Cooperation, Contributor Types, and Control Questions*

Luke Boosey† R. Mark Isaac‡ Douglas Norton§ Joseph Stinn¶

Abstract

A large number of experimental studies use the strategy method procedure introduced by Fischbacher, Gächter, and Fehr (Econ. Lett. 71:397-404, 2001) to measure individuals' attitudes towards cooperation. The procedure elicits subjects' strategic-form decisions in a one-shot game and classifies each subject as one of several different contributor types. In this paper, we examine the robustness of the procedure and its capacity to help explain the pattern of contributions observed in a separate, repeated game setting. Overall, we show that the elicited contributor types can explain behavior fairly well in the repeated game. Free-rider types contribute less than conditional cooperators, although we observe evidence consistent with strategic cooperation in the early periods of the repeated game. Nevertheless, by the last period, classified free-rider types converge to pure free-riding behavior. We highlight a methodological concern related to the use of control questions, showing that while they do not affect the distribution of classified types in our experiment, their inclusion leads to substantial wait times for many subjects. This raises concerns about the psychological impact of long wait times, through boredom, frustration, or spite, on the cooperative behavior of subjects in the laboratory setting.

Keywords: Public goods, experimental methodology, strategy method, conditional cooperation, free-riding, control questions

JEL classification codes: C72, C92

*We would like to thank Angela de Oliveira, Ragan Petrie, Alex Smith, and seminar participants at UMass Amherst, Florida State University, and the 2017 SEA Meetings in Tampa for providing many helpful comments and suggestions on earlier versions of this paper. The paper was previously circulated under the title "Eliciting contributor types to explain behavior in repeated public goods games".

† *Corresponding author*; Department of Economics, Florida State University, 113 Collegiate Loop, Tallahassee, FL 32306-2180, USA; E-mail: lboosey@fsu.edu

‡Department of Economics, Florida State University, Tallahassee, FL 32306-2180

§Department of Economics, Florida State University, Tallahassee, FL 32306-2180

¶Farmer School of Business, Miami University, Oxford, OH 45056

1 Introduction

The vast majority of experimental research on social dilemmas indicates significant heterogeneity in individuals' attitudes towards cooperation. Evidence suggests that many individuals are conditionally cooperative, with a relatively smaller fraction of individuals who follow the dominant strategy incentive to free ride. In a seminal contribution, Fischbacher, Gächter and Fehr (2001) (hereafter, FGF) introduced an incentive-compatible procedure, using a variant of the strategy method (Selten, 1967), for measuring individuals' attitudes towards cooperation. The procedure elicits subjects' strategic-form decisions in a one-shot, linear public goods game and classifies subjects into different classes of contributor types.

In their experiment, FGF classified 50% of their subjects as conditional cooperators and 30% of their subjects as pure free-riders. Several subsequent studies have found similar evidence of conditional cooperation using the FGF procedure.[1] In an extension of the original experiment, Fischbacher and Gächter (2010) used the classifications from the FGF procedure to help explain the decline in cooperation observed in a multi-period sequence of one-shot public goods games implemented using a random matching protocol. Their results support the argument that the well-documented decay in average contributions is driven primarily by an average preference for imperfect conditional cooperation, though also still in part by the downward adjustment of beliefs about the contributions of others.

In this paper, we examine whether the FGF procedure can be similarly used to explain the patterns of behavior observed in *repeated* public goods games. That is, in contrast to Fischbacher and Gächter (2010), we use a partners matching protocol such that players interact in fixed groups across periods. In this setting, there are several factors that may affect behavior other than the individual's preference for cooperation. In particular, reputation concerns may influence the use of repeated game strategies, even in finitely repeated games. If selfish types believe there are enough conditionally cooperative types in their group, they may strategically behave similarly to a conditional cooperator, especially in the early rounds of the repeated game (see, e.g., Kreps and Wilson, 1982; Kreps et al., 1982; Andreoni, 1988; Croson, 1996; Andreoni and Samuelson, 2006; Ambrus and Pathak, 2011). Thus, ex-ante, it is not clear whether the types elicited using the one-shot FGF procedure will help to explain behavior as successfully as in Fischbacher and Gächter (2010).

Given its popularity among researchers studying cooperation in social dilemmas, we set out to design a careful investigation of the FGF procedure's applicability to the repeated game setting. In preparation for previous work (Boosey, Isaac and Norton, 2016), the authors elicited

[1] Although not a comprehensive list, see, e.g., Burlando and Guala (2005), Gächter and Thöni (2005), Chaudhuri and Paichayontvijit (2006), Kocher et al. (2008), Muller et al. (2008), Gächter and Herrmann (2009), Herrmann and Thöni (2009), Fischbacher and Gächter (2010), Fischbacher, Gächter and Quercia (2012), and Martinsson, Pham-Khanh and Villegas-Palacio (2013). See also the recent article by Thöni and Volk (2018), who review 17 replications of the original study and offer a refinement of the classification criteria.

contributor types and collected data from a repeated public goods game as part of a larger experiment with several unrelated treatments. Their full experiment consisted of 30 periods, broken into two blocks. In the first block (periods 1 to 10), subjects played a standard repeated public goods game with partners matching. During the second block (periods 11 to 30), they introduced several different treatments that modified the game.[2] The authors had intended to use the FGF classifications and the resulting data from the first block of their experiment to inform their analysis of the data in the second block. However, the results were so surprising that they set aside those intentions.[3] In looking at the apparent inability of the FGF procedure to appropriately predict behavior in the first 10 periods, the current authors wondered whether there is some organic inability of the FGF protocol to predict behavior in the repeated public goods game with partners matching. However, there were two limiting features of the experimental design implemented by Boosey, Isaac, and Norton, that gave pause to making such a sweeping conclusion. Our current design addresses these two features in order to provide clean evidence regarding the appropriateness of using the FGF procedure to help explain behavior in a repeated (partners) public goods game.

The first limiting feature of the design used in Boosey, Isaac and Norton (2016) was that during the first block of 10 periods, the subjects knew that there would be an additional 20 periods during which they would interact (in some capacity) with the same group of players. Even though subjects did not know any specific details about the decisions they would need to make in subsequent periods, they were aware from the beginning that the experiment would last for a total of 30 periods. Thus, one might be concerned that the anticipation of future interaction could distort subject behavior during the first 10 periods (and especially in the 10th period) of the experiment.[4] In our study, we made it clear to the subjects at the beginning of the experiment that there would not be any further interaction beyond period 10.

The second limiting feature of the design in Boosey, Isaac and Norton (2016) concerns the use of control questions in the instructions phase of the experiment. Recent evidence suggests that behavior in public goods games may be sensitive to the format and length of instructions (Ramalingam, Morales and Walker, 2018). Similarly, there have been a number of studies to investigate the extent to which confusion plays a role in public goods experiments (Burton-Chellew, El Mouden and West, 2016). In their earlier experiments, Boosey, Isaac, and Norton used a streamlined version of instructions that did not include any structured control questions.

[2]We refer the interested reader to Boosey, Isaac and Norton (2016) for further details, since the treatments are not relevant for the current paper.

[3]To see an example of the unexpected results that concerned them, see Figure A.1 in Appendix A. The figure shows that free rider types (i) contributed more, on average, than conditional cooperator types in all but one of the first 10 periods, and (ii) contributed 30% of the endowment in period 10.

[4]This would be especially true if, for instance, free rider types were to behave strategically by contributing during the early rounds of the game, in the same fashion as suggested by Kreps et al. (1982) or Ambrus and Pathak (2011), and considered period 10 to still be relatively early. Indeed, the 10th period of a 30 period interaction is only one-third of the way through the full interaction.

In our current study, we implement two treatments, distinguished by whether or not participants were required to complete a set of control questions before the FGF procedure was introduced. To keep things as consistent as possible with the way others have implemented the FGF protocol, we used the original set of control questions chosen by Fischbacher, Gächter and Fehr (2001), adapted to the particular parameters used in our experiment.

Our study makes several contributions. First and foremost, we find considerable support for the applicability (and robustness) of the FGF procedure to repeated game settings with partners matching. Behavior in our 10-period repeated games is largely in line with the participants' elicited types, after accounting for the possibility of strategic behavior. Subjects classified as free-rider types contribute positive amounts in the early rounds, but nothing in the final round. Conditional cooperator types contribute more than free-riders, but exhibit a similar decline over time. We further find that groups with more conditional cooperators (and fewer free-riders) contribute more, on average, across the 10 periods of the repeated game.

Our second contribution is more methodological. Overall, we conclude that the presence or absence of control questions does not affect the distribution of classified types. We do observe slightly more free-rider types and slightly fewer conditional cooperator types when control questions are included, but the difference between distributions is not statistically significant. However, sessions that included control questions lasted 35 minutes longer (on average) than sessions without control questions, which raises concerns about the impacts of boredom, frustration, and fatigue on subject behavior. Taken together, these observations suggest that control questions of the same length and time demands as in Fischbacher, Gächter and Fehr (2001) can lead to important session costs, both in terms of time and behavioral distortions, that must be weighed against their capacity to improve the subjects' understanding of the game. Although our experiment is not designed to detect whether the introduction of control questions affects contributor types through improved understanding or through frustration or boredom (see, e.g., Jensenius, 2016), we provide some evidence that subjects classified as free-riders endured significantly longer wait times than subjects in the other classification types.

Finally, we also examine the robustness of the FGF procedure to different classification criteria, including the original FGF criteria and the statistical classification algorithm introduced by Kurzban and Houser (2005) (hereafter, KH). While both approaches have been well received by subsequent research, we are not aware of any study that has examined the robustness of the FGF elicitation mechanism to the choice of classification criteria. The results regarding the distributions of contributor types classified using the KH criteria are qualitatively similar to those obtained using the FGF criteria. That is, while there are no significant differences overall, we observe slightly more pure free-rider types and slightly fewer conditional cooperators when the control questions are included. However, in terms of explaining the patterns of behavior in the repeated game, we find that using the KH criteria is slightly inferior to using the relatively stricter FGF criteria.

Our paper is closely related to two recent studies by de Oliveira, Croson and Eckel (2015) and Cotla and Petrie (2015). In their study, de Oliveira, Croson and Eckel (2015) examine the effect of group composition (in terms of contributor types) on contributions in a repeated game setting. However, their classification criteria are far less stringent than the original FGF criteria and they concentrate mostly on the effect of free-riders ("bad apples") on cooperation.[5] Cotla and Petrie (2015) also examine the consistency between social preferences elicited by the one-shot FGF procedure and behavior in a repeated linear public goods game. However, rather than use the FGF classification criteria, they use the KH statistical classification algorithm to group participants into classes of contributor type. They find, among other things, that social preferences are relatively stable across the one-shot and repeated environments. Moreover, they argue that the pattern of contributions in the repeated game is best explained by a combination of the effects of classified contributor type on first-period contributions with a simple model of payoff-based reinforcement learning.

In the next section, we describe the experimental design and procedures used in our experiments, including a review of the FGF procedure. In Section 3, we present the main findings. We first provide a comparison of the type classifications across experiments, using various classification approaches. We then examine whether behavior in the repeated game component of the experiment is consistent with the classified contributor types. We integrate a discussion of our findings into Section 3, then provide a few concluding remarks in Section 4.

2 Experimental Design & Procedures

2.1 The game

The basic decision situation in our experiment is a linear public goods game with a fixed endowment ω, to be allocated between a private good and a public good. Thus, the payoff to player i is given by

$$\pi_i(y) \quad = \quad \omega - y_i + 0.5 \sum_{j=1}^{5} y_j \tag{1}$$

where $y = (y_j)_{j=1}^5$ is the profile of contributions made to the public good by the members of i's group.

In all sessions, participants were randomly seated at private computer terminals and provided with a set of written instructions describing the linear public goods setting and the payoff function in (1). An experimenter read the instructions aloud and privately answered any ques-

[5] They also systematically vary the composition of the groups and examine the effect of providing information about the elicited types to group members.

tions. Our two treatments are labeled NC10 and C10.[6] In the NC10 experiment, subjects were then provided with instructions for the FGF elicitation procedure. In contrast, in C10, subjects were required to complete the ten control questions used by Fischbacher, Gächter and Fehr (2001), adapted to our experimental parameters, before the elicitation procedure was introduced. We include the list of ten control questions in Appendix B, along with a sample of the instructions for C10 and NC10.

The control questions were presented across three screens using z-Tree (Fischbacher, 2007). Subjects were required to wait until everyone had correctly answered all questions on the current page before advancing to the next page of questions. As noted in Section 3.1, this led to long wait times for the majority of subjects in each session.

2.2 The FGF procedure

In order to keep our design as close as possible to the original implementation of the FGF procedure, the instructions were written in virtually identical language to those used in Fischbacher, Gächter and Fehr (2001). The only substantive difference between our experiments and their original design is that we use groups of five players with a marginal per capita return (MPCR) of 0.5 instead of groups of four with an MPCR of 0.4. Given that we used these parameters in Boosey, Isaac and Norton (2016), we opted to keep the same game structure in our two new designs, C10 and NC10, rather than revert to the parameters used in Fischbacher, Gächter and Fehr (2001). In this part of the experiment, subjects were given an endowment of $\omega = 20$ tokens.

Subjects then made two types of decisions referred to as their *unconditional investment* and their *investment table*. For the unconditional investment, each subject made a single decision about how many of their 20 tokens to invest in a project (the public good). Then, for the investment table, subjects indicated their desired investment for each of the 21 possible (rounded) average investment levels of the other four players. Thus, the investment table elicits an investment schedule, using a variant of the strategy method.

In order to make both decisions potentially payoff relevant, subjects were told that, at the end of the experiment, one player in each group would be randomly selected to have their decision determined by their investment table and the average unconditional investment made by the other four players in the group. To make the details of this randomized mechanism clear to the subjects, we included two examples written in the same language as the examples included in the instructions used by Fischbacher, Gächter and Fehr (2001). At the end of the experiment, payoffs were calculated according to equation (1) and the outcome of the random mechanism, then converted into US dollars at the exchange rate of 20 tokens = \$1. In particular, note that subjects were not informed about the results from this part until after all parts of the

[6] We use C and NC to denote the inclusion and exclusion of control questions, respectively, and 10 to denote the length of the repeated game interaction. Previous versions of the paper compared the results against a subset of the data collected by Boosey, Isaac, and Norton for the 2016 paper, which was labeled NC30 for comparison.

experiment were completed.

Thus, the only difference between our experiments in the instruction phase was the inclusion of control questions (about the linear public goods game) before the FGF procedure was introduced in C10. In particular, the instructions for the elicitation procedure itself were held constant. Furthermore, the existence of and instructions for the repeated game part of the experiments were not known to subjects until after the FGF procedure was completed.

2.3 Session procedures

After the FGF procedure was completed, subjects were randomly rematched into new groups of five players for the second part of the experiment. This part consisted of a standard 10-period repeated linear public goods game, played with a partners matching protocol. Subjects received $\omega = 100$ tokens in each period, and the MPCR was 0.5. At the end of the game, the subjects' earnings were calculated by adding together their earnings from all 10 periods and converting the total into US dollars using the exchange rate of 150 tokens = \$1. These dollar earnings were added to the dollar earnings from the FGF procedure and the \$10 show-up fee.

We conducted four sessions for C10, each with 25 subjects (5 groups per session) for a total of 100 participants (20 groups). Each session lasted approximately 90 minutes, with subjects earning \$22.31, on average, including the \$10.00 show-up fee. We also conducted four sessions for NC10, with 20 subjects each in two sessions (4 groups per session) and 25 subjects each in the other two sessions (5 groups per session) for a total of 90 participants (18 groups). These sessions lasted for approximately 55 minutes, with subjects earning \$22.11, on average, including the \$10.00 show-up fee.

All sessions were run in the XS/FS laboratory at Florida State University (FSU) using z-Tree (Fischbacher, 2007). Subjects were randomly recruited via ORSEE (Greiner, 2015) from a subpopulation of FSU undergraduate students who pre-registered to receive announcements about participating in upcoming experiments. No subject participated in more than one experiment or more than one session.

3 Results

3.1 Classification of contributor types

We begin by summarizing the classification of contributor types in our data. We consider both the original FGF classification criteria and the KH classification approach. Therefore, it is instructive to outline the two approaches before presenting the results.

FGF Classification Criteria. The original criteria used by Fischbacher, Gächter and Fehr (2001) classify each subject using their investment table decisions as follows. A subject is classified as (1) a *pure free-rider* (F) if she entered 0 in every cell of the investment table; (2) a *conditional cooperator* (CC) if either the entries in her investment table are weakly monotonically increasing *or* the Pearson correlation coefficient between the subject's desired investment and the corresponding average investment of others is significantly positive (at the 1% level of significance); (3) a *triangle contributor* (T) if the entries in her investment table are increasing with the average investment of others up to a point, and decreasing thereafter; (4) *other* (O) if she cannot be classified as any of the three previous types.

KH Classification Criteria. The main alternative approach to classifying contributor types is the statistical classification algorithm introduced by Kurzban and Houser (2005). Following this approach, for each subject, we estimate the slope coefficient and intercept for an OLS regression of conditional investment on others' average investment (i.e., the subject's investment table entries) in the FGF elicitation procedure. These two coefficient estimates are used to construct a linear contribution profile (LCP) for each subject. Then we apply the following criteria. A subject is classified as (1) a *free-rider* (F) if the LCP is (strictly) less than half of the endowment everywhere; (2) an *unconditional cooperator* (U) if the LCP is (strictly) greater than half of the endowment everywhere; (3) a *conditional cooperator* (CC) if the LCP has a positive slope, starts below half of the endowment, and ends above half of the endowment; and (4) a *noisy contributor* or *other* (O) if she cannot be classified as any of the previous three types. Note that these criteria are much weaker than the FGF criteria.

Comparison of Distributions. Table 1 summarizes the number of subjects classified into each category for each of the experimental designs. The top panel uses the FGF criteria, while the bottom panel uses the KH criteria. First, note that we observe a higher percentage of pure free-rider (F) types and a lower percentage of conditional cooperator (CC) types in C10 than in NC10, whether we use the FGF criteria or KH criteria. However, the overall distributions of types are not significantly different between the two treatments under either approach (Fisher-Exact test, $p = 0.347$ using FGF criteria, $p = 0.173$ using KH criteria).

Compared with Fischbacher, Gächter and Fehr (2001), we find a much lower percentage of pure free-riders in our experiments (6.7% and 12%, compared to 29.5% in FGF) and a much higher percentage of conditional cooperator types (73% and 81%, compared to 50% in FGF). However, these percentages are more in line with those obtained in replication studies that followed FGF. For instance, Kocher et al. (2008) compare the distributions of classified types across three continents and find that the percentage of conditional cooperators (free-riders) is 80.6% (8.3%) in the United States, much higher (lower) than in Austria (44.4% conditional cooperators, 22.2% free-riders) and Japan (41.7% conditional cooperators, 36.1% free-riders). Similarly, Herrmann and Thöni (2009) report that across different subject pools, the percentage of free-rider types varies between 2% and 11%, while the percentage of conditional cooperator

Table 1: Classification of subjects by treatment.

FGF Criteria	C10	NC10
Pure free-rider (F)	12.0	6.7
Conditional cooperator (CC)	73.0	81.1
Triangle (T)	6.0	0.0
Other (O)	9.0	12.2
%	100.0	100.0
N	100	90
KH Criteria	C10	NC10
Pure free-rider (F)	33.0	25.6
Conditional cooperator (CC)	64.0	70.0
Unconditional cooperator (U)	1.0	4.4
Other (O)	2.0	0.0
%	100.0	100.0
N	100	90

Note: Frequencies are reported as percentages.

types varies between 48% and 60%.

For robustness, we also consider two variations on the original FGF criteria.[7] Note that Fischbacher, Gächter and Fehr (2001) require a strict definition for a pure free-rider (F), but a relatively more relaxed definition for conditional cooperators (CC). We examine the natural alternatives, whereby either both definitions are strict or both definitions are relaxed (weak).[8] The frequencies are reported in Table A.1 in Appendix A. Using either alternative approach, we still find no significant differences between the distributions for C10 and NC10.

We summarize these results as follows.

Finding 1 *The exclusion of control questions during the initial instructions phase does not significantly affect the distribution of classified contributor types.*

Wait times. While we do not find evidence of significant differences between the distributions of types, the inclusion of control questions substantially increased the length of the session. This

[7] Another alternative classification approach is suggested in recent work by Thöni and Volk (2018).

[8] Formally, our Strict criteria modify the requirements for a conditional cooperator (CC) to have an investment table that is weakly monotonically increasing *and* that exhibits a significantly positive Pearson correlation coefficient between desired investment and average others' investment (using the 1% level of significance). In contrast, for the Weak criteria, we modify the original FGF criteria for a pure free-rider (F) by requiring only that the entries be less than or equal to 2 (10% of the endowment) in every cell of the investment table. In our view, consistency in the level of stringency is more easily justified than enforcing a strict criterion for one type and a weak criterion for another, as in the original FGF classification.

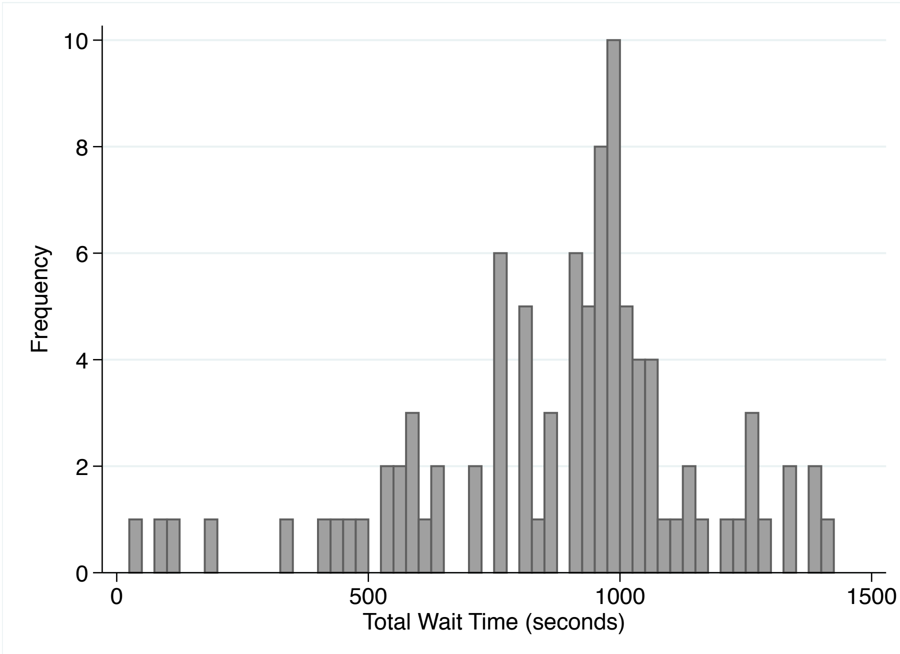


Figure 1: Frequency histogram for total wait time (C10)

raises a concern about the potential effects of long wait times on behavior. Recall that subjects who correctly completed the current page of control questions were required to wait until all subjects in the room completed the page before the next set of questions were revealed. The impact of boredom, frustration or other negative emotional responses to the wait times imposed by the slowest subjects might manifest in less cooperative behavior in the responses to the elicitation task.

In Figure 1, we plot a frequency histogram for the total time subjects were forced to wait while others were completing the control questions in C10. The 10 control questions were spread over three screens, and each subject was required to wait until everyone had correctly completed the current screen before moving to the next screen. Thus, we calculated the total wait time by first determining the longest time taken for each page in the session, taking the difference between this longest time and the time taken by each subject to complete the page, and summing the wait times from each page.

The horizontal axis in Figure 1 represents wait times in seconds. From the right tail of the histogram, we can compute that one subject was required to wait for (in total) 23 minutes and 45 seconds while others completed the control questions. The mean amount of total wait time was 14 minutes and 45 seconds (or an average of just under 5 minutes waiting per page of control questions). During this time, subjects were required to sit quietly and wait.

Although our experiment is not designed to test explicitly for the effect of wait times on type classification, we can explore the correlation between wait time and classified contributor

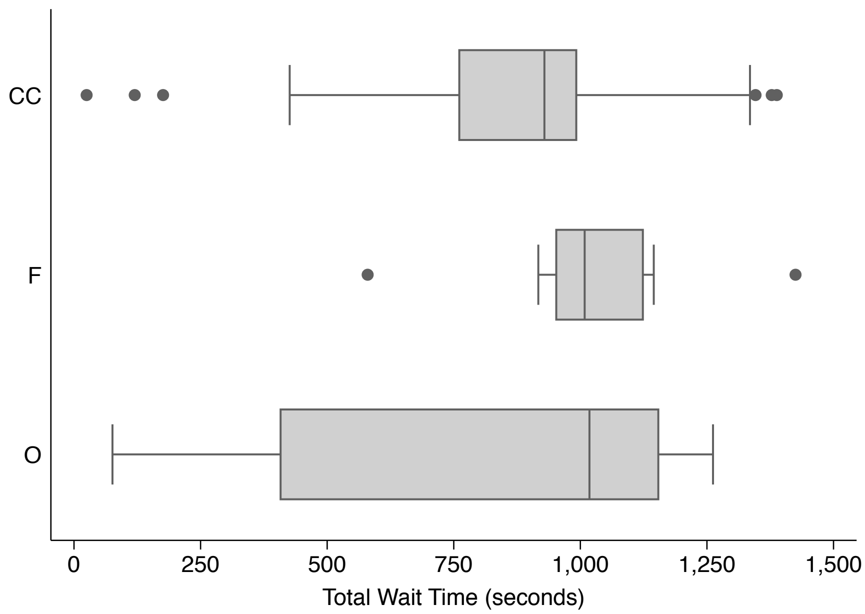


Figure 2: Boxplots for total wait time (C10)

type. In Figure 2, we provide boxplots of the total wait time for conditional cooperators (CC), pure free-riders (F), and others (O). As illustrated by the figure, the median total wait time for conditional cooperators (CC) is lower than the vast majority of the total wait times for free riders (F). Furthermore, while the median wait time for the unclassified others (O) is similar to the median for the free riders (F), the range of total wait times is considerably wider. The mean wait time for free-riders was 1023s (17m 3s), with a standard deviation of 195s. In contrast the mean wait time for conditional cooperators (CC) was 871s (14m 31s, std. dev. 267s) and for others (O) was 798s (13m 18s, std. dev. 440s). These apparent differences are also strongly supported by a Kruskal-Wallis test comparing the classified CC, F, and O subjects ($p = 0.0001$). The pairwise comparison also supports strongly significant differences between the populations of free-riders and conditional cooperators ($p = 0.0001$).

Thus, we find some intriguing evidence to suggest that the free-rider classification is correlated with longer total wait times. However, we cannot yet distinguish between alternative explanations for this correlation. For example, if free-rider types better understand the incentives of the game to begin with, they may be faster to complete the control questions, and thus endure longer wait times. Alternatively, as we conjecture above, the longer wait times may influence some conditional cooperators (or reciprocal players) to instead adopt a free-riding strategy due to frustration, boredom, or spite. One possible design choice to help distinguish between these potential explanations would be to introduce exogenous wait times. We leave this open for future research.

3.2 Average contributions in the repeated game

In this section, we analyze the observed behavior in the 10-period repeated public goods game.

FGF Criteria. In Figure 3, we compare the average contributions in the repeated game by conditional cooperators and free-rider types in C10 and NC10, where types are classified using the FGF criteria. Consistent with their dominant strategy, free-rider types contribute *exactly* zero in the last period in NC10. Similarly, the average contribution of the free-rider types in C10 is indistinguishable from zero in the last period (with a mean of 2.5 and a median of 0).[9] In both C10 and NC10, average contributions of free-riders are non-zero in the early periods, consistent with strategic behavior of selfish types, but decay over time. However, the average drops much earlier in NC10 than in C10, falling and remaining below 10 by period 4. We also observe the familiar pattern of decay over time in the average contributions for conditional cooperator types, from between 50% and 60% of the endowment in period 1, to approximately 30% of the endowment in period 10.

Restricting attention to the first period of the repeated game, we find that average contributions are significantly higher for CC types than for F types, in both C10 ($p = 0.091$, Wilcoxon ranksum test) and NC10 ($p = 0.008$, Wilcoxon ranksum test).[10] In addition, average contributions are lower for F types than for the CC types in all 10 periods, consistent with the idea that free-riders strategically contribute positive, but below-average amounts in early periods in order to manipulate the beliefs of the conditional cooperator types.

We further examine whether the composition of types within a group, based on the classification criteria, explains the variation in group level investment. Using the mean total investment across all 10 periods for a group as a single observation, we regress total investment on *numCC* (the number of conditional cooperators in the group). Pooling the two treatments together, we have 38 observations (20 groups in C10 and 18 groups in NC10). We also estimate the model separately for each treatment. The results are reported in Table 2.

When the two treatments are pooled, we find a significant, positive effect of the number of conditional cooperators on mean total investment (see Column (1), $p = 0.063$). However, when we consider the treatments separately, the effect is only present in the NC10 treatment (see Column (3), $p = 0.013$). Thus, in groups with more conditional cooperator (CC) types (and correspondingly, fewer free-rider (F) and other (O) types), mean total investment across all rounds is significantly higher, but only in the NC10 treatment.

The results of a similar set of regressions, using *numF* (the number of free-riders in the group) are reported in Table 3. While they are all in the expected direction (negative sign), none of the estimated coefficients from *numF* are statistically significant. Still, the coefficient

[9] All except one of the free-riders contributed exactly zero in the last period.

[10] If we pool the data from C10 and NC10, the difference is also strongly significant, with $p = 0.003$, Wilcoxon ranksum test.

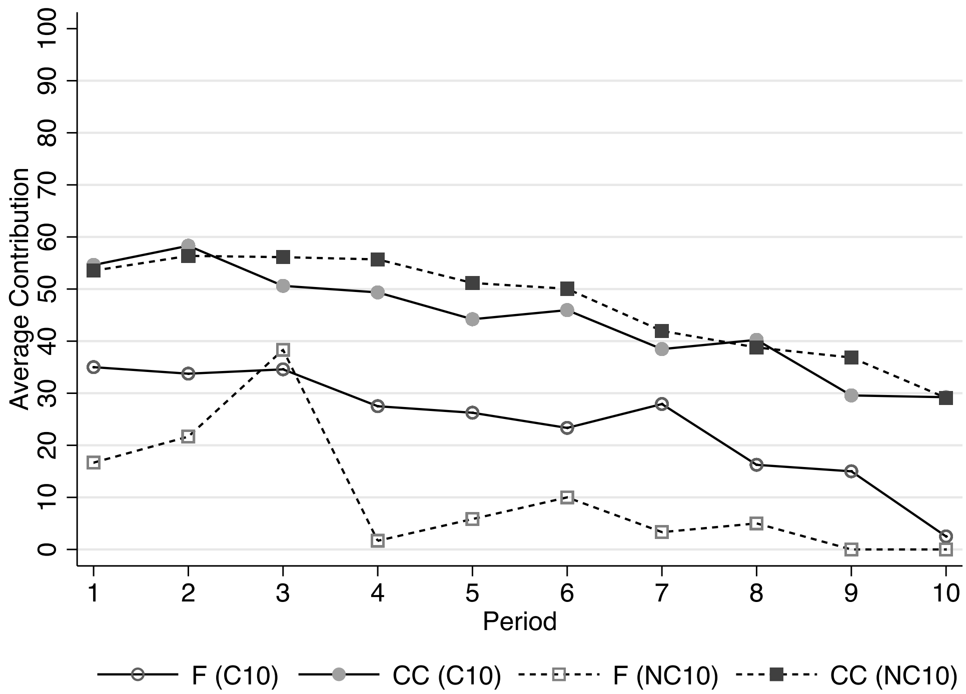


Figure 3: Average contributions by period according to FGF type classification.

Table 2: OLS regression results for the effect of the number of conditional cooperator types in the group on mean total investment.

Dependent variable: Mean total investment

	Pooled (1)	C10 (2)	NC10 (3)
numCC	28.480*	4.432	65.852**
	(14.817)	(18.929)	(23.511)
Intercept	93.344	179.835 **	-56.790
	(58.681)	(71.704)	(97.413)
Observations	38	20	18
R-squared	0.093	0.003	0.329

Standard errors are in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

for the NC10 treatment is the closest to being significant at the 10% level (see Column (3), $p = 0.106$).

One plausible explanation for the differences between C10 and NC10 regarding the effects of group composition is that free-rider types in C10 engage in more strategic cooperation, and do so for more periods of the repeated game. Consistent with Figure 3, the average investment by

Table 3: OLS regression results for the effect of the number of free-rider types in the group on mean total investment.

Dependent variable: Mean total investment

	Pooled (1)	C10 (2)	NC10 (3)
numF	-32.159	-7.559	-57.421
	(21.375)	(28.915)	(33.569)
Intercept	218.002***	200.546 ***	229.418***
	(17.681)	(25.862)	(25.021)
Observations	38	20	18
R-squared	0.059	0.004	0.155

Standard errors are in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

classified free-riders remains substantially higher in C10 than in NC10 all the way through the first nine periods. If free riders do cooperate strategically for longer in C10, it would be likely that the effect of group composition on average contributions would be mitigated, compared to NC10, where free-rider types dropped to lower contributions earlier in the repeated game.

KH Criteria. If we use the KH classification criteria, the results are slightly different, although the predictions for the different types, which are more loosely defined, are also not as sharp. The results for NC10 are very similar to those obtained using the FGF criteria, as illustrated by the dashed lines in Figure 4. Free-riders contribute less than conditional cooperators in the first period ($p < 0.001$, Wilcoxon ranksum test) and converge to less than 10% of the endowment in the last period.

However, the results for C10 are quite different. As illustrated by Figure 4, average contributions for free-riders and conditional cooperators are indistinguishable in all periods. Free-riders contribute almost 30% of the endowment in the last period, and first-period contributions are not statistically different for free-riders and conditional cooperators ($p = 0.715$, Wilcoxon ranksum test). This is perhaps less surprising than at first glance, given the relatively loose classification of free-riders under the KH classification criteria. Thus, we are more inclined to attribute the differences in the pattern of behavior for free-riders in C10 to be a consequence of the misclassification of subjects who are "impure" free-riders, as opposed to the pure free-riders identified by the FGF criteria.

Turning to the effects of group composition, using the KH classification procedure, we find results very similar to those in Table 2 for the effect of the number of conditional cooperators. As shown in Table 4, more conditional cooperators in the group are correlated with higher mean total investment, but only in the NC10 treatment. Furthermore, with the looser classification

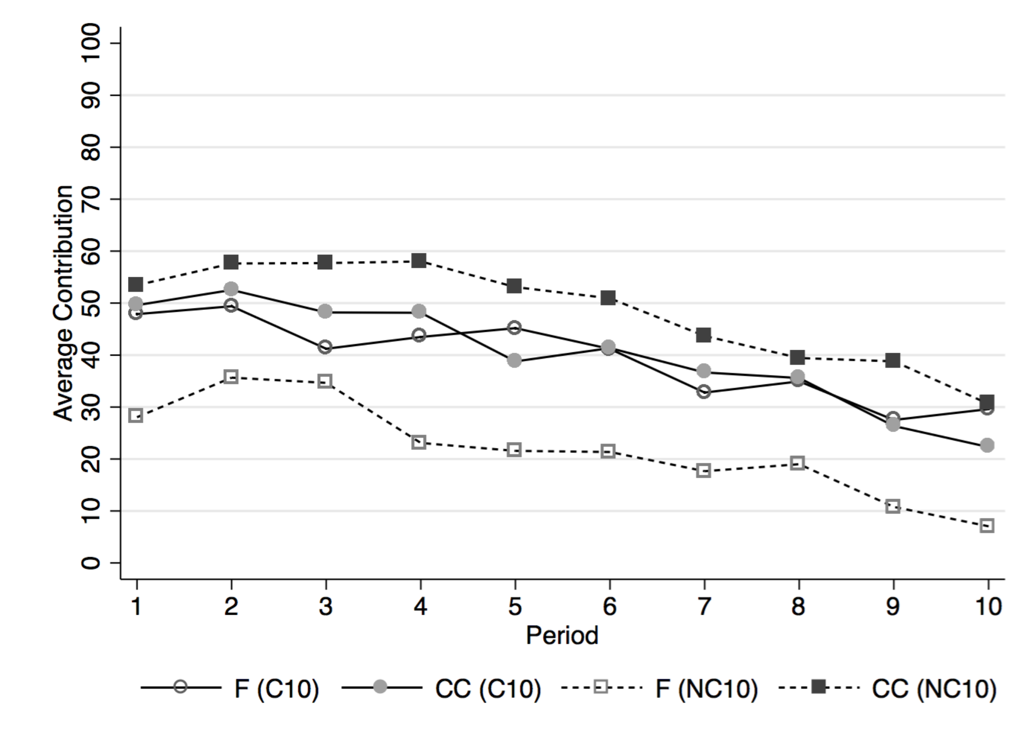


Figure 4: Average contributions by period according to KH type classification.

Table 4: OLS regression results for the effect of the number of KH conditional cooperator types in the group on mean total investment.

Dependent variable: Mean total investment

	Pooled (1)	C10 (2)	NC10 (3)
numCC	39.935**	5.058	83.162***
	(17.358)	(23.624)	(24.153)
Intercept	69.301	179.826 **	-80.789
	(59.668)	(77.993)	(86.524)
Observations	38	20	18
R-squared	0.128	0.003	0.426

Standard errors are in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

criteria for free-riders used by the KH approach, we also find a significant negative effect of the number of KH-classified free-rider types on mean total investment – but again only in the NC10 treatment (see Table 5). That is, in NC10, the mean total investment is significantly lower in groups with a greater number of KH-classified free-riders.

We summarize our results as follows.

Table 5: OLS regression results for the effect of the number of KH free-rider types in the group on mean total investment.

Dependent variable: Mean total investment

	Pooled (1)	C10 (2)	NC10 (3)
numF	-45.365**	4.839	-90.987***
	(19.457)	(29.368)	(24.333)
Intercept	269.623***	188.026 ***	326.539***
	(31.878)	(52.123)	(35.818)
Observations	38	20	18
R-squared	0.131	0.002	0.466

Standard errors are in parentheses.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finding 2 *Average contributions in the repeated game are broadly consistent with the predicted pattern of contributions for the main contributor types (using FGF criteria) in C10 and NC10. Using the KH criteria to classify contributor types somewhat weakens the correspondence between predicted and observed behavior.*

Finding 3 *The composition of the group has a significant impact on the mean total investment over all rounds, but only in the NC10 treatment. Groups with more conditional cooperators contribute higher amounts (on average) than groups with more free-riders or unclassified others. In C10, the impact of group composition on mean total investment appears to be mitigated by stronger (and more persistent) strategic cooperation on the part of the classified free-rider types.*

4 Conclusion

In this paper, we examine the validity of the FGF procedure as a tool for explaining behavior in repeated public goods games. First, we find that the elicitation of contributor types is robust to alternative classification criteria and to the inclusion or exclusion of control questions in the instructions phase of the experiment. Second, we discover that including control questions imposes significant wait times on subjects, extending the sessions by more than half an hour. We find correlations between subject classifications and the time spent waiting for others to finish. In particular, subjects classified as free-rider types experienced much longer wait times than those classified as conditional cooperators or others. One important question that we leave open for future research is whether subject behavior is affected by these long wait times, as has been suggested by Jensenius (2016).

Finally, our experiments provide some validation for the applicability of the FGF procedure, even when used to explain behavior in a repeated public goods game (with partners matching), where reputation and strategic considerations may confound the elicited type classification. Both at the individual level (based on classified types) and the group level (based on group composition), we observe contribution patterns consistent with the FGF classification procedure. Thus, we conclude that the procedure is also fairly robust for explaining the behavior of subjects in repeated game settings.

In future work, the classification procedure could be combined with various forms of belief elicitation in order to better understand the behavior of conditionally cooperative types and strategic free-rider types in repeated game settings. Furthermore, in connection with the impact of wait times on cooperative behavior, an important next step would be to determine carefully whether control questions that generate lengthy wait times for experimental subjects distort behavior due to the psychological impacts of boredom, frustration, or spite. Such considerations would need to be weighed against any purported improvements to subject understanding generated by the control questions.

A final area for further research returns to our original surprising findings that helped to motivate the current paper. The fact that subject behavior in the first 10 of 30 periods in our earlier research did not appear to be well explained by the FGF classifications remains perplexing. On the one hand, it is possible that the behavior may be consistent with extended strategic cooperation by free rider types who are playing the game as if they still have 20 periods to go. On the other hand, it could also be that the FGF classifications cannot explain behavior as well in very long periods of repeated interaction. One way to distinguish between these two possibilities would be to conduct a similar experiment in which subjects are informed that they will interact for 30 periods.

References

Ambrus, Attila, and Parag A Pathak. 2011. "Cooperation over finite horizons: A theory and experiments." *Journal of Public Economics*, 95(7): 500–512.

Andreoni, James. 1988. "Why free ride?: Strategies and learning in public goods experiments." *Journal of public Economics*, 37(3): 291–304.

Andreoni, James, and Larry Samuelson. 2006. "Building rational cooperation." *Journal of Economic Theory*, 127(1): 117–154.

Boosey, Luke A, R Mark Isaac, and Douglas A Norton. 2016. "Passionate Providers and the Possibility of Public Commitment." In *The WSPC Reference of Natural Resources and Environmental Policy in the Era of Global Change.* Vol. 4, , ed. Ariel Dinar. World Scientific Publishing.

Burlando, Roberto M, and Francesco Guala. 2005. "Heterogeneous agents in public goods experiments." *Experimental Economics*, 8(1): 35–54.

Burton-Chellew, Maxwell N, Claire El Mouden, and Stuart A West. 2016. "Conditional cooperation and confusion in public-goods experiments." *Proceedings of the National Academy of Sciences*, 113(5): 1291–1296.

Chaudhuri, Ananish, and Tirnud Paichayontvijit. 2006. "Conditional cooperation and voluntary contributions to a public good." *Economics Bulletin*, 3(8): 1–14.

Cotla, Chenna Reddy, and Ragan Petrie. 2015. "Stability of Social Preferences and Learning in Public Goods Games." Working Paper.

Croson, Rachel TA. 1996. "Partners and strangers revisited." *Economics Letters*, 53(1): 25–32.

de Oliveira, Angela, Rachel Croson, and Catherine Eckel. 2015. "One bad apple? Heterogeneity and information in public good provision." *Experimental Economics*, 18(1): 116–135.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10(2): 171–178.

Fischbacher, Urs, and Simon Gächter. 2010. "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments." *American Economic Review*, 100(1): 541–556.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters*, 71(3): 397–404.

Fischbacher, Urs, Simon Gächter, and Simone Quercia. 2012. "The behavioral validity of the strategy method in public good experiments." *Journal of Economic Psychology*, 33(4): 897–913.

Gächter, Simon, and Benedikt Herrmann. 2009. "Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518): 791–806.

Gächter, Simon, and Christian Thöni. 2005. "Social learning and voluntary cooperation among like-minded people." *Journal of the European Economic Association*, 3(2-3): 303–314.

Greiner, Ben. 2015. "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1(1): 114–125.

Herrmann, Benedikt, and Christian Thöni. 2009. "Measuring conditional cooperation: a replication study in Russia." *Experimental Economics*, 12(1): 87–92.

Jensenius, John. 2016. "On Inactivity in the Lab." Unpublished PhD Dissertation, Florida State University.

Kocher, Martin G, Todd Cherry, Stephan Kroll, Robert J Netzer, and Matthias Sutter. 2008. "Conditional cooperation on three continents." *Economics Letters*, 101(3): 175–178.

Kreps, David M, and Robert Wilson. 1982. "Reputation and imperfect information." *Journal of economic theory*, 27(2): 253–279.

Kreps, David M, Paul Milgrom, John Roberts, and Robert Wilson. 1982. "Rational cooperation in the finitely repeated prisoners' dilemma." *Journal of Economic theory*, 27(2): 245–252.

Kurzban, Robert, and Daniel Houser. 2005. "Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations." *Proceedings of the National Academy of Sciences of the United States of America*, 102(5): 1803–1807.

Martinsson, Peter, Nam Pham-Khanh, and Clara Villegas-Palacio. 2013. "Conditional cooperation and disclosure in developing countries." *Journal of Economic Psychology*, 34: 148–155.

Muller, Laurent, Martin Sefton, Richard Steinberg, and Lise Vesterlund. 2008. "Strategic behavior and learning in repeated voluntary contribution experiments." *Journal of Economic Behavior & Organization*, 67(3): 782–793.

Ramalingam, Abhijit, Antonio J Morales, and James M Walker. 2018. "Varying experimental instructions to improve comprehension: punishment in public goods games." *Journal of behavioral and experimental economics*, 73: 66–73.

Selten, Reinhard. 1967. "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments." Vol. 1, 136–168. Tübingen:Mohr.

Thöni, Christian, and Stefan Volk. 2018. "Conditional cooperation: Review and refinement." *Economics Letters*, 171: 37 – 40.

A Robustness of Classifications to Strict and Weak Criteria

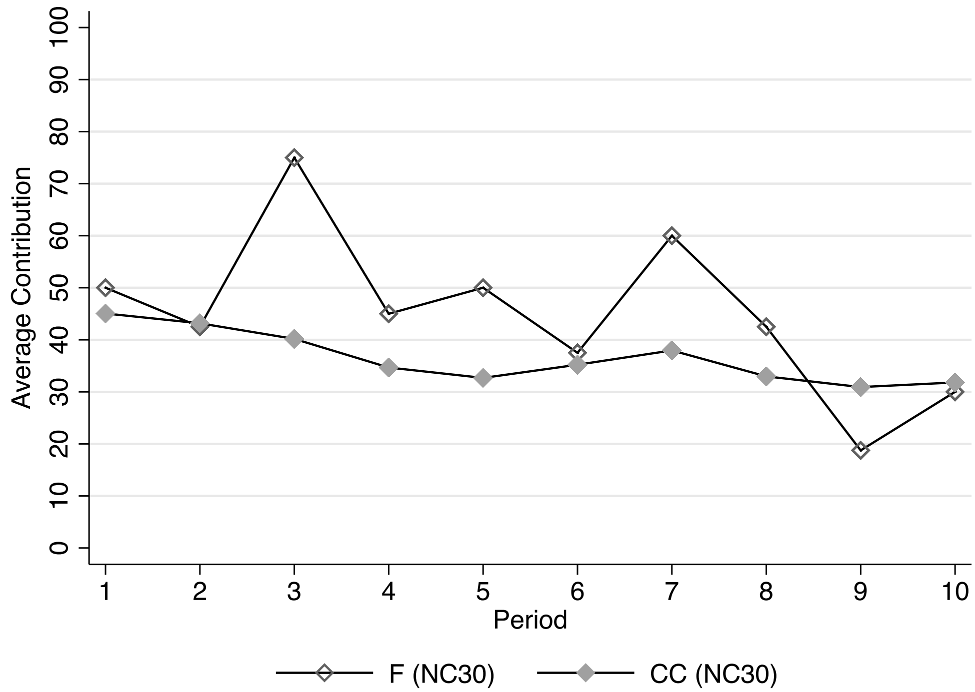


Figure A.1: Average contribution by period according to FGF classified type using the first 10 periods of data collected (but not reported in the paper) by Boosey, Isaac and Norton (2016).

Table A.1: Classification of subjects in C10 and NC10 using the Strict and Weak criteria.

	Strict		Weak	
	C10	NC10	C10	NC10
Pure free-rider (F)	12.0	6.7	13.0	8.9
Conditional cooperator (CC)	55.0	64.4	73.0	81.1
Triangle (T)	6.0	2.2	6.0	0.0
Other (O)	27.0	26.7	8.0	10.0
%	100.0	100.0	100.0	100.0
N	100	90	100	90

Note: Frequencies are reported as percentages.

B Experimental Instructions for C10 and NC10

Each experimental session consisted of two parts. First, the instructions for Part 1 were distributed and read aloud by the same person in every session. After the basic decision situation was described, subjects in C10 were given time to answer the control questions on the screen. The experiment did not advance until all subjects were able to correctly answer each control question. Subjects were permitted to ask for assistance or clarification from the person reading the instructions. After all subjects had correctly answered each control question, the remaining instructions were read aloud and subjects made their Part 1 decisions (the FGF strategy method). Subjects in NC10 proceeded directly from the description of the basic decision situation to the remaining instructions and then made their Part 1 decisions (the FGF strategy method). Only after Part 1 was completed, were the subjects informed about Part 2 of the experiment. Instructions for Part 2 were distributed and read aloud, then subjects made their decisions. At the end of the experiment, subjects were paid privately by check.

In this appendix, we first reproduce the experimental instructions (for both Part 1 and Part 2) for C10. The instructions for NC10 were identical except for the part labeled **Control Questions**, which was removed. After the instructions, we provide screenshots showing the 10 control questions that were asked in C10.

Experimental Instructions

Thank you for participating in this experiment. Please give us your full attention and follow the instructions carefully. Please do not attempt to communicate with other subjects, or engage in any other activities during the course of the experiment.

At the end of the experiment, you will be paid privately, based on the payoffs you earn and the show-up fee of $10. How much you earn depends partly on your own decisions, and partly on the decisions of others. Your earnings in the experiment will be denominated in **tokens**. At the end of the experiment, these earnings will be converted to US dollars according to the exchange rate **20 tokens = 1 US Dollar**.

All participants will be divided into groups of five members. **Other than the experimenters, nobody knows the identity of the other members in their group.**

B.1 The decision situation

You will learn how the experiment will be conducted in a moment. We first introduce you to the basic decision situation. After the decision situation is described, you will have the opportunity to answer some control questions that will be displayed on your screen that help you to understand the decision situation.

You will be a member of a group consisting of **5 people**. Each group member will be given **20 tokens** and must decide how to allocate these 20 tokens between a **group account** and an **individual account**. Your investment to your individual account can be any integer from a minimum of 0 up to the maximum of 20 tokens. Likewise, your investment to the group account can be any integer from a minimum of 0 up to a maximum of 20 tokens. However, the sum of your investments into the two accounts must be exactly 20 tokens. Thus, if you invest some number x tokens into the individual account, the other $20 - x$ tokens will be invested in the group account.

B.1.1 Your income from your individual account

You will earn one token for each token you put into your individual account. For example, if you put 20 tokens into your individual account (and therefore put 0 tokens into the group account) your income from the individual account will be 20 tokens. If you put 6 tokens into your individual account your income from the individual account will be 6 tokens. **Only you can earn income from your individual account.**

B.1.2 Your income from the group account

Each group member will profit equally from the amount you invest into the group account. At the same time, you will also earn some income from the other group members' investments in the group account. Specifically, the income **for each group member** from the group account will be determined as follows:

Income from the group account
= (*sum of all tokens invested in the group account*) $\times 0.5$

If, for example, the sum total of all investments in the group account is 60, then you and the other members of your group will each earn $60 \times 0.5 = 30$ tokens from the group account. If the sum total of all investments in the group account is 10, then you and the other members of your group will each earn $10 \times 0.5 = 5$ tokens from the group account.

B.1.3 Total Income

Your total income is the sum of your income from your individual account and your income from the group account:

Total Income
= *Income from your individual account* + *Income from the group account*
= (20 - your investment in group account) + $0.5 \times$ (sum total of all tokens invested in the group account)

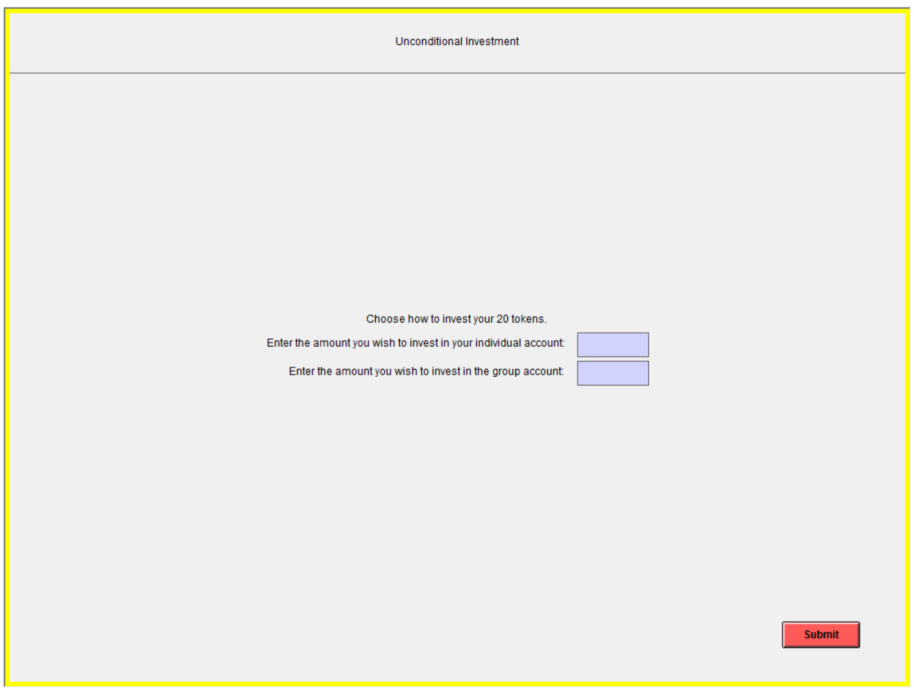
B.1.4 Control Questions

Please answer the control questions that appear on your screen. They will help you to gain an understanding of the calculation of your income, which varies with your decision about how to allocate your 20 tokens. *Once everyone has correctly answered all of the questions, we will proceed with the instructions.*

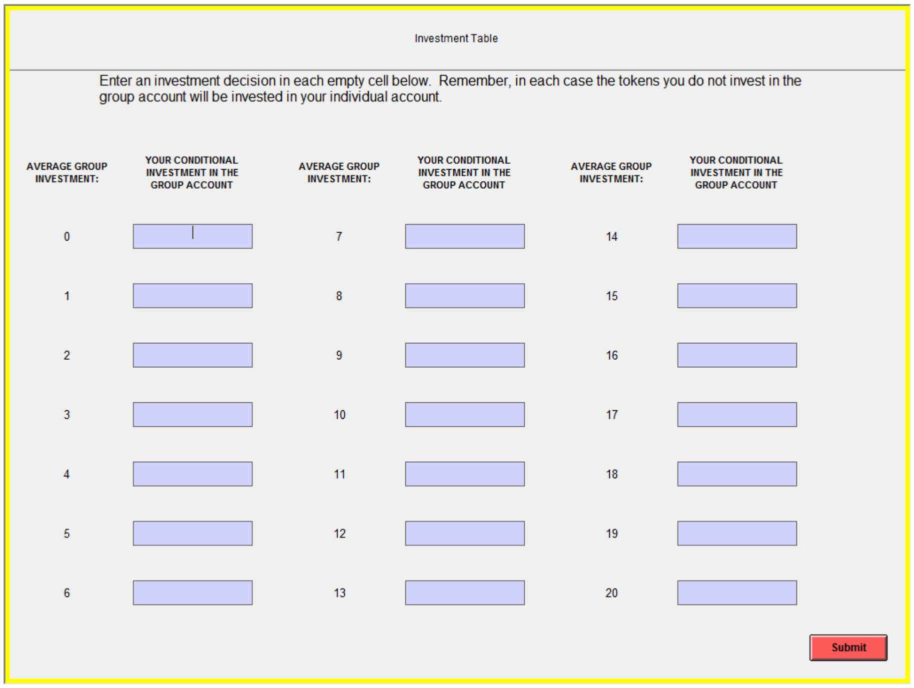
B.2 Instructions for the experiment

Part 1 of the experiment includes the decision situation just described to you. You will be paid at the end of the experiment based on the decisions you make. The experiment will only be conducted **once**. You will have 20 tokens to allocate. You can invest them into your individual account or into the group account. Each group member has to make **two types** of decisions in this experiment, which we will refer to below as the **"unconditional investment"** and **"investment table"**.

- You will first decide how many of the 20 tokens you want to invest into your individual account and how many tokens you want to invest into the group account. Your investment into the group account will be called your **unconditional investment**. You will indicate your decision in the following computer screen:



- Your second task is to fill in an **"investment table"** where you indicate how many tokens **you want to invest** in the group account, **for each possible average investment of the OTHER group members** (rounded to the next integer). Thus, you can condition your investment on the average investment made by the other group members. This should become more clear to you if you look at the table in the screenshot below.



In the table, the numbers in each column are all of the possible (rounded) average investments that could be made by the **other** group members to the group account. Your task is to insert, into each input box, the number of tokens you wish to invest in the group account, if the average investment chosen by the other group members is the amount listed to the left of that input box.

You need to make an entry into each input box. You can insert **any integer numbers from 0 to 20** in *each* input box. Note that, for each input box, you only need to enter the amount you would like to invest into the group account. The remaining tokens will be automatically placed into your individual account.

After all participants in the experiment have made an unconditional investment and have filled in their investment table, a random mechanism will select **one group member** from every group. For the randomly selected subject in your group, the **payoff-relevant decision will be determined by their investment table**. For the other subjects in your group, the payoff-relevant decision will be their **unconditional investment**.

You do not know whether the random mechanism will select you when you make your unconditional investment decision or when you fill in the investment table. You will therefore have to think carefully about both types of decisions because either one can become payoff-relevant for you. Two examples should make this clear.

Example 1. Assume that **the random mechanism selects you. This means that your payoff-relevant decision will be determined from your investment table.** The relevant decisions for the other four group members will be their unconditional investments.

Suppose that the others in your group chose unconditional investments of 0, 2, 4, and 6 tokens into the group account. Then the average investment of these four **others** is 3 tokens. If you indicated in your investment table that you will invest 2 tokens if the others contribute 3 tokens on average, then the total investment in the group account is given by $0+2+4+6+2 = 14$ tokens. Therefore, all the members of your group will earn $0.5 \times 14 = 7$ tokens in income from the group account, in addition to their respective incomes from their individual accounts.

Example 2. Assume that **the random mechanism did not select you. This means that your payoff-relevant decision is your unconditional investment.** Likewise, for three of the other group members (who were not selected by the random mechanism), the payoff-relevant decisions are their unconditional investments.

Suppose that your unconditional investment is 14 tokens, while the unconditional investments for the other three are 16, 18, and 20 tokens. Thus, the average investment in the group account made by you and these three other players is 17 tokens. For the other remaining group member, who is selected by the random mechanism, the payoff-relevant decision will be determined from their investment table. If the randomly selected group member indicated in their investment table that they will invest 18 tokens if the average investment by the others is 17 tokens, then the total investment in the group account is given by $14+16+18+20+18 = 86$ tokens. Thus, all group members will earn $0.5 \times 86 = 43$ tokens from the group account, plus their respective income from their individual accounts.

B.3 A second experiment

We will now conduct another experiment. For this experiment, the exchange rate will be **150 tokens = US $1**. This experiment lasts for **10 periods**, in which you and the other members of a group have to make decisions.

As in the other experiment, every group consists of **5 people**. Your group in this experiment will be different from your group in the other experiment. That is, before this experiment begins, you will be randomly rematched into a new group of five subjects. However, this is the only time you will be rematched. Your group will consist of the <u>same</u> people in all 10 periods. There are no additional experiments after these 10 periods.

The decision situation is the same as that described in Section 1 of these instructions. However, instead of having 20 tokens to allocate, each member of the group has to decide how to allocate **100 tokens** between their individual account and a group account. Your income will be determined in the same way as before, accounting for this one main difference. That is,

Total Income
= *Income from your individual account* + *Income from the group account*
= (100 - your investment in the group account) + 0.5 × (sum of all tokens invested in the group account)

The decision screen, which you will see in every period, looks like this:

In every period, you face the same decision situation. On the decision screen, in the input boxes provided, you must indicate how many of your 100 tokens you want to invest in your individual account, and how many tokens you want to invest in the group account. Your investments can be any integer numbers from 0 to 100, with the restriction that the sum of your investments must be exactly 100 tokens.

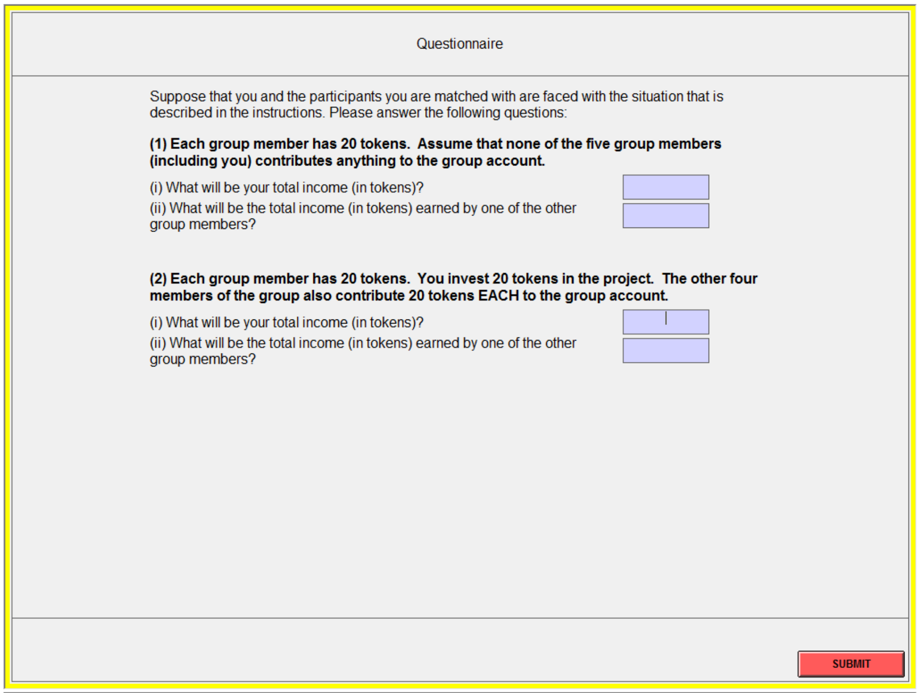
B.4 Overall Earnings

After the 10 periods of the second experiment are over, the whole experiment is finished and your overall earnings will be calculated as follows.

> **Overall Earnings** = Total Income from the first experiment (in dollars) + Total Income from all 10 periods of the second experiment (in dollars) + **Show-up Fee (which is $10)**.

Control Questions

Below, we reproduce screenshots of the control questions used in C10.


Questionnaire

Suppose that you and the participants you are matched with are faced with the situation that is described in the instructions. Please answer the following questions:

(1) Each group member has 20 tokens. Assume that none of the five group members (including you) contributes anything to the group account.

(i) What will be your total income (in tokens)?

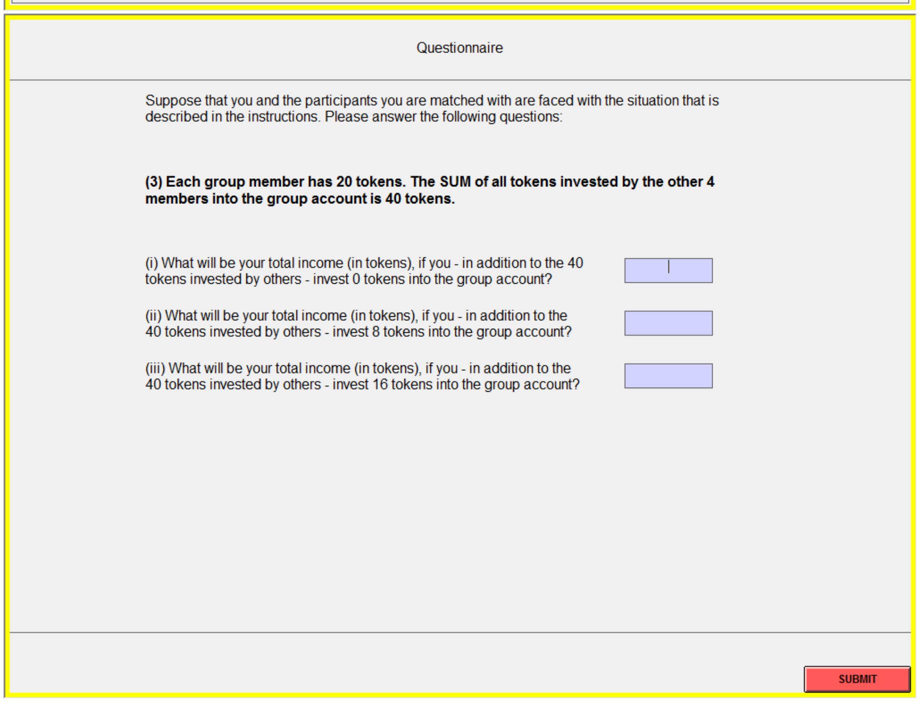
(ii) What will be the total income (in tokens) earned by one of the other group members?

(2) Each group member has 20 tokens. You invest 20 tokens in the project. The other four members of the group also contribute 20 tokens EACH to the group account.

(i) What will be your total income (in tokens)?

(ii) What will be the total income (in tokens) earned by one of the other group members?

SUBMIT


Questionnaire

Suppose that you and the participants you are matched with are faced with the situation that is described in the instructions. Please answer the following questions:

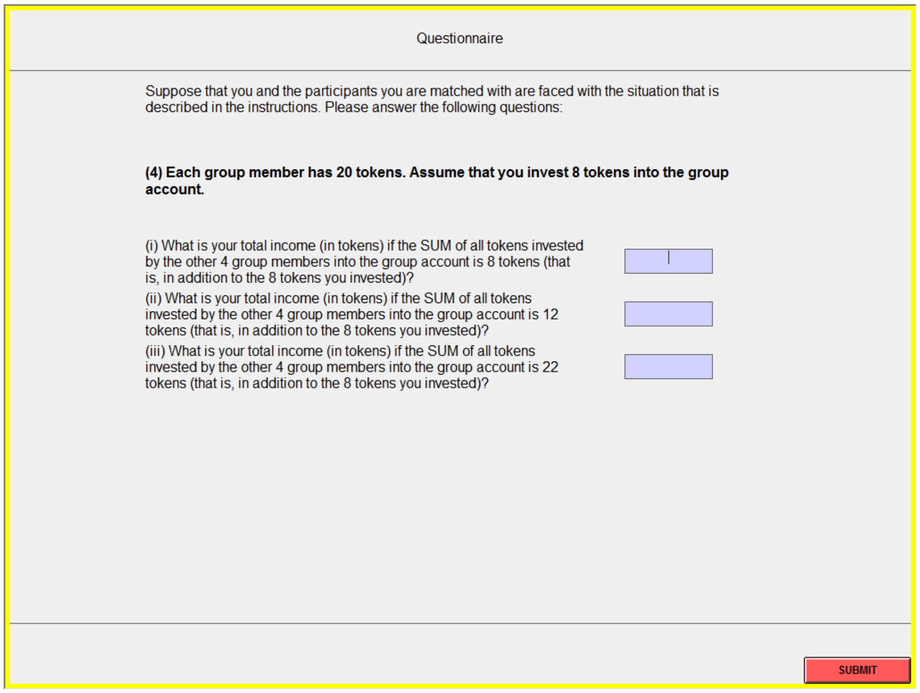
(3) Each group member has 20 tokens. The SUM of all tokens invested by the other 4 members into the group account is 40 tokens.

(i) What will be your total income (in tokens), if you - in addition to the 40 tokens invested by others - invest 0 tokens into the group account?

(ii) What will be your total income (in tokens), if you - in addition to the 40 tokens invested by others - invest 8 tokens into the group account?

(iii) What will be your total income (in tokens), if you - in addition to the 40 tokens invested by others - invest 16 tokens into the group account?

SUBMIT


Questionnaire
Suppose that you and the participants you are matched with are faced with the situation that is described in the instructions. Please answer the following questions:
(4) Each group member has 20 tokens. Assume that you invest 8 tokens into the group account.
(i) What is your total income (in tokens) if the SUM of all tokens invested by the other 4 group members into the group account is 8 tokens (that is, in addition to the 8 tokens you invested)?
(ii) What is your total income (in tokens) if the SUM of all tokens invested by the other 4 group members into the group account is 12 tokens (that is, in addition to the 8 tokens you invested)?
(iii) What is your total income (in tokens) if the SUM of all tokens invested by the other 4 group members into the group account is 22 tokens (that is, in addition to the 8 tokens you invested)?
SUBMIT