game situations with no potential for reputational gains.⁷

The application of reciprocity to public goods contributions is that when others contribute, it benefits one, so a reciprocator will wish to return the favor by reciprocating also (see Fehr and Gächter, 2000b). One could logically suppose that reciprocity would not impact contributions to a public good unless contributions were sequential or the same group of individuals interacted repeatedly and could thus respond to one another’s previous moves. The alternative approach, which we follow here, is to assume that reciprocity matters even for simultaneous moves, with partners’ expected moves taking the place of past moves in determining own current actions. If reciprocators are optimistic about others’ contributions at the outset and, in addition, trust that fellow players will continue to act as they have acted thus

⁵ Other experiments providing evidence of positive and negative reciprocity include gift exchange games (for example, Fehr, Gächter and Kirchsteiger, 1997 and Fehr and Gächter, 1998), the extended form games studied by McCabe, Rassenti and Smith (1996), and the two stage dictator game of Ben-Ner *et al.* (2004).

⁶ Examples include Boyd and Richerson, 2002, Henrich and Boyd, 2001, Cosmides and Tooby, 1989, Rabin, 1993, Sethi and Somanathan, 2003, Gintis 2000, Guttman, 2000, Guttman, 2003, McCabe, Rassenti and Smith, 1996, Fehr and Gächter, 2002b, and Ben-Ner and Putterman (2002).

⁷ Examples include engaging in punishment of low contributors in the perfect stranger conditions of the VCM-with-punishment experiments of Fehr and Gächter (2000a) and Anderson and Putterman (2003). Gintis, Bowles, Boyd and Fehr (2005) call this propensity “strong reciprocity” to distinguish it from the kind of reciprocity motivated by self-interest.

far, then they behave in the VCM like tit-for-tat players, contributing on the first round and continuing to contribute in each subsequent one provided that others have done so thus far. By the same token, two individuals with the same degree of reciprocity might act differently owing to different beliefs, which at the start of play may be based on different past experiences.

Negative reciprocity also has a part to play in a public goods game. Not contributing when others contribute constitutes failure to reciprocate others' kindness, which can trigger negative reciprocity—that is, a desire to punish the free rider. In the basic VCM without punishment stage, negative reciprocity can only take the form of reducing one's contributions, a blunt instrument since there is no way to direct it differentially against free riders without also hurting high contributors in the group. By contrast, in the VCM with punishment stage as introduced by FG, negative reciprocity can be more distinctly manifested in the form of punishing low contributors, an action that doesn't prevent the reciprocator from continuing to reciprocate the contributions of high contributors in the same group. Although positive and negative reciprocity have been presented as two sides of the same coin, we see as an open matter, to be investigated empirically, how closely the strengths of the two tendencies are correlated.

Analysis of past VCM-with-punishment experiments indicates that one other preference must also be accounted for to understand the ecology of interacting types. Cinyabuguma, Page and Putterman (hereafter CPP, 2004) have demonstrated that about 20% of punishment in VCM-with-punishment experiments is aimed at high, rather than low, contributors, and that this is a major reason why earnings (unlike contributions) fail to be higher in treatments that include a punishment option.⁸ Most of this “perverse” punishment seems explicable either as attempts to retaliate for the punishment the agent has herself received, or as attempts to raise the punisher's *relative* earnings at the cost of his *absolute* earnings, a motivation called “spite” by Saijo and Nakamura (1995). In a repeated game with fixed group composition, retaliatory punishment could be a self-interested response intended to make it safe to continue free riding with less likelihood of being punished again. But some retaliatory and spiteful punishment appears to stem from a preference type in its own right. Perverse punishment is observed even in the last period of play, and in perfect stranger designs.

⁸ They analyze the original data used in FG (2000), provided by those economists, as well as the data of the cooperation-with-punishment experiments reported in Bochet, Page and Putterman (forthcoming,) and PPU. Anderson and Putterman (2004) find even more perverse punishment in a set of perfect stranger treatment cooperation with punishment experiments with which they study they impact of differing punishment costs.

Although we frequently simplify by referring to discrete types of agents, we view individuals as in principle distinguishable by preferences that can take on wide ranges of values, making possible a potentially infinite number of types. We use the terms “type” and “types” to refer to differing preference-based inclinations to contribute to a public good, to punish free riding, and to punish perversely. We assume that those inclinations come into the experiment with the subject, while the degree to which they persist over time is a matter to be determined experimentally. In principle, we should be able to predict an individual i 's contribution C_i to a public good by reference to the three preference parameters (a) strength of reciprocity, (b) strength of negative versus positive reciprocity, and (c) degree of spitefulness (propensity toward perverse punishment). i 's initial beliefs about the distribution of types in the population and the evolution of those beliefs in response to others' observed choices will also affect C_i , but are conceptually distinct from i 's preferences. The amount of costly punishment i gives to some $j \neq i$ can in principle be predicted by the same factors plus the cost of punishment.⁹

b. The ecology of types

If *individuals* can be characterized in terms of their beliefs and preferences, it may be possible to predict the outcomes of *group* interactions in a finitely repeated VCM-with-punishment game by knowing what types of individuals constitute the group. This is straightforward in a few simple cases: for example, if a group is composed entirely of reciprocators who begin with optimistic beliefs about one another's types, they will establish from the outset and maintain even without communication an equilibrium of high contributions, since each will begin with a high contribution and will continue to contribute, having had her expectations validated. For some heterogeneous groups, too, outcomes may be fairly predictable. Consider a group of four with the following composition: two agents are payoff maximizers; two agents are strong reciprocators, one inclined more strongly towards positive than negative reciprocity, the other inclined more strongly towards negative than positive reciprocity. Suppose all four agents believe there to be a 50% chance of encountering a reciprocator, and that each believes the others to have the same beliefs as herself. Though our prediction might lack precision, we could be fairly confident that this group would begin with a range of positive although probably not maximal contributions, and that it would exhibit

⁹ Carpenter (2003) and Anderson and Putterman (2003) find that the amount of punishment purchased by the punisher is a decreasing function of its cost to her.

relatively high contributions after a few periods of play, since the subject with strong negative reciprocity would from the outset punish any low contributions, and the payoff maximizers will adjust their contributions upwards to avoid being punished.

Now change the composition of this hypothetical group so that instead of two payoff maximizers there is one payoff maximizer and one agent inclined towards low contributions and perverse punishment (the other two members are as before). The group outcome is now more difficult to predict, since the negative reciprocator will tend to punish the perverse punisher, but the latter will punish back. Since the retaliator resists the pressure to contribute more, so might the payoff maximizer. The positive reciprocator too will refrain from making maximal contributions, since some others are failing to do so. A good deal of punishment might end up being wasted without inducing a rise in contributions. Without being more precise, we can predict that this group will exhibit lower average contributions and earnings than the counterpart group that differs in the type of one member only, thus illustrating how different group compositions will be associated with different group outcomes.

c. Types and environments

These examples illustrate that it is not only individuals' preferences and beliefs, but also the way in which the latter interact with the preferences and beliefs of the others with whom they are grouped, that will influence their individual choices and their groups' outcomes in continuing interactions including laboratory experiments. We find helpful a notion of type/environment interactions that parallels (without precisely corresponding to) the concepts of genotype and phenotype common in biology. An individual enters an encounter, in life or in the laboratory, as a bearer of *preferences* that constitute her type, similar to the genotype in biology. The individual's *behavior*—biology's phenotype—remains the same or changes over the course of the interaction depending on the actions of the others she encounters.¹⁰ Our working hypothesis, which we test in several ways, is that there is some persistence of type. We try to contribute, however modestly, to the broader investigation of just how much type persistence individuals display and how much their behaviors are altered by the environments they encounter.

¹⁰ One reason the analogy is imperfect is that it leaves out the a range of influences intermediate between the very early influences of genes and socialization, on the one hand, and events in the lab, on the other. The human behavioral phenotype is a work in progress throughout life, with the experience in our lab being one of tens of thousands of behavior-shaping experiences.

2. Experimental design

In our experiment, subjects play a repeated linear public goods game with contribution and punishment stages in groups of four. Every period, each subject i has 10 experimental dollars¹¹ of which he contributes an integer amount, C_i , possibly 0, to a group account, and retains $(10 - C_i)$, giving provisional earnings

$$y_{i,p} = (10 - C_i) + (0.4)\sum_{\text{all } j} C_j \quad (3)$$

where the summation is taken over all four group members, i included. After individual contribution decisions are revealed, subjects can use current period provisional earnings to reduce the earnings of other group members at a cost of 0.25 to the punisher for each experimental dollar lost by the person targeted. i 's final earnings for the period are thus

$$y_{i,f} = \max\{[(10 - C_i) + (0.4)\sum_j C_j - (0.25)\sum_j R_{ij} - \sum_j R_{ji}], 0\} \quad (4)$$

where R_{ij} is the number of dollars by which i reduces j 's earnings, and conversely for R_{ji} .¹² As in FG, other group members are identified to i by letters B, C and D, which switch randomly each period, to minimize vendettas.¹³

In each of the twelve sessions constituting the present set of experiments, sixteen subjects, undergraduates at Brown University drawn from all disciplines, made twenty five sets of contribution and punishment decisions. While they were unable to communicate with others and unable to tell which other individuals constituted their own group, each could see that fifteen other subjects were present, and the experimenter truthfully informed them that they would be put in one group for the first period, a possibly different group whose members would not change during periods 2 – 5, another fixed group for periods 6 – 15, and a final fixed group for periods 16-25, with some overlap of membership among these four groupings being possible but

¹¹ Experimental dollars were converted for payment at the end of the session at the rate of one experimental dollars = \$.05. Subjects also received a \$5 show-up fee. Earnings for a 90-110 minute session averaged about \$25.

¹² Earnings were set to zero if they would otherwise be negative so that subjects could be guaranteed at least their \$5 show-up fee and each period is strictly independent of the others from a budgetary standpoint.

¹³ The choices and consequences each period are identical to the treatments in BPP and PPU with punishment and without communication or endogenous group formation. They are also the same as in FG except that (a) in our experiment, the cost of punishment to the punisher is a constant fraction of its cost to the person punished, and punishment is imposed in dollar terms, rather than in percentages of pre-punishment earnings, and (b) our experiments do not also include a VCM-without-punishment condition which FG use for within-subject comparison. Although our fixed per unit cost of punishment differs from that in FG (2000), it is consistent with the mechanisms used in Fehr and Gächter (2002) and in Sefton, Shupp and Walker (2002).

not certain. No information was provided about the basis of group formation. We analyze three treatments, two of which implement purposeful grouping algorithms in differing orders, the third of which uses random grouping and as a baseline for comparison.

The two purposeful grouping treatments attempted first to identify subject “types” by allowing them to play the contribution and punishment game for five periods in groups of similarly diverse composition. When these five “diagnostic” periods ended, the subjects played in homogeneous groups made up of members of similar “type” to themselves for ten periods, and in randomly formed groups of no special composition for ten periods. In the HR (homogeneous-random) treatment, play in homogeneous groups occurred in periods 6-15 and that in random groups in periods 16-25, while in the RH (random-homogeneous) treatment, this order was reversed. Four sessions of an RR (random-random) treatment were also conducted as a baseline for comparison.

The five diagnostic periods worked as follows. Each subject first decided on an amount to contribute to the group account in period 1. The computer then showed the subject the contributions of the other three members of her group, each group having been made similarly diverse by being assigned as members one of the four highest first-round contributors in the session, one of the four lowest contributors, and so on, without the subjects’ knowledge.¹⁴ After the subject made her first-period decisions on punishment of her three counterparts, there was a second placement into groups again meant to be diverse, this time in terms *both* of contribution *and* of punishment behavior. The computer did this by calculating a *punishment index* which is a larger positive number the more the subject engaged in punishment of low contributors, a larger negative number the more she punished high contributors, and zero if she didn’t punish at all.¹⁵ Subjects were given ranks from 1 to 16 both for their contribution level (with 1 representing the

¹⁴ Group formation also follows decisions in Gunthorsdottir *et al.* In neither case is it necessary to mislead subjects; the matter is simply not addressed in the instructions. Since even in experiments like BPP and PPU, subjects are shown others’ contribution amounts only after all have submitted their decisions, and since the computer makes the group assignment in a fraction of a second, a subject could see no difference between an experiment in which groups are formed before contribution decisions and one like the present one in which groups are formed in period 1 only after the first contribution decisions.

¹⁵ The index is defined as $PI_i = -0.5R_{i1} + 0.33R_{i2} + 0.33R_{i3} + 0.5R_{i4}$ where the R ’s are the number of experimental dollars by which subject i reduced the earnings of another subject, 1 references the subject who made the highest contribution to the group account in that group and period, 2 references the subject who made the next highest contribution, etc., and where the weights $-0.5, 0.33, \text{etc.}$, capture the average change in the recipient’s next period contribution due to receiving one dollar of reductions, as estimated in CPP (2004) using the data in BPP by a method similar to that in Table 3 below. For a more complete description, see Appendix A of our working paper.

highest level, 16 the lowest) and their punishment index (1 representing the most punishment of free riders, 16 the least punishment of free riders and most punishment of high contributors), the two rank numbers were added together, and the four subjects with the lowest combined ranks (who tended to be both high contributors and strong punishers of free riders) were assigned to *different* groups, the four with the next lowest combined ranks to different groups, and so on.¹⁶ This assured, to the extent possible, that every group had both high and low contributors, both aggressive punishers of low contributors and non-punishers or perverse punishers.

At the end of period 5, the subjects are ranked by their average contribution over periods 1 to 5 as a whole, and by their average punishment index over periods 1 to 5 as a whole. These ranks for behavior in periods 1 to 5 were the basis of their group assignments during the homogeneous grouping portion of the session, regardless of whether that was periods 6-15 (as in HR) or periods 16-25 (as in RH).¹⁷ In both the period 2 group assignment and the homogeneous (period 6 or 16) group assignment stages, the contribution and reduction ranks were added together. For homogeneous groupings, the four subjects with the lowest summed rank, and so on, were placed in the *same* group, exactly the opposite of heterogeneous grouping for diagnostic play, where apparently like subjects were dispersed among *different* groups.¹⁸ For purposes of analysis, we refer to the four subjects in each session with the “best” ranks for contributions and punishment of low contributors (i.e., the apparently high contributors and high punishers of free riders) as that session’s “Group 1,” those with the next best ranks as “Group 2,” and so on.

We placed subjects into groups that came as close as possible to being equally diverse during the diagnostic periods because only by doing so could we hope to distinguish between differences in behavior attributable to differences in the kind of group one found oneself in and differences in behavior attributable to subjects’ “types.” In particular, with homogeneous or randomly formed groups, one subject might chose a lot of punishment for low contributors while

¹⁶ If two or more subjects were tied in contributions or in punishment indices, they received the same rank, for example if two tied for second place, both were treated as having the rank 2.5. This way the ranks of all 16 subjects always summed to 16!, assuring that the contribution and punishment ranks had equal weight when added together. If there were ties with respect to the combined rank, for example between the fourth and the fifth ranked subjects, the computer broke them randomly. The four lowest ranked, the next four lowest ranked subjects, and so on, were allocated among the four groups in a random order, so that it was *not* the case that one group always included the subjects with combined ranks 1, 5, 9, and 13 while another included those ranked 2, 6, 10 and 14, and so on.

¹⁷ That is, even in the RH treatment, only behaviors in periods 1 to 5 were used in identifying the “type” of the subject, behaviors in the additional ten periods (6-15) that elapsed before “homogeneously grouped” playing no part in period 16 – 25 group assignments.

¹⁸ We considered a different design, in which the highest contributors are put in one group, the most vigorous punishers of free riders in another, etc. Preliminary results showed contribution and punishment tendencies to be correlated (see Result 10), however, so we decided not to pursue this idea in this first exploration of exogenous grouping for contributions and punishment.

a second did not do so only because the first saw both high and low contributions in his group whereas the second saw only high or only low ones. We chose not to regroup subjects between periods 2 and 5, even though such regrouping might have aided in maintaining heterogeneity, because our goal was to study the kind of ongoing interactions represented by partner groups during periods 6-15 and 16-25, and we thus thought it better to have our diagnostic reading on types be based on behaviors in a partner play environment.¹⁹

In the RR treatment, subjects were placed in possibly different groups in periods 1, 2, 6 and 16, always by a strictly random grouping process. As in the HR and RH treatments, group membership remained fixed during periods 2 – 5, 6 – 15, and 16 – 25. Exactly the same instructions were given and subjects’ tasks were identical in all three treatments, so subjects were unaware of the basis of grouping, unaware that different treatments existed, unaware that they were participating in one treatment rather than another, and unaware when they had been placed in a high contributor or a low contributor group relative to other groups in their session (subjects never saw the decisions made in groups other than their own). We conducted four sessions of each treatment using a total of 192 subjects. The three treatments are summarized in Table 1. Instructions are provided in the Appendix, available on line.

Table 1. Summary of treatments and grouping procedures.

Treatment				Overall
Period(s)	HR	RH	RR	
1	Heterogeneous Groups ^a	Heterogeneous Groups ^a	Random Groups	
2 – 5	Heterogeneous Groups ^b	Heterogeneous Groups ^b	Random Groups	
6 – 15	Homogeneous Groups ^c	Random Groups	Random Groups	
16 – 25	Random Groups	Homogeneous Groups ^c	Random Groups	
Number of Sessions	4	4	4	
Number of Subjects	64	64	64	192

Notes: a. Grouping based on initial contribution rank. b. Grouping based on initial contribution and initial punishment rank. c. Grouping based on combined contribution and punishment ranks of periods 1 – 5.

¹⁹ It should be noted that with these desirable features of the diagnostic portion of our experiment came a much steeper challenge to the persistence of subject type than occurs in the experiment of Gunnthorsdottir *et al.* In that experiment, subjects were grouped with like contributors from period 1 onwards, and were regrouped as often as necessary to preserve homogeneity, which provided each subject with immediate reinforcement of his initial inclination and gave little or no opportunity to observe contrary behavior. Therefore, if subjects persisted in making characteristically high or low contributions after five periods of playing with heterogeneous others, in our HR treatment, or after fifteen periods of playing with heterogeneous and then randomly assigned others, in our RH treatment, this is stronger evidence of persistence of heterogeneous types than the already impressive evidence in Gunnthorsdottir *et al.*

3. Results, Part I: General description

Figure 2 plots the average contribution of subjects in the HR sessions, by period, while Figure 3 does the same for the RH sessions. For purposes of comparison, the average contribution of subjects in the (baseline) RR sessions is also shown in each figure. In the figures, contributions are averaged over all subjects in periods in which there is no basis for distinguishing the 16 groups involved, but for those ten periods in which grouping was intendedly homogeneous by diagnosed type, a separate line is shown for the average contribution of “Group 1,” etc. These separate lines average the contributions in the respective groups over the four sessions of the treatment in question (HR or RH).

Result 1. *Contributions began at high levels and initially increased with repetition.*

The first thing to be observed is that in all three treatments, contributions to the group account began at relatively high levels, about 65% of endowment in HR, 74% in RR and 80% in RH,²⁰ and that unlike in VCM experiments without punishment but like ones with punishment (FG, Bochet et al., PPU, etc.), contributions do not show a tendency to decline with repetition. (Summary statistics, including average contribution levels in various periods, are provided in Appendix Table 1, available on line.) These two features suggest (a) that punishment of low contributors was anticipated even before it was observed by subjects, and (b) that the threat of punishment, perhaps combined with reciprocators’ abilities to continue contributing even while punishing free riders (FG), kept most subjects contributing at relatively high levels, although

²⁰ First period contributions in all three treatments are higher than in the no punishment treatments in Page *et al.* (PPU, forthcoming), which use identical parameters. Mann-Whitney tests show the differences in first period contributions in comparison to treatments without punishment (those in PPU) to be statistically significant. Because instructions, parameters, and experimental procedures were identical for the RH as for the HR treatment during periods 1 – 5, there is no reason why contributions should have differed systematically between them, nor between those treatments and RR, which differed only in that early group formation was strictly random rather than being intendedly heterogeneous. The apparent differences—which are found to be statistically significant both in period 1 and in periods 1 – 5 as a whole using either Mann-Whitney or Kruskal-Wallis tests—can thus only be attributed to unmeasured differences in subjects’ personalities, moods, weather, etc. A possible explanation for contributions *remaining* somewhat higher in the RH and RR than in the HR treatment after period 5, and for contributions to vary more among groups in the homogeneous grouping periods (6 – 15) of HR than those (16-25) of RH, is that because vigorous punishers of free riders tended to be present in all RH and RR groups during periods 6 – 15, while some HR groups were intendedly assigned no such punishers in those periods, expectations of punishment and thus norms of high contribution were more thoroughly established in the RH and RR treatments than in the HR treatment. (We return to this theme in the concluding section.) Also, as Appendix Table 2 shows, the HR sessions appear to have had more subjects inclined toward perverse punishment: HR sessions had a total of 8 subjects with negative punishment indices during periods 1 – 5, among whom two had PI values below -3 and a third a PI value close to -2, whereas RH sessions had only 6 subjects with negative punishment indices, of whom none had a PI value below -2 and only two had values below -1.5. These differences again cannot be attributed to treatment. Participants were in both cases inexperienced and drawn from the same general undergraduate subject pool.

small end-game effects are apparent in periods 5 and 15 and a larger one in the approach to period 25. Discussion of the differences among groups in Figures 1 and 2 is postponed until Section 4.

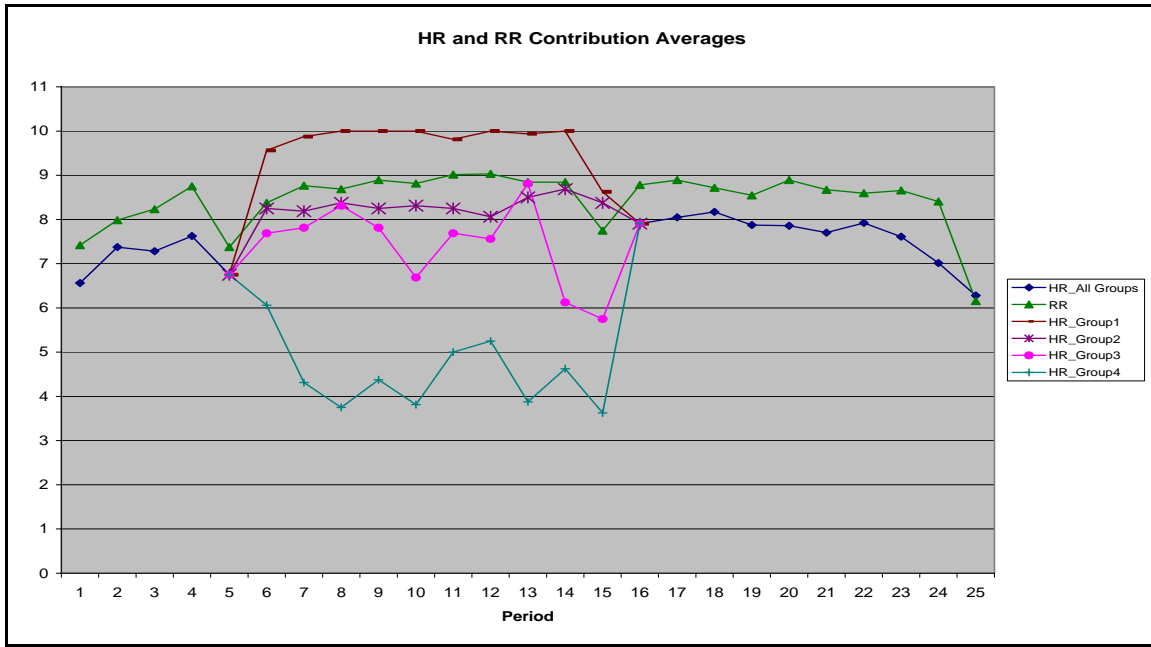


Figure 1. HR and RR (Baseline) average contribution by period and group.

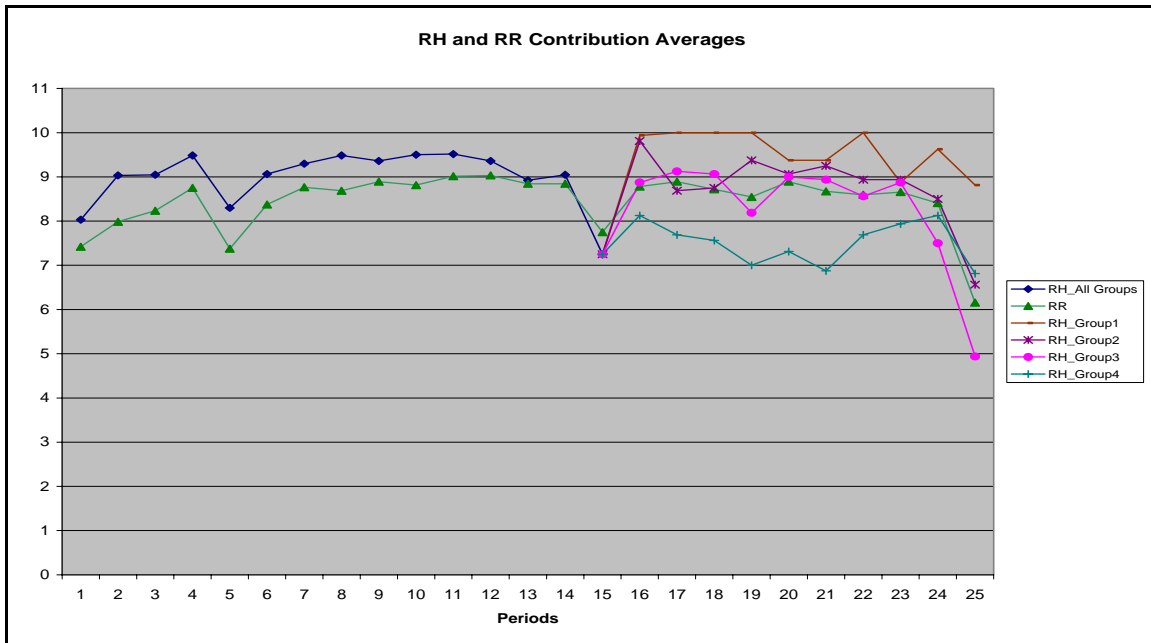


Figure 2. RH and RR (Baseline) average contribution by period and group.

Punishment was common in the experiment, including the last period, with 15%, 11%, and 11% of each subject's three opportunities to reduce others' earnings each period being utilized in the HR, RH, and RR treatments, respectively—equivalent to an average subject punishing some other subject in 1/3 or more of all periods. Most punishment was directed at low contributors, suggesting the presence of negative reciprocity, but 35% of punishment dollars in the HR treatment, 38% in the RH treatment, and 10% in the RR treatment were perversely aimed at groups' highest contributors of the period, suggesting the presence of spite or some other motivator of perverse punishment, as well. Last period punishment is addressed further, below.

Result 2. *As in other such experiments, the less an individual contributed to the group account relative to others, the more he or she tended to be punished. But in the HR and RR treatments, one was also more likely to be punished the more one exceeded the average, a clear indication of perverse punishment.*

Table 2 shows estimates of a regression that follows a specification in FG, where the dependent variable is the number of dollars of reductions targeted at subject j , and the explanatory variables are (a) the *absolute negative deviation* of j 's contribution, defined as the difference between it and the average of other group members' contributions in the period, if j contributed less than the average, and as zero otherwise, (b) j 's *absolute positive deviation*, defined conversely, and (c) the others' average contribution.²¹ Like FG, group and period fixed effects are also included but not shown.²² For the HR and RR treatments, both deviation terms have highly significant coefficients, although the magnitude is higher for negative deviation, indicating that between six and eight times as much punishment was received per dollar below the average as per dollar above the average. Still, it is noteworthy that one would have been “sticking one's neck out” to contribute “too much” in some groups. The presence of this much perverse punishment is an obvious disincentive to efficiency and thus a reason why one might want to “exclude” perverse punishers, as is achieved in many groups during our periods of

²¹ All of the regressions are reported with robust (Huber-White) standard errors calculated using the robust command in Stata.

²² Although the dependent variable is the amount of punishment received by a given subject, say j , in a given period, individual fixed effects are inappropriate because the decision-makers determining this variable were the other three members of j 's group, who could not distinguish j from one period to the next from other members of their group (due to the random reassignment of letter codes). Observations are assigned the dummy of the *group* that the punished subject belonged to during the period in question. For each treatment there are a total of 64 group dummies, that is four per session in each of four sets of periods (1, 2-5, 6-15 and 16-25).

homogeneous grouping (assuming persistence of past behaviors).²³ For the RH treatment, only the negative deviation term is significant (as in FG), in part because the average contribution in RH was so high that there was little margin above it most of the time. The point estimate of the coefficient on absolute positive deviation is in this case negative.

	HR	RH	RR
Constant	0.782 ** (0.36)	1.314 ** (0.57)	-0.947 (0.65)
Absolute Positive Derivation	0.086 ** (0.04)	- 0.060 (0.07)	0.118 * (0.06)
Absolute Negative Derivation	0.548 *** (0.04)	0.535 *** (0.06)	0.914 *** (0.06)
Average Others' Contribution	-0.062 ** (0.03)	-0.088 (0.05)	0.077 (0.07)
Number of Observations	1600	1600	1600
R²	0.410	0.377	0.539
F Value	11.98 ***	9.32 ***	11.80 ***

Dependent Variable: Punishment Received
(an increase is a positive value)

(*) Significant in 10 %

(**) Significant in 5%

(***) Significant in 1 %

Numbers in parentheses are Huber/White/Sandwich adjusted standard errors. Same Group Dummies (16) and Period dummies (25) used.

Table 2. Punishment received as a function of contribution deviations and average.

Result 3. *Subjects responded predictably to being punished.*

In the regressions in Table 3, one for each treatment, the dependent variable is the change in subject j 's contribution from period t to period $t+1$. The independent variables are formed by multiplying the amount of punishment received by j in period t by two dummy variables, the first being 1 if j contributed an amount greater than or equal to the group's average in the period, 0 otherwise; the second being 1 if j contributed less than the group average, 0 otherwise. The

²³ Since perverse punishers constitute about a third of our subject pool and they obtain low rankings in periods 1 – 5, they should be almost entirely absent from groups 1 and 2 and mainly found in groups 3 and 4. Supportive evidence is given in Result 6.

coefficients can be interpreted as the impact of one dollar of punishment received on j 's change of contribution conditional on j 's relative contribution falling in the indicated range. Because 24 contribution changes are observed for each subject and reactions might differ with time, the regressions include subject and period fixed effects (not shown). The estimated coefficients show that on average, a subject increased her contribution by somewhere between 59 and 72 cents for every dollar of punishment she received if she had been contributing less than the group's average, which is consistent with the subject interpreting the punishment as an indication of displeasure for free riding or as a warning that more punishment would be forthcoming if the behavior continued. By contrast, subjects who were contributing the average or more when they received punishment tended to reduce their contributions by somewhere between 14 and 25 cents for each dollar of punishment received, which demonstrates the efficiency reducing impact of perverse punishment.

	HR	RH	RR
Constant	-2.804 ** (1.11)	1.259 *** (0.44)	-0.018 (0.52)
Pun Received as High Contributor	-0.142 ** (0.06)	-0.245 *** (0.07)	-0.200 (0.17)
Pun Recieved as Low Contributor	0.721 *** (0.05)	0.708 *** (0.05)	0.587 *** (0.04)
Number of Observations	1536	1536	1536
R²	0.244	0.277	0.295
F Value	3.65 ***	3.63 ***	3.42 ***

Dependent Variable: Change in Contribution
(an increase is a positive value)

(*) Significant in 10 %

(**) Significant in 5%

(***) Significant in 1 %

Numbers in parentheses are Huber/White/Sandwich adjusted standard errors without clusters. Subject fixed effects (64) and period fixed effects (24) applied.

Table 3. Change in contribution as a function of punishment received.

Result 4. *There is strong evidence of negative reciprocity, and milder evidence of spiteful preferences, in the form of last period punishing that cannot be explained by strategic motivations.*

A payoff maximizing subject would never punish in the last period. We can thus investigate whether free riders and high contributors were punished mainly to try to get them to change their behaviors or whether underlying preferences were manifested in the form of punishing in the last period of a group's interaction.²⁴ We can confirm by a simple count that there were similar amounts of punishment in period 25 (and the last periods for specific groups, periods 1, 5, and 15) as in other periods. However, a more rigorous test can be done by estimating regression equations similar to those in Table 2 but including interaction terms to check whether the likelihood of being punished for positive or negative deviations was any different in final periods. In one set of regressions, in Table 4, we multiply the absolute positive and absolute negative deviation by a dummy variable, DUMMY LAST, which takes the value 1 in periods 1, 5, 15 and 25 and is 0 otherwise. In another set of regressions, we use one dummy variable, DUMMY 1,5,15, for periods that represent the last time a particular group plays together but not the last period in the session as a whole, and a separate dummy variable, DUMMY25, for the last period of the session, to test whether there is a stronger reduction of punishing in that period. All regressions include group and period dummies, not shown.²⁵

Inspecting Table 4, we find no qualitative change with respect to the non-interacted deviation variables. The DUMMY LAST interaction, covering all four last periods of partner groups, shows no significant difference except in the RH treatment, where it indicates that there was significantly *less* perverse punishment in last periods for that treatment. In the other specification, there are also no significant interaction dummies except for the interactions of DUMMY1,5,15 with absolute positive deviation in the RH treatment. Consistent with the corresponding DUMMY LAST term for RH, these coefficients show there to be significantly *less* perverse punishment in last periods of that treatment. The interactions with absolute negative deviation are always insignificant, and for DUMMY25 they have positive values. These results support FG's belief that punishment of low contributors is not mainly strategic in nature, but rather is due to a preference (see also Falk, Fehr and Fischbacher, 2001). The result for the interaction with positive deviation in the RH treatment raises the possibility that the impulse to punish perversely is largely strategic, but since this effect is not observed in the other

²⁴ Although figures 1 and 2 show that there were substantial contributions to the group account in period 25 in all three treatments, contributing can't be taken as definitive proof of *positive* reciprocity, because even payoff maximizing subjects had reason to contribute in the last period if they believed they would otherwise have been heavily punished.

²⁵ The method of assigning group dummies and the number of those dummies is the same as in Table 2.

two treatments, with the end period interactions actually having positive point estimates for the RR treatment, our assumption that some subjects hold a preference for punishing perversely is supported for two out of three treatments.²⁶

	HR		RH		RR	
Dependent Variable Punishment Received						
Constant	0.904 ** (0.43)	-0.204 (0.32)	1.259 ** (0.62)	0.761 (0.94)	-1.088 (0.72)	-1.597 ** (0.77)
Absolute Positive Deviation	0.088 ** (0.04)	0.090 ** (0.04)	0.004 (0.07)	0.004 (0.07)	0.120 * (0.07)	0.110 (0.07)
Absolute Negative Deviation	0.558 *** (0.05)	0.559 *** (0.05)	0.479 *** (0.06)	0.478 *** (0.06)	0.900 *** (0.06)	0.898 *** (0.06)
Others Contribution	-0.061 ** (0.03)	-0.061 ** (0.03)	-0.084 (0.05)	-0.082 (0.05)	0.079 (0.07)	0.070 (0.07)
Dummy Last * Abs. Pos Dev	-0.018 (0.07)		-0.148 * (0.08)		0.012 (0.05)	
Dummy Last * Abs. Neg Dev	-0.056 (0.10)		0.154 (0.14)		0.067 (0.17)	
Dummy 25* Abs. Pos Dev		0.068 (0.09)		-0.124 (0.19)		0.157 (0.13)
Dummy 25*Abs. Neg Dev		0.185 (0.17)		0.404 (0.32)		0.192 (0.30)
Dummy 1,5,15*Abs. Pos. Dev.		-0.081 (0.10)		-0.138 * (0.08)		-0.034 (0.05)
Dummy 1,5,15*Abs. Neg. Dev.		-0.161 (0.12)		0.063 (0.12)		0.017 (0.19)
Number of Observations	1600	1600	1600	1600	1600	1600
R²	0.410	0.414	0.384	0.390	0.540	0.541
Dummy Significance (F)	11.47 ***	11.24 ***	8.94 ***	8.52 ***	11.62 ***	11.29 ***

(*)Significant in 10 %

(**)Significant in 5 %

(***)Significant in 1 %

Numbers in parentheses are Huber/White/Sandwich adjusted standard errors. Group and period dummies are included but not shown.

Table 4. Punishment received as a function of contribution deviations.

²⁶ The substantial perverse punishment exhibited in the perfect stranger experiment of Anderson and Putterman (forthcoming) is also inconsistent with interpreting perverse punishment as mainly a strategy to increase payoff by warning high contributors not to punish one's free riding in the future, and instead consistent with there being a preference—possibly due to spite, possibly some other cause—for punishing perversely.

4. Results, Part II: Differences among groups and persistence of behaviors.

In this section, we investigate the central question motivating our experimental design: were group behaviors differentiated as would be predicted if we correctly identified subject types and if types were persistent?

Result 5: During the homogeneous grouping periods of both the HR and RH treatments, the ordering of contributions is as would be expected assuming correct identification of types and persistence of the associated behaviors.

Figure 3 graphs the average levels of contributions during periods of homogeneous grouping in the HR treatment sessions. Each of the left two and right two bars represents the average contribution in a set of groups—the Group 1 bar, for example, being the average contribution in the groups composed of periods 1 – 5’s highest contributors and most vigorous non-perverse punishers in the four sessions of the treatment. As is easily seen, the ordering is exactly as would be predicted based on diagnostic period behaviors. For comparison, the middle bar represents the average contribution of treatment RR subjects during the same ten periods (6 – 15). Using Mann-Whitney tests, we find that average contributions by groups of first-ranked subjects in the HR treatment significantly exceeded those of RR treatment counterparts during the same periods (6 – 15). The HR treatment’s fourth-ranked groups contributed significantly less than RR groups.²⁷ These results follow expectations, since first-ranked groups, consisting of diagnostic period high contributors who punished free riding, are expected to contribute more than randomly formed groups, which may contain not only low contributors and non-punishers but also perverse punishers. On the other hand, fourth-ranked groups, which are likely to contain diagnostic period perverse punishers and low contributors, would be expected to contribute less, on average, than randomly formed groups.²⁸ A Kruskal-Wallis test finds the difference in mean contribution among groups to be significant at the 5% level.²⁹

Figure 4 parallels Figure 3 but graphs the average contributions by group during homogeneous grouping periods 16-25 of the RH treatment sessions and the average contributions of RR subjects during the same periods of their sessions. Once again, the rank order of the contribution averages is as would be predicted assuming correct identification and persistence

²⁷ Third-ranked groups also contributed significantly less than RR subjects, according to these tests.

²⁸ We list the average contribution and average punishment index of each member of each intendedly homogeneous group in Appendix Table 2A.

²⁹ See Appendix Table 3A.

Figure 3

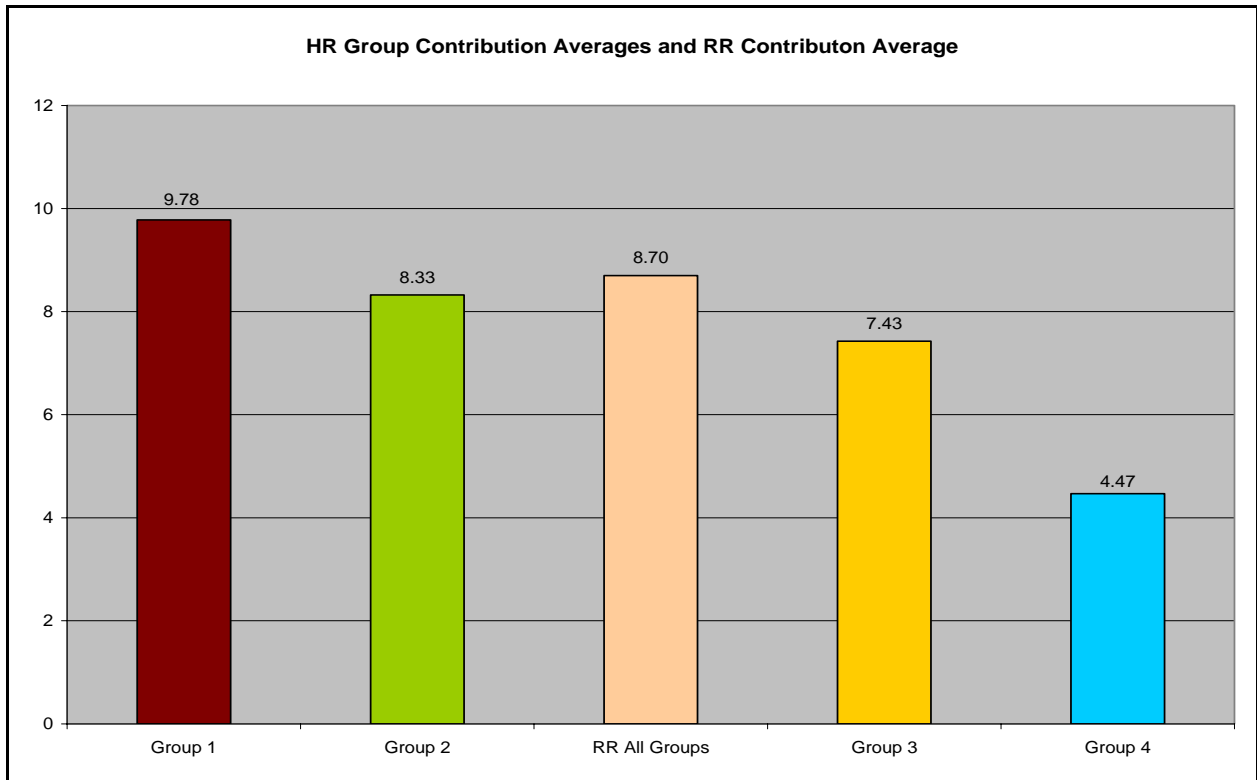
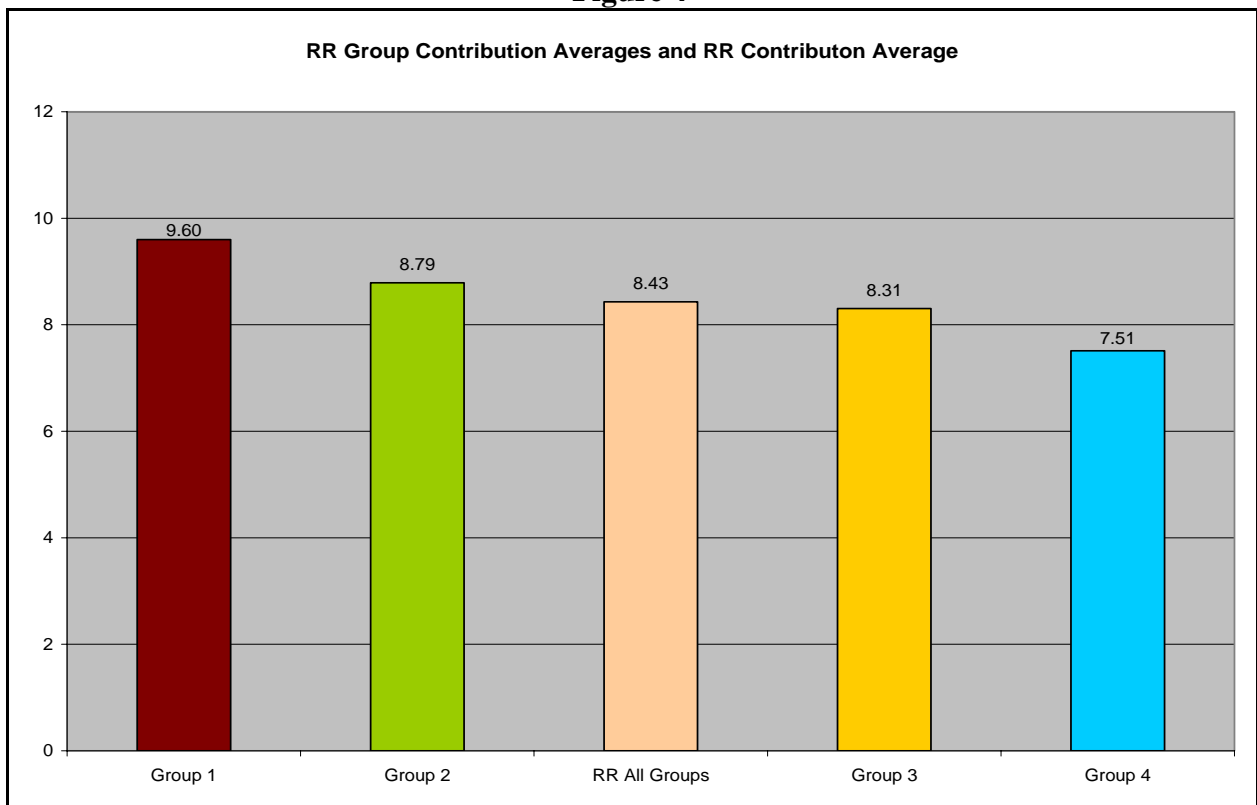


Figure 4



of types—this despite the fact that the inclination to contribute and to punish low contributors must in this case persist not only through the five 5 periods of play in heterogeneous groups but also through the next 10 periods of contribution and punishment decisions in randomly-formed groups.³⁰ Once again, the comparison with average contributions in the RR treatment meets basic expectations, although only the dominance of Group 1 over RR contributions is significant according to a Mann-Whitney test, and the differences among groups are in this case not significant according to a Kruskal-Wallis test.³¹

Persistence of type is also evident during the relevant periods as shown in Figures 1 and 2, respectively. There are more periods during which the “ideal” group order is violated in the RH than in the HR treatment, which is perhaps not surprising given the longer time lapse between classification (after period 5) and homogeneous grouping (beginning period 16). Nonetheless, in both treatments, average Group 1 contributions exceed average Group 4 contributions in every period.³²

Result 6: *During the homogeneous grouping periods of both the HR and RH treatments, probabilities that low and high contributors will receive punishment are also broadly ordered as would be expected assuming correct identification and persistence of tendencies to punish.*

Table 5 shows for each group during the periods of homogeneous grouping the ratio of the number of times a low contributor (contributor of less than the group average) was punished by another group member to the total possible number of such events—dubbed “normal punishment”—and likewise for punishment of contributors of more than the group average, dubbed “perverse punishment.”³³ In both treatments, the proportion of subjects who chose to punish a low contributor when that was possible was highest, as would be expected if there were

³⁰ The exact composition of the randomly-formed groups in terms of the diagnostic period average contribution levels and punishment indices of the group members can be seen in Appendix Table 2B.

³¹ For the Kruskal-Wallis test, see Appendix 3B. The Mann-Whitney tests discussed in this paragraph and the previous one take the behaviors of each group of 4 averaged over ten periods as its units of observation. More details are provided in our working paper.

³² An especially stringent test of the ordering effect entails checking the rank order of contributions by groups 1, 2, 3 and 4 in each individual session. See our working paper, in which we show that the ordering of groups by ranking corresponds to the ordering by actual contributions at the individual session level significant at the 0.1% level during the homogeneous grouping periods in the HR treatment, but that the correspondence is not statistically significant at the session level for the homogeneous grouping periods of the RH treatment.

³³ Each time a group’s four members contributed different amounts, there was at least one contributor of less than the average and at least one contributor of more than the average each of whom could be punished by up to three other group members. The denominator thus counts as 3 the number of possible punishments for each such instance while the numerator includes the number of fellow group members who in fact punished and targeted some punishment at the subject in question, that is 0, 1, 2 or 3.

persistence, in Group 1—which was selected partly based on its members’ propensity to punish low contributors during periods 1 – 5. Correspondingly, the proportion that chose to punish a high contributor when that was possible was highest in Group 4—likewise partly selected for such behaviors during periods 1 – 5. The orderings by group are imperfect, but in line with what would be expected in enough respects to conclude that broadly speaking—e.g., comparing groups 1 to groups 4—the behaviors displayed in periods 1 – 5 are somewhat predictive of behaviors in periods 6 – 15 of the HR treatment and periods 16 – 25 of the RH treatment.³⁴ Further evidence of the persistence of tendencies to punish is given at the individual level in Result 9.

HR	Group1	Group2	Group3	Group4
Normal Punishment	10/33 0.43	16 / 90 0.18	55 / 156 0.35	44 / 234 0.19
Perverse Punishment	2 / 54 0.04	2 / 174 0.01	21 / 264 0.08	43 / 234 0.18
RH				
Normal Punishment	23 / 33 0.70	10 / 66 0.15	22 / 105 0.21	55 / 129 0.43
Perverse Punishment	0 / 87 0.00	0 / 92 0.00	6 / 183 0.03	45 / 207 0.22

Table 5. Ratio of actual to possible punishment events during periods of homogeneous grouping, by group and treatment. The number at left on the first line is the number of times any individual *i* punished any individual *j*, while the number on the right is the total number of times such an event could have occurred, and the number on the line below shows the ratio in decimal form, for convenience. If *j* contributed less (more) than the average contributed by other group members in the period in question, up to three normal (perverse) punishment events are possible. Cases in which *j*’s contribution equaled the average (e.g., all contributed the same amount) are not counted.

³⁴ Another way of investigating differences in punishment behaviors is to estimate regression equations like those of Table 2 using the data of only Group 1 subjects, of only Group 2 subjects, etc. Results of such an exercise are shown in Appendix Table 4. Generally, the differences in the effects of deviations on punishment are not as great as might be expected: absolute negative deviation has a positive effect in all groups, significant except in HR Group 4, while the effect of positive deviations is insignificant in most cases. There are, nonetheless, some differences which are in line with our expectation of more vigorous normal punishment in lower-numbered groups and more perverse punishment in higher-numbered groups: in particular, (1) the one case in which negative deviations do not seem to be punished significantly is in Group 4 of HR; (2) the point estimates of the coefficient on negative deviation are higher in Group 1 than Group 4 in both treatments, and (3) the coefficient on positive deviation is significant and negative for Group 1 of HR (contributing more means being punished less, even given above-average contributions) but significant and positive for Group 3 of that treatment. The main exception to expectations is a significant positive coefficient on positive deviation in Group 1 of HR, a group that would have been expected to have little or no perverse punishment. Investigation shows that that result is due to a single subject acting in only one period inconsistently with the remainder of her behavior. Leaving out the punishments given by that subject in that one period, the coefficient becomes negative and significant at the 10% level, qualitatively matching the result for Group 1 of the RH treatment. (Because this might raise concerns about Table 2’s regression for the HR treatment as a whole, we rechecked that regression and found that exclusion of the same observations makes no qualitative difference: the coefficient on absolute positive deviation becomes a little smaller in absolute value but is still positive and significant at the 5% level.)

5. Results, Part III: “Type” and experience at the individual level

Having shown at the group level that there are differences in outcomes that are consistent with the successful identification and persistence of subject types, we now examine the matter more closely at the individual level. We note at the outset that persistence of early-exhibited tendencies does not rule out that experience in the course of play may also affect behaviors. We attempt to study the roles both of type and of experience in results 7 - 10.

Result 7. *Contributions during the diagnostic periods are a significant predictor of contributions in later periods in the HR and RH treatments, although there is a last period anomaly for RH.*

Table 6 summarizes a set of regressions at the individual level in each of which the subject’s average contribution during diagnostic periods 1-5 is the sole explanatory variable apart from a constant and group fixed effects (not shown).³⁵ The dependent variables, from left to right, are the same subject’s average contribution in periods 6-15, in period 16 alone, in periods 16-20, in periods 21-24, and in period 25.

The regressions for HR treatment subjects show that own contributions of periods 1-5 positively predict own later contributions, significant at the 1% level, even after controlling for group effects, except in period 25. In that period, the effect of own initial

		Ci,6-15	Ci,16	Ci,16-20	Ci,21-24	Ci,25
HR	Ci,1-5	0.633 ***	0.579 **	0.357 ***	0.425 ***	0.235
	Std.Err	(0.14)	(0.22)	(0.13)	(0.08)	(0.18)
	No. obs.					
	R ²	0.811	0.371	0.712	0.842	0.449
RH	Ci,1-5	0.398 ***	0.361	-0.156	0.048	-1.522 **
	Std.Err	(0.11)	(0.53)	(0.32)	(0.34)	(0.60)
	No. obs.					
	R ²	0.668	0.374	0.673	0.691	0.419

Table 6. Own early contributions as a predictor of own later contributions.

Note: OLS regressions with robust standard errors. Each regression included a constant and group dummy variables, not shown. * = significant at 10% level, ** = signif. at 5% level, *** = signif. at 1% level.

contributions falls short of significance at the 10% level, although 45% of the overall variance is explained by it in combination with group fixed effects, according to the regression R-squared.

For RH treatment subjects, the coefficient on own period 1-5 contribution is significant only for

³⁵ As before, subjects are assigned the group dummy of the group to which they belonged during the periods covered by the dependent variable.

period 6-15 contributions. That result, significant at the 1% level, suggests persistence of type, since it occurs despite the subjects having been placed in a group of players selected for their heterogeneity during the first five periods and then playing periods 6-15 among what are also generally heterogeneous groups, randomly selected. The insignificant results for the other periods and the unexpected significant negative result for period 25 clearly fail to support expectations of persistence. It is necessary to recall, however, that our experimental design, in which punishment is possible in all periods including the last, never affords us a clear view of subjects' inclinations to cooperate independent of their beliefs about the likelihood and probable severity of punishment. While a definite theoretical prediction can be made about last period *punishing*, if types persist, no such prediction can be made about *contributions*. We test for persistence of punishing behaviors in Result 9, but first we investigate how experience and type combine to explain contribution decisions.

Result 8. *The combination of own “type” measures with measures of experience with others (“environment”) adds, in some cases substantially, to the explained portion of the variation in late contributions by individuals.*

Table 6 reports a series of regression equations in which a subject's own contributions in various periods are the variables to be explained, while that subject's later contributions, the contributions of other group members, the subject's history of punishment, and group dummies, are included as explanatory variables.³⁶ The first five regressions are for HR subjects and the second five for RH subjects. The first column for each treatment tests whether the initial diagnosis of type based on first period decisions, on which the period 2 – 5 assignment is based, actually persists into those four periods. In the HR and RH treatments, first contribution is a positive predictor of period 2 – 5 average contribution, significant at the 5% and 1% level respectively, despite the heterogeneous decisions of the others with whom subjects were grouped. The average contribution of others (subscript $-i$) in one's initial group shows no significant effect.

³⁶ As before, subjects are assigned the dummy variables of the group they belonged to during the periods in which the actions of the dependent variable were taken.

	HR					RH				
	Ci,2t5	Ci,6-15	Ci,16	Ci,21-24	Ci,25	Ci,2-5	Ci,6-15	Ci,16	Ci,21-24	Ci,25
Constant	4.847 (3.39)	1.062 (1.62)	-0.587 (2.44)	-0.721 (1.01)	-2.217 (4.00)	6.243 ** (2.38)	5.623 ** (2.88)	2.205 (8.19)	70.528 (60.00)	23.089 (14.46)
Ci,1	0.465 ** (0.18)					0.200 *** (0.07)				
C-i,1	-0.229 (0.48)					0.070 (0.21)				
Ci,1-5		0.591 *** (0.14)	-0.200 (0.18)	0.215 (0.15)	0.095 (0.44)		0.404*** (0.12)	-0.107 (0.38)	0.493 * (0.28)	-1.656 ** (0.67)
C-i,1-5		0.255 ** (0.12)	-0.003 (0.22)	-0.067 (0.13)	-0.100 (0.45)		0.024 (0.23)	-0.228 (0.75)	-0.711 (0.42)	-1.605 (1.12)
Ci,6-15			1.232 *** (0.16)	0.056 (0.12)	0.300 (0.45)			0.830 ** (0.36)	-0.326 (0.23)	1.145 (0.80)
C-i,6-15			-0.157 (0.13)	-0.117 (0.08)	-0.132 (0.25)			0.131 (0.36)	0.387 * (0.20)	0.110 (0.88)
Ci,16-20				0.458 *** (0.15)					-1.012 (1.49)	
C-i,16-20				0.338 ** (0.15)					-4.932 (4.51)	
Ci,16-24					0.498 (0.54)					0.411 (0.33)
C-i,16-24					0.467 (0.44)					-0.027 (0.50)
PH,1-15			0.169 (0.66)					0.524 (0.40)		
PL,1-15			0.545 (0.42)					-0.627 (0.89)		
PH,1-20				-0.010 (0.30)					-0.367 (0.41)	
PL,1-20				-0.427 (0.36)					1.457 ** (0.69)	
PH,1-24					0.931 (0.84)					-1.013 (1.60)
PL,1-24					0.729 (0.97)					2.303 (2.43)
No. obs.	64	64	64	64	64	64	64	64	64	64
R²	0.566	0.824	0.746	0.902	0.536	0.383	0.669	0.546	0.813	0.495

Table 7. “Type” and “Environment” as Predictors of Later Contributions

Note: Regressions with group fixed effects (not shown) and adjusted standard errors. * = significant at 10% level, ** = signif. at 5% level, *** = signif. at 1% level.

The second regression for each group resembles the first column of Table 6 except that the average contributions of others’ in one’s period 1 and periods 2 – 5 groups are entered along with one’s own average contribution in those periods to explain own average contribution during the third grouping, periods 6 – 15. For both the HR treatment, in which subjects are grouped

with like others in those periods, and the RH treatment, in which periods 6 – 15 are played in a random grouping, own period 1 – 5 contribution has a highly significant positive coefficient. The coefficient on others' contributions is positive and significant in the HR but not significant in the RH regression.

The remaining three regressions for each treatment attempt to explain own later contributions as a function of both (a) own and group members' earlier contributions and (b) experience as a recipient of punishment. The variables labeled PH (for "punished high") measure the average dollars of punishment that i received per period when contributing as much or more than the average contributed by others in his or her group, while those labeled PL ("punished low") measure the average dollars of punishment that i received per period when contributing less than the average contributed by those others. These averages are calculated for periods 1 – 15, periods 1 – 20, or periods 1 – 24, always terminating before the period or periods whose contributions are being explained. The results show that, except in period 25, at least one of the own past contribution terms is a significant positive predictor of one's later contribution(s), with more recent periods being significant more often than are earlier ones. Usually, the contributions of others in one's groups do not exert an independently significant effect, although the effect shown is a positive one in the two cases that are significant. Only one of sixteen coefficients on the punishment terms, this being for the RH treatment and of the predicted sign (more past punishment when a low contributor leads to higher later contributions), achieves statistical significance. However, the R-squares of most of the regressions are somewhat higher in Table 7 than in the corresponding specification in Table 6, suggesting that interactions with others explain more of the variance in own later contributions than does own initial contributions.

For period 25, the overall explanatory power of the regressions are lower, as in Table 6. The otherwise unexpected significant negative coefficient on own earliest contributions appears again for the RH treatment, almost identical to the Table 6 result. All other own contribution terms are positive, however, and for RH, the signs on the coefficients of past punishment received are as expected, although neither achieves significance. The estimate for PL, although imprecise, suggests that for each additional dollar of average past punishment for free riding, final period contribution rose by 2.3 dollars.³⁷

³⁷ Since the punishment variable measures the average punishment per period over 24 past periods, the estimate implies that *total* past punishment would have to be $E\$24$ higher to induce a $E\$2.30$ increase in period 25 contribution.

Result 9. *Individuals' tendencies to punish low or high contributors are persistent into final periods, suggesting that their choices on punishment are based on stable preferences rather than strategic calculation or chance.*

Since there is no possible influence on others' future choices, decisions to reduce others' earnings at a cost to oneself in a final period of play should in principle reflect only underlying preferences. To test whether the tendency to punish low or high contributors in early periods (when future interactions are expected) persisted into later periods (in which groups' interactions were ending), we estimated a series of regressions in which the individual's average punishment indices in periods 2 – 4 are an explanatory variable and the same individuals' punishment indices in period 5, period 15, period 25, or their average punishment index in all three of those periods, is the dependent variable.³⁸ We also included constants and group dummy variables in the regressions.³⁹ Table 8 shows the results, which indicate that the early tendency to punish low (high) contributors is in all cases positively related to the tendency to do the same in groups' last periods, including the known last period of the experiment as a whole.⁴⁰ Period 2 – 4 punishment behavior is a statistically significant predictor of both period 5 and period 25 punishment behaviors for both HR treatment and RH treatment subjects, and punishment behavior in the three end periods (5, 15 and 25) averaged together is significantly predicted by early punishment behavior for the HR treatment (and in alternate specifications, not shown, for the RH treatment as well).⁴¹ This is fairly strong evidence that the tendency to punish was not only persistent but also that it reflected a genuine taste—the punisher prefers to give up some

³⁸ Periods 2 – 4 are chosen because these were periods of diagnostic play, on which subjects' classifications for later homogeneous grouping were based, but unlike periods 1 and 5, they were also periods in which at least one future interaction with the same players would be taking place.

³⁹ As before, each individual is assigned the group dummy corresponding to the group with which he or she interacted during the period corresponding to the dependent variable; for example, for period 15 regression, it is the group of periods 6 – 15 that applies, and so on.

⁴⁰ In periods 5 and 15 subjects knew that there would be additional periods of play in possibly different groups, which did not entirely rule out the possibility that some members of a subject's future group would have belonged to her current group also. The cleanest test of a pure preference for punishment is thus that focusing on period 25, which subjects knew to be the last in their sessions.

⁴¹ Since individuals belonged to a different group in period 5 than in period 15 and to still another group in period 25, we would have had to use 48 group dummy variables to cover all possible group influences on subjects' choices. In order not to reduce degrees of freedom to this extent, we chose to estimate the regressions for average behavior in periods 5, 15 and 25 using only one set of dummy variables. The estimates shown in Table 7 use the period 25 groups as control. We also separately estimated these regressions once assigning group dummies according to period 5 group membership and once assigning dummies according to period 15 group membership. For both HR and RH samples, the coefficient on the period 2 – 4 index was positive and significant at the 5% level or better in all four of these additional estimates.

income if doing so causes the person targeted to lose even more—rather than being mainly calculated to deter others from future free-riding.⁴²

	HR				RH			
	PunIn5	PunIn15	PunIn25	PunInLast*	PunIn5	PunIn15	PunIn25	PunInLast*
Constant	3.668 ** (1.76)	-0.920 (0.94)	0.598 (0.64)	0.650 (0.50)	-0.401 (0.82)	2.096 ** (0.80)	0.436 (0.64)	0.436 (0.54)
Avg. PunIn 2-4	0.661 ** (0.28)	0.136 (0.16)	0.448 * (0.23)	0.394 ** (0.16)	1.026 ** (0.39)	0.777 (0.48)	0.748 ** (0.31)	0.352 (0.39)
No. obs.	64	64	64	64	64	64	64	64
R²	0.485	0.306	0.401	0.494	0.404	0.355	0.313	0.358

Table 8. End period(s) punishment index as a function of average punishment index for periods 2 – 4. Regressions with group fixed effects (not shown) and adjusted standard errors. PunInLast is the average of the individual’s punishment indices for periods 5, 15 and 25. * = significant at 10% level, ** = signif. at 5% level, *** = signif. at 1% level.

***Result 10.** The tendency to contribute or not to a public good and the tendency to punish or not free riders or high contributors, are positively related for subjects in both the HR and RH treatments, with the relationship being statistically significant in the HR treatment.*

Implicit in our design choice of grouping subjects with equal weight on contribution and punishment tendencies was the plausibility of the assumption of FG and others that the propensity to contribute if others do so and the propensity to punish free riders go together, being two facets of a single preference called “reciprocity.” In Section 1, however, we noted that we see as an open matter, to be investigated empirically, how closely the strengths of the two tendencies are correlated. Table 9 reports the results of estimating regression equations in which the individual’s average contribution during periods 1 – 25 is the dependent variable and the individual’s average punishment index for periods 1 – 25 is an explanatory variable.⁴³ The results show a significant positive relationship between the two behaviors among HR subjects and an insignificant positive relationship among RH and RR subjects. The R² statistic is close to zero in the RH and RR regressions and only 0.09 in the HR regression, indicating that even where the relationship is statistically significant, it explains less than 10% of the variance in

⁴² Further evidence of persistence is found in the fact that a total of 11 subjects punished low contributors in period 25 in the HR treatment, and of these, six had been assigned to group 1 (high contributors, normal punishers), three to group 2 (next highest), and none to group 4 (low contributors, non- and perverse-punishers). Equally remarkably, there were only 3 instances of a low contributor punishing a high one in period 25 of the HR treatment, and 2 such instances in period 25 of the RH treatment. In every case, the punisher had been assigned to group 4 based on period 1 – 5 behaviors.

⁴³ No dummy variables are used because there is only one observation per individual and the 64 group dummies that would have been required to control for each of the four groups that each individual could be assigned to in each session would cut too deeply into degrees of freedom, while the expedient adopted in Table 8 (see footnote 41) is less applicable here.

contributions. Thus, our theoretical conjecture that the taste for positive reciprocity and the taste for negative reciprocity may not be perfectly correlated receives support from our data. Although the notion that both tastes are aspects of an overall trait of reciprocity is not entirely invalidated,⁴⁴ it would be interesting to design future grouping experiments that make more use of the imperfect correlation between the traits.

Dependent Variable: Contribution Averages Period 1 to 25	HR	RH	RR
Constant	7.115 *** (0.44)	8.712 *** (0.19)	8.262 *** (0.31)
Punishment Index in Periods 1 to 25	1.385 ** (0.92)	0.528 (0.46)	0.411 (0.30)
Number of Observations	64	64	64
R²	0.091	0.000	0.020

(*) Significant in 10 %
 (**) Significant in 5 %
 (***) Significant in 1 %

Table 9. Contribution behavior as a function of punishing behavior.

Note: OLS regressions with adjusted standard errors.

6. Discussion and Conclusion

This paper investigates the hypothesis that people differ in persistent ways in their inclinations to support collective action, and that the way in which a given group of people will respond to a collective action problem or social dilemma therefore depends partly on the types of people who comprise the group. Some people are more willing than others to cooperate provided they see others doing so also; some are more inclined than others to punish non-cooperation; a few are especially resistant to such punishment and actually punish cooperators.

We tested the proposition that individuals exhibit persistent differences by seeing whether we could identify types from behaviors in similarly heterogeneous initial groups. Having made our identifications, we placed subjects in groups of seemingly like type and confirmed that tendencies to contribute more or less to a public good persisted. In particular, the

⁴⁴ An important theoretical issue for biologists and social scientists is to understand the evolutionary relationship between the two tendencies. Whatever promoted an increase in the proportion of pro-social punishers in human populations would have favored propensities toward cooperation, while more cooperative populations would have reduced the relative costliness to the individual of being a pro-social punisher (Gintis *et al.*, 2005).

rank ordering of contributions among intendedly homogeneously formed groups matched the ordering of subjects by early contributions both when subjects played in homogeneous groups immediately after the diagnostic periods (in the HR treatment) and when they did so only considerably later in their sessions (in the RH treatment). Persistence of contribution tendency was also confirmed by individual-level regressions. Although initial inclinations or types evidently differed among subjects in ways that persisted, experience in the series of interactions also influenced later behaviors.

In addition to our findings about contributions, we found that early exhibited tendencies to punish free riders, to refrain from punishment, or to (perversely) punish high contributors, persisted into later periods, and in particular into periods in which any strategic motive for punishing was absent. Punishment of free riding was no less prevalent in the final period of play, supporting the notion that it reflects a taste (negative reciprocity). Because last period contributions could be motivated by fear of punishment while no such motive applies to punishing itself, our evidence on type persistence is most definitive where punishing is concerned.

The “social engineering” or “organizational design” implications of our experiment should be noted. Although we demonstrated that one can with a reasonable degree of replicability put together groups that will significantly exceed average levels of cooperation—something one might want to do, for instance, to create a more successful business partnership or team—this came, in our homogeneous grouping periods and especially in our HR treatment, at the cost of also creating some relatively uncooperative groups. The persistence of significantly higher average contributions in the RH and in the RR treatments suggest that for better average results, low contributors and punishers of low contributors should be put together, rather than squandering the efficiency-enhancing potential of the punishers by grouping them with already cooperative types. Whether one wants to achieve pockets of excellence or as good as possible an average result depends upon the problem at hand. For some purposes, the best approach seems likely to be to constitute as many groups as possible out of a mix of strong positive reciprocators, strong negative reciprocators, and more neutral or payoff maximizing types while isolating the few strongly perverse punishers in groups that must either be treated as a “lost cause,” or policed

by some external mechanism.⁴⁵ Other mechanisms for limiting the negative impact of anti-social types can also be designed.⁴⁶

We conclude that understanding agent heterogeneity is important to understanding and improving the solution of collective action problems. Cooperation doesn't decline with time in public goods experiments because a representative agent, a payoff maximizer, learns the iterated dominant solution with experience. Instead, cooperation usually declines when there is no way to control group membership or punish free riding, because in such a situation more cooperative subjects find no other way to protect themselves from free riding. If cooperative subjects teamed up with strategic-minded payoff maximizers can exclude or punish free riders, high cooperation can be sustained at least until the final periods of interaction. In the real world, where there is rarely a known and commonly shared last period, this may be an adequate solution to many problems.

⁴⁵ In a sense, this is exactly what societies do when they consign their most anti-social individuals to prisons! For an experiment in which free-riders are ejected from the main body of subjects leading to highly cooperative performance by those who remain, see Cinyabuguma, Page and Putterman (forthcoming).

⁴⁶ See Ertan, Page and Putterman for an experiment in which choice of rules by majority vote neutralizes the impact of the least cooperative subjects.

References

- Ahn, T.K., Elinor Ostrom, and James Walker. 2003. "Heterogeneous Preferences and Collective Action." *Public Choice*. 117: 295-314.
- Anderson, Christopher M. and Louis Putterman, forthcoming, "Do Non-strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism," *Games and Economic Behavior* (in press).
- Andreoni, James, 1988, "Why Free Ride? Strategies and Learning in Public Goods Experiments," *Journal of Public Economics* 37: 291-304.
- Andreoni, James, "Cooperation in Public-Goods Experiments: Kindness or Confusion?" *American Economic Review* 85 (4) Sept. 1995, 891-904.
- Ben-Ner, Avner and Louis Putterman, 1998, "Values and Institutions in Economic Analysis," pp. 3-72 in Ben-Ner and Putterman, eds., *Economics, Values and Organization*. New York: Cambridge University Press.
- Ben-Ner, Avner and Louis Putterman, 2002, "On Some Implications of Evolutionary Psychology for the Study of Preferences and Institutions," with Avner Ben-Ner, *Journal of Economic Behavior and Organization* 43: 91-99, 2000.
- Ben-Ner, Avner and Louis Putterman with Fanmin Kong and Dan Magan, 2004, "Reciprocity in a Two Part Dictator Game," *Journal of Economic Behavior and Organization* 53 (3): 333-52.
- Bochet, Olivier, Talbot Page and Louis Putterman, forthcoming, "Communication and Punishment in Voluntary Contribution Experiments," *Journal of Economic Behavior and Organization* (in press).
- Boyd, Robert and Peter Richerson, 1985, *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- Boyd, Robert and Peter Richerson, 2002, "Group Beneficial Norms can Spread Rapidly in a Cultural Population," *Journal of Theoretical Biology* 215: 287-96.
- Carpenter, Jeffrey, 2003, "The Demand for Punishment," unpublished paper, Department of Economics, Middlebury College, January.
- Carpenter, Jeffrey and Peter Matthews, 2002, "Social reciprocity," Middlebury College Department of Economics Working Paper #29.
- Charness, Gary and Matthew Rabin, 2002, "Understanding Social Preferences with Simple Tets," *Quarterly Journal of Economics* 117 (3): 817-69.
- Cinyabuguma, Matthias, Talbot Page and Louis Putterman, forthcoming, "Cooperation Under the Threat of Expulsion in a Public Goods Experiment," *Journal of Public Economics* (in press).

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

Cosmides, Leda and John Tooby, 1989, "Evolutionary Psychology and the Generation of Culture, Part II, Case Study: A Computational Theory of Social Exchange," *Ethology and Sociobiology* 10: 51-97.

Cox, James and Daniel Friedman, 2002, "A Tractable Model of Reciprocity and Fairness," working paper, Department of Economics, University of California Santa Cruz.

Davis, Douglas D. and Charles A. Holt, 1993, *Experimental Economics*. Princeton: Princeton University Press.

Durham, William H., 1991. *Coevolution: Genes, Culture, and Human Diversity*. Stanford, CA: Stanford University Press.

Falk, Armin, Ernst Fehr, and Urs Fischbacher, 2001, "Driving Forces of Informal Sanctions," Working Paper No. 59, Institute for Empirical Research in Economics, University of Zurich, September.

Fehr, Ernst, Simon Gächter, and Georg Krichsteiger, 1997, "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica* 65(4): 833-60.

Fehr, Ernst and Simon Gächter, 1998, "How Effective are Trust- and Reciprocity-Based Incentives?" pp. 337-63 in A. Ben-Ner and L. Putterman, eds., *Economics, Values and Organization*. New York: Cambridge University Press.

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

Fehr, Ernst and Simon Gächter, 2000b, "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives* 14 (3): 159-81.

Fehr, Ernst and Simon Gächter, 2002, "Altruistic Punishment in Humans," *Nature* 415: 137-40.

Fischbacher, Urs, Simon Gächter and Ernst Fehr, 2001, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Economics Letters* 71: 397-404.

Gintis, Herbert, 2000, "Strong Reciprocity and Human Sociality," *Journal of Theoretical Biology* 206: 169-179.

Gintis, Herbert, Samuel Bowles, Robert Boyd and Ernst Fehr, eds., 2005, *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*. Cambridge, MA: MIT Press.

Gunnthorsdottir, Anna, Daniel Houser, Kevin McCabe, and Holly Ameden, 2002, "Disposition, History and Contributions in a Public Goods Experiment," unpublished manuscript, Department of Economics and Economic Science Laboratory, University of Arizona.

Guttman, Joel, 2000, "On the Evolutionary Stability of Preferences for Reciprocity," *European Journal of Political Economy*, 16: 31-50.

Guttman, Joel, 2003, "Repeated Interaction and the Evolution of Preferences for Reciprocity," *Economic Journal* 113, no. 489: 631-656.

Henrich, Joseph and Robert Boyd, 2001, "Why People Punish Defectors: Weak Conformist Transmission can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas," *Journal of Theoretical Biology* 208: 78-89.

Hoffman, Elizabeth, Kevin McCabe and Vernon Smith, 1998, "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry* 36: 335-52.

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

Kurzban, Robert and Daniel Houser, 2001, "Individual Differences in Cooperation in a Circular Public Goods Game," *European Journal of Personality* 15 (S1): S37-S52.

Ledyard, John, 1995, "Public Goods: A Survey of Experimental Research," pp. 111-94 in John Kagel and Alvin Roth, eds., *Handbook of Experimental Economics*. Princeton: Princeton University Press.

Masclot, David, Charles Noussair, Steven Tucker and Marie-Claire Villeval, 2003, "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism," *American Economic Review* 93: 366-80.

McCabe, Kevin, Stephen Rassenti and Vernon Smith, 1996, "Game Theory and Reciprocity in Some Extensive Form Bargaining Games," *Proceedings of the National Academy of Science*, November, 13421-28.

Offerman, Theo, Joep Sonnemans and Arthur Schram, 1996, "Value Orientations, Expectations, and Voluntary Contributions in Public Goods," *Economic Journal* 106: 817-45.

Ones, Umut and Louis Putterman, 2004, "The Ecology of Collective Action: A Public Goods and Sanctions Experiment with Controlled Group Formation," Working Paper No. 2004-01, Department of Economics, Brown University.

Page, Talbot, Louis Putterman and Bulent Unel, forthcoming, "Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency," *Economic Journal* (in press).

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

Putterman, Louis with Matthias Cinyabuguma, Ioannis Garos and Theodore Marr, in process, "On Perverse Punishment in Decentralized Sanction Regimes," unpublished paper, Department of Economics, Brown University.

Rabin, Matthew, 1993, "Incorporating Fairness into Game Theory and Economics," *American Economic Review* 83: 1281-1302.

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

Schelling, Thomas, 1971, "On The Ecology of Micromotives," *The Public Interest* 25: 61-98.

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

Sethi, Rajiv and E. Somanathan, 2003, "Understanding Reciprocity," *Journal of Economic Behavior and Organization* 50(1): 1-27.