

Cooperation over finite horizons: a theory and experiments*

Attila Ambrus[†] and Parag A. Pathak[‡]

This Version: July 2006

Abstract

This paper proposes a theory of cooperation over finite horizons, focusing on public good contribution games, that implies the broadly documented feature of decreasing cooperation over time. The central assumption is that there are two types of players: those who only care about their own material payoffs, and those who reciprocate others' contributions. The main result is that if reciprocity functions satisfy some regularity conditions, then generically there is a unique perfect equilibrium, in which contributions are decreasing. In this equilibrium, selfish players contribute to induce subsequent contributions by reciprocal players, and this incentive diminishes as the end of the play approaches. The model explains the puzzling restart effect and is consistent with various other empirical findings. In one-shot games, the model predicts no contributions.

We also report the results of a series of experiments, using a probabilistic continuation design in which after each round, the game is restarted with low probability. The results support the implications of our model that the restart effect is present even with experienced players, whereas, in one-shot games, contributions disappear with experience. We show that experienced players correctly foresee the pattern of contributions, suggesting that the declining pattern comes from equilibrium play. We also identify the presence of conditional reciprocity among experienced players, and document that selfish players (identified exogenously) stop contributing earlier than reciprocal players, as implied by the model.

*We would like to thank Satoru Takahashi for a careful reading of the paper. We would also like to thank Drew Fudenberg, Imran Rasul, Al Roth, and participants of the 2005 SITE meeting on Experimental Economics for useful comments. The experiment was programmed and conducted with z-Tree; we are grateful to Urs Fischbacher for making this software available. Pablo Guillen and the staff of the Computer Lab for Experimental Research at Harvard Business School were helpful in conducting the laboratory sessions. For financial support, Pathak thanks the National Science Foundation and the Division of Research at Harvard Business School.

[†]Department of Economics, Harvard University, Cambridge, MA 02138, ambrus@fas.harvard.edu, <http://post.economics.harvard.edu/faculty/ambrus/ambrus.html>.

[‡]Department of Economics, Harvard University, Cambridge, MA 02138, ppathak@fas.harvard.edu, <http://www.people.fas.harvard.edu/~ppathak>.

1 Introduction

There are many economic and social relationships in which participants interact with each other repeatedly, but for only a commonly known finite time. For instance, consider manager-employee relationships in which the employee has a nonrenewable fixed period employment contract, or faces mandatory retirement at a known time. Other examples include individuals working together as part of a team, where the team is only together for a pre-specified amount of time, or the owner-tenant relationship, when it is commonly known that tenants will not renew their lease. In these games, if at every stage participants face a myopic incentive to choose a noncooperative action, then no matter how long the interaction lasts, backwards induction implies that the unique subgame perfect equilibrium outcome involves playing the noncooperative action every period.

In contrast to this theoretical prediction, there is considerable experimental evidence that repeated interaction facilitates cooperation, even with a known finite horizon. Moreover, in many different environments, the same pattern of play emerges: there is a relatively large amount of cooperation in early rounds, which eventually breaks down, and by the end of the interaction, there is very little cooperation. This phenomenon is documented in finitely repeated prisoner's dilemma games (Selten and Stoecker [86], Andreoni and Miller [93], Kahn and Murningham [93]), repeated duopoly games (Selten et al. [97]), centipede games (McKelvey and Palfrey [92]), and repeated voluntary public good contribution games (in many papers, starting with Isaac et al. [84] and Kim and Walker [84]). Outside the laboratory setting, Bandiera et al. [06] find evidence from a field experiment that temporary workers working together for a fixed period of time cooperate under a relative incentive scheme by holding back their efforts, provided that they can monitor each other's decisions.

In this paper, we present a theory that explains the documented pattern of cooperation in games with finite horizon, and we report the results of a series of experiments designed to test the validity of the theory. We restrict the formal analysis to the context of voluntary public good contribution games, although our ideas can be applied to the other contexts mentioned above.

We focus on investigating voluntary contributions to a public good in a group setting because it has been a remarkably active research topic in experimental

economics, as well as in various other disciplines such as sociology and anthropology.¹ The experiments typically involve games in which players decide how much of their endowment to allocate to a public good and how much they retain for private investment. Payoffs are such that, in a one-shot game, any contribution to the public good is a dominated strategy, but the aggregate benefit of contributing to the public good exceeds the individual benefit of a private investment.

In experiments, average contribution levels are observed to be positive both in a one-shot contribution game and in repeated versions of the game. In finitely repeated versions, the decreasing cooperation phenomenon is prevalent: at the end of the game, contributions are very small.² The latter observation also holds for games played by experienced subjects.³ Furthermore, it remains true even in games that follow a “surprise restart” announcement. Namely, at the end of a repeated public good contribution game, if a surprise announcement is made that the same group of subjects will play another repeated game, contributions initially jump back to a relatively high level and then decrease again over time. This “restart effect” was first reported by Andreoni [88] and was later confirmed by several other studies.⁴ In contrast, existing evidence suggests that in one-shot games, contributions diminish when players become more experienced.⁵

These findings seem to be incompatible with the assumption that all subjects are rational in the traditional sense, that they care only about their own material payoffs.⁶ Standard game theoretic solution concepts (Nash equilibrium, subgame perfect Nash equilibrium, correlated equilibrium) predict that players contribute zero to the public good in every period. Theories proposed so far to explain the observed patterns of contributions can be divided into three rough categories: (i) players make mistakes, but they learn over time and therefore contribute less; (ii) players do not just maximize their own monetary payoffs, but their preferences involve altruism, reciprocity, or “warm-glow effects” (Andreoni

¹For an early survey, see Ledyard [95]. Fiske [92] and Field [02] are relatively recent references in sociology and anthropology, respectively.

²According to Fehr and Schmidt [02], in the final period, roughly 75% of the subjects contribute nothing to the public good, and the rest contribute very little.

³See Isaac et al. [88].

⁴See for example Croson [96].

⁵See Isaac et al. [84] and Andreoni [88].

⁶Fudenberg and Levine [97] argue that in several well-known experiments, subjects’ observed behavior is consistent with self-confirming epsilon-equilibrium, but they mention that in voluntary contribution experiments, players’ losses seem to be too large for this to be the case.

[89]); (iii) reputational considerations along the lines of Kreps and Wilson [82], i.e. making the other players believe that one is not a selfish utility maximizer, play a big role at the beginning of the game, but over time, reputation wears out. However, none of the above theories can explain all the robust findings of the experiments. Theories of altruism, reciprocity, or warm-glow effects can explain positive contribution levels, but not in any natural way the decreasing pattern of contributions in repeated games. Learning or reputation can explain positive contribution levels and the decreasing level of contributions, but not the restart effect. If players learn not to contribute by the end of the game, then contributions should not jump back to a high level.

We present a theory that explains the initially positive but decreasing pattern of contributions as an equilibrium phenomenon. The basic feature of the model is that we assume the presence of both selfish players (in the sense of maximizing only their own material payoffs), and players who reciprocate contributions by others. This assumption is motivated by the empirical finding that most subjects in public good contribution experiments can be classified into these two categories. According to the observations of Fischbacher et al. [01], Brandts and Scram [01], Palfrey and Prisbey [97], Ledyard [95], and Saijo and Yamaguchi [92], roughly half of the subjects in public good experiments maximize individual payoffs, while 40-50% of them are conditional cooperators.⁷ In a study of real-life collective action by a group of Shuar hunter-horticulturalists in Ecuador, Price [06] finds that group members accurately distinguished intentional cooperators and noncooperators. Moreover, several papers suggest the importance of investigating the nature of strategic interaction between different types of players. Fehr and Schmidt [02] argue that “the interaction between fair and selfish individuals is the key to the understanding of the observed behavior in strategic setting” and that “even if we do not yet have a fully satisfactory model of fair behavior, one can probably go a long way with simple models that take into account the interaction between selfish and fair types.” But to the best of our knowledge, our paper is the first one that provides a formal investigation of this strategic interaction in a dynamic setting.

Motivated by existing experimental evidence, we assume that the reciprocal players reciprocate both past realized contributions and current expected con-

⁷There is overwhelming evidence that very few people behave unconditionally altruistically in these experiments. Rabin [94] writes: “If people do not think that others are doing their fair share, then their enthusiasm for sacrificing for others is greatly diminished.”

tributions by others. Sonnemans et al. [99] (also Keser and Winden [00]) find both forward-looking and backward-looking (adaptive) behavior in public good experiments.⁸ The reciprocal preferences we adopt can be derived from various underlying motives, including fairness considerations, conditional altruism, or following some social norm. Furthermore, since we allow reciprocity functions to treat different types of players differently, our model is compatible with the assumption that reciprocal players care not only about how much each other player contributes to the public good, but also about what motivates another player to contribute (intentions).

Our central claim is that dynamic equilibrium interaction between selfish and reciprocal players implies all the robust features of contribution paths observed in experiments. Formally, we show that if reciprocity functions satisfy some regularity conditions, there is at least one selfish player in the game, and after every period, players are informed of every other group member's contribution, then generically the game has a unique subgame perfect Nash equilibrium. This equilibrium exhibits a decreasing pattern of contributions. Moreover, the same pattern is implied by a perfect equilibrium in a game in which, after each period, only the aggregate contribution of the group is revealed, although in this context the equilibrium is not necessarily unique. The intuition behind these results is that selfish players can influence future contributions of reciprocal players, and the more periods are left, the higher the increment they can induce on these contributions. This makes it worthwhile for them to contribute more of their endowment to the public good at the beginning of the game. In equilibrium, reciprocal players correctly anticipate these high contribution levels in early periods, which induces them to also contribute. As the game progresses, selfish players have less incentive to contribute, and in equilibrium, their contributions to the public good decrease. In particular, their contribution is always zero in the last period of the game. Lastly, decreasing contributions by the selfish players imply decreasing contributions by the reciprocal players, although the rate of decrease in their contribution may be smaller, since they might reciprocate past contributions as well, besides current expected ones.

Our model explains both the decreasing pattern of contributions and the

⁸More direct evidence comes from Keser [00]. In her experiment, players played multiple sessions of repeated contribution games and selected repeated game strategies from a large list of choices. The most popular strategy chosen by experienced players started out contributing, then reciprocated others' average contribution in the previous period and did not contribute anything in the last periods.

restart effect. Since a major factor in determining equilibrium contributions is the number of remaining periods in the game, a surprise announcement of playing additional periods increases equilibrium contributions. In addition, our model predicts a much smaller restart effect if the experimental design is such that subjects are reshuffled and play the game in a different group after each period, which is confirmed both by Andreoni [88] and Croson [96]. In a similar vein, it predicts no restart effect if the restart does not come as a surprise (if the subjects know all along that a second session will be played), which is confirmed by Burlando and Hey [97]. If the restart effect was a pure psychological phenomenon (a “new beginning”), with no strategic considerations involved, then presumably the restart effect should have the same magnitude in the above treatments. Related to this point, the model we present implies that in a longer game, contributions to the public good are more persistent (as opposed to the possibility that in a longer game, there are just more periods at the end where players contribute very little), another empirical result reported by several papers.⁹ Finally, the model implies the well-documented fact that a relative increase in the private returns of public contribution increases equilibrium contributions, and is consistent with the observations that neither an increase in the number of players nor an increase in the monetary stakes seems to decrease contribution levels.¹⁰

Our model also implies that as long as there is at least one selfish player in the game, and reciprocal players do not “overreciprocate” others’ contributions (an extra unit of contribution by other players is never reciprocated by more than a unit), then contributions are zero in the unique Nash equilibrium of a one-shot game.

We abstract away from learning and assume complete information. It is clear from existing data that learning plays an important role in the experiments. However, the features of positive initial contributions and subsequent decreasing pattern are shown to apply to games in which players are both experienced and may have learned about each other (for example, because the current game follows a surprise restart, i.e., the same group of players already played a round of repeated game), which suggests that a fully satisfactory theory of cooperation over finite horizon should also apply to complete information settings.

⁹See for example Isaac et al. [94].

¹⁰See Isaac and Walker [88] and Isaac et al. [90].

The experiments we conducted were partly designed to test whether the positive initial contributions are an equilibrium phenomenon, and partly to test more specific implications of our model. Furthermore, we directly addressed whether positive contributions can be an equilibrium phenomenon in one-shot games. We focused on games that best approximated the complete information assumption: games in which players already played the contribution game for several rounds beforehand, and also had a chance to learn about each other (for example, by playing a previous game together that was followed by a surprise restart). To be able to conduct multiple “surprise” restarts, we used a design in which after each round, it was randomly decided whether the group stayed together and played a “restarted” game (25% probability) or whether players in the group were randomly reshuffled and played the next round of game with a new group (75% probability).¹¹

Our experimental results confirmed that in ten period contribution games, there is a significant restart effect even when players are experienced, and that the contribution pattern in these restarted games is decreasing. In sessions where individual histories were revealed, in the last two rounds of the experimental sessions, the mean contribution in the first round of a restarted game was 10.69, and only 29% of subjects contributed nothing. Meanwhile, in the 10th period of the preceding games, the mean contribution was 1.56, and 88% of subjects contributed zero. The patterns were similar in sessions in which only average contributions were revealed to group members. In all sessions, the restart effect was significant both with respect to mean contributions and to the proportion of players contributing a positive amount. Furthermore, period-10 mean contributions in the same restarted games were only 0.6 – 1.3, confirming that the decreasing pattern of contributions holds for these games. In sharp contrast to these results, in the session involving one-shot games, play in later periods involved 94% of subjects contributing zero in restarted games. There was no significant restart effect in later rounds, and contributions in restarted games were not significantly different from zero.

In the experiments, we obtained two pieces of evidence that the observed play in restarted games in the last two rounds of experimental sessions approximated equilibrium play. In some sessions, we solicited players’ expectations

¹¹We set the probability that a group stays together to play another round of game relatively low, to mitigate repeated game considerations.

on others' subsequent contributions before the start of a new repeated game. Experienced players' expectations on average closely tracked the average of actual play, with the median expectation tightly near the median of actual play. In particular, even before the start of the game, the overwhelming majority of players correctly anticipated that at the end of the game, contributions would be close to zero; despite this, most players started out by contributing significant positive amounts. The other piece of evidence came from testing whether the average contribution path in restarted games stabilized by the last two rounds. We found that contribution levels at the beginning and at the end of 10 times repeated games were the same in the last two rounds of the experiments as in the preceding two rounds, while contributions in the middle of the game (period 5) were marginally significantly lower in the last two rounds. This suggests that the pattern of contributions stabilized by the end of the experiment, although we did not get conclusive evidence that the rate at which contributions go to zero reached a steady state.

We also demonstrated the presence of conditional reciprocity even among experienced players by regressing players' second-period contributions on others' first-period contributions. In restarted games of the last two rounds, a unit increase in average contributions of others increased a player's second-period contribution by 0.4 units, significantly different from 0. This is also compatible with the implication of the model that in equilibrium, completely self-interested players start out contributing.

Finally, we conducted sessions in which the contribution games were preceded by a sequence of gift-giving games. We used these gift-giving games to identify selfish and reciprocal players. Looking at the behavior of these players in contribution games, we found that games with more reciprocal players average contributions were higher. Furthermore, selfish players started contributing 0 significantly earlier than reciprocal players, which is one of the implications of our model.

2 Related literature

The dynamic interaction between players who only care about their own material payoff and players who are conditionally reciprocal is a relatively unexplored area. Fehr and Schmidt [99], in Section IV of their paper, consider a two-stage

game played by selfish and fairness-motivated players. The first stage is a public good contribution game. In the second stage, after observing the outcome of the first stage, players can engage in costly “retaliation.” They show that for some parameter values, there are equilibria in which cooperation can be sustained in the public good contribution game. Levine [98] assumes that players’ utilities depend both their own and on others’ payoffs, and the weight attributed to another player’s payoff depends on how much that player cares about others. The model is then calibrated using data from ultimatum games experiments, and its predictions are confronted with data from other experiments, including a one-shot public good contribution game. Andreoni and Samuelson [06] analyze a twice-played prisoner’s dilemma in which players some time prefer to cooperate, and players are heterogeneous in their tastes for cooperation. Duwfenberg and Kirchsteiger [04] extend reciprocity equilibrium (Rabin [93]) to extensive form games. The examples they provide suggest that typically there is a multiplicity of such equilibria. Anderson et al. [98] provide a model with both altruism and decision error in public good provision games. Their model does not allow for strategic interaction across periods. Offerman et al. [96] present a model of a one-shot step-level (the public good is provided if there is enough aggregate contribution) public contribution game with heterogeneous individuals. Although the model focuses on the interaction of selfish players and cooperators, most of the issues we investigate in this paper are not present since the game has only a single period.

There are various explanations of cooperation over finite horizon in which the focus is not on the dynamic interaction of different types of players. Radner ([80] and [86]) shows that cooperation can be maintained for a while in a repeated oligopoly game and in a repeated prisoner’s dilemma if players only care about maximizing their payoff up to epsilon precision, even for small values of epsilon. A somewhat similar argument is presented by Klumpp [04] for repeated public good contribution games. He shows that even a relatively small amount of altruism can generate relatively large contributions at the beginning of the game. Furthermore, contributions are decreasing over time, because the amount of sustainable cooperation decreases as the number of periods left to play decreases. The scope of these explanations are limited by the fact that in games with discrete action spaces, small departures from maximizing individual payoffs cannot explain any amount of cooperation, unless the number of periods is very large, while large deviations do not explain the breakdown of cooperation

in the end. Furthermore, these theories would imply that relative contributions go to zero as monetary stakes increase, which does not hold empirically. Neyman [85] shows that cooperation can be achieved in equilibrium of a finitely repeated prisoner’s game if players can only use strategies with bounded complexity. Again, the argument is more appealing for games in which the number of repetitions is large. Finally, there are various explanations which relax the assumption that the fundamentals of the game are common knowledge among players. Kreps et al. [82], Sobel [85], and Fudenberg and Maskin [86] show how a small amount of uncertainty about payoffs (reputation) can induce cooperation in games with finite horizon, while Neyman [99] points out that a small departure from the assumption that the length of the game is commonly known can lead to cooperative outcomes. These arguments, however, face a difficulty in explaining why cooperation is achieved even by very experienced players, who already played the game multiple times with the same set of opponents.

3 The model

3.1 General specification

We consider a T -period public good contribution game with $N \geq 2$ players. We also use N to denote the set of players whenever it does not cause confusion. In each period, each player has an endowment of 1 unit. Players in each period simultaneously decide how much of their endowment to contribute for public investment and how much to retain for private investment. Let $x_i^t \in [0, 1]$ denote player i ’s contribution to the public investment in period t .

We will consider two information environments, corresponding to the two most popular rules used in experimental settings. In the first one, after every period, each player observes the contribution choices of every other player. In the second version, after every period, only the total contribution to the public investment is revealed to the players.

The material payoff of player i in period t is the amount of endowment she retains for herself plus her share from the aggregate returns to the public

investment:¹²

$$(1 - x_i^t) + \frac{A}{N} \sum_{j \in N} x_j^t, \quad \forall i \in R.$$

Public investment yields a constant marginal return A , which is divided equally to all players. We assume $A > 1$ but $\frac{A}{N} < 1$.

Players $i = 1, \dots, S$ are rational in the traditional sense of maximizing the sum of per period material payoffs. From now on, we refer to them as selfish players. Let S also denote the set of selfish players and R denote the rest of the players.

Players $R = \{S + 1, \dots, N\}$ are reciprocal. Their payoffs are determined through their reciprocity functions. It is convenient to think about these functions as specifying target contribution levels. The arguments of f_i^t , the period- t reciprocity function of player $i \in \{S + 1, \dots, N\}$, are past and current contributions to the public good by others. We assume that every reciprocity function is nondecreasing in all other players' contributions and takes values in $[0, 1]$.

To keep the analysis tractable, we only consider reciprocity functions which are additively separable with respect to contributions made at different periods, and which, within the same period, are additively separable with respect to contributions made by different players:

$$f_i^t((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^t)_{j \in N/\{i\}}) = \sum_{k=1}^t \sum_{j \in N/\{i\}} f_{i,j}^{t,k}(x_j^k)$$

where $f_{i,j}^{t,k}(\cdot)$ is nondecreasing for $i \neq j$ and $f_{i,j}^{t,k}$ is defined only for $k \leq t$.¹³ Furthermore, throughout the paper, we assume that $f_{i,j}^{t,k}$ is concave and differentiable for every $i, j \in N$ and $t, k \in \{1, \dots, T\}$.

The t -period payoff of player i is $g(x_i^t - f_i^t()) = -(x_i^t - f_i^t())^2$.¹⁴ Reciprocal

¹²All our results can be extended to the case when the returns from private investment are $r(1 - x_i^t)$, where r is any concave function.

¹³Allowing for an additional constant term c_i^t , which only depends on the number of the period and not others' contributions, would not change any of our main results. We omit this term partly to simplify the exposition and partly because there is a lot of evidence that players do not behave unconditionally altruistically in public good experiments.

¹⁴Instead of the quadratic loss function specified above, we could consider any strictly quasiconcave function $g()$ which attains its maximum at 0.

players maximize the sum of these per period payoffs.¹⁵ Note that we allow for a great deal of heterogeneity in the preferences of reciprocal players, since they can have different reciprocity functions.

The game is a standard extensive form game: in particular, reciprocal players are standard utility maximizers. They differ from the selfish players in that their payoffs do not just depend on their own material payoffs. We assume that the game is of complete information. Players know how many selfish and how many reciprocal players there are in the group, and they know the reciprocity functions.¹⁶ This corresponds to our intention to analyze play in games in which the players are experienced and become familiar with each other. On the other hand, the game is not of perfect information, because in each period, players decide on their current contributions simultaneously. Furthermore, if only aggregate contributions are revealed to players after each round, then the game is not a multi-stage game with observed actions; hence, after each history, players have to form beliefs on what happened so far in the game.

3.2 Interpretation of reciprocity functions

Reciprocity functions provide a simple and tractable way to model players whose preferences are for one reason or another influenced by how much other players contribute. The underlying motivation behind reciprocal preferences can come from many sources. One possibility is that reciprocal players are conditionally altruistic: the payoffs of those who contribute to the public good enter their utility functions. Another source can be a desire to follow social norms: a player with such considerations would contribute more if she thinks others contribute more (or if she observes that others contributed a lot in past rounds). Yet another possibility is that reciprocal players care about fairness or equality, as in Fehr and Schmidt [99] or Bolton and Ockenfels [00]. A player who is concerned

¹⁵In the above formulation, the extent to which reciprocal players care about their own monetary payoffs is embedded in their reciprocity functions. Alternatively, we could assume that a reciprocal player's utility in each round is given by the sum of her monetary payoff and a reciprocity term, which is a function of the difference between her contribution and the value specified by her reciprocity function. In this formulation, reciprocal players might try to induce reciprocal contributions from other reciprocal players too, just like selfish players (they might contribute more than what is specified by their reciprocity functions). Despite this, it can be shown that all our results remain valid in this alternative formulation, provided that a reciprocal player's disutility from contributing less than the value of the reciprocity function is large enough.

¹⁶More precisely, all this information is common knowledge among players.

about fairness reciprocates others' contributions in a public good contribution game because a contribution increases others' flow payoff at the expense of one's own payoff.

Our model specification is also consistent with the consideration that reciprocal players care not only about how much others contribute, but also about the types, or preferences, of the other players (recall that we assume that players know each others' types). In particular, we allow reciprocity towards other players to be asymmetric. For example, reciprocity functions are allowed in our model to be more responsive to other reciprocal players' contributions than to contributions made by selfish players. Therefore, as long as intentions of a player depend only on his or her preferences, our model allows for reciprocity to depend on the intentions behind contributions.

4 Example: reciprocal players reciprocate the simple time average of others' contributions

In this section, we consider a concrete specification of the public good contribution game and explicitly compute contributions in subgame perfect Nash equilibrium. The example is useful for understanding the main features of dynamic strategic interaction between selfish and reciprocal players, and for providing intuition for the more general results in the next section.

In this specification, individual contributions are revealed after each round, and reciprocity functions are such that every reciprocal player targets to contribute the simple time average of other player's per-period contribution up until and including the current period:

$$f_i^t(\cdot) = \frac{1}{t(N-1)} \left(\sum_{k=1, \dots, t} \sum_{j \in N/\{i\}} x_j^k \right) \quad \forall \quad i \in R \quad \text{and} \quad t \in \{1, \dots, T\}.$$

We also assume that there is at least one selfish player ($S \geq 1$).

Claim 1 below establishes that in the above specification, for almost every value of A , there is a unique subgame perfect Nash equilibrium. Furthermore, along the equilibrium path, the contribution of every player to the public good is

weakly decreasing over time. In particular, selfish players start out contributing their full endowment to the public investment, but at a certain period, they switch to contributing zero. Reciprocal players contribute all their endowment up until this point, too, and then gradually decrease their contributions.

For the formal statement, it is convenient to define the following term:

$$C(t) = \sum_{k=t+1}^T c(t, k),$$

where

$$c(t, t+1) = \frac{N-S}{S+t(N-1)}$$

$$c(t, k) = \frac{N-S}{S+(k-1)(N-1)} \left[1 + \sum_{l=t+1}^{k-1} \left(\frac{N-S-1}{N-S} \right) c(t, l) \right] \text{ for } k > t+1.$$

$C(t)$ is the marginal cumulative impact of a contribution at t on the future contribution of reciprocal players. The term includes both the direct impact of a unit contribution for the future periods as well as the indirect effect of a unit contribution in subsequent periods on future periods.

Claim 1: Suppose $C(t) \neq 1 - \frac{A}{N} \forall t \in \{1, \dots, T\}$ and $S \geq 1$. Then the above game has a unique subgame perfect Nash equilibrium, which exhibits the following contribution path. If $C(1) < 1 - \frac{A}{N}$, then every player contributes 0 at every period. Otherwise, let T^* be the the largest integer between 1 and $T-1$ such that $C(t) > 1 - \frac{A}{N}$. Then every selfish player contributes 1 in periods 1 to T^* and 0 afterwards. Meanwhile, reciprocal players contribute 1 until period T^* , and their contributions are strictly decreasing afterwards.

It is easy to see that the property $C(t) \neq 1 - \frac{A}{N} \forall t \in \{1, \dots, T\}$ in the claim holds generically for A , for every N and S .

For the formal proof, as well as for proofs of all subsequent theorems, see the Appendix. Here we provide a sketch of the argument. First note that a contribution at any period does not influence how much other players contribute in the same period, since contributions are decided simultaneously. This implies that at the last period, all selfish players have to contribute 0 in any subgame

perfect Nash equilibrium. Furthermore, it can be shown that in subgame perfect Nash equilibrium, reciprocal players have correct expectations concerning the contributions of others at T , and that there is only one continuation profile after any $(T - 1)$ -long history such that reciprocal players have correct expectations and contribute the amount specified by their reciprocity functions. This implies that after every $(T - 1)$ -long history, the continuation profile is uniquely pinned down in subgame perfect Nash equilibrium. Consider now any $(T - 2)$ -long history. At period $T - 1$, selfish players' contributions to the public good depend only on the extent to which they influence reciprocal players' contributions at T . Because of the linearity of reciprocity functions, the marginal impact of period- $(T - 1)$ contributions on period- T contributions of reciprocal players is constant and independent both of the history and of how much other players contribute at $T - 1$.¹⁷ Then for generic values of A , either all selfish players contribute 1 at $T - 1$ to the public good (if the marginal impact on contributions at T is larger than $1 - \frac{A}{N}$, the difference between marginal returns of the private and public investments), or all of them contribute 0 (if the marginal impact is smaller than $1 - \frac{A}{N}$). Then it can be shown that after every $(T - 2)$ -long history, the continuation profile is uniquely pinned down in subgame perfect Nash equilibrium.

An iterative argument then establishes that (generically) selfish players' contributions at each period are independent of history, and they are always either 0 or 1, depending on their impact on future contributions of reciprocal players. The latter impact is larger in earlier periods, for two reasons: first, a unit contribution at earlier periods has a larger impact on the average contribution, and second, the contribution has an impact on more future periods. This leads to selfish players contributing 1 at the beginning and 0 later.

5 General theoretical results

In this section, we show that the main insights from the previous example (selfish players contribute in order to induce future contributions by reciprocal players, decreasing pattern of contributions in equilibrium) extend to much more general versions of reciprocity.

¹⁷Note though that it is not $1/((N - 1)T)$, the direct effect of the contribution, unless there is only one reciprocal player. If there are more than one reciprocal player, then the direct effect is multiplied by a factor depending on T and $N - S$ due to the fact that reciprocal players reciprocate each others' period- T contributions.

5.1 General assumptions on reciprocity

First, we impose that reciprocity functions are linear with respect to other reciprocal players' contributions. This assumption guarantees that the impact of a marginal contribution by a selfish player on subsequent contributions by reciprocal players does not depend on the history of contributions. Then the amount of contribution by every selfish player in each period can be determined solely on the basis of the future impact of the contribution on contributions by reciprocal players. This both greatly simplifies the analysis and helps avoid multiplicity of equilibria.

A1: *Linear Reciprocity towards Reciprocal Players*

$$f_{i,j}^{t,k}(x_j^k) = \alpha_{i,j}^{t,k} x_j^k \quad \forall i \in R, j \in R/\{i\} \text{ and } t, k \in \{1, \dots, T\}, \text{ where } k \leq t.$$

The second assumption requires that a reciprocal player treat selfish players symmetrically: contributions by one selfish player are reciprocated exactly the same way as by another one. Note that we do not impose symmetric reciprocity towards other reciprocal players. The motivation is that reciprocity towards another player should only depend on the latter player's preferences (and all selfish players have the same preferences, but reciprocal players can have different reciprocity functions, hence different preferences).

A2: *Symmetric Reciprocity toward Selfish Players*

$$f_{i,j}^{t,k}(\cdot) = f_{i,j'}^{t,k}(\cdot) \quad \forall i \in R, j, j' \in S \text{ and } t, k \in \{1, \dots, T\}, \text{ where } k \leq t.$$

The next two assumptions require that the extent of reciprocity not grow over time. Assumption 3 is a condition on the level of reciprocity: it states that a reciprocal player does not reciprocate in a strictly increasing manner a nonincreasing sequence of contributions by any other player. This assumption in particular implies that reciprocity toward a given contribution decreases over time.

A3: *Nonincreasing Total Impact of Contributions over Time*

Suppose $i \in R$, $j \in N/\{i\}$, $t \in \{1, \dots, T-1\}$, and x_j^1, \dots, x_j^{t+1} is such that $x_j^k \geq x_j^{k+1} \quad \forall k \in \{1, \dots, t\}$. Then $\sum_{k=1}^{t+1} f_{i,j}^{t+1,k}(x_j^k) \leq \sum_{k=1}^t f_{i,j}^{t,k}(x_j^k)$.

Assumption 4 is a statement on marginal reciprocity: it states the marginal impact of a contribution on reciprocity k periods later (a contribution at t on

reciprocity at $t + k$, a contribution at $t + 1$ on reciprocity at $t + k + 1, \dots$) does not increase over time.

A4: Nonincreasing Marginal Impact of Contributions over Time

For any $x \in [0, 1]$, $i \in R$, $j \in N/\{i\}$, $t < t'$, $k \geq 0$, and $t' + k \leq T$, $\frac{\partial f_{i,j}^{t+k,t}(x)}{\partial x} \geq \frac{\partial f_{i,j}^{t'+k,t'}(x)}{\partial x}$.

We note that although the above assumptions on the time structure of reciprocity are sufficient to establish our main results, the reciprocity functions that are most appealing to us are ones in which reciprocity is in some sense constant over time. One way to formalize constant reciprocity over time is strengthening A3 by requiring that $\sum_{k=1}^{t+1} f_{i,j}^{t+1,k}(x_j^k) = \sum_{k=1}^t f_{i,j}^{t,k}(x_j^k)$ whenever $x_j^k = x_j^{k+1} \forall k \in \{1, \dots, t\}$ and $j \in N/\{i\}$. A simple example of a reciprocity function satisfying this requirement is when contributions in the current and the preceding period get reciprocated, at a constant rate:

$$f_i^1(\cdot) = (\alpha_0 + \alpha_1) \sum_{j \in N/\{i\}} x_j^1, \quad \text{and}$$

$$f_i^t(\cdot) = \alpha_0 \sum_{j \in N/\{i\}} x_j^t + \alpha_1 \sum_{j \in N/\{i\}} x_j^{t-1} \text{ if } t \geq 2$$

The next assumption imposes that reciprocity is initially strictly positive toward every other player. If reciprocal players are nonresponsive to selfish players' contributions, then there is no interesting dynamic strategic interaction in our model.

A5: Positive Initial Reciprocity

For every $i \in R$, $j \in N/\{i\}$, and $t \in \{1, \dots, T\}$, there exists $t' \in \{1, \dots, t\}$ such that $\left. \frac{\partial f_{i,j}^{t,t'}(x)}{\partial x} \right|_{x=0} > 0$.

We also assume that the reciprocal players do not “overreciprocate” others' contributions in the sense that at any period t , a unit increase in contributions by other players up until t increases the value of a reciprocity function by not more than a unit. This, besides being a natural requirement, is imposed in order to avoid multiplicity of equilibria resulting from reciprocal players either

having optimistic expectations with respect to each others' contributions and contributing more, or having pessimistic expectations and contributing less.

A6: No Overreciprocation

$$\sum_{k=1}^t \sum_{j \in N/\{i\}} \frac{\partial f_{i,j}^{t,k}(x_j^k)}{\partial x_j^k} \leq 1 \quad \forall t \in \{1, \dots, T\}, i \in R, \text{ and } (x_j^k)_{j \in N/\{i\}}^{k=1, \dots, t} \in [0, 1]^{(N-1)t}.$$

5.2 Decreasing pattern of contributions in equilibrium

The arguments used in analyzing the concrete model specification of Section 4 generalize to all specifications that satisfy the assumptions made in Section 5, for games in which individual contributions are revealed after each round. In particular, since reciprocity functions are concave, no overreciprocation and positive initial reciprocity imply that the impact of a contribution on future contributions by reciprocal players is uniquely defined. Linearity in other reciprocal players' contributions implies that this impact is independent of contributions made in other periods or by other players. Then for generic values of A (if reciprocity functions are strictly concave in selfish players' contributions, then for all values of A), selfish players' contributions are uniquely determined in subgame perfect equilibrium at every period. Nonincreasing marginal impact of contributions over time then implies that the marginal return of contributions at earlier periods, when more periods are left to be played, is higher. This establishes that selfish players' contributions are weakly decreasing over time. Finally, nonincreasing total impact of contributions over time implies that the reciprocal players' contributions are weakly decreasing over time, too. The difference between this more general specification and the linear setting of Section 4 is that selfish players' per period contributions are not restricted to be either 1 or 0. Depending on the reciprocity functions, the contributions of selfish players can follow any weakly decreasing pattern.¹⁸

Theorem 1: If $S \geq 1$, then for generic A , the game in which individual contributions are revealed after every period has a unique subgame perfect Nash equilibrium, and this equilibrium exhibits a weakly decreasing pattern of con-

¹⁸For example, if reciprocity functions towards selfish players are such that marginal reciprocities towards initial contributions are high enough, but they decrease fast enough for additional amounts of contributions, then selfish players' contributions in equilibrium take intermediate values (strictly between zero and the whole endowment) in all rounds but the last.

tributions. If reciprocity functions are strictly concave in selfish players' contributions, then the above statement holds for all A .

As opposed to the game in which individual contributions are revealed after each period, the game in which only the average contribution of the group is revealed typically has many different perfect Bayesian Nash equilibria (PBNE). The reason is that in this information environment, reciprocal players' out-of-equilibrium beliefs are not uniquely determined in PBNE. After a deviation, players are only informed that the total contributions to the public good differ from what the equilibrium specifies. Then for different beliefs by the reciprocal players concerning who deviated, their continuation strategies after the deviation are different. This indeterminacy of out-of-equilibrium beliefs can result in many different equilibria: in particular, ones in which selfish players' contributions are asymmetric, and ones in which contributions are nonmonotonic over time. However, we show that there is always a PBNE which yields the same outcome as the unique subgame perfect Nash equilibrium of the game in which individual contributions are revealed. Furthermore, this contribution path is implied by every pure strategy PBNE where selfish players contribute the same after any history (the equilibrium is strongly symmetric with respect to selfish players), and the following feature of beliefs is common certainty among players: after any observed deviation, all reciprocal players believe with probability one that it was one of the selfish players who deviated whenever this belief is possible. We consider the above equilibria focal, but we do not argue that they are the only plausible equilibria in the game in which only aggregate contributions are revealed after each period.

Let $\bar{x}^k = \sum_{i \in N} x_i^k$. For any t -period long history $h^t = (x_i^1)_{i \in N}, \dots, (x_i^t)_{i \in N}$, let $\omega_i(h^t)$ denote the beliefs that player i has over other players' action choices so far in the game (note that beliefs can only depend on the observed part of the history: $\omega_i((x_i^1)_{i \in N}, \dots, (x_i^t)_{i \in N}) = \omega_i((y_i^1)_{i \in N}, \dots, (y_i^t)_{i \in N})$ if $(\bar{x}^k, x_i^k)_{k=1, \dots, t} = (\bar{y}^k, y_i^k)_{k=1, \dots, t}$). Let $s^{t+1}(h^t)$ denote the period- $(t+1)$ action profile specified by s after h^t . If s is strongly symmetric with respect to selfish players, then let $s_S^{t+1}(h^t)$ denote the period- $(t+1)$ action specified by s for a selfish player after h^t .

Definition: For a strategy profile s that is strongly symmetric with respect to selfish players, *reciprocal players think that a selfish player is responsible for*

any deviation if for any $i \in R$ and any history $h^t = (x_j^1)_{j \in N}, \dots, (x_j^t)_{j \in N}$ which satisfies that $x_j^{l+1} = s_j^{l+1}(h^t) \forall j \in R$ and that

$$\bar{x}^{l+1} \in \left[\sum_{j \in N} s_j^{l+1}(h^l) - s_S^{l+1}(h^l), \sum_{j \in N} s_j^{l+1}(h^l) + 1 - s_S^{l+1}(h^l) \right],$$

$\omega_i(h^t)$ allocates positive probability only to histories $(y_j^k)_{j \in N}^{k=1, \dots, t}$ which satisfy that for every $l \in \{1, \dots, t-1\} \exists i' \in S$ such that $y_{i'}^{l+1} = s_{i'}^{l+1}((x_j^1)_{j \in N}, \dots, (x_j^l)_{j \in N}) + \bar{x}^{l+1} - \sum_{j \in N} s_j^{l+1}((x_{j'}^1)_{j' \in N}, \dots, (x_{j'}^l)_{j' \in N})$ and $y_j^{l+1} = (s_j^{l+1}(x_{j'}^1)_{j' \in N}, \dots, (x_{j'}^l)_{j' \in N}) \forall j \in N/\{i'\}$.

Theorem 2: Assume $S \geq 1$ and that A is such that in the game in which individual contributions are revealed after each period, there exists a unique subgame perfect Nash equilibrium. Then in the game in which only aggregate contributions are revealed in each period, there exists a perfect Bayesian Nash equilibrium which yields the same outcome. Furthermore, this outcome is implied by every pure strategy perfect Bayesian Nash equilibrium which is (i) strongly symmetric with respect to selfish players, and in which (ii) it is commonly believed that reciprocal players think that a selfish player is responsible for any deviation.

6 Properties of the equilibrium contribution path

Our model is consistent with a series of comparative statics results in the existing literature on public good contribution experiments, and it can explain the restart effect. We discuss these and explore some additional implications of our theory.¹⁹ In what follows, we assume that individual contributions of the players are revealed after every period, and focus on the (generic) parameter values for which the game has a unique subgame perfect Nash equilibrium. The same results apply to PBNE with the properties stated in Theorem 2 for games in which only aggregate contributions are revealed.

6.1 Comparative statics

Private Return from Contributing

¹⁹See the next section for testing these implications in a laboratory setting.

Increasing the return of the public investment (A) brings the individual return from contributing to the public good ($\frac{A}{N}$) closer to the return from private investment (1). It is well-documented in experimental settings that this increases players' contributions to the public good (see for example Isaac and Walker [88], and Isaac et al. [94]). The next theorem shows that this is in accordance with our model: an increase in A (weakly) increases contributions by all players in all periods.

Theorem 3: Assume $S \geq 1$. Let A and \hat{A} be such that that the games in which the return of private investment are A and \hat{A} have unique subgame perfect Nash equilibria s and \hat{s} . Then $A < \hat{A}$ implies $s_i^t \leq \hat{s}_i^t \forall t \in \{1, \dots, T\}$ and $i \in N$.

Number of Periods

Increasing the number of periods in public good contribution games is shown to result in a longer period of positive contributions and in higher aggregate contribution levels (see Isaac et al. [94]). This effect is implied by our model, too. In a longer game, selfish players have higher incentives to contribute, because there are more future periods in which reciprocal contributions are affected. Then in equilibrium, all players end up contributing more.

Theorem 4: Assume $S \geq 1$. Let A be such that the games with T and \hat{T} number of periods have unique subgame perfect Nash equilibria s and \hat{s} . Then $T < \hat{T}$, $f_i^t = \hat{f}_i^t$ implies $s_i^t \leq \hat{s}_i^t \forall t \in \{1, \dots, T\}$ such that $t \leq \min(T, \hat{T})$ and $i \in N$.

Endowment Level

The effect of an increase in the endowment levels depends on whether the domain of reciprocity functions is relative or absolute levels of contributions.²⁰ In the former case, changing the endowment level does not change relative contributions (changes absolute amounts of contributions proportionally to the endowment). If reciprocity functions are defined on absolute amounts of contributions, then an increase in endowments increases contributions, but in general not proportionally (only linear reciprocity functions imply that). In any case, our model is consistent with the findings of Fehr and Tougareva [95], Hoffman et al. [96], Slonim and Roth [97,] and others (although without further assumptions does not imply it) that even if monetary stakes are high, voluntary contributions in experiments are high.

²⁰Note that we normalized per period endowment to 1; hence we could put this issue aside beforehand.

Number of Players

Changing the number of players leads to ambiguous comparative statics in our model. We give the intuition for this using the symmetric linear specification of Section 4.

Consider first an increase in N such that both $\frac{S}{N-S}$, the ratio of selfish and reciprocal players, and $\frac{A}{N}$, the private benefit from public contribution, are held fixed. It is easy to show that the marginal effect of a unit contribution at t by a selfish player on total contributions at $t+1$, $\frac{N-S}{(N-1)t+S}$, decreases in N if $\frac{S}{N-S}$ is held fixed. However, the effect of increased contributions by $(N-S)$ reciprocal players at $t+1$ on contributions at $t+2$, $\frac{(N-S-1)}{(N-1)t+S} \frac{1}{(N-1)(t+1)+S} (N-S)$, increases in N if $\frac{S}{N-S}$ is held fixed. The intuition behind these relationships is that for a fixed $\frac{S}{N-S}$, a larger N implies that from a selfish player's point of view, the ratio of reciprocal players among others is higher, while from a reciprocal player's point of view, the ratio of reciprocal players among others is lower.²¹ Overall, an increase in N , depending on the parameters of the model, can lead to either higher or lower equilibrium contributions.

This ambiguity corresponds to the mixed empirical findings concerning group size. Isaac and Walker [88] and Isaac et al. [94] report that increasing group size increases contributions to the public good. On the other hand, Bagnoli and McKee [91] find that increasing group size has a negative effect on contributions, especially in early periods. The latter finding fits our results particularly well. In any case, the model we propose is consistent with the fact that even in large groups, contributions to the public good can be large.

If N is increased such that A (and $\frac{S}{N-S}$) is held fixed, again comparative statics are ambiguous, but it becomes more likely that aggregate contributions decrease, because of the decrease in $\frac{A}{N}$. This is consistent with the findings of Isaac and Walker [88] and Isaac et al. [94]. Another prediction that our model gives is that the variance of contributions decreases with group size. It is because for any ratio of selfish and reciprocal players in the population, a larger group size implies that the ratio of players in the given game is more likely to be close to the population average. As far as we know, this aspect of contributions has not been tested yet.

²¹A concrete example is that if half of the players are reciprocal and $N = 4$, then from a selfish player's point of view, $2/3$ of the other players are reciprocal, while if $N = 10$, then only $5/9$ of them are. From a reciprocal player's point of view, $1/3$ of the other players are reciprocal if $N = 4$, and $4/9$ of them are if $N = 10$.

6.2 Relative contributions by selfish and reciprocal players

In this subsection, we investigate how equilibrium contributions depend on group composition.²² The equilibrium contribution pattern has the same qualitative feature (weakly decreasing) for any group composition that involves at least one selfish player, but the ratio of selfish to reciprocal players influences the level and time pattern of contributions. If a group consists of only selfish players, then the unique subgame perfect Nash equilibrium of the game involves zero contributions in every period (unique perfect Bayesian Nash equilibrium if only aggregate contributions are revealed). More generally, for any A and T , there is a critical ratio such that if the fraction of selfish players exceeds this ratio, then the unique subgame perfect Nash equilibrium involves zero contribution. An increase in the number of reciprocal players increases selfish players' incentives to contribute, leading to a larger number of periods with positive contributions and to a higher amount of contribution by every selfish player. However, reciprocal players' equilibrium contributions are not necessarily monotonic in the ratio of selfish to rational players, even if the distribution of reciprocal types is held constant.²³

Theorem 5: Consider two public good contribution games \mathcal{G} and $\widehat{\mathcal{G}}$ in which individual contributions are revealed after each round, and in which $N = \widehat{N}$, $T = \widehat{T}$, $S > \widehat{S}$, $f_i^t(\cdot) = \widehat{f}_i^t(\cdot) \forall i \in \{S+1, \dots, N\}$, and $t \in \{1, \dots, T\}$. Assume \mathcal{G} and $\widehat{\mathcal{G}}$ have unique subgame perfect Nash equilibria s and \widehat{s} . Then $s_i^t \leq \widehat{s}_i^t \forall i \in \{1, \dots, \widehat{S}\}$ and $t \in \{1, \dots, T\}$.

Our model does not give a clear prediction for the ratio of total contributions by a selfish and a reciprocal player. In the specification of Section 4, selfish players contribute strictly less in equilibrium than reciprocal players, but for different reciprocity specifications, selfish players might end up contributing more. The intuition behind this is that although selfish players only contribute in equilibrium to induce future contributions by reciprocal players, the condition for a unit contribution to be optimal is not that it induces more than a unit of

²²We state the formal theorems for subgame perfect Nash equilibria of games in which individual contributions are revealed, but again the results apply to all perfect Bayesian equilibria that satisfy the conditions of Theorem 2 for games in which only aggregate contributions are revealed.

²³In the specification of Section 4, a decrease in the number of selfish players always increases reciprocal players' contributions, too.

total contributions by each reciprocal player, but that in total, it induces at least as much contribution in the future as $1 - \frac{A}{N}$, the difference between the returns of the private investment and the public contribution. However, there is a general implication of the model with respect to the timing of contributions of selfish and reciprocal players: every reciprocal player contributes positive amounts for at least as many periods as any selfish player. The reason is that selfish players have purely forward-looking considerations in contributing, while reciprocal players are partly backward-looking in deciding how much to contribute; hence selfish players' contributions tend to be relatively more concentrated on earlier periods than reciprocal players' contributions.

Theorem 6: Consider the game in which individual contributions are revealed after each period, and assume there is a unique subgame perfect Nash equilibrium s . Then $s_i^t > 0$ for some $i \in S$ implies $s_j^k > 0 \forall j \in R$ and $k \in \{1, \dots, t\}$.

6.3 The restart effect

Consider a surprise restart announcement, as in Andreoni [88], that the same group of players are to play another session of repeated contribution games. If players treat the restarted game as a new game, then our model immediately explains the restart effect. If the conditions of Theorem 1 hold, then there is a unique subgame perfect Nash equilibrium, so the contribution pattern in the restarted game is expected to be the same as in the first game. This equilibrium implies a decreasing pattern of contributions; thus, in the first period of the restarted game, contributions are higher than those in the last period of the preceding game. Even if players do not treat the new session as a new game, but as an extension of the first game (and therefore aggregate contributions in the previous game become a history in a longer game), the model is compatible with the restart effect. This is because selfish players' contributions in equilibrium depend on the number of periods left to be played. If it is unexpectedly revealed that more periods are to be played than previously thought, selfish players have increased incentives to contribute. This unambiguously increases the contributions of selfish players, which has a positive impact on reciprocal players' contributions as well. Whether reciprocal players contribute more in the first period of the restarted game than in the last period of the game before depends on the concrete specification of reciprocity functions (how much reciprocal players care about the "distant past").

6.4 One-shot games

In a one-shot game, conditions A1-A6 guarantee that if there is at least one selfish player, then there is a unique Nash equilibrium of the public good contribution game, in which all players contribute zero.²⁴ The result applies to restarted one-shot games as well. The intuition behind the result is simple. Since there is no continuation, all selfish players contribute zero. Then A5 (positive initial reciprocity toward any player) and A6 (no overreciprocation) imply that the only fixed point of the expectations of reciprocal players is when they expect zero contributions from each other.

Theorem 7: If $T = 1$ and $S \geq 1$, then the game has a unique Nash equilibrium, which involves all players contributing 0 to the public good.

This result corresponds to the empirical finding that although initially subjects contribute significantly positive amounts in one-shot games in an experimental setting, contributions seem to go to zero with learning. In a setting in which after each round of play, subjects get randomly assigned to a new group (called the “strangers” treatment in the literature), contributions to the public good diminish over time, according to both Andreoni [88] and Croson [96].²⁵ This suggests that in these games, contribution is not an equilibrium phenomenon. In Section 7, we provide further experimental evidence that contributions diminish in one-shot games as players get more experienced.

7 Experimental design

7.1 Hypotheses

The purpose of the laboratory experiments is to test some features of our model in a voluntary public good contribution context. Some of these implications

²⁴This result holds both for the game in which individual contributions are revealed and for the game in which only aggregate contributions are revealed. Note that in a one-shot game, the above change in the information environment is immaterial for strategic considerations, since the game is over by the time contributions are revealed.

²⁵The results in Andreoni [88] and Croson [96] differ in whether initially subjects contribute more in a strangers treatment or in a partners treatment (the latter corresponds to the same set of players staying together and playing the game multiple times). Since we argue that contributions in the strangers treatment are due to inexperience and disappear with learning, the above debate is irrelevant for the points that we make.

apply not only to our model, but to any theory that predicts that play approximates an equilibrium as players get more experienced. Other features are more intimately connected to the ideas we have presented. Our model assumes no uncertainty; therefore we focus on players who are experienced with the game (already played multiple rounds of the same game) and also have some information about how others in the group play. The latter can be either because they already played the game with the same set of opponents (it is a restarted game, where the restart was surprise) or because they received some information from the experimenter about their opponents' past play.

The main hypotheses we wish to examine are:

- H1: With experienced players in restarted games, there is a declining pattern of contributions in 10 times repeated games. Furthermore, with experienced players, the restart effect is significant.
- H2: With experienced players, the restart effect is not significant, and contributions are not statistically different from zero in one-shot games.
- H3: The average contribution path in restarted 10 times repeated games stabilizes as players get more experienced.
- H4: Players' expectations approach actual play in restarted 10 times repeated games as they become more experienced.
- H5: With experienced players, selfish players start contributing zero earlier than reciprocal players.
- H6: Contributions positively affect other players' subsequent contributions, even with experienced players.

H1 is an implication of Theorems 1 and 2, which provide conditions under which the contribution of every player to the public good is weakly decreasing. What we test here is that the restart effect and the declining nature of contributions are valid even for players who have become experienced with both the game and each other. H2 follows directly from Theorem 7. H3 and H4 test the hypothesis that play is approximately in equilibrium if players are experienced. H5 is intended to test Theorem 6. Finally, H6 examines the existence of conditional reciprocity when players are experienced.

7.2 Treatments

Table 1 describes the eight experimental treatments we ran. In all sessions involving play of the public good game, players played the stage game in groups of four, with per-period endowment of 20 tokens. The average public contribution was multiplied by a scale factor of 1.6, and players received this amount plus the remainder of their endowment after their contribution. The sessions varied on the number of repetitions of the stage game, the way in which players were reshuffled and restarted, the information solicited from participants before playing a set of ten games, and the information revealed to subjects after each game. In each session, subjects were asked to answer two control questions as a check on their understanding of the payoffs.²⁶ At the end of the session, tokens were converted to dollar amounts.

The first treatment, Session O, involved only one-shot games, with the intention to test H2. In the first part, subjects played one-shot public good contribution games and were reshuffled after each period. Finally, after the 19th period, there was a surprise restart and subjects played with the same group. In the second phase of the experiment, subjects played one-shot games, and after each period, the likelihood of being reshuffled was 0.75. With the remaining probability, the group stayed together to play an extra round. Subjects were informed in advance about these probabilities. In the second part, subjects played a total of 25 one-period games. The advantage of this design is that it makes possible to study the behavior of subjects after multiple restarts while still maintaining that restart comes as a surprise (the chance for the latter is only 25%) in order to mitigate repeated game effects.

Sessions A, B, C, as well as B-IH and C-IH, were used to examine H2-H4 and H6. In Session A, we first asked players to play a 10 times repeated public good game, where the group composition stayed the same for ten periods. Players were only informed that the stage game would be played for ten periods. At the end of each period, players were informed the average contribution of the group. After the tenth period, there was a surprise restart and the players were asked to play ten more games with the same group. This concluded part I of Session A.

In part II of Session A and in all of Session B, we used the same probabilistic

²⁶The actual instructions and control questions are available from the authors upon request.

continuation design (after any newly started 10-period game, reshuffling the group with probability 75%) as in part II of Session O.

Session C differed from Session B in only one way. At the beginning of every round of a ten-period repeated game, we asked subjects to report what they expected the average contribution of their three opponents to be in the first, fifth, and tenth period of the coming game. This session was intended to test H4.

Sessions B-IH and C-IH maintained the same structure as Sessions B and C, with the exception that after each period, the individual contribution of each group member was revealed to players. In a group, the opponents of a given player were randomly labeled A, B, and C, and the labeling stayed fixed for the ten-period game (and also if that game was restarted). After a player made a contribution decision, s/he was informed about what player A, B, and C contributed in that same period.

Finally, Session T was designed to test H5. Part I consisted of a gift exchange game. In this game, players were randomly paired, with each of them taking the role of first proposer and second proposer three times. No subjects were ever paired with the same opponent more than once. In all games, the first proposer started the game by deciding on what fraction of her endowment of 10 tokens to give to the second proposer. This amount was doubled and was given to the second proposer. The remaining portion of the first proposer's endowment was kept by her (but was not doubled). Following this first offer, the second proposer responded by deciding how much of her endowment of 10 tokens to give to the first proposer, who received double this amount. The remaining portion second proposer (but was not doubled). Play in the gift exchange games was used to identify the degree of reciprocity in players. We constructed a measure of reciprocity based on how players responded to positive proposals when playing as the second proposer. The index of reciprocity was the ratio of their response to the first proposer's offer when the first proposer's offer was nonzero. The higher this ratio was on average, the more reciprocal we regarded a player. In particular, we labeled the eight subjects with the highest reciprocity ratios as "reciprocal players" and labeled the eight subjects with the lowest reciprocity ratios as "selfish players." We labeled the rest of the players in the experiment "unidentified."

Subjects were not made aware of the second part of Session T until after completion of the first part. In part II, we asked players to play two sets of ten-period public good games in groups of four, where groups were randomly reshuffled, so that they become experienced with the game. Then, we sorted players into groups with three reciprocal players and one selfish player, and groups with one reciprocal player and three selfish players.²⁷ Before the start of the ten-period game, subjects were made aware of the histories of all their opponents as second proposers in the gift exchange game. After the tenth game, subjects were resorted into different groups of three reciprocal players and one selfish player, and groups with one reciprocal player and three selfish players, and played another round of the ten-period repeated game. Again the histories of everyone in the group as second proposers in part I were shown to the players before the game.

Session T-IH mimicked the design of Session T, with the exception that individual contributions of group members were revealed to players after every period.

8 Experimental results

All eight sessions were conducted with zTree (Fischbacher [99]) at the Computer Lab for Experimental Research at Harvard Business School. Subjects were recruited from the greater Boston area, with a large fraction of university students from Harvard, MIT, BU, and Northeastern. No subjects were allowed to participate in multiple sessions.

Table 2 summarizes the treatment conditions and payment profile of the subjects. Each session lasted approximately 1.5 hours. In the first session, subjects were paid on average of \$25.00, and the average number of tokens received was 1,660. Given the show-up fee of \$10.00, this corresponds to a ratio where 22 tokens is approximately 20 cents. In the other treatments, the conversion ratio was similar. The sessions were run between September 2005 and March 2006.²⁸

²⁷The unidentified players played among each other in the remaining groups.

²⁸We also ran earlier pilot studies in July 2005. We do not report the results of these sessions here.

In the following, when we discuss play in restarted games, we will pool play in restarted games from Session A, B, and C to have enough observations of play in restarted games. Similarly, we will pool observations from B-IH and C-IH together. Since there were some differences in the designs of these sessions, we tested for differences in play across the treatments. We found that play in the last four rounds (the observations on which we focus) in Session A, B, and C displayed the same general patterns, and we could not reject the hypothesis that the contributions in the first game and tenth game were equal across sessions. Similarly, we could not reject the hypothesis that contributions in the first and tenth game were equal across Session B-IH and C-IH.²⁹ In particular, soliciting expectations did not seem to change contribution patterns.

To examine the first hypothesis, Panel A of Table 3 shows the mean and median contributions in restarted games for the two rounds before the last two rounds. In all sessions, the mean contribution in the first period is positive and statistically different from zero, while in the tenth game, an overwhelming majority of players contribute zero. In Session A, B, and C, mean contributions are not strictly decreasing, though the general trend is downward and any non-monotonicities are less than one.³⁰ In B-IH and C-IH, the mean contributions are strictly declining.

Panel B of Table 3 shows the mean and median contributions in restarted games in the last two rounds. These are the games in which we regard players as experienced. In all sessions, contributions are significantly positive in the first game, and the majority of participants contribute zero in the last game.

Figure 1 plots the average contribution paths in different rounds in restarted games for Sessions A, B, and C pooled together. Darker lines refer to earlier rounds. With the exception of Round 6, which involved only one group of four in Session B, all other contribution patterns are declining. Figure 2 plots the average contribution paths in different rounds for Sessions B-IH and C-IH. The

²⁹The average contribution in the first game in the last four rounds in unrestarted games in Session A was 10.3, in Session B was 9.0, and in Session C was 9.6. For the tenth game, the average contribution in Session A was 2.4, in Session B was 0.5, and in Session C was 1.4. A joint test for the equality of contributions in the first and tenth game was not rejected at conventional significance levels. Similarly, a joint test for the equality of contributions in the first and tenth game between Session B-IH and Session C-IH was not rejected at conventional levels.

³⁰We only allow subjects to enter integer amounts as contributions.

overall trend in contribution patterns is Declining, and there are no noticeable patterns across rounds.

Figure 3 compares the pattern of play for a group that is restarted by plotting the average contribution path in the ten periods before the restart and in the ten periods after the restart. The figure demonstrates that there is a restart effect in these games, as the average contribution in the last period of the previous game is much smaller than play in the first period of the restarted game. In Figure 4, we see the same pattern in games in which individual contributions by players are displayed.

In Table 4, we report a formal test of the restart effect. Panel A displays average contributions before and after restarts in the last two rounds of the sessions and in two rounds before the last two rounds. We find that in the period before a restart, the mean contribution is 1.66 while the median is 0. Over three-fourths of players contribute zero. In contrast, in the first period of the restarted game, the mean contribution is 7.75 and only 29% of players contribute zero. Both the distribution of contributions and the fraction who contribute zero are statistically different. The same pattern also exists two rounds before the last two rounds. The mean contribution in the first period of the restarted game is 6.20, and 43% of contributions are zero in contrast to the last period. Interestingly, the positive restart effect is even greater in the last two rounds than in the two rounds before the last two rounds. This shows that even as players get experienced, they continue to display a significant restart effect.

In Panel B, we present the analogous set of results for Sessions B-IH and C-IH. We find that between 75% and 88% of players contribute zero in the tenth game, while between 13%-23% contribute zero in the restarted first game. Moreover, the average contribution jumps from between 1.56-2.57 to 10.16-10.69. As in the former treatments, both of these differences are statistically significant.

Conclusion 1. Experienced players contribute a statistically significant lower amount in the last period before a restarted game than in the first period of the restarted game. Furthermore, the average contribution in the first period of a restarted game is significantly higher than in the last period.

Table 5 describes play in part II of Session O. The table reports the mean and median contributions in restarted games, as well as the fraction of players

contributing zero in these games. For all of the games, 79% of subjects contribute zero, with the fraction who contribute zero increasing from period 5-9. In the last 5 periods, 94% of players in restarted games contribute zero. These patterns imply no restart effect in one-shot games with experienced players.

Conclusion 2: Experienced players' contributions in one-shot games are not significantly different from zero, and their play does not exhibit a restart effect.

Summarizing the previous two results, positive initial contributions and the restart effect continue to hold for experienced players in 10 times repeated games, but they disappear with experience in one-shot games.

Figure 1, which shows the evolution of average contribution across rounds, suggests that expected play stabilizes by the last two rounds. Table 6 presents a formal test of this, by comparing the average contribution in periods 1, 5, and 10 across sessions between the last two rounds and the two rounds prior in restarted games. In Panel A, we report that in period 1, the average contribution in the last two rounds was 7.75, while the average contribution in restarted games in two previous rounds was 6.20. The p -value of 0.20 shows that we cannot reject a hypothesis that play was from the same distribution. For game 5, in the last two rounds, the average contribution was 3.48 versus 5.66 in the two previous rounds. This difference was statistically different. Finally, in game 10, the average contribution in the last two rounds was 0.25 versus 1.41 two rounds prior. The F -test row of the table indicates that the joint of the equality of all three distributions of play across early and later rounds cannot be rejected at conventional significance levels. Panel B presents the same exercise for Sessions B-IH and C-IH. The table shows that average play in each of periods 1, 5, and 10 appears to be from the same distribution, and the joint test cannot reject the hypothesis that play in each game is from the same distribution. Panel C pools the data from the first two panels. The pattern that emerges supports the hypothesis that play in game 1 and game 10 is the same in the last two rounds and in the previous two rounds. Play in game 5 appears to be marginally statistically different. The joint test cannot reject the hypothesis that play is from the same distribution for each game.

To summarize, average contributions at the beginning of the game seem to stabilize after a few rounds of play, but we do not have conclusive evidence on whether the rate at which contributions go down to zero gets close to a steady

state during the same time.³¹ This is very similar to the findings of Selten and Stoecker [86], who examined play in 25 repetitions of a 10-period prisoner’s dilemma. They observed in their data that the cooperation appeared to stop earlier in the sequence of ten games in their later repetitions at a slow rate. Even with their considerably larger number of repetitions, they were unable to conclude whether cooperation would eventually terminate or whether it would converge to a limit.³²

Conclusion 3: Average contributions at the beginning and at the end of restarted games stabilize by the last two rounds of the experiments. It is unclear whether the rate at which contributions go to zero stabilizes by the same stage.

Table 7 examines the how closely expectations solicited before playing a sequence of ten games match actual play. The table focuses on restarted games in the last two rounds of Sessions C and C-IH. For the first game, players expect the average contribution to be 6.44, and on average, the actual average contribution of the three opponents is 6.90. For Session C-IH, the average expected play is 10.87, while the average actual play is 10.44. Similarly, for the fifth game and for the tenth game, the average expected play closely tracks the average actual play. The fraction of players overestimating others’ contributions is similar to the fraction who underestimate them. The median difference for Sessions C and C-IH together in game 1 is -1.17 , in game 5 is 0, and in game 10 is 0. This means that in games 5 and 10, remarkably, the median player’s expectations are exactly correct.

The last three columns show estimates from a regression of expected contribution on average contribution. The regression estimates show that the average contribution is correlated almost one-to-one with expected contribution, and the amount of variation explained by the average contribution is high. The pattern is the same for both the first game and the fifth game. For the tenth game, a large fraction of observations are 0. This prevents us from running the regression but indicates that a large fraction of subjects correctly anticipate at the beginning of the game that although initial contributions in the game will

³¹A further investigation of this would require sessions with a higher number of rounds. Given that each round of 10-period game takes up a considerable amount of time, the appropriate design would have to address how to maintain the concentration of subjects after playing the same game over and over again for a long period.

³²For more recent work on the rate at which cooperation decays, see Meyer and Roth [06].

be high, contributions at the end of the game will be close to zero. In short, players clearly foresee the declining pattern of contributions.

Although there are some differences between expected and actual play, taken together, these regressions show that there seems to be no large systematic error in expectations for experienced players in restarted games.

Conclusion 4: Experienced players on average correctly anticipate the pattern of contributions in a 10 times repeated game. In particular, they foresee that contributions will be close to zero by the end of the game.

Table 8 reports the results from sessions T (Panel A) and T-IH (Panel B). The first two rows of Panel A show the average contribution per player for ten games for the group with more reciprocal players and for the group with more selfish players. The more reciprocal groups tend to contribute a much higher amount per player than the more selfish groups. Also, in the groups with a majority of reciprocal players, the last positive contribution is always by a reciprocal player. In groups with more selfish players than reciprocal players, the last positive contribution is by a reciprocal player half of the time. The next two rows of Panel A report play in the more reciprocal groups. Since there is relatively little cooperation in groups with three selfish players, we only focus on groups with three reciprocal players when calculating the average contribution per player. The table reports that the average selfish player in a reciprocal group contributes 44 tokens over the span of ten games, while the average reciprocal player contributes 124 tokens over the span of ten games. Thus, the average contribution per game of the selfish player is 4.4, while for the reciprocal player it is 12.4.

In Panel B, we find that the average contribution with individualized histories is smaller in groups with more selfish players. We also find that selfish players on average contribute less than reciprocal players in groups with many reciprocal players.

In both Panels A and B, the last positive contribution is made by a reciprocal player in all of the groups with three reciprocal players, and half of the time, it is made by the sole reciprocal player in the groups with three selfish players. Panel C compares when selfish and reciprocal players stop contributing. The table shows that selfish players on average stop contributing between the 4th

and 5th game,³³ while reciprocal players stop contributing in their 6th game. The average difference between the stopping times is 1.4, and it is statistically significant at the 10% level, despite the small number of observations and the presumably noisy method of identifying types.

Conclusion 5: In 10 times repeated games, selfish players stop contributing earlier than reciprocal players.

The last table considers a measure of the importance of the strategic incentive to contribute. For the 10 times repeated public good games in Sessions A, B, C, B-IH, and C-IH, the table reports regressions on the correlation of the average first period contribution on subsequent play. The three columns regress an individual's contribution in the second game on the average contribution in the first period. The table reports the estimated coefficient and T-statistic in brackets of the impact of the average first period contribution. Each row corresponds to a different specification: OLS is simply an ordinary least squares regression with an intercept; session fixed effects include a fixed effects for whether the public good game was played in Session A, B, C, B-IH, or C-IH; and round fixed effects include fixed effects for the round of play. The fixed effects specifications control for the differences across sessions and time within a session. The three columns correspond to both restarted and not restarted games, only not restarted, and only restarted games. There is no difference across the type of game. The three columns all show a robust pattern: a unit increase in the first period contribution of a player increases others' contributions in the next period by around 0.4 unit.

This establishes the existence of conditional reciprocity even when players are experienced. We note that the amount of responsiveness we found in second period contributions is not enough by itself to induce selfish players to contribute in the first period. However, presumably first-period contributions have an effect on contributions in periods after the second one, too, increasing the selfish players' incentives to contribute at the beginning. We cannot estimate the latter effects from our data, because players' contributions are endogenous in all previous contributions of others and therefore in their own contributions up until two rounds preceding the current round.³⁴

³³This implies that the period of their last positive contribution is the 3rd and 4th game - 3.53.

³⁴The correlation coefficient between the total number of units a player contributes in rounds 2-10 and the average number of units the others contribute in round 1 is 4.5.

Conclusion 6: For experienced players, contributions in the first round positively affect contributions in the subsequent round.

9 Conclusion

This paper shows that all documented findings from finitely repeated public good contribution games can be explained by a model in which there are two types of players: ones who only care about their own material payoff, and ones who are conditionally reciprocal. The model can be extended to various other games, like centipede games and finitely repeated oligopoly games, in which selfish players have an incentive to cooperate at early stages if they anticipate that the cooperation might be reciprocated. One caveat is that in different settings, the domain and the range of reciprocity might change. For example, in centipede games in which at every stage, players face a binary decision whether to continue or terminate the game, reciprocity might be defined on continuation probabilities. We hope to return to investigate these issues in future work.

10 Appendix: Proofs

Lemma 1: Let $t \in \{0, \dots, T - 1\}$, $(x_i^1)_{i \in N}, \dots, (x_i^t)_{i \in N}$ be a length- t contribution history and $(x_i^{t+1})_{i \in S}, \dots, (x_i^T)_{i \in S}$ be a sequence of contributions by the selfish players after period t . Assume A1, A5 and A6 hold. Then there is a unique sequence of contributions by the reciprocal players after period t , $(x_i^{t+1})_{i \in R}, \dots, (x_i^T)_{i \in R}$ such that $x_i^k = f_i^k((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^k)_{j \in N/\{i\}}) \forall i \in R$ and $k \in \{t+1, \dots, T\}$, and $x_R^k = (I - M^k)^{-1} \hat{x}_R^k$, where $\hat{x}_i^k = \sum_{k'=1}^{k-1} \sum_{j \in N/\{i\}} f_{i,j}^{k,k'}(x_j^{k'}) + \sum_{j \in S} f_{i,j}^{k,k}(x_j^k) \forall k \in \{t+1, \dots, T\}$ and $i \in R$, $\hat{x}_R^k = (\hat{x}_i^k)_{i \in R}$, and M^k is the $(N - S) \times (N - S)$ matrix whose diagonal elements are 0 and its (m, n) -th element is $\alpha_{S+m, S+n}^{k,k}$.

Proof of Lemma 1: Assume that $k \in \{t+1, \dots, T\}$ is such that after any length- k history $(x_i^1)_{i \in N}, \dots, (x_i^k)_{i \in N}$ and any sequence of period $k+1$ to period T contributions $(x_i^{k+1})_{i \in S}, \dots, (x_i^T)_{i \in S}$ by the selfish players there is a unique

sequence of period $k + 1$ to period T contributions $(x_i^{k+1})_{i \in S}, \dots, (x_i^T)_{i \in S}$ by the reciprocal players such that

$$x_i^l = f_i^l((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^l)_{j \in N/\{i\}})$$

$\forall i \in R$ and $l \in \{k + 1, \dots, T\}$. Note that the above trivially holds for $k = T$.

Consider now any length $(k - 1)$ history $(x_i^1)_{i \in N}, \dots, (x_i^{k-1})_{i \in N}$ and any sequence of period k to period T contributions $(x_i^k)_{i \in S}, \dots, (x_i^T)_{i \in S}$ by the selfish players. By definition, if for some $(y_i^k)_{i \in R}$ it holds that

$$y_i^k = f_i^k((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^{k-1})_{j \in N/\{i\}}, ((x_j^k)_{j \in S}, (y_j^k)_{j \in R}))$$

$\forall i \in R$ then $x_i^k = \widehat{x}_R^k + M^k x_i^k$. Assumptions A5, A6 and $S > 1$ imply that $\sum_{j \in R/\{i\}} \alpha_{i,j}^{k,k} < 1 \forall i \in R$. Then by a well-known theorem (see Takayama (1985, p. 381)) $I - M^k$ is invertible and therefore the solution to $x_i^k = \widehat{x}_R^k + M^k x_i^k$ is unique and satisfies $x_R^k = (I - M^k)^{-1} \widehat{x}_R$. The claim then follows by induction. QED

Proof of Claim 1: First note that $\frac{A}{N} < 1$ implies that in any subgame perfect Nash equilibrium all selfish players contribute 0 after any $(T - 1)$ -length history $(x_i^1)_{i \in N}, \dots, (x_i^{T-1})_{i \in N}$. Furthermore, in any subgame perfect Nash equilibrium $x_i^T = E_i f_i^T((x_i^1)_{i \in N}, \dots, (x_i^{T-1})_{i \in N}, (x_j^T)_{j \in N/\{i\}}) \forall i \in R$, where the expectation is taken with respect to player i 's beliefs concerning $(x_j^T)_{j \in N/\{i\}}$ after history $(x_i^1)_{i \in N}, \dots, (x_i^{T-1})_{i \in N}$. Lemma 1 then implies that there is a unique continuation strategy profile $(x_j^T)_{j \in N}$ after $(x_i^1)_{i \in N}, \dots, (x_i^{T-1})_{i \in N}$ in subgame perfect Nash equilibrium, and in this continuation profile

$$x_i^T = \frac{N - S}{S + (T - 1)(N - 1)} \sum_{t'=1}^{T-1} \sum_{j \in S} x_j^{t'} +$$

$$\sum_{t'=1}^{T-1} \sum_{j \in R/\{i\}} \frac{NT - T}{S - N + 2T + NT + ST + NST - T^2 - N^2T + N^2T^2 + 1} x_j^{t'} +$$

$$\sum_{t'=1}^{T-1} \frac{N - 1 - S}{S - N + 2T + NT + ST + NST - T^2 - N^2T + N^2T^2 + 1} x_i^{t'}$$

$\forall i \in R$. Note that only the first term depends on selfish players' contributions.

Let $t \in \{1, \dots, T-1\}$ and assume that for every $k \in \{t, \dots, T-1\}$ and for every length- k history $(x_i^1)_{i \in N}, \dots, (x_i^k)_{i \in N}$ all subgame perfect Nash equilibria specify the same continuation profile, which is history-independent for selfish players and satisfies:

$$x_i^{k+1} = \frac{N-S}{S+k(N-1)} \sum_{t'=1}^k \sum_{j \in S} x_j^{t'} +$$

$$\sum_{t'=1}^k \sum_{j \in R/\{i\}} \frac{(N-1)(k+1)}{S-N+1+(k+1)(2+N+S+NS-N^2)+(k+1)^2(N^2-1)} x_j^{t'} +$$

$$\sum_{t'=1}^k \frac{N-1-S}{S-N+1+(k+1)(2+N+S+NS-N^2)+(k+1)^2(N^2-1)} x_i^{t'}$$

for every $i \in R$. Then a marginal contribution by any $i \in S$ at t has zero impact on contributions of any $j \in S$, and its marginal impact on the total future contributions of any $j \in R$ is $C(t)$. By assumption $C(t) \neq 1 - \frac{A}{N}$. Then independently of history, in any subgame perfect Nash equilibrium all selfish players contribute 1 at t if $C(t) > 1 - \frac{A}{N}$ and 0 if $C(t) < 1 - \frac{A}{N}$. Note that the starting assumption implies that in any subgame perfect Nash equilibrium, after any length- t history all reciprocal players get a per period payoff of 0 in periods $t+1, \dots, T$. Then after any length- $(t-1)$ history $(x_i^1)_{i \in N}, \dots, (x_i^{t-1})_{i \in N}$, it has to hold that $x_i^t = E_i f_i^t((x_i^1)_{i \in N}, \dots, (x_i^{t-1})_{i \in N}, (x_j^t)_{j \in N/\{i\}}) \forall i \in R$, where the expectation is taken with respect to player i 's beliefs concerning $(x_j^t)_{j \in N/\{i\}}$ after history $(x_i^1)_{i \in N}, \dots, (x_i^{t-1})_{i \in N}$. Then Lemma 1, together with the starting assumption, implies that for any length- $(t-1)$ history, there is a unique continuation profile in subgame perfect Nash equilibrium, which is history-independent for selfish players and satisfies:

$$x_i^k = \frac{N-S}{S+(k-1)(N-1)} \sum_{t'=1}^{k-1} \sum_{j \in S} x_j^{t'} +$$

$$\sum_{t'=1}^{k-1} \sum_{j \in R/\{i\}} \frac{(N-1)k}{S-N+1+k(2+N+S+NS-N^2)+k^2(N^2-1)} x_j^{t'} +$$

$$\sum_{t'=1}^{k-1} \frac{N-1-S}{S-N+1+k(2+N+S+NS-N^2)+k^2(N^2-1)} x_i^{t'}$$

for every $k \in \{t, \dots, T\}$ and $i \in R$. The claim then follows by induction. QED

Proof of Theorem 1: The same arguments as in the proof of Claim 1 establish that in every subgame perfect Nash equilibrium at period T , all selfish players contribute 0 after any $(T-1)$ -length history, and that the continuation strategy of reciprocal players after any $(T-1)$ -length history $(x_i^1)_{i \in N}, \dots, (x_i^{T-1})_{i \in N}$ is uniquely determined in subgame perfect Nash equilibrium and satisfies $x_i^T = f_i^T((x_j^1)_{j \in N}, \dots, (x_j^{T-1})_{j \in N}, ((0)_{j \in S}, (x_j^T)_{j \in R/\{i\}}) \forall i \in R$.

Let $t \in \{1, \dots, T-1\}$ and assume that for almost every A (if $f_{i,j}^{k,l}$ is strictly concave for every $i \in R, j \in S$ and $k, l \in \{1, \dots, T\}$ then for every A), the following hold: for every $k \in \{t, \dots, T-1\}$ and for every length- k history $(x_i^1)_{i \in N}, \dots, (x_i^k)_{i \in N}$, all subgame perfect Nash equilibria specify the same continuation profile, which is history-independent for selfish players and satisfies that $x_i^l = f_i^l((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^l)_{j \in N/\{i\}}) \forall i \in R$ and $l \in \{k+1, \dots, T\}$. Consider now an arbitrary $(t-1)$ -length history $(x_i^1)_{i \in N}, \dots, (x_i^{t-1})_{i \in N}$. Note that any period- t contribution by a selfish player cannot influence period- t contributions by other players. Furthermore, by the starting assumption, selfish players' contributions from $t+1$ on are generically uniquely pinned down in subgame perfect Nash equilibrium. Therefore a period- t contribution by a selfish player can only influence reciprocal players' contributions from $t+1$ on. Let $M^{k,k'}$ be the $(N-S) \times (N-S)$ matrix whose diagonal components are 0 and whose (m,n) -th component is $\alpha_{S+m, S+n}^{k,k'}$. Then Lemma 1, together with A1, implies that contribution x_i^t by $i \in S$ induces a total of $\mathbf{1}(I - M^{t+1})^{-1}(f_{j,i}^{t+1,t}(x_i^t))_{j \in R}$ contributions at $t+1$, a total of $(I - M^{t+1})^{-1}[M^{t+2,t+1}((I - M^{t+1})^{-1}(f_{j,i}^{t+1,t}(x_i^t))_{j \in R} + (f_{j,i}^{t+2,t}(x_i^t))_{j \in R})]$ contributions at $t+2$, and in general for $l \in \{t+1, \dots, T\}$ a total which is equal to a linear nonnegative combination of $(f_{j,i}^{t+1,t}(x_i^t))_{j \in R}, \dots, (f_{j,i}^{l,t}(x_i^t))_{j \in R}$ at l . The aggregate impact, and therefore the aggregate change in the flow future payoffs of i , $\Delta_i^{t+}(x_i^t)$ is given by a linear nonnegative combination of $(f_{j,i}^{t+1,t}(x_i^t))_{j \in R}, \dots, (f_{j,i}^{T,t}(x_i^t))_{j \in R}$. Since $f_{j,i}^{l,t}$ is concave, differentiable, and increasing for every $j \in R$ and $l \in$

$\{t + 1, \dots, T\}$, Δ_i^{t+} is also concave, differentiable, and increasing. Furthermore, if $f_{j,i}^{l,t}$ is strictly concave for every $j \in R$ then Δ_i^{t+} is strictly concave. Meanwhile, the net change in the flow payoff of i at t is $(\frac{A}{N} - 1)x_i^t$, a linear and decreasing function. Since Δ_i^{t+} is concave and increasing, there can only be a countable set of parameter values for A such that $\frac{d\Delta_i^{t+}}{dx_i^t} = 1 - \frac{A}{N}$ has multiple solutions in $[0, 1]$, and if Δ_i^{t+} is strictly concave then there are no parameter values like that. Therefore, for almost all values of A (for any A if $f_{j,i}^{l,t}$ is strictly concave for every $j \in R$ and $l \in \{t + 1, \dots, T\}$), the contribution of i is uniquely pinned down after $(x_j^1)_{j \in N}, \dots, (x_i^{t-1})_{j \in N}$. Since this holds for any $i \in S$ and length- $(t - 1)$ history $(x_j^1)_{j \in N}, \dots, (x_i^{t-1})_{j \in N}$, the induction assumption implies that for almost all values of A (for any value of A if reciprocity functions towards selfish players are strictly concave), the continuation strategy after $t - 1$ is uniquely pinned down for all selfish players in subgame perfect Nash equilibrium.

Note that the induction assumption implies that in any subgame perfect Nash equilibrium, after any length- t history, all reciprocal players get a flow payoff of 0 in periods $t + 1, \dots, T$. Therefore, after any length- $(t - 1)$ history $(x_j^1)_{j \in N}, \dots, (x_j^{t-1})_{j \in N}$ and any subgame perfect Nash equilibrium s , it holds that the action specified by s_i after $(x_j^1)_{j \in N}, \dots, (x_j^{t-1})_{j \in N}$ is

$$E_i f_i^T((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^T)_{j \in N/\{i\}})$$

$\forall i \in R$, where the expectation is taken with respect to player i 's beliefs concerning $(x_j^T)_{j \in N/\{i\}}$ after history $(x_i^1)_{i \in N}, \dots, (x_i^{T-1})_{i \in N}$. Lemma 1 then implies that after $(x_j^1)_{j \in N}, \dots, (x_j^{t-1})_{j \in N}$ the period t contribution of i is uniquely pinned down in subgame perfect Nash equilibrium, and

$$x_i^t = f_i^T((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^T)_{j \in N/\{i\}}).$$

This concludes that for almost every A (if $f_{i,j}^{k,l}$ is strictly concave for every $i \in R$, $j \in S$ and $k, l \in \{1, \dots, T\}$ then for every A), the following hold: for every $k \in \{t - 1, \dots, T - 1\}$ and for every length- k history $(x_i^1)_{i \in N}, \dots, (x_i^k)_{i \in N}$, all subgame perfect Nash equilibria specify the same continuation profile, which is history-independent for selfish players and satisfies:

$$x_i^l = f_i^l((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^l)_{j \in N/\{i\}})$$

$\forall i \in R$ and $l \in \{k + 1, \dots, T\}$. By induction then, for almost every A (if $f_{i,j}^{k,l}$ is strictly concave for every $i \in R$, $j \in S$ and $k, l \in \{1, \dots, T\}$ then for

every A), there is a unique subgame perfect Nash equilibrium, which is history-independent for selfish players and satisfies:

$$x_i^l = f_i^l((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^l)_{j \in N/\{i\}})$$

$\forall i \in R$ and $l \in \{1, \dots, T\}$.

A4 implies that $\frac{d\Delta_i^{t+}(x)}{dx} \geq \frac{d\Delta_i^{(t+1)+}(x)}{dx}$ for every $i \in S$, $x \in [0, 1]$ and $t \in \{1, \dots, T-1\}$; therefore, for generic parameter values, the contributions of selfish players are weakly decreasing in subgame perfect Nash equilibrium. A3 then implies that along the equilibrium path, the contributions of all players are weakly decreasing. QED

Proof of Theorem 2: Consider any pure strategy perfect Bayesian Nash equilibrium s that satisfies the properties stated in the theorem, and any length- $(T-1)$ history $h^{T-1} = (x_j^1)_{j \in N}, \dots, (x_j^{T-1})_{j \in N}$ which is such that at every period, at most one selfish player deviated from the action profile specified by s . The same arguments as in the proof of Theorem 1 establish that s specifies $x_i^T = 0$ after any length- $(T-1)$ history for every $i \in S$. Let $h^l = (x_j^1)_{j \in N}, \dots, (x_j^l)_{j \in N}$ for $l = \{1, \dots, T-2\}$ and let h^0 be the null history. Then the assumption that it is common certainty that reciprocal players think that a selfish player is responsible for any deviation implies that after $(x_j^1)_{j \in N}, \dots, (x_j^{T-1})_{j \in N}$ s specifies a contribution level of $E_i f_i^T((y_j^1)_{j \in N/\{i\}}, \dots, (y_j^{T-1})_{j \in N/\{i\}}, (x_j^T)_{j \in N/\{i\}})$ at T for $i \in R$, where

$$(y_j^l)_{j \in N} = (s_1^l(h^{l-1}) + \sum_{j \in N} x_j^l - \sum_{j \in N} s_j^l(h^{l-1}), s_{-1}^l(h^{l-1}))$$

$\forall l \in \{1, \dots, T-1\}$, and the expectation is taken with respect to i 's belief concerning $(x_j^T)_{j \in N}$. Then the same arguments as in the proof of Theorem 1 establish that after $(x_j^1)_{j \in N}, \dots, (x_j^{T-1})_{j \in N}$ all subgame perfect equilibria specify the same continuation profile, which satisfies:

$$x_i^T = f_i^T((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^{T-1})_{j \in N/\{i\}}, (x_j^T)_{j \in N/\{i\}})$$

$\forall i \in R$. This implies that after any history like the above, the flow payoff of any $i \in R$ is 0 in s . Furthermore, the above implies that if for two perfect Bayesian Nash equilibria satisfying the properties in the theorem it holds that along the equilibrium path they specify the same actions in periods $\{1, \dots, T-1\}$, then

they also specify the same actions in period T .

Suppose now that for some $t \in \{2, \dots, T\}$ and $k \geq t - 1$, it holds that, for any perfect Bayesian Nash equilibrium satisfying the properties in the theorem, it is true that after any length- k history $(x_j^1)_{j \in N}, \dots, (x_j^k)_{j \in N}$ which is such that at every period at most one selfish player deviated from the action profile specified by the equilibrium, the continuation profile is such that $x_i^{k+1} = f_i^T((x_j^1)_{j \in N/\{i\}}, \dots, (x_j^{T-1})_{j \in N/\{i\}}, (x_j^T)_{j \in N/\{i\}}) \forall i \in R$. Suppose also that if s and s' are two perfect Bayesian Nash equilibria satisfying the properties in the theorem and $(x_j^1)_{j \in N}, \dots, (x_j^k)_{j \in N}$ and $(y_j^1)_{j \in N}, \dots, (y_j^k)_{j \in N}$ are two length- k histories which are such that at every period at most one selfish player deviated from the action profile specified by the equilibrium, then s and s' specify the same actions along the continuation equilibrium path for all $i \in S$, and that this is the same action sequence as implied by the unique subgame perfect Nash equilibrium of the game in which individual contributions are revealed. Then the same arguments as in the proof of Theorem 1 establish that the same claims hold for all length- $(t-2)$ histories for generic A (for all A if $f_{i,j}^l$ is strictly concave $\forall l \in \{1, \dots, T\}$, $i \in R$ and $j \in S$).

Since the null history satisfies that at every period, at most one selfish player deviated from the action profile specified by the equilibrium, by induction for generic A it holds that every perfect Bayesian Nash equilibrium satisfying the properties in the theorem implies the same equilibrium path, which is equal to the contribution path implied by the unique subgame perfect Nash equilibrium of the game in which individual contributions are revealed.

A perfect Bayesian Nash equilibrium satisfying the properties in the theorem can be constructed by specifying the above action choices for histories satisfying that at every period, at most one selfish player deviated from the action profile specified by the unique subgame perfect Nash equilibrium of the game in which individual contributions are revealed, and arbitrary continuation equilibria that are strongly symmetric with respect to selfish players after other histories. QED

Lemma 2: Consider two contribution games \mathcal{G} and $\widehat{\mathcal{G}}$ with individual contributions revealed after rounds, in which the players are the same: $N = \widehat{N}$, $S = \widehat{S}$, and $f_i^t = \widehat{f}_i^t \forall i \in R$ and $t \in \{1, \dots, \min(T, \widehat{T})\}$. Suppose \mathcal{G} and $\widehat{\mathcal{G}}$ have unique subgame perfect Nash equilibria s and \widehat{s} . If $s_i^t \leq \widehat{s}_i^t \forall i \in S$, then $s_i^t \leq \widehat{s}_i^t \forall i \in N$.

Proof of Lemma 2: Let $T^* = \min(T, \widehat{T})$. Let $\widehat{y}_i^k = \sum_{k'=1}^{k-1} \sum_{j \in N/\{i\}} f_{i,j}^{k,k'}(\widehat{s}_j^{k'}) +$

$\sum_{j \in S} f_{i,j}^{k,k}(\widehat{s}_j^1)$, and let $y_i^k = \sum_{k'=1}^{k-1} \sum_{j \in N/\{i\}} f_{i,j}^{k,k'}(s_j^{k'}) + \sum_{j \in S} f_{i,j}^{k,k}(s_j^1) \forall i \in R$ and $k \in \{1, \dots, T^*\}$. Suppose that for some $k \in \{1, \dots, T^*\}$, it holds that $\widehat{s}_i^t \geq s_i^t \forall i \in N$ and $t \in \{1, \dots, k-1\}$, and $\widehat{s}_i^k \geq s_i^k \forall i \in S$. Note that this holds for $k=1$. Lemma 1 implies that $\widehat{s}_R^k = (I - M^k)^{-1} \widehat{y}_R^k$ and $s_R^k = (I - M^k)^{-1} y_R^k$, where $\widehat{y}_R^k = (\widehat{y}_i^k)_{i \in R}$, $y_R^k = (y_i^k)_{i \in R}$, $\widehat{s}_R^k = (\widehat{s}_i^k)_{i \in R}$, $s_R^k = (s_i^k)_{i \in R}$, and M^k is the $(N-S) \times (N-S)$ matrix whose diagonal elements are 0 and its (m,n) -th element is $\alpha_{S+m, S+n}^{k,k}$. Since $\sum_{j \in S} f_{i,j}^{k,k}(s_j^k)$ is increasing in $s_j^k \forall i \in R$ and $j \in S$, and $(I - M^k)^{-1} y_R^k$ is increasing in y_R^k in the relevant nonnegative range, the above establishes that $\widehat{s}_i^k \geq s_i^k \forall i \in R$. Then $k < T$ implies $\widehat{s}_i^t \geq s_i^t \forall i \in N$ and $t \in \{1, \dots, k\}$, and $\widehat{s}_i^{k+1} \geq s_i^{k+1} \forall i \in S$. The claim then follows by induction. QED

Proof of Theorem 3: The proof of Theorem 1 establishes that the equilibrium contribution of any player $i \in S$ in any period $t \in \{1, \dots, T\}$ is given by $x_i^t = \arg \max_{x \in [0,1]} ((\frac{A}{N} - 1)x + \Delta_i^{t+}(x))$ if the return to contributing is A , and it is given by $x_i^t = \arg \max_{x \in [0,1]} ((\frac{\widehat{A}}{N} - 1)x + \Delta_i^{t+}(x))$ if the return to contributing is \widehat{A} , where Δ_i^{t+} is a term that increases in x . $\widehat{A} > A$ then implies that $\widehat{s}_i^t \geq s_i^t \forall i \in S$ and $t \in \{1, \dots, T\}$, where \widehat{s} is the unique subgame perfect Nash equilibria of the game in which the return to contributing is \widehat{A} and s is the unique subgame perfect Nash equilibria of the game in which the return to contributing is A . The claim then follows from Lemma 2. QED

Proof of Theorem 4: Theorem 1 establishes that the equilibrium contribution of any player $i \in S$ in any period $t \in \{1, \dots, T\}$ is given by $x_i^t = \arg \max_{x \in [0,1]} ((\frac{A}{N} - 1)x + \Delta_i^{t+}(x))$, where Δ_i^{t+} is a term that increases in x and increases in T . This implies $\widehat{s}_i^t \geq s_i^t \forall i \in S$ and $t \in \{1, \dots, T\}$. The claim then follows from Lemma 2. QED

Proof of Theorem 5: Let $i \in \{1, \dots, \widehat{S}\}$ and $t \in \{1, \dots, T-1\}$. For any $k \in \{t+1, \dots, T\}$ and $x_i^t \in [0,1]$ let $c_j^k(x_i^t)$ denote the marginal impact in \mathcal{G} of a contribution at t by i on the period- k contribution of j , assuming that the contribution does not effect period- k contributions of selfish players, and that $x_j^k = f_j^k((x_{j'}^1)_{j' \in N/\{j\}}, \dots, (x_{j'}^{k-1})_{j' \in N/\{j\}}) \forall j \in R$ and $k \in \{t+1, \dots, T\}$. Let $\widehat{c}_j^k(x_i^t)$ be the corresponding marginal impact in $\widehat{\mathcal{G}}$. Let

$\underline{f}_i^{t+1,t}(x_i^t) = (f_{S+1,i}^{t+1,t}(x_i^t), \dots, f_{N,i}^{t+1,t}(x_i^t))$ and $\widehat{f}_i^{t+1,t}(x_i^t) = (\widehat{f}_{\widehat{S}+1,i}^{t+1,t}(x_i^t), \dots, \widehat{f}_{N,i}^{t+1,t}(x_i^t))$.
For any $j \in R$ let $f_{i,j}^{t+1,t}(x_i^t)$

By Lemma 1, for any $j \in R$, $c_j^{t+1}(x_i^t) = \frac{\partial((I-M^{t+1})^{-1}f_i^{t+1,t}(x_i^t)_{j-s})}{\partial x_i^t}$ and $\widehat{c}_j^{t+1}(x_i^t) = \frac{\partial((I-\widehat{M}^{t+1})^{-1}\widehat{f}_i^{t+1,t}(x_i^t)_{j-s})}{\partial x_i^t}$, where M^{t+1} is the $(N-S) \times (N-S)$ matrix whose diagonal elements are 0 and its (m,n) -th element is $\alpha_{S+m,S+n}^{t+1,t+1}$, and \widehat{M}^{t+1} is the $(N-\widehat{S}) \times (N-\widehat{S})$ matrix whose diagonal elements are 0 and its (m,n) -th element is $\widehat{\alpha}_{\widehat{S}+m,\widehat{S}+n}^{t+1,t+1}$. Since $\widehat{f}_{j,i}^{t+1,t}$ is an increasing function for every $j \in \{\widehat{S}+1, \dots, S\}$ and $\widehat{f}_{j,j'}^{t+1,t+1}$ is an increasing function for every $j, j' \in \{\widehat{S}+1, \dots, S\}$, it follows that $\widehat{c}_j^{t+1}(x_i^t) \geq c_j^{t+1}(x_i^t) \forall j \in \{\widehat{S}+1, \dots, N\}$. Suppose now that for some $k \in \{t+1, \dots, T-1\}$, it holds that $\widehat{c}_j^k(x_i^t) \geq c_j^k(x_i^t) \forall j \in \{\widehat{S}+1, \dots, N\}$. Then since $\widehat{f}_{j,j'}^{k+1,l}$ is an increasing function for every $j, j' \in \{\widehat{S}+1, \dots, S\}$ and $l \in \{t+1, \dots, k+1\}$, it follows that $\widehat{c}_j^{k+1}(x_i^t) \geq c_j^{k+1}(x_i^t) \forall j \in \{\widehat{S}+1, \dots, N\}$. Then by induction $\sum_{j \in \widehat{R}} \sum_{k=t+1}^T \widehat{c}_j^k(x_i^t) \geq \sum_{j \in R} \sum_{k=t+1}^T c_j^k(x_i^t)$, that is

$$\frac{\partial \widehat{\Delta}_i^{t+}(x_i^t)}{\partial x_i^t} \geq \frac{\partial \Delta_i^{t+}(x_i^t)}{\partial x_i^t}. \quad (*)$$

Since \mathcal{G} and $\widehat{\mathcal{G}}$ both have unique subgame perfect Nash equilibria, in both s and \widehat{s} a marginal contribution at t by i does not effect future contributions of selfish players, and for every $k \in \{t+1, \dots, T\}$ it holds that $s_j^k = f_j^k((s_{j'}^1)_{j' \in N/\{j\}}, \dots, (s_{j'}^{k-1})_{j' \in N/\{j\}}) \forall j \in R$ and

$$\widehat{s}_j^k = f_j^k((\widehat{s}_{j'}^1)_{j' \in N/\{j\}}, \dots, (\widehat{s}_{j'}^{k-1})_{j' \in N/\{j\}}) \forall j \in R.$$

Furthermore, $s_i^t = \arg \max_{x \in [0,1]} ((\frac{A}{N} - 1)x + \Delta_i^{t+}(x))$ and

$$\widehat{s}_i^t = \arg \max_{x \in [0,1]} \left(\left(\frac{A}{N} - 1 \right) x + \Delta_i^{t+}(x) \right).$$

Then (*) implies $\widehat{s}_i^t \geq s_i^t$ (note that (*) holds for any $x_i^t \in [0,1]$). QED

Proof of Theorem 6: By Theorem 1, $s_i^t > 0$ for some $i \in S$ implies $s_i^k > 0 \forall k \in \{1, \dots, t\}$. A5 then implies the claim. QED

Proof of Theorem 7: For any $i \in S$, contributing 0 is a dominant strategy, therefore $s_i^1 = 0$ for any Nash equilibrium s . By Lemma 1, there is a

unique vector of contributions x_{S+1}, \dots, x_N by the reciprocal players such that $f_i(0, \dots, 0, x_{S+1}, \dots, x_N) = x_i \forall i \in R$. Since $f_i(0, \dots, 0) = 0 \forall i \in R$, the unique Nash equilibrium of the game is then $s_i^1 = 0 \forall i \in N$. QED

11 References

- ANDERSON, S., J. GOEREE and C. HOLT (1998): "A theoretical analysis of altruism and decision error in public goods games," *Journal of Public Economics*, 70(2), 297-323.
- ANDREONI, J. (1988): "Why free ride? Strategies and learning in public goods experiments," *Journal of Public Economics*, 37(3), 291-304.
- ANDREONI, J. (1989): "Giving with impure altruism: applications to charity and Ricardian equivalence," *Journal of Public Economics*, 37(3), 291-304.
- ANDREONI, J. and J. MILLER (1993): "Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence," *Economic Journal*, 103, 570-585.
- ANDREONI, J. and L. SAMUELSON (2006): "Building rational cooperation," *Journal of Economic Theory*, 127, 117-154.
- BAGNOLI, M. and M. McKEE (1991): "Voluntary contribution games: efficient private provision of public goods," *Economic Inquiry*, 29, 351-66.
- BANDIERA, O., I. BARANKAY and I. RASUL (2006): "Social preferences and the response to incentives: evidence from personnel data," *Quarterly Journal of Economics*, forthcoming
- BOLTON, G. E. and A. OCKENFELS (2000): "ERC: A theory of equity, reciprocity, and competition," *American Economic Review* 90, 166-193.
- BRANDTS, J. and A. SCHRAM (1996): "Cooperative gains or noise in public goods experiments," *mimeo University of Amsterdam*.
- 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, M. and HEY, J. (1997): "Do Anglo-Saxons free-ride more?," *Journal of Public Economics*, 64, 41-60.
- CROSON, R. (1996): "Partners and strangers revisited," *Economics Letters* 25, 25-32.
- DUWFENBERG, M. and G. KIRCHSTEIGER (2004): "A theory of sequential reciprocity," *Games and Economic Behavior*, 47, 268-298.
- FEHR, E. and K. SCHMIDT (1999): "A theory of fairness, competition, and cooperation," *Quarterly Journal Of Economics* 114, 817-868.
- FEHR, E. and K. SCHMIDT (2002): "Theories of fairness and reciprocity - evidence and economic applications," IN *Advances in Economics and Econometrics - 8th World Congress, Econometric Society Monographs*, Cambridge, Cambridge University Press.
- FEHR, E. and E. TOUGAREVA (1995): "Do high monetary stakes remove reciprocal fairness? Experimental evidence from Russia," *Institute for Empirical Research in Economics, University of Zürich, Working Paper No. 120*.
- FIELD, A. (2002): "Altruistically inclined? The behavioral sciences, evolutionary theory and the origins of reciprocity," Ann Arbor: UMichigan Press.
- FISCHBACHER, U. (1999): z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual, Working Paper 21, Institute for Empirical Research in Economics, University of Zurich.

FISCHBACHER, U., S. GÄCHTER and E. FEHR (2001): "Are people conditionally cooperative? Evidence from a public goods experiment," *Economics Letters*, 71, 397-404.

FISKE, A. P. (1992): "The four elementary forms of sociality: framework for unified theory of social relations," *Psychological Review*, 99 (4), 689-723.

FUDENBERG, D. and D. LEVINE (1997): "Measuring players' losses in experimental games," *Quarterly Journal of Economics*, 112 (2), 507-536.

FUDENBERG, D. and E. MASKIN (1986): "The folk theorem in repeated games with discounting and incomplete information," *Econometrica*, 54: 533-554.

HOFFMAN, E., K. McCABE and V. SMITH (1996): "On expectations and monetary stakes in ultimatum games," *International Journal of Game Theory*, 25, 289-301.

ISAAC, M. and J. WALKER (1988): "Group size effects in public goods provision: the voluntary contributions mechanism," *Quarterly Journal of Economics*, 103(1), 179-99.

ISAAC, M., J. WALKER and S. THOMAS (1984): "Divergent evidence on free riding: an experimental evidence of possible explanations," *Public Choice*, 43(1), 113-49.

ISAAC, M., J. WALKER and A. WILLIAMS (1994): "Group size and the voluntary provision of public goods: experimental evidence utilizing large groups," *Journal of Public Economics*, 54, 1-36.

KAHN, L. and J. MURNINGHAM (1993): "Conjecture, uncertainty, and cooperation in the finitely repeated prisoner's dilemma: experimental evidence," *Journal of Economic Behavior and Organization*, 22, 91-117.

KESER, C. (2000): "Strategically planned behavior in public good experiments," CIRANO Working Paper.

KESER, C. and F. WINDEN (2000): "Conditional cooperation and voluntary contributions to public goods," *Scandinavian Journal of Economics*, 102(1), 23-39.

KIM, O. and M. WALKER (1984): "The free rider problem: experimental evidence," *Public Choice*, 18, 3-24.

KLUMPP, T. (2004): "Finitely repeated provision of a public good," *mimeo Indiana University*.

KREPS, D. and R. WILSON (1982): "Reputation and imperfect information," *Journal of Economic Theory*, 27, 253-279.

LEDYARD, J. (1995): "Public goods," IN: *Handbook of Experimental Economics*, ed. by Kagel, J. and A. Roth, Princeton University Press.

LEVINE, D. (1998): "Modeling altruism and spitefulness in experiments," *Review of Economic Dynamics*, 1, 593-622.

McKELVEY, R. and T. PALFREY (1992): "An experimental study of the centipede game," *Econometrica*, 60, 803-836.

MEYER, Y. and A. ROTH (2006): "The speed of learning in noisy games: partial reinforcement and the sustainability of cooperation," *American Economic Review*, forthcoming.

- NAGEL, R. and F. TANG (1998): "An experimental study on the centipede game in normal form an investigation on learning," *Journal of Mathematical Psychology*, 42, 356-384.
- NEYMAN, A. (1985): "Bounded complexity justifies cooperation in the prisoner's dilemma," *Economic Letters*, 19: 227-229.
- NEYMAN, A. (1999): "Cooperation in repeated games when the number of stages is not commonly known," *Econometrica*, 67: 45-64.
- OFFERMAN, T., J. SONNEMANS, & A. SCHRAM (1996): "Value orientations, expectations, and voluntary contributions in public goods," *Economic Journal*, 106, 817-845.
- PALFREY, T. and J. PRISBEY (1997): "Anomalous behavior in linear public goods experiments: how much and why?," *American Economic Review*, 87 (5), 829-46.
- PRICE, M. (2006): "Monitoring, reputation, and 'greenbeard' reciprocity in a Shuar work team," *Journal of Organizational Behavior*, 27, 201-219.
- RABIN, M. (1993): "Incorporating fairness into game theory and economics," *American Economic Review*, 83, 1281-1302.
- RABIN, M. (1994): "Incorporating behavioral assumptions into game theory," in James Friedman (ed.), *Problems of Coordination in Economic Activity*, Norwell, MA: Kluwer Academic Publishers.
- RADNER, R. (1980): "Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives," *Journal of Economic Theory*, 22: 136-154.
- RADNER, R. (1986): "Can bounded rationality resolve the prisoner's dilemma?" in Andreu Mas-Colell and W. Hildenbrand (eds.), *Contributions to Mathematical Economics*, North-Holland: Amsterdam, 387-399.
- ROTH, A. E., V. PRASNIKAR, M. OKUNO-FUJIWARA and S. ZAMIR (1991): "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: an experimental study," *American Economic Review*, 81, 1068-95.
- SAIJO, T. and H. YAMAGUCHI (1992): "The 'spite' dilemma in voluntary contribution mechanism experiments," *University of Tsukuba mimeo*.
- SEFTON, M. and R. STEINBERG (1996): "Reward structures in public good experiments," *Journal of Public Economics*, 61, 263-287.
- SELTEN, R., M. MITZKEWITZ and G. UHLICH (1997): "Duopoly strategies programmed by experienced players," *Econometrica*, 65, 517-555.
- SELTEN, R. and R. STOECKER (1986): "End behavior in sequences of finite prisoner's dilemma supergames: a learning theory approach," *Journal of Economic Behavior and Organization*, 7, 47-70.
- SLONIM, R. and A. ROTH (1997): "Financial incentives and learning in ultimatum and market games: an experiment in the Slovak Republic," *Econometrica*, 65, 569-596.
- SOBEL, J. (1985): "A theory of credibility," *Review of Economic Studies*, 52: 557-573.
- TAKAYAMA, A. (1985): "Mathematical economics," *Cambridge University Press*.

Figure 1: Average Contribution in Restarted Games by Round, Session A, B, and C

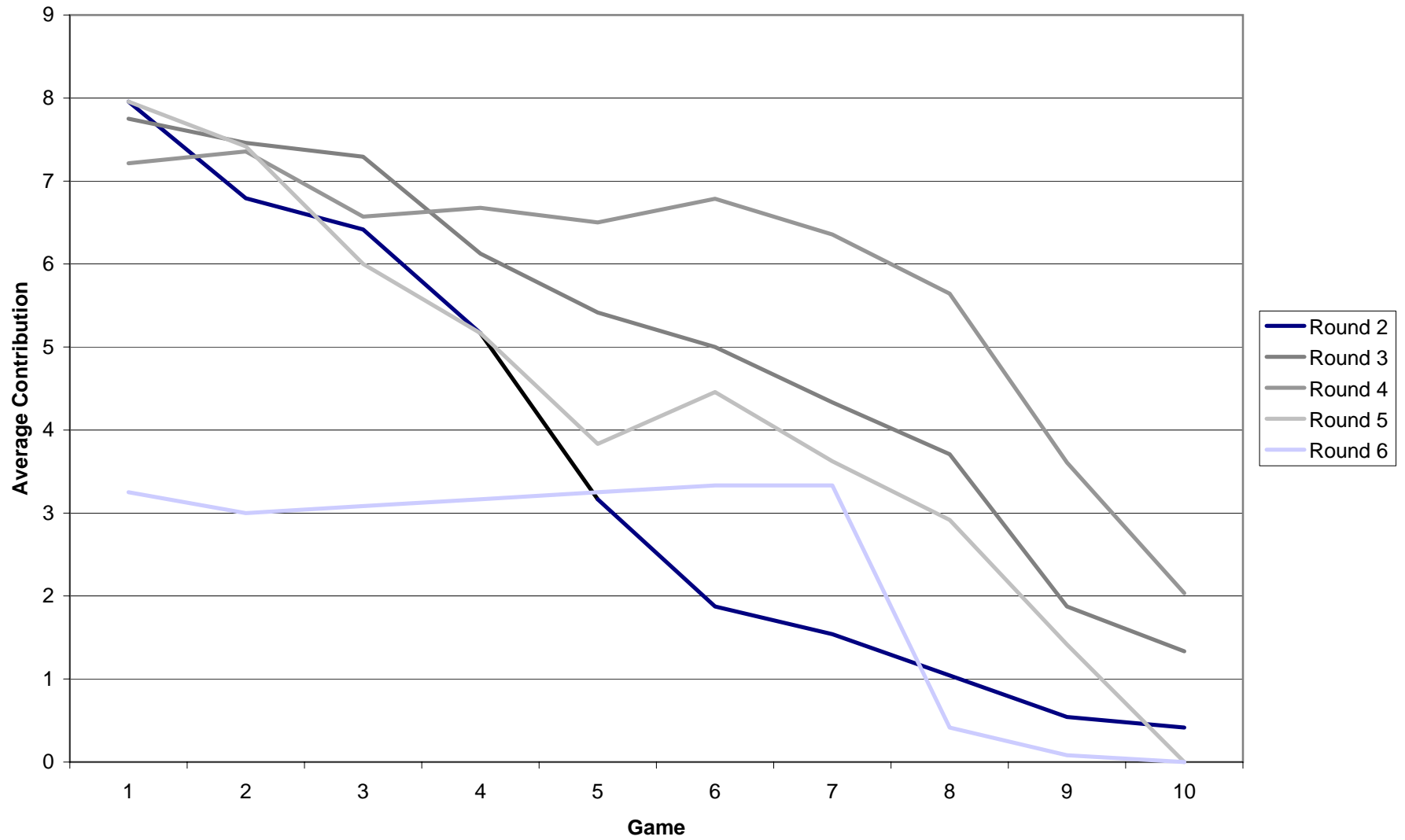


Figure 2: Average Contribution in Restarted Games by Round, Session B-IH and C-IH

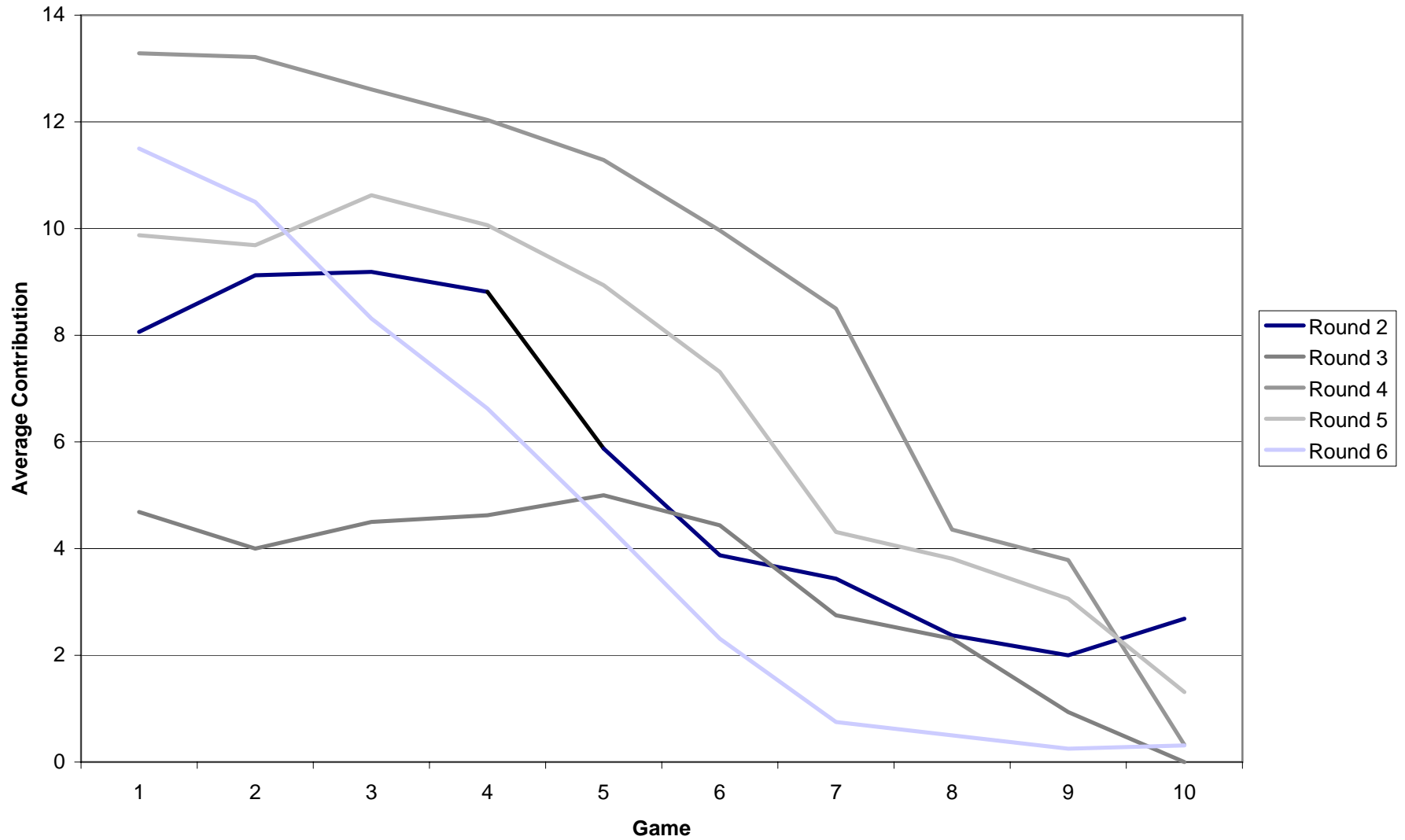


Figure 3: Average Contribution Pattern within Restarted Groups in Sessions A, B, and C

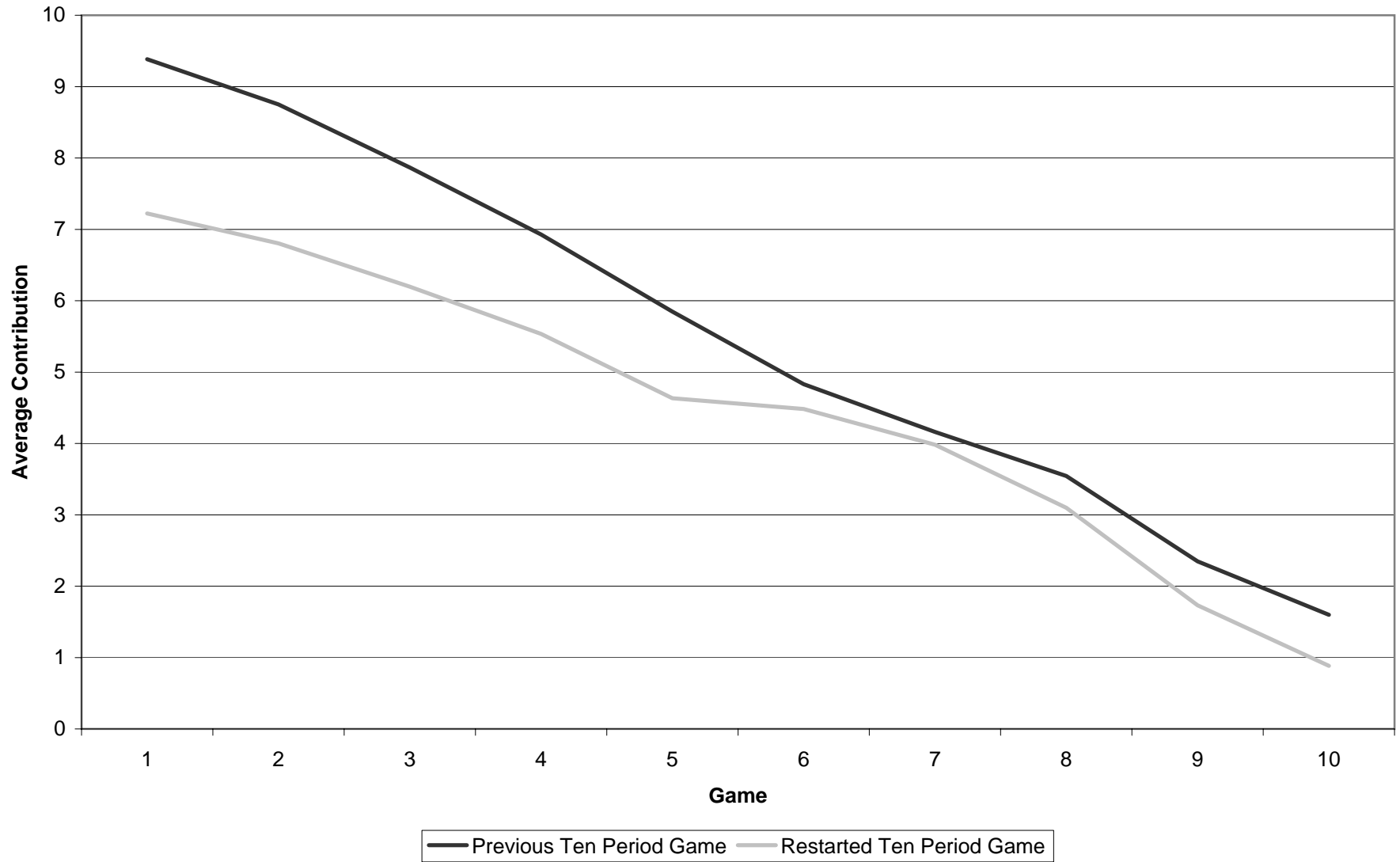


Figure 4: Average Contribution Pattern within Restarted Groups in Sessions B-IH and C-IH

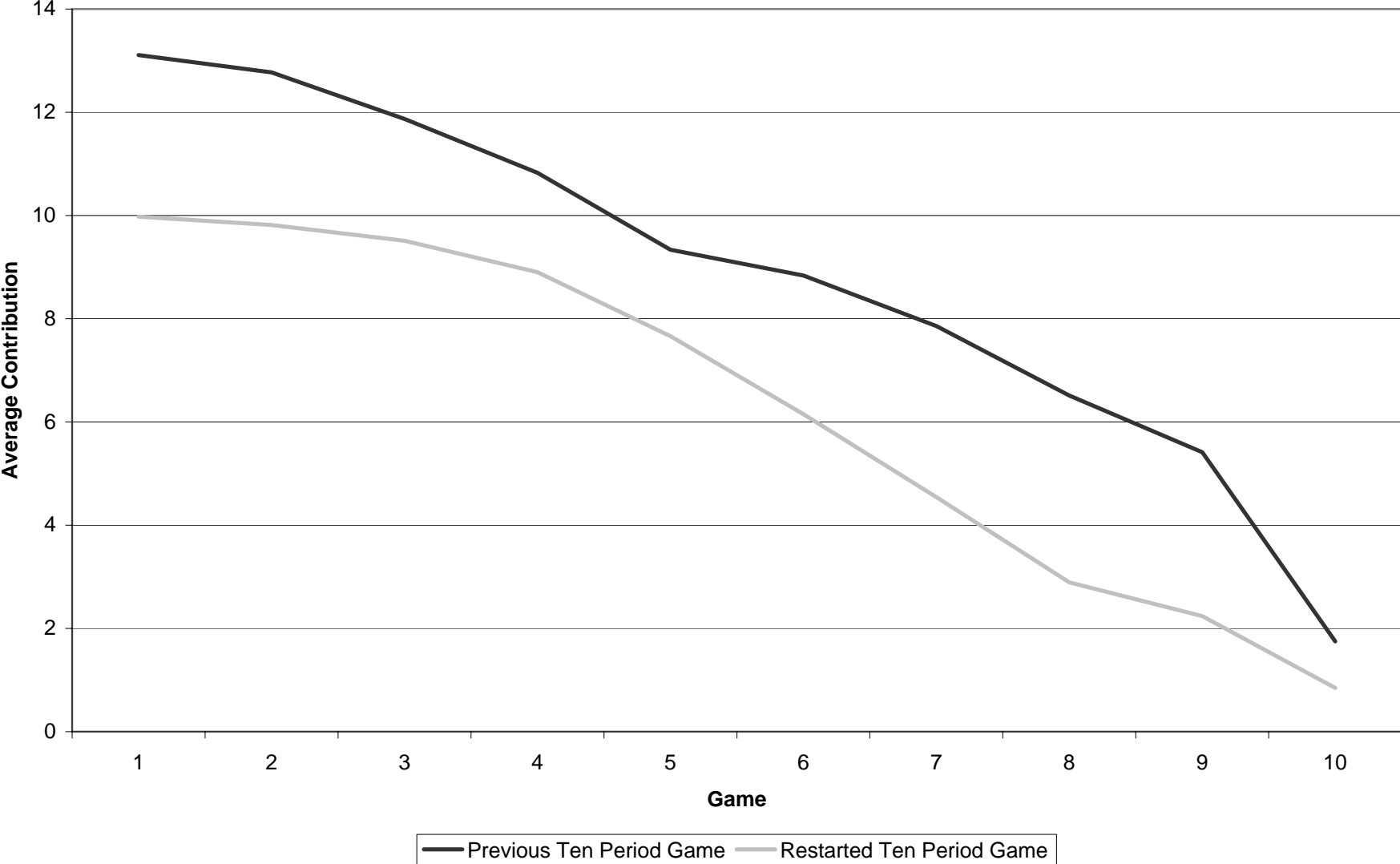


Table 1— Treatments

<u>1-period (Session O)</u> 19 1-period games, random group Surprise restart One 1-period game, same group	<u>10-period (Session A)</u> One 10-period game, random group Surprise restart One 10-period game, same group
25 1-period games, groups reshuffled with known probability 0.75 after each set	Five 10-period games, groups reshuffled with known probability 0.75 after each game
<u>10-period (Session B)</u> Six 10-period games, groups reshuffled with known probability 0.75 after each set	<u>10-period (Session B-IH)</u> Same as Session B with contributions of all group members shown after each game
<u>10-period (Session C)</u> Six 10-period games, groups reshuffled with known probability 0.75 after each set At the start of each set of games, ask for expectations	<u>10-period (Session C-IH)</u> Same as Session C with contributions of all group members shown after each game
<u>Identifying Types (Session T & T-IH)</u>	
Six gift exchange games (play as first and second proposer equal number of times) Two 10-period games, random group composition Create groups based on gift exchange play; Display group history One 10-period game Create second group based on gift exchange play; Display group history One 10-period game	
Session with average group contribution and one with contributions of all group members revealed	

Table 2— Treatment Conditions^a

Treatment	Session A	Session B	10-period games			1-period	Identifying Types	Identifying Types-IH
			Session B-IH	Session C	Session C-IH			
Number of Participants	28	36	36	32	32	28	36	36
Average Earnings	\$15.00	\$14.05	\$10.94	\$15.00	\$13.06	\$13.00	\$15.00	\$11.97
Show Up Fee	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00	\$10.00
Range	\$21-30	\$21-28	\$17-25	\$20-32	\$20-26	\$14-24	\$20-31	\$19-26
Date	10/11/05	9/30/05	3/10/06	12/8/05	3/6/06	9/30/05	12/8/05	3/6/06

^aNotes: All sessions performed at Computer Lab for Experimental Research (CLER) at the Harvard Business School. Subject were recruited among a pool of participants that include students from the Boston area. No subjects were allowed to participate in a session multiple times.

Table 3— Contribution Pattern in Restarted 10-Period Games in Multiple Restart Treatment^a

Panel A: Two Rounds Before Last Two Rounds						
Period	<u>Session A</u>		<u>Sessions B & C</u>		<u>Sessions B-IH & C-IH</u>	
	Mean	Median	Mean	Median	Mean	Median
1	5.9	4.5	7.8	7.5	10.2	10.0
2	6.0	2.5	7.8	6.5	9.9	10.0
3	5.1	2.0	7.3	6.0	9.7	10.0
4	3.3	0.5	6.3	5.0	9.3	10.0
5	1.7	0.0	5.7	5.0	9.0	5.0
6	1.6	0.0	5.3	3.5	8.0	1.5
7	0.6	0.0	4.7	0.0	6.4	1.0
8	1.6	0.0	4.0	0.0	3.6	0.0
9	0.7	0.0	2.3	0.0	2.8	0.0
10	0.7	0.0	1.4	0.0	0.2	0.0

Panel B: Last Two Rounds						
Period	<u>Session A</u>		<u>Sessions B & C</u>		<u>Sessions B-IH & C-IH</u>	
	Mean	Median	Mean	Median	Mean	Median
1	6.0	4.5	6.2	4.0	10.7	10.0
2	5.4	6.0	5.9	5.0	10.1	10.0
3	3.1	4.0	4.8	1.0	9.5	10.0
4	2.0	2.0	4.3	1.0	8.3	5.5
5	1.3	0.0	3.5	0.0	6.7	5.0
6	1.9	0.0	4.1	0.0	4.8	4.0
7	2.3	0.0	3.6	0.0	2.5	0.0
8	3.1	0.0	2.5	0.0	2.2	0.0
9	2.3	0.0	1.3	0.0	1.7	0.0
10	1.3	0.0	0.9	0.0	0.6	0.0

^aNotes: Statistics tabulated by author.

Table 4— Restart Effect in 10 Period Games Across Sessions^a

	Game	Number of Games	Mean	Median	Fraction who contribute zero
Panel A: Sessions A, B, C					
Last Two Rounds	Last Before Restart	14	1.66	0.0	77%
	First After Restart	14	7.75	7.5	29%
p-value			< 0.01		< 0.01
Two Rounds Before Last Two Rounds	Last Before Restart	10	1.63	0.0	85%
	First After Restart	10	6.20	4.0	43%
p-value			< 0.01		< 0.01
Panel B: Sessions B-IH & C-IH					
Last Two Rounds	Last Before Restart	8	1.56	0.0	88%
	First After Restart	8	10.69	10.0	13%
p-value			< 0.01		< 0.01
Two Rounds Before Last Two Rounds	Last Before Restart	11	2.57	0.0	75%
	First After Restart	11	10.16	10.0	23%
p-value			< 0.01		< 0.01

^aNotes: Statistics tabulated by author.

Table 5— Contribution Pattern in One-Shot Games^a

Periods	Type of Game	Number of Games	Mean	Median	Fraction who Contribute Zero
All (1-25)	Restarted	35	0.91	0.00	79%
1-4	Restarted	4	0.50	0.00	75%
5-9	Restarted	9	1.36	0.00	61%
10-14	Restarted	7	1.14	0.00	79%
15-19	Restarted	7	0.46	0.00	86%
20-25	Restarted	8	0.78	0.00	94%

^aNotes: Statistics tabulated by author.

Table 6— Evidence of Stabilized Play in Restarted Games^a

Periods	Number of Observations	Average in Game 1	Average in Game 5	Average in Game 10
Panel A: Sessions A, B, and C				
Two Rounds Before Last Two Rounds	56	6.20	5.66	1.41
Last Two Rounds	40	7.75	3.48	0.25
p-value Pr(F>value) ^b		0.20	0.02	0.08 0.11
Panel B: Sessions B-IH & C-IH				
Two Rounds Before Last Two Rounds	44	10.16	9.00	0.20
Last Two Rounds	32	10.69	6.72	0.81
p-value Pr(F>value)		0.84	0.47	0.68 0.49
Panel C: Sessions A, B, B-IH, C, & C-IH				
Two Rounds Before Last Two Rounds	100	8.81	7.13	0.88
Last Two Rounds	72	8.19	4.91	0.50
p-value Pr(F>value)		0.52	0.08	0.26 0.13

^aNotes: Statistics tabulated by author.

^bJoint tests that play in games 1, 5 and 10 are from the same distribution.

**Table 7— Expectations and Actual Play
in Sessions C and C-IH (Restarted)^a**

Session	Round	Number of Games	Expected Play		Actual Play		Regression ^b		
			Mean	Median	Mean	Median	β	T-stat	R^2
First Game									
C	5-6	4	6.44	5.00	6.90	6.00	0.92	8.28	0.82
C-IH	5-6	4	10.87	10.00	10.44	10.83	0.92	5.06	0.63
Fifth Game									
C	5-6	4	4.81	0.00	5.38	2.00	1.00	12.60	0.91
C-IH	5-6	4	7.06	5.00	4.82	4.17	1.05	4.12	0.53
Tenth Game									
C	5-6	4	0.94	0.00	0.00	0.00	- ^c	-	-
C-IH	5-6	4	2.19	0.00	1.63	0.33	- ^d	-	-

^aNotes: Statistics tabulated by author.

^bThis column presents the estimated coefficient, t-statistic and R^2 of the linear regression:

$$\text{expected play} = \beta \cdot \text{actual play} + \epsilon.$$

All regressions are without intercepts.

^cAll observations of average actual play are zero.

^d7 observations of average actual play are zero, while 11 observations of expectations are zero.

Table 8— Individual Behavior in Identifying Types Sessions^a

	Number of Players	Average Contribution Per Player	Last Positive Contribution Is By Reciprocal
Panel A: No Individualized Histories			
Group			
(3R,1S)	16	104	100%
(3S,1R)	16	27	50%
Within (3R, 1S)			
Selfish	4	44	-
Reciprocal	12	124	-
Panel B: Individualized Histories Shown			
(3R,1S)	16	77	100%
(3S,1R)	16	70	50%
Within (3R, 1S)			
Selfish	4	70	-
Reciprocal	12	79	-

^aNotes: Statistics tabulated by author. Group (3R,1S) means a group with 3 reciprocal players and 1 selfish player and group (1R,3S) means a group with 1 reciprocal player and 3 selfish players.

Table 8— Individual Behavior in Identifying Types Sessions (cont.)^a

	Number of Players	Average Period of Last Positive Contribution	p-value ^b
Panel C: Period of Last Positive Contribution ^c			
Selfish	32	3.53	
Reciprocal	32	4.97	0.06

^aNotes: Statistics tabulated by author.

^bFrom Wilcoxon two-sample test.

^cWhen a player never contributes a positive amount, the period of the last positive contribution is 0.

**Table 9— Impact of First Period Average Contribution^a
of Opponents on Future Play**

Dependent variable:	<u>Contribution in 2nd Game</u>			
	Pooled	Not restarted	Restarted	Restarted (Last Two Rounds)
Value of β : ^b :				
OLS	0.49 [9.85]	0.41 [6.69]	0.61 [6.53]	0.44 [2.61]
session fixed-effects	0.38 [7.27]	0.26 [4.14]	0.50 [4.87]	0.30 [1.54]
round fixed-effects	0.49 [9.63]	0.41 [6.63]	0.56 [5.86]	0.43 [2.43]
N	956	752	204	72

^aNotes: Statistics tabulated by author. T-statistics are in brackets under estimated coefficients. All 10-period games in Sessions A, B, B-IH, C and C-IH are included.

^bThe regression equation is contribution in the second game = $\beta \cdot$ avg contrib of opponents in first game + controls + ϵ_{it} . The controls in the OLS specification are an intercept, in the session fixed-effects specification they are session dummies, and in the round fixed-effects specification they are round dummies.