Experimental Economics, 8:35–54 (2005)

Heterogeneous Agents in Public Goods Experiments*

ROBERTO M. BURLANDO
Department of Economics, University of Torino, Italy and School of Psychology, University of Exeter, UK
email: roberto.burlando@unito.it

FRANCESCO GUALA
Department of Sociology and Philosophy, University of Exeter, UK
email: f.guala@ex.ac.uk

Received May 28, 2003; Revised April 22, 2004; Accepted November 24, 2004

Abstract

We explore by purely experimental means a heterogeneous agents scenario in experimental public goods games, assuming the existence of at least three types of player: free riders, cooperators, and reciprocators. We identify the various types by means of four classification methods, and then play the public goods game with homogeneous groups. We observe that (1) the average contribution level is enhanced in this setting; (2) the decay phenomenon is replicated in groups of 'pure' free riders, whereas in groups of cooperative and reciprocating players the contribution is high and fairly stable throughout the game.

Keywords: economic experiment, voluntary contribution, heterogeneous agents, reciprocation

JEL Classification: H41, C92

1. Introduction

'Overcontribution' in linear Public Goods (PG) experiments is by now a well-established phenomenon. It is also established that the level of contribution tends to diminish with repetition. This phenomenon of declining contribution is sometimes referred to as 'decay'. However, overcontribution does not disappear completely, even after up to 60 rounds (Ledyard, 1995).

Standard economic theory based on the model of a selfish and perfectly rational representative agent is unable to account for overcontribution and decay. The most obvious departure from the standard approach consists in relaxing the selfishness and the full rationality assumptions by introducing altruism and error parameters (Palfrey and Prisbey, 1996, 1997; Anderson et al., 1998). Empirical estimates of these augmented models however

*The experiments reported in this paper were financed by the Cognitive Science Laboratory and the Computational and Experimental Economics Laboratory of the University of Trento. We would like to thank Marco Tecilla, Macrina Marchesin and Dino Parisi for their help in running the experiments, and Luigi Mittone, director of CEEL, for letting us use the laboratory's resources. Paul Webley, Robert Moir, seminar participants at the Universities of Pisa and Trento, Charles Holt and two anonymous referees provided very useful comments on a previous draft.

have discovered a considerable heterogeneity among the subjects participating in PG experiments (cf. e.g. Goeree and Holt, 2002).

Although explanations based on agents' heterogeneity have a long history in economic and social psychology and in microeconometrics (see e.g. Kelley and Stahelski, 1970; Johnson and Nohrem-Hebeisem, 1979; Webley et al., 2001, Ch. 5) they have entered mainstream economic literature only recently. The implications and relevance of heterogeneity have been highlighted with respect to various policy issues by Heckman (2001a, 2001b), and recent discussions of PG experiments pointing in this direction include, among others, Andreoni (1995b), Offerman et al. (1996), Burlando and Hey (1997), Burlando and Webley (1999), Fischbacher et al. (2001), Andreoni and Miller (2002), Fehr and Fischbacher (2002), Goeree and Holt (2002). In this paper we follow their lead and try to establish the existence of different types of agent and its relevance for the decay of overcontribution in a non-theoretical way, or by purely experimental means.

Suppose that the decay phenomenon resulted from the combination of different types of players. Suppose also that the experimental sample were composed of *free riders* and *cooperators*, as well as *reciprocators* willing to give conditional on the others' contribution. Free riders, roughly speaking (a more precise empirical definition will be given later in the paper), try to exploit the contribution of other members of the group to their own advantage. Pure or unconditional cooperators invest either in the region of the Pareto-optimal solution, regardless of the behaviour of other members of the group, or at least significantly more than the group average. Reciprocation, finally, is here a broad category that includes 'matching' behaviour of various kinds and is in principle compatible with several possible motivations: 'sucker aversion', 'conformity' or 'miming', 'achievement orientation', and the well-known 'tit-for-tat' strategy.[1]

Such a scenario has an important implication: it makes decay dependent on the composition of the experimental groups. If there are enough free riders in the experimental population, uniformly distributed across the groups, the decay phenomenon gets triggered. Reciprocators, in fact, try to offer cooperation by contributing to the public good in the early rounds of the game; their attempts however are partly frustrated by the free riders in the group. By 'matching' the average group contribution, reciprocators cause the average to fall at even lower levels, until it gets trapped in the area of the Nash equilibrium (more precisely, a few pure cooperators may sustain it slightly above Nash). This story has been told several times in the literature, but has never been tested directly. In this paper we try to do so by exploring the dependence of the decay phenomenon on the composition of the groups: if we could somehow isolate free riders from players with a cooperative or reciprocating attitude, then we should observe a quick decay towards Nash in groups of free riders, vs. a more stable contribution rate in groups made up entirely of reciprocators and/or 'pure' cooperators.

We implement a within-subjects design where the same individuals participate first in a repeated linear PG experiment with groups of *heterogeneous* subjects, and then in a repeated linear PG experiment with *homogeneous* groups. Between the first and the second phase of the experiment, the subjects are classified according to the above categories. We show that (1) players of different type clearly differ in their contribution level, and (2) homogeneous groups contribute on average more than heterogeneous groups. We see these results as a first step towards the construction of a more adequate theoretical explanation of behaviour in PG and similar experimental games.

The most difficult task in running such an experiment is to identify and classify subjects according to types. This holds in particular for reciprocators, a type of agent that is remarkably difficult to model theoretically.[2] In this experiment we try to obtain a data-driven classification that is not dependent on any specific theoretical model,[3] by combining four sources of evidence: the ‘Strategy Method’ used by Fischbacher et al. (2001), the ‘Decomposed Game Technique’ used by Offerman et al. (1996), various measures of behaviour in a repeated linear PG game (along the lines of Burlando and Webley, 1999), and a questionnaire. All these methods have been used before in the literature, but none of them seems to be entirely adequate to identify the categories we are interested in. By triangulating between them, however, we hope to come up with a classification that is better than the one that would be provided by each method applied in isolation. Obviously such an approach can be criticised for its reliance on intuitive rather than theoretically-founded measures. However, it’s important to stress that *the classification itself can receive direct support from the experimental evidence*: if the categories were totally arbitrary or ill-defined, we should not expect to find any significant difference between the levels of cooperation in Session 1 (with ‘heterogeneous’ groups) and in Session 2 (with ‘homogeneous’ groups).

To see why, imagine an extreme case in which our classification method failed completely: in such a case every *true* cooperator, say, would have an equal chance of being assigned to any of the groups. This would amount to a mere random reallocation of the subjects to different groups, and therefore we should expect in Session 2 to observe approximately the same phenomenon of decaying overcontribution as in Session 1. In contrast, the stronger the difference between the two sessions, the more our classification method turns out to be corroborated.

Another piece of evidence corroborating the classification method is the convergence of the results obtained by means of different techniques. As microbiologists know very well, it is not necessary that we are 100% confident in the reliability of a microscope in order to obtain valid empirical data. We can use different instruments (e.g., a light microscope, an electronic microscope, etc.) to observe the same specimen and then use an argument from coincidence of the following sort: it would be a true miracle if all these different instruments reported the same, mistaken, phenomenon. Their convergence suggests that the phenomenon must be real, rather than an artifact of the observation procedure. Our use of different classification techniques (or ‘triangulation’) follows the same logic.

As a matter of fact, we do observe a highly significant difference in the contribution patterns of Sessions 1 and 2 in our experiment; and we do observe a remarkable convergence between the different methods of classification. Hence, we can conclude with a high degree of confidence that the classification procedure does the job it was designed for. Section 2 is devoted to illustrating these techniques and the experimental design in general. Section 3 contains the results of the experiment, and Section 4 concludes.

2. Experimental design

The experiment consisted of two sessions, with exactly a one week interval between them. The first session consisted of four different tasks: a Strategy Game, a Decomposed Game, a repeated linear PG game, and a questionnaire. The second session consisted of another repeated linear PG game. In the first session the subjects were allocated randomly to the various groups, whereas in the second one they were matched so as to create homogeneous groups.[4]

Between Sessions 1 and 2 we classified every single subject into types. Each subject was initially marked with four labels, one for each task. Overall, there was a remarkable degree of convergence between the four classification tasks, but obviously some cases of disagreement (on one or more dimensions) remained. In order to resolve them, we assigned weights to the four classification methods. Given that the ultimate goal is to understand behaviour in a repeated PG game, we decided to put more weight on the PG game data, according to the following formula: Repeated PG game 40%; Strategy Method 20%; Decomposed Game 20%; Questionnaire 20%. When no classification reached the 50% threshold (and therefore in all cases of tie) we assigned the subject to a 'noisy' group.

The experiment was run at the University of Trento (Italy) in May 2002, and involved 92 subjects recruited by means of flyers (mostly, but not entirely, undergraduates from the School of Economics and Business). Overall, subjects spent about 80 minutes in the lab (50 for the first session and 30 for the second one) and earned on average 20.50 Euros.[5] The experiment was run entirely by computerised means, with up to 20 subjects sitting in the same room, in front of terminals isolated by means of partitions.[6] The experimental currency was expressed in 'tokens', with 1 token = 1 cent of Euro. The first session consisted of four different tasks. The subjects were given a sheet of instructions for the first task. Each subject was instructed to press a key when she had finished reading the instructions; when all the subjects had done so, the experimenter asked if anybody wanted to ask any question. Then, the first task began. (The same procedure was followed before each experimental task.)

2.1. *Session 1, task 1: Strategy Method*

The so-called 'Strategy Method' was used by Fischbacher et al. (2001) in an attempt to observe the phenomenon of reciprocation (or 'conditional cooperation') directly. The environment for this task is a linear PG game with payoff function

$$\pi_i = 200 - g_i + 0.5 \sum_{j=1}^{4} g_j, \tag{1}$$

where 200 is the total number of tokens to be shared between a 'private' (200 – g) and a 'public' account (g).[7] Once subjects have been made familiar with the situation, they are asked to take two decisions: first, they are asked to make an 'unconditional contribution', i.e. to decide how much they would like to contribute in a standard one-shot PG game where each player, at the moment of taking her decision, doesn't know how much the other players have contributed. Secondly, subjects are asked to fill in a (conditional) 'contribution table', i.e. to indicate how much they *would* be willing to contribute *if* they knew that the other members of their group had, on average, contributed to the public good a given amount. This question is iterated, varying the amount hypothetically contributed by the 'others' (in discrete intervals) from 0 to 200. In other words, subjects are asked to make a series of 'conditional contributions' in addition to the unconditional one just indicated.

Clearly the data from the 'contribution table' are particularly useful in order to identify reciprocating players. Participants know that after the decisions have been made subjects will be randomly allocated to groups of 4 players, one of which will be selected at random as the one who will actually play the conditional contribution task, based on the other three players' unconditional decisions. The actual rewards are then calculated (and communicated to the players) according to the payoff function above. This way, both decisions (conditional and unconditional) are made relevant for the final result, and monetary incentives are provided for all members of the group.[8]

2.2. *Session 1, task 2: Decomposed Game*

The Decomposed Game technique has been widely used by psychologists and (more recently) economists in order to measure attitudes towards cooperation. Subjects are asked to make 24 choices between pairs of allocations. Each subject knows that she has been paired with another participant who will remain the same, but unknown, throughout the game. Each allocation consists of a number of tokens paid to yourself and another sum paid to the other player. The token amounts can be positive or negative. A typical choice may involve, e.g., a combination A = (75, −130) vs. B = (39, −145), where one must choose between gaining either 75 or 39 tokens, with related losses on the other's part of either 130 or 145 tokens. The total sum ('own' plus 'other') allocated is not constant over the 24 combinations. (The full list of A, B combinations is standard and can be found in Appendix 1.) There is no feedback concerning the other's choices. The final payoff is obtained by combining the 24 choices of each subject with those of the other player.[9]

At the end of the experiment we take the 24 vectors chosen by each subject and add them up to obtain his or her own 'motivational vector'. We use the standard classification criteria used in the previous literature:[10] the motivational vector is placed on the standard 'value orientation circle' (see Appendix) and classified accordingly. The length of the motivational vector is used as a measure of a subject's consistency. A perfectly consistent subject (i.e. a subject who always chooses the alternative lying closest to her own motivational vector) should have a vector length of 300. Subjects with a vector length of 150 or less (less than 50% of the maximal length) thus display a considerable degree of inconsistency or confusion.

2.3. *Session 1, task 3: Repeated linear PG experiment (heterogeneous groups)*

The third task is a repeated linear PG experiment, with the same payoff function as in the first task (the Strategy Method), except that the individual endowment at each round is 20 tokens instead of 200. A new set of instructions reminds subjects about the environment, and states clearly that the groups (of 4 players) will be different from those of tasks one and two. The PG game is played for 23 rounds, with the first three rounds for training (the payoffs do not count) and 20 rounds for real. After each round the subjects are given feedback about their total earnings, their earnings in the previous round, and the average contribution level of their group. The number of rounds that have already been played is displayed at the bottom of the screen.

2.4. Session 1, task 4: Questionnaire

Before leaving the room, subjects were required to fill in a questionnaire (on their PCs). The first four questions were open, with enough space to write an articulate answer: (1) "What were you trying to do in the experiment (in other words: what were your goals or objectives)?" (2) "Did you achieve your objectives?" (3) "What were the other members of your group trying to do (what were their objectives)?" (4) "What was the scope of this experiment (in other words, what were the experimenters trying to discover)?"

Then, each subject was asked to indicate her level of agreement with four statements, on a 1–7 scale: (5) "This experiment requires a great concentration effort"; (6) "The rules of the game were explained clearly and were understandable"; (7) "In this experiment one must try to work together with others, in order to have everyone end up with more money"; (8) "In this experiment, everyone's earnings depend on the decisions of all members of the group". The relevant questions for us were (1), (2), (3), and (7). Question (7) was used by Brandts and Schram (2001) as an indicator of individual attitude towards cooperation; by means of the first three questions (taken from Burlando and Webley, 1999) we hoped to collect useful data about the strategies implemented during the first three tasks of the experiment.

2.5. Classification criteria

The classification of players within the repeated PG experiment is a complex matter, for it should take into account various aspects of behaviour. On top of the 'triangulation' argument outlined in Section 1, there may also be good psychological reasons to follow a multi-method approach, as the different tasks may tap on different aspects of a complex psychological construct that a single approach might overlook.[11] We used data from the four tasks described above in order to classify individual players. For the purposes of our research, we needed to divide our subjects into three main categories: free riders (F), cooperators (C), and reciprocators (R). However, in order to account for ambiguous or borderline cases we also formed a residual group of 'noisy' players (N) (or, more precisely, subjects with 'noisy' data that are not easily interpretable).[12]

The Strategy Method provides two main parameters: the conditional contribution table and the unconditional contribution. We decided to classify as 'reciprocators' all players whose conditional contribution functions approximate the 'perfect reciprocation function', with a margin of variation of $\pm 10\%$. For example: a player willing to contribute 0 if the others contribute 0, 0 if they contribute 1, 1 if they contribute 2, 2 if the contribute 3, and so on, would be classified as a reciprocator (albeit with a slight 'individualistic' bias). In contrast, a subject willing to give 0 if the others give 0, 0 if they give 1, 0 if they give 2, 0 if they give 3, 1 if they give 4, 2 if they give 5, and so on, would count as a free rider according to our classification. (A symmetrical criterion is used in order to identify cooperators.)

Although the great majority of conditional contribution functions display a coherent pattern (they are either flat or increasing with the group contribution-level), some seem to follow random walks or other non-easily interpretable patterns.[13] We classified cases like this as 'noisy'. The conditional contribution table was then compared with the unconditional contribution of the same subject. The idea is that subjects classified as 'cooperative' according to the conditional contribution table should give high unconditional contributions, and the opposite should hold for free riders. The behaviour of reciprocators, as usual, is more difficult to pin down, for their unconditional contribution will depend decisively on their expectations about the members of their group. Here we relied on their answer to the question: 'how much do you think the other members of the group have contributed, on average?' which was inserted between the unconditional and the conditional contribution tasks. For genuine reciprocators, the answer to the latter question should be highly correlated with the unconditional contribution. Free riders and 'pure' cooperators, in contrast, should act pretty much independently of their expectations on others' behaviour. When these simple rules were violated, the subject was labeled as 'noisy'.

The Decomposed Game provides a tight classification of subjects into 5 categories, according to where the 'motivational vector' lies on the 'value orientation circle'. They are: 'aggressive' if it lies between degree −112.5 and −67.5; 'competitive' if between −67.5 and −22.5; 'individualistic' if between –22.5 and 22.5; 'reciprocating' if between 22.5 and 67.5; 'cooperative' if between 67.5 and 112.5.[14] For our purposes, we aggregated subjects belonging to the first three categories ('aggressive', 'competitive' and 'individualistic') into one single category, that of 'free riders'. Subjects with a low level of coherence (below 50%) were classified as 'noisy'.

In the PG game, the obvious parameter to start from is the simple average individual contribution over the 20 rounds played 'for real' (IA). This measure of course is the variable under examination and hence is unable to explain itself, but nevertheless can be legitimately used for taxonomic/predictive purposes.[15] A more serious problem is that IA is likely to depend on group behaviour: a subject with an average low contribution, for instance, may be either a pure free rider, or a reciprocator which happens to be in a group of free riders. Similarly, a high average contribution may signal either a pure cooperator, or a reciprocator in a group of reciprocating/cooperating players. Hence, IA needs to be complemented with a second parameter: its difference with respect to average group contribution (or average difference, AD).

In interpreting PG data we follow a two-steps algorithm. Our algorithm for IA is:

(1) C or R if $IA > 12$;
(2) R if $12 \geq IA \geq 8$;
(3) F or R if $IA < 8$.

Whenever we face an ambiguous case (1 or 3 above), we use the second measure, the difference between individual and group average contribution (AD). The algorithm in this case is:

(4) C if $AD \geq 1$;
(5) R if $-1 < AD < 1$;
(6) F if $AD \leq -1$.

Finally, we used evidence from the questionnaire to complement the data from the three previous tasks. Questionnaires are not highly valued in experimental economics, but we found them of some use as an aid in the interpretation of the PG game evidence.[16] The questionnaire included subjects' own *post-hoc* explanation (rationalisation) of the strategies they followed in the repeated PG task. The answers being 'open', the quality of the data varied from subject to subject.[17]

2.6. *Example*

In order to give an idea of the classification process, we shall here illustrate one particular instance taken from our experimental data. The behaviour of subject number 51 in the PG game, vis-à-vis the average contribution level of the members of her group, is represented in figure 1. Her individual contribution is almost invariably below the group average—subject 51 was pretty clearly free riding. (Quantitatively, IA = 3.15 and AD = −3.55.) In the Decomposed Game, the subject emerged as a 'competitive' individual (hence a 'free rider' in our terminology), but with a low level of coherence (35%). As a consequence we classified her as 'noisy' in this task. The data from the Strategy Method, conditional contribution task, are represented in figure 2. The conditional contribution function lies unequivocally below the 45% (or 'perfect reciprocation') line. This is confirmed by the other two parameters of the Strategy Method: the subject expected other players to contribute on average 100 tokens, but she herself contributed (in the 'unconditional' task) only 50. According to this classification criterion, she is a free rider. Finally, Subject 51 gave the following answer to the question 'What were you trying to achieve in this game?':

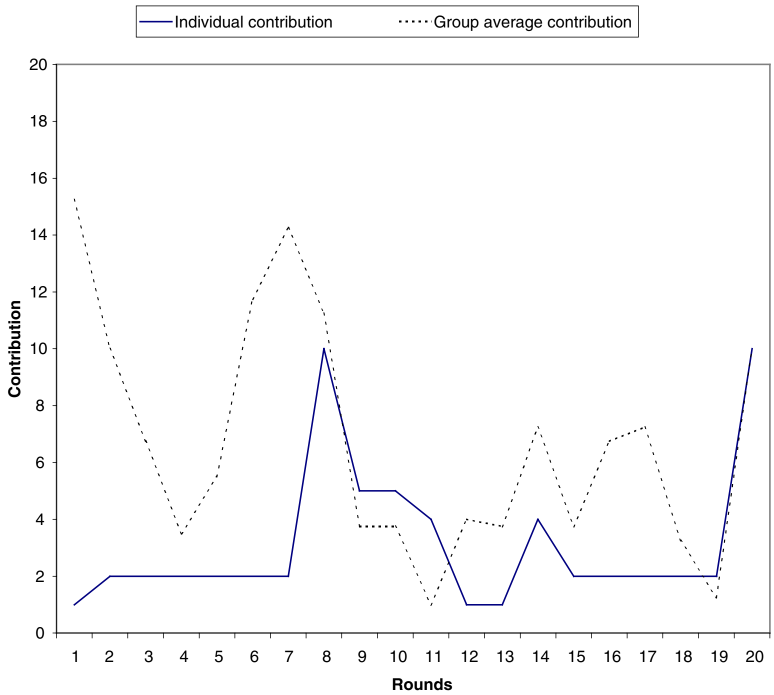

Figure 1. Subject 51, Session 1, PG game.

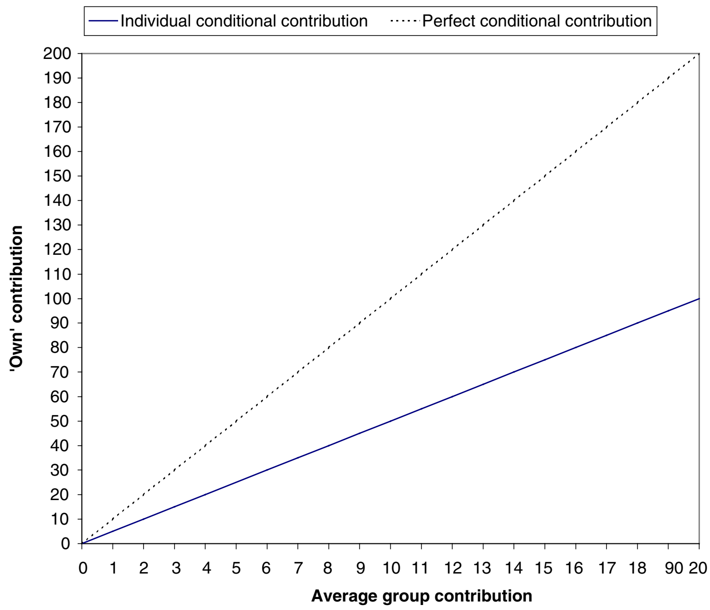

Figure 2. Subject 51, Session 1, Strategy Method.

> "My main objective was to invest little money in the project so as to obtain a sure gain from the money I had put in the individual account and also to benefit from the gains obtained from the project thanks to the investments made by the other components of the group."

So, to sum up, Subject 51 was classified in the following way:

PG Game: Free rider (40%)
Decomposed Game: Noisy (20%)
Strategy Game: Free rider (20%)
Questionnaire: Free rider (20%)
Aggregate classification: Free rider.

Of course other cases were less straightforward than this one but, as we shall see very shortly, our ad hoc algorithms seem with hindsight to have performed quite well.

2.7. Session 2: Repeated linear PG experiment (homogeneous groups)

All subjects were classified during the week that elapsed between Session 1 and Session 2. The number of subjects falling in each category is shown in Table 1.

As expected the numbers in each category are not multiples of four, which caused some inconvenience in creating the homogeneous groups. In order to achieve maximum

Table 1. Classification.

	Free riders	Reciprocators	Cooperators	Noisy	Total
n	29	32	17	14	92

homogeneity and in order not to lose too many subjects, we devised a system of substitutes to replace subjects who might unexpectedly fail to show up at Session 2. Eventually, we managed to create 6 groups of free riders, 6 groups of reciprocators, 4 groups of cooperators, and 3 groups of 'noisy' players. We also formed two non-homogeneous groups with the remaining players. One group turned out to be made of three reciprocators and one free rider (as we shall see, the results from this group are quite surprising); the other group of one free rider, two reciprocators, and one 'noisy' player. For obvious reasons, these non-homogenous groups will not be considered in the main process of data-analysis below. Eight subjects (three free riders, three reciprocators, one cooperator, and one 'noisy' player) did not participate in Session 2, either because they failed to attend, or because there were no other players available to form a group.[18]

Three days after the first session subjects were contacted individually (by email) and confirmed the exact time of the second experimental session. When they arrived, subjects were allocated to their terminals. The instructions were identical to those of Session 1, Task 3, except that it was made clear that the composition of the groups differed from that of the first session. Once again, subjects played 3 rounds for training and 20 for real. At the end of the experiment we combined their earnings from both sessions, converted into Euros, and paid them accordingly.

3. Results

The core of our analysis is based on the behaviour of the 84 subjects who attended and played both experimental sessions, and especially the 76 subjects who played the second session in homogeneous groups. We will analyse separately the behaviour of the 4 homogeneous groups of cooperators, the 6 groups of free riders, and the 6 groups of reciprocators. Towards the end of this section, however, we shall also comment briefly on the behaviour of 'noisy' players (3 groups) and on the two non-homogeneous groups.

Two implications of the heterogeneous agents scenario turn out to be strongly corroborated: (1) in the second PG game the average level of contribution is significantly higher than in the first one; (2) in the second PG game groups of ('pure' and 'conditional') cooperators display a very high and fairly constant level of contribution. In groups of free riders, in contrast, the contribution starts lower and quickly jumps to a low level, reaching in the last round the expected Nash equilibrium of zero contribution.

The first result is apparent from figure 3: the average contribution level in Session 2 is always above that of Session 1. Statistically, the hypothesis that the two sets of data (contribution in the PG Game, Session 1, and contribution in the PG Game, Session 2) come from the same population is rejected at the 1% level in all rounds but six, where it is rejected at 5% (Wilcoxon signed ranks test). The question then arises of which subjects effectively

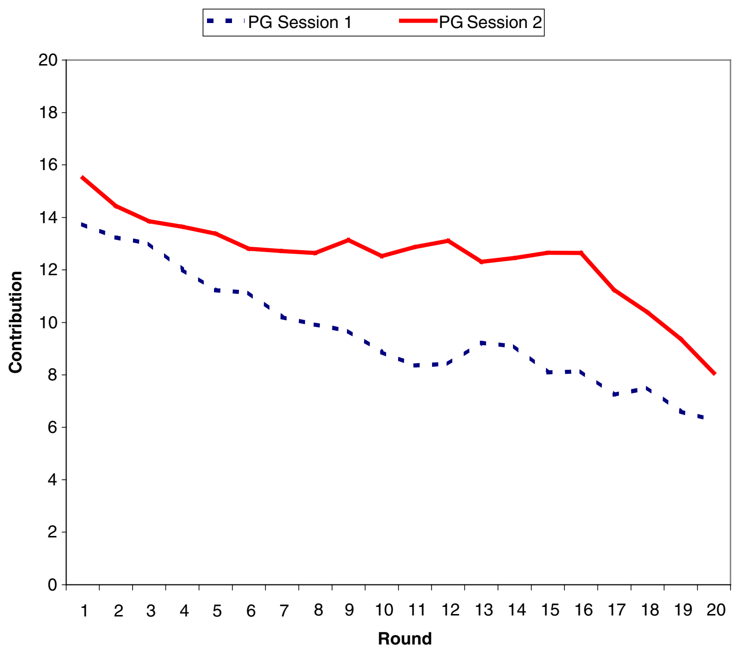

Figure 3. Contribution level in the PG Games: Session 1 vs. Session 2 ($n = 76$).

contributed to the upward shift of the contribution curve. Table 2 shows that reciprocators are mostly responsible for the shift: on average, their contribution jumped from 10.38 to 18.76 tokens. This abrupt increase is to be expected in the light of the heterogeneous agents scenario. What the latter does not necessarily imply is that cooperators should also raise their contribution level once placed in homogeneous groups (+1.87), nor that free riders should end up free riding more (−2.09). One plausible explanation for this is that our groups of cooperators and free riders were 'infiltrated' by a certain amount of reciprocators– i.e. they were less 'pure' than we would have liked them to be. Or, perhaps, even cooperative players are partly influenced by the group they happen to be in. In the case of free riders, it is also likely that in the first session some of them played strategically, trying to profit by contributing less than the group average but still above the Nash equilibrium (in order

Table 2. Average contribution by type and session.

Type	Session 1	Session 2	Both sessions
Cooperators	14.58	16.45	15.52
Free Riders	4.17	2.26	3.22
Reciprocators	10.38	18.66	14.52
Noisy	8.74	10.10	9.42
Overall	9.53	12.49	11.01

not to discourage others' contributions too early in the game). In the second session, they might have quickly realised that such a strategy was not profitable, and provoked a more rapid decay.

Notice that Sessions 1 and 2 are not exactly comparable because subjects could benefit from increased experience while playing the PG game the second time. According to the widely shared opinion that experience leads to a marked decrease in the level of cooperation (Isaac et al., 1984; Ledyard, 1995: 143; Plott, 1996), experience should bias *downwards* the results of Session 2, which in turn suggests that the 'pure' effect of heterogeneity may in fact be stronger than the (already remarkable) one observed in our experiment. In light of the above discussion, however, the received view should probably be qualified: while experience seems to cause a decrease in the level of cooperation of free riders and probably also reciprocators in heterogeneous groups, this is clearly not true in the case of cooperators and reciprocators playing in homogeneous groups.

Figures 4(a–c), 5 and 6 show the average contribution of each type of player during the 20 rounds of play, in Sessions 1 and 2. Notice how in Session 1 (figure 5) the contributions of reciprocators lie between those of free riders and cooperators, but then shift upwards to the top of the graph in Session 2 (figure 6). Part of the explanation is that in Session 2 the variance is greater among cooperators (which is probably due to the fact that we were not entirely successful in forming perfectly homogeneous groups for this category).

As was to be expected, in Session 2 free riders start with a low level of contribution, which tends to decay very quickly (the average is 5 tokens after only two rounds) and never manages to recover. Cooperators and reciprocators start with a very high level of contribution, which remains high throughout most of the game. The behaviour of reciprocators is particularly impressive, with constant (and almost full) contribution until round 16. There is evidence of an 'end-effect' (Andreoni, 1988; Keser, 1996) in the last 3–4 rounds, with abrupt decay to a level of about 12 tokens. (Such an 'end-effect', incidentally, explains why the difference between the contribution levels in the first and second sessions is greater in the middle of the game—cf. figure 3.) The presence of an end-effect in reciprocating players casts further doubts on explanations based on pure altruism. Many reciprocating players do defect at the end of the game, which suggests that their behaviour is indeed prompted either by selfish motives or at least by fear of being exploited. It is likely that reciprocators do not consider the equilibrium strategy to be rational (they probably cannot bring the backward induction argument to its radical and counterintuitive consequences).

We ran a Kruskal-Wallis non-parametric test on the contribution in the second session of the four samples (free riders, reciprocators, cooperators, and 'noisy' players). The results show that the hypothesis that the four samples of data come from the same generation mechanism can be rejected with near certainty, both when the data from all rounds are analysed jointly (.000 asymptotic significance) and when a round-by-round comparison is made as reported in Table 3 below. We also ran Mann-Whitney tests on pairs (free riders vs. cooperators, cooperators vs. reciprocators, etc.), and the results are as expected. The difference between free riders and, respectively, cooperators and reciprocators, is highly significant both when the data from different rounds are analysed jointly (.000) and when analysed on a round-by-round basis as in Table 4. The difference in the behaviour of cooperators and reciprocators in the second session is smaller: the two groups cannot be

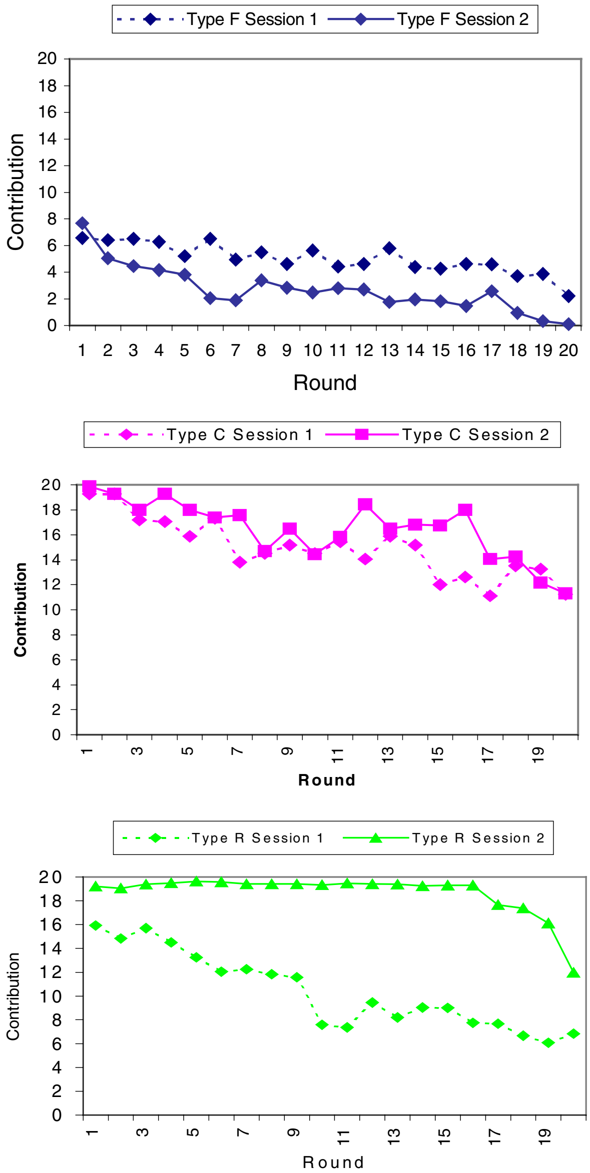

Figure 4. (a–c): Contribution in PG Game, Sessions 1 vs. 2, per type.

easily distinguished statistically on a round-by-round comparison, but a joint Mann-Whitney test on all rounds provide an asymptotic significance of .018.

We haven't said much about 'noisy' players. In fact, it is difficult to characterise them except for the fact that their choices show a much greater variance than those of the other groups.[19] On average their contribution levels are above those of free riders in both sessions,

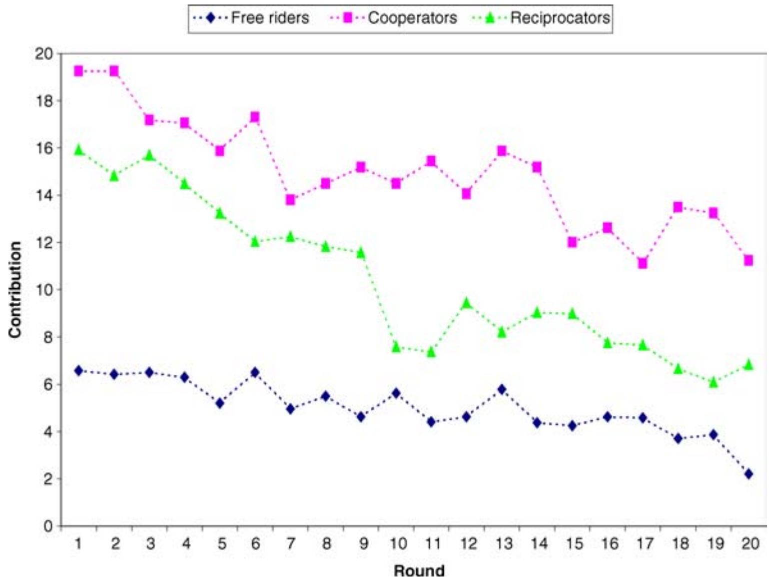

Figure 5. Contribution level by type of player: Session 1.

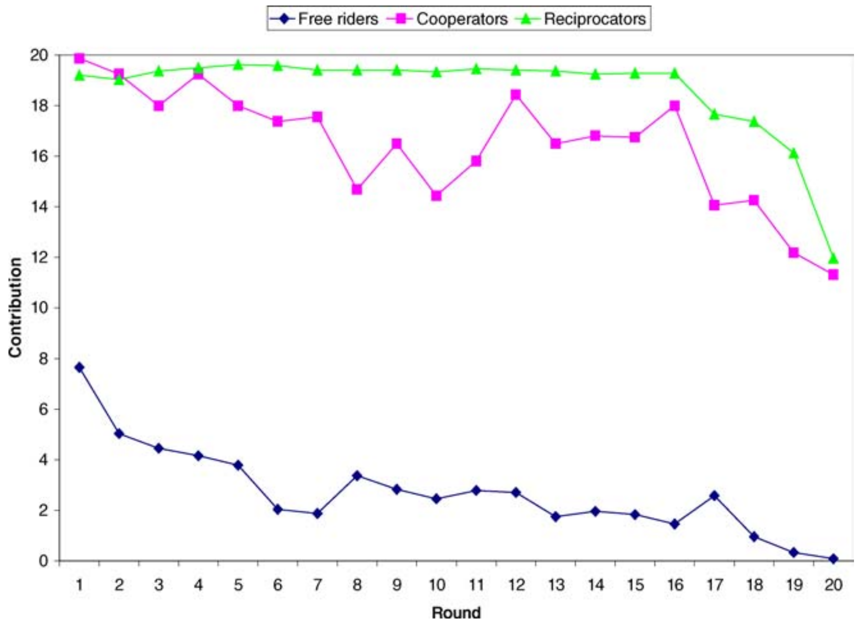

Figure 6. Contribution level by type of player: Session 2.

Table 3. Kruskal-Wallis test on all groups (F, C, R, N), Session 2; asymptotic significance.

rd1	rd2	rd3	rd4	rd5	rd6	rd7	rd8	rd9	rd10	rd11	rd12	rd13	rd14	rd15	rd16	rd17	rd18	rd19	rd20
.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

Table 4. Mann Whitney test, Session 2; exact significance.

	rd1	rd2	rd3	rd4	rd5	rd6	rd7	rd8	rd9	rd10	rd11	rd12	rd13	rd14	rd15	rd16	rd17	rd18	rd19	rd20
C vs. F	.000	.000	.000	.000	.000	.000	.000	.003	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.002	.002
C vs. R	.467	.570	.968	.885	.682	.517	.644	.375	.435	.435	.500	.781	.404	.781	.552	.926	.227	.295	.188	.721
F vs. R	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

and below those of cooperators. This is not surprising, given that the ‘noisy’ group was used as a ‘catch-all’ and therefore probably included players of all types. The two non-homogeneous groups engaged in fairly strange behaviour, which we cite only for curiosity and the sake of completeness. The group composed of three reciprocators and one free rider managed to sustain a decent level of cooperation – below the average of reciprocators but well above that of free riders. The group with two reciprocators, one noisy player and one cooperator displayed a lower, and swinging, level of cooperation throughout the game.

4. Conclusion

Our evidence strongly supports the heterogeneous agents scenario. A first implication for the debate on repeated PG experiments is that the divide between models of self-interested agents on the one hand and models of altruistic or cooperative players on the other, might never be resolved simply because there are agents of both types. In fact it seems more interesting and fruitful to recognise not only the existence of these two types of players, but also their influence on another (large) category of players: reciprocators (or ‘conditional cooperators’).

This does not mean that representative agent-type of models are useless: in many circumstances they may be perfectly adequate for predictive purposes (especially the ‘augmented’ models with altruism). Ultimately, however, any such model is likely to lack explanatory depth and will fail to capture the important mechanisms that sustain cooperation. Overall, we agree with Andreoni and Miller that “a model that predicts well in the aggregate may not help us understand the behaviour of individual actors” and “capturing the variety of choices among individuals and then aggregating their behaviour will lead to better understanding of both individuals and markets” (2002: 750).

Another important finding is that, contrary to Houser and Kurzban’s (2002) recent claim, frustrated attempts at reciprocation play a major role in the decay of contribution in PG experiments. Since reciprocators constitute a large portion of the experimental population (at least in our sample, but see also Fischbacher et al., 2001),[20] it is possible to raise the overall level of contribution by forming homogeneous groups of players with similar attitudes towards cooperation. This is consistent with the results of PG experiments with *endogenous* group formation: letting people choose their partners may be an effective mean to promote cooperating behaviour. *How much* effective this can be, however, depends on at least two factors: the quantity and quality of the information that is publicly available for the identification of players, and the possibility of 'locking' homogeneous groups so as to prevent free riders from infiltrating the more cooperative ones (Ehrhart and Keser, 1999; Page et al., 2002). Our experiment with *exogenous* group formation may be taken as an example of what happens when a relatively great amount of information is available, and the mobility between groups is zero (We have recently become aware of a similar experiment with exogenous group formation, performed by Simon Gachter and Christian Thöni (2005)). Exogenous group formation may also be useful in certain management contexts: for example when the members of a team must exert effort with public goods characteristics, a manager might want to prevent the spread of free riding behaviour by forming homogeneous production units.[21] In this sense homogeneous group formation performs a similar function to cooperation-sustaining mechanisms such as the punishment of free riders (Fehr and Gachter, 2000).

Appendix

Table 1A. Decomposed Game: choice of allocations.

	Option A		Option B	
Question	Self	Other	Self	Other
1	+150	0	+145	+39
2	+144	−39	+130	−75
3	+130	−75	+106	−106
4	+106	−106	+75	−130
5	+75	−130	+39	−145
6	+39	−145	0	−150
7	0	−150	−39	−145
8	−39	−145	−75	−130
9	−75	−130	−106	−106
10	−106	−106	−130	−75
11	−130	−75	−145	−39
12	−145	−39	−150	0
13	−150	0	−145	+39
14	−145	+39	−130	+75
15	−130	+75	−106	+106
16	−106	+106	−75	+130
17	−75	+130	−39	+145
18	−39	+145	0	+150
19	0	+150	+39	+145
20	+39	+145	+75	+130
21	+75	+130	+106	+106
22	+106	+106	+130	+75
23	+130	+75	+145	+39
24	+145	+39	+150	0

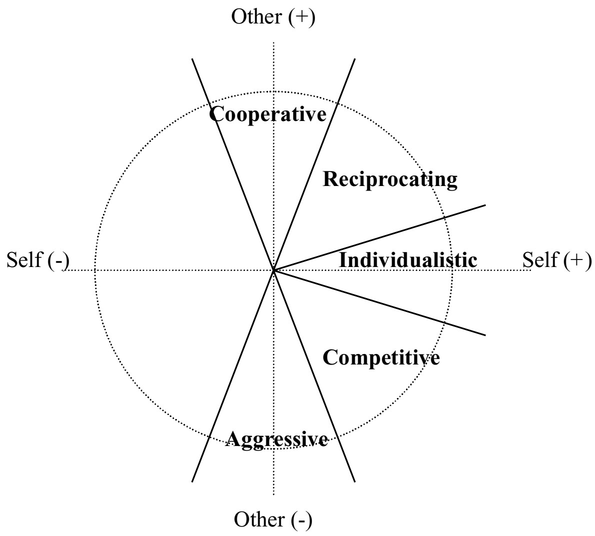

Figure 1A. Decomposed Game: the ‘value orientation circle’.

Notes

1. ‘Sucker aversion’ is the attempt to avoid being taken advantage of; ‘conformity’ is the deliberate choice of being the same as others; ‘miming’ is the thoughtless copying of others; ‘achievement orientation’ is the setting of an earning goal and, when others free ride and put you short of that goal, respond to free riding with more free riding; ‘tit-for-tat’ is the mechanical replay of the previous move of the opponents. We are particularly grateful to an anonymous referee for highlighting these distinctions.
2. For some attempts in this direction cf. Sugden (1984), Rabin (1993), Levine (1998), Charness and Rabin (2002), Falk and Fischbacher (2002), Dufwenberg and Kirchsteiger (2004).
3. This does not mean that we are against a theoretically-based approach, but given the difficulties of modelling behaviour (especially reciprocating behaviour) in PG experiments, it seems reasonable to start from a set of purely empirical observations and *then* engage with the theoretical side at a later stage of research.
4. Of course the participants did not know they were playing in pre-determined groups. Subjects, moreover, were not revealed the results of the various tasks until the very end of the experiment.
5. This is more than what an average Italian student can earn in a part-time job. Subjects were told in advance that payment was strictly conditional on participating to both sessions, and that earnings from both sessions would be paid immediately after the second one. Those who didn’t show up at the second session lost their earnings from the first session, and their data were disregarded.
6. The software for the experiment was created by Marco Tecilla at CEEL.
7. In this phase we used (with minor adaptations) instructions and control questions provided by Urs Fischbacher, whom we would like to thank. With respect to the Fischbacher et al. experiment we raised the marginal payoff function of contributions to the public goods from 0.4 to 0.5 tokens. After the unconditional contribution task, we also asked subjects ‘how much do you think the other members of the group have contributed to the public good, on average?’
8. See Fischbacher et al. (2001) for a more detailed analysis of this procedure. Unlike in the original one, in our experiment the random selection mechanism was run by the computer itself.
9. For example: suppose subject i accumulates a total payoff of 500 (‘own’) and allocates to the ‘other’ a total payoff of 100; his counterpart j, in contrast, accumulates a total payoff of 550 (‘own’) and allocates to the other a total payoff of -40. In the end, i gets $500 - 40 = 460$, whereas j receives $550 + 100 = 650$ tokens.

10. Cf. Griesinger and Livingston (1973), Liebrand (1984), Offerman et al. (1996).
11. A classic defence of this approach can be found in Rushton (1980, Ch. 4).
12. This is in line with evidence presented, among others, by Andreoni (1995b) and Palfrey and Prisbey (1996, 1997).
13. A significant portion of subjects display puzzling 'hump-shaped' functions: their contribution grows with the group contribution up to a point (typically, 50% of the tokens), and then declines towards zero. See also Fischbacher et al. (2001).
14. Notice that here, and everywhere else throughout this paper, our use of the terms 'reciprocator' and 'cooperator' differs from the conventional terminology in the standard literature on the Decomposed Game. We use the term 'reciprocation' (or 'conditional cooperation') where the standard literature uses the term 'cooperation' and we use 'cooperation' for behaviour that is termed 'altruistic' there. This is more in line with the terminology used in the literature on PG experiments.
15. In this respect, our research differs from previous attempts to observe individual attitudes or motives in PG-like environments (e.g. Offerman et al. 1996; Brantds and Schram 2001; Andreoni 1995a). Our categories are more 'behavioural' and less psychological in character than those used in these studies.
16. We don't believe, in other words, that questionnaire data can be used as the only or even primary source of evidence, but often they allowed to discriminate between, say, a 'mild' cooperator and a 'noisy' player, or between a cooperator and a reciprocator, etc.
17. Many answers were simply too short or too obscure to be of any use. In order to control as much as possible for our own interpretative bias, each experimenter separately evaluated the answers and classified subjects accordingly; then, we compared our results and resolved the (rare) cases of disagreement.
18. The subjects who did not show up in both sessions were not paid at all. Those who did show up but could not participate in the experiment in Session 2, received their earnings from Session 1, plus a flat fee equal to the average earnings of the other players in Session 2.
19. In the last few rounds the variance increases also among cooperators and reciprocators (the end-effect is probably involved here), while it remains small and decreases to zero in the last round among free riders.
20. In this experiment we probably underestimated the number of reciprocators, but it is always difficult to find 'pure' types, especially at the extremes of the distribution. In principle it should be possible to distinguish among reciprocators with a more 'altruistic' attitude and others that are simply fearful of being exploited or are even inclined to free riding. We attempted a more detailed analysis, identifying three sub-types for each category of players, but we didn't have enough participants to allow us to do this properly. And the payoff of introducing finer distinctions is quite dubious anyway.
21. We are indebted towards a referee for this suggestion.

References

Anderson, S.P., Goeree, J.K., and Holt, C.A. (1998). "A Theoretical Analysis of Altruism and Decision Error in Public Goods Games." *Journal of Public Economics*. 70, 297–323.

Andreoni, J. (1988). "Why Free Ride? Strategies and Learning in Public Goods Experiments." *Journal of Public Economics*. 37, 291–304.

Andreoni, J. (1995a). "Warm Glow vs. Cold Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments." *Quarterly Journal of Economics*. 110, 1–21.

Andreoni, J. (1995b). "Cooperation in Public Goods Experiments: Kindness or Confusion?" *American Economic Review*. 85, 891–904.

Andreoni, J. and Miller, J. (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*. 70, 737–753.

Brandts, J. and Schram, A. (2001). "Cooperation and Noise in Public Goods Experiments: Applying the Contribution Function Approach." *Journal of Public Economics*. 79, 399–427.

Burlando, R.M. and Hey, J.D. (1997). "Do Anglo-Saxons Free Ride More?" *Journal of Public Economics*, 64. 41–60.

Burlando, R.M. and Webley, P. (1999). “Individual Differences and Long-run Equilibria in a Public Good Experiment.” In *Inquiries into the Nature and Causes of Behavior: Proceedings of the 24th IAREP Annual Colloquium*. Belgirate: Italy.

Charness, G. and Rabin, M. (2002). “Understanding Social Preferences with Simple Tests.” *Quarterly Journal of Economics*. 117, 817–869.

Dufwenberg, M. and Kirchsteiger, G. (2004). “A Theory of Sequential Reciprocity.” *Games and Economic Behavior*. 47, 268–298.

Ehrhart, K.M. and Keser, C. (1999). “Mobility and Cooperation: On the Run.” Working Paper 99s-24. CIRANO, Montreal.

Falk, A. and Fischbacher, U. (2000). “A Theory of Reciprocity.” Working Paper no. 6. Institute for Empirical Research in Economics, University of Zurich.

Fehr, E. and Fishbacher, U. (2002). “Why Social Preferences Matter—The Impact of Non-selfish Motives on Competition, Cooperation and Incentives.” *Economic Journal*. 112, C1–C33.

Fehr, E. and Gachter, S. (2000). “Cooperation and Punishment in Public Goods Experiments.” *American Economic Review*. 90, 980–994.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). “Are People Conditionally Cooperative? Evidence from a Public Goods Experiment.” *Economics Lette*rs. 71, 397–404.

Gächter, S. and Thöni, C. (2005). “Social Learning and Voluntary Cooperation among Like-minded People.” *Journal of the European Economic Association. 3*, forthcoming.

Goeree, J. K. and Holt, C.A. (2002). “Private Costs and Public Benefits: Unravelling the Effects of Altruism and Noisy Behavior.” *Journal of Public Economics*. 83, 257–278.

Griesinger, D.W. and Livingston, J.W. (1973). “Towards a Model of Interpersonal Motivation in Experimental Games.” *Behavioral Science*. 18, 173–188.

Heckman, J.J. (2001a). “Accounting for Heterogeneity, Diversity and General Equilibrium in Evaluating Social Programmes.”*Economic Journal*. 111, F654–F699.

Heckman, J.J. (2001b). “Microdata, Heterogeneity, and the Evaluation of Public Policy.” *Journal of Political Economy*. 109, 673–748.

Houser, D. and Kurzban, R. (2002). “Revisiting Kindness and Confusion in Public Goods Experiments.” *American Economic Review*. 92, 1062–1069.

Isaac, R. M., Walker, J.M., and Thomas, S.H. (1984). “Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations.” *Public Choice*. 43, 113–149.

Johnson, D.W. and Norem-Hebeisen, A.A. (1979). “A Measure of Cooperative, Competitive and Individualistic Attitudes.” *Journal of Social Psychology*. 109, 253–261.

Kelley, H. and Stahelski, A. (1970). “Social Interaction Basis of Cooperators’ and Competitors’ Beliefs About Others.” *Journal of Personality and Social Psychology*. 16, 66–91.

Keser, C. (1996). “Voluntary Contributions to a Public Good When Partial Contribution is a Dominant Strategy.” *Economics Letters*. 50, 359–366.

Ledyard, J.O. (1995). “Public Goods: A Survey of Experimental Research.” In J. Kagel and A. Roth (eds.), *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Levine, D. (1998). “Modelling Altruism and Spitefulness in Experiments.” *Review of Economic Dynamics*. 1, 593–622.

Liebrand, W.B.G. (1984). “The Effect of Social Motives, Communication and Group Size on Behavior in an N-person Multi-stage Mixed-motive Game.” *European Journal of Social Psychology*. 14, 239–264.

Offerman, T., Sonnemans, J., and Schram, A. (1996). “Value Orientations, Expectations, and Voluntary Contributions in Public Goods.” *Economic Journal*. 106, 817–845.

Page, T., Putterman, L., and Unel. B. (2002). “Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry, and Efficiency.” Unpublished paper, Brown University.

Palfrey, T.R. and Prisbey, J.E. (1996). “Altruism, Reputation and Noise in Linear Public Goods Experiments.” *Journal of Public Economics*. 61, 409–427.

Palfrey, T.R. and Prisbey, J.E. (1997). “Anomalous Behavior in Public Goods Experiments: How Much and Why?” *American Economic Review*. 87, 829–846.

Plott, C.R. (1996). "Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis." In K.J. Arrow, E. Colombatto, M. Perlman and C. Schmidt (eds.), *The Rational Foundations of Economic Behaviour*. London: Macmillan.

Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*. 83, 1281–302.

Rushton, J.P. (1980). *Altruism, Socialisation and Society*. Englewood Cliffs: Prentice Hall.

Sugden, R. (1984). "Reciprocity: The Supply of Public Goods Through Voluntary Contributions." *Economic Journal*. 94, 772–787.

Webley, P., Burgoigne, C., Lea, S.E.G., and Young, B. (2001). *The Economic Psychology of Everyday Life*. Philadelphia: Psychology Press.