The Economic Journal, **115** (*October*), 1032–1053. © Royal Economic Society 2005. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

VOLUNTARY ASSOCIATION IN PUBLIC GOODS EXPERIMENTS: RECIPROCITY, MIMICRY AND EFFICIENCY*

Talbot Page, Louis Putterman and Bulent Unel

We find that a process of voluntary association where individuals express a preference about whom they want to be associated with can create strong incentives to increase efficiency and contributions in provision of a public good. This process of endogenous group formation perfectly sorted contributions by the order of group formation. Comparison of middle and last period behaviour suggests that a majority of the subject population are conditional cooperators, with a minority of monetary payoff maximisers. The experiment illustrates that under favourable conditions, where the opportunities of entry and exit are symmetrically balanced, a process of voluntary association can mitigate the free-rider problem.

In the filming of the Coen brothers' 'The Man Who Wasn't There', a new production company of three hundred people came into being. Nine months later, when the filming was finished, the company, 'Pompadour Pictures', disappeared. Parts of it regrouped into production companies for a next generation of films, along with individuals from other newly extinct production companies. This pattern, which is standard in the film industry, makes sense because the next production might call for a different mix of tasks and skills. The process of regrouping also makes sense in its incentives – individuals have an incentive to establish reputations for being productive and cooperative workers, giving them a chance to join better groups later. The more experienced and established people from a just liquidated production company, for example a director of the art department, are typically given slots to fill, and they fill them with the most promising assistants from the old production company.

The film industry is an extreme example of associations that form for a limited time and then are reconstituted. With varying degrees of voluntary entry and exit, and also different degrees of reliance upon internal sanctions against shirking group members, other examples include law partnerships, academic departments, athletic teams and virtually all firms. In response to the risk of free-riding, teams attempt to select members with good reputations and weed out others who fail to pull their weight. In practical life, endogenous group formation is a much used mechanism to mitigate free-rider problems.

The free-rider problem is a central problem in the theory of incentives and mechanism design (Green and Laffont, 1979). While standard theory and the

*The research reported here was supported by a grant from the MacArthur Foundation Norms and Preferences Network and N.S.F. grant SES-0001769. We thank Samuel Bowles, Robert Boyd, Ernst Fehr and Herbert Gintis for suggestions regarding research design. We thank George Athanassacopoulos, Yeoul Bae, Vicki Bogan, Wouter Buitenhuis, Eric Cao, Mark Faktorovich, Chris Horn, Robert Letzler, John Mello and Xiaotong Wang for their supporting roles on our research team. We thank Jeffrey Carpenter, James Walker and Martin Sefton for comments and suggestions, and Ernst Fehr and Simon Gächter for their numerous courtesies and kind encouragement. We thank the editor and referees for insightful questions and comments that helped us to improve the paper considerably.

elimination of dominated strategies predicts zero cooperation in a finitely repeated prisoners' dilemma, a Bayesian approach (Kreps *et al.*, 1982) shows that cooperation, in principle, is possible among self-interested expected utility maximising individuals if they believe that there may be (a few) others who are conditional cooperators and likely to increase their contributions in response to the self-interested maximisers' contributions. This suggests that there might be an incentive for self-interested maximisers to build a reputation of cooperation to increase the chance of being grouped with the potentially high contributing conditional cooperators.

This article analyses an experimental study of how people's ability to choose their associates affects free-riding in groups or organisations. We find the following main results for an *n*-person finitely repeated prisoners' dilemma with a periodic opportunity to choose new associates and regroup. The endogenous regrouping treatment:

(*i*) significantly increased efficiency and contributions to a public good, compared with a baseline treatment without regrouping
(*ii*) had significantly higher efficiency than a comparison treatment with monetary sanctions
(*iii*) perfectly sorted the groups in the order of their average levels of contribution to the public good
(*iv*) provided a surprisingly high estimate of the proportion of cooperative preference types in the experimental population.

There are two main contributions of the article. First, it demonstrates experimentally that when permitted to have a say over with whom they interact, individuals are significantly more likely to cooperate in a public goods game, and they sort themselves perfectly into groups displaying different degrees of cooperation. To our knowledge, this is the first experimental study to achieve these results. Second, the article contributes to a shift in perspective in the study of the free-rider problem, from solution concepts of subgame perfection and the successive elimination of dominated strategies toward a Bayesian view emphasising signalling, the building of reputation and heterogeneity of preference types.[1] The self-sorting of our subjects into higher and lower contributors, and the observed impact on last period behaviours in particular, suggest that a large fraction of subjects are conditionally willing to cooperate when they expect others to also do so, even though they lower their own payoffs in the process.

Much of the experimental literature studying the free-rider problem uses voluntary contribution mechanisms (VCMs), which are finitely repeated *n*-person prisoners' dilemmas where individuals have iterated dominant strategies to contribute nothing. But instead of no contributions, experimentalists found substantial contributions in the early periods, declining in the later periods.[2] These

[1] Other papers providing experimental support for heterogeneity of types include Fischbacher *et al.* (2001), Offerman *et al.* (1996), Park (2000) and Brosig (2002), the last three of which combine social psychologists' value orientation tests with public goods experiments and find correlations between measured value orientations and contributions to the public good.

[2] See Ledyard (1995) and Davis and Holt (1993) for reviews of the early literature.

results by themselves are consistent with iterated dominance, which can take time to learn. However, when added to the finding of a 'restart' effect when a second finitely repeated game is played (Andreoni, 1988; Isaac and Walker, 1988) and to findings of more cooperation in partner than in stranger groups (Croson, 1996; Keser and van Winden, 2000), the body of VCM findings have suggested Bayesian interpretations with signalling and reputation building to researchers. These interpretations are consistent with the theoretical work of Kreps *et al.* (1982) and Guttman (2003), who developed formal models of finitely repeated prisoners' dilemmas and computed sequential Bayesian equilibria in which individuals initially cooperated and then defected toward the end. McKelvey and Palfrey (1992) studied cooperation in the centipede game, undertaking both theoretical and experimental analysis. They computed Bayesian equilibria with a similar pattern of initial cooperation and bailing out toward the end, and this prediction was confirmed experimentally.

Bayesian interpretations depend on the existence of heterogeneous preference types, or at least the belief in the existence of heterogeneous types. Moreover, it appears that it is hard to explain the experimental results without an interaction of types, for example an interaction of payoff maximising types and conditionally cooperative types. For this reason, VCMs are increasingly recognised to be a useful tool for studying 'non-standard' preferences such as cooperation and altruism.

While most VCM experiments have assigned subjects to groups randomly, a few have adopted procedures that give explicit attention to who plays with whom. Gunnthorsdottir *et al.* (2002), Burlando and Guala (2005), and Ones and Putterman (forthcoming) found evidence of subject heterogeneity and its persistence by having the experimenter sort subjects into groups. Their studies used *exogenous* regrouping to identify and study the persistence of heterogeneous types. Our experiment differs from theirs by making the regrouping process *endogenous.*

Ehrhart and Keser (1999) were the first, to our knowledge, to use endogenous regrouping in a VCM experimentally, allowing subjects to move freely from one group to another. In their design, however, subjects could move unilaterally without agreement of those whose groups they join. They found unbridled chasing of high contributors by free riders, with unsustained sorting between high and low contributing groups, and a declining trend in contributions. In a public goods game played in groups of two, Coricelli *et al.* (2004) let subjects bid for the right to choose partners, in one treatment unidirectionally, in another bidirectionally, always in six sequences of five games. In their unilateral partner selection treatment they found, like Ehrhart and Keser, that free riders showed a tendency to chase high contributors, often succeeding by making the highest bids for the right to choose a partner. Coricelli *et al.* found significantly higher contributions in their unilateral partner selection treatment compared with random rematching, when the first five and the last five periods are excluded. Our study differs from Ehrhart and Keser's and from Coricelli *et al.*'s unidirectional treatment in our regrouping process, which provides enough protection to cooperators (high contributors) from free riders so that the two types are almost never grouped together unless the latter first work hard at feigning cooperation. Unlike Ehrhart and Keser, our mechanism maintains a constant group size and hence a constant incentive structure. Bohnet and Kübler (2005) attempted to sort cooperators from non-cooperators in two-person binary-choice prisoners' dilemma games by letting subjects bid for the right to play one of two games with differing payoffs.[3] Similar to Ehrhart and Keser, they found that sorting was not sustainable and that the auction price accordingly approached zero in the last period.

Changing group membership by regrouping (or by expelling some group members, see Cinyabuguma, *et al.* (2005)) is one way to mitigate free riding; another is to sanction free riders while they continue to be in the group. Fehr and Gächter (2000 and 2002), Ostrom *et al.* (1992), and others[4] have investigated the effects of monitoring and endogenous punishment in public goods and common pool resource games with fixed or randomly changing groups, with impressive findings. Given the opportunity, many subjects in Fehr and Gächter's modified VCM incurred costs to impose punishment on free-riders, and contributions accordingly rose rather than fell with repetition. The right to punish adds a degree of freedom to the original prisoners' dilemma, as does the power to form new groups in our experiment. To study the two forms of sanctioning – non-exclusionary punishment, and exclusion of free-riders – side by side and together we included in our experiment a treatment with punishment and a treatment that combines both punishment and endogenous group formation. We found that endogenous grouping and punishment behave similarly in their abilities to increase and to stem the fall-off in contributions, but, in our experiment, endogenous grouping is more efficient because there is a higher resource cost to punishment.

The remainder of the article proceeds as follows. Section 1 describes the design of our experiment; Section 2 discusses empirical results of the experiment; Section 3 concludes.

1. Experimental Design

In this Section we list our design goals, explain our interpretative framework, and then operationalise our experiment.

1.1. *Design Goals*

Our main goal was to identify conditions in which an endogenous process of group formation would induce sustained cooperation and sorting of subjects into those more versus those less cooperative in their behaviour. More specifically we incorporated the following design characteristics:

[3] A related contribution to the experimental prisoners' dilemma literature is Orbell and Dawes (1993).

[4] Other studies include Carpenter and Matthews (2002), Masclet *et al.* (2003), Sefton *et al.* (2002), and Bochet *et al.* (forthcoming).

- *A simple and symmetric regrouping process* We chose a regrouping process that was symmetric in the sense that each subject's preference regarding with whom to be grouped (and with whom not to be grouped) had equal weight.
- *Easily understood information* The subjects need some information on the others' behaviour in order to form preferences over whom to be grouped with in the next regrouping. After several pilots we decided on quite simple and limited information: the average of each of the other subjects' previous contributions.
- *An incentive structure that is invariant over regroupings* For this goal we fixed the group size over regroupings.
- *Comparison of the effects of regrouping with those from direct punishment* This goal led to a 2×2 design.

1.2. *The Interpretive Framework: Direct Tests and Indirect Evidence*

To highlight possible effects of signalling and sorting, we followed a quite standard idea in the VCM literature. The idea is to design the experiment so that if everyone is of a payoff maximising type, the Nash equilibrium will be for no one to contribute (i.e. complete free-riding) from the first period onward. This happens automatically in the baseline treatment because it is a finitely repeated prisoners' dilemma; it happens in the punishment treatment by making punishment costly; and it happens in the regrouping treatment by making the participation in the regrouping process (specifically, the ranking of prospective associates) costly. In each of the three cases the finitely repeated game unravels from the end. Thus, if we observe substantial amounts of contributions, punishment, participation in regrouping, and sorting by contribution level, these observations support a Bayesian interpretation, which includes signalling and building of reputations. Since such an interpretation depends on heterogeneous types, this provides indirect evidence of these types.

In much of the VCM literature, the observed results are interpreted in terms of two preference types: a payoff maximising type (*Homo economicus*) and what for the moment we will simply call a cooperator type (which may take various particular forms including conditional cooperator, altruist, and warm glow contributor). The idea is that in a population of both types, if a payoff maximising type believes there may be cooperator types in the population, it may be worthwhile for her to mimic a cooperator type, by contributing and signalling that she is a cooperator type, thereby increasing other cooperator types' contributions and indirectly benefiting herself. (Mimicking strategies are very common in Bayesian games.) In previous studies, theorists found that a quite small fraction of cooperators in the population could justify the mimicking behaviour among the payoff maximising types.

It appears that the experimental results below are most easily explained by the existence of conditional cooperator preference types and from time to time we will focus on this version of cooperation. However, because of the indirect and limited nature of the evidence, we do not want to make very strong claims concerning which specific preference type best explains the experimental observations below, and it is likely that several such preferences exist with shifting boundaries.[5]

Sorting by contribution levels over a series of regroupings is not direct evidence of sorting by preference type because payoff maximising mimickers and cooperator types may be grouped together. Experimentally, it can be hard to distinguish between cooperative and payoff maximising types mimicking cooperator types, since both types act in much the same way in the middle periods of the game. In contrast, last period behaviour can distinguish between mimickers and cooperators better than averages can over the whole repeated game. In the last period, payoff maximisers, having a dominant strategy to contribute nothing, have no further incentive to mimic, while conditional cooperators, depending on their beliefs about others' last period behaviour, may continue to contribute. The sorting of subjects by differences in contribution behaviour that occurs in the earlier periods of our experiment can translate into different expectations about fellow subjects' types in different groups and, if those in high contributing groups contribute to the public good even in the last period, there would be strong evidence of their conditional preference for cooperating.

So much for design considerations. In the actual experiment we developed a rough estimate of the fraction of the population that were conditional cooperators. This estimate was surprisingly high, compared with previous estimates.

1.3. *Operationalisation*

Table 1 summarises the 2 × 2 design of our experiment consisting of a baseline finitely repeated VCM treatment, a treatment with endogenous group formation, a treatment with punishment (sanctioning) opportunities, and a treatment with both endogenous group formation and punishment opportunities. The experiment as a whole consisted of 16 sessions involving a total of 256 subjects. In each experimental session, 16 inexperienced subjects drawn from the general undergraduate population at Brown University played 20 periods of a finitely repeated VCM in groups of four. Subjects did not know with whom they were grouped, and interacted only through computer terminals.[6]

[5] Other kinds of preferences that have been suggested as ways of explaining positive contributions in VCM experiments include inequality aversion, altruism and 'warm glow' from the act of giving independent of its effect on others. For work along these lines see Andreoni (1995), Falk *et al.* (2001), Palfrey and Prisbrey (1997) and Kurzban and Houser (2001). Our notion of conditional cooperation is roughly the same as what Hoffman *et al.* (1998) and Fehr and Gächter (2000) call reciprocity, except that we are agnostic about the extent to which positive and negative forms of reciprocity are associated in the same individuals.

[6] Subjects were recruited mainly through the distribution of flyers to the campus mailboxes of all Brown undergraduates, with the experiments being identified as being conducted by researchers in the Economics Department. A brief post-experiment debriefing questionnaire shows that 11% of the subjects were economics concentrators, almost identical to the percentage of graduates completing this concentration in 2000. Widely distributed over all classes from freshmen to seniors, 42% of subjects had taken one or more economics classes, the average number of economics courses taken being just over 1. 58% of the subjects were female, slightly higher than the share of females among Brown undergraduates as a whole.

Table 1

2 × 2 Design, with the Number of Sessions and Subjects in Each of the 4 Treatments

	No regrouping	Regrouping
No punishment	Baseline 4 sessions, 64 subjects	Regrouping 4 sessions, 64 subjects
Punishment	Punishment 4 sessions, 64 subjects	Combined 4 sessions, 64 subjects

1.3.1. *Baseline treatment*

We discuss the baseline treatment first, since the other treatments are built on the baseline design. In this treatment, groups are formed randomly in the first period and remain fixed for the 20 periods of an experimental session.[7] The members of one group have no information on the behaviours in other groups of the same session.

In each period, each subject was provided (electronically credited) with 10 experimental dollars, with one experimental dollar converting at the end of the session to $0.07 in real money. In the baseline treatment, the subject's only task was to divide her endowment, in integer amounts, between a private and a group account before learning the others' contributions. Money assigned to the private account became personal earnings, while an amount equal to 0.4 times the sum of contributions to the group account was distributed to each group member, regardless of what he or she contributed to that account – hence, its 'public good' character. The earnings of subject i in a given period were accordingly

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

where C_i is the amount i assigns (contributes) to the group account and the summation is taken over all members of i's group, i included.

In each period, after each subject assigned an amount to the group account, the computer informed everyone in each group how much each of the other members of the group contributed to the group account. To protect anonymity the other group members were identified by the letters B, C, and D, with the letters assigned randomly, and reassigned randomly each period. After learning the others' contribution levels, each subject was shown his or her earnings for the period.

[7] A referee suggested that randomly regrouping subjects every three periods would constitute a better baseline treatment, because new groups form every three periods in our endogenous regrouping treatment and the likely increases in contributions every three periods due to restart effects would lead to a more stringent test of the effectiveness of endogenous group formation. An opposing factor noted by another referee, however, is that partner groups have usually been found to sustain higher contributions than stranger groups, so the restart 'upticks' might be more than offset by a steeper declining trend in contributions. Thus, a baseline with random regroupings could make it *easier*, rather than harder, to show endogenous grouping's advantage.

At the end of the experiment, the sum of the earnings in all 20 periods was converted into real dollars and each subject was paid a $5 participation fee. Experimental sessions lasted from one to two hours, including an instruction and practice period. Real earnings, including the participation fee, averaged around $25.

1.3.2. *Regrouping treatment*

The regrouping treatment is the same as the baseline treatment, except that at the end of *each third period* (periods 3, 6, 9, 12, 15 and 18) there is a regrouping decision. Each subject is shown a list, without other identifying information and in a random order, of each of the other 15 subjects' average contribution to their group accounts over the experiment so far. Subjects were then given the opportunity to express a preference among possible future partners by ranking them according to the following procedure.

If a subject chose to rank others, she typed a number in a box next to the information about each other subject. Potential ranking numbers ran from 1 to 15, with 1 standing for the most preferred prospective partner. The same ranking number could be assigned to two or more subjects, allowing ties. If a subject did not choose to rank others, she clicked on a 'don't rank' box. In this case, the computer assigned the number 8 to every subject, interpreting the 'don't rank' decision as fifteen-way tie.[8] When all subjects had completed this process, the computer assigned subjects to groups by searching, first, for that group of four individuals the sum of whose mutual ranks of one another was the lowest among the universe of potential groups, then repeating this process over the remaining subjects, to form the second and third groups, leaving the last four subjects in the fourth group. After new groups were formed, subjects resumed play without information about whom they had been grouped with, a matter on which only indirect inference could be made from observing one's three partners' contributions once play had resumed.

Subjects were charged 25 experimental cents for the first ranking decision of a period and 5 experimental cents for each additional ranking decision that period. The instructions (see our working paper) were read to the subjects and shown on the computer screens before the experiment began, and the subjects worked several practice problems to make sure they understood the process.

Both the frequency of regrouping and the information provided were important design decisions. Ranking 15 other subjects is time consuming and we believed that if there were too many periods of ranking, the subjects might tire, losing focus. Our choice of ranking every third period was a compromise between having enough regroupings to reveal a regrouping effect if it was there and not too many to maintain interest among the subjects. The same consideration suggested providing simple and readily understood information. In a series of pilots we tried

[8] The rankings were scaled so that each person's rankings summed to $1 + 2 + \cdots + 15 = 120$, with ties set equal to the average ranking within the ties, for a constant sum of point votes of 120. For example, in a 2 way tie for the highest rank the first and second rankings would be 1.5 and 1.5.

differing amounts and forms of information, and we chose the average of past contributions because it led to the most cooperation.[9]

1.3.3. *Punishment treatment*

The punishment treatment is the same as the baseline treatment, except that at the end of *each period*, after everyone in a group learns the contribution levels of the others in their group, each subject has an opportunity to reduce the earnings of other group members at a cost of 25 experimental cents for each experimental dollar of reductions. In a given period, earnings of subject i after reductions are thus

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

where R_{ij} is the number of dollars by which i reduces j's income, and conversely for R_{ji}. In the punishment treatment as in the baseline, groups are formed randomly in the first period and remain fixed for the 20 periods.[10]

Punishment every period is a standard design in earlier experiments with a punishment treatment and, with only three others in one's group, the punishment decision for each of them each period is not over complicated or time consuming.

1.3.4. *Combined regrouping and punishment treatment*

In every period of the combined treatment there is a contribution stage, followed by a punishment stage, and at the end of every third period (after the punishment stage) there is a ranking and regrouping stage. In the ranking stage, subjects receive the same information as in the regrouping only treatment.[11] The combined treatment thus merges the main elements of the regrouping and punishment treatments, allowing us to check whether the effects of the two treatments are complementary (as in Ostrom *et al.*'s combination of 'swords' [punishment] and covenants [face-to-face communication]), redundant or conflicting.

[9] A referee suggested that it would have been better to show subjects information about others' contributions in the last three periods only, or in all past periods but disaggregated into three period blocks. The idea was that with a subject carrying her full past history into each ranking, it would be difficult to escape the reputation attained early on, so there would be less incentive to try to establish a good reputation and sorting would be based more on the accidents of early decisions than on true type. But in a pilot experiment in which only three periods of past contributions were shown to others at ranking time, we found that this led to short-run strategies in which some subjects contributed the maximum for three periods, got into a good group, engaged in free riding for three periods, then repeated the cycle, leading to lower contributions overall. We also found in a compromise pilot with eight periods of past contributions averaged at ranking time that contributions were lower than when the subjects were informed of the contributions averaged over all preceding periods. We eventually settled on the overall average contribution levels because they led to higher contributions. With regard to mobility, see footnote 18.

[10] As in the baseline and all other treatments, subjects were identified to one another only by letters *B*, *C* and *D* which switched randomly from period to period. This feature was adopted by us, following Fehr and Gächter, in order to reduce the opportunity for vendettas by preventing subjects from tracking the behaviour of other group members over time.

[11] In some pilot sessions, we showed subjects information both about one another's contribution behaviour and about one another's punishment behaviour at each ranking stage. Showing the punishment information seemed to have the effect of inducing subjects to punish free riders more, but rather than increasing efficiency, this actually reduced it, because the money spent by punishers and lost by targeted subjects was not fully offset by gains from increased contributions. To give the combined treatment its best shot, we therefore omitted the punishment information.

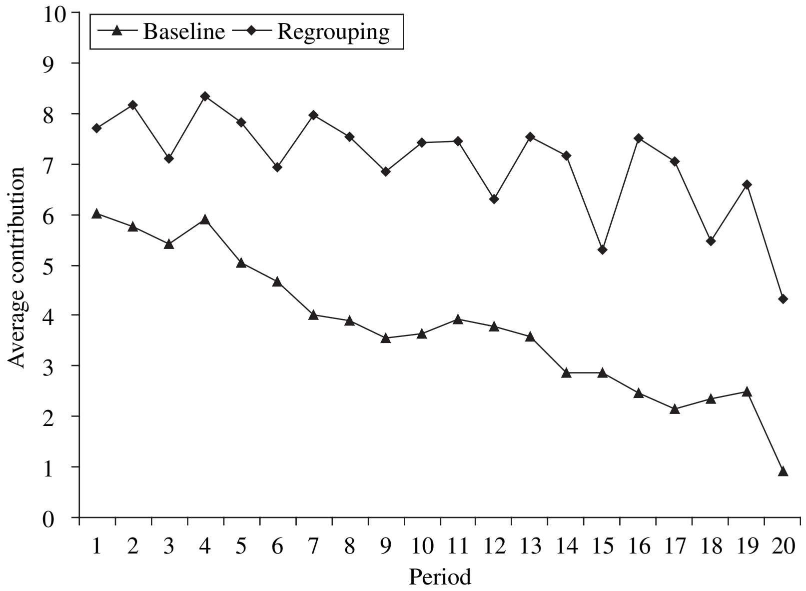

Fig. 1. *Average Contribution by Period in Baseline and Regrouping Treatments*

2. Results

2.1. *Efficiency and Contribution Levels*

Figure 1 compares average contributions of the regrouping treatment by period with those of the baseline treatment. The upper line shows that average contributions are higher with regrouping than without, decline more gradually and have an increase after each regrouping (at the end of periods 3, 6, 9, 12, 15, and 18) consistent with the restart effect mentioned in the Introduction. Over the course of the experiment as a whole, in the regrouping treatment the average contribution to the group account is 70% of endowment, compared with 38% in the baseline treatment, and there is a significant difference in contributions according to a Mann-Whitney test.[12] The contribution levels and their decline in the baseline

[12] We conducted this test two ways. The first way was to include one observation per partner group in the baseline treatment and one observation per 16-person experiment session in regrouping treatments. With 16 baseline observations and 4 regrouping observations, the Mann-Whitney test statistic for average contributions is 6, which has a p-value less than 0.01 (one-tailed test p-value is 0.006); the corresponding statistic for earnings is 7, with a p-value less than 0.01 (one-tailed test p-value is 0.008). In the second way we took each baseline experiment as a single observation. While this reduces the baseline sample to 4 observations only, it can be considered more appropriate because we then average the contributions of 16 different subjects when looking at both baseline and regrouping treatments. Thus we avoid differences in variance due to differences in numbers of subjects per group. (The reason we do not study groups of 4 separately for the regrouping treatment is that some subjects move from group to group and all subjects get some information about the others in their session at ranking stages. Thus actions in each group are potentially influenced by actions in other groups over the course of a session.) Under this alternative method, the Mann-Whitney test statistic for average contributions is 6, which has a p-value of 0.015 in a one-tailed test; the corresponding statistic for earnings is also 6, with the same one-tailed test p-value, 0.015.

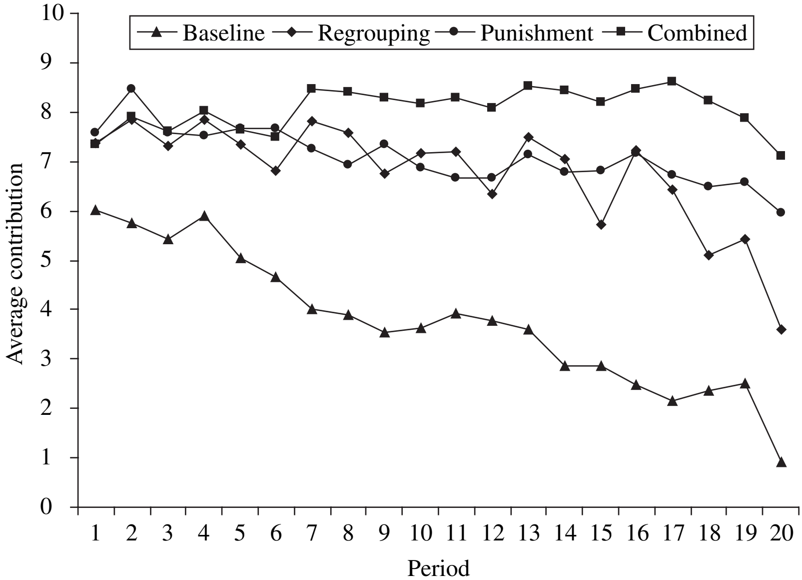

Fig. 2. *Average Contribution by Period in Baseline, Regrouping, Punishment and Combined Treatments*

are consistent with other baseline findings in earlier VCM experiments (Davis and Holt, 1993).

Figure 2 shows the average contributions of the regrouping treatment in comparison with the punishment treatment, the baseline treatment and the combined treatment (regrouping and punishment). Average contributions in the regrouping treatment (diamond marks) are about the same as in the punishment treatment (circular marks) but with a smaller last-period decline in contributions in the punishment treatment. Average contributions in the combined regrouping and punishment treatment (square marks) are higher than either the regrouping treatment or the punishment treatment alone but not significantly so. The contribution levels and their relative stability in the punishment treatment are similar to those reported in the literature,[13] and contributions are significantly higher than in the baseline, as in earlier studies. As in Fehr and Gächter, some subjects show a willingness to engage in costly punishment even in the last period of play, which has implications for our understanding of preferences, to which we return in Section 2.4.

Table 2 summarises the average contribution levels, the average earnings levels and the efficiency for each of the four treatments. Efficiency is higher in regrouping than in punishment (with a p-value of 0.055 for a one tailed Mann-Whitney test of the null hypothesis that there is no difference). The lower efficiency in the punishment treatment, compared with regrouping, is explained by differences in cost. Punishment has direct monetary costs for both the punisher

[13] Fehr and Gächter found a rising trend in contributions in their punishment treatment, but with a slightly lower starting point.

Table 2

Summary of the Average Contribution, Earnings and Efficiency by Treatment (efficiency is defined as observed earnings divided by the maximum possible earnings)

	Baseline	Regrouping	Punishment	Combined
Average contribution	3.8	7.0	7.1	8.1
Average earnings	12.3	14.1	12.9	13.8
Efficiency	77%	88%	81%	86%

and the punished, while ranking has direct monetary costs to only the persons doing the ranking.[14]

In sum, the subjects' abilities to influence with whom they are grouped has a demonstrable positive effect on cooperation and efficiency in our experiment. Both the opportunity to form new groups periodically on the basis of costly mutual rankings and the opportunity to reduce others' earnings at a cost, as well as the two together, raise contributions and stem their decay relative to a VCM without punishment or regrouping. In our experiment, adding punishment opportunities alone does not significantly increase earnings and thus efficiency, whereas either regrouping alone or regrouping with punishment does this. On a cautionary note, this comparison between regrouping and punishment depends upon the cost of punishment and other details of experimental design. But our results concerning punishment are similar to those of Fehr and Gächter (2000), who used a non-linear cost-of-punishment function, differing from the linear one we used and from other results in the literature.[15]

2.2. *Ranking and Sorting*

The above results suggest that there is an incentive effect from endogenous regrouping, which of course depends on individuals expressing their preference rankings. We found that 94% of the subjects chose to rank at some time, with an

[14] The punishment opportunity also allows subjects to reduce the earnings of high as well as low contributors and about 20% of punishments in our punishment treatment are aimed at high contributors. In a regression analysis we found that punishment of high contributors leads to lower contributions by them in subsequent periods and to lower efficiency. Similar frequencies of punishing high contributors are found by Ostrom *et al.* (1992), Fehr and Gächter (2000) and Falk *et al.* (2001).

[15] In Fehr and Gächter (2000), each punishment point reduces the pre-reduction earnings of the individual targeted by 10% and there is an increasing marginal cost of purchasing punishment points. By contrast, a fixed cost of punishment per unit loss to the person targeted is used in Fehr and Gächter (2002), Sefton *et al.* (2002), Bochet *et al.* (forthcoming) and Anderson and Putterman (forthcoming), as in the present article. We analysed the original data of Fehr and Gächter (2000), kindly provided to us by those authors, and found that while earnings tend to be higher with punishment than without punishment in later periods, averaged over all periods, earnings are lower in the punishment than in the baseline condition (for all treatments – partner, stranger, and perfect stranger), just as in our experiment. That earnings are lower, or else are not significantly higher with punishment than without, is also the finding of Sefton *et al.* (2002) and Carpenter and Matthews (2002). Anderson and Putterman (forthcoming) found that the demand for punishment varied predictably with its cost and that, when punishment was less costly and thus more utilised, contributions were higher but earnings lower. But in Casari and Plott (2003)'s common pool resource experiment, punishment is in the form of fines and thus a transfer rather than a deadweight loss, so that efficiency can approach 100%. Masclet *et al.* (2001) also increased efficiency with sanctions simply in the form of signalling disapproval, avoiding the cost of monetary punishment.

average of 79% choosing to rank in any given ranking period. The subjects tended to give lower (more favourable) rankings to those who had previously contributed more,[16] so that for each regrouping, subjects with the highest previous contributions tended to be grouped together in the first-formed group. Of the remaining subjects, those with highest previous contributions tended to be grouped together in the second-formed group and so on for the remaining subjects, down to the last-formed group, which tended to include those with the lowest previous contributions.[17] But mutual selection of one another by those with the highest past contributions to form the first group does not assure that the first-formed group will end up displaying the highest contributions, unless those who were high contributors in the past continue to be high contributors following new group formation. We checked whether this was the case by looking at the ordering of contributions by order of group formation.

Figure 3 shows the average contribution, in each period, in the first, second, third and fourth formed groups (aggregated over the four sessions) in the regrouping treatment. The first-formed groups, which change membership from time to time as low contributors are weeded out and higher contributors substituted,[18] show little or no decline in average contribution prior to the final regrouping, after which an

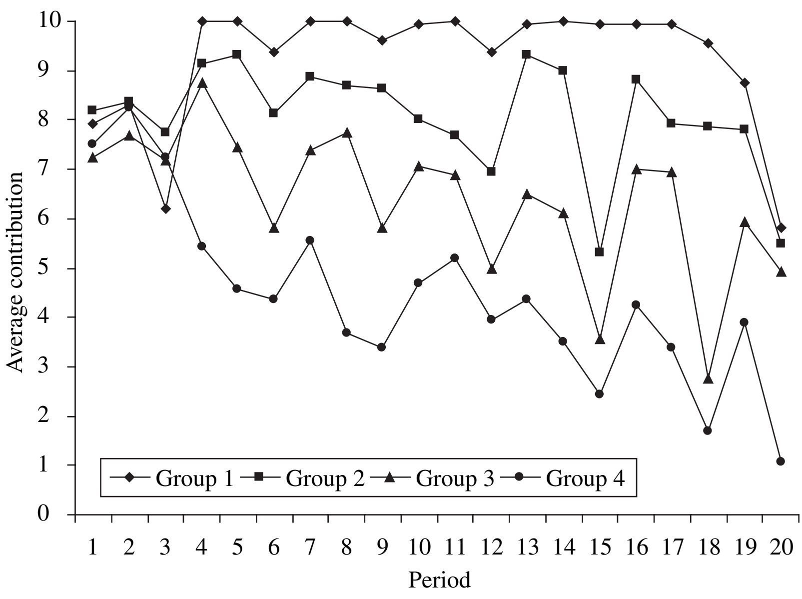

Fig. 3. *Average Contribution Disaggregated by Group in the Regrouping Treatment*

[16] Ordinary least squares regressions show that rankings are significantly negatively correlated with the past contributions displayed to the group account of the person being ranked. For each ranking stage, we regressed average ranking on average contribution across all individuals and found a statistically significant negative relationship in each case.

[17] There was a mild tendency for lower contributing subjects also to engage in less costly ranking of others, a tendency that strengthens as play proceeds, suggesting a recognition that they could not succeed in being grouped with the high contributors on whose contributions they would have preferred to free ride (because high contributors were unlikely to give them favourable rankings).

[18] 19% of subjects who started in first-formed groups, 37% of those who started in second groups, 50% of those who started in the third groups and 38% of those who started in the fourth groups moved to groups of other ranks during their sessions.

end-game effect appears. The second and third-formed groups show some decline but prior to the final regrouping much of the decline in Figure 1 for the regrouping treatment is attributable to the fourth-formed groups. Notice also that the order of average contributions coincides with the order of group formation in every single period from period 4 to period 20.[19]

A formal test of endogenous sorting by contribution behaviour is as follows. For each session of each of the two treatments where there is regrouping, we aggregated the average contributions of the first, second, third and fourth formed group over the 17 periods after the first regrouping begins (periods 4–20). In the regrouping treatment, in each of the four sessions the order of average contributions is the same as the order of group formation (see Figure 4).[20] The probability of observing this perfect sorting under the null hypothesis that group formation was random is less than 0.000003.[21] In the combined treatment, sorting is not perfect, but fairly close. For this treatment the p-value is less than 0.004.[22]

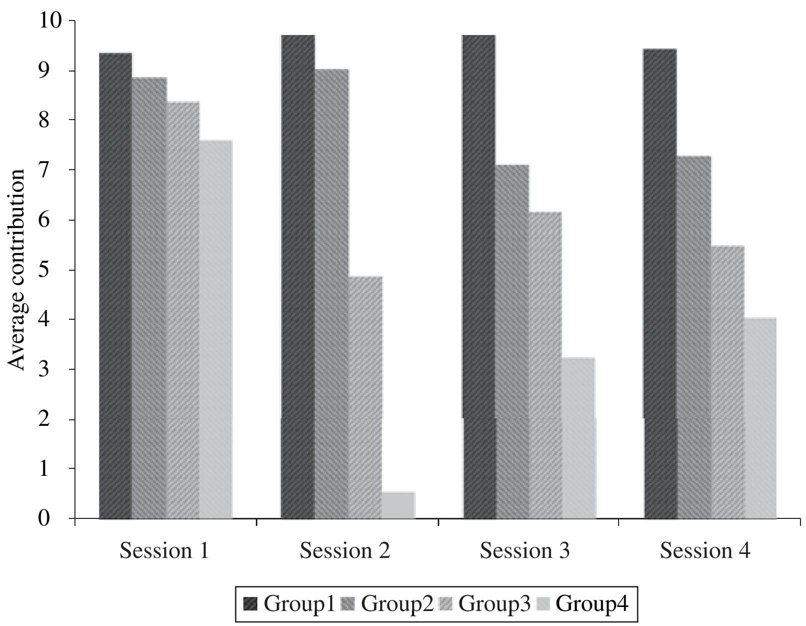

Fig. 4. *Average Contribution by Group Formation Order, Grouped by Session*

[19] Recall that in the regrouping treatment, group membership is assigned randomly for periods 1–3.

[20] More specifically, to find the height of the left-most bar in the graph, we calculated the average contribution in the first formed group of periods 4–6 in session 1, the average contribution in the (possibly different) first-formed group of periods 7–9 in session 1 etc., then averaged over the averages for the six sets of periods (4–6,7–9,10–12,13–15,16–18,19–20) in session 1 and similarly in the other groups and sessions.

[21] 24 orderings of height are possible for each group of four bars. The probability that a perfect tall-to-short ordering would occur by chance on a single trial (session) is therefore 1/24. The probability that this would happen on four out of four trials is $(1/24)^4$, or just under 0.000003.

[22] Two of the combined treatment sessions had perfect sorting of contributions, the third session had a single tie, and the fourth had two single step exchange permutations. By this we mean the ordering for the first session was 1,2,3,4, for the second session 1,2,3,4, for the third 1, $2\frac{1}{2}$, $2\frac{1}{2}$,4, and for the fourth 2,3,1,4.

As a second test, consider that if the regrouping process sorts subjects by their contribution behaviour, we should find less variation of contributions within groups in the regrouping treatment than in the baseline treatment. To test this prediction we calculated the coefficient of variation to measure within-group differences while controlling for differences in the average contributions level across groups. We found using a Mann-Whitney test that the average coefficient of variation is significantly lower for endogenously formed groups in the regrouping treatment, compared with the randomly formed groups in the baseline treatment.

As noted in Section 1.2, persistence of contributing behaviour after regrouping does not by itself prove that subjects actually differed in preference type. Up to period 18, it is possible that payoff maximising types were mimicking cooperator types to get into better groups. In the next Section we focus on last period behaviour, which provides strong evidence that not all of the earlier contributing was due to decisions to mimic cooperativeness.

2.3. *Individual Behaviour and Heterogeneous Preference Types*

Figure 5(*a*) shows a histogram of the frequency of each contribution level in the baseline treatment, in the last period, for all subjects in all the sessions. In this period, only 4.7% of baseline subjects contribute all of their endowment and only about 12.5% contribute anything. This is consistent with the interpretation that there are a small number of true cooperators, but with many mimickers, there can still be substantial contributions in earlier periods. In the histogram, it appears to be evidence of at least 12.5% being some form of cooperator (some conditional cooperators might contribute nothing in the last period if they believed that others would be likely to contribute nothing).

In comparison, the histogram of the frequency of each contribution in the last period of the regrouping treatment in Figure 5(*b*) suggests a much higher proportion of cooperators. This histogram shows 28.1% contributing their entire endowment, and 59.4% contributing at least something, in the last period. Since rational payoff maximisers have a dominant strategy to contribute nothing in the last period, this suggests that at least 59% of subjects in the regrouping treatment were conditionally or unconditionally cooperative – the 59% figure being a lower bound, since conditional cooperators might also contribute nothing depending on their expectations of others' behaviour.[23]

The disaggregated histograms of the regrouping treatment, Figure 5(*c*) – (*f*), strengthen this interpretation. In the first formed groups (Group 1), 50% of the subjects contributed their entire endowment in period 20; in the second formed groups (Group 2), 43% contributed their entire endowment; in the third formed groups, 18%, and in the last formed group, no one contributed their entire endowment. The patterns of last period contributions are also consistent with expectations in that contributions become more spread over the spectrum from 0

[23] The data in Gunnthorsdottir *et al.* also seem to show surprisingly high proportions of subjects contributing something in the last period. But the authors do not provide an explicit analysis of last period behaviour.

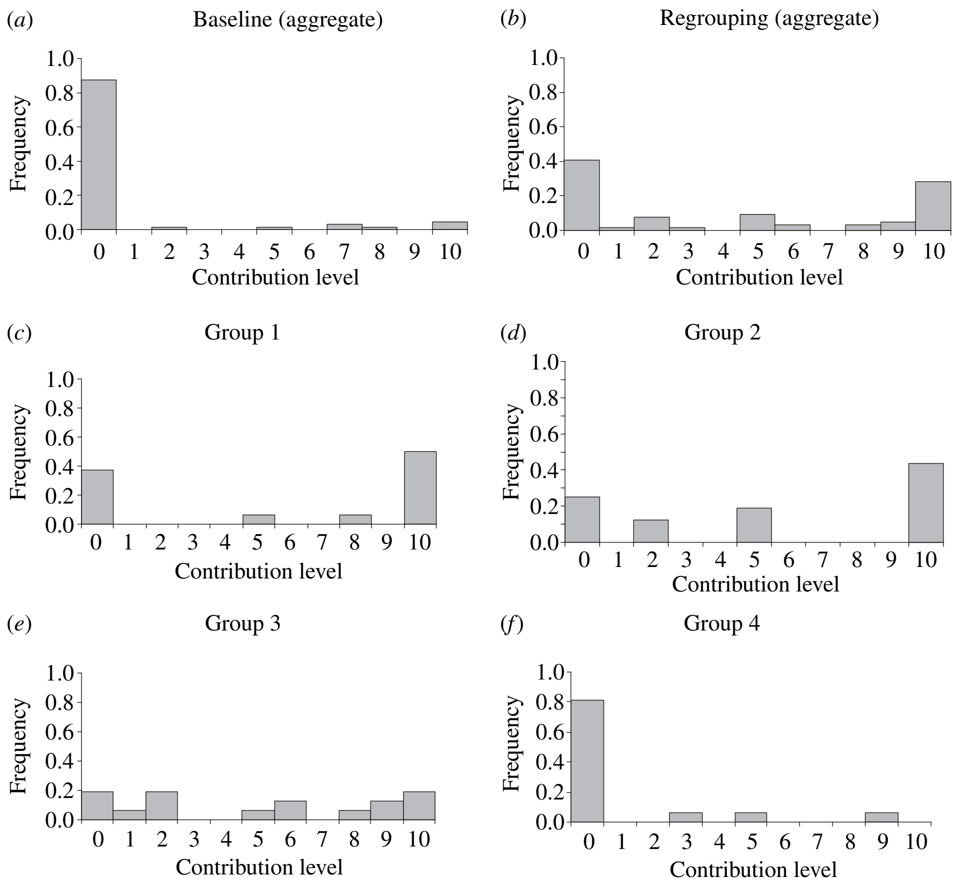

Fig. 5. *Frequency of Last Period Contributions by Treatment, Group, and Contribution Level.* Panels (a) and (b) are frequencies for all 64 subjects in the baseline treatment and for all 64 subjects in the regrouping treatment. Panels (c) to (f) are frequencies for the 16 subjects (4 subjects per session) in the relevant regrouping treatment group as of the last period.

to 10 as one moves from Group 1 to Group 2 and on to Group 3, while the modal contribution is zero in Group 4.

Based on the last period histograms for the baseline and endogenous grouping treatments, a case can be made that not just a few but in fact most subjects are conditional cooperators – that is, that they are willing to contribute if they expect others to do so, even though it lowers their earnings. Suppose, as seems reasonable, that the expectation (on the part of the subjects themselves) of the number of last period contributors of each possible amount in each group is an increasing function of past contributions in that group, and define a conditional cooperator as someone who will contribute more the more she expects others to contribute. Then if conditional cooperators are common, we should see higher period 20 contributions in those groups (e.g., the Group 1s of each endogenous grouping session) with high and stable contributions during periods 4 to 19, and lower contributions in those groups (e.g., Group 4s and almost all baseline treatment groups) with low and declining contributions during those periods. Indeed, viewing our baseline and endogenous grouping treatments as a laboratory for generating differing expectations as subjects proceed toward the last period, subjects' last period contributions are consistent with the conjecture that most if not all subjects are conditionally cooperative. If all subjects were payoff maximising, there would be no last period contributions and, if most subjects were unconditionally cooperative, then the levels of last period contributions would be unrelated to the levels of other group members' recent contributions.[24]

By forming groups where conditional cooperators tend to be matched together, even with substantial numbers of mimickers, the regrouping process tends to increase and stabilise the conditional cooperators' contribution levels, and reveal their type in the last period, where the mimickers drop away.

2.4. *Some Evidence on Positive and Negative Reciprocity*

Incurring a cost to punish free riders may be a strategically rational move if group members are to continue playing together and if this signal is expected to induce them to raise their contributions. But like contributing to a public good in a one shot game or in the last period of a finitely repeated one, punishing free riders in the last period is never payoff maximising to the punisher. Yet, as in Fehr and Gächter (2000), in our experiment there is no sign that the tendency to punish free riders declines in period 20 of the punishment and combined treatments. Like the willingness to contribute when others do so, discussed in Section 2.3, a willingness to incur a cost to punish a free rider on other people's contributions is evidently also part of the preferences of many subjects – a part that might dramatically change the dynamics of many group interactions, see, for example, Fehr and Schmidt (2002).

Fehr and Gächter have suggested that positive reciprocity – including the willingness to contribute to a public good if others do so – and negative reciprocity – including the willingness to incur a cost to punish a free rider – are two sides of a single preference type which they call reciprocity. We looked at the combined treatment to explore the interpretation that subjects with a stronger preference to cooperate also have a stronger preference for punishing free riders. If the two propensities closely coincide and if it is mainly high contributors who punish low ones, then we should expect to see punishment concentrated mainly in second or third-formed groups, since first-formed groups tend to have little free riding to punish, fourth-formed groups few if any high contributors and thus few to punish their free riders, and the two middle-

[24] As to why contributions decline with repetition in randomly formed groups without special incentives such as regrouping or punishment, we agree with Fehr and Gächter's interpretation that the main reason is that in ordinary VCMs like our baseline treatment, conditional cooperators who find themselves grouped with free riders have no way to protect themselves against free riding or to punish free riders other than to reduce their own contributions. Since it is impossible for other conditional cooperators to tell whether fellow members are reducing their contributions to punish or reducing their contributions due to 'learning that free riding is the most profitable action', conditional cooperation within the ordinary VCM group is steadily eroded.

formed groups tend to have both some free-riding and some high contributors (willing, on this interpretation, to punish the free riders).

In the last period behaviour in the combined treatment, we found more punishment in the second and third-formed groups than in the first and fourth-formed groups, an observation supporting these conjectures. (In the last period, payoff maximising types, who were previously mimicking cooperators, have a dominant strategy of neither contributing nor punishing, so that there is likely to be less confounding between payoff maximisers and other types). The situation is less clearcut in the combined treatment in periods 4 to 18 from the time of the first regrouping to the time of the last one. In these middle periods we found relatively little punishment in the first-formed group but we also found the most punishment in the fourth-formed group, where there were the lowest contributions. Over these periods, in fact, we found punishment to be almost perfectly ordered with the order of group formation. For the treatment as a whole, and at session level for three out of four sessions, the first-formed group had the lowest level of punishment, the second-formed group the next lowest, the third-formed group the second highest level of punishment and the fourth-formed group (where contributions are lowest) the highest level of punishment. This ordering suggests that while there may be reciprocator types who both contribute to the public good and punish low contributors, low contributing free riders also frequently punish in our subject population.

2.5. *The Anticipation Effect*

Comparing first period contributions in the regrouping treatment with first period contributions in the baseline suggests an anticipation effect (see Figure 1). In the regrouping treatment, groups are formed randomly before the first period, three periods before the first possible sanction by regrouping (recall that the baseline groups are also formed randomly before the first period, then remain fixed throughout the twenty periods). But first period contributions were significantly higher in the regrouping treatment than in baseline.[25] The higher initial contributions suggest that there was an incentive effect, namely that some subjects contributed more in period 1, before there was any endogenous regrouping, in anticipation of earning a reputation that would help them get into a good group. Conditional cooperators might also have contributed more from the outset because the mechanism's incentive structure led them to expect higher contributions, on average, from others.

Figure 2 shows that the anticipation effect of the regrouping treatment is about the same magnitude as the anticipation effect of the punishment treatment (the anticipation effect of the punishment treatment is similar to earlier findings in the literature for the punishment treatment).[26] Figure 2 also shows an anticipation

[25] Because each individual's decision is independent of the others in the first period, we compared all 64 decisions in the regrouping treatment to all 64 decisions in the baseline treatment. The p-value of a one-tailed Mann-Whitney test was 0.025.

[26] The p-value for a one-tailed Mann-Whitney test of the 64 first period contributions in punishment compared with the 64 first period contributions in baseline treatment is 0.01.

effect of the combined regrouping and punishment treatment, again of approximately the same magnitude.[27]

3. Concluding Discussion

Despite the standard Nash equilibrium prediction that a society of narrowly self-interested individuals will free ride, people are often observed to cooperate effectively in work groups, in local public goods provision and in other settings. As a way of explanation, theorists have found that the existence of heterogeneous preference types, including a relatively small number of cooperative types along with a substantially larger number of narrowly self-interested types, could sustain substantial levels of cooperation in a repeated game. Unfortunately, theorists so far have only been able to compute Bayesian equilibria for fairly simple games, such as two-person prisoners' dilemmas, and not for more complicated games such as four-or-more person VCMs with 10 or more periods, eleven or more contribution alternatives and added complications such as punishment or endogenous regrouping. The difficulty of the problem at a theoretical level increases the importance of carefully designed experiments as a tool for studying the determinants of cooperation and free riding.

Our experiment was designed to illuminate the role that having a say over who one interacts with plays in the mitigation of free rider problems. Our design also offered new ways of investigating the presence and the impacts of heterogeneous types. If subjects chose their actions believing there to be a significant likelihood of encountering cooperative individuals or being seen as cooperative by others, then our endogenous grouping treatment could increase the level of cooperation, because rational subjects would have incentives to mimic cooperators and because any conditionally cooperative subjects actually present would be more likely to encounter other cooperators or seeming cooperators, which would sustain their own willingness to contribute.

A subtle aspect of our regrouping treatment is that until the final period, any sorting of subjects that occurs cannot safely be presumed to reflect real preferences. Logically, the dramatic sorting exhibited in the treatment between periods 4–18 indicates only an unknown mix of true cooperators, payoff-maximising mimickers and less sophisticated or less optimistic payoff-maximisers. Until period 20, indeed, we observe sorting by behaviour, but not necessarily sorting by true type. In the last period, however, the logic of the game demands that 'the chickens come home to roost'. The surprisingly high proportion making non-zero contributions (including contributions of the full endowment) in the last period, suggests that many – indeed most – members of higher-contributing groups had actual willingnesses to contribute if they expected others to do so, contrary to the requirement of payoff maximisation. Contrasted with the low final contributions in almost every group in our baseline treatment and in other finitely-repeated VCM experiments, this

[27] The p-value for a one-tailed Mann-Whitney test of the 64 first period contributions in the combined treatment compared to the 64 first period contributions in the baseline treatment is 0.013.

constitutes strong evidence that there are substantial numbers of conditional cooperators.[28]

Our experimental results are, roughly speaking, consistent with rational choice (in the sense of expected utility maximisation) but we make no claim that each subject is a perfect Bayesian game player.[29] After all, the VCMs of this literature are sufficiently complicated so that no one, to our knowledge, has calculated Bayesian equilibria for them. Nonetheless, the stylised facts of Bayesian games that have been solved are in accord with the observed results of this experiment. It appears that subjects believe that there are at least a few cooperators who contribute highly, and the subjects see confirmation that there are many high contributors in the experiment (who may include both cooperators and those who decide to act like cooperators).[30] The apparent presence of so many conditional cooperators cannot be bad news for the business of economic modelling – nor is it bad news for designers of and actors in real social institutions.

Brown University
Brown University
University of Texas at Austin

Submitted: 3 June 2003
Accepted: 3 September 2004

References

Anderson, Christopher M. and Putterman, Louis. (forthcoming). 'Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism', *Games and Economic Behavior*.

Andreoni, James (1988). 'Why free ride? Strategies and learning in public goods experiments', *Journal of Public Economics*, vol. 37, pp. 291–304.

Andreoni, James (1995). 'Cooperation in public-goods experiments: kindness or confusion?', *American Economic Review*, vol. 85 (4) (September), pp. 891–904.

[28] Another experiment that provides evidence that there are many true reciprocator types (akin to the cooperator types in our setting), as well as evidence of the presence of mimickers who act like reciprocators but cease to do so in the final period, is the gift exchange experiment reported in Gächter and Falk (2002). We are indebted to a referee for bringing this similarity to our attention.

[29] A question posed by an editor of this JOURNAL is: why do there appear to have been mimickers in every group (put differently, if mimickers were rational, why did they not all seek membership in first-formed groups)? As a practical matter, we suspect that subjects differ in their degrees of strategic sophistication, calculating abilities and rates of learning. And even if they were equally sophisticated and rational in the sense of expected utility maximisation, mimickers did not know at the beginning of the experiment how many cooperators and other mimickers there were, and they are likely to differ in their beliefs as to the benefits of getting into a good group and what it takes to get into one. It also appears that there were too many cooperators and mimickers to fit into the top group or groups. And we think it likely that the stylised types of cooperator, mimicker, and free rider have substantial gradations within them, and are not entirely stable. In other words, we think the observed perfect sorting of group contribution level by the order of group formation comes from imperfect efforts of utility maximisation with heterogeneity, rather than from hyper-rationality.

[30] It is a legitimate question, of course, to ask how conditionally cooperative preferences could have come into being and survived in an evolutionary sense. While this takes us beyond the scope of our article, note that that question is currently being treated in thoughtful and rigorous ways, and that in much of the literature concerned, questions of assortative matching that bear a relation to the endogenous group formation studied in this paper play a prominent role; see, for instance, Bergstrom (2002), Sethi and Somanathan (2003) and the papers in Gintis *et al.*, (2005).

Bergstrom, Theodore (2002). 'Evolution of social behavior: individual and group selection', *Journal of Economic Perspectives*, vol. 16, pp. 67–88.

Bochet, Olivier, Page, Talbot and Putterman, Louis (forthcoming). 'Communication and punishment in voluntary contribution experiments', *Journal of Economic Behavior and Organization*.

Bohnet, Iris and Kübler, Dorothea (2005). 'Compensating the cooperators: is sorting in the prisoner's dilemma possible?', *Journal of Economic Behavior and Organization*, vol. 56, pp. 61–76.

Brosig, Jeannette (2002). 'Identifying cooperative behavior: some experimental results in a prisoner's dilemma game', *Journal of Economic Behavior and Organization*, vol. 47, pp. 275–90.

Burlando, Roberto and Guala, Francesco (2005). 'Heterogeneous agents in public goods experiments', *Experimental Economics*, vol. 8, pp. 35–54.

Carpenter, Jeffrey and Matthews, Peter (2002). 'Social reciprocity', Middlebury College Department of Economics Working Paper no. 29.

Casari, Marco and Plott, Charles R. (2003). 'Decentralized management of common property resources: experiments with a centuries-old institution', *Journal of Economic Behavior and Organization*, vol. 51 (2), pp. 217–47.

Cinyabuguma, Matthias, Page, Talbot and Putterman, Louis (2005). 'Cooperation under the threat of expulsion in a public goods experiment', *Journal of Public Economics*, vol. 89, pp. 1421–35.

Coricelli, Giorgio, Fehr, Dietmar and Fellner, Gerlinde (2004). 'Partner selection in public goods experiments', *Journal of Conflict Resolution*, vol. 48 (3), pp. 356–78.

Croson, Rachel (1996). 'Partners and strangers revisited', *Economic Letters*, vol. 53, pp. 25–32.

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

Ehrhart, Karl-Martin and Keser, Claudia (1999). 'Mobility and cooperation: on the run', Working Paper 99s-24, CIRANO, Montreal.

Falk, Armin, Fehr, Ernst and Fischbacher, Urs (2001). 'Driving forces of informal sanctions', Working Paper No. 59. Institute for Empirical Research in Economics, University of Zurich, May.

Fehr, Ernst and Gächter, Simon (2000). 'Cooperation and punishment', *American Economic Review*, vol. 90, pp. 980–94.

Fehr, Ernst and Gächter, Simon (2002). 'Altruistic punishment in humans', *Nature*, no. 415, pp. 137–40.

Fehr, Ernst and Schmidt, Klaus (2002). 'Theories of fairness and reciprocity – evidence and economic applications', in (M. Dewatripont, L. Hansen and S. Turnovsky, eds.), *Advances in Economics and Econometrics – 8th World Congress, Econometric Society Monographs*, Cambridge: Cambridge University Press.

Fischbacher, Urs, Gächter, Simon and Fehr, Ernst (2001). 'Are people conditionally cooperative? Evidence from a public goods experiment', *Economics Letters*, vol. 71, pp. 397–404.

Gächter, Simon and Falk, Armin (2002). 'Reputation or reciprocity?', *Scandinavian Journal of Economics*, vol. 104, pp. 1–26.

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

Green, Jerry and Laffont, Jean-Jacques (1979). *Incentives in Public Decision-Making*, New York: North-Holland.

Gunnthorsdottir, Anna, Houser, Daniel, McCabe, Kevin and Ameden, Holly (2002). 'Distribution, history and contributions in public goods experiments', unpublished manuscript, Department of Economics and Economic Science Laboratory, University of Arizona, September.

Guttman, Joel (2003). 'Repeated interaction and the evolution of preferences for reciprocity', Economic Journal, vol. 113 (489), pp. 631–56.

Hoffman, Elizabeth, McCabe, Kevin and Smith, Vernon (1998). 'Behavioral foundations of reciprocity: experimental economics and evolutionary psychology', *Economic Inquiry*, vol. 36, pp. 335–52.

Isaac, R. Mark and Walker, James M. (1988). 'Group size effects in public goods provision: the voluntary contributions mechanism', *Quarterly Journal of Economics*, vol. 103, pp. 179–200.

Keser, Claudia and van Winden, Frans (2000). 'Conditional cooperation and voluntary contributions to public goods', *Scandinavian Journal of Economics*, vol. 102(1), pp. 23–39.

Kreps, David, Milgrom, Paul, Roberts, John and Wilson, Robert (1982). 'Rational cooperation in finitely repeated prisoners' dilemma', *Journal of Economic Theory*, vol. 27, pp. 245–52.

Kurzban, Robert and Houser, Daniel (2001). 'Individual differences in cooperation in a circular public goods game', *European Journal of Personality*, vol. 15 (S1), pp. S37–52.

Ledyard, John (1995). 'Public goods: a survey of experimental research', in (John Kagel and Alvin Roth, eds.), *Handbook of Experimental Economics*, pp. 111–94, Princeton: Princeton University Press.

Masclet, David, Noussair, Charles, Tucker, Steven and Villeval, Marie-Claire (2003). 'Monetary and non-monetary punishment in the voluntary contributions mechanism', *American Economic Review*, vol. 93 (1), pp. 366–80.

McKelvey, Richard and Palfrey, Thomas (1992). 'An experimental study of the centipede game', *Econometrica*, vol. 60(4), pp. 803–36.

Offerman, Theo, Sonnemans, Joep and Schram, Arthur (1996). 'Value orientations, expectations, and voluntary contributions in public goods', ECONOMIC JOURNAL, vol. 106, pp. 817–45.

Ones, Umut and Putterman, Louis (forthcoming). 'The ecology of collective action: a public goods and sanctions experiment with controlled group formation', *Journal of Economic Behavior and Organization*.

Orbell, John and Dawes, Robyn (1993). 'Social welfare, cooperators' advantage, and the option of not playing the game', *American Sociological Review*, vol. 58 (6), pp. 787–800.

Ostrom, Elinor, Walker, James and Gardner, Roy (1992). 'Covenants with and without a sword: self governance is possible', *American Political Science Review*, vol. 86 (2), pp. 404–16.

Palfrey, Thomas and Prisbrey, Jeffrey (1997). 'Anomalous behavior in public goods experiments: how much and why?' *American Economic Review*, vol. 87(5), pp. 829–46.

Park, Eun-Soo (2000). 'Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding', *Journal of Economic Behavior and Organization*, vol. 43, pp. 405–21.

Sefton, Martin, Shupp, Robert and Walker, James (2002). 'The effect of rewards and sanctions in provision of public goods', Working Paper, University of Nottingham and Indiana University.

Sethi, Rajiv and Somanathan, E. (2003). 'Understanding reciprocity', *Journal of Economic Behavior and Organization*, vol. 50(1), pp. 1–27.