Testing Static Game Theory with Dynamic Experiments:

A Case Study of Public Goods

by

Anabela Botelho, Glenn W. Harrison, Lígia M. Costa Pinto & Elisabet E. Rutström †

June 2008

Working Paper 05-25, Department of Economics,
College of Business Administration, University of Central Florida, 2005

ABSTRACT

Game theory provides predictions of behavior in many one-shot games. On the other hand, most experimenters usually play repeated games with subjects, to provide experience. To avoid subjects rationally employing strategies that are appropriate for the repeated game, experimenters typically employ a "random strangers" design in which subjects are randomly paired with others in the session. There is some chance that subjects will meet in multiple rounds, but it is claimed that this chance is so small that subjects will behave as if they are in a one-shot environment. We present evidence from public goods experiments that this claim is not always true.

Corresponding author: Glenn Harrison, Department of Economics, College of Business, University of Central Florida, Orlando, FL 32816, USA. E-mail: gharrison@research.bus.ucf.edu.

JEL Classification Codes: C72, C92, H41
Keywords: Game theory, experiments, public goods

† Department of Economics, University of Minho and NIMA (Botelho and Pinto) and Department of Economics, College of Business, University of Central Florida (Harrison and Rutström). E-mail: botelho@eeg.uminho.pt, gharrison@research.bus.ucf.edu, pintol@eeg.uminho.pt, and erutstrom@bus.ucf.edu. Harrison and Rutström thank the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/HSD 0527675 and NSF/SES 0616746; Botelho and Pinto thank the Fundação para a Ciência e Tecnologia for sabbatical scholarships SFRH/BSAB/489/2005 and SFRH/BSAB/491/2005, respectively. We are grateful to Ryan Brosette, Linnéa Harrison, James Monogan and Bob Potter for research assistance, and to R. Mark Isaac and a referee for comments. All data, instructions, and statistical code is available at the *ExLab* Digital Library at http://exlab.bus.ucf.edu.

Game theory provides predictions of behavior in many one-shot games. On the other hand, when testing one-shot games many experimenters conduct dynamic sequences of multiple games with subjects, to provide experience and to collect a larger set of observations. We consider the difficulty of drawing inferences about static game theory using repeated game experiments. Theoretically, the restrictions under which the predictions of a repeated game are identical to its one-shot version are very strong, involving common knowledge of the rationality of the other player as well as the absence of psychological valuations affecting the subjective payoffs. Behaviorally, both of these assumptions are likely violated in many experimental situations designed to test the theory.

To avoid subjects rationally employing strategies that are appropriate for the repeated game, and inappropriate for the one-shot game, experimenters often employ techniques that minimize or eliminate the probability that subjects meet more than once. In a "Random Strangers" design subjects are randomly paired with others in the session. There is some chance that subjects will meet multiple times, but it is believed that this chance is so small that subjects will behave as if they are in a one-shot environment. In fact, this belief is so strong that it has been referred to as a "repeated single-shot" design (Andreoni and Croson [2008]). In a "Perfect Strangers" design a matching algorithm is used that guarantees that subjects meet only once. The polar opposite is a "Partners" design where the same subjects are pitted against each other for each round in a repeated game. We examine whether subjects perceive a Random Strangers experiment the same way that they perceive a Perfect Strangers experiment, such that the former can be used to reliably implement static games in the laboratory.

Comparisons of Partners and Random Strangers designs have been common since Andreoni [1988] reported the counter-intuitive result that the latter generates a greater amount of cooperation than the former. Most of the replications and variations of this experiment have found no significant difference in behavior, and a few even report the opposite, more intuitive, pattern. All of these studies assume that the Random Strangers design is a good proxy to a Perfect Strangers design. Nevertheless, to date there has been no systematic test of Random Strangers versus Perfect Strangers, which is surprising given the popularity of the former. We provide evidence from public goods experiments that shows that *the assumption that Random Strangers is the same as Perfect Strangers should be treated with considerable caution*. We find that the fraction of subjects that play the game strictly by the non-cooperative Nash Equilibrium prediction is significantly higher in Perfect Strangers. The difference is 13 to 30 percentage points depending on the group size.

The Random Strangers design is by far the most popular operational counterpart of a one-shot environment in experiments. N subjects are recruited into a session in which subjects are paired into groups of K<N in each round. We assume without loss of generality that N is an integer multiple of K. In the Random Strangers design the K subjects in each group are picked at random from the N subjects. This random selection is repeated in each period, and there is some chance, varying in N and K, that any one subject will see the same partner in a later period providing that the subjects are not in the last round. This chance gets small very quickly as N increases in relation to K, and on the basis of this arithmetic it is usually assumed to be an adequate procedure that ensures that subjects behave as if the chance is actually zero. Occasionally, subjects are told that these chances are very small.

The Perfect Strangers design, on the other hand, picks pairings in a way that ensures that no subject will *ever* be paired with the same person in later rounds. The only problem with the Perfect Strangers design is that it is "subject hungry." For K>2, and N around 20 or so, it becomes difficult to run sessions with more than a few rounds. But the Perfect Strangers design is the one that literally matches the one-shot notion that underlies the theories being tested in the lab.

Experimenters know all of this, and yet it is surprising to find virtually no studies that systematically compare behavior in Perfect Strangers and Random Strangers settings. The applications in which the Random Strangers has been employed are important ones, involving public goods and auctions. The implications of the ability of the Random Strangers to implement a controlled one-shot environment may therefore be quite significant. If the design does not sufficiently implement the required one-shot environment, inferences drawn based on experimental tests may be inappropriate.

We test the assumption that Random Strangers implement a one-shot environment, in the sense of Perfect Strangers, using the classic public goods voluntary contribution game. We find that *the assumption that subjects treat Random Strangers designs as if they were one-shot experiments is false.* Our subjects behave in a systematically different manner in the Perfect Strangers design. In fact, we can show that the Perfect Strangers design is associated with more subjects adopting a strict free riding behavior consistent with the one-shot theory, rather than with subjects simply providing smaller contributions conditional on making some contribution. Thus the use of the Perfect Strangers design seems to *encourage a qualitative change in the way subjects view the game*, with more of them thinking the game through in the strategic manner assumed by game theory. We also re-analyze data from two of the previous experiments that have examined Partners and Random Strangers, and show that their results are consistent with these conclusions.

1. Partners, Random Strangers, and Perfect Strangers

A. Theoretical Issues

Why do we worry about Strangers designs at all, let alone whether they are Random Strangers or Perfect Strangers? If the game has a finite and known number of repetitions, if the stage game in any single round has only one Nash Equilibrium (NE), and if it is common knowledge that all players can backward induct for the horizon of the game,[1] then the NE of the repeated game for rational players is just a "degenerate" succession of NE of the stage games and reputation plays no role. Nevertheless, if one relaxes any of these three conditions, then there may be many NE of the repeated game that differ from degenerate, successive plays of the NE of the stage game.[2] Additionally, even in finitely repeated games with unique equilibria reputation effects can be generated by relaxing the common knowledge assumption of rationality, as shown by Kreps, Milgrom, Roberts and Wilson [1982].

The public good games considered in the experimental literature are virtually identical in form to the prisoners' dilemma games considered in repeated game theory. In the standard form of both games there is invariably a single NE of the stage game. Most experiments provide subjects with a known and finite horizon. Some experiments leave the final horizon indeterminate, which can generate many of the same effects as having an infinite horizon. But the one thing we cannot easily control in experiments is the knowledge that subjects have about the other players. If the common knowledge assumption of rationality does not hold, the one-shot theory prediction is not the appropriate one to use since players may strategically want to create reputations. Thus, if the Random Strangers design introduces reputation effects, or even just perceived reputation effects, the appropriate theoretical domain is not one of static non-cooperative games.

We therefore hypothesize that behavior in Perfect Strangers will be significantly more like the

---

[1] Chess reminds us that backward induction is not an "all or nothing" thing behaviorally.

[2] For textbook expositions, see Fudenberg and Tirole [1991; ch.4,5] or Binmore [1992; ch.8]. Obviously repeated games are interesting in their own right. Our concern is with the difficulty of drawing inferences about static game theory using repeated game experiments. The task of drawing inferences about repeated game strategy choices from observed actions in repeated game experiments is actually a delicate one: see Engle-Warnick and Slonim [2006] for a discussion of the issues and a proposed methodology.

non-cooperative one-shot NE prediction than Random Strangers. In addition we predict that increasing the size of the cohort from which subjects are matched in Random Strangers (i.e. we vary N) will cause behavior to approach that of Perfect Strangers as the probability of being re-matched with the same person declines.

B. Previous Experiments

The experimental literature on public goods has a long tradition of being concerned with the strategic importance of differentiating between "Partners and Strangers." However, it is striking that virtually all of the Strangers designs have been what we call Random Strangers.

The earliest public goods experiments were conducted exclusively with a Partners design.[3] However, the presumption was that the game would be viewed by subjects as a finite-horizon repeated game in which the sole NE was the same outcome as the NE of the stage game. For example, Isaac and Walker [1988; p.195] state this position clearly:

> The results across all periods are not supportive of the multi-period Nash equilibria prediction of zero contribution in every period (based upon a backwards induction argument). Instead, the experiments uniformly begin with positive contributions [...] followed by a tendency for contributions to decay. This decay pattern is consistent with the experimental results cited by Kreps et al. [1982], and it suggests that the incomplete information models should also be a fruitful line of theoretical inquiry for public goods research.

Of course, an alternative is to consider the effects of experimental designs that mitigate the role of reputation effects under incomplete information.

Andreoni [1988] initiated this approach in the experimental literature, explicitly contrasting what he termed Partners and Strangers. The Strangers design in his experiments were Random Strangers: 20 subjects were randomly assigned to 4 groups of 5 in each of 10 rounds. He reports that

[3] The same is true of the extensive early experimental literature on first-price sealed-bid auctions.

Random Strangers contributed more, on average, than Partners. This counter-intuitive result generated a flurry of interest, as discussed below. At an "eyeball" level the average contributions in each treatment are compared, round by round (Table 1, p.296). Partners contribute an average of 16.6 tokens over 10 rounds, and Random Strangers contribute an average of 20.7 tokens, for a difference of 4.1 tokens. The problem, noted by Croson [1996; p.30] and Palfrey and Prisbey [1996; p.413], is that there is a significant standard deviation in contributions, around 16 tokens per round in each treatment. The test used by Andreoni [1988; p.296, fn. 9] is a median test applied to all individual contributions over all rounds. The null hypothesis here is that the Partners and Random Strangers samples arise from populations with the same median, and he reports that this null can be rejected with a *p*-value of less than 0.01. Of course, this test assumes that the samples are random (Conover [1980; p.171]), and this is violated by the temporal dependence between rounds and subjects. We examine this hypothesis later, using an econometric specification which accounts for some of the features of these data and that is comparable to the analysis of our own data. To anticipate, we find that there is no statistically significant evidence of differences between Partners and Random Strangers in the data from Andreoni [1988].

Weimann [1994] undertook a replication of the Andreoni [1988] conclusion, but his experiments also changed the design in a way that makes them hard to compare. In his Strangers experiments the subjects were contacted by phone, after receiving instructions and a record form in the mail, rather than in some common setting. This has the advantage of complete anonymity, of course. Unfortunately, two of the variations on the baseline experiments employed deception, clouding the credibility of inferences. In any event, Weimann [1994] concludes that he did not replicate the conclusion of Andreoni [1988].

Croson [1996] replicated the Andreoni [1988] design and also found different results.

Although she does not report the average contributions, inspection of her Figure 1 (p.28) indicates that contributions in the Partners treatments were roughly 4 to 5 tokens *higher* than those in the Random Strangers treatments, relative to an endowment of 25 tokens in each round. But the same concern with the variation of individual contributions arises. She reports (Table 2, p.30) standard deviations in each treatment around 8 tokens in each round. Using a Wilcoxon test, despite its assumption that observations are independent, she rejects the hypothesis that the contributions are the same, in favor of higher contributions in the Partners treatment. We also re-consider these results later, using an econometric specification to help us identify the sources of these differences, and verify these conclusions.

Palfrey and Prisbey [1996] conduct an experiment that compares Partners and Random Strangers, along with other treatments. Their subjects participated in 40-round games, broken into 4 treatments. In each round the subject received a random "exchange rate" that would convert their tokens into points. Each subject received a different exchange rate each round, and the subjects in the same group of 4 received different exchange rates. These exchange rates were drawn uniformly at random as integers between 1 and 20. In the first 20 rounds each subject received a fixed group return, and then a new, fixed group return in the last 20 rounds. These returns were "low" and then "high." In each 10-round sequence each subject received a random private return. Subjects were either in a Partners treatment for the entire 40 rounds or in a Random Strangers treatment for the entire 40 rounds. They find average contributions of 3.46 tokens in the Partners treatment and 3.71 in the Random Strangers treatment, out of endowments of 9 tokens. The standard deviation of each is 3.68 tokens and 3.60 tokens. One cannot reject the null hypothesis of identical mean contribution levels using a *t*-test (*p*=0.13),[4] nor can one reject the null hypothesis of identical variances in contributions using an F-test (*p*=0.51).[5] These results are even stronger if one only considers the responses in the first round, although sample sizes become very small since there were only 24 subjects in each of the main treatments.

Burlando and Hey [1997] also fail to replicate the conclusion claimed by Andreoni [1988]. They find no significant difference between Partners and Random Strangers overall, although there are some minor interaction effects depending on the national location of the experiments and the sequencing.[6]

Keser and van Winden [2000] provide convincing evidence that Partners contribute more than Random Strangers in stationary public good experiments. A key feature of their design was simply to increase the number of replications in each treatment, so that they had 6 sessions with the Random Strangers treatment and 10 sessions with the Partners treatment, spanning 160 subjects.

[4] Andreoni and Croson [2008; Table 1] and Keser and van Winden [2000; p. 24] claim that these data show that Strangers contribute more than Partners, but this may just be due to them being willing to accept a *p*-value this high.

[5] Keser and van Winden [2000; p. 24] claim that these data show that Strangers have a higher variance in *contributions* than Partners, but this is likely due to a misreading of a claim by Palfrey and Prisbey [1996; p.424] about the dispersion in the fitted *parameter* of a specific model estimated from these treatments.

[6] The classification of these results in Andreoni and Croson [2008; Table 1] does not match the conclusions of the original study. They classify the British subjects as contributing more on the Strangers design compared to the Partners design, whereas the original study finds no difference; they classify the Italian subjects as contributing more in the Partners design compared to the Strangers design, but this is due to some differences following a restart, rather than in initial rounds of behavior. Burlando and Hey [1997; p.53-4] note that "... for the UK subjects the partners percentage was 86.64 as compared to 85.65 for the strangers – a difference that is not statistically significant (*p*=0.2073); for the Italians, the partners percentage was 70.62 as compared with 73.38 for the strangers – a difference that is significant according to a [Wilcoxon rank-sum test] at 1% (*p*=0.0051). Interestingly, this difference is largely driven by the difference in behavior between the first and second sub-sessions among the Italian subjects – in the second sub-session partners free-rode much less than strangers [...]. Perhaps by then they had learned that co-operation was a good thing?" The percentages they refer to are the percentages of the *bad* that was placed in the public domain. So "dumping" *more* in the public domain here amounts to free riding *more* and contributing *less*. Note that what may be interpreted as differences across nations may be attributable to differences in individual characteristics rather than nationality, as noted by Botelho, Harrison, Hirsch and Rutström [2005] in the context of cross-national bargaining experiments. Unfortunately, raw data on individual characteristics was not collected for the British subjects (Burlando and Hey [1997; p.47, fn.10]).

They found that average contributions were 1.9 tokens and 4.53 tokens, out of an endowment of 10, across the two treatments.

Brandts and Schram [2001] report a comparison of Partners and Random Strangers, using a modified design in which subjects submitted a contribution function in each round instead of a single contribution. This function was defined over 10 states of nature in which there were different rates of return on the public contribution, and subjects knew that one state of nature would actually apply in that round and be selected at random. They find no difference between Partners and Random Strangers, and note (p. 408) that the "... lack of difference in behavior between partners and strangers is not consistent with the original evidence of Andreoni [1988], where strangers contribute more than partners. His result has not always been replicated, however."

Andreoni and Croson [2008] review the literature on public goods contributions with Partners and Random Strangers. They discuss additional studies examining these treatments, but in which there was some other design change.

Fehr and Gächter [2000; fn.3] report the only evidence we know of comparing Random Strangers and Perfect Strangers in a public goods experiment. They note briefly that the results of a Perfect Strangers replication of their design generated essentially the same results as their Random Strangers experiments. However, they only considered one sequence of regimes (Punishment followed by Non-Punishment), and did not maintain the Perfect Strangers treatment after the first regime of 6 periods. In other words, subjects that were matched only once in rounds 1-6 might have been matched again in rounds 7-12, thereby reducing the Perfect Strangers control. Moreover, for the Non-Punishment game that participants always played followed the Punishment game, one would have to control for the history generated by the first game, as we do below, to be able to draw any inferences about the effects of Perfect Strangers rather than Random Strangers.

Generally, the comparisons of behavior in Partners and Random Strangers in this literature has led to interpretations that involve preferences with social arguments such as altruism or "warm glow." We introduce the alternative hypothesis that some subjects did not perceive the Random Strangers experiments as one-shot game environments, but strategically based their decisions on the presence of reputation effects.

2. Experimental Design

Each subject participated in an experimental session in which there were 10 rounds of a traditional voluntary contribution public goods game. We use data from two experiments. In our main experiment subjects participate in groups of 4 in each round.[7] To test for robustness of our findings we also analyze data from another experiment where subjects participate in groups of 2 in each round. We explain to subjects how we ensure that there is no chance that they will meet the same person in any other round. We describe the design of the main experiment, and then discuss differences in the second experiment.

A. Main Experiment

Table 1 summarizes the experimental design. We recruited 172 subjects at the University of Central Florida in 2008 who were randomly assigned to one of the 5 sessions shown in Panel A of Table 1. Two sessions consisted of 60 subjects and 56 subjects, respectively, so that we could effect a Perfect Strangers design for 10 periods and have 4 subjects per game. These two sessions were each conducted in one large room. The other three sessions were Random Strangers designs in which we

[7] Most public goods experiments use 4 subjects per group, although the effect of larger group sizes has been studied by Isaac and Walker [1988] and others. Harrison and Hirshleifer [1989] and Goeree, Holt and Laury [2002] employed groups of 2 in their public goods experiments.

varied the number of subjects in the session from 28 down to 16 and 12. We hypothesize that increasing the size of the cohort in Random Strangers will reduce the perceived reputation effect and make behavior more similar to Perfect Strangers.

Each subject received an endowment of 20 tokens at the outset of each round, and each token was worth 5 cents. Each subject was told the aggregate level of contributions at the end of each round. The return to investments in the public good was 0.8. We used a linear payoff schedule which was constant for all contributions, so the dominant strategy is simple: a subject that only seeks to maximize individual earnings in a single period should contribute nothing to the public good.[8] The average subject earned $28 in these experiments, including a standard $5 show-up fee. No session lasted more than 2 hours, and most were at least 1½ hours in length.

Subjects were randomly assigned to each session, with no prior knowledge of the parameters or treatments. Instructions were provided in written form and orally, and the experiment was implemented using version 2.1.4 of the *z-Tree* software developed by Fischbacher [1999].[9] The oral instructions also utilized a large-screen display that could be easily seen by all subjects, to ensure that certain information was common knowledge. Training rounds were included prior to each regime, to ensure that subjects understood the task.

B. Second Experiment

In order to investigate the robustness of our findings in the main experiment we also analyze data from an experiment with a reduced group size of 2. There is some evidence that behavior in the

[8] Alternative assumptions about the factors motivating subjects to contribute in public goods experiments have long been studied. See, in particular, Palfrey and Prisbrey [1996][1997], Goeree, Holt and Laury [2002] or Keser and van Winden [2000].

[9] All instructions, scripts, and software are available at http://exlab.bus.ucf.edu. The latest version of the *z-Tree* software and documentation is available at http://www.iew.unizh.ch/ztree/index.php.

N=2 case can be qualitatively different than the N=4 case in closely related experimental games (Huck, Norman and Oechssler [2004]). In all respects the sessions were conducted using the same procedures as in the main experiment.

Thirteen sessions were conducted in this experiment shown in panel B of Table 1. The first 4 used Perfect Strangers designs, and the last 9 used Random Strangers designs. We varied the size of the cohort in the Random Strangers design from 6 up to 16 participants, although we kept the group size at 2 in every cohort. Subjects were recruited from the University of Central Florida, and 180 participated in this experiment in 2005. Average earnings were $39 including the $5 showup fee.

In this experiment there was also some other task either after the initial 10 rounds, or in a prior 10 rounds. This task was a "sanctions" public goods game in the spirit of Fehr and Gächter [2000]. We will refer to the case where the sanctions game is played as a "punishment game" and the case where there are no sanctions as a "no-punishment game" and use the labels "p_np" and "np_p" to indicate the game order for each session. The results of those tasks are not of interest here, but we control for the order of the tasks in the statistical analysis.[10] In the treatments where it was used, the Perfect Strangers design was maintained across both tasks: that is, no subject ever encountered the same subject twice in a given session.

In two sessions we used a relatively low return on contributions to the public good, and in all other sessions we used a relatively high return. The low return was 0.6 of token: hence every token contributed to the public good by one subject would decrease their private endowment by 1 token and return 0.6 of a token for herself. Of course, it would also generate 0.6 of a token for the other player, so the social return was 1.2 tokens for every 1 token invested. In the high return treatment we

---

[10] The substantive results from that task are evaluated in Botelho, Harrison, Pinto and Rutström [2005].

changed the public good return from 0.6 to 0.8, thereby increasing the social return from 20% to 60%. The objective of this treatment was to see the effects of making the environment more rewarding to any strategy that would increase contributions to the public good. In terms of the marginal per capita return (MPCR) to contributing, which is just the ratio of the return from the public good contribution to the return from the implicit private good contribution, these are 0.60 and 0.80, respectively.[11]

3. Results

A. Main Experiment

One of the advantages of the simple design of the experiment is that, with some caveats, we can just look at the raw data to see the effect of the treatment. Table 2 collates the raw frequency for each possible contribution level, pooling over all 10 periods. The Perfect Strangers design clearly increases the fraction of responses in which zero is contributed, and does not appear to have any systematic effect on other levels of contribution. We easily reject the null hypothesis of no association between treatment and contribution level, using a Pearson χ^2 test with a *p*-value less than 0.001.

We confirm these findings using a structural econometric estimation approach. Our analysis employs a likelihood function that is constructed to be appropriate for this type of experiment. Theory tells us that there may be some individuals that gravitate to one particular contribution level: contribute zero. The raw data also flag this as a "spike" that needs to be addressed explicitly, resulting in an evident mode at zero, even for responses in the first period. Our statistical analysis therefore considers the process by which some subject decides to contribute zero or some positive amount as

[11] Isaac and Walker [1988] carefully discuss the relationship between changes in group size and the implied MPCR. They use MPCR values of 0.3 and 0.75, and refer to the latter as high. So our values tend to be "high" in relation to the ones they consider.

separate from the process by which the subject decides how much to contribute.

The natural specification to capture this intuition from theory and the raw data is a "hurdle model." This specification is common in health economics, for example, where it is used to capture the idea that the factors that cause someone to seek medical care are distinct from the factors that cause the doctor and patient to decide how much to spend.[12] In this case going to the doctor is the hurdle that must be passed before expenditures would be positive. In our case the subject has to decide whether to contribute any amount at all, and only then does the process determining the positive contribution level apply.

The likelihood function for the overall hurdle model is constructed as the product of two likelihoods.[13] The first component is the likelihood that the subject contributed zero or not, and uses a standard probit specification defined over an index function $x_i\alpha$, where α is a parameter vector to be estimated and x_i is a vector of explanatory variables for observation i. The second component is the *conditional* likelihood that the subject contributed a certain fraction of the endowment. This likelihood function is constructed using the specification developed by Papke and Wooldridge [1996] for fractional dependant variables, since the dependant variable in this case is the fraction of the endowment contributed (conditional on any positive contribution).[14] Thus the log-likelihood of observation i is defined as $l_i(\beta) = C_i \times \log[G(x_i\beta)] + (1-C_i) \times \log[1-G(x_i\beta)]$ for contribution fraction C_i, parameter vector β, and some convenient cumulative distribution function $G(\cdot)$. We use the standard normal cumulative distribution function $G(z) = \Phi(z)$. Thus the overall likelihood function for the hurdle model requires the estimation of α and β. Since our data is a panel we use a

[12] See Coller, Harrison and McInnes [2002] for an application.

[13] As is well known, one can alternatively estimate the two parts of the hurdle model separately and obtain consistent and efficient estimates (McDowell [2003]).

[14] In order to compare the estimation results based on our data to that of Andreoni [1988] and Croson [1996] we use fractions, since the initial endowments used across these three studies differ.

"clustering" specification that treats observations as independent across subjects but not within.[15]

Explanatory variables include individual demographics and treatment effects. In addition to a binary dummy variable for the Perfect Strangers designs (Pstrangers), we also include dummy variables for the size of the cohort conditional on the use of a Random Strangers design (Csize).[16] If differences in behavior between Perfect Strangers and Random Strangers are due to perceived reputation effects, we expect a smaller cohort in Random Strangers to be correlated with a stronger difference in behavior from Perfect Strangers. Demographics include a measure of age in years (Age), binary indicators for sex (Male), race (Black, Asian, Hispanic or Other Race), academic major (Business), class standing (PreSenior), cumulative GPA below 3¼ (GPAlow), cumulative GP above 3¾ (GPAhigh), number of people in the subject's household (Hhsize), and a binary indicator of those that work part-time or full-time (Work). Table 3 lists descriptive statistics for these variables. Finally, we include binary dummy variables for the sequence of periods within each treatment, to correct for any temporal patterns while remaining agnostic about appropriate learning models.

Table 4 provides maximum likelihood estimates of the hurdle model for these data. All estimates for the α parameter represent the calculated marginal effect of that variable on the probability of contributing. The reported estimates for the β are the marginal effects in terms of the

[15] The use of clustering to allow for "panel effects" from unobserved individual effects is common in the statistical survey literature. Clustering commonly arises in national field surveys from the fact that physically proximate households are often sampled to save time and money, but it can also arise from more homely sampling procedures. The procedures for allowing for clustering allow heteroskedasticity between and within clusters, as well as autocorrelation within clusters. Wooldridge [2003] reviews some issues in the use of clustering for panel effects, in particular noting that significant inferential problems may arise with small numbers of panels. Under our null hypothesis, that there is no difference between Perfect Strangers and Random Strangers, we can assume independent responses across subjects in the same session. If that hypothesis is rejected, and one wants to further model behavior within Random Strangers sessions, then some correction for clustering within session should be included.

[16] This variable takes on the value 0 for the Perfect Strangers treatment and the size of the cohort (from Table 1) for the Random Strangers treatments. Thus it can be viewed as an interaction between the Perfect Strangers treatment and cohort size.

positive fraction of tokens contributed.

The focus variable is the Pstrangers binary dummy since it measure the marginal effect of switching from a Random Strangers to a Perfect Strangers design. It clearly has a statistically significant effect on the decision to contribute something or nothing, and no impact on the level of positive contributions. This is striking evidence that the *Perfect Strangers treatment affects qualitative behavior, in the sense that it elicits more subjects to focus on the zero contribution response*. Subjects are on average 13 percentage points more likely to be free riders in the Perfect Strangers treatments, and this effect is statistically significant (*p*-value = 0.025). This is clear evidence that a Random Strangers environment does not elicit the same behavior as a comparable Perfect Strangers environment, and that the direction of the change in behavior is consistent with the Perfect Strangers environment being more conducive to subjects viewing the stage game as one-shot. The marginal effect of the treatment on the level of contributions, conditional on there being any, is small and statistically insignificant (*p*-value = 0.36).

There is no effect on either the propensity to free ride or on the amounts contributed when increasing the size of the cohort in the Random Strangers design (Csize). In addition, the propensity to free ride is reasonably stable over the periods but conditional contributions increase slightly during early rounds.

These estimates control for individual demographics and period effects, but essentially the same results obtain if we do not control for these factors.[17] There are some clear demographic effects on contributions. It is interesting to note that men free ride more than women, but are actually more generous once they decide to contribute.

[17] Since sessions A and B are each large, there is a risk for session effects; we checked for this by estimating just using session B responses, and found the same results.

In these experiments the effect of increasing the size of the cohort from which subjects are matched in the Random Strangers treatment is to lower the probability of contributing by only 0.17 of a percentage point for every extra member of the cohort. Moreover, this effect is not statistically significantly different from zero (*p*-value = 0.64). Although we can formally calculate the cohort size in a Random Strangers design that results in the equivalent behavior observed in a Perfect Strangers design of 56 or 60 subjects, and it is 96 subjects, the reliability of this calculation is too low to provide a guide to practice.

B. Second Experiment

Our analysis of the second experiment with group size 2 confirms these findings, although the effect on free riding from the Perfect Strangers design is even stronger. Table 5 shows the maximum likelihood estimates of the hurdle model for this experiment. We include a dummy for the low reward sessions in the Perfect Strangers design (Low) to capture the marginal effect of further lowering incentives to contribute. Additionally, we include a dummy for the order of the game relative to the sanctions game during periods 1-10 (np_p).[18]

The treatment effect from Perfect Strangers is even stronger here – a reduction in free riding behavior by 30 percentage points, and we also see that in the Random Strangers design contributions fall as the cohort size increases and the probability or being rematched with another player drops. The effect of increasing the size of the cohort from which subjects are matched is to lower the probability of contributing by 2.4 percentage points for every extra member of the cohort. The difference between Random Strangers and Perfect Strangers is therefore diminishing in the size of

[18] We do not include the data from the sanctions game in our analysis here, but subjects who participated in the sanctions game first may have been affected by that experience so it is important to control for that experience.

the Random Strangers cohort.[19] Based on the estimated model we predict the cohort size in the Random Strangers treatment that would result in the same propensity to contribute as that observed in the Perfect Strangers treatment. We find this cohort to be of size 14, i.e. 7 groups of 2 members each.

These estimates control for individual demographics and period effects, but essentially the same results obtain if we do not control for these factors. There are some clear demographic effects on contributions, particularly in terms of the fraction of contributions conditional on making any contribution. We confirm the finding from the main experiment that men are actually more generous once they decide to contribute, even though there is an offsetting (insignificant) effect on the decision to contribute.

We also find that the ordering of the experiments, as first or second in the game sequence, also makes a large difference to contributions. When the experiment comes before the other task (variable np_p equals 1), the probability of being a free rider is 0.25 *lower* on average. [20]

Decreases in the reward for contributing to the public good in the Perfect Strangers design are associated with significant increases in free riding, and significant decreases in the amount contributed when someone does contribute something. This is consistent with previous observations in public goods experiments, as is the fact that the passage of time increases the likelihood of a subject becoming a free rider. The marginal effects of the period dummies on α in Table 5 show a

[19] This effect is only weakly significant since the estimated standard error is 1.5 percentage points, and the *p*-value is 0.12.

[20] Perhaps this is related to the fact that subjects in the Random Strangers experiment may underestimate the total number of rounds during which they will be randomly rematched within their cohort, thus leading to a lower perceived reputation effect than when they have already played 10 rounds. This is a possibility since in the np_p treatment the experiment comes before the other task and subjects are not made explicitly aware of an additional 10 rounds in the second task, although they know of the second task itself. In the alternative treatment the experiment comes after the other task, so subjects know for sure that they are matched up with people from their cohort over a total of 20 periods.

significant and steady decline after period 3.[21]

The period effects are more significant in these data and we see an increase in free riding behavior over time, although no effect in conditional contributions. We separately tested for any interaction between this temporal effect and the matching protocol and found none. The increase in free riding over time is therefore common to both the Perfect Strangers and Random Strangers designs.

4. Comparisons to the Previous Experiments

Our econometric model of contributions allows us to re-examine data from the previous experiments of Andreoni [1988] and Croson [1996] using statistical methods that are comparable to our own. Using data on individual contributions,[22] we estimate the same hurdle model with controls for the key Partners versus Random Strangers treatment.

Our estimated model based on the data from Andreoni [1988] finds no significant difference between Partners and Random Strangers. There is generally no difference in terms of whether subjects decide to contribute anything at all, or in terms of what level of contribution they would make if positive. We do find statistically significant round effects, but these are common to the Partners and Random Strangers treatments.[23]

The data from the experiments of Croson [1996] tell an even stronger story. Estimating the

[21] The size of the marginal effect for each period appears to be too large, but is appropriate given that each dummy has an average sample value of 0.1, and the effects are each measured relative to period 1.

[22] Generously provided by James Andreoni and Rachel Croson.

[23] The only effect that we observe is a fascinating one in terms of the underlying static game theory: an end-period effect. Andreoni [1988; p.295] explains that "... we expect that giving by Partners will be greater than giving by [Random] Strangers, especially early in the game (before the Partners begin to 'bail out'). In the tenth round, however, both Partners and Strangers are playing an end-game, hence both are predicted to free ride." However, in the last round we find that Random Strangers are 17 percentage points *more* likely to contribute some amount, and this effect is statistically significant (p-value = 0.028). This end-game effect is not sufficient overall to offset the conclusion that Random Strangers and Partners are behaving similarly in these experiments.

same statistical model with her data, over all 20 rounds, we find a large effect from Random Strangers on the propensity to free ride, compared to Partners, and no effect at all on the level of contributions conditional on making any. Random Strangers are 42.5 percentage points less likely overall to make any contribution, and this is a significant effect (*p*-value = 0.004). They are estimated to contribute 11.1 percentage points more conditional on making any contribution at all, but this is not statistically different from zero (*p*-value = 0.31).

Thus, these statistical results show a qualitative effect on behavior such that Random Strangers elicits more subjects to focus on the zero contribution response, but with no effect on the conditional contributions.

5. Conclusions

We find a significant effect on the propensity to free ride in public goods settings from the use of an experimental design that ensures a zero probability of re-encounters between subjects, the Perfect Strangers design, as compared to the more commonly used Random Strangers.

We also observe some reductions in contributions, conditional on not free riding, but only in the data from the experiments with group size 2. This effect depends on the size of the cohort with contributions decreasing as the probability of re-encounters decreases. This finding is consistent with participants perceiving the presence of reputation effects in smaller cohorts. We find no such effect with a group size of 4, however, leading us to conclude that Random Strangers may not approximate Perfect Strangers even when cohorts are large.

It appears, based on our data and that of Croson [1996], that the fraction of subjects that free ride is larger in Random Strangers than in Partners, but even larger in Perfect Strangers. In addition, this fraction sometimes increases as the cohort size increases in the Random Strangers design. We conjecture that Perfect Strangers designs will have comparable effects in other strategic experimental tasks in which there may be effects from the subjects behaving as if in a repeated game.

The implications for testing game theory are significant. Testing one-shot theories may require substantial increases in cohort sizes, at least if subjects sufficiently care about being rematched with at least one other group member. Under some circumstances increasing the cohort size may not be sufficient to allow a Random Strangers design to approximate a Perfect Strangers one. The language used to explain the matching protocol might also be usefully examined in light of our findings: our instructions used what we view as standard text, but we know that seemingly small changes in such things can, on occasion, generate large behavioral effects. At the very least, the assumption that Random Strangers designs sufficiently controls for the influence of reputation effects must be treated with considerable caution.

Table 1: Experimental Design

A. Main Experiment

Each experiment had 10 rounds of the "no punishment" regime, with groups of 4 per game

Session	Return to Public Good	Anonymity	N in Session	History
A	High	Perfect	60	NP
B	High	Perfect	56	NP
C	High	Random	28	NP
D	High	Random	16	NP
E	High	Random	12	NP

B. Second Experiment

Each experiment had 10 rounds of one regime (punishment or no punishment), followed by 10 rounds of the other regime and groups of 2 per game

Session	Return to Public Good	Anonymity	N in Session	History
F	Low	Perfect	26	NP-P
G	Low	Perfect	24	P-NP
H	High	Perfect	26	NP-P
I	High	Perfect	26	P-NP
J	High	Random	10	P-NP
K	High	Random	16	P-NP
L	High	Random	8	P-NP
M	High	Random	6	P-NP
N	High	Random	8	NP-P
O	High	Random	6	NP-P
P	High	Random	6	NP-P
Q	High	Random	8	NP-P
R	High	Random	10	NP-P

Table 2: Raw Tabulation of Results in Main Experiment

Contribution	Frequency			Column Percentages		
	RS*	PS**	Total	RS*	PS**	Total
0	31	214	245	5.5	18.5	14.2
1	11	12	23	1.9	1.0	1.3
2	19	22	41	3.4	1.9	2.4
3	7	14	21	1.3	1.2	1.2
4	21	28	49	3.7	2.4	2.8
5	11	57	68	2.0	4.9	4.0
6	31	15	46	5.5	1.3	2.7
7	6	22	28	1.1	1.9	1.6
8	41	38	79	7.3	3.3	4.6
9	0	9	9	0	0.8	0.5
10	58	94	152	10.4	8.1	8.8
11	4	6	10	0.7	0.5	0.6
12	20	24	44	3.6	2.1	2.6
13	2	15	17	0.4	1.3	1.0
14	14	12	26	2.5	1.0	1.5
15	48	81	129	8.6	7.0	7.5
16	24	34	58	4.3	2.9	3.4
17	9	14	23	1.6	1.2	1.3
18	33	51	84	5.9	4.4	4.9
19	7	18	25	1.3	1.55	1.5
20	163	380	543	29.1	32.8	31.6
Total	560	1160	1720	100	100	100

*   RS indicates the Random Strangers treatment
** PS indicates the Perfect Strangers treatment

Table 3: Descriptive Statistics for Explanatory Variables

Variable	Description	N	Mean	Std. Dev.	Minimum	Maximum
			A. Main Experiment			
Pstrangers	Perfect strangers	172	0.67	0.47	0	1
Csize	Cohort size for RS*	172	6.88	10.71	0	16
Age	Age	172	21.71	3.78	18	49
Male	Male	172	0.60	0.49	0	1
Black	Black	172	0.08	0.27	0	1
Asian	Asian	172	0.08	0.27	0	1
Hispanic	Hispanic	172	0.14	0.35	0	1
Business	Business major	172	0.53	0.50	0	1
PreSenior	Pre-senior	172	0.63	0.48	0	1
GPAlow	Low GPA	172	0.50	0.50	0	1
GPAhigh	High GPA	172	0.10	0.30	0	1
HHsize	Size of household	172	1.63	1.25	1	7
Work	Any work	172	0.62	0.49	0	1
			B. Second Experiment			
Pstrangers	Perfect strangers	180	0.57	0.50	0	1
Csize	Cohort size for RS*	180	4.20	5.34	0	16
np_p	Game Ordering	180	0.50	0.50	0	1
Low	Low reward for contributions	180	0.28	0.45	0	1
Age	Age	180	21.51	2.65	18	36
Male	Male	180	0.64	0.48	0	1
Black	Black	180	0.08	0.28	0	1
Asian	Asian	180	0.08	0.28	0	1
Hispanic	Hispanic	180	0.13	0.33	0	1
White	White	180	0.66	0.48	0	1
OtherRace	Other Race	180	0.04	0.21	0	1
Business	Business major	180	0.43	0.50	0	1
PreSenior	Pre-senior	180	0.47	0.50	0	1
GPAlow	Low GPA	180	0.47	0.50	0	1
GPAhigh	High GPA	180	0.15	0.36	0	1
HHsize	Size of household	180	1.65	1.26	1	7
Work	Any work	180	0.72	0.45	0	1

* RS indicates the Random Strangers Treatment

Table 4: Maximum Likelihood Estimates of the Hurdle Model of Contributions for the Main Experiment

Marginal effects for α and β parameters

N=1720 responses from 172 subjects; Wald test of H_0: $\alpha=\beta=0$ has $\chi^2_{22} = 87.5$ (p-value<0.001)

Parameter	Variable	Description	Estimate	SE	*p*-value	95% Confidence Intervals	
α	Pstrangers	Perfect strangers	-0.132	0.059	0.025	-0.246	-0.017
	Csize	Cohort size for Random strangers	-0.002	0.004	0.639	-0.009	0.005
	Age	Age	0.021	0.007	0.002	0.008	0.034
	Male	Male	-0.107	0.028	0.000	-0.163	-0.051
	Black	Black	0.068	0.033	0.037	0.004	0.132
	Asian	Asian	0.034	0.042	0.415	-0.048	0.117
	Hispanic	Hispanic	0.039	0.033	0.241	-0.026	0.104
	Business	Business major	-0.051	0.029	0.078	-0.108	0.006
	PreSenior	Pre-senior	0.101	0.043	0.019	0.017	0.186
	GPAlow	Low GPA	0.019	0.031	0.538	-0.042	0.081
	GPAhigh	High GPA	-0.029	0.064	0.650	-0.154	0.096
	HHsize	Size of household	0.025	0.011	0.030	0.002	0.047
	Work	Any work	0.046	0.033	0.158	-0.018	0.110
	PeriodNP2	Period 2	0.039	0.017	0.021	0.006	0.072
	PeriodNP3	Period 3	0.015	0.021	0.478	-0.027	0.057
	PeriodNP4	Period 4	0.016	0.020	0.435	-0.024	0.055
	PeriodNP5	Period 5	0.021	0.022	0.350	-0.023	0.065
	PeriodNP6	Period 6	0.010	0.024	0.685	-0.037	0.057
	PeriodNP7	Period 7	0.002	0.023	0.923	-0.043	0.048
	PeriodNP8	Period 8	0.013	0.020	0.519	-0.027	0.053
	PeriodNP9	Period 9	-0.014	0.025	0.574	-0.062	0.034
	PeriodNP10	Period 10	-0.031	0.028	0.261	-0.085	0.023
β	Pstrangers	Perfect strangers	0.087	0.096	0.364	-0.101	0.275
	Csize	Cohort size for Random strangers	0.001	0.004	0.787	-0.007	0.009
	Age	Age	-0.001	0.005	0.785	-0.012	0.009
	Male	Male	0.136	0.038	0.000	0.062	0.210
	Black	Black	0.032	0.067	0.633	-0.099	0.162
	Asian	Asian	0.038	0.069	0.580	-0.098	0.174
	Hispanic	Hispanic	0.023	0.061	0.703	-0.097	0.143
	Business	Business major	-0.070	0.037	0.061	-0.143	0.003
	PreSenior	Pre-Senior	0.047	0.042	0.258	-0.035	0.129
	GPAlow	Low GPA	-0.044	0.041	0.284	-0.124	0.036
	GPAhigh	High GPA	-0.097	0.079	0.220	-0.252	0.058
	HHsize	Size of household	-0.014	0.016	0.382	-0.046	0.018
	Work	Any work	0.035	0.041	0.402	-0.046	0.116
	PeriodNP2	Period 2	0.035	0.018	0.056	-0.001	0.071
	PeriodNP3	Period 3	0.088	0.019	0.000	0.051	0.126
	PeriodNP4	Period 4	0.064	0.020	0.002	0.024	0.103
	PeriodNP5	Period 5	0.095	0.023	0.000	0.049	0.141
	PeriodNP6	Period 6	0.071	0.024	0.003	0.024	0.117
	PeriodNP7	Period 7	0.020	0.024	0.406	-0.027	0.067
	PeriodNP8	Period 8	0.020	0.024	0.417	-0.028	0.068
	PeriodNP9	Period 9	0.027	0.025	0.275	-0.022	0.076
	PeriodNP10	Period 10	0.036	0.028	0.189	-0.018	0.090

Table 5: Maximum Likelihood Estimates of the Hurdle Model of Contributions for the Second Experiment

Marginal effects for α and β parameters

N=1800 responses from 180 subjects; Wald test of H_0: $\alpha=\beta=0$ has $\chi^2_{25} = 115.5$ (*p*-value<0.001)

Parameter	Variable	Description	Estimate	SE	*p*-value	95% Confidence Intervals	
α	Pstrangers	Perfect strangers	-0.299	0.156	0.055	-0.605	0.007
	Csize	Cohort size for Random strangers	-0.024	0.015	0.119	-0.054	0.006
	np_p	Sequence	0.255	0.063	0.000	0.132	0.378
	Low	Low reward for contributing	-0.373	0.070	0.000	-0.511	-0.236
	Age	Age	0.021	0.012	0.089	-0.003	0.045
	Male	Male	-0.041	0.071	0.565	-0.181	0.099
	Black	Black	0.082	0.125	0.509	-0.162	0.327
	Asian	Asian	-0.041	0.103	0.694	-0.243	0.162
	Hispanic	Hispanic	0.130	0.082	0.111	-0.030	0.291
	OtherRace	Other Race	0.109	0.137	0.425	-0.159	0.378
	Business	Business major	-0.092	0.068	0.176	-0.225	0.041
	PreSenior	Pre-senior	0.088	0.067	0.186	-0.043	0.219
	GPAlow	Low GPA	0.180	0.069	0.009	0.045	0.315
	GPAhigh	High GPA	-0.081	0.096	0.401	-0.270	0.108
	HHsize	Size of household	-0.021	0.025	0.415	-0.071	0.029
	Work	Any work	0.078	0.064	0.222	-0.047	0.204
	PeriodNP2	Period 2	-0.028	0.037	0.454	-0.100	0.045
	PeriodNP3	Period 3	-0.081	0.037	0.031	-0.153	-0.008
	PeriodNP4	Period 4	-0.109	0.038	0.004	-0.183	-0.035
	PeriodNP5	Period 5	-0.133	0.038	0.001	-0.208	-0.058
	PeriodNP6	Period 6	-0.195	0.042	0.000	-0.278	-0.111
	PeriodNP7	Period 7	-0.239	0.041	0.000	-0.319	-0.159
	PeriodNP8	Period 8	-0.212	0.043	0.000	-0.296	-0.128
	PeriodNP9	Period 9	-0.257	0.040	0.000	-0.336	-0.178
	PeriodNP10	Period 10	-0.301	0.040	0.000	-0.378	-0.223
β	Pstrangers	Perfect strangers	-0.141	0.100	0.160	-0.338	0.056
	Csize	Cohort size for Random strangers	-0.015	0.010	0.130	-0.034	0.004
	np_p	Sequence	0.045	0.050	0.374	-0.054	0.143
	Low	Low reward for contribution	-0.288	0.055	0.000	-0.395	-0.181
	Age	Age	-0.003	0.013	0.842	-0.028	0.023
	Male	Male	0.139	0.045	0.002	0.050	0.228
	Black	Black	-0.016	0.095	0.863	-0.202	0.170
	Asian	Asian	-0.043	0.080	0.589	-0.199	0.113
	Hispanic	Hispanic	-0.104	0.074	0.159	-0.248	0.041
	OtherRace	Other Race	-0.009	0.077	0.905	-0.161	0.142
	Business	Business major	0.074	0.051	0.142	-0.025	0.174
	PreSenior	Pre-Senior	-0.005	0.055	0.930	-0.112	0.103
	GPAlow	Low GPA	-0.087	0.050	0.083	-0.185	0.012
	GPAhigh	High GPA	-0.042	0.079	0.600	-0.197	0.114
	HHsize	Size of household	0.015	0.019	0.429	-0.022	0.051
	Work	Any work	0.118	0.063	0.061	-0.005	0.241
	PeriodNP2	Period 2	0.026	0.025	0.290	-0.022	0.075
	PeriodNP3	Period 3	0.026	0.030	0.378	-0.032	0.085

PeriodNP4	Period 4	-0.027	0.029	0.341	-0.083	0.029
PeriodNP5	Period 5	-0.027	0.033	0.422	-0.092	0.038
PeriodNP6	Period 6	-0.029	0.034	0.392	-0.095	0.037
PeriodNP7	Period 7	-0.023	0.037	0.533	-0.094	0.049
PeriodNP8	Period 8	-0.006	0.037	0.876	-0.077	0.066
PeriodNP9	Period 9	-0.036	0.038	0.347	-0.110	0.039
PeriodNP10	Period 10	-0.039	0.038	0.305	-0.114	0.036

References

Andreoni, James, "Why Free Ride? Strategies and Learning in Public Goods Experiments," *Journal of Public Economics*, 37, 1988, 291-304.

Andreoni, James, and Croson, Rachel T.A., "Partners versus Strangers: Random Rematching in Public Goods Experiments," in C.R. Plott and V.L. Smith (eds.), *Handbook of Experimental Economics Results* (North-Holland: Amsterdam, 2008).

Binmore, Ken, *Fun and Games: A Text on Game Theory* (D.C. Heath & Co.: Lexington, MA, 1992).

Botelho, Anabela; Harrison, Glenn W.; Hirsch, Marc A., and Rutström, Elisabet E., "Bargaining Behavior, Demographics and Nationality: What Can the Experimental Evidence Show?" in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics* (Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 10, 2005, 337-372).

Botelho, Anabela; Harrison, Glenn W.; Pinto, Lígia M. Costa, and Rutström, Elisabet E., "Social Norms and Social Choice," *Working Paper 05-23*, Department of Economics, College of Business Administration, University of Central Florida, 2005.

Brandts, Jordi, and Schram, Arthur, "Cooperation and Noise in Public Goods Experiments: Applying the Contribution Function Approach," *Journal of Public Economics*, 79, 2001, 399-427.

Burlando, Roberto, and Hey, John D., "Do Anglo-Saxons Free-Ride More?" *Journal of Public Economics*, 64, 1997, 41-60.

Coller, Maribeth; Harrison, Glenn W., and McInnes, Melayne Morgan, "Evaluating the Tobacco Settlement: Are the Damages Awards Too Much or Not Enough?"*American Journal of Public Health*, 92(6), June 2002, 984-989.

Conover, W.J., *Practical Nonparametric Statistics* (New York: Wiley, Second Edition, 1980).

Croson, Rachel T.A., "Partners and Strangers Revisited," *Economics Letters*, 53, 1996, 25-32.

Engle-Warnick, Jim, and Slonim, Robert L., "Inferring Repeated-Game Strategies from Actions: Evidence from Trust Game Experiments," *Economic Theory*, 28, 2006, 603-632.

Fehr, Ernst, and Gächter, Simon, "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90(4), September 2000, 980-994.

Fischbacher, Urs, "z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual," *Working Paper Nr. 21*, Institute for Empirical Research in Economics, University of Zurich, 1999.

Fudenberg, Drew, and Tirole, Jean, *Game Theory* (MIT Press: Cambridge, MA, 1991).

Goeree, Jacob K.; Holt, Charles A., and Laury, Susan K., "Private costs and public benefits: unraveling the effects of altruism and noisy behavior," *Journal of Public Economics*, 83, 2002, 255–276.

Harrison, Glenn W., and Hirshleifer, Jack, “An Experimental Evaluation of Weakest-Link/ Best-Shot Models of Public Goods,” *Journal of Political Economy*, 97, February 1989, 201-225.

Huck, Steffen; Normann, Hans-Theo Normann, and Oechssler, Jörg, “Two are few and four are many: Number effects in experimental oligopoly,” *Journal of Economic Behavior and Organization*, 53, 2004, 435-446.

Isaac, R. Mark and Walker, James M., “Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism,” *Quarterly Journal of Economics*, 53, 1988, 179-200.

Keser, Claudia, and van Winden, Frans., “Conditional Cooperators and Voluntary Contributions to Public Goods,” *Scandinavian Journal of Economics*, 102, 2000, 23-39.

Kreps, David M.; Milgrom, Paul; Roberts, John, and Wilson, Robert, “Rational Cooperation in the Finitely-Repeated Prisoners’ Dilemma,” *Journal of Economic Theory*, 27, 1982, 245-252.

McDowell, Allen, “From the help desk: hurdle models,” *Stata Journal*, 3(2), 2003, 178-184.

Palfrey, Thomas R., and Prisbrey, Jeffrey E., “Altruism, Reputation, and Noise in Linear Public Goods Experiments,” *Journal of Public Economics*, 61, 1996, 409-427.

Palfrey, Thomas R., and Prisbrey, Jeffrey E., “Anomalous Behavior in Linear Public Goods Experiments: How Much and Why?” *American Economic Review*, 87, 1997, 829–846.

Papke, Leslie E., and Wooldridge, Jeffrey M., “Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates,” *Journal of Applied Econometrics*, 11, 1996, 619-632.

Sonnemans, Joep; Schram, Arthur, and Offerman, Theo, “Strategic Behavior in Public Good Games: When Partners Drift Apart,” *Economics Letters*, 62, 1999, 35-41.

Weimann, Joachim, “Individual Behaviour in a Free Riding Experiment,” *Journal of Public Economics*, 54, 1994, 185-200.

Wooldridge, Jeffrey M., “Cluster-Sample Methods in Applied Econometrics,” *American Economic Review (Papers & Proceedings)*, 93, May 1993, 133-138.

Appendix: Instructions (NOT FOR PUBLICATION)

A. Perfect Strangers Sessions

WELCOME TO THE EXPERIMENT

The experiment you are about to take part in consists of one task. Instructions for the task will be given in a moment and you will have a chance to practice the task and ask questions before performing it for money. All earnings will be paid at the end of the experiment in cash. We expect the experiment to last up to two hours. Please make sure that you can stay until the end. In addition to your earnings, you will also receive our standard $5 participation fee. This fee is guaranteed independent of the outcome of the experiment.

Payments will be made in private at the end of each session. You will be called up one by one to get paid.

The instructions which we have distributed to you are solely for your private information. **You may not communicate with the other participants during the session**. Should you have any questions during the actual paid task, please raise your hand and we will come to your desk and help you out in private.

During the session your earnings will be calculated in tokens. At the end of the session the total amount of tokens you have earned will be converted to dollars at the following rate:

1 token = 5¢

We will now go over the instructions for the task.

Instructions for the task

The task that will be described to you next will be repeated 10 times. Each repetition will be referred to as a period. In each period you will be matched with new 3 persons in the room. You will not know who these persons are, nor will they know when and if they are matched with you. All matchings are done anonymously. You will only meet each person one time. Actually, everyone in the room will only meet every other person one and only one time.

In order to make sure that each person is matched up with every other person only one time, we have used a computer algorithm. We have created a handout for you so that you can verify for yourself that it is possible to make sure that people are only meeting once.

The handout has as many rows as there are people in the room, and it has 10 columns, one for each period. Each row is labeled with a letter. Each letter represents a computer in this room, and therefore a person sitting at that computer. By reading a row across the columns from left to right you can see that a person, as represented by the letter, is matched with different persons in each period. Nobody meets the same person twice. Please watch the display of the table at the front of the lab and we will look at a couple of examples.

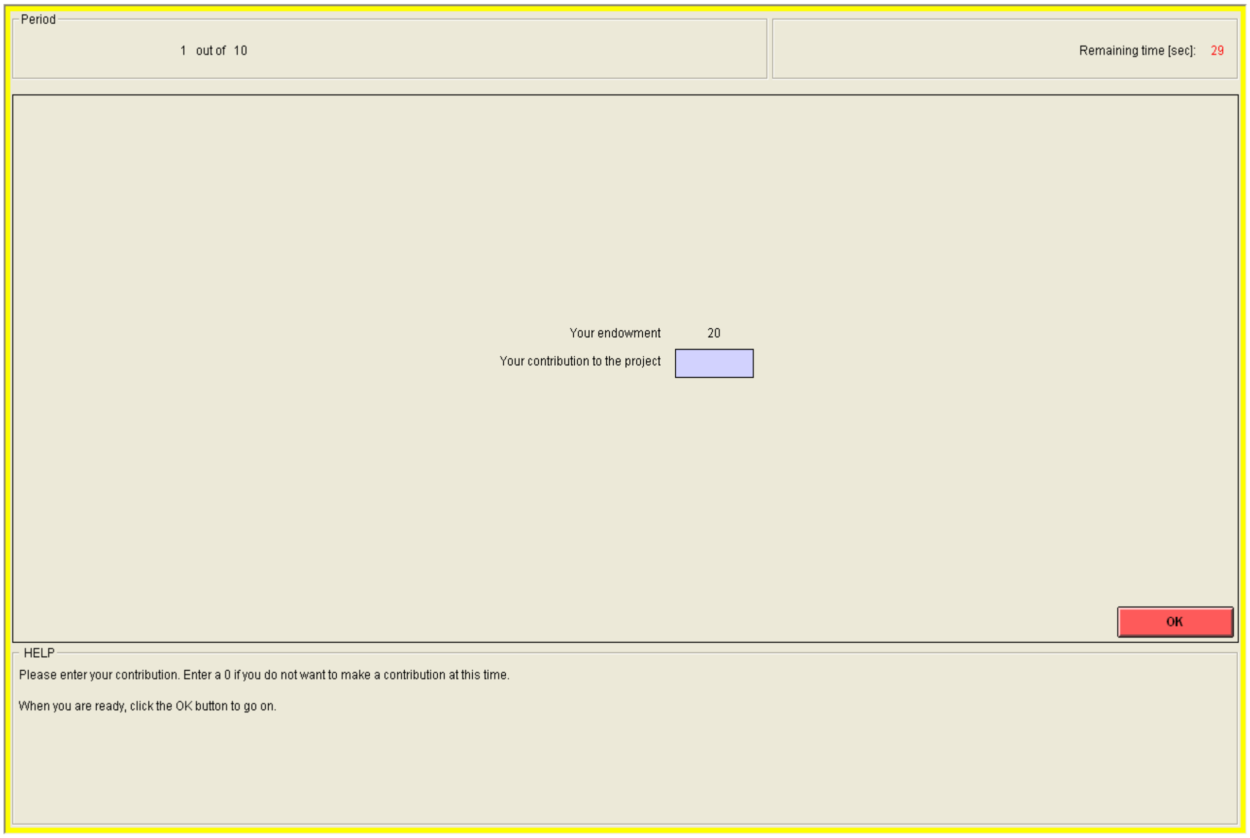
All matchings are done anonymously so you will not know which persons you are matched with. The same thing is true in reverse – nobody else will know when and if they are being matched with you either.

At the beginning of each period you will be given 20 tokens, and you have to decide how many tokens you would like to contribute to a project. The following pages describe what this project is about.

Detailed information on the task

At the beginning of each period each participant receives 20 tokens. We call this your endowment. Your task is to decide how to use your endowment. You have to decide how many of the 20 tokens you want to contribute to a project and how many of the 20 tokens to keep for yourself. The consequences of your decision are explained in detail below.

At the beginning of each period the following input screen will appear:

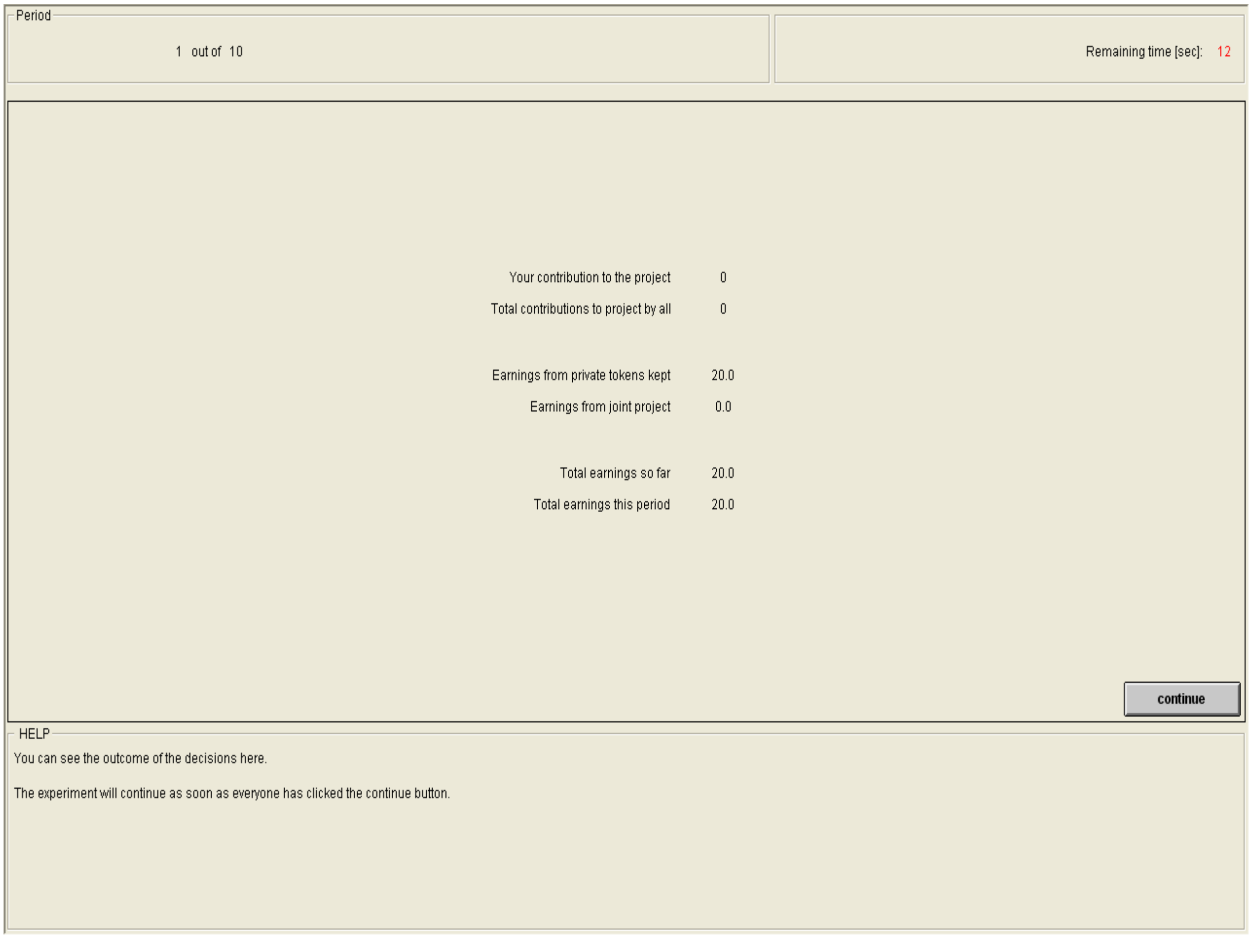


The number of the period appears in the top left corner of the screen. In the top right corner you can see how many more **seconds** remain for you to make your decision. Your decision must be made before the time displayed reaches 0 seconds.

Your endowment in each period is 20 tokens. You have to decide how many tokens you want to contribute to the project by typing a number between 0 and 20 in the input field. This field can be reached by clicking it with the mouse. As soon as you have decided how many tokens to contribute to the project, you have also decided how many tokens you keep for yourself, this is: **20 tokens minus your contribution**. After entering your contribution you must press the OK button with the mouse. Once you have done this, your decision can no longer be revised.

After both you and the 3 persons you are matched with have made your decisions, the following earnings screen will show you the total amount of tokens contributed to the project by the four of you. This screen also shows you how many tokens you have earned during the period.

The Earnings Screen:



Your earnings consist of two portions:

1) The tokens which you have kept for yourself ("Earnings from private tokens kept")
2) The "earnings from joint project". This is calculated as follows:
Your earnings from the project = 0.8 × the sum of the contributions of the four of you who are matched up this period.

Your earnings in tokens for each period is therefore: (20 - your contribution to the project) + 0.8 × (sum of the contributions to the project)

The earnings **from the project** to each of the four of you is calculated in the same way, which means that each group member receives the same earnings from the project. Suppose the sum of the contributions is 30 tokens. In this case each of you receives earnings from the project of: 0.8 × 30 = 24 tokens. If the total contribution to the project is 3 tokens, then each member of the group receives earnings of 0.8 × 3 = 2.4 tokens **from the project**.

Each token which you keep for yourself is **added to your earnings**. Suppose you contributed this token to the project instead, then your earnings **from the project** would rise **by 0.8 × 1=0.8 tokens**. However the earnings of each of the other persons would also rise by 0.8 tokens, so that your combined earnings from the project would rise by 0.8×4=3.2 tokens. Your contribution to the project therefore also raises the earnings of the other persons. On the other hand you also get earnings for each token contributed by the other members to the project. For each token contributed by the other members you earn 0.8 × 1=0.8 tokens.

To summarize: For each token you contribute to the project, your earnings are reduced by the token you contribute and increased by the project earnings of 0.8 tokens. In addition, your earnings are increased by 0.8 tokens for each token contributed by the other group members. You have 20 seconds to view the earnings screen. If you are finished before the time is up, please press the continue button by using the mouse. The period is then over and the next period commences.

Do you have any further questions?

We now ask you to respond to the questionnaire on the screen to ensure that you have understood the directions. After that you will be able to practice making these decisions for two practice periods.

B. Random Strangers Sessions

Modifies the instructions above only with respect to the page on “Instructions for the task” which in the RS sessions is as follows:

Instructions for the task

The task that will be described to you next will be repeated 10 times. Each repetition will be referred to as a period. In each period the computer will randomly and anonymously match you with 3 persons in the room. You will not know who these persons are, nor will they know when and if they are matched with you. The same thing is true in reverse – nobody else will know when and if they are being matched with you either. There is some chance that you will meet the same persons more than one time, although you will not know when.

At the beginning of each period you will be given 20 tokens, and you have to decide how many tokens you would like to contribute to a project. The following pages describe what this project is about.