DISCUSSION PAPER SERIES

IZA DP No. 2011

Heterogeneous Social Preferences and the Dynamics of Free Riding in Public Goods

Urs Fischbacher
Simon Gächter

March 2006

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Heterogeneous Social Preferences and the Dynamics of Free Riding in Public Goods

Urs Fischbacher
University of Zurich

Simon Gächter
University of Nottingham,
CESifo and IZA Bonn

Discussion Paper No. 2011
March 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 2011
March 2006

ABSTRACT

Heterogeneous Social Preferences and the Dynamics of Free Riding in Public Goods*

We provide a direct test of the role of social preferences in voluntary cooperation. We elicit individuals' cooperation preference in one experiment and make a point prediction about the contribution to a repeated public good. This allows for a novel test as to whether there are "types" of players who behave consistently with their elicited preferences. We find clear-cut evidence for the existence of "types". People who express free rider preferences show the most systematic deviation from the predicted contributions, because they contribute in the first half of the experiment. We also show that the interaction of heterogeneous types explains a large part of the dynamics of free riding.

JEL Classification: C91, C72, H41, D64

Keywords: public goods games, experiments, voluntary contributions, conditional cooperation, free riding

Corresponding author:

Simon Gächter
School of Economics
Sir Clive Granger Building
University Park
Nottingham NG7 2RD
United Kingdom
Email: simon.gaechter@nottingham.ac.uk

* This paper is part of the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms and the Research Priority Program of the University of Zurich on "The Foundations of Human Social Behavior: Altruism versus egoism". Support from the EU-TMR Research Network ENDEAR (FMRX-CT98-0238) is gratefully acknowledged. Sybille Gübeli, Eva Poen, and Beatrice Zanella provided very able research assistance. We also thank Nick Bardsley, Rachel Croson, Eva Eppler, Armin Falk, Ernst Fehr, Björn Frank, Sally Gschwend, Sybille Gübeli, Radosveta Ivanova-Stenzel, Martin Kocher, Michael Kosfeld, Charles Noussair, Eva Poen, Ekkehart Schlicht, Martin Sefton, Daniel Seidmann, and seminar participants at CES Munich, the Max Planck Institute in Jena, the Universities of Bonn, Copenhagen, Hannover, Konstanz, Magdeburg, and Munich, the Wissenschaftszentrum Berlin, the ESA conference in Amsterdam and the PET04 conference in Beijing for their helpful comments. Simon Gächter thanks the University of St. Gallen, CES Munich and CMPO Bristol for the hospitality he enjoyed while working on this paper.

I. Introduction

Numerous public goods experiments have shown that many people contribute more to the public good than pure self-interest can easily explain. However, an equally important observation is that free riding becomes paramount in repeated public goods experiments (Ledyard 1995).

Explanations of these findings differ in the weight they attach to "social preferences" or "confusion" and "trial-and-error learning". One explanation is that the experimental situation triggers some cues ("social norms" or "heuristics"), which induce inexperienced subjects to make positive contributions initially. Over time, however, subjects become less confused and learn what their best strategy is and free riding – or some minimal contributions due to "warm glow", "altruism", or some other-regarding motives – prevail (e.g., Palfrey and Prisbrey 1996; 1997; Houser and Kurzban 2002; Binmore, forthcoming). By contrast, others argue that social preferences must play an important role in explaining cooperation, some confusion notwithstanding (e.g., Andreoni 1995; Keser and van Winden 2000; Brandts and Schram 2001; Croson 2002; Ashley, Ball and Eckel 2005). However, these latter studies only *indirectly* come to the conclusion that social preferences matter in voluntary contribution games, i.e., by observing data patterns that are apparently inconsistent with income-maximizing behavior and confusion.[1]

In this paper we provide a *direct* test of the role of social preferences in voluntary contribution games. Our direct test involves measuring people's cooperation preferences in one public good environment and making point predictions about behavior in other comparable public good games. Such a direct test is important in our view when one wants to assess the relevance of social preferences for voluntary contributions.

Our direct test consists of two main steps which constitute our two contributions to the literature. Our first contribution is to test whether there are "types" of players in the sense that people's expressed cooperation preferences and actual cooperation behavior are consistent in the following sense: a person expressing the preference for free riding will opt for a free ride in various social dilemma situations where this is in his or her self-interest, while a person with a preference for conditional cooperation will cooperate conditionally in response to others' contributions in comparable public goods situations. To our knowledge, not much is

[1] We will discuss all studies mentioned in this paragraph in more detail in Section IV.

known about the consistency of expressed preferences and actual behavior in different games.[2]

Our second contribution is to quantify the extent to which the typically observed decay in contributions is the result of the interaction of different types of players. The argument which we will test is as follows. If people differ systematically in their contribution preferences and their contribution behavior is consistent with their expressed preferences, then conditional cooperators might start out contributing to the public good but will eventually realize that they get "suckered" by the free riders and will then reduce their contributions to prevent being duped. Thus, contributions to the public good unravel. Notice that this argument contrasts with the trial-and-error learning argument which has no role for social preferences.

The environment of our study is the standard linear public good game. We conduct two different experiments. In the first (which we call the P-experiment), we measure people's preferences toward voluntary contributions in an incentive-compatible way. The P-experiment is a one-shot game to avoid contamination with strategic incentives. Our instrument is a variant of the strategy method that uses the same strategy set as the standard public good game. The main concept behind the P-experiment is to ask subjects how much they will contribute to the public good conditional on *each* of the other group members' possible average contribution (rounded to integers). In the second experiment (the C-experiment), people actually contribute to a public good with random matching (repeated ten times). We also elicit subjects' beliefs about other group members' contributions in the C-experiment. This allows us (i) to assess the relationship between one's own contribution and the expected contributions of others and (ii) to make a point prediction how much this individual will contribute in the C-experiment, given his or her expressed preferences in the P-experiment and stated beliefs in the C-experiment. Our design permits us to assess consistency of expressed preferences and behavior because we elicit people's preferences and observe the

[2] Some previous studies combined questionnaires and experiments or observed people across different games. For instance, Ashraf, Bohnet and Piankov (2004) observe people in trust and dictator games. In the context of public goods provision, Offerman, Sonnemans and Schram (1996), Park (2000) and van Dijk, Sonnemans and van Winden (2002) elicit social value orientations and compare them to behavior in public good environments. They find that the social value orientation is positively correlated with contributions to public goods. Brandts and Schram (2001) use questionnaires to classify people as free riders and cooperators. Carpenter and Seki (2005) observe fishermen in laboratory experiments and in their daily activities and find that social preferences as measured in the experiments and fishing productivity are positively linked. Benz and Meier (2005) observe subjects' donation decisions in the laboratory and of the same persons in a naturally-occurring environment. However, all these studies do not make point predictions about behavior, which is what we do.

same person in another comparable environment. Our design will also allow us to see whether there are systematic deviations from predicted contributions which are specific to preference-type.

Our results confirm the important role of social preferences in voluntary cooperation. First, roughly half of our subjects are conditional cooperators and a quarter of the subjects are free riders. The rest shows more complicated patterns. Interestingly, subjects who experienced the C-experiment before we elicited their preferences in the P-experiment express on average the same preferences as do inexperienced subjects whose preferences we elicited before they played the C-experiment. We see this as supporting evidence for our interpretation of the P-experiment as a measurement of cooperation preferences. Second, we find clear-cut evidence for the existence of types. People classified as conditional cooperators in the P-experiment also behave as conditional cooperators in the C-experiment. People classified as free riders contribute significantly less than do all others. Quite surprisingly, however, some of them did contribute to the public good in the contribution game, but basically only in the first half of the experiment. Third, we show with the help of the elicited preferences from the P-experiment and the beliefs from the C-experiment that the often observed decay in cooperation is indeed to a very large extent due to the interaction of heterogeneously motivated types of players. Thus, trial-and-error learning and reduced confusion are not the sole reasons why cooperation collapses. The decay in contributions results from learning about how others behave *and* the players' social preferences which determine the players' response to the observed behavior of other players. We also find that consistency between expressed cooperation preferences and actual contributions increases over time.

Our paper is structured as follows. We describe our design in Section II and the results in Section III. We discuss our contribution with respect to the related literature in Section IV. We believe that our results also have interesting implications for theory and policy, and discuss them briefly in our concluding Section V.

II. Design and procedures

Recall that we want to (i) measure subjects' cooperative preferences, (ii) use these preferences to predict a subject's contribution to a public good, and (iii) assess the role of preference interaction in explaining the decay of cooperation. To achieve our goals, the main novel feature of our experimental design is a combination of two different but comparable experiments. Table 1 provides an overview of our experimental design.

The first type of experiment (the "P-experiment") applies a variant of the strategy method to elicit people's contribution *preferences* in a public goods game. By contrast, subjects in the second type of experiment actually make *contribution choices* in a standard linear public goods environment (labeled "C-experiment") for ten rounds in the random matching mode ("Strangers" – see Andreoni 1988). All subjects play both types of experiments, but not in the same sequence. For example, subjects first go through the preference elicitation experiment in the P-C sessions before making their contribution choices in an ordinary linear public goods game. Our C-P sessions counterbalance the order of experiments to control for possible sequence effects. The C-P sequence allows for a particularly strong test of measured preferences because people experience ten rounds of decisions in the C-experiment before their cooperation preferences are elicited in the P-experiment.

Table 1: Design overview

	1st experiment	2nd experiment
P-C sessions	Preference elicitation (P) (strategy method)	Ten rounds of linear public good with random matching (C) (Actual contribution choices)
C-P sessions	Ten rounds of linear public good with random matching (C) (Actual contribution choices)	Preference elicitation (P) (strategy method)

We now describe the two experiments in detail. We start with the P-experiments. Our preference elicitation study intends to directly elicit subjects' willingness for conditional cooperation. To what degree are subjects willing to cooperate given other subjects' degrees of cooperation? Being able to observe contribution preferences without using deception requires observing contributions that can be contingent on others' contributions. Fischbacher, Gächter and Fehr 2001 (henceforth FGF) introduced an experimental design that accomplishes this task.[3] The central idea of FGF is to apply a variant of the so-called "strategy method" (Selten 1967). The subjects' main task in the experiment is to indicate how much they want to contribute to the public good *for each average contribution level of other group members*. Before we describe how exactly this is done, we first outline the basic decision situation.

[3] The FGF paper is a methodological contribution that was evaluated with a rather small number of subjects (n=44). We adopt the FGF method here and use it with our n=140 subjects. Ockenfels (1999) developed a similar design independently of FGF. Other researchers follow similar goals but devised other instruments; we discuss them in Section IV.

The basic decision situation is a standard linear public goods game (Ledyard 1995). The subjects are randomly assigned to groups of four people. Each subject is endowed with 20 tokens, which she can either keep for herself or contribute to a "project", the public good. The payoff function is given as

$$\pi_i = 20 - g_i + 0.4\sum_{j=1}^{4} g_j \ , \qquad (1)$$

where the public good is equal to the sum of the contributions of all group members. Contributing a token to the public good yields a private marginal return of 0.4 and the social benefit is 1.6. Standard assumptions therefore predict that all subjects free ride completely, i.e., $g_j = 0$ for all j.

The public good problem was explained to the subjects in the instructions (see Appendix A). Since we want to measure subjects' preferences as accurately as possible, we also took great care to ensure that they understood both the rules of the game and the incentives. Therefore, after subjects had read the instructions, they had to answer ten control questions. The questions tested their understanding of the comparative statics properties of (1), to ensure that subjects are aware of their selfish incentives and the dilemma situation. We did not proceed until all subjects had answered all questions correctly. We can thus safely assume that people understood the game.

Subjects in the P-experiment were asked to make two types of contribution decisions, one as a so-called "unconditional contribution", and the other by filling in a "contribution table". Subjects only had to make a *single* decision on a particular contribution to the public good in the "*unconditional contribution"* (i.e., they had to choose a number between 0 and 20), while they had to indicate their contribution dependent on the average contribution of the other three group members (rounded to integers) in the *contribution table*. Specifically, subjects were shown a table of the 21 possible values of the average contribution of the other group members (from 0 to 20) and were asked to state their corresponding contribution for *each* of the 21 possibilities. Subjects were under no time pressure when they made their decision.

After the subjects had made their unconditional and conditional contribution decisions, we asked them on a new screen to estimate the average of the *unconditional* contributions of their three other group members. We gave the subjects financial incentives for correct estimates.[4]

[4] If their estimation was exactly right, subjects received 3 experimental money units in addition to their other experimental earnings. They received 2 (1) additional money units if their estimation deviated by 1 (2) point(s) from the other group members' actual average contribution, and no additional money if their estimation was off the actual contribution by more than three points.

After all subjects had made all entries, including the estimate, a throw of a die selected one subject of the group, with each subject having the same probability of being chosen. The subjects who were not selected contributed their unconditional contributions, while the selected subject contributed in accordance with his or her contribution table, based on the unselected subjects' average contributions. This random device made both decisions potentially outcome relevant. Hence, subjects had an incentive to take both decisions seriously.[5]

What are the incentives in the P-experiment? It helps to think of our experiment in terms of the following extensive form game played with the strategy method: nature chooses three players who make their contribution decisions simultaneously. The fourth player learns the (rounded) average contribution of the other players and then decides how much to contribute. Since we play this game in the strategy method, all subjects have to make contribution decisions in the role of the fourth player, which is what they do when they fill in the contribution table. All players learn whether they actually are the fourth player or not. If they are not chosen to be the fourth player, however, they do not learn *who* is chosen. The following prediction applies for rational and selfish players: it is optimal for the fourth player to contribute zero – independent of the contributions of the other players. Thus, rational and selfish players should only show "0" entries in their contribution tables under the strategy method. Assuming common knowledge of rationality and selfishness, the players who have to make simultaneous contribution decisions will also contribute zero to the public good. If we lift the assumption of *common knowledge* of rationality, however, the latter prediction no longer necessarily holds. If players assume that a "fourth player" is a "conditional cooperator" who displays a pattern of increasing contributions in her schedule, then a "non-zero" *unconditional* contribution may be optimal. However, only rationality and selfishness are assumed for the prediction of the *conditional* contribution. We are only interested in the contribution schedule and not in the unconditional contribution in this paper.

[5] An example illustrates this mechanism. Assume that the four group members make an unconditional contribution of 4, 6, 8, and 10 tokens, respectively. Assume that the random device determines that the contribution table becomes the payoff-relevant decision for the fourth subject, whose unconditional contribution is 10 tokens, while the unconditional contributions are relevant for the other three group members. Hence, the average of their unconditional contribution is six tokens. Assume that the contribution table of the fourth subject says that she will contribute 5 tokens in case the others contribute 6 tokens, then her contribution to the public good was taken to be five tokens. Thus, the sum of contributions in this example is 23 tokens. Individual payoffs can now be calculated according to payoff function (1).

The experiment was only played *once*, i.e., there were no repetitions and the subjects knew this. The rationale is that we wanted to elicit subjects' preferences, without intermingling preferences with strategic considerations. For example, if a subject chooses a contribution table that is increasing in the average contribution of others in the P-experiment, this choice cannot be due to reputation formation or any kind of repeated game consideration. Instead, we can take it as an unambiguous measure of the subject's willingness to be conditionally cooperative. Since we identify the "type" of a subject with his or her preference, the P-experiments will allow us to classify our subjects into types and see how consistently they behave across the two games.

Subjects in the P-C sessions were informed after finishing the P-experiment that they would play another experiment (at the time they revealed their preferences in the P-experiment they did not know about a further experiment). Subjects were then told that the second experiment (the C-experiment) simply involved playing the basic decision situation ten times, where each group member simultaneously makes his or her contribution choice. We emphasized that the groups of four would be randomly reshuffled in each period. In addition to their contribution decisions, subjects also had to indicate their *beliefs* about the average contribution of the other three group members in the current period. In addition to their earnings from the public good experiment, we also paid subjects based on the accuracy of their estimates. Incentives were the same in each period as in the P-experiment.

We elicited beliefs for two reasons. First, we can assess the correlation between beliefs and contributions, which we expect to differ between types of players. For instance, free riders are expected to have a zero correlation between their beliefs about what others contribute and their own contribution (which is predicted to be zero). In contrast, conditional cooperators are expected to have a positive correlation of beliefs and contributions. Second, as we will explain in detail below, we can use the beliefs and the elicited schedules from the P-experiment to make *point predictions* about an individual's contributions in the C-experiment.

The sequence of experiments was simply reversed in the C-P sessions. Subjects first played ten rounds of the C-experiment in the random matching mode and were then introduced to the P-experiments. The comparison of results from the P-experiments in the C-P sequence with those of the P-C sequence allow us to assess the relevance of experience with the public goods game for elicited cooperation preferences.

All experiments were computerized, using the software z-Tree (Fischbacher 1999). The experiments were conducted in the computer lab of the University of Zurich. Our participants

were undergraduates from various disciplines (except economics) from the University of Zurich and the Swiss Federal Institute of Technology (ETH) in Zurich. We conducted six sessions (three in the P-C sequence and three in the C-P sequence). In five sessions we each had 24 subjects and in one 20 subjects. A post-experimental questionnaire confirmed that participants were largely unacquainted with one another. Our 140 subjects were randomly allocated to the cubicles in each session, where they took their decisions in complete anonymity from the other subjects. On average, subjects earned 35 Swiss Francs (roughly $30, including a show-up fee of 10 Swiss Francs). Each session lasted roughly 90 minutes.

III. Results

The relationship between elicited contribution preferences in the P-experiment and actual cooperation behavior in the C-experiment is the center of our analysis. We therefore organize the presentation of our results as follows. We start our analysis by investigating how expressed cooperation preferences and actual contributions are related at the aggregate level. Our second step is to disaggregate cooperation preferences and to investigate the degree of heterogeneity existing in our subject pool. We then turn to a description of actual cooperation behavior in our third step. The fourth step relates actual cooperation behavior to the expressed preferences at a disaggregate level. Our final step will investigate the dynamics of free riding as a function of heterogeneous preferences.

A. *Expressed cooperation preferences and actual contributions at the aggregate level*

Fig. 1 illustrates our first main result.[6] On average, we find that contributions increase in the other group members' contributions. This is true for both the P-experiment, which elicits contribution preferences as a function of other group members' contributions, and the C-experiment, where we observe actual contributions (over all ten periods) as a function of the belief about others' contributions. In other words, people are on average "conditional cooperators" who (i) express a preference for contributing more the more others' contribute in the P-experiment and (ii) who actually contribute more the more they believe others contribute in the C-experiment. Strikingly, we also find that mean contributions in the P-experiment and the C-experiment are almost identical. This holds in particular in the range of

[6] Since beliefs above 11 occur in less than 10 percent of the cases, we summarizes these observation in a category ">11".

beliefs up to seven tokens, which comprises two-thirds of all observations. For beliefs above seven tokens, actual contributions exceed the expressed contribution preferences.

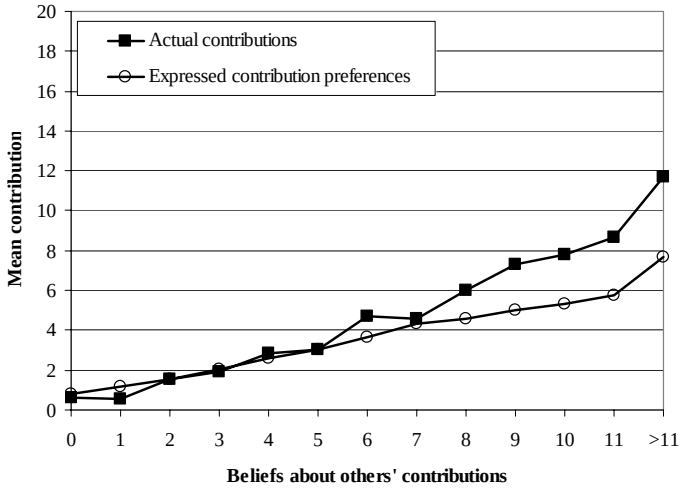


Fig. 1: Expressed contribution preferences (P-experiment) and actual contributions (C-experiment).

We summarize these findings in our first main result.

Result 1: *On average, people are conditional cooperators who contribute more the more they believe others contribute. Expressed contribution preferences and actual contribution behavior coincide in the empirically relevant range of beliefs.*

The significance of Result 1 is two-fold in our view. First, expressed contribution preferences and actual contribution behavior are on average largely consistent with one another. This has not yet been shown before to our knowledge. Result 1 provides a good starting point for our analysis of whether there are "types" of players who behave consistently with their expressed preferences. Later in the paper, we will show that Result 1 is not just an artifact of aggregation but reflects real consistency. Second, Result 1 provides support on a methodological level for the strategy method, which we used to elicit cooperation preferences. The strategy method may be criticized as being psychologically unrealistic since research suggests that a so-called "hot-cold empathy gap" exists in many decisions (see Loewenstein 2000, for example). Decisions may depend on whether they are made in a "cold" state, as in our P-experiments, where subjects have to think about what they do if others contributed a certain amount, or whether they are, as in the C-experiments, actually confronted with contributions that others in fact made in the previous period (the "hot" situation). Result 1

suggests that no "hot-cold empathy gap" exists on average in our context (see Muller, Sefton, Steinberg and Vesterlund 2005 for a similar finding in a related experiment).[7]

In the remainder of the paper, we investigate the relationship between expressed preferences and actual behavior in detail. Our next step in the analysis is to look at the distribution of cooperation preferences.

B. *Heterogeneous preferences*

Recall that we have a complete contribution schedule from each subject that indicates how much he or she is prepared to contribute as a function of others' contribution. A first simple way of characterizing heterogeneity is to look at the slope (of a linear regression) of the schedule and the mean contribution in the schedule. For instance, a free rider's schedule consists of zero contributions for all contribution levels of other group members. Therefore, his slope and mean contribution are zero. An unconditional cooperator, who contributes 20 tokens for all others' contribution levels, has a mean contribution of 20 and a slope of zero. A perfect conditional cooperator, who contributes exactly the amount others contribute, has a slope of one and a mean contribution of 10 tokens. Fig. 2 depicts the results separately for the C-P and the P-C experiments. The *x*-axis shows the slope of the schedules and the *y*-axis the average contribution in the schedule. The dots in Fig. 2 correspond to individual observations, and the size of a dot to the number of observations it represents.

Fig. 2 conveys two main messages. First, there is a large degree of heterogeneity. Free riders (located at 0-0) and perfect conditional cooperators (at 1-10) are relatively the largest group of subjects. We also find a few subjects who contribute an unconditional positive amount (along the *y*-axis, at x=0). A large number of subjects has a positive mean contribution and a positive slope; a few subjects have a negatively-sloped schedule (i.e., they contribute more the less others contribute). Second, the distribution between the C-P and the P-C sessions is very similar. Kolmogorov-Smirnov tests do not allow rejecting the null hypotheses that both means and slopes are equally distributed between the treatments ($p>0.6$).[8] Thus, elicitation of preferences before subjects actually experienced contributions to

[7] The evidence on the hot-cold empathy gap in strategy method experiments and actual game-playing experiments is mixed. For instance, Brandts and Charness (2000) find no difference, whereas Brosig, Weimann and Yang (2003) report differences between methods.

[8] In Fig. 2 we looked at slope and mean contribution of a subject's schedule. However, qualitatively, we get very similar results if we look at Spearman rank order correlation coefficients, linear correlation coefficients, and slopes and intercepts of linear regressions. In all cases p-values of Mann-Whitney tests that compare the C-P and the P-C experiments yield $p>0.275$.

the public good (in the P-C sessions) or after (in the C-P sessions) did not affect the elicited preferences. We see this as a first robustness check.

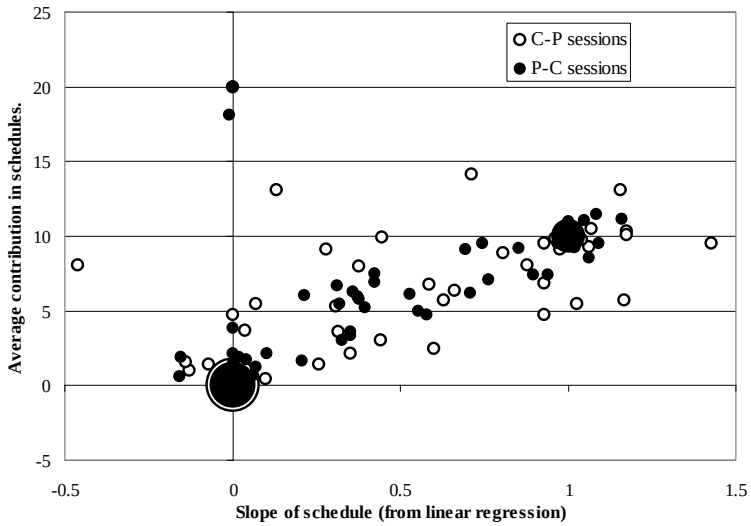


Fig. 2. Heterogeneous contribution preferences: Dots show observations of an individual's slope of a linear regression (x-axis) and mean contribution of his or her schedule in the P-experiment. Filled (open) dots correspond to the P-experiment in the P-C (C-P) sessions. The size of dots corresponds to the number of observations.

For expositional ease, it is also helpful to group subjects with similar patterns. The precise rules for this classification follow FGF and are as follows: All subjects who show either a monotonic pattern with at least one increase or have a positive Spearman rank correlation that is significant at the 1%-level are classified as "conditional cooperators". All subjects who choose to contribute 0 in any case are classified as "free riders". We designate subjects who have a significantly increasing scheme up to some maximum and a significantly decreasing scheme thereafter as "triangle contributors", again using the Spearman rank test at the 1% level as the criterion (FGF call this pattern "hump-shaped contributions"). All subjects who cannot be classified this way fall into the category "others".[9] We find that the distribution of preference types is the same across all six independent sessions (χ^2-test, p=0.510). We therefore pool the data from all sessions. Fig. 3 depicts – separately for each preference type – the mean contributions as a function of the average contribution level of other group members. Table 2 summarizes the results of the classification into preference types separately for the two sequences.

[9] The category "others" contains two "unconditional cooperators" who always contribute 20 and one "negatively conditional cooperator". We include them in the category "others" in order to focus our analysis on the major groups we classified.

A first robustness test is a comparison with FGF. As a comparison of columns "pooled" and "FGF" shows, we replicate FGF's results quite closely. A χ^2-test cannot reject the null hypothesis of an equal distribution of types (p=0.729). This also holds for all separate pair-wise Fisher exact tests (see last column, where all p-values > 0.4).

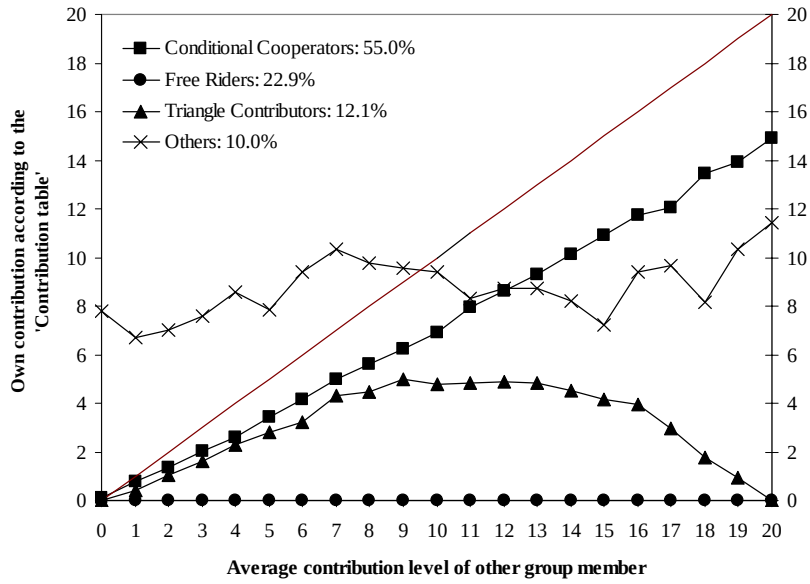


Fig. 3. Average own contribution level for each average contribution level of other group members (diagonal = perfect conditional cooperator).

Subjects only played the P-experiment in FGF. A novel advantage of the present design is that we can test the impact of experience with a C-experiment (in the C-P sessions) on elicited preferences by comparing the C-P and the P-C sessions. We find that the relative sequential order of the P-experiment does not affect the distribution of types under P-C and C-P, respectively (χ^2-test; p=0.481). Pair-wise Fisher exact tests, performed separately for each preference type, all return p-values > 0.30, i.e., the null hypothesis of an equal distribution of types under C-P and P-C cannot be rejected (see penultimate column). This is an important finding for our interpretation that the P-experiment elicits cooperation preferences. It shows that subjects in the C-P sessions who have experienced actual contribution behavior do not express different cooperation preferences than do subjects in the P-C sessions who are inexperienced in actual game playing when they express their preferences. Apparently, our P-experiment, even in the P-C sessions, allowed our subjects immediate access to their cooperation preferences. This finding suggests that our subjects did not need the experience of the C-experiment to discover their preferences (as argued by the "discovered preference hypothesis" (Plott 1996, see also Cooper and Kraker Stockman 2003)). Likewise, the

argument that one-shot environments like our P-experiment do not allow for sufficient learning for eliciting meaningful responses (for instance, Binmore, forthcoming) does not apply to our experiments.

Table 2: Classifying expressed preferences (percentages)

	Our experiments			FGF	p-values of pair-wise Fisher exact tests	
	P-C (n=72)	C-P (n = 68)	Pooled (n=140)	(n=44)	P-C vs. C-P	Pooled vs. FGF
Conditional Cooperator	56.9	52.9	55.0	50.0	0.734	0.772
Free Rider	19.4	26.5	22.9	29.6	0.421	0.419
Triangle Contributor	15.3	8.8	12.1	13.6	0.305	0.512
Other	8.3	11.8	10.0	6.8	0.579	0.686

Note: For the exact definition of types see the main text.

We summarize our findings in our second main result.

Result 2: *We find a considerable degree of heterogeneity in people's contribution preferences. An absolute majority of 55 percent of people are "conditional cooperators"; 22.9 percent are "free riders", and the rest show more complicated patterns. The distribution of preferences is stable between the C-P and the P-C sessions.*

Result 2 is consistent with existing experimental evidence and recent theories of social preferences which also suggest that people differ in these preferences (see, e.g., Camerer 2003, Chap. 2, for an overview of results in experimental economics, and Kelley and Stahelski 1970, for a psychological approach). We conclude from Result 2 that our measurement of preferences is robust and therefore provides a sound basis for investigating behavior in the C-experiments, to which we turn next.

C. *Heterogeneous contribution behavior*

The aggregate results are strongly consistent with previous evidence (e.g., Isaac, Walker and Thomas 1984; Andreoni 1988; Weimann 1994; Laury, Walker and Williams 1995; Croson 1996, 2002; Keser and van Winden 2000; Fehr and Gächter 2000). Temporal patterns and mean contributions in the C-experiments are very similar in the C-P and the P-C sessions (5.2 and 4.5 tokens, respectively). Contributions start between 7 and 9 tokens and decline smoothly and almost identically to an average contribution of two tokens. Beliefs about others' average contribution also decline steadily from 9.8 tokens in period 1 to 3.3 tokens in

period 10. We will look at the temporal development of contributions in detail in subsection III.E. We now concentrate on heterogeneous contribution behavior.

Figure 4 shows a scatter plot of individual slopes of linear regressions of contributions on beliefs on the *x*-axis and average contributions in the C-experiment on the *y*-axis. The dot size corresponds to the number of observations. Thus, the construction of Fig. 4 is similar to Fig. 2, which looks at heterogeneous contribution preferences.

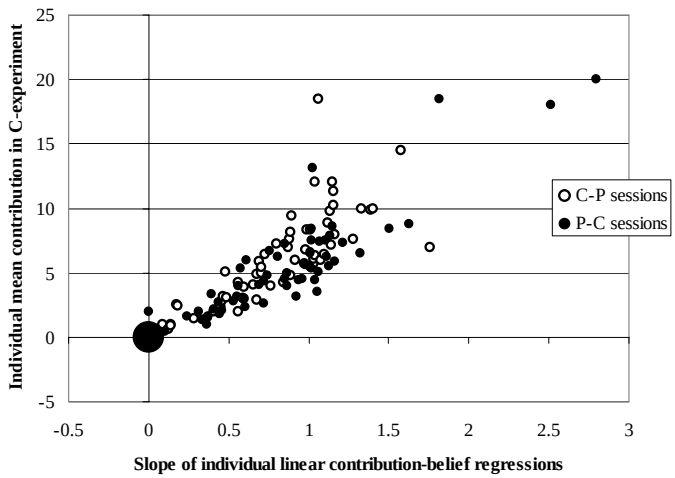


Fig. 4. Heterogeneous contribution behavior: Dots show observations of an individual's slope of a linear regression (x-axis) of contribution_t on belief_t and mean contribution over all periods *t* in the C-experiment. Filled (open) dots correspond to the C-experiment in the P-C (C-P) sessions. The size of dots corresponds to the number of observations.

As in Fig. 2, we distinguish between the C-P and the P-C sessions. We find no sequence effect, neither with respect to average contributions nor with respect to slopes (Kolmogorov-Smirnov tests, $p > 0.27$). Thus, there is no sequence effect in our data and we will therefore pool the sequences for the rest of our analysis.

Fig. 4 reveals considerable heterogeneity in contribution behavior. Individual average contributions (depicted on the *y*-axis) vary between 0 and 20, although most subjects contribute less than ten tokens on average. Twelve percent of all subjects contribute exactly zero in all ten periods. The figure also shows that the strength of conditional cooperation (as measured by the slope) and average contributions are positively correlated. We also find that the individual estimated slopes of the schedules from the P-experiment (see Fig. 2) and the slopes of individual linear contribution-belief regressions in the C-experiments (Fig. 4) are highly significantly positively correlated (Spearman's $\rho=0.39$, $p=0.0000$). We interpret this as a first piece of evidence that expressed cooperation preferences and actual cooperation

behavior are not only correlated at the aggregate level (see Result 1) but also at the individual level. We turn to a systematic investigation of this issue in the following subsection.

D. *Consistency of preferences and contribution behavior*

The next, and indeed central, step in our analysis concerns the contribution behavior of the different preference types in the C-experiments, given their beliefs and their elicited cooperation preferences in the P-experiment. A high degree of consistency of preferences and actual contribution behavior for the different types of preferences (not just at the aggregate level as shown by Result 1) would support the hypothesis that there are indeed "types" of players who behave consistently across different but comparable games.

We begin this analysis with a preliminary observation on first-period beliefs. We investigate whether the different preference types hold different first-period beliefs. First-period beliefs are particularly interesting because they express a subject's intuitive estimate of how others will behave before a subject actually has made any observation about others' behavior. One hypothesis we examine comes from the "false consensus effect" and the observation, made by psychologists (e.g., Kelley and Stahelski 1970), that people have a tendency to believe that others behave similarly to themselves. If this were the case, we should observe that free riders hold consistently lower beliefs than do conditional cooperators, for example. An alternative hypothesis is that people may differ in their perceptions but, since our subjects had an incentive to estimate their group members' contribution correctly, we predict that there should be no difference in beliefs between preference types. The results reject the hypothesis of a "false consensus effect" (see also Engelmann and Strobel 2000) and favor the alternative hypothesis that there are no systematic differences in first-period beliefs between preference types (Kruskal-Wallis test, p=0.4531).

We now look at the contribution behavior of the different preference types. We find that contributions already differ strongly between types in the first period (Kruskal-Wallis test, p=0.0134). Free riders start out with much lower contributions (less than 5 tokens) than all other preference types, whose starting contributions amount to roughly 9 tokens; the other preference types are statistically indistinguishable (Kruskal-Wallis test, p=0.8911). The difference in initial contributions between free riders and all other preference types is highly significant according to a Mann-Whitney test (p=0.0012, two-tailed). Across all periods free riders contribute an average of 2.5 tokens, which is less than all others. Free riders also make the lowest average contributions in all six sessions. Their contributions are significantly lower than those of the other preference types (p=0.0039, Mann-Whitney test, session averages as

observations). Seventy percent of all free rider contributions are exactly zero. By contrast, conditional cooperators contribute on average 5.6 tokens; triangle contributors spend 4.9 tokens on the public good and "others" invest 5.7 tokens. Free riders contribute almost nothing to the public good in the second half of the C-experiment (periods 6-10). More than 80 percent of them contribute exactly zero; their average contribution is 1 token. The average contributions of the conditional cooperators, the triangle contributors, and "others" in periods 6-10 are 4.1, 2.2, and 5.0 tokens, respectively.

Table 3 reports econometric support for the conclusion that the different preference types make different contributions to the public good. The dependent variable is "Contribution in period *t*" and the explanatory variables are (i) dummies for the different preference types (with free riders providing the reference group) and (ii) the belief about others' average contribution in period *t*.

Table 3: Mean contributions of the different preference types (Tobit estimates)

	(1) All periods	(2) Periods 1-5	(3) Periods 6-10	(4) Period 1	(5) Period 10
Belief	1.149	1.084	1.154	1.292	1.224
	(0.115)***	(0.143)***	(0.163)***	(0.159)***	(0.355)***
Conditional Cooperator	6.281	5.293	7.455	5.759	6.788
	(1.451)***	(1.794)***	(1.446)***	(2.102)***	(2.511)***
Triangle Contributor	4.747	4.985	4.673	8.083	4.523
	(1.619)***	(2.035)**	(1.566)***	(2.258)***	(2.622)*
Others	5.900	3.982	8.005	3.746	6.841
	(2.263)***	(2.430)	(2.273)***	(3.152)	(3.110)**
Constant	-9.385	-7.930	-10.346	-9.926	-10.940
	(1.474)***	(1.819)***	(1.615)***	(2.265)***	(2.685)***
Observations	1400	700	700	140	140
$\chi^2(4)$	122.80*	74.96***	73.45***	89.39***	19.95***

Note: Tobit regression with robust standard errors (clustered on subjects). Numbers in parentheses are standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%. Free Riders are the reference group. Dependent variable is "Contribution in period *t*".

This model thus measures to what extent the different types differ in their average contribution behavior when we control for the belief they hold. We estimate five models. The first model comprises all periods. The second and third models look at the first five and last five periods, respectively. The fourth and the fifth model investigate the beginning and the end of the experiment, i.e., period 1 and period 10, respectively.

We find in all models that the higher the belief about the other group members' contribution, the higher is, *ceteris paribus*, the own contribution. Moreover, holding beliefs constant, all types contribute significantly more on average than free riders (the exception are "others" in models (2) and (4)). Model (4) looks at period 1. Period 1 is interesting because it

tells us something about how contributions are determined initially, when subjects have not yet made any observation about others' contributions. We find that the higher beliefs are the higher are initial contributions. The preference-type composition of the subject pool also matters: conditional cooperators and triangle contributors contribute significantly more for a given belief than free riders; "others" do not contribute significantly more than free riders. Finally, model (5) demonstrates that even in the last period subjects contribute more the higher their belief about others' contribution; and, given their belief, all types contribute more than free riders. A further remarkable result from comparing all five models is that the estimated parameters are very similar. This holds in particular for "Belief", "Conditional Cooperators", and the constant. We interpret this as evidence for the stability of the relationship between beliefs and type-specific contributions. Given this stable relationship, contributions decline because on average subjects adjust their beliefs downwards as they learn about other subjects' contributions. We will investigate this mechanism in detail in subsection III.E.

After having established that the different preference types exhibit a different contribution behavior, our next step is to contrast the predicted relationship between beliefs and contributions with the actual contributions. Remember that we asked subjects in the P-experiment how much they would like to contribute *if* the other group members contribute a certain amount *x*. A contribution in the C-experiment, given a subject's belief *x* about others' contribution, is consistent with the preferences in the P-experiment when a subject contributes the same amount he or she indicates in her schedule given others' contribute *x*. Hence, on the basis of Fig. 3 we predict a significantly positive correlation between beliefs and contributions for the conditional cooperators, and no correlation for the free riders, apart from some random deviations. The triangle contributors should have a hump-shaped relation between their stated beliefs and their actual contributions. Beliefs and contributions should be unrelated for "others".

Table 4 reports the Tobit estimates of actual contributions regressed on the beliefs about other group members' contribution.[10] Since we predict this relationship to be hump-shaped for triangle contributors, we also include "Belief squared" as a regressor. We estimate this model separately for each type and for periods 1 – 5 and 6 – 10, respectively.

We find that, with the exception of triangle contributors, beliefs and contributions are significantly correlated for all types in the first half of the experiment. This correlation

[10] In Appendix B we document the raw data that underlie these estimations in form of a scatter plot of contributions on beliefs for the different preference types and for periods 1-5 and 6-10, respectively.

remains positive for all types in the second half, with the exception of free riders. Squared beliefs only matter for triangle contributors and "others" in the second half of the experiment.

Fig. 5 plots the predicted contributions – given the stated beliefs – that follow from the estimation model of Table 4.[11] Again we distinguish between the first and the second half of the experiment. We contrast the predicted contributions in the C-experiment with the predictions gained from the P-experiment.

Table 4: The correlation between own belief and actual contribution of the different types in periods 1-5 and 6-10, respectively (Tobit estimates). Dependent variable is actual contributions.

	Conditional Cooperators		Free Riders	
	Period 1-5	Period 6-10	Period 1-5	Period 6-10
Belief about others' mean contribution	0.853	1.301	3.140	1.955
	(0.505)*	(0.421)***	(1.339)**	(1.799)
Belief squared	0.014	-0.008	-0.095	-0.072
	(0.304)	(0.026)	(0.065)	(0.153)
Constant	-1.851	-3.266	-19.803	-14.516
	(2.032)	(1.272)**	(7.505)***	(5.889)***
Observations	385	385	160	160

	Triangle Contributors		Others	
	Period 1-5	Period 6-10	Period 1-5	Period 6-10
Belief about others' mean contribution	0.244	1.779	1.607	5.697
	(0.657)	(0.472)***	(0.698)***	(1.709)***
Belief squared	0.035	-0.095	-0.031	-0.271
	(0.050)	(0.032)***	(0.039)	(0.083)***
Constant	1.413	-4.393	-5.455	-16.246
	(2.039)	(1.642)***	(3.860)	(6.130)***
Observations	85	85	70	70

Note: Tobit regression with robust standard errors (clustered on subjects). Numbers in parentheses are standard errors. ** significant at 5%; *** significant at 1%. Dependent variable is "Contribution in period *t*".

The estimated contributions for conditional cooperators, given their stated beliefs, are almost identical between the first and the second half of the experiment. However, the estimated contributions exceed the predicted contributions as derived from their expressed preferences in the P-experiment. We find that the estimated relationship between beliefs and contributions for free riders is positive but becomes much flatter in the second half of the

[11] Specifically, we calculate the predicted contribution for each empirically observed belief using a post-estimation command ("predict" in Stata 8.2). We average over those beliefs and plot them in Fig. 5. Since in a few cases we did not observe a belief (e.g., free riders never expressed a belief between 16 and 19 in the C-experiment), there are some gaps in the predicted contributions of free riders, triangle contributors, and "others".

experiment. This finding supports the argument that some cooperation by the free riders is (misperceived) strategic cooperation.[12]

Triangle contributors and "others" have a positively increasing relationship between beliefs and contributions in the first half of the experiment. This relationship becomes hump-shaped in the second half of the experiment for both types. The estimated contributions for the triangle contributors follow the predicted contributions from the P-experiment quite well.

Our next step is to look at single decisions and to assess how they deviate from the predicted choice. Since we have each subject's schedule from the P-experiment and since we also have his or her beliefs in the C-experiment, we can calculate a predicted contribution that follows from a particular subject's schedule and compare this point prediction to his or her actual contribution.

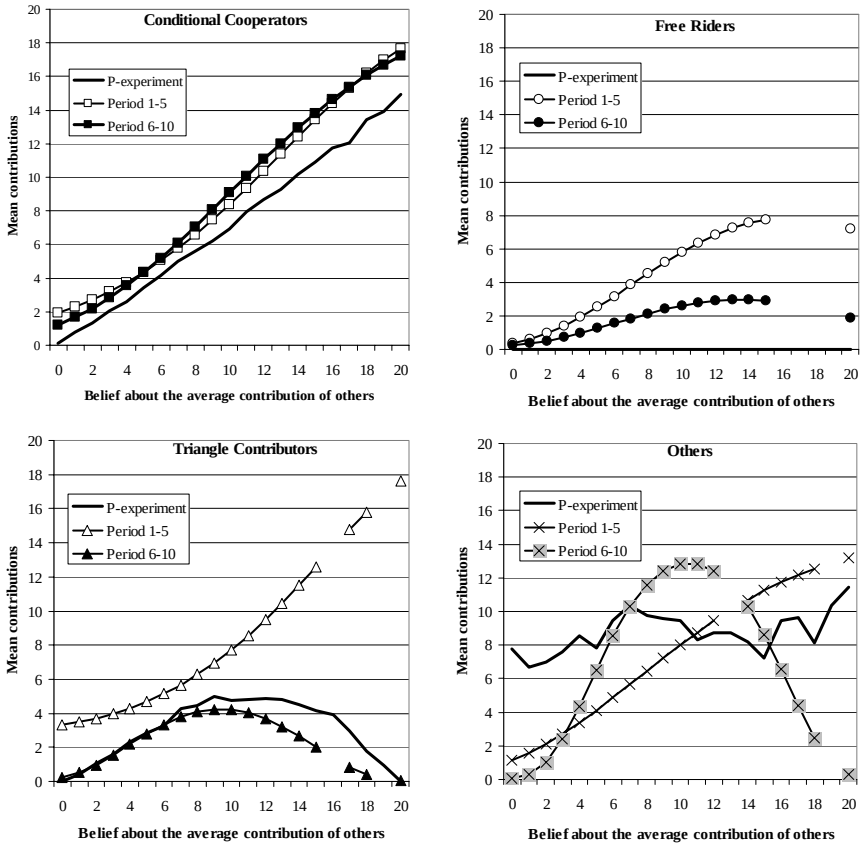


Fig. 5. Estimated contribution (predicted values from a Tobit estimation) for a given belief in periods 1-5 (filled symbols) and periods 6-10 (open symbols) of the C-experiment, and predicted contributions from the P-experiment (solid lines).

[12] See also Sonnemans et al. (1999) for an analysis of strategic behavior by free riders in public goods experiments.

Fig. 6 depicts the distribution of the deviation of actual choices from predicted choices separately for each preference type. We also distinguish between the first and the second half of the experiment. We find that for all types, the modal choice equals the predicted choice, except for the “others”. This holds true for both halves of the experiment.

Fig. 6 also shows that choices are closer to the predicted choice in the second half than in the first half of the experiment for all types (except for “others”). Free riders – for whom consistency is certainly easiest – make the most consistent choices of all types in the sense that 60 to 80 percent of their choices (between periods 1-5 and 6-10, respectively) conform exactly to their predicted choice. The corresponding percentage of cases is between 26 and 38 percent (21 and 31 percent) for conditional cooperators (triangle contributors). In periods 1-5 and 6-10, respectively, “others” exhibit consistent choices in 11 and 8 percent of the cases.

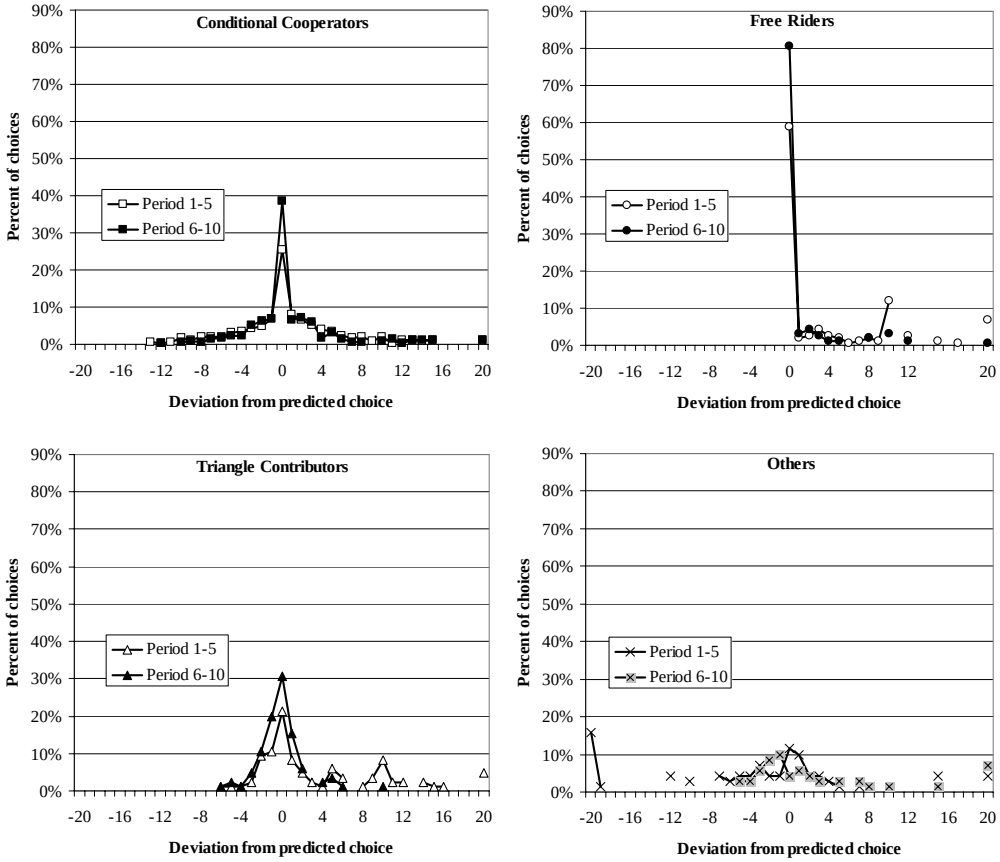


Fig. 6. Percent of choices that deviate from predicted choices.

Insisting on a perfect coincidence of predicted and actual choices is certainly a very demanding criterion, in particular for the non-free riders. If we relax this criterion and call a subject consistent if he or she deviates at most by ±2 tokens (10 percent of the endowment), then consistency rates increase from 52, 63 and 54 percent in periods 1-5 to 65, 88, and 82

percent in periods 6-10 for conditional cooperators, free riders, and triangle contributors, respectively. “Others” demonstrate consistent choices in 34 and 33 percent of the cases, respectively. Thus, the median choice of all types is consistent with the expressed preferences of this type, with the exception of “others”.

Conditional cooperators deviate rather symmetrically from their exact predicted choices. On average, they slightly overcontribute: they deviate by 0.73 tokens (0.76 and 0.69 for periods 1-5 and 6-10, respectively) from their predicted choice. Triangle contributors overcontribute on average by 1.81 tokens. Yet, they become considerably more consistent over time: their average deviation in periods 1-5 drops from 3.67 tokens to -0.06 tokens in periods 6-10. The free riders' overcontribution drops from 3.94 tokens in periods 1-5 to 1.04 tokens in periods 6-10.

Our last question concerns the consistency of *individuals* rather than single choices: How are consistent choices distributed across individuals? We call a choice consistent in this analysis if it deviates at most between ±2 tokens from the predicted choice. Since each individual plays ten rounds in the C-experiment, he or she has ten opportunities to be consistent. Fig. 7 plots the cumulative frequency distribution of individuals who make a consistent contribution in x out of 10 periods. Thus, an individual for whom $x = 10$ makes a consistent choice in ten out of ten rounds; an individual for whom $x = 0$ never makes a consistent choice. As a benchmark we also include the cumulative frequency distribution if people were to choose randomly.

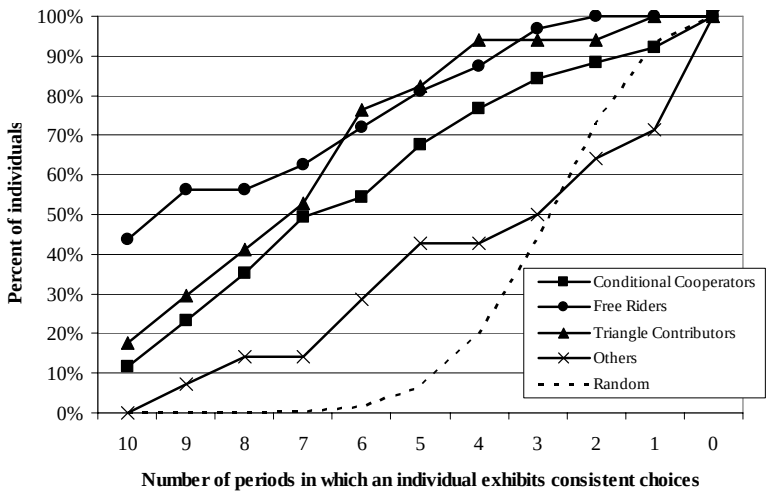


Fig. 7. Distribution of individuals who behave consistently with their expressed preferences for x numbers of periods.

The results show that free riders are the most consistent people: 44 percent of them are consistent (i.e., free ride) in ten out of ten periods. Only relatively few conditional cooperators and triangle contributors (12 and 17 percent, respectively) make a consistent choice in each of the ten rounds. However, except for "others" (who are closest to random choice), the clear majority of individuals (between 55 and 77 percent) makes a consistent choice in at least six periods of the C-experiment. Thus, in the majority of periods the median individual behaves consistently with his or her expressed preferences.

We summarize our main finding in this section in our third main result.

Result 3: *We find a high degree of consistency between expressed preferences and actual contribution behavior for all preference types, with the exception of "others". Specifically, we find* (i) *that for a given belief free riders contribute significantly less than all others;* (ii) *the estimated relationship between beliefs and actual contribution conforms to the predicted relationship for all types, in particular in the second half of the experiment;* (iii) *the modal and median choices are consistent with the predicted choices; and* (iv) *the median individual behaves consistently with his or her expressed preferences in the majority of periods. The strongest* systematic *deviation from predicted choices occurs by free riders, who contribute strategically in the first half of the C-experiment.*

We see Result 3 as evidence for the existence of different "types" of economic agents, who are characterized by different social preferences and behave consistently across different situations. Our final step is to quantify to what extent the interaction of different types can explain the downward trend in contributions.

E. *Heterogeneous preferences and the dynamics of free riding*

Fig. 3 already suggests that contributions will decline over time, because the schedules of all types, with the exception of "others", are consistently below the diagonal. FGF and Fehr and Fischbacher (2004), for instance, argue that this can in principle explain the decay in cooperation. The implicit assumption is that there are types of players who are consistent across games. Result 3 supports this assumption. However, even if there are types of players, this does not say anything about the actual path of contributions and the speed and asymptote of the decline. Thus, our goal in this subsection is to provide rigorous evidence on the role of heterogeneous types in explaining the decay in cooperation.

Since we have the schedules of all subjects, we can assess the type-interaction induced decay in cooperation by simulating the contribution path that would occur if all participants would strictly adhere to their schedules. In other words, we take the first-period contributions as the starting values in our simulation and let contributions evolve as determined by the schedules of our participants, given the exact matching structure that was in place in the C-experiment. We use the average contribution generated in the previous period for the period-by-period updating of simulated beliefs. Fig. 8 shows – separately for each of the six sessions – the resulting average contributions of this simulation, along with the average actual contribution paths. The simulated contributions decrease in all six sessions and largely follow the actual contribution paths.

It is tempting to attribute the difference between the simulated and the actual contribution solely to errors. After all, the identifying assumption of the simulation is that people do not make any errors. The rationale of this argument is that the schedules measure people's error-free contribution *preferences*, as a function of other group members' contributions.[13] Therefore, if we keep the expressed preferences fixed and let them interact we get the "error-free" contribution path. Consequently, the difference between the actual contribution path and the simulated path measures the extent of contributions that are due to errors. However, this is a premature conclusion, because it might overestimate the extent of errors. If, for instance, a free rider contributes more than zero, then this contribution will influence the conditional cooperators' beliefs, for example, about the average contribution level of the other group members. Since the contribution schedule of a conditional cooperator is increasing in the belief about what others contribute, the predicted contribution of a conditional cooperator increases in the free rider's erroneous contribution. Thus, the predicted contribution of all types, except the free riders, is generally endogenous and, via the beliefs, influenced by both the contributions due to the other types' preferences and their "errors". The line with the open diamonds in Fig. 8 depicts the mean predicted contribution over time that results if we evaluate the schedules given the expressed beliefs.

As we see, the predicted contribution that is endogenous to the other players' errors falls as well, but is closer to the actual contribution level than the simulated contribution that assumes errors away. The mean absolute deviation (MAD) of the simulated contributions

[13] This interpretation is justified given the extensive training, the absence of time pressure in the P-experiment, and the absence of sequence effects in elicited preferences in the C-P and P-C experiments.

from the actual contributions is 1.4 tokens, whereas it is 1.1 tokens for the predicted contributions.[14]

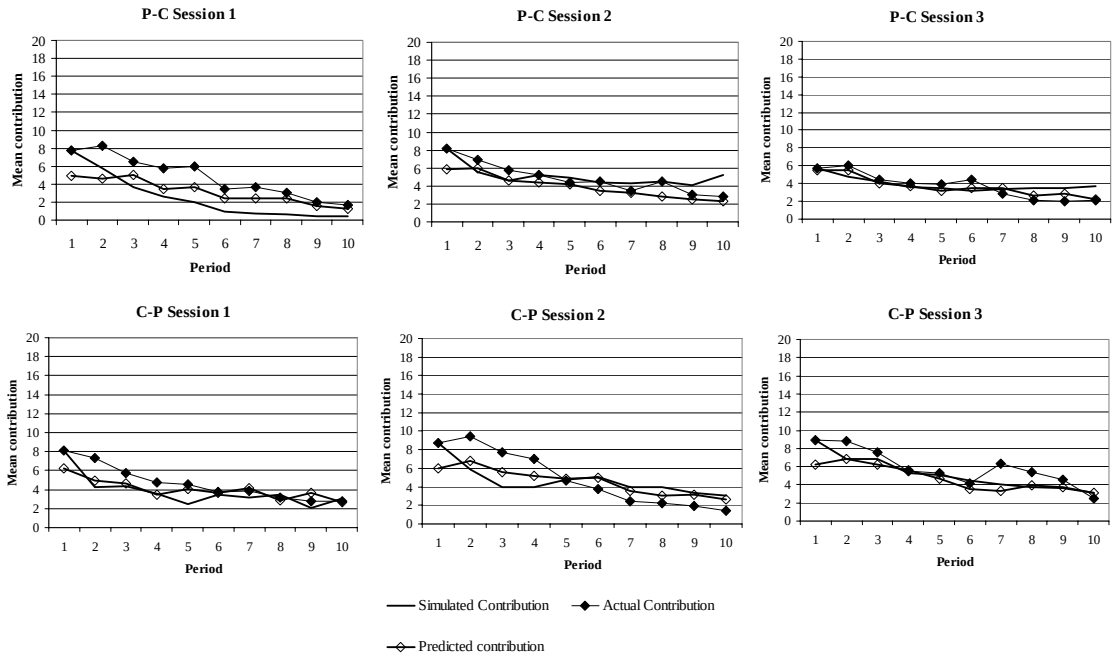


Fig. 8. Mean actual, mean predicted contribution (derived from beliefs and schedules) and mean simulated contribution (resulting from interaction of schedules) across all types over time.

The final step in our analysis concerns the development of the MAD of the predicted contributions from the actual contributions of the four types. Fig. 9 plots the resulting time series for each type and the total MAD.

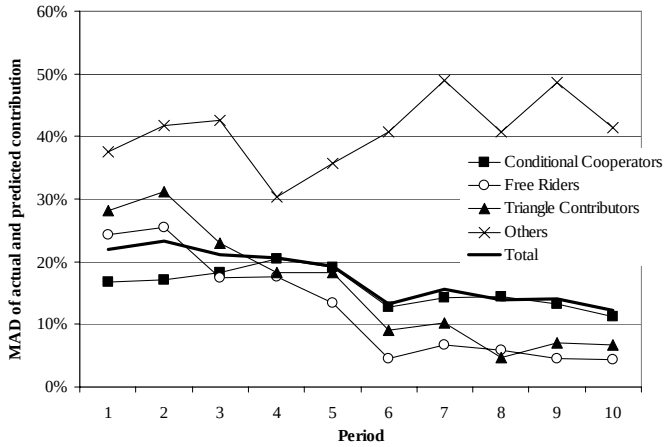


Fig. 9. MAD of actual and predicted contribution in percent of endowment per type and period.

[14] These figures are the averages across all periods. For the simulated contributions, we calculate MAD for each period except period 1, which by construction has a MAD of zero.

MAD clearly decreases over time, except for "others", who start at a MAD level in excess of all others and which decreases over time. By contrast, the consistency of conditional cooperators, triangle contributors, and free riders is very similar over time; ultimately, all become gradually more consistent. The MAD of the free riders and the triangle contributors are already persistently below those of conditional cooperators after period 4.

We now turn to an econometric analysis of the development of the MAD over time. We are particularly interested in estimating the starting and asymptotic values of the deviations of actual and predicted contributions. We estimate the following model:[15]

$$\begin{aligned} AD_{ijt} &= \beta_{1C} D_C (1/t) + \beta_{1F} D_F (1/t) + \beta_{1T} D_T (1/t) + \beta_{1O} D_O (1/t) + \\ &+ \beta_{2C} D_C (t-1)/t + \beta_{2F} D_F (t-1)/t + \beta_{2T} D_T (t-1)/t + \beta_{2O} D_O (t-1)/t + u \end{aligned}$$

where AD_{ijt} is the absolute deviation of predicted and actual contribution in period t of individual i who is of type j. D_j is a dummy for type $j \in \{$*C*onditional cooperator, *F*ree rider, *T*riangle contributor, *O*thers$\}$. The coefficient β_{1j} measures the starting value of the absolute deviation of type j while the coefficient β_{2j} measures the asymptote. As t increases the weight of β_{1j} gets small because $1/t$ goes to zero, while the weight of β_{2j} gets large because $(t-1)/t$ approaches 1. The estimation approach is OLS. We calculate robust standard errors by clustering on subjects. As a benchmark we also estimate a model where we pool the types:

$$AD_{it} = \beta_1 (1/t) + \beta_2 (t-1)/t + \nu \, .$$

The findings are documented in Table 5. The econometric results confirm the impression from Fig. 9. We start with the pooled model. The estimated initial value of AD is 25 percent of the endowment, which is highly significantly different from zero. Absolute deviations are reduced almost by half over time and approach 14.4 percent.

When we distinguish between types, we find that both starting values and asymptotic values of the absolute deviations differ between types. F-tests confirm these conclusions for both the starting values [F(4,5)=30.53; p=0.0008] and the asymptotic values [F(4,5)=28.75; p=0.0000]. The goodness of fit as measured by the R^2 also increases from 0.34 to 0.42. All types have significantly positive starting values. The absolute deviation of free riders and triangle contributors declines asymptotically, as β_2 for these types is an order of magnitude smaller than the starting values. The asymptotic absolute deviation rates are 5.3 and 7.7

[15] The approach is inspired by Noussair, Plott and Riezman (1995, p. 473).

percent of the endowment, respectively. Deviations of conditional cooperators are only slightly (but significantly) reduced; their initial deviation rate is 18.6 percent, which declines modestly to an asymptotic deviation rate of 14.6 percent of the endowment. Only for "others" do deviation rates increase from an initial rate of 36.9 percent to an asymptotic value of 42.5 percent.

Table 5: Development of absolute deviation over time

	Dependent variable: |actual contribution – predicted contribution| in percent of endowment	
	Starting value β_1	Asymptote β_2
Pooled data (no distinction between types)	0.250*** (0.0259)	0.144*** (0.0170)
Observations	1400	
R-squared	0.34	
Distinguish between types:	Starting value β_{1j}	Asymptote β_{2j}
Conditional Cooperators	0.186***	0.146***
	(0.0212)	(0.0188)
Free Riders	0.298**	0.053***
	(0.0725)	(0.0172)
Triangle Contributors	0.349**	0.077**
	(0.0900)	(0.0153)
Others	0.369**	0.425***
	(0.1032)	(0.0954)
Observations	1400	
R-squared	0.42	

Note: Robust standard errors (clustering on subjects) in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

We summarize these findings in our final main result:

Result 4: *The interaction of heterogeneously motivated types explains a large part of the decay of cooperation. We also find that the consistency of expressed cooperation preferences and actual behavior increases over time.*

Result 4 together with the estimated initial contributions (model (4) of Table 3) provide an explanation for the observed contribution pattern in public goods experiments with random matching. Initial contributions are determined by the type composition of the subject pool and the initial beliefs subjects hold; the higher their initial beliefs are the higher are their initial contributions. After subjects observe others' contributions they adapt their beliefs and determine their new contributions given their cooperation preferences. Experience matters in

two ways: first, subjects learn how others behave and on average adjust their beliefs downwards, and second, subjects become more consistent with their preferences over time.

IV. Discussion and related literature

We see our paper as a contribution to the literature on the motivations behind voluntary contributions. In this section, we discuss our contribution relative to some previous studies that share some of the goals of this paper.

Sugden (1984) argued early on for the importance of reciprocity ("a conditional obligation to contribute if others contribute") in voluntary contributions. Our first main result vindicates his theoretical argument by showing that people actually are conditionally cooperative on average. A number of studies, including FGF, have already suggested conditional cooperation as an important motivation in public goods experiments. For instance, Keser and van Winden (2000) and Ashley, Ball and Eckel (2005) find that subjects reduce their contributions when they find out that they have been contributing above average and (slightly) increase their contribution if they were below average. We find similar results in the data of our C-experiment, and such a reaction is consistent with conditional cooperation. Tighter evidence for conditional cooperation comes from studies that elicit beliefs about others' contributions (e.g., Croson 2000, 2002; Neugebauer, Perote, Schmidt and Loos 2005). Like our study (see Result 1), these papers find a positive correlation between beliefs and contributions. These studies, however, only examine aggregate data and draw no inferences about individual differences, which we find to be essential (see Result 2 and Fig. 4). More importantly, the problem with beliefs is that they evolve endogenously and are thus beyond the control of the experimenter. Moreover, a pessimistic conditional cooperator and a free rider exhibit the same behavior, but for different motivational reasons. Thus, the correlation between beliefs and contributions is only an imperfect indicator of preferences. The strategy method that we apply to measure preferences avoids this problem because the contributions of others are fixed by design.

Several studies try to estimate heterogeneity in the population. One of the first studies to do this is Palfrey and Prisbrey (1996, 1997). The main idea is to assign subjects different marginal rates of substitution (MRS) between the public good and the private good. The MRS are changed from round to round. The design allows estimating individual response functions from which altruism, "warm glow" and confusion can be separated. Altruism (i.e., maximizing group payoff) does not explain behavior, whereas "warm glow" (a willingness to contribute a certain constant amount independently of others' contribution – see also Andreoni 1990) is significant and differs between people. Considerably less that half of all

subjects show a significantly positive warm glow effect, however. Goeree, Holt and Laury (2002) find more evidence for altruism than for warm glow and also report considerable individual heterogeneity with respect to altruism. Still, the models by Palfrey and Prisbrey and Goeree et al. (2002) do not consider conditional cooperation, which – according to our Results 2 and 3 – characterizes contribution preferences and behavior of more than 50 percent of the subjects.

Brandts and Schram (2001), FGF, and the present study use the strategy method to draw inferences about heterogeneous motivations. Brandts and Schram (2001) developed a "contribution function approach" which asks subjects to make contributions for ten different MRS. These MRS are such that they define three generic situations: (i) it is a dominant and efficient strategy to contribute; or (ii) it is dominated and inefficient to contribute, or (iii) it is efficient but dominated to contribute. The latter situation is the standard one; the former situations allow identifying erroneous contributions, because all preference types should either contribute (in situation (i)) or not (in situation (ii)). Brandts and Schram (2001) argue that confusion alone cannot explain their data. They categorize 40.6 percent of their subjects as "individualists" because according to their contribution function individualists only contribute in situation (i); 31.8 percent of the subjects make positive contributions in situations (i) and (iii) and are therefore classified as "cooperators"; the rest goes unclassified. The main difference between Brandts and Schram (2001) and the present study is that Brandts and Schram (2001) used the strategy method for different MRS, whereas the present study kept the MRS constant and used the strategy method to elicit how preferences depend on others' contributions.

Several recent studies tried to assess heterogeneity with different methods. Bardsley and Moffatt (2005) use the "conditional information lottery" to vary other's contribution behavior exogenously. They find that 25 percent are free riders, 39 percent are strategists (who only cooperate strategically), 29 percent are reciprocators and 6 percent are altruists. Kurzban and Houser (2005) measure types as follows. First subjects simultaneously make initial contributions and are then repeatedly given the opportunity to revise their contributions for an unspecified (but randomly determined) number of opportunities. Subjects play several such rounds with varying numbers of opportunities for revising contributions. Subjects are then classified statistically by calculating each subject's "linear conditional-contribution profile". Kurzban and Houser (2005) find 20 percent free riders, 13 percent cooperators, and 63 percent reciprocators. Burlando and Guala (2005) use a mixture of methods to classify types: They use the strategy method, value orientation tests, experimental choices, and

questionnaires. They find 32 percent free riders, 35 percent reciprocators, 18 percent cooperators, and 15 percent "noisy" types. Finally, the subjects in Muller et al. (2005) play a two-stage public goods experiment (using a variant of the strategy method). Muller et al. (2005) classify 35 percent as "selfish subjects" who give nothing in the second stage irrespective of the first stage contribution of the other players. 38 percent are "conditional cooperators" who condition their second stage contribution positively on the first stage contribution of the other players.

The frequency distributions of types are hardly comparable because designs differ in too many details. Yet, the fact that all studies find substantial heterogeneity supports one of our main findings, namely that there is heterogeneity in subjects' contribution behavior. The main difference between approaches is that we (and Muller et al. 2005) conceptually distinguish between preferences and actual behavior. To do this, we developed two experiments to measure preferences and actual behavior separately, whereas the other approaches infer motivations from observing patterns of actual contributions. Kurzban and Houser (2005), for instance, infer types by observing how subjects revise their contributions, if given the opportunity. Since subjects did not know how many opportunities for revision they have, decisions might not only reflect preferences, but strategic behavior as well. Thus, the distinction between preferences and actual contributions is not as straightforward as in our design.

Why do contributions in public goods experiments decline over time? Specifically, in addition to investigating the role of learning (e.g., Andreoni 1988; Hichri and Kirman 2004), research explaining the decay has evolved around the issue of separating motivations from mistakes. One of the first papers on this issue is Andreoni (1995). He developed a design that allows him to distinguish "kindness" from "confusion" and concludes that "cooperation is about half kindness and half confusion" (p. 898). Given this result, he explains the decline in cooperation "to be due to frustrated attempts at kindness, rather than learning the free-riding incentives" (p. 900). Our observation on the importance of conditional cooperation (Result 1 and in particular Result 4), supports this conclusion and shows how exactly "frustrated attempts at kindness" lead to the decline in cooperation. Houser and Kurzban (2002) replicated Andreoni's (1995) results by measuring confusion in an experiment where subjects play against a computer and comparing contributions to a public good with human contributors (for a related comprehensive experiment see Ferraro and Vossler 2005). Yet, Houser and Kurzban (2002) and Palfrey and Prisbrey (1996, 1997) attribute the decay in cooperation largely to reduced confusion. Our Result 4 certainly supports the importance of

increased consistency over time (see in particular Fig. 9 and Table 5). Learning about others' behavior *and* preference heterogeneity, and not just reduced confusion alone, explains the decay in our data to a large extent, however. The experiments by Neugebauer, Perote, Schmidt and Loos (2005), Kurzban and Houser (2005), and Gächter and Thöni (2005) support this conclusion.

First, Neugebauer et al. (2005) report the results from two treatments, one in which subjects are informed about others' contributions in each round, and one in which they receive no information. Conditional cooperation and learning about the dominant strategy are possible in the first treatment, while only learning is possible in the second. The authors hypothesize that if learning largely explains the decay of contributions, this decay should also appear in the no-information treatment as well. Their results show, however, that contributions only decline with information but remain roughly constant in the absence of information. Second, Kurzban and Houser (2005) find that the type composition of a group predicts the actual (declining) contribution pattern rather well. Third, an implication of the preference heterogeneity explanation of the decay of cooperation is that conditional cooperators who would be among other *cooperators* and would know this should be able to maintain high contributions, i.e., be able to prevent the decay. Gächter and Thöni (2005) test this argument as follows. Subjects first play a one-shot experiment and are then regrouped according to their cooperativeness. Subjects are then informed about this regrouping and told that they will play the public goods game for ten rounds with the same new people. Gächter and Thöni (2005) find that cooperative subjects who know that they are among other "like-minded" cooperators indeed cooperate at almost the efficient level and without any substantial decay.[16]

A final issue concerns the stability of preferences. We measured the preferences only once in our P-experiment. This raises the issue of whether preferences would be stable against repeated measurement. Muller et al. (2005) address a similar issue in their two-stage experiment, in which subjects play against five different opponents. Muller et al. find that the distribution of strategies is stable but about 28 percent of their subjects change the type of strategy from period to period. It is an open question whether the same kind of instability occurs also in an experiment in which preferences are elicited as in our P-experiment.

[16] See Gunnthorsdottir, Houser, McCabe and Ameden (2001) for an early related experiment and Ones and Putterman (2004) for a more comprehensive discussion of group composition effects.

V. Summary and conclusions

Our goal in this study was to provide a direct test of the role of social preferences in voluntary cooperation. We achieved this by eliciting preferences in one specially-designed game (the "P-experiment") and predicting behavior in ten standard public goods games with random matching (the "C-experiments"). Our most important results are as follows:

First, consistent with previous evidence, we find a high degree of heterogeneity in people's cooperation preferences as well as actual contributions. Our novel finding is that there are "types" of players in the sense that expressed cooperation preferences (as elicited in the P-experiments) and actual contributions (as observed in the C-experiments) are largely consistent with each other. This holds in particular for conditional cooperators. Free riders (and triangle contributors) show the most systematic deviation from their expressed cooperation preferences in the first half of the C-experiment, which is likely due to a (misplaced) strategic attempt to induce others to contribute more. Actual contributions are strongly consistent with predicted contributions in the second half of the C-experiment.

Second, the interaction between heterogeneously motivated types explains the decay often observed in contributions to the public good. This effect is quantitatively important and vindicates previous speculative arguments by FGF and Fehr and Fischbacher (2004) that cooperation is fragile due to preference heterogeneity.

We see our result, namely that there are types of players who are consistent across different but related games, as a first step in understanding heterogeneous motivations. Our research strategy was to make the P-experiments and the C-experiments comparable. This feature allowed us to make *point predictions* about how a consistent individual ("a type") should behave; any other design would have made the consistency check more ambiguous. Moreover, consistency between closely related but different games is a necessary first step to take if one wants to establish the existence of types of players; if we had found largely inconsistent behavior between comparable games, then the existence of types of players would have been in doubt from the start. The broader methodological importance of this result is that one can measure social preferences in one game and predict behavior in other games.

We believe that our findings have significance for theoretical and policy issues. Take the theoretical reasons first. Our finding of a high degree of consistency of preferences and behavior is relevant for the interpretation of recent theories of social preferences (see Sobel 2005 for a recent comprehensive overview). These theories also allow for heterogeneous

preferences but assume that the relevant preference parameters for a given individual are fixed across games. Result 3 supports this crucial assumption.

Second, the results support modeling assumptions that assume different types of players in an incomplete information game. Our findings suggest that, for practical modeling purposes, a concentration on two sorts of types – the free riders and the conditional cooperators, suffices because triangle contributors are very similar to conditional cooperators in the relevant range of actual contributions (and therefore beliefs). Even "others" behave conditionally cooperatively in the C-experiments (compare Fig. 5).

Third, Result 4 reinforces the theoretical prediction by recent models of social preferences that in the absence of punishment opportunities (see, e.g., Fehr and Gächter 2000; Masclet et al. 2003; Carpenter, forthcoming), the co-existence of heterogeneously motivated agents will lead to almost universal free riding behavior. This holds despite the fact that a majority of people is *not* motivated by selfishness. We think this result is important in view of arguments that deviations from income-maximizing Nash-equilibria are largely due to inexperience and lack of trial-and-error learning. Binmore (forthcoming) makes this viewpoint very clear in a recent comment on social preference explanations of experimental results:

> "… There is a huge literature which shows that adequately rewarded laboratory subjects learn to play income-maximizing Nash equilibria in a wide variety of games – provided they have gained sufficient experience of the game and the way that other subjects play".

After Binmore has cited the standard result of eventual free riding in repeated public goods experiments in support of the above argument he argues:

> "I emphasize the standard results in Public Goods games because the orthodox view among mainstream economists and game theorists (…) is not that the learning or trial-and-error adjustment that might take place during repeated play (against a new opponent each time) is a secondary phenomenon to which conclusions may or may not be sensitive. On the contrary, the fact that laboratory subjects commonly adapt their behavior to the game they are playing as they gain experience is entirely central to our position."

Our results from the P-experiment and the C-experiment, which separate preferences and behavior, uncover how experience in the public goods game matters. First, experience matters in the sense that over time contributions become more consistent with expressed preferences for all types (except "others"). Second, experience about others rather than just the game itself matters in particular for conditional cooperators and triangle contributors. These types make their contributions contingent on what they believe others will do; this holds even in the final period of the C-experiment. Since experience tells the non-free rider types that there are free riders types who behave by and large in a selfish income-maximizing way, the non-free rider

types adjust their contributions accordingly. Thus, deviations from the income-maximizing Nash equilibrium are not solely due to inexperience but to social preferences. Yet, after some time all types *behave* like income-maximizing free riders, even though only the free riders types are motivated by selfish income-maximization. Thus, the income-maximizing Nash equilibrium of free riding is attained not only because people "adapt their behavior to the game they are playing as they gain experience" but because the non-free rider types experience being duped by the free riders.

Our observations on the existence of different types of players may also shed light on interesting social issues. For instance, expectations on how others vote can shape voting behavior (Tyran 2004). The work morale that co-workers exhibit might influence behavior at the work place (e.g., Falk and Ichino, forthcoming), while motivations similar to conditional cooperation and free riding might also affect corruption (Abbink, Irlenbusch and Renner 2002). Similarly, conditional cooperation and free rider incentives might also shape the tendency to follow a leader's example in team work (Gächter and Renner 2004; Güth, Levati, Sutter, and van der Heijden 2004; Potters, Sefton and Vesterlund 2004), and others' actions might also govern contributions to charity (e.g., Falk 2004; Frey and Meier 2004; Shang and Croson 2005).

Policy consequences might differ if, in fact, most people are conditional cooperators or free riders rather than if they contribute for reasons of a "warm glow" or errors (see Gächter 2005 for a more comprehensive discussion). The latter are motives that are independent of other people's contribution behavior, whereas conditional cooperation is contingent behavior. For instance, observers of welfare state policies point out that many people hold reciprocity norms that are akin to the conditional cooperation observed in our experiments. Fong et al. (2004) even argue that "people support the welfare state because it conforms to deeply held norms of reciprocity and conditional obligations to others". People resist paying welfare to individuals who could support themselves but who instead cheat or free ride. Norms of reciprocity and conditional cooperation might also influence tax morale; too many cheaters can spoil tax morale. People might be more likely not to cheat on their taxes if others honestly pay theirs (e.g., Besley and Coate 1992; Andreoni et al. 1998; Feld and Tyran 2002). Policies not only affect the incentives people face, but also influence their beliefs about others, which, as our results show, also strongly shape behavior.

References

Abbink, Klaus; Irlenbusch, Bernd, and Renner, Elke. "An experimental bribery game." *Journal of Law, Economics and Organization* 18(2), 2002, pp. 428-454.

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

Andreoni, James. "Impure altruism and donations to public goods: A theory of warm-glow giving." *Economic Journal* 100, 1990, pp. 464–477.

Andreoni, James. "Cooperation in public-goods experiments: kindness or confusion?" *American Economic Review* 85 (4), 1995, pp. 891–904.

Andreoni, James; Erard, Brian, and Feinstein, Jonathan. "Tax compliance." *Journal of Economic Literature* XXXVI, 1998, pp. 818-860.

Ashley, Richard; Ball, Sheryl, and Eckel, Catherine. "Motives for Giving: A Reanalysis of Two Classic Public Goods Experiments." Mimeo, Dept. of Economics, Virigina Tech 2005.

Ashraf, Nava; Bohnet, Iris, and Piankov, Nikita. "Is trust a bad investment?" Mimeo, Kennedy School of Government, Harvard University 2004.

Bardsley, Nicholas, and Moffatt, Peter. "The experimetrics of public goods: Inferring motivations from contributions." CeDEx Working Paper 2005-09, June 2005, University of Nottingham.

Benz, Matthias, and Meier, Stephan. "Do people behave in experiments as in real life? Evidence from donations." Working paper No. 21, 1999, Institute for Empirical Research in Economics, University of Zürich.

Besley, Tim, and Coate, Stephen. Understanding taxpayer resentment. *Journal of Public Economics* 48, 1992, pp. 165-183.

Binmore, Ken. "Economic man – or straw man?" *Behavioral and Brain Sciences.* forthcoming.

Brandts, Jordi, and Charness, Gary. "Hot vs. cold: Sequential responses and preference stability in experimental games." *Experimental Economics* 2, 2000, pp. 227-238.

Brandts, Jordi, and Schram, Arthur. "Cooperation and noise in public goods experiments: applying the contribution function approach." *Journal of Public Economics* 79, 2001, pp. 399–427.

Brosig, Jeanette, Weimann, Joachim, and Yang, Chun-Lei. "The hot versus cold effect in a simple bargaining experiment." *Experimental Economics* 6, 2003, pp. 75-90.

Burlando, Roberto and Guala, Francesco. "Heterogeneous Agents in Public Goods Experiments." *Experimental Economics* 8(1), 2005, pp. 35-54.

Camerer, Colin. *Behavioral Game Theory*. Princeton: Princeton University Press 2003.

Carpenter, Jeffrey. P. "Punishing free-riders: how group size affects mutual monitoring and the provision of public goods." Forthcoming, *Games and Economic Behavior*.

Carpenter, Jeffrey P. and Seki, Erika. "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay." IZA Discussion Paper No. 1697, July 2005.

Cooper, David, and Kraker Stockman, Carol. "The Formation of Social Preferences: An Experimental Study." Mimeo, Case Western Reserve University 2003.

Croson, Rachel. "Partners and strangers revisited." *Economics Letters* 53, 1996, pp. 25-32.

Croson, Rachel. "Thinking like a game theorist: factors affecting the frequency of equilibrium play." *Journal of Economic Behavior and Organization* 41, 2000, pp. 299-314.

Croson, Rachel. "Theories of altruism and reciprocity: evidence from linear public goods games." Mimeo, University of Pennsylvania, 2002.

Engelmann, Dirk, and Strobel, Martin. "The False Consensus Effect Disappears if Representative Information and Monetary Incentives are Given." *Experimental Economics* 3, 2000, pp. 241-260.

Falk, Armin. "Charitable giving as a gift exchange. Evidence from a field experiment." Discussion Paper IZA-DP 1148, 2004.

Falk, Armin, and Andrea Ichino. "Clean evidence on peer pressure". *Journal of Labor Economics*, forthcoming.

Fehr, Ernst, and Fischbacher, Urs. "The nature of human altruism." *Nature* 425, 2004, pp. 785-791.

Fehr, Ernst, and Gächter, Simon. "Cooperation and punishment in public goods experiments." *American Economic Review* 90(4), September 2000, pp. 980-994.

Feld, Lars, and Tyran, Jean-Robert. "Tax evasion and voting: An experimental analysis." *Kyklos* 55(2), 2002, pp. 197-222.

Ferraro, Paul J. and Vossler, Christian, A. "The Dynamics of Other-regarding Behavior and Confusion: What's Really Going on in Voluntary Contributions Mechanism Experiments?" Experimental Laboratory Working Paper Series #2005-001, Dept. of Economics, Andrew Young School of Policy Studies, Georgia State University.

Fischbacher, Urs. "z-Tree, Toolbox for readymade economic experiments: Experimenter's manual." Working paper No. 21, 1999, Institute for Empirical Research in Economics, University of Zürich.

Fischbacher, Urs; Gächter, Simon, and Fehr, Ernst. "Are people conditionally cooperative? Evidence from a public goods experiment." *Economics Letters* 71, 2001, pp. 397-404.

Fong, Christina; Bowles, Samuel, and Gintis, Herbert. "Reciprocity, self-interest, and the welfare state." Forthcoming in: Ythier, J.M., Kolm, S. and Gerard-Varet L.-A., *Handbook on the Economics of Giving, Reciprocity and Altruism.* Amsterdam: Elsevier 2004.

Frey, Bruno S., and Meier, Stephan. "Social comparisons and pro-social behavior. Testing 'conditional cooperation' in a field experiment." *American Economic Review* 94(5), December 2004, pp. 1717-1722.

Gächter, Simon. "Conditional cooperation: Behavioral regularities from the lab and the field and their policy implications". Paper presented at the CESifo Summer Institute on "Psychology and Economics", Venice 18-19 July 2005.

Gächter, Simon, and Renner, Elke. "Leading by example in the presence of free rider incentives." Mimeo, University of Nottingham 2004.

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

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

Gunnthorsdottir, Anna, Houser, Daniel, McCabe, Kevin, and Ameden, Holly. Dispositions, History and Contributions in Public Goods Experiments. Working Paper 99-02, 2001, University of Arizona.

Güth, Werner, Levati, Vittoria, Sutter, Matthias, and van der Heijden, Eline. "Leadership and cooperation in public goods experiments." Mimeo, Max Planck Institute Jena 2004.

Hichri, Walid and Kirman, Alan. "Do individuals learn to play Nash? Evidence from Public Goods Games. Mimeo, IDEP-GREQAM, University of Aix-Marseille III, November 2004.

Houser, Daniel, Kurzban, Robert. "Revisiting kindness and confusion in public goods experiments." *American Economic Review* 92(4), September 2002, pp. 1062-1069.

Isaac, Mark; Walker, James, and Thomas, Susan. "Divergent evidence on free riding: An experimental investigations of possible explanations." *Public Choice* 43, 1984, pp. 113-149.

Kelley, Harold and Stahelski Anthony. "Social interaction basis of cooperators' and competitors' beliefs about others." *Journal of Personality and Social Psychology* 16, 1970, pp. 190-197.

Keser, Claudia, and van Winden, Frans. "Conditional cooperation and voluntary contributions to public goods." *Scandinavian Journal of Economics* 102(1), 2000, pp. 23-39.

Kurzban, Robert, and Houser, Daniel. "Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations." *PNAS* 102(5), 2005, pp. 1803-1807.

Laury, Susan; Walker, James, and Williams, Arlington. "Anonymity and the voluntary provision of public goods." *Journal of Economic Behavior and Organization* 27(3), 1995, pp. 365-380.

Ledyard, John. "Public Goods: A survey of experimental research." In: Kagel, John & Roth, Alvin E. (Eds.), *Handbook of Experimental Economics*. Princeton: Princeton University Press 1995.

Loewenstein, George. "Emotions in economic theory and economic behavior." *American Economic Review* 90(2), May 2000, pp. 426-432.

Masclet, David; Noussair, Charles; Tucker, Steven and Villeval, Marie-Claire. "Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review*, 93(3), June 2003, pp. 366-380.

Muller, Laurent; Sefton, Martin, Steinberg, Richard, and Vesterlund, Lise. "Strategic Behavior and Learning in Repeated Voluntary-Contribution Experiments." CeDEx Working Paper No. 2005-13, July 2005, University of Nottingham.

Neugebauer, Tibor, Perote, Javier, Schmidt, Ulrich and Loos, Malte. "Self-biased conditional cooperation: On the decline of cooperation in repeated public goods experiments". Mimeo, University of Hanover 2005.

Noussair, Charles, Plott, Charles, and Riezman, Raymond. "An experimental investigation of the patterns of international trade." *American Economic Review* 85(3), June 1995, pp. 462-491.

Ockenfels, Axel. *Fairness, Reziprozität und Eigennutz – ökonomische Theorie und experimentelle Evidenz.* Die Einheit der Gesellschaftswissenschaften, Bd. 108, Tübingen: Mohr Siebeck 1999.

Offerman, Theo; Sonnemans, Joep, and Schram, Arthur. "Value orientations, expectations, and voluntary contributions in public goods." *Economic Journal* 106, 1996, pp. 817–845.

Ones, Umut and Putterman, Louis. "The Ecology of Collective Action: A Public Goods and Sanctions Experiment with Controlled Group Formation." Department of Economics Working Paper 2004-01, Brown University.

Palfrey, Thomas R., and Prisbrey, Jeffrey E. "Altruism, reputation and noise in linear public goods experiments." *Journal of Public Economics* 61, 1996, pp. 409-427.

Palfrey, Thomas R., and Prisbrey, Jeffrey E. "Anomalous behavior in public goods experiments: How much and why?" *American Economic Review* 87(5), December 1997, pp. 829–846.

Park, Eun-Soo. "Warm-glow versus cold-prickle: a further experimental study of framing effects on free-riding." *Journal of Economic Behavior and Organization* 43, 2000, pp. 405-421.

Plott, Charles. "Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis." In: K. Arrow, E. Colombatto, M. Perleman and C. Schmidt (eds), *Rational Foundations of Economic Behavior* (pp. 225-250). London: Macmillan 1996.

Potters, Jan; Sefton, Martin, and Vesterlund, Lise. "Leading-by-Example in Voluntary Contribution Games: An Experimental Study." CeDEx Working Paper No. 2004-01, January 2004, University of Nottingham.

Selten, Reinhard. "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes." In: Sauermann, H. (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung* (pp. 136-168). Tübingen: J.C.B. Mohr (Paul Siebeck) 1967.

Shang, Jen, and Croson, Rachel. "Field experiments in charitable contribution: The impact of social influence on the voluntary provision of public goods." Mimeo, Wharton School, University of Pennsylvania 2005.

Sobel, Joel. "Interdependent Preferences and Reciprocity." *Journal of Economic Literature* XLIII, June 2005, pp. 392-436.

Sonnemans, Joep, Schram, Arthur, and Offerman, Theo. "Strategic behavior in public good games: when partners drift apart." *Economics Letters* 62, 1999, pp. 35-41.

Sugden, Robert. "Reciprocity: the supply of public goods through voluntary contributions." *Economic Journal* 94, 1984, pp. 772-787.

Tyran, Jean-Robert. "Voting when money and morals conflict: an experimental test of expressive voting." *Journal of Public Economics* 88, 2004, pp. 1645-1664.

van Dijk, Frans, Sonnemans, Joep, and van Winden Frans. "Social ties in a public good experiment." *Journal of Public Economics* 85, 2002, pp. 275-299.

Weimann, Joachim. "Individual behavior in a free riding experiment." *Journal of Public Economics* 54, 1994, pp. 185-200.

Appendix A: Instructions for the experiment

This is a translation of the original German version. We present the instructions of the P-C experiments here; those of the C-P experiments were adapted accordingly. They are available upon request.

Instructions for the P-Experiment

You are now taking part in an economics experiment financed by the Swiss Science Foundation. If you read the following instructions carefully, you can – depending on your decisions – earn some more money in addition to the 10 Francs, which you can keep in any case. The entire amount of money which you earned with your decisions will be added up and paid to you in cash at the end of the experiment. These instructions are solely for your private information. **You are not allowed to communicate during the experiment.** If you have any questions, please ask us. Violation of this rule will lead to the exclusion from the experiment and all payments. If you have questions, please raise your hand. A member of the experimenter team will come to you and answer them in private.

We will not speak of Francs during the experiment, but rather of points. Your whole income will first be calculated in points. At the end of the experiment, the total amount of points you earned will be converted to Francs at the following rate:

1 point = 35 centimes.

All participants will be divided in groups of four members. **Except for us - the experimenters - no one knows who is in which group.**

We describe the exact experiment process below.

The decision situation

You will learn how the experiment will be conducted later. We first introduce you to the basic decision situation. You will find control questions at the end of the description of the decision situation that help you to understand the decision situation.

You will be a member of a group consisting of **4 people**. Each group member has to decide on the allocation of 20 points. You can put these 20 points into your **private account** or you can invest them **fully or partially** into a project. Each point you do not invest into the project, will automatically remain in your private account.

Your income from the private account:

You will earn one point for each point you put into your private account. For example, if you put 20 points into your private account (and therefore do not invest into the project) your income will amount to exactly 20 points out of your private account. If you put 6 points into your private account, your income from this account will be 6 points. **No one except you earns something from your private account.**

Your income from the project

Each group member will profit equally from the amount you invest into the project. On the other hand, you will also get a payoff from the other group members' investments. The income for each group member will be determined as follows:

Income from the project *= sum of all contributions × 0.4*

If, for example, the sum of all contributions to the project is 60 points, then you and the other members of your group each earn 60 × 0.4 = 24 points out of the project. If four members of the group contribute a total of 10 points to the project, you and the other members of your group each earn 10 × 0.4 = 4 points.

Total income:

Your total income is the sum of your income from your private account and that from the project:

Income from your private account (= 20 – contribution to the project) *+ Income from the project (= 0.4 × sum of all contributions to the project)*
Total income

Control questions:

Please answer the following control questions. They will help you to gain an understanding of the calculation of your income, which varies with your decision about how you distribute your 20 points. *Please answer all the questions and write down your calculations.*

1. Each group member has 20 points. Assume that none of the four group members (including you) contributes anything to the project.
   What will *your* total income be? _____
   What will the total income of the *other* group members be? _____

2. Each group member has 20 points. You invest 20 points in the project. Each of the other three members of the group also contributes 20 points to the project.
   What will *your* total income be? _____
   What will the total income of the *other* group members be? _____

3. Each group member has 20 points. The other 3 members contribute a total of 30 points to the project.

   a) What will *your* total income be, if you – in addition to the 30 points – invest 0 points into the project?
   ***Your* Income** _____

   b) What will *your* total income be, if you – in addition to the 30 points – invest 8 points into the project?
   ***Your* Income** _____

   c) What will *your* total income be, if you – in addition to the 30 points – invest 15 points into the project?
   ***Your* Income** _____

4. Each group member has 20 points at his or her disposal. Assume that you invest 8 points to the project.

   a) What is your total income if the other group members – in addition to your 8 points – contribute another 7 points to the project?
   ***Your* Income** _____

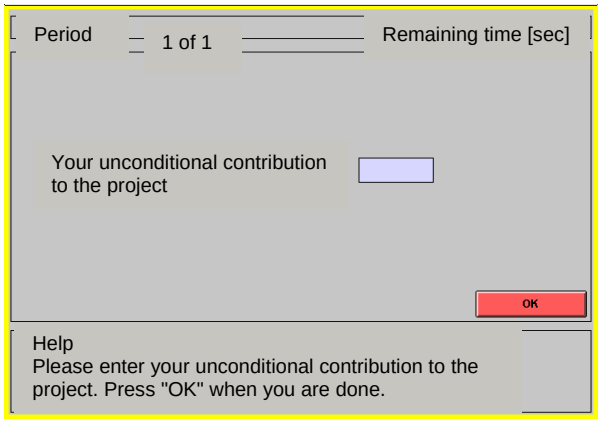
   b) What is your total income if the other group members – in addition to your 8 points – contribute another 12 points to the project?
   ***Your* Income** _____

   c) What is your income if the other group members – in addition to your 8 points – contribute another 22 points to the project?
   ***Your* Income** _____

The Experiment

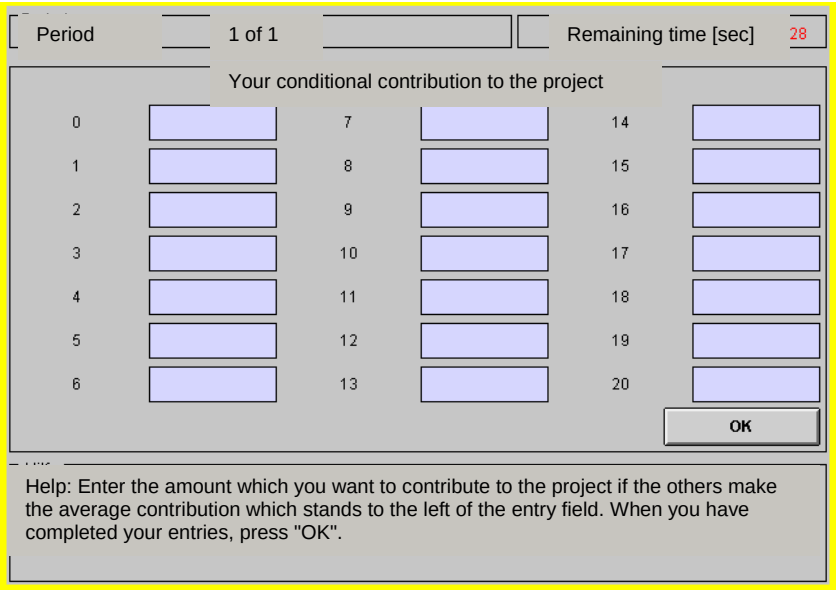
The experiment includes the decision situation just described to you. You will be paid at the end of the experiment based on the decisions you make in this experiment. The experiment will only be conducted **once**. As you know, you will have 20 points at your disposal. You can put them into a private account or you can invest them into a project. Each subject has to make **two types** of decisions in this experiment, which we will refer to below as the **"unconditional contribution"** and **"contribution table"**.

- You decide how many of the 20 points you want to invest into the project in the **unconditional** contribution. Please indicate your contribution in the following computer screen:



After you have determined your unconditional contribution, please click "OK".

- Your second task is to fill in a "contribution table" where you indicate how many tokens **<u>you</u> want to contribute** to the project **for each possible average contribution of the <u>other</u> group members** (rounded to the next integer). You can condition your contribution on that of the other group members. This will be immediately clear to you if you take a look at the following table. This table will be presented to you in the experiment:



The numbers are the possible (rounded) average contributions of the **other** group members to the project. You simply have to insert how many tokens you will contribute to the project into each input box – conditional on the indicated average contribution. **You have to make an entry into each input box**. For example, you will have to indicate how much you contribute to the project if the others contribute 0 tokens to the project, how much you contribute if the others contribute 1, 2, or 3 tokens, etc. You can insert **any integer numbers from 0 to 20** in each input box. Once you have made an entry in each input box, click "OK".

After all participants of the experiment have made an unconditional contribution and have filled in their contribution table, a random mechanism will select a group member from every group. Only **the contribution table will be the payoff-relevant decision for the randomly determined subject**. Only the **unconditional contribution** will be the payoff-relevant decision for the **other three group members** not selected by the random mechanism. You obviously do not know whether the random mechanism will select you when you make your unconditional contribution and when you fill in the contribution table. You will therefore have to think carefully about both types of decisions because both can become relevant for you. Two examples should make this clear.

<u>EXAMPLE 1</u>: Assume that **the random mechanism selects you. This implies that your relevant decision will be your contribution table**. The unconditional contribution is the relevant decision for the other three group members. Assume they made unconditional contributions of 0, 2, and 4 tokens. The average contribution of these three group members, therefore, is 2 tokens. If you indicated in your contribution table that you will contribute 1 token if the others contribute 2 tokens on average, then the total contribution to the project is given by 0+2+4+1=7 tokens. All group members, therefore, earn 0.4×7=2.8 points from the project plus their respective income from the private account. If, instead, you indicated in your contribution table that you would contribute 19 tokens if the others contribute two tokens on average, then the total contribution of the group to the project is given by 0+2+4+19=25. All group members therefore earn 0.4×25=10 points from the project plus their respective income from the private account.

<u>EXAMPLE 2:</u> Assume **that the random mechanism did not select you,** implying that **the unconditional contribution is taken as the payoff-relevant decision** for you and two other group members. Assume your unconditional contribution is 16 tokens and those of the other two group members are 18 and 20 tokens. Your average unconditional contribution and that of the two other group members, therefore, is 18 tokens. If the group member whom the random mechanism selected indicates in her contribution table that she will contribute 1 token if the other three group members contribute on average 18 tokens, then the total contribution of the group to the project is given by 16+18+20+1=55 tokens. All group members will therefore earn 0.4×55=22 points from the project plus their respective income from the private account. If, instead, the randomly selected group member indicates in her contribution table that she contributes 19 if the others contribute on average 18 tokens,

then the total contribution of that group to the project is 16+18+20+19=73 tokens. All group members will therefore earn 0.4×73=29.2 points from the project plus their respective income from the private account.

The random selection of the participants will be implemented as follows. Each group member is assigned a number between 1 and 4. As you remember, a participant, namely the one with the number 11, was randomly selected at the very beginning of the experiment. This participant will throw a 4-sided die **after** all participants have made their unconditional contribution and have filled out their contribution table. The resulting number will be entered into the computer. If participant 11 throws the membership number that was assigned to you, then your contribution table will be relevant for you and the unconditional contribution will be the payoff-relevant decision for the other group members. Otherwise, your unconditional contribution is the relevant decision.

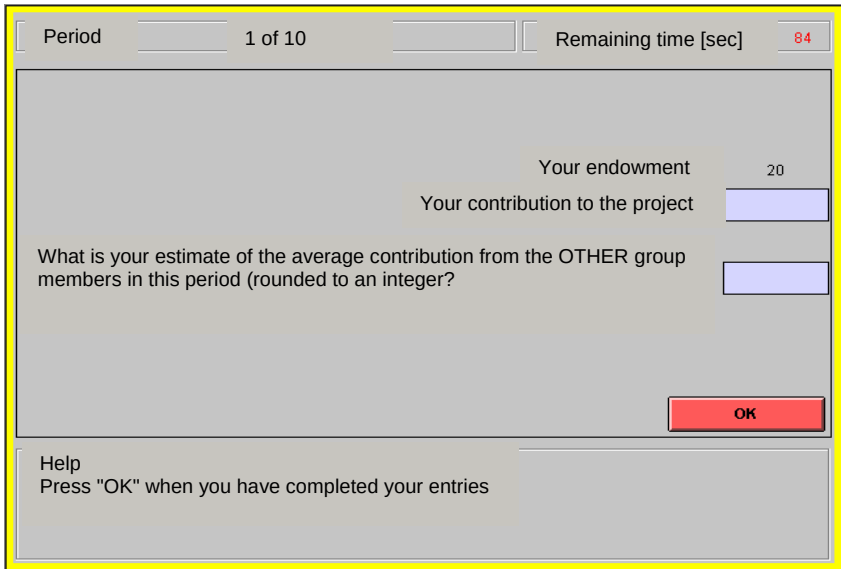
Instructions for the C-Experiment

We will now conduct another experiment. This experiment lasts **10 periods**, in which you and the other group members have to make decisions. As in the other experiment, every group consists of **4 people**. The formation of the group changes at random after every period. **So your group consists of <u>different</u> people in all 10 periods.** The whole experiment is finished after these 10 periods,.

The decision situation is the same as that described on page 2 of the instructions of the previous experiment. Each member of the group has to decide about the usage of the 20 points. You can put these 20 points into your private account or you can invest them fully or partially into a project. Each point you do not invest into the project is automatically placed into your private account. Your income will be determined in the same way as before. Reminder:

Income from your private account (*= 20 – contribution to the project*)
+ Income from the project (*= 0.4 × sum of all contributions to the project*)
Total income
1 point = 7 centimes!

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



As you can see, you have to make two inputs:

1. First you have to **decide on your contribution to the project**, that is, you have to decide how many of the 20 points you want to contribute to the project, and how many points you want to put into your private account. This decision is the same as the unconditional contribution of the previous experiment. You only make unconditional decisions in this experiment. There is **no contribution table**.

2. Afterwards you have to estimate the average contribution to the project (rounded to an integer) of the other three group members of this period. You will be paid for the accuracy of your estimate:

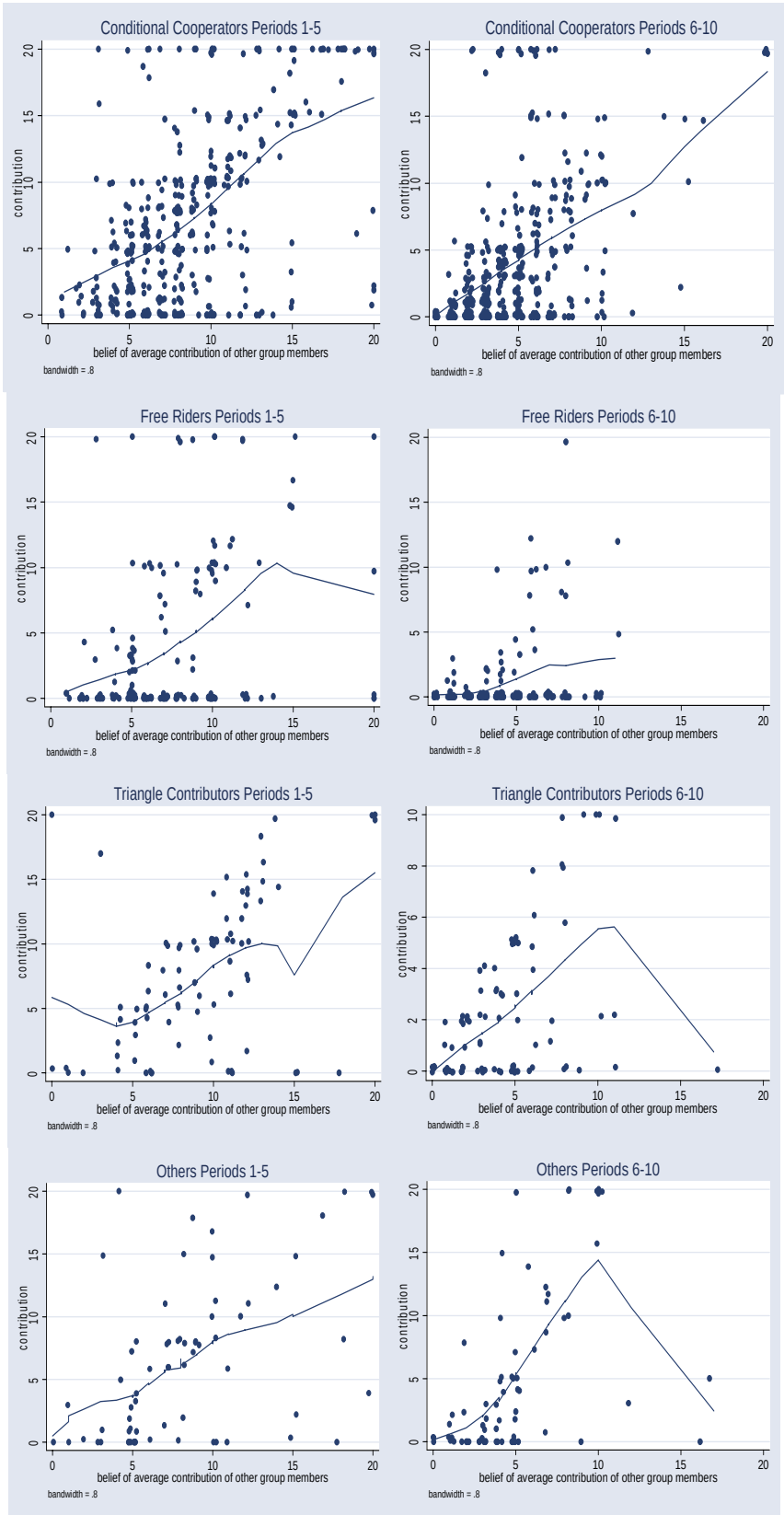
   - If your estimate is exactly right (that is, if your estimate is **exactly** the same as the actual average contribution of the other group members), you will get **3 points** in addition to your other income from the experiment.
   - If your estimate deviates by one point from the correct result, you will get 2 additional points.
   - A deviation by 2 points still earns you 1 additional point.
   - If your estimate deviates by 3 or more points from the correct result, you will not get any additional points.

After these 10 periods are over, the whole experiment is finished and you will receive:

+ your income from the first experiment
+ your income from the second experiment (including your income from your correct estimates)
= total income from both experiments

+ 10 Francs show up fee !

Appendix B: Scatterplots



Note: Solid lines are locally weighted regressions of contributions on beliefs (with least-squares smoothing) using the "lowess" command in Stata 8.2. Observations are jittered (using a jitter-factor of 3).