CONDITIONAL COOPERATION AND GROUP DYNAMICS: EXPERIMENTAL EVIDENCE FROM A SEQUENTIAL PUBLIC GOODS GAME*

BY

DANIEL HOUSER[a] & ROBERT KURZBAN[b]

[a]INTERDISCIPLINARY CENTER FOR ECONOMIC SCIENCE
DEPARTMENT OF ECONOMICS
GEORGE MASON UNIVERSITY

[b]DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF PENNSYLVANIA

June 17, 2003

Abstract: We design a novel sequential public goods experiment to study reciprocity, or conditional cooperation. In contrast to the standard simultaneous contribution game, our sequential design provides direct evidence on how subjects condition their own contributions on the contributions of other subjects in the experiment. We develop a simple but useful behavioral-type classification procedure and use it to analyze the data from this design. Our results inform two fundamental hypotheses: (1) subjects' types are persistent over an experiment; and (2) the types of subjects included in a group affects a group's ability to sustain cooperation. These hypotheses are often assumed in the public goods literature, yet neither has been directly supported. We find support for both hypotheses. Moreover, we provide a simple summary statistic that, we show, predicts group cooperative dynamics remarkably well.

*We gratefully acknowledge that this research was supported by grants from the International Foundation for Research in Experimental Economics to both authors as well as a grant from the Russell Sage Foundation to the second author.

Introduction

Over the last two decades, research in economics and psychology has begun to take seriously the idea that individual differences among participants might be important in understanding and modeling behavior and dynamics in experimental games. In the experimental public goods literature, one focus of attention is the interaction of reciprocators, or "conditional cooperators," with non-cooperators or "free-riders." One reason for this attention is that a group composed of both conditional cooperators and free-riders would likely exhibit diminishing aggregate contributions to the public good over time, and this is consistent with frequently-replicated behavior in public goods experiments. Another important reason is that these strategies have firm theoretical justification: conditional cooperation (henceforth, CC) emerges as a rational behavior under many models of social giving that people have posited (see e.g., Rabin (1993), Andreoni (1995a), or Palfrey and Prisbey (1997)), while free-riding (henceforth, FR) can typically be supported as a Nash equilibrium. In addition, a number of models designed to address the puzzle of how large-scale non-kin based cooperation in humans evolved (Boyd & Richerson, in press) suggest that populations will come to equilibrate at a mix of free riders and reciprocal types (e.g., Gintis, 2000).

Ultimately though, two fundamental hypotheses provide the foundation for interest in CC and FR behavior in the experimental public goods literature. One is that a person who is classified as a conditional cooperator or free-rider will behave that way systematically. We denote this as the Persistence of Type (POT) Hypothesis. Note that type persistence does not require that a subject's contributions must remain constant over time, but rather that his or her contribution decisions should be well summarized, and well predicted, by his or her type classification. The POT hypothesis was advanced by Ledyard, among others, who conjectured that there are three types of subjects in public goods experiments, including free-riders and those that are socially motivated (Ledyard, 1995, pp. 171-2). Moreover, personality psychologists and behavior geneticists, whose work is predicated on the existence of stable individual differences in traits and predispositions, have long endorsed some version of the POT hypothesis (see, e.g., Funder, 2001).

The second foundational hypothesis is that the number of CC and FR types within a given group affects that group's cooperative outcomes. We call this the Types Affect Outcomes (TAO) Hypothesis. Again, this hypothesis is far from new. Andreoni (1995), for example, attributes about half of all cooperative decay in public goods experiments to the interaction of free-riders and other types of subjects (see also Houser and Kurzban, 2002). Many other researchers have suggested that a group's type composition might play a role in explaining their results (see, e.g., Andreoni and McGuire (1993), Keser and van Winden (2000), Brandts and Schram (2001), Gunthorsdottir et. al. (2001) and Frey and Meier (2003)).

While they are fundamental, and perhaps largely accepted, neither POT nor TAO is well informed by existing evidence from the experimental public goods literature. Although there is a large and recent literature that touches on conditional cooperation (see the above cited papers, and e.g., Croson (1998), and Kurzban, McCabe, Smith, and Wilson (2000)), to the best of our knowledge only Fishbacher, Gachter and Fehr (2001) (henceforth, FGF) provides direct laboratory evidence on whether people actually adopt conditionally cooperative strategies in group environments.[1] We are not aware of any evidence on whether conditional cooperation is played consistently within groups even within an experimental session.[2] In addition, FGF are only able to speculate about the implications of their results for the TAO hypothesis. In this paper we present data from a novel public goods experiment that is designed to provide direct evidence on both hypotheses POT and TAO.[3]

Our research builds on FGF (2001). Their design uses a variant of the strategy method (Selten, 1967) in an effort to investigate conditionally cooperative play directly. In particular, FGF asked each subject to indicate, hypothetically, the amount that he/she would contribute to the public good, for each possible mean amount contributed by the subject's group members.[4] Each subject's responses were plotted to create his or her "contribution profile," and these profiles were visually grouped to create the four distinct behavioral types mentioned above. They found that 50% of their subjects were conditional cooperators, one-third free-riders and the remaining either "Hump-Shaped" or "Others."

In this paper, we aim to design an environment in which we can observe a large number of contribution decisions for each subject and, importantly, the way in which each subject conditions his or her decisions on information about group members'

---

[1] Field experiments on charitable giving have found support for conditionally cooperative strategies (see, e.g., Frey and Meier, 2003, and List and Lucking-Reiley, 2002)

[2] Camerer and Fehr (2003) cite FGF (2001) and Croson (1998) to support the claim that "about half the subjects [in public goods experiments] are conditional cooperators." FGF (2001) argue that existing experimental evidence on conditional cooperation, including Croson (1998), is only indirect (fn. 2, p. 398).

[3] Evidence on the existence and dynamics of conditionally cooperative play has been provided within the context of two-person games (see, e.g., Clark and Sefton (2001) and McCabe, Rigdon and Smith (2002)). See also Heiner (2002) and Congleton and Vanberg (2001) for closely related theory studies.

[4] To make responses payoff-salient, a public goods game was actually played one time. Subjects were randomly placed into groups of four. Then, three of the four subjects contributed simultaneously any amount they wished to the public good, and then the (randomly chosen) fourth was required to contribute according to the contributions of the other group members and his or her expressed contribution profile.

decisions. Also, we want to use an alternative, complementary approach to the strategy method, and to incorporate group dynamics. To accomplish these goals we design a "sequential" version of the Voluntary Contribution Mechanism, previously described in Kurzban and Houser (2002), that proceeds as follows.

At the beginning of each round of play, participants simultaneously make an initial allocation of tokens between their private and group accounts. Subsequently, the current aggregate contribution to the group account is revealed to exactly one player, and that same player is allowed to change his or her contribution. Play continues with each subject in turn told the aggregate contribution and then given a chance to change his or her contribution, until the game ends at a randomly chosen point that is unknown to participants. Payoffs are determined by the allocation of tokens to the group and private accounts at the point when the game ends.

An advantage of our design is that it allows us to look at reciprocity in a way that most other public goods games do not. In particular, we are able to plot a subject's contributions against the average contribution to the group account that she observed before making her own contribution. So, for example, conditional cooperators' contributions should cluster around the forty-five degree line on this plot, while a free rider's contribution should be small regardless of the average contribution he or she observes.

We use our sequential public goods experiment to inform POT and TAO as follows. First, the game provides a way to elicit each subject's (conditional) contribution profile, and this can in turn be used to determine each subject's type. The statistical procedure that we introduce to determine a subjects type is novel, simple and, it turns out, quite useful. Whether the CC or FR types are present is directly informed by application of this procedure.[5] We provide evidence on POT by investigating whether subjects' decision strategies are similar across games played with different counterparts. Then, because subjects are typed, each group can itself be typed according to the composition of its members. This enables us to shed light on the validity of TAO.

Our results are both broadly consistent with and extend FGF (2001), and provide support for both POT and TAO. We classify 63% of our subjects as conditional cooperators and 20% as free-riders. With respect to POT, there is strong evidence of continuity in subjects' individual behavioral strategies across the many games they play with different counterparts. With respect to TAO, our evidence suggests that groups primarily composed of cooperative types sustain cooperation better than groups primarily composed of free-riders. While this result was expected, we were surprised to find that a group's cooperative *dynamics* can be remarkably well predicted by a single, easily calculated statistic that summarizes the cooperative propensities of its members.

2. Experimental Procedures and Design

We ran eighty-four participants in three groups of 24 and one group of 12. We found no statistically significant difference in behavior between the 12 and 24 subject sessions, and data are pooled to obtain the results reported below. Participants were recruited from the undergraduate population using the recruitment system in place at the University of Arizona's Economic Science Laboratory as well as through undergraduate classes at the University of Arizona.

[5] The approach to behavioral type classification that we propose below is a natural extension of the visual inspection approach used by FGF (2001). Because our data are collected with greater noise than is present when the strategy method is used, the visual inspection approach with our data is not straightforward.

Participants arrived in the laboratory and were seated at computer terminals that are divided by partitions so that it is not possible to see anybody else's screen. Except for the instructions, which were printed on paper, the entire experiment was conducted via computer. Once seated, participants were told that they had already earned their show-up payment and that their decisions and earnings would be kept confidential. Subsequently, the instructions for our sequential voluntary contribution mechanism, which closely followed Andreoni (1995) and are reported in Appendix A, were distributed.

The instructions informed participants that they would be assigned to a series of groups, each consisting of four people, and that the members of these groups would be shuffled randomly over an unspecified number of games. Participants were told that they would receive an endowment of 50 tokens per game to divide between two exchanges, and that tokens in the individual exchange earned one cent per token, while tokens in the group exchange earned half of one cent for each player. The marginal per capita return was thus 0.5.

Each game consisted of all players simultaneously making an initial allocation of tokens between the two accounts. After this initial decision, there were a number of rounds during which exactly one player in each group was provided the current aggregate contribution to the group exchange, and given the opportunity to change his or her allocation to the two accounts. Each person in the group was given a similar decision until the game ended at a point unknown to the participants. Participants were told that each player would have at least one chance to change their contribution decision. It was emphasized that payoffs each round would be determined by the final allocation of all group members' tokens to the two exchanges. The composition of the groups, the order of play within groups, and the length of each game was generated randomly before the experiment and kept constant for all three sessions. Game lengths were generated by assigning a probability of .05 that the game would end after any given player's opportunity to change their allocation. This process generated the following game lengths, measured in terms of the total number of individual contributions (including the four initial simultaneous contributions): 20, 11, 27, 36, 36, 38, 8, 21, 35, 12.

Previous research indicates that there might be significant amounts of confusion in understanding instructions in public goods games (see, e.g., Andreoni, 1995, and Houser and Kurzban, 2002); consequently participants completed a ten question quiz that had to be answered correctly before players could proceed. Most participants were able to complete the quiz quickly and successfully, and the experimenter privately assisted those who required additional help. The first game began after everyone had completed the quiz correctly, and subsequent games proceeded automatically after all groups had reached the end of the preceding game.

Participants were paid their experimental earnings privately, roughly $20 on average, and dismissed when the experiment concluded. Subjects were in the laboratory for about 90 minutes on average.

3. Results

Each experimental session included at least seven games. Some sessions proceeded slightly faster, and included as many as 10 games. Figure 1 describes the average contribution to the group exchange by round, where the averages are taken over all of the games that subjects played during the course of our experiment. Note that round 1 in this figure indicates the amount in the public account after the initial, simultaneous contribution by all subjects. Round two shows the average contribution after the first subject is given a chance to revise their contribution, round three the average contribution after the second subject has a chance to update, and so on. Note also that because different games end at different points, the averages are over different numbers of rounds. For example, all games included a round one, so its average is over 186 observations, while there are only 21 observations to include in the round 35 average because it was reached only in game six. Figure 1 shows the decay in contributions typically found in public goods experiments. In particular, average contributions decayed over time from 60% to about 35% of the subjects' endowment.

The averages described by Figure 1 mask substantial heterogeneity in behavior between groups. Figure 2 displays the contribution path realized by two groups over different 35 round games. One group's contribution decays from about 150 to zero after 19 rounds, and remains very small from that point forward. The other group achieved perfect cooperation from the fifth through the final round of the game. Although we selected highly disparate group dynamics for Figure 2, there is substantial variation between groups in their dynamic contribution paths in general. Hypothesis TAO suggests that this variation might be systematically related to variation in groups' "type" compositions.

3.1. Player Types

There is clear variation in decision strategies between subjects in our experiment. Figure 3 describes the empirical contribution profile (ECP) for three subjects. The horizontal axis indicates the mean contribution of a subject's group members that he/she observed before making a contribution decision,[6] and the vertical axis indicates the amount of his or her own contribution conditional on that information. Note that the these contribution profiles, and all of the data underlying contribution profiles that we report and analyze throughout this paper, exclude subjects' initial simultaneous contributions. The reason, of course, is that these contributions are made before any information on the contributions of others is observed.[7]

The three subjects that we highlight in Figure 3 demonstrate substantial differences in their ECPs. The subject marked with solid squares always contributes all of his or her endowment to the public good, regardless of others' contributions. A second subject, marked with diamonds, never contributes to the public good. A final subject, marked with solid triangles, roughly matches the mean contribution of his or her group members. These behaviors exemplify the three contribution strategies that we focus on below: cooperators, conditional cooperators and free-riders.

Our empirical contribution profiles include more "noise" than the profiles reported by Fishbacher, Gachter and Fehr (2001). For example, the subject marked by triangles in Figure 3 contributed, on different occasions, 18 and 43 when his or her group's mean contribution was 25. One reason this might occur is that a subject's contribution strategy depends on more than his or her group's current contribution. A contribution strategy could depend on previous own contributions, or on the path by which others' current mean contribution had been achieved. An alternative explanation

[6] Throughout this paper, the mean contribution of a subject's group members is always net of his or her own contribution.

[7] The initial, simultaneous contribution is also different from the other contributions in that subjects were told that they would have at least one opportunity to change it.

for this variation is that the contribution strategy is played with error. In this paper we propose a type classification procedure, described below, that is based on a simple behavioral framework that models a subject's contribution decision as a stochastic function of the mean contribution of his or her group counterparts.

3.1.a. A Fast and Reliable Type Classification Procedure

Our approach to behavioral type classification is to pre-specify a set of behaviors of interest, and then assign one from this set to each subject. This sort of approach is used, for example, by El-Gamal and Grether (1995) in their well-known behavioral typing algorithm (see also Houser and Winter, 2002). The behaviors that interest us are contributing little most of the time (free-riding), contributing a great deal most of the time (cooperating), and contributing an amount roughly equal to the contributions of others (conditional cooperation). These are the strategies most often studied in the experimental public goods literature, and we do not view our decision to focus on them here as controversial.[8] Our procedure is to use a subject's *linear conditional-contribution profile* to classify them as one of these types.

We define each subject's *linear conditional-contribution profile* (LCP) as the outcome of an ordinary least squares regression of his or her contribution decisions on the mean contribution that he/she observed immediately prior to making a contribution decision. Formally, let $y_n(x_j)$ denote the contribution of subject n when he/she observed mean contribution x_j. We estimate with ordinary least squares the model

[8] Alternative approaches to type classification that allow for the nature of types to be determined endogenously are available (e.g., Houser, Keane and McCabe, 2002), but unlike the approach we adopt here, these are not necessarily fast or easy.

$$y_n(x_j) = \beta_{0n} + \beta_{1n} x_j + \varepsilon_{nj} \tag{1}$$

for each subject n, where we assume ε_{nj} is independently and identically distributed across j for each n, and is allowed to have different variances across subjects.

An important advantage of equation (1) is that it allows us to summarize each subject's contribution strategy with two parameters: the intercept and slope of his or her LCP. The intercept provides a measure of how willing a subject is to cooperate even when his or her group counterparts contribute little to the public good. The slope measures a subject's responsiveness, both in direction and magnitude, to other's contributions. We classify each subject according to their estimated LCP, as follows.

Consider first a subject who contributes little to the public good, regardless of others' contributions. A strong free-rider, who always contributes nothing to the public good, is such a type. More generally, we classify a subject as a free-rider if their LCP implies contributions of less than 50% of his or her endowment regardless of the mean contribution of others. Because our subjects are endowed with 50 tokens per period, this means a subject is classified as a free-rider if and only if the graph of his or her LCP lies everywhere below 25. Similarly, a subject will be classified as a "cooperative" type if and only if the graph of their LCP lies everywhere at or above 25. The decision to use 50% of the endowment as the distinguishing point is obviously arbitrary, and different cut-off points might lead to somewhat different classifications. We will see below that the 50% point leads to reasonable results, in that the average behavior of subjects classified as free-riders and cooperators conforms well to the behavior researchers likely expect these types of subjects to exhibit.

Consider next the "conditionally cooperative" type. This type of subject should tend to contribute about the amount that their group counterparts contribute. In this sense, the 45-degree line reflects ideal conditional cooperation. More generally, we classify a subject as a conditionally cooperative type if the graph of his or her LCP has a positive slope and lies both above and below the 50% line.

LCPs with negative slopes that lay both above and below the 50% line form the "residual" group not included in one of the three types we model. Such a player would tend to contribute a lot when others contribute little, and little when others contribute a lot. To the best of our knowledge, this strategy has received no attention in the public goods literature, and indeed we find that such LCPs are infrequent in our sample. Based on their decisions in the first seven games, our classification procedure assigns 17 of 84 subjects (20%) to the free-rider type, 11 of 84 (13%) to cooperator, 53 of 84 (63%) to conditional cooperator, and places just 3 subjects (4%) in the residual group. Note that these frequencies correspond reasonably well to those reported by FGF (2001), and this lends support to using the strategy method approach in this environment.[9]

Finally, it is worth emphasizing that our goal with the typing procedure is not to find the model that best fits, in some sense, the contribution decisions of our subjects. Rather, it is to provide a fast and accurate method for inferences about behavioral types, after which any analysis of interest may be conducted.

3.1.b. Contribution Profiles by Type

[9] See also Burlando and Guala (2002) for an investigation of the accuracy of the strategy method within the FGF (2001) environment.

Our type classification results are based on data from the seven games that all subjects were able to complete. We treat available additional games as a holdout sample, as discussed below.[10]

Figures 4, 5 and 6 describe the contribution profiles for our three types of subjects. Each of these figures contains the same five pieces of information, the first three of which are: 1) a solid 45-degree line; 2) solid circles that describe how a type's mean contribution to the public good (rounded to the nearest integer) varies with the observed contribution of others; and 3) a vertical bar that extends two standard errors above and below each mean (when there are multiple observations).

Fourthly, the translucent squares describe the fit of a Tobit regression model given by equation (1), which is estimated by pooling the (non-rounded) data from all subjects of a given type, and under the constraints that individual contributions can be neither above 100% nor below 0% of the endowment. This is a parsimonious non-linear procedure to compare and contrast the contribution decisions of different behavioral types. (In the Figures' "A" panels the model is fit using the first seven games' data, while in the "B" panels it is based on the holdout sample.) The spacing of the squares, located wherever there is at least one data point, is the last piece of information, and provides an impression of the underlying sample that was used to estimate the model. Comparing the patterns in Figures 4A and 5A, for example, reveals that the distribution of the observed

[10] When comparing holdout sample behavior against a baseline, it is critical that the baseline be well estimated. This is the reason that we use a relatively large part of our sample, seven games, to establish baseline behavioral type characterizations. In some studies, particularly research with large field panel data sets, it is possible to use half of the observations (or more) as a holdout sample, because the baseline can be precisely estimated regardless. The limited amount of data available from our laboratory experiment precludes this possibility.

contributions of others includes more weight at lower contribution levels for free-riders than for cooperators.

Figure 4A describes the mean decisions of the 17 subjects that were classified as free-riders. It is clear from visual inspection that free-riders tend to contribute less than their counterparts' average at nearly all contribution levels. The difference is of course small and statistically insignificant when their counterparts contribute very little, because contributions are bounded below by zero. Figure 5A provides analogous information for the cooperator type. The cooperator contribution profile tends to imply contribution amounts that are greater than their counterpart's mean contributions.

Visual inspection of Figure 6A suggests that, as a first pass, the simple heuristic "contribute what others on average contribute" seems to characterize conditional cooperator contribution decisions quite well. It is worthwhile to emphasize that this heuristic is not implied by our classification approach. Recall that to be typed as a conditional cooperator requires only that a subject's contribution profile be upward sloping, below 25 at zero and above 25 at 50. Hence, for example, a subject with a relatively flat profile, ranging from 20 at zero to 30 at 50, would be classified as a conditional cooperator.

With respect to the fit of the Tobit regression, visual inspection indicates that the model fits the major features of the data well in all cases, and seems to differ substantially across behavioral types. Indeed, statistical tests reveal that these differences are statistically significant.[11] However, because these tests are based on the data used to

[11] Recall that we estimate a Tobit model for a particular type by pooling the data for all subjects of that type, and then estimating (1) under the restriction that contributions are bounded. Our interest is in testing whether the resulting intercept and slope estimates are jointly statistically significantly different between behavioral types. We ran these tests pairwise, which generate under the null hypothesis a statistic that is

create the types, they are also biased in favor of finding evidence of differences. Appropriate statistical tests based our holdout sample data are reported below.

Tables 2 and 3 provide information about the distribution of earnings and final contributions, respectively, for each of the three types over the seven games that were played by all subjects.

3.2. Evidence on POT and TAO Hypotheses

3.2.a. Evidence on the Persistence of Type Hypothesis

The persistence of type hypothesis is that there is temporal continuity to the behavioral strategies people adopt, at least within an experimental session. To test this hypothesis, consider first free-riders. Of the 17 subjects classified as free-riders based on their play in the first seven games, 12 are observed in holdout sample games. Figure 4B, which is analogous in information content to Figure 4A, describes the contribution profile of the 12 subjects in these holdout games. It is apparent that, on average, these 12 subjects continued to contribute less than their group members during the final games that they played. However, in visual comparison with Figure 4A, contributions seem relatively higher. This visual impression is confirmed by comparing the results of the Tobit analysis on the in and out of sample data. Table 1 shows that free-riders are estimated to increase their contribution by about 23 cents for every dollar contributed, on average, by

---

asymptotically distributed according to an F(2,n) distribution, where n is the degrees of freedom for the test. The possibility of individual-specific effects in the data means the number of degrees of freedom available for our tests is likely less than the number of observations in our data set (over three thousand). A conservative number of degrees of freedom to use is 28. This is the smallest number of distinct subjects included in any pairwise comparison (11 cooperators and 17 free-riders.) Hence, we use an F(2,28) as a conservative sampling distribution of our test statistics under the null hypothesis. The values of these statistics are 107 (CO = CC), 246 (FR = CC) and 349 (FR = CO), all of which are highly significant (P<0.001).

their group counterparts during the first seven games. Table 1 also shows that this estimate increases to 0.87 in the holdout sample, and the difference between the estimates is statistically significant (P-value<0.001). Hence, while the free-rider type tends to contribute less than their counterparts' mean both in and out of sample, it seems that free-riders contribute somewhat more in the holdout sample than their behavior during the first seven games might have led one to predict.

Figures 5B and 6B, and the second and third columns of Table 1, report the results of similar analyses for cooperators and conditional cooperators, respectively. The results for the 11 cooperators that we observed out of sample are analogous to what we found for free-riders: they continued to contribute generally more than their counterparts, but statistically significantly differently than they did in the first seven games. However, the 35 conditional cooperators observed in the holdout sample made statistically and visually identical decisions to those that they made in the first seven games.[12]

We take two main points from this data. First, while a strong from of the POT hypothesis is not supported, there is certainly evidence of continuity in subjects' decision making strategies. In particular, although there are statistical differences in behavior between the first seven games and the holdout sample by the FR and CO types, it is nevertheless true that each type's holdout sample is more similar to its own in-sample data than to that of any other type.[13] Also, differences in decisions between types are in

[12] Specifically, we tested the null hypothesis that the coefficient estimates based on the first seven games were identical to the estimates obtained from the data from the holdout sample. The test statistic is asymptotically distributed F(2,n) where n is the number of degrees of freedom. A conservative value of N to use is 12, which is the smallest number of distinct subjects in any sample comparison. The test statistics' values are 15.7 (P<0.001), 5.7 (P=0.02), and 0.2 (P=0.82) , for free-riders, cooperators and conditional cooperators, respectively.

[13] We ran the test described in footnote 10 using the holdout sample data, and obtained test statistics values of 11 (CO=CC), 25 (FR=CC) and 40 (FR=CO) which are significant at the 1% level for any n>5.

the expected direction, in that free-riders continue to contribute less than their counterparts on average, while cooperators tend to continue to contribute more.

Second, conditional cooperation seems to be an extremely stable strategy, and both FR and CO decisions move in the direction of conditional cooperation, on average, in the holdout sample. This suggests that even the strong types are somewhat facultative, changing their behavior as they gather information about the environment. We find this result comforting. In particular, from an evolutionary standpoint, it would be unusual to find utter inflexibility in an aspect of behavior as important as cooperation.

3.2.b. Evidence on the Types Affect Outcomes Hypothesis

The TAO hypothesis is that groups composed of more cooperative types tend to achieve more cooperative outcomes. Because subjects are assigned types, groups are distinguished by their type compositions. We assign a score to each group according to the number of each type that it includes. In particular, we assign a score of zero to each free-rider in a group, and scores of one and two to each conditional cooperator and cooperator, respectively. A group's score is then the sum of the scores of its members. Each of our groups contains four members, so that scores can range between a possible low of zero (all free-riders) to a high of eight (all cooperators), with higher scores suggesting greater cooperativeness.

We obtain group scores ranging from one to seven, because the random matching process did not generate any group that included only cooperators or only free-riders. The composition of groups with scores of one and seven can be easily deduced from the scoring procedure. Groups with other scores have multiple possible compositions. For example, a score of three can occur if there are three conditional cooperators and a free-rider, or if there is one cooperator, one conditional cooperator, and two free-riders. It turns out that in this particular experiment, 33 of 36 groups assigned a score of three were of the former type.[14]

The TAO hypothesis is informed by investigating how a group's score is related to its propensity to sustain cooperation. One reasonable measure of whether cooperation is sustained is the final amount contributed by a group to the public good, and over the first seven games we find that there is a positive relationship between this measure and our readily calculated group score.[15] Moreover, a regression analysis reveals that this relationship is statistically significant, and implies that a one point increase in the group cooperativeness score is associated with about a 20 token increase in the amount of the final contribution to the public good.[16] A similar analysis run on the groups observed in the holdout sample reveals a nearly identical relationship, and one cannot reject the hypothesis that the group score has the same effect on final group contributions both in and out of sample.[17]

---

[14] A detailed listing of the compositions we obtained in this experiment are available from the authors on request.

[15] As pointed out in section 2, the games lasted different numbers of rounds (between 8 and 36). Clearly, game length will also contribute to the amount of cooperative decay that a group experiences. When the number of observations on a particular group score is large, the game length effect is mitigated by the randomization procedures embedded in our design. For some group scores we have a quite small number of observations (see Figure 8). In these cases, the game length effect will tend to obscure the relationship between the group's cooperativeness score and the final group contribution, and therefore will tend to make it more difficult for this simple analysis to provide support for the TAO hypothesis. In this light, we find it remarkable that TAO is so strongly supported by our results.

[16] We ran a regression of the final contribution amount on an intercept and the group cooperativeness score. The coefficients' point estimates are 11 and 21, respectively, with only the 21 statistically significant (P-value < 0.001).

[17] The same regression run on the holdout data generated coefficient estimates of 10 and 25, respectively, with only 25 statistically significant (P-value < 0.01). The null hypothesis that the intercept and slope parameters are jointly identical for the first seven games and holdout sample cannot be rejected at standard significance levels (P-value = 0.59).

Figure 7 details the ability of our simple group score to predict a group's final public contribution. This figure provides the mean final group contribution (over the first seven games) for each group score, and a two standard error interval above and below that mean. It also plots the realized final group contribution for those groups observed in the holdout sample. The absolute difference between the in-sample means and the holdout sample means are close, with the greatest deviation being about 26 tokens (when the cooperativeness score is six.) Moreover, the holdout sample means are generally within, or nearly within, two standard errors of the in-sample means. Of course not all of these standard errors are tight, because, as indicated in the figure, the number of observations for some group scores is quite small (e.g., two and three observations during the first seven games on groups with scores of seven and one, respectively.) Nevertheless, this evidence suggests that our simple cooperativeness score predicts final group contributions well.

Figures 8A through 8D provide additional detail on the cooperative dynamics of in-sample and out-of-sample groups in our sessions. On these figures, the horizontal axis indicates the round, and the vertical axis the equally-weighted (over groups) mean contribution to the public good. We include only groups with cooperativeness scores between two and five, because the number of observations on groups with other scores is too low (either one or two) to allow reasonable inferences with respect to dynamics.

The pattern of mean contributions over rounds in the first seven games is a non-parametric prediction of cooperative dynamics in the holdout games. The vertical bars extend two standard errors above and below each in-sample mean. It is visually evident that the cooperative dynamics in our holdout sample are strikingly well predicted by in-sample behavior. In particular, for groups with scores of 2, 4 and 5, nearly all of the holdout sample's mean contributions are within two standard errors of the predicted means. The fit is least strong for groups with scores equal to three, but even in this case visual inspection suggests that the predictions are "good," and still over half of all holdout means are within a two-standard error band of the predicted value. This is further evidence in support of the TAO hypothesis, and that our naïve group scoring procedure, which is clouded by game length effects and the fact that multiple compositions are possible for a single score, is reasonably able to predict outcomes.

These findings have implications for the POT hypothesis. The fact that we find within session consistency in behavior is comforting, but POT will be more valuable to the extent that there is between-session consistency. Future behavioral type-classification research might profitably be directed towards investigations where subjects participate in the same experiment over a period of days or weeks to assess consistency over more substantial periods of time. Research investigating "social value orientation" is encouraging in this respect (e.g., Van Lange & Semin Goossens, 1998). Moreover, it would be useful to obtain evidence of type-persistence between games. The type construct would be strengthened by finding that, for example, CC players in our sequential public goods game tend to be tit-for-tat players in a prisoners' dilemma game.

4. Conclusion

Conditional cooperation, or reciprocity, is the focus of considerable theoretical and empirical attention, and rests on two hypotheses, which we labeled "Persistence of Type" and "Types Affect Outcomes," both of which are in need of direct empirical investigation. The novel sequential public goods game described in this paper provides a new opportunity to observe players' contribution decisions conditional on information about others' current contributions, and provides a clean way to test the two fundamental hypotheses.

The Persistence of Type hypothesis is that an individual's behavioral "type" well characterizes his or her decisions within an experimental session (at least). We developed a simple and reliable type classification algorithm and used it to assign types to our subjects. Over half of our subjects were classified as conditional cooperators based on their behavior early in the experiment. Many of those same subjects were observed in a holdout sample, and we found their decision strategy in the holdout sample to be statistically identical to the strategy they used earlier in the experiment. Similarly, subjects who were classified as free-riders (cooperators) and observed in the holdout sample tended to contribute less (more) than their group counterparts in both cases.

The existence of discernable types resonates with models of the evolution of cooperation. Though there is not complete consensus in this area, some models suggest that populations can stabilize at equilibria in which there is heterogeneity in terms of altruistic tendencies, either because agents play mixed strategies, or because there are different kinds of agents (see, e.g., Gintis, 2000, Maynard Smith, 1982). It is plausible that clarifying the frequency of different types in the population can help to inform these models.

The second hypothesis is that "Types Affect Outcomes," or that a group's cooperative dynamic is affected by its type composition. We found support for this, in that groups that contained more cooperative types tended to cooperate more. We then suggested a simple score statistic to summarize a group's type composition. We found that this score was statistically significantly correlated with a group's ability to sustain cooperation, and moreover, that a group's score statistic is a reliable predictor of the quantitative extent to which a group will sustain cooperation in our environment.

Note that the ability to predict contributions in a VCM experiment is not necessarily surprising. It is regularly found, for example, that subjects in a standard linear VCM (strangers condition) contribute about half of their endowment, on average, and after 10 rounds decay to around 10% of their endowment. We do find it surprising, however, that our simple group score statistic, based on our simple individual type-classification algorithm, has a remarkable ability both to separate and to predict group cooperative dynamics.

In the concluding section of their stimulating article, Fishbacher, Gachter and Fehr (2001) begin to speculate tentatively on the implications of their results (based on a one shot game) for group cooperative dynamics. Having found evidence of both free-riding and conditionally cooperative strategies, they point out that groups that include both types might be expected to experience cooperative decay and convergence to a non-cooperative equilibrium. They go on to speculate that, "The speed of convergence depends on the actual composition of the group." (Section 4, p. 403). The results of this paper provide the first direct evidence in support of these and other closely related hypotheses that have been previously advanced only speculatively in the public goods literature. More broadly, because the theoretical efficiency of an institution usually relies on some form of behavioral continuity, evidence that individuals' behavioral rules exhibit temporal stability is valuable to any interested in the design of new mechanisms.

Appendix A. Transcript of Subjects' Instructions

1 Welcome. Instructions and messages will appear in this part of the screen. Since we have now begun, please keep your attention on your own computer screen and stay silent throughout this experiment unless otherwise instructed. To go to the next instruction screen click on the 'forward' button below.

2 This experiment is a study of group and individual investment behavior. Reading and following these instructions carefully will help you to make good investment decisions and, possibly, to earn a considerable amount of money. All of your earnings will be paid to you privately, in cash, at the end of the experiment. A research foundation has provided the funds for this study.

3 You will be in a group consisting of four players. The other players in your group will be people sitting in this room.

4 Each player in your group will be given a certain number of tokens to invest. You will choose to divide your tokens between two investment opportunities: The Individual Exchange and the Group Exchange. Any tokens invested into these exchanges will be turned into cash, as described below, and paid to you at the end of the experiment. Any tokens that are not invested have zero value at the end of the experiment.

5 THE INDIVIDUAL EXCHANGE Every token you invest in the Individual Exchange will earn you a return of one cent at the end of the experiment. (Examples are given to the right)

Example: Suppose you invested 28 tokens in the Individual Exchange. Then you would earn $0.28 from this exchange. Example: Suppose you invested 42 tokens in the Individual Exchange. Then you would earn $0.42 from this exchange. Example: Suppose you invested 0 tokens in the Individual Exchange. Then you would earn nothing from this exchange.

6 THE GROUP EXCHANGE Investments in the Group Exchange benefit everybody in your group. Every token you invest in the group exchange returns 0.5 cents to every member of your group including yourself. Therefore, the total amount you earn from the Group Exchange depends on the total number of tokens that you and the other three members of your group invest in the Group Exchange. The more the group invests in the Group Exchange, the more it returns to each member of the group. The process is best explained by a number of examples, again given to the right.

EXAMPLE: Suppose that you decided to invest no tokens in the Group Exchange, but that the three other members invested a total of 50 tokens. Then your earnings from the Group Exchange would be $0.25. Everyone else in your group would also earn $0.25 _

EXAMPLE: Suppose that you decided to invest 20 tokens in the Group Exchange, and the three other members invested a total of 40 tokens. This makes a total of 60 tokens. Then your earnings from the Group Exchange would be $0.30. Everyone else in your group would also earn $0.30 EXAMPLE: Suppose that you decided to invest 30 tokens in the Group Exchange, but that the three other members invest nothing. Then you, and everyone else in the group, would get a return from the Group Exchange of $0.15.

7 THE GROUP EXCHANGE (continued) Every token invested in the Group Exchange will earn one half of a cent for every member of the group including the person who made the investment. It does not matter who invests in the group exchange. Everyone will get a return from every token invested whether they invested in the group exchange or not. The table below has been provided for your reference during the experiment. The table lists what your payoff would be from the Group Exchange given the number of tokens in the exchange.

8 THE INVESTMENT DECISION Your task is to decide how many of your tokens to invest in the Group Exchange. You are free to put some tokens into the Individual Exchange and some into the Group Exchange. Alternatively, you can put all of them into the Individual Exchange or all of them into the Group Exchange.

9 STAGES OF INVESTMENT The game begins with each member making an initial investment into the Group Exchange. After each initial investment has been made the game proceeds by giving each person in your group the opportunity, in turn, to change their investment decision. When it is your turn to change your investment decision you will be told the average amount that each other player in your group has contributed to the group exchange. The opportunity to change the amount invested in the group exchange continues to pass from person to person in your group until at some randomly determined point the game ends and payoffs are made. You will have at least one opportunity to change your investment in the group account before the game ends, and you may have several. Your earnings are determined by the amount you have invested in the Group and Private Exchanges at the point when the game ends.

10 Remember that your earnings when the game ends are the sum of the returns from the tokens you placed in your Individual Exchange plus the return from the total number of tokens placed in the Group Exchange.

11 You will play this game several times. The members of your group will change from game to game. The point at which the game ends will also change from game to game.

12 At the beginning of each game you are given an Endowment of 50 tokens. You are to indicate by filling in the blank below your investment in each account. You simply enter the number of tokens you want to place in the Group Exchange. The number of tokens in the Individual Exchange will automatically be entered so that the sum of your investments equals your endowment, 50 tokens. The other players in your group will also have 50 tokens to invest. You must make your investment decisions without knowing

what the other players in your group are investing. When you have made your investment decisions you will click on the red ‘submit’ button.

13 When the game ends you will be told the total number of tokens currently invested in the Group Exchange as well as your earnings from the game.

14 You have now completed the instructions. If you have any questions, raise your hand and ask the experimenter. Otherwise, click on the large green ‘start’ button. _ When you do so, you will exit the instructions. Please be sure that you have understood the instructions before continuing.

INITIAL ROUND:

Please choose an initial investment amount for the Group Exchange. When you are done click on the red ‘submit’ button.

BETWEEN ROUNDS, THE INSTRUCTIONS WERE:

It is now your turn to change your investment to the Group Exchange if you would like to do so. When you are done making your changes, if any, click on the red 'submit' button.

References

Andreoni, J., 1995. Cooperation in public-goods experiments: kindness or confusion? American Economic Review 85:4,September, 891-904.

Andreoni, J. and M. McGuire. 1993. Identifying the free riders: A simple algorithm for determining who will contribute to a public good. Journal of Public Economics. July. 51:3. 447-54.

R. Boyd and P. J. Richerson. Forthcoming. Solving the Puzzle of Human Cooperation, In: *Evolution and Culture*, S. Levinson ed. MIT Press: Cambridge MA.

Brandts, J. and A. Schram. 2001. Cooperation and noise in public goods experiments: applying the contribution function approach. Journal of Public Economics, 79. 399-427.

Burlando, R. and F. Guala. 2002. Conditional cooperation: New evidence from a public goods experiment. Manuscript, University of Torino.

Camerer, C. and E. Fehr. 2003. Measuring social norms and preferences using experimental games: A guide for social scientists. In Henrich, Boyd, Bowles, Camerer, Fehr, Gintis and McElreath (Eds.), Foundations of Human Sociality – Experimental and Ethnographic Evidence from 15 Small-Scale Societies.

Clark, K. and M. Sefton. 2001. The sequential prisoner's dilemma: Evidence on reciprocation. Economic Journal, 111:3. March, 357-62.

Congleton, R. and V. Vanberg. 2001. Help, Harm or Avoid? On the personal advantage of dispositions to cooperate and punish in multilateral PD games with exit. Journal of Economic Behavior and Organization. 44:2, February. 145-67.

Croson, R. 1998. Theories of altruism and reciprocity: evidence from linear public goods games. Manuscript, University of Pennsylvania.

El-Gamal, M. and D. Grether. 1995. Are people Bayesian? Uncovering behavioral strategies. Journal of the American Statistical Association, 90:432, December. 1137-45.

Fischbacher, U., S. Gachter and E. Fehr. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters. 71, 397-404.

Funder, D.C. 2001. The Personality Puzzle (2nd Edition). New York: Norton.

Gintis, H. 2000. Strong reciprocity and human sociality. Journal of Theoretical Biology, 206:2, September 169-179.

Gunnthorsdottir, A., D. Houser, K. McCabe and H. Ameden. 2003. Disposition, History and Contributions in Public Goods Experiments. Manuscript. George Mason University.

Heiner, R. 2002. Robust evolution of contingent cooperation in pure one-shot prisoners' dilemmas (part I and part II). Working paper, George Mason University.

Houser, D. 2003. Classification of Types of Dynamic Decision Makers, in Nadel, L. (ed.), Encyclopedia of Cognitive Science, Vol. 1, Nature Publishing Group, London. 1020-1026.

Houser, D. and R. Kurzban. 2002. Revisiting kindness and confusion in public goods experiments. American Economic Review. 92:4, September. 1062-69.

Houser, D. and J. Winter. 2003. How do behavioral assumptions affect structural inference? Evidence from a laboratory experiment. Journal of Business and Economic Statistics. Forthcoming.

Houser, D., M. Keane and K. McCabe. 2002. Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. Manuscript, George Mason University.

Isaac, M. and J. Walker. 1988. Group size effects in VCM provision: The voluntary contribution mechanism. Quarterly Journal of Economics, 179-199.

Keser, C. and F. van Winden. 2000. Conditional cooperation and voluntary contributions to public goods. Scandinavian Journal of Economics. 102:1, 23-39.

Kurzban, R. and D. Houser. Individual Differences in Cooperation in a Circular Public Goods Game. *European Journal of Personality*. **15**, 37-52.

Kurzban, R., K. McCabe, V. Smith and B. Wilson. 2000. Incremental commitment and reciprocity in a real-time public goods game. Personality and Social Psychology Bulletin. 27:12. 1662-1673.

Ledyard, J. 1995. Public goods: a survey of experimental research. In Kagel, J. and Roth, A. (Eds.), Handbook of Experimental Economics. Princeton University Press. Princeton.

Maynard Smith, John. 1982. Evolution and the Theory of Games. Cambridge: Cambridge University Press.

McCabe, K., M. Rigdon and V. Smith. 2002. TITLE? Journal of Economic Behavior and Organization. Forthcoming.

Rabin, M. 1993. Incorporating Fairness into Game Theory and Economics. American Economic Review, 83:5, December. 1281-1302.

Palfrey, T.R. and J.E. Prisbrey. 1997. Anomalous behavior in public goods experiments: how much and why. American Economic Review 87:5, December, 829-846.

Selten, R. 1967. Die Strategiemethode zuer Erforschung des eingeschrankt rationalen Verhaltens im Rahmen eines Oligopolexperimentes. In Sauermann, H. (Ed.), Beitrage zur experimentellen Wirtschaftsforschung. J.C.B. Mohr (Paul Siebeck), Tubingen. 136-168.

Van Lange, P. A. M., and A. Semin-Goossens, A. 1998. The boundaries of reciprocal cooperation. European Journal of Social Psychology, 28: 847-854.

Table 1

Contribution Profile Estimates by Type

	Free Riders	Conditional Cooperators	Cooperators
Mean Net Group Contribution First Seven Games	**0.22**	**1.15**	**0.69**
Standard Error	***0.07***	***0.04***	***0.21***
Mean Net Group Contribution Holdout Sample	**0.83**	**1.21**	**2.18**
Standard Error	***0.12***	***0.09***	***0.43***
First Seven Games Dummy	**-4.33**	**-1.45**	**46.73**
Standard Error	***1.86***	***1.24***	***6.48***
Holdout Dummy	**-11.09**	**-3.05**	**-2.51**
Standard Error	***3.49***	***2.48***	***13.09***

Table 2

Earnings Distributions by Type over First Seven Games

	Distribution of Earnings per Game				
Type	**mean**	**SD**	**median**	**min**	**max**
FR	**76.45**	**18.10**	**77.5**	**45**	**125**
CO	**70.44**	**24.18**	**72.5**	**25**	**110**
CC	**71.69**	**19.13**	**70.0**	**30**	**125**

Table 3

Distribution of Final Contribution by Type over First Seven Games

	Contribution				
Type	**mean**	**SD**	**median**	**min**	**max**
FR	**7.34**	**10.75**	**1**	**0**	**50**
CO	**36.68**	**19.20**	**50**	**0**	**50**
CC	**24.48**	**19.10**	**25**	**0**	**50**

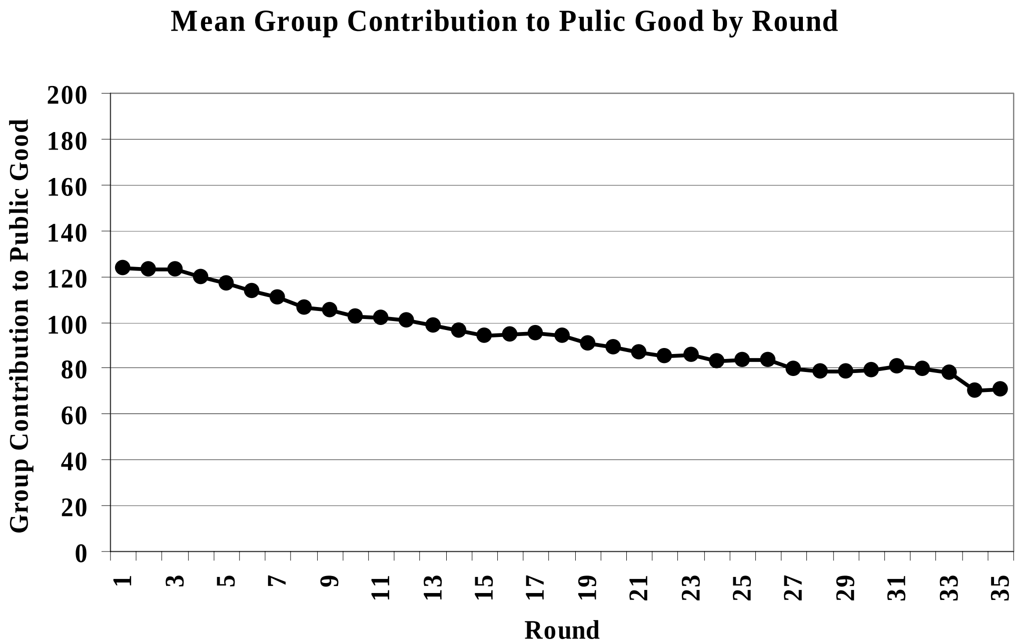

Figure 1.

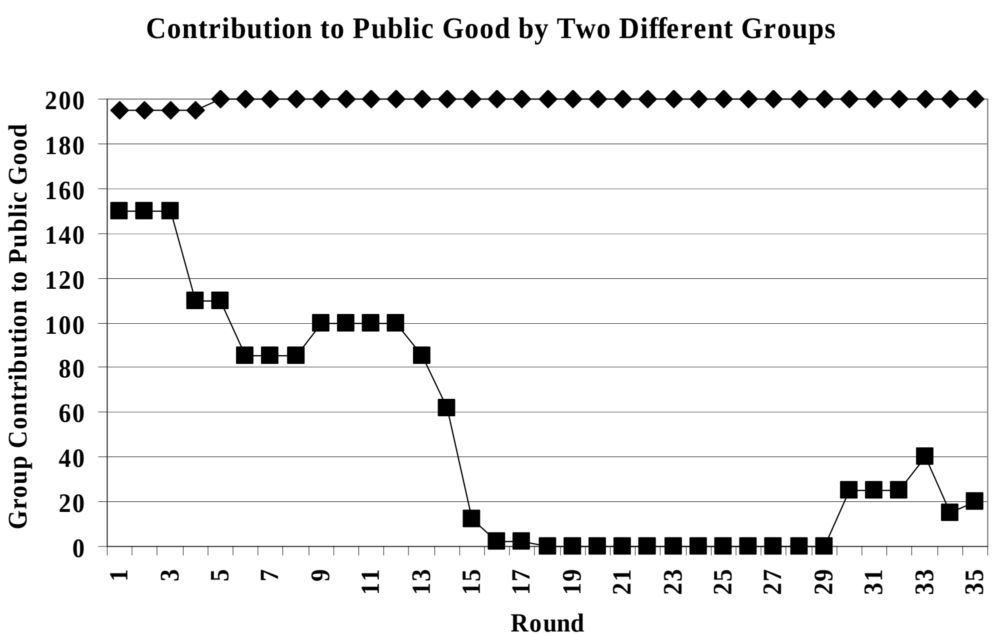

Figure 2.

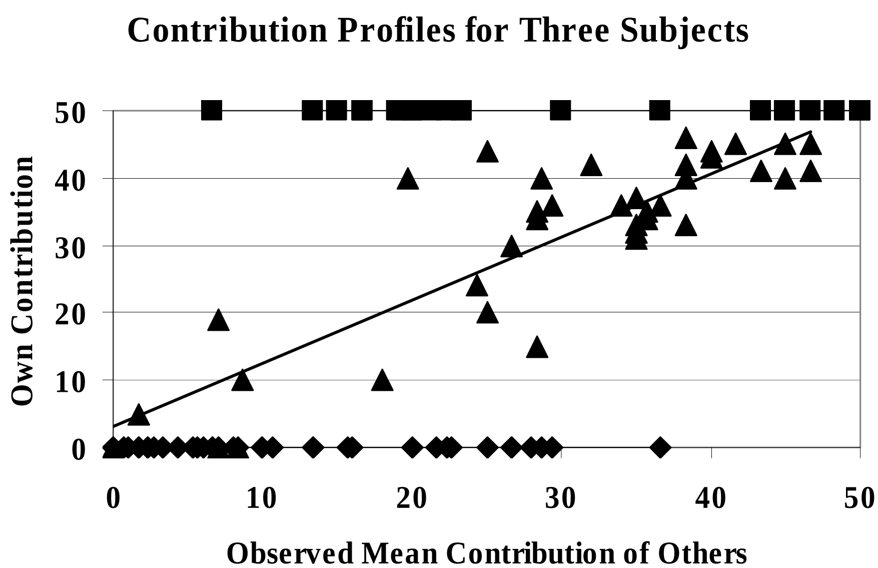
Contribution Profiles for Three Subjects
Own Contribution
50
40
30
20
10
0
0
10
20
30
40
50
Observed Mean Contribution of Others

Figure 3.

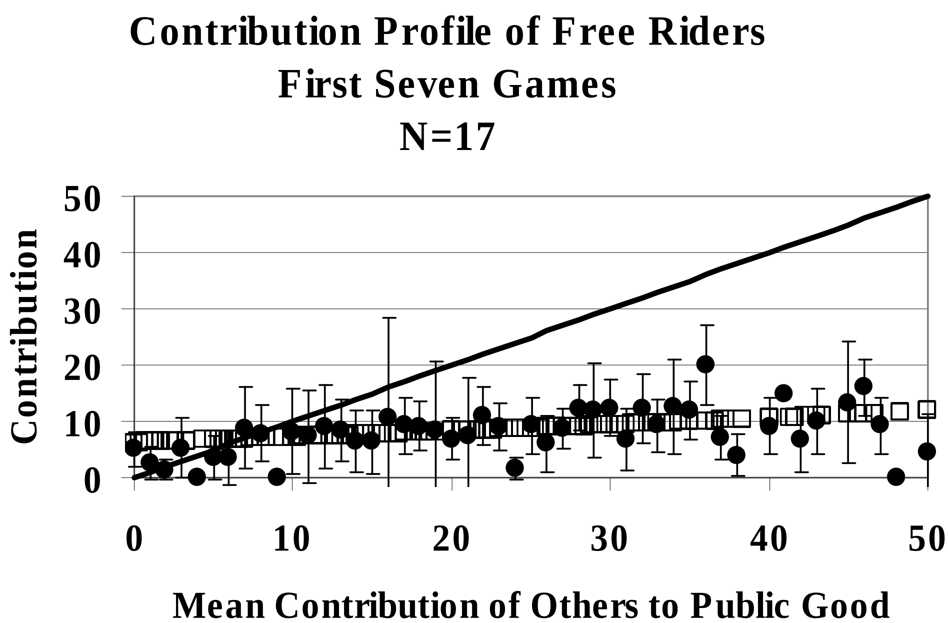

Figure 4A.

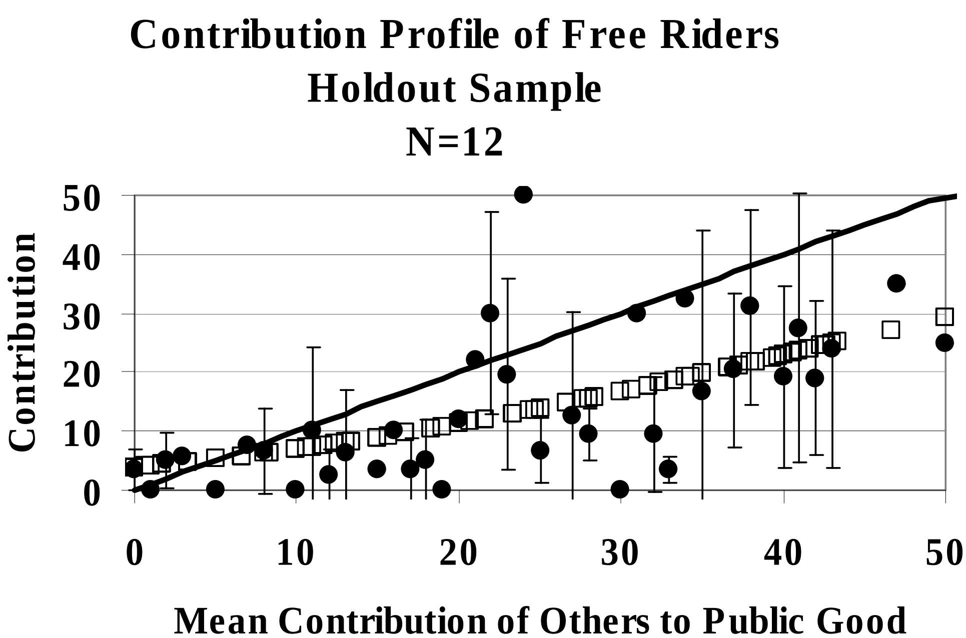

Figure 4B.

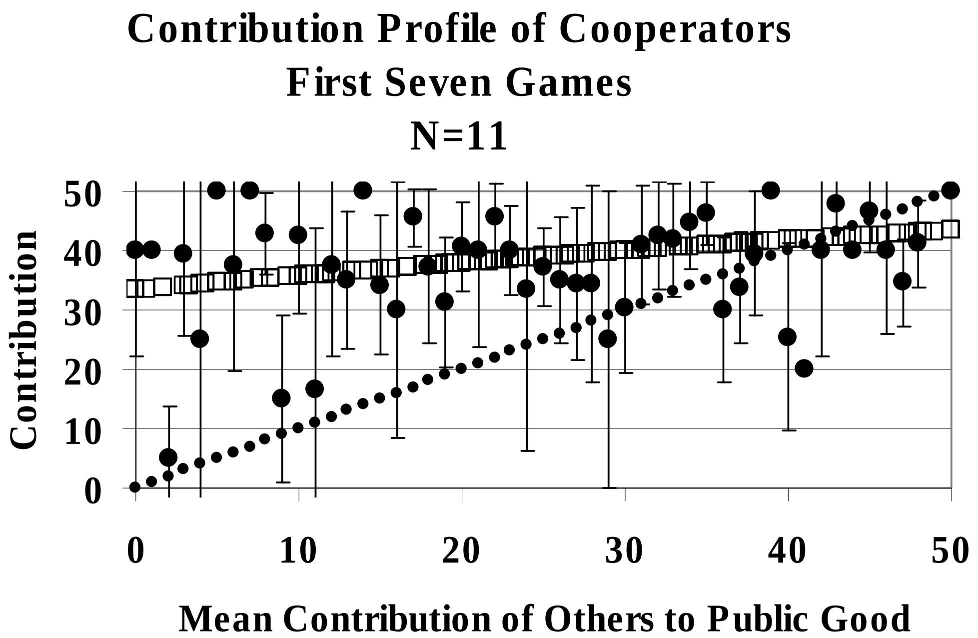

Figure 5A.

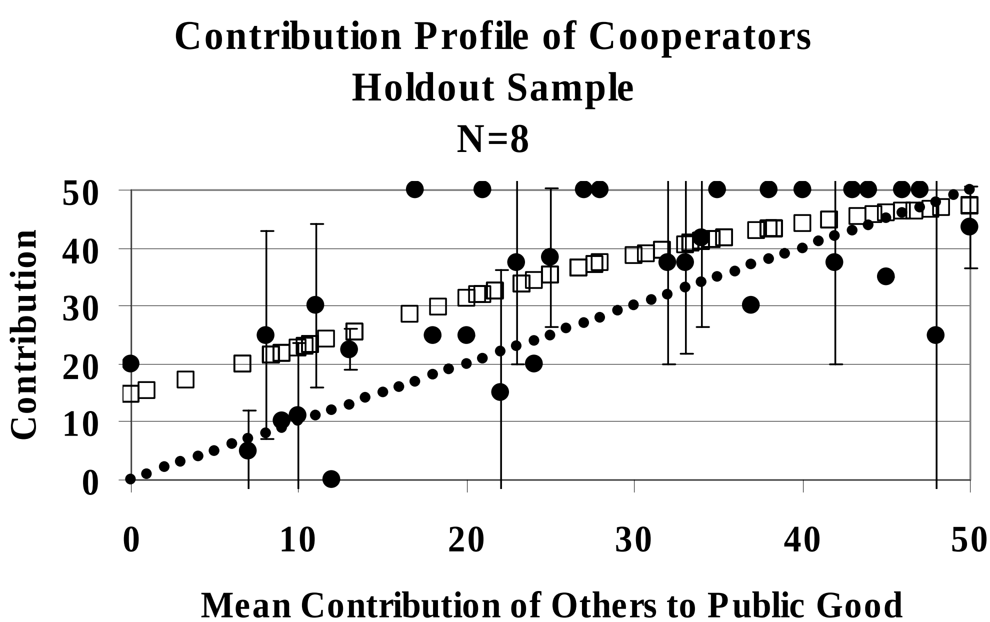

Figure 5B.

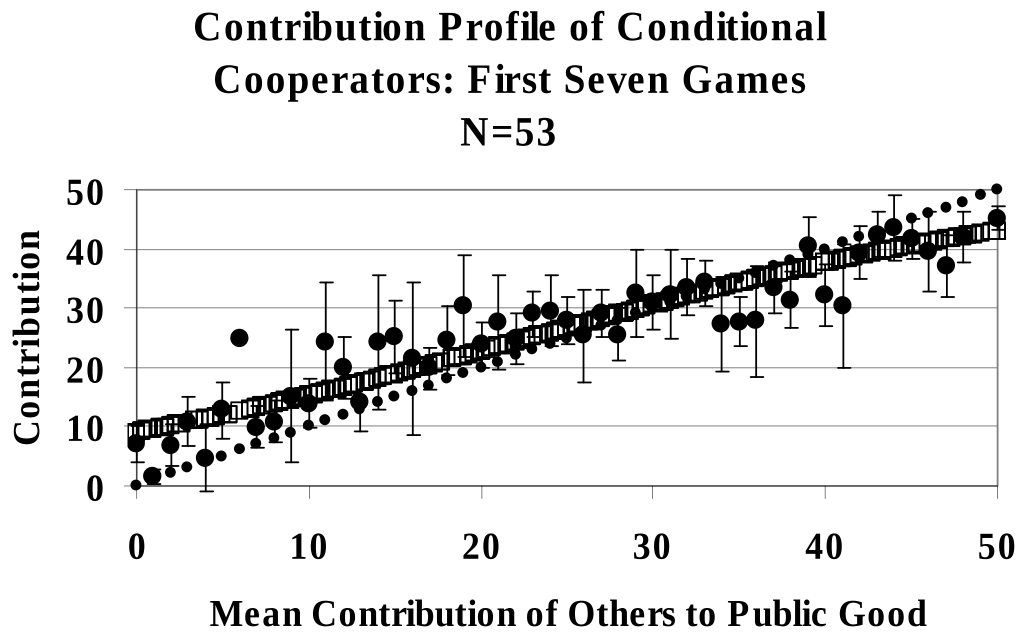

Figure 6A.

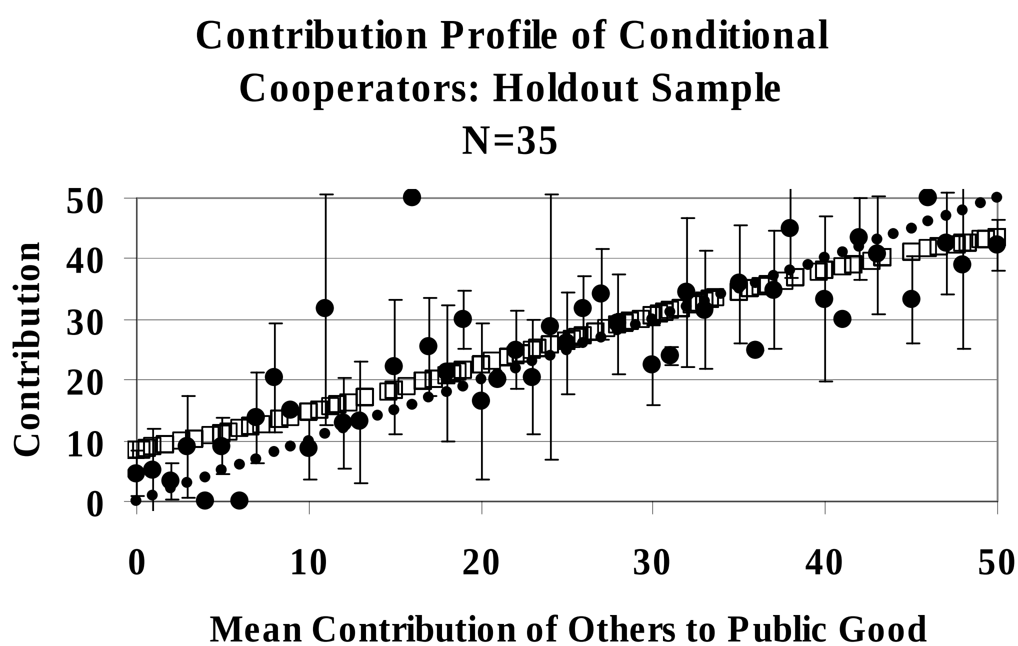

Figure 6B.

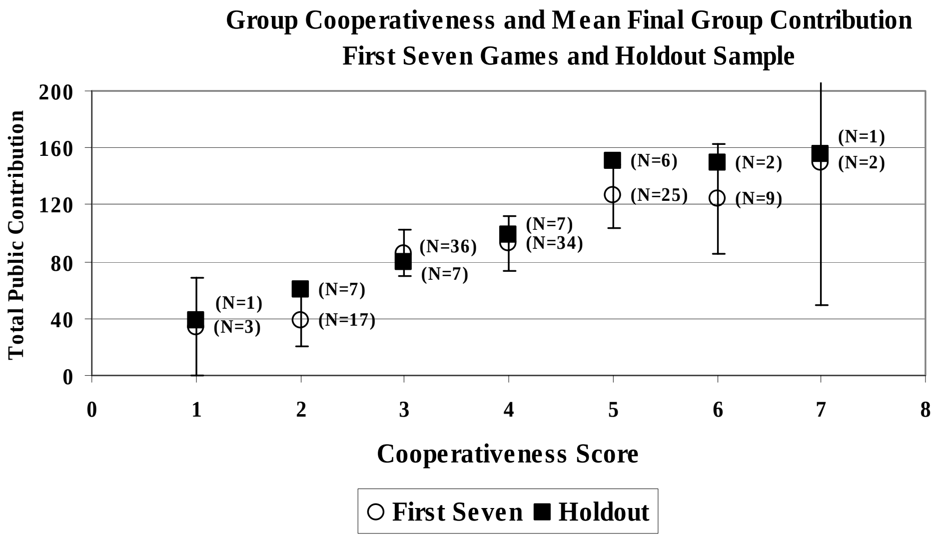
Group Cooperativeness and Mean Final Group Contribution
First Seven Games and Holdout Sample
Total Public Contribution
200
160
120
80
40
0
0
1
2
3
4
5
6
7
8
Cooperativeness Score
(N=1)
(N=3)
(N=7)
(N=17)
(N=36)
(N=7)
(N=7)
(N=34)
(N=6)
(N=25)
(N=2)
(N=9)
(N=1)
(N=2)
First Seven
Holdout

Figure 7.

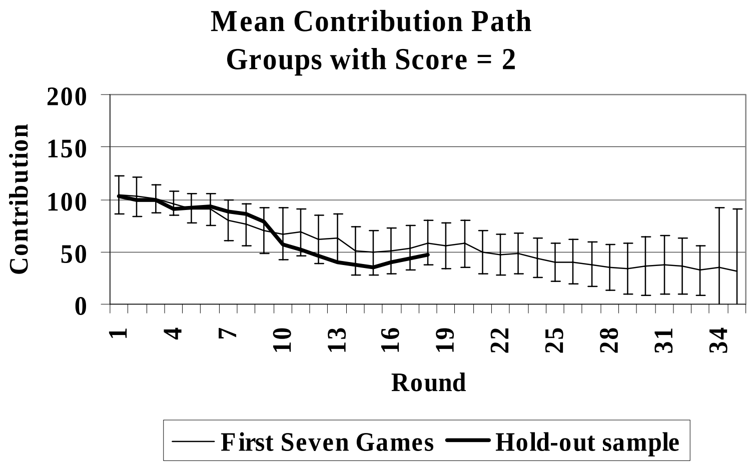

Figure 8A.

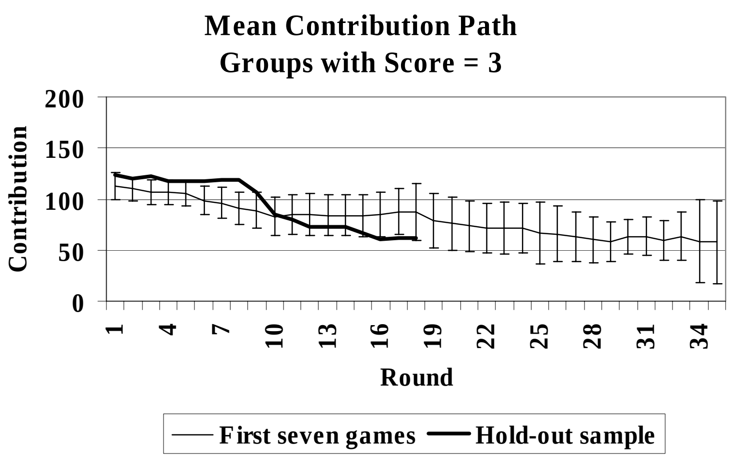

Figure 8B.

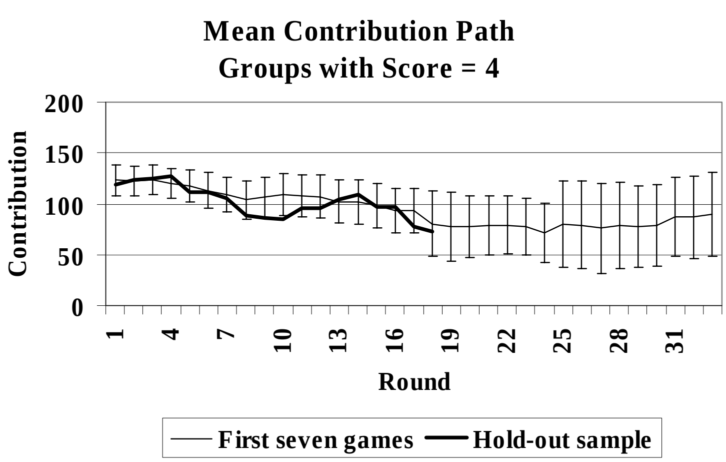

Figure 8C.

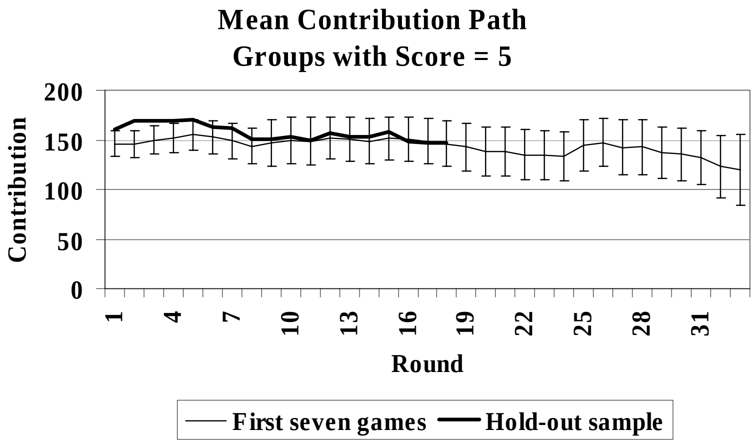
Mean Contribution Path
Groups with Score = 5
Contribution
200
150
100
50
0
1
4
7
10
13
16
19
22
25
28
31
Round
First seven games
Hold-out sample

Figure 8D.