No. 88

Irenaeus Wolff

When best-replies are not in equilibrium: understanding cooperative behaviour

Research Paper Series
Thurgau Institute of Economics and Department of Economics
at the University of Konstanz

When best-replies are not in equilibrium: understanding cooperative behaviour§

Irenaeus Wolff

Thurgau Institute of Economics (TWI) / University of Konstanz
Hauptstrasse 90, 8280 Kreuzlingen, Switzerland
wolff@twi-kreuzlingen.ch

This version: 16th December, 2013

Abstract:
To understand cooperative behaviour in social-dilemma experiments, we need to understand the game participants play not only in monetary but in preference terms. Does a Nash-prediction based on participants' actual preferences describe their behaviour in a public-good experiment well? And if not, where does the observed behaviour diverge from the prediction? This study provides an environment which allows to answer these questions: when making their contribution decision, participants are informed about their co-players' priorly-elicited conditional-contribution preferences. This induces common knowledge of preferences and thereby leads to direct experimental control over the game participants play. Results show that most people play best-responses to their beliefs. At the same time, beliefs in a third of the cases do not correspond to an equilibrium prediction that is based on the elicited conditional-cooperation preferences. Moreover, more often than not, beliefs are empirically inaccurate. This holds true even in a treatment that gives participants the option to look up the set of equilibria of their game.

Keywords: Public good, social dilemma, Nash-equilibrium, rational beliefs, conditional cooperation, social preferences.
JEL: C72, C92, D83, H41

§I am particularly grateful to Urs Fischbacher for many helpful discussions, the data of earlier experiments, and for the idea for the algorithm to calculate the game-theoretic equilibria in virtually no time during the experiment. Furthermore, I would like to thank the lively research group at the Thurgau Institute of Economics (TWI) and Alexander K. Wagner for helpful comments all along the way, as well as the participants of the ESA European Meeting 2012 and the 15th International Conference on Social Dilemmas for the fruitful discussions.

1 Introduction

In many situations, the Nash-equilibrium is a good predictor for behaviour, while in many other situations, it is not.[1] This paper addresses the question of whether the Nash-equilibrium is a good predictor in public-good experiments once we base the Nash-prediction on participants' actual preferences. This is fundamental to our understanding of cooperative behaviour because virtually all theoretic accounts of cooperative behaviour are based on the notion of Nash-equilibrium.[2]

In one-shot linear public-good experiments, the Nash-equilibrium's predictive power is low if we assume players care only about their own monetary payoff: even in these anonymous one-shot situations, observed average contribution rates are typically around half of players' endowments rather than being close to the predicted zero-contributions.[3] One explanation for this finding could be that contributing participants have some form of other-regarding preferences which may give rise to positive-contributions equilibria. In this case, the Nash-equilibrium may still describe behaviour well, particularly if we take into account participants' uncertainty over their co-players' preferences. Alternatively, it may be that even accounting for participants' preferences and for their beliefs regarding their co-players' preferences, behaviour diverges from the Nash-prediction. As of today, it is widely acknowledged that most people care about others to some extent.[4] What remains to be answered is the question of whether behaviour corresponds to a Nash-prediction that is based on participants' actual preferences, which I will call the *revealed-preference Nash-equilibrium* (RPNE) in the following.

To answer this question, we have to know which game participants play in a typical one-shot public-good experiment. The first thing we need to know is their preferences. Following a long tradition from social psychology as well as behavioural economics, I measure preferences for conditional cooperation: how much participants are willing to contribute to the public good depending on others' contributions.[5] The second condition that is necessary for answering whether participants act according to a RPNE is that we know participants' beliefs about their co-players' preferences. This is because the definition of a game requires

[1] Examples are oligopolies with more than two firms on the one hand (e.g., Huck, Normann and Oechssler, 2004), and common-value auctions (e.g., Bazerman and Samuelson, 1983; Kagel and Levin, 1986), on the other.

[2] E.g., see the recent approaches of Ambrus and Pathak (2011) and Klumpp (2012), and the references cited therein. A notable exception is the theoretic model of sociologist Dijkstra (2012) in which only individual rationality is assumed.

[3] E.g., Ledyard (1995).

[4] E.g., Binmore and Shaked (2010).

[5] Cf. the many references provided in Fischbacher and Gächter (2010), or Fischbacher, Gächter and Quercia (2012). Conditional-cooperation preferences may be a type of social preferences in their own right, or a manifestation of underlying preferences, e.g., for reciprocity.

common knowledge at least of the preference-type distribution. In most public-good experiments, participants face a number of unknown other group members whose preferences they do not know. They may not even know the distribution of preference-types in the population. Scholars have circumvented this issue by looking at late-round behaviour from repeated games, postulating that participants will have had enough interaction experience to learn what kind of preferences others have (e.g., Ambrus and Pathak, 2011). However, the latter is an assumption that has not been subjected to thorough empirical testing.[6] Unfortunately, few studies combine an elicitation of preferences with many repetitions. Hence, we typically do not know whether the equilibrium predictions to be tested rely on the correct assumptions with respect to participants' preferences—not even speaking of participants' beliefs over others' preferences.[7]

This study uses a different approach. By inducing common knowledge of preferences, I control for participants' beliefs over others' preferences directly. Prior to the game of interest, I elicit participants' conditional-contribution preferences, that is, their best-response vectors for all contingencies of the game, using the approach of Fischbacher, Gächter and Fehr (2001) as refined by Cheung (forthcoming). In the subsequent simultaneous linear public-good situation, I reveal each participant's best-response vector to all other members of the participant's group. Thereby, I create an environment with common knowledge of conditional-contribution preferences that allows to test the pure-strategy RPNE predictions in a highly controlled way.[8]

In the experiment, I control for further relevant aspects by design. To minimise the amount of confusion about the situation, I invite only experienced participants to the study. And to control for computational complexity, I add a treatment in which participants may look up the set of pure-strategy RPNE. The second measure sheds further light on why participants may not behave like prescribed by the RPNE prediction.

The experimental data shows that participants deviate less from the RPNE when preferences are common knowledge compared to when they are not. At the same time, average contributions are higher than predicted even under common knowledge of preferences. This is due to frequent positive contribution-levels when the

[6] Ambrus and Pathak (2011) justify this assumption by reporting that experienced participants predict the pattern of average contributions over a 10-period public-good game sufficiently well.

[7] Possibly the only study where a preference-elicitation part is combined with a *long* public-good interaction is Burlando and Guala (2005), with 20 periods of a public-good situation with constant groups that is repeated a week later. They focus on the dynamic pattern of contributions in the second super-game and do not contrast the behavioural outcomes to game-theoretic predictions.

[8] Note that common knowledge of preferences also is an assumption that is made regularly in models involving other-regarding preferences (e.g., Ambrus and Pathak, 2011; Fehr and Schmidt, 1999; Klumpp, 2012). In light of this fact, examining the effects of common knowledge of preferences may have a value in its own right.

only RPNE strategy profile would be omnilateral full defection. While most participants best-respond to their beliefs and choose RPNE actions, many do not believe others will act according to the RPNE prediction. In fact, most participants' beliefs deviate from the empirically accurate contributions combination of their fellow group members. The latter is particularly surprising because participants may state multiple guesses on their co-players' contributions. In sum, these results suggest that in public-good experiments, participants act rationally upon their preferences for conditional cooperation. However, they also suggest that the belief-formation process does not lead to equilibrium beliefs even for experienced participants under common knowledge of preferences. Contrary to what we might think, this is not due to the complexity of calculating the set of RPNE: there is no meaningful difference between the data obtained in the main treatment and the data from the control treatment with optional RPNE-set disclosure.

As may be expected, we find a higher average percentage of inaccurate beliefs in groups with multiple RPNE compared to groups with a single, full-defection RPNE. An additional exploratory analysis reveals that this does not seem to be due to a coordination problem stemming from the multiplicity of equilibria. Rather, the higher belief inaccuracy can be explained merely by two factors: the combination of the preference types within the respective group, and the simple fact that the behaviour of defectors seems to be easier to predict than the behaviour of other types.[9] Surprisingly, as many as one-third of the participants seem to expect substantial contributions even from defectors in groups with a single, full-defection RPNE—who virtually always adhere to the equilibrium strategy.

The remainder of the paper is organised as follows: Section 2 briefly reviews the experimental literature; Section 3 presents testable research hypotheses that form the basis of the data analysis; Section 4 presents the experimental design, Section 5 contains the data analysis and Section 6 concludes.

2 A brief review of the literature

As mentioned above, there is a huge literature on behaviour in public-good settings. Excellent reviews of this literature can be found in Ledyard (1995), Gächter (2007), Gächter and Herrmann (2009), and Chaudhuri (2011). The latter three reviews explicitly address the literature on conditional-cooperation preferences up to the time of their respective publication. More recently, Fischbacher, Gächter and Quercia (2012) show that the preference-elicitation method introduced in Fischbacher, Gächter and Fehr (2001) is behaviourally valid in the sense that contributions in a simultaneous public-good game can be predicted by the preferences elicited in conjunction with participants' beliefs. Reverting the argument and

[9]Different player types do not differ in their ability to predict others' behaviour.

putting it into the context of the present study, Fischbacher, Gächter and Quercia show that participants on average best-respond to their beliefs.

This study adds to Fischbacher, Gächter and Quercia (2012) and the earlier literature on cooperation because so far, no study has compared behaviour or beliefs to the Nash-equilibrium benchmark (nor to the predictions of any other solution concept) induced by the preferences revealed.[10] This comparison is fundamental because much of the results reported in experimental studies stems from a conjoint test of a set of participant preferences and assumptions about how these preferences translate into behaviour, which are often equilibrium assumptions. In interpreting the results, it is then assumed that these assumptions do hold. Naturally, the present study also relies on a conjoint test, but arguably, the assumptions made here are weaker. I only need to assume that participants know how they like to respond to the contributions of others, and that this does not change between two consecutive situations.[11]

Probably closest in approach is the working paper by Healy (2011). He examines the question of how participants' preferences and beliefs over preferences translate situations into games and subsequently, how they determine beliefs and behaviour. This is done for a selection of five different game forms, one of which is a prisoner's dilemma in monetary terms. Healy (2011) finds that often, participants have different games in mind when they face the same game form, in clear violation of the assumption of common knowledge of the preference-type distribution. Obviously, this cannot happen in the main treatments of the present study. Another important difference to this study is that the preferences elicited in Healy (2011) do not include reciprocity motives, which are a central motive in social dilemmas (e.g., Croson, 2007).

3 Research hypotheses

The experiment sets out to identify which conditions of an elicited-preference-based Nash-equilibrium describe participant behaviour well, and which do not.

[10]A working paper by Breuer and Hüwe (2013) addresses this same issue in a slightly different way, resting on a reciprocity model that incorporates over-optimistic beliefs. They report strong experimental support for their model, singling out a 'false-consensus effect' as a main driver for participants' over-optimism. Note that in our main treatments, a 'false-consensus effect' should have no role as preference types are common knowledge.

[11]Recent studies such as Brosig, Riechmann and Weimann (2007) or Blanco, Engelmann and Normann (2011) seem to suggest that the assumption of stable preferences over different games as well as over time may not be warranted under all circumstances. It can be argued, however, that given the overwhelming majority of participants in the present study are playing a best-response to their beliefs in the second game (where best-responses are defined by the first game), the assumption is reasonable in this particular case.

In line with preceding studies such as Fischbacher and Gächter (2010), I use a reduced-form approach to preferences: I look at conditional-contribution preferences, that is, preferences over action profiles. Behind these proximate preferences, there may be concerns over payoffs, inequality, reciprocity, efficiency, or even other concerns participants may have. A pure-strategy *revealed-preference Nash-equilibrium* (RPNE) of the game then is a contribution profile in which each player chooses a contribution in line with her conditional-contribution preferences given her belief on the other players' contributions, a belief that is empirically correct.

To give some examples for RPNE, suppose that three payoff-maximising players are facing a one-shot three-person linear public-good situation. Then, the unique RPNE of the game is the well-known full-defection equilibrium in which no player contributes anything, and this is expected by all three players. Suppose now that the three group members all have fully altruistic preferences. Then, the unique RPNE would be a full-contribution equilibrium in which all players contribute their full endowment, and all three would expect full contributions by the respective others. Finally, suppose the three group members are perfectly conditionally cooperative players.[12] Then, one of the pure-strategy RPNE would be that all contribute some fraction k of their full endowment, $0 \leq k \leq 1$, expecting the others to do the same (which the respective others do). Note that whether the three conditional cooperators are motivated by strong inequity aversion or by reciprocal concerns is irrelevant for the RPNE prediction as long as the best-response correspondences are the same.

The experimental design presented below allows to make precise predictions for participants' contributions, as well as set predictions for their beliefs. This is accomplished as follows: I present participants with a situation in which each player's utility maximum for each contribution combination of the other players is known to all players, a novelty within the domain of games in which social preferences play a role. The elicitation of participants' preferences also enables me to calculate all pure-strategy RPNE of the game players face, and thus to make precise set predictions for each player's contribution and belief. Finally, eliciting players' beliefs allows to make a clear prediction about players' actions.

From here, the general research question is obvious, namely whether participants play a pure-strategy RPNE. As I imagine that participants generally do not play in line with a RPNE, the more intriguing—and potentially more important—questions are those of which equilibrium conditions hold and which do not. To be more precise, I will address whether participants play best-responses to their beliefs; whether they play equilibrium actions; whether their beliefs correspond to

[12] A *perfectly conditionally cooperative* player is defined as a player who always wants to match exactly her fellow group members' average contributions, cf. Fischbacher, Gächter and Fehr (2001).

some RPNE action of the other players; and whether participants' beliefs are empirically correct. Given these research questions, the hypotheses to be tested are clear and derived from basic game theory:

H 1. Participants play a best-response to their belief.

H 2. Participants play RPNE actions.

H 3. Participants' beliefs correspond to the actions their co-players should take according to the RPNE prediction.

H 4. Participants' beliefs are empirically correct.

If hypotheses **H 1–4** hold, the conjoint hypothesis will also hold:

H 5. Participants play a *revealed-preference Nash-equilibrium* of the game.

4 Experimental Design

My main research interest lies in understanding cooperative behaviour in a one-shot public-good situation. Therefore, the central experiment of each session will be a non-repeated, simultaneous linear public-good situation I call the SIMPG-experiment. As I want to create an environment in which participants have common knowledge of preferences in some treatments of the SIMPG-experiment, I need to elicit participants' preferences beforehand. For this purpose, prior to the SIMPG-experiment participants go through a public-good situation with the same parameters but using the strategy method (as introduced by Fischbacher, Gächter and Fehr, 2001, and refined by Cheung, forthcoming). I call this the PREFS-experiment. Finally, in order to prevent participants from signalling in the PREFS-experiment, I include four additional experiments.[13] Hence, participants make decisions in six distinct experiments, with new interaction partners in each of them. Participants are paid for only one randomly chosen experiment and do not get any feedback about others' behaviour before the SIMPG-experiment. Further, each experiment is explained only as soon as it begins. I discuss the precautions taken to prevent signalling behaviour in full detail in Section 4.3.

As this study focuses on the potential reasons for why players may not act in line with the equilibrium prediction, I recruited *experienced* participants. This has

[13]These experiments are a dictator game, two different Ultimatum-Reciprocity-Measure games (Nicklisch and Wolff, 2012), and a reciprocity-willingness-to-pay game. In some of the later sessions, the second Ultimatum Reciprocity Measure was replaced by a trust game. The game parameters are summarised in Table B.16 in Appendix B, for more details, please refer to the screenshots in Appendix C.

two advantages: it should eliminate inexperience as a reason for non-equilibrium play and it should increase participants' understanding of the game. Participants in the experiment had participated in at least one public-good experiment and at least four additional other experiments, with no upper limits. In the following, I describe the SIMPG-experiment and the PREFS-experiment in due detail.

4.1 The SIMPG-experiment

The SIMPG-experiment consists of a simultaneous three-player linear public-good situation with an MPCR $= 0.5$ and an endowment of 20 Euros. Each player has to choose a contribution to the public good from the set $\{0, 4, 8, 12, 16, 20\}$ Euros, which is multiplied by 1.5 and divided equally among the three members of the group, regardless of each group member's own contribution.

The SIMPG-experiment is played under three information conditions. In NOINFO, a benchmark treatment, no further information is given to participants. The INFOTYPE treatment induces common knowledge of best-reply correspondences to test whether the pure-strategy *revealed-preference Nash equilibrium* (RPNE) accurately describes participant behaviour. To this end, participants see the complete conditional-contribution vector from the PREFS-experiment of both other members of their SIMPG group before choosing their public-good contribution. In addition, they are offered a 'calculator device': they can choose repeatedly either of the other players, enter contributions for themselves and the remaining other player, and have the conditional contribution of the chosen player displayed to them (as opposed to finding this information in the 21-row table; see Figure C.II.27 in Appendix C). As a further control treatment, in the INFOEQM treatment participants may additionally click on a button to have the 'stable contribution combinations' (i.e., all pure-strategy RPNE) displayed to them.[14] This treatment controls if potential deviations from the theoretic prediction in INFOTYPE stem from participants' inability to calculate the equilibria.

After the simultaneous public-good situation, participants in all treatments are asked to specify their beliefs on what the other group members have contributed in the SIMPG-experiment. When stating their beliefs, participants are shown their co-players' conditional-contribution vectors from the PREFS-experiment (again). Specifically, they can state up to four contribution combinations. The fewer combinations they state and the further up in their list of stated combinations, the more they earn in case of a hit, For a correct specification of a single combination, they earn another 20 Euros; if the first out of four combinations is correct, they earn 10 Euros.[15] This procedure has two advantages. It allows participants

[14]See Figure C.II.26 in Appendix C for the explanation provided to the participants on what a 'stable contribution combination' stands for.

[15]The full list is: {20}, {15, 10}, {12.5, 8, 6}, {10, 7.5, 5, 3}. The belief-elicitation mechanism was

to express their degree of uncertainty about others' behaviour, and it should limit the impact of the potential hedging problem. This is because participants can state both their true belief and their hedge, in conjunction with a somewhat 'lenient' data analysis in the sense that I will not focus on whether participants *exclusively* state RPNE-beliefs.[16]

4.2 The PREFS-experiment

In the PREFS-experiment, participants face the same three-player linear public-good payoff structure with an MPCR $= 0.5$ and an endowment of 20 Euros as in the SIMPG-experiment. However, the PREFS-experiment differs from the SIMPG in that the former is a sequential game in which two group members move first and the third group member moves last, being informed of the others' choices.

Participants have to decide in either role. First, they specify their first-mover contribution to the public good that is implemented if they are not (randomly) chosen to be the last-moving player. Then, I elicit their last-mover choices using the strategy method: they are presented with all possible combinations of first-mover contributions and asked to specify their 'conditional' contributions.[17] It is because of the importance to elicit the full conditional-contribution vector that contributions are restricted to the set $\{0, 4, 8, 12, 16, 20\}$ Euros in both the SIMPG- and the PREFS-experiments.[18] It is essential to elicit responses to all *combinations* because the players' response to a combination of, e.g., (8,8) may be very different from their response to (0,16).[19] This would be the case if, for example, a participant's utility function conformed to the model of Fehr and Schmidt (1999), with an advantageous-inequality parameter that is strong enough to ensure the participant matches others' contributions when they are equal.[20] I hold the conditional-

introduced already in the dictator-game experiment at the beginning of the session to acquaint participants with the procedure, cf. Table B.16 in Appendix B and Figure C.II.4 in Appendix C.

[16]For an indication that this measure may have helped, note that about 10% of the participants in the INFO treatments included both 0–0 and 20–20 in their list of beliefs (which, of course, does not mean they must be hedging).

[17]The order of the combinations were randomised individually for all players, for two reasons: (i) to make each decision as salient as possible, (ii) to elicit 'smooth' response-patterns only in case preferences gave rise to them.

[18]This design was pioneered by Cheung (forthcoming). The original design by Fischbacher, Gächter and Fehr (2001) elicited contributions conditional on others' contribution average.

[19]Cf. the evidence in Cheung (forthcoming); this could explain why in Fischbacher and Gächter (2010), only half of all participants best-respond to their stated belief in both the first and the last period, compared to 80% in our design. I am grateful to the authors for sharing their data with me.

[20]In our example, the player would choose 8 in response to 8–8 but 0 in response to 0–16. By definition, the parameter of disadvantageous inequality is at least as high as the parameter for advantagous inequality, and so the player will always match the *minimum* of the other players' contributions in a three-player public-good game.

contribution schedules from the PREFS-experiment to be a direct expression of participants' (proximate) preferences. Therefore, I equate schedules and best-response correspondences for the remainder of this article.

4.3 Addressing the signalling-vs.-deception issue

Any study that requires the display of a participant's choices in one situation to this participant's (new) interaction partner in another situation faces a dilemma when there are potential gains from signalling. This dilemma cannot be resolved in principle. Either the participant is informed of the subsequent revelation of his actions—in which case there is a signalling incentive, however large or small—or he has to be deceived in the sense that he might be facing a rude surprise if, for example, a selfish choice is revealed that he took under the conviction that this choice would remain anonymous.

In order not to deceive participants, in all three treatments the instructions included the sentence that "your behaviour from one of the earlier parts will possibly be displayed to other participants in a later part." To address the resulting signalling issue that participants may change their behaviour in the PREFS-experiment for a better SIMPG-outcome, I took the following measures:

(i) as pointed out above, there are 6 experiments so that it is unclear which of them is the experiment from which behaviour may become publicly known (if any). This should reduce strongly the signalling incentive;

(ii) each experiment is described only once the preceding experiment has finished; in particular, participants do not know what situation they would face in the SIMPG-experiment when choosing their actions in the PREFS-experiment. This means that participants would have to play a very undirected signalling strategy if they wanted to do so at all;

(iii) only one experiment is chosen for payment by the public roll of a die.[21] This should make a signalling strategy prohibitively risky: in five out of six cases, they would be signalling for a payoff-irrelevant situation. Also it puts an emphasis on each individual experiment, as do the assignment of new interaction partners between experiments and the choice to explain each individual experiment only once it has started;

(iv) in all six experiments they face a decision where the maximisation of their own payoff stands against socially-oriented choices. This should prevent participants from guessing the experiment in which to signal. Also, because

[21]For a discussion of the theoretic and behavioural properties of this random-lottery mechanism, see, e.g., Bardsley et al. (2009).

they were not acquainted with some or all of the earlier tasks, understanding these tasks should occupy sufficient cognitive resources to prevent participants from devising a signalling strategy over the whole session;

(v) the instructions stress that "the average payoff to be expected from each of the parts is the same."[22] This should make it clear that the potential expected returns from signalling behaviour are not enormous, as would be the case if earlier experiments paid far less than the experiment in which earlier behaviour would be revealed.

The evidence presented in Section 5.1 suggests that these measures successfully cancelled out any signalling incentives.

4.4 Procedures

On the day of the experiment, participants were welcomed and asked to draw lots in order to assign them to a cabin. There, they would find some general explanation on the general structure of the experiment and on the selection of the payoff-relevant experiment (and role, if applicable). The instructions for each experiment were displayed directly on their screen at the beginning and during the corresponding part. The (translated) general and on-screen instructions are gathered in Appendix C.

In the experiment, participants earned on average 27.66 Euros (USD 36) for 75-90 minutes; this included a 2-Euro flat payment for the completion of a post-experimental questionnaire. Altogether, 10 sessions of 24 participants each were conducted at the LakeLab of the University of Konstanz, between January 2012 and January 2013.[23] Unfortunately, there was a severe no-show problem in one of the sessions. In order not to cancel the complete session, additional participants were recruited on the spot. The data from the 4 participants who were not recruited by the standard recruitment procedure are excluded from the ensuing analysis; see Table 1 for an overview.

5 Results

This section is organized as follows: before I put my research hypotheses to a test, I present evidence that the experimental design successfully prevented signalling

[22] This target (at 25 Euros) was close to being met, with average earnings per experiment (before selection) of 20.84 Euros, 26.21 Euros, 27.06 Euros, 22.82 Euros, 23.04 Euros, and 28.37 Euros for experiments one through six.

[23] Two further INFOTYPE-like sessions that did not yet contain the 'calculator device' described above were discarded from the analysis. The data from these sessions does not convey any additional insights and is included in the supplementary materials.

	noInfo	infoType	infoEqm
Number of sessions	2	4	4
Non-excluded participants	48	96	92

Table 1: Overview of the sessions by treatment

in Section 5.1 and give a brief overview of the data of both the prefs- and the simPG-experiments in Section 5.2. In Section 5.3, I present the evidence on whether participants play best-responses to their beliefs (**H 1**) and specify how this depends on the existing types of pure-strategy *revealed-preference Nash-equilibria* (rpne). Then, I answer the question of whether participants play according to the rpne prediction (**H 2**), whether their beliefs correspond to any pure-strategy rpne of the game (**H 3**), and whether they are empirically correct (**H 4**). Taken together, the results show that the predictive power of the rpne is limited in the context examined in this paper, both individually and on the group level (**H 5**). In Section 5.4, I categorise players by their type as proposed by Fischbacher, Gächter and Fehr (2001). This sets the stage for an explorative analysis of how the accuracy of beliefs relates to group composition and individual players' preference types, in terms of both the belief-forming player and the player about whom the belief is formed, in Section 5.5.

5.1 Evidence for successful signalling prevention

Anticipating three observations from Section 5.3, I argue that the provisions to prevent signalling have been sufficient. First, contributions tends to deviate *positively* from the equilibrium predictions based on the participants' conditional-contribution preferences. However, to reap the benefits from others' increased cooperativeness in reaction to signalling behaviour, participants should have inflated their conditional-contribution preferences in the prefs-experiment. In that case, the rpne predictions for the simPG-experiment—calculated using the conditional-contribution preferences from the prefs-experiment—should have shifted upward. Hence, we would expect contributions in the simPG-experiment (which would be based on true preferences) to deviate *negatively* from the predictions under a signalling hypothesis. This is the opposite of what we see.

Second, on a micro-level, in the simPG-experiment the vast majority of the participants play a best-response to their beliefs according to the conditional-contribution preferences elicited in the prefs-experiment. In contrast, had the conditional-contribution preferences been inflated, then the predicted simPG-contribution frequently should deviate from the observed contribution, as contributions in the simPG-experiment are determined by participants' true—and therefore, non-inflated—preferences.

Third, players do not mis-estimate systematically others' contributions. This could mean two things: either there is no signalling going on and participants have a rough idea of what others will contribute; or participants have inflated their conditional-contribution schedules in the PREFS-experiment *and* adjust rationally for the amount of signalling. The observation that SIMPG-behaviour seems to be at least as pro-socially oriented as the conditional-contribution vectors seems to lend support to the first explanation: under signalling in conjunction with rational adjustment of beliefs, we once again should expect a downward-shift in pro-sociality in the SIMPG- compared to the PREFS-experiment, which we do not observe. Hence, participants also do not seem to believe in systematic signalling attempts by others.

As final remark, note that the *ex-post* incentive of a completely 'selfish' player to mimic the behaviour of a conditional cooperator would have been slightly negative *even if the selfish player had known which behaviour would be displayed to others in some of the treatments.*[24] In other words, actual signalling incentives were strongly negative given participants did *not* know whether behaviour would be disclosed, and if so, which.

5.2 Background information: general data overview

5.2.1 Conditional-contribution preferences elicited in the PREFS-experiment

What do participants' conditional-contribution preferences look like? For a first impression, I display average conditional contributions from the PREFS-experiment in Figure 1. As can be seen from the figure, the data are similar to those reported in the literature: on average, participants react to others' contributions by (imperfect) conditional cooperation, and for a given average contribution by the other play-

[24]A simple ordinary-least-squares regression of the sum of others' contributions on my-preference-type dummies yields a coefficient of 5.6 for the conditional-cooperator dummy ($p = 0.004$; 'selfish' being the baseline). Given MPCR $= 0.5$, displaying a conditional-cooperator's rather than a full-defector's contribution schedule yields 2.80 Euros. Because the SIMPG-experiment is chosen with probability $\frac{1}{6}$, and because players are in one of the INFO treatments with probability $\frac{96+96}{48+96+96} = \frac{4}{5}$ (cf. Section 4.4), the cooperative schedule yields 0.38 Euros in expectation. To calculate the signalling costs, note that the average first-mover contribution in the PREFS-experiment was just above 8 Euros, while the average conditional-contributor schedule (cf. Fig. 4 in Section 5.4) can be described roughly by *own contribution* $= 0.9 \cdot$ *others' average contribution*. So, to effectively signal being a conditional contributor, the 'selfish' player would have had to invest an average of 7.2 Euros as the last-moving player in the PREFS-experiment. This would happen in $\frac{1}{3}$ of the cases if the PREFS-experiment were selected (with probability $\frac{1}{6}$), so that expected signalling costs of a completely selfish participant, conditional on knowing when to signal, amounted to 0.40 Euros, which is 2 Cents above the expected benefit.

ers, a lower variance leads to higher contributions (cf. Cheung, forthcoming).[25] At the same time, the average seems to be the stronger criterion: the average conditional-contribution vector is monotonic in the others' average irrespective of the associated variance.

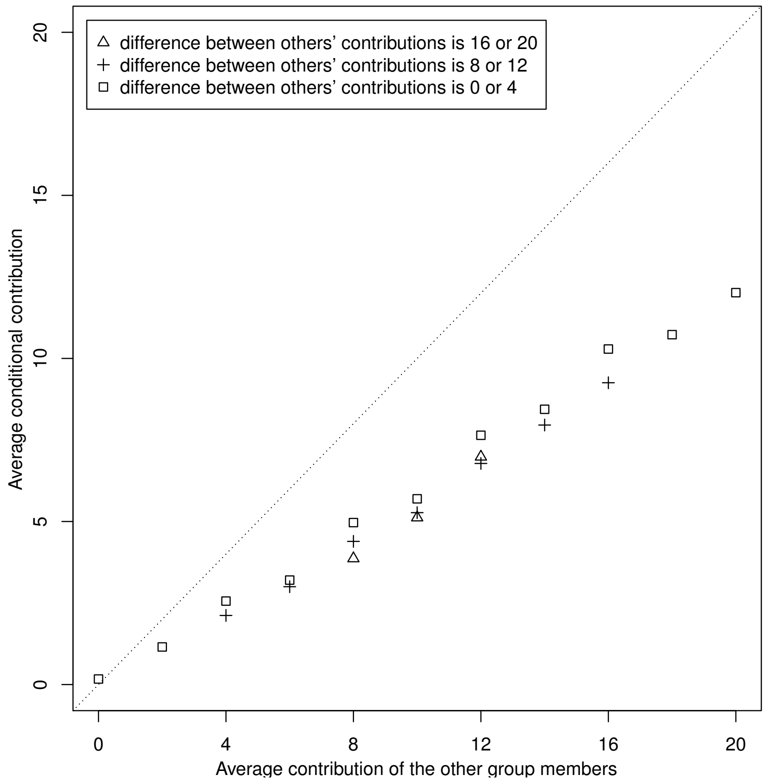

Figure 1: Average contribution conditional on others' average contribution. Note that the difference between others' contributions can be only 0, 8, or 16 if the others' average contribution is divisible by 4, and it can be only 4, 12, or 20, if it is not.

From the literature, we know there is considerable heterogeneity in conditional-contribution preferences. Table 2 shows the distribution of preference-types as introduced by Fischbacher, Gächter and Fehr (2001; for the classification procedure, see Section 5.4), alongside the corresponding distribution in each of the treatments of this study. Except for a somewhat higher (lower) fraction of conditional cooperators (defectors), the preference-type distribution in the PREFS-experiment is similar to Fischbacher, Gächter and Fehr's data.[26]

[25]The only clear exception is for an average contribution of 12: in this case, the combination 4–20 yields slightly higher conditional contributions than the combination 8–16.

[26]χ^2-tests under the hypotheses that any two type distributions (including the study by Fischbacher, Gächter and Fehr, 2001) are from the same underlying distribution yield $p > 0.3$. A

Percentage of...	Conditional cooperators	Defectors	Triangle cooperators	Others
noInfo	58	27	13	2
infoType	67	20	8	5
infoEqm	54	24	14	8
Average, all treatments	60	23	11	6
Fischbacher et al. (2001)	50	30	14	7

Table 2: Distribution of player types.

5.2.2 Contributions and rpne sets in the simPG-experiment

In the simPG-experiment, the average contributions to the public good were 30% in the noInfo treatment, 31% in infoType, and 22% in infoEqm. These contributions have to be seen against the background of the respective *revealed-preference Nash-equilibria* (rpne). I calculate the set of rpne faced by a group in the simPG-experiment in the two following conceptual steps. First, as mentioned in Section 4, I equate the individual group members' conditional-contribution vectors elicited in the prefs-experiment with the best-response correspondence that derives from their underlying preferences. Second, I identify all contribution profiles in which each simPG group member chooses the contribution specified by her best-response correspondence, evaluated at the point defined by her fellow group members' contributions. Hence, every group member plays a best-reponse to the contributions of the other group members.

To categorise the pure-strategy rpne sets for each group from the simPG-experiment, I use the respective level of predicted contributions. The following four categories fully capture the rpne sets that occurred in the simPG-experiment: (i) the only pure-strategy rpne is full-defection, (ii) the rpne set contains a full-defection equilibrium as well as 'moderate-contributions equilibria' (defined by average contributions of at most 8 Euros out of 20); (iii) the rpne set contains a full-defection equilibrium, and at least one 'high-contributions equilibrium' (defined by average contributions being at least two thirds of endowment). These rpne sets have 3.6 equilibria on average, and roughly two thirds contain a full-contributions equilibrium (in 14 out of 23 groups);[27] (iv) the rpne set contains 'only high-contributions equilibria' with average contributions of 8 to 20 Euros in all rpne. Table 3 reports the distribution of rpne sets over the different treatments. Perhaps most importantly, all groups had a non-empty set of pure-strategy rpne,

χ^2-test under the hypothesis that the pooled data from this experiment is drawn from the same distribution as Fischbacher, Gächter and Fehr's data yields $p = 0.658$.

[27] Four groups faced rpne sets that only comprised a full-defection and a full-contributions equilibrium.

and in all but one group, this set included the full-defection equilibrium.

Percentage of groups with...	noInfo	infoType	infoEqm
..only a full-defection equilibrium	75	44	69
..full-defection to moderate-contributions equilibria	0	16	3
..full-defection to high-contributions equilibria	25	41	25
..only high-contributions equilibria	0	0	3

Table 3: Distribution of rpne sets over treatments.

As can be seen from the table, the outcome of the random-matching procedure lead to a fraction of groups with only a full-defection rpne that was clearly lower in infoType (44%) compared to both noInfo (75%) and infoEqm (69%). These differences explain the higher contribution average in infoType when compared to infoEqm, as well as the lack of a difference in contribution averages between noInfo and infoType (more frequent 'over-contributions' in noInfo offset the larger fraction of groups with only a full-defection equilibrium, cf. Figure 2 below). At the same time, they are not a problem for the focus of this study, as except for two observations, this study is not about treatment comparisons, and because I will always condition on the type of rpne set.

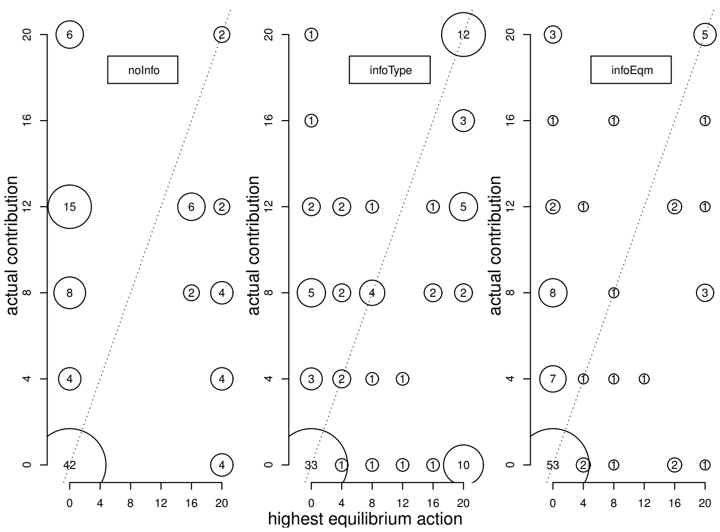

Figure 2: Actual contributions (y-axis), compared to the highest rpne action of the respective player (x-axis), by treatment. Numbers in the bubbles indicate rounded percentages.

Figure 2 contrasts participants' actual contributions with their highest pure-strategy-rpne action, which allows to put participants' contributions in perspective. The figure's left-hand panel refers to the noInfo treatment, the middle panel to infoType, and the right-hand panel to infoEqm. What Figure 2 shows is that ‘over-contributions’ are well and alive when players face low-contribution equilibria only (in particular, when there is only a full-defection equilibrium, see the left-most bubbles in each panel of Figure 2). At the same time, it makes no sense to speak of systematic ‘over-contributions’ when the highest pure-strategy-equilibrium action of a given player is at least 8.[28] Figure 2 also seems to suggest that common knowledge of player types reduces the frequency of ‘over-contributions’ when players face a full-defection equilibrium only.

5.2.3 Beliefs on others’ contributions in the SIMPG-experiment

From Section 5.2.1, we know that a substantial fraction of the participants are neither unconditional defectors nor unconditional cooperators. Hence, in addition to participants preferences, beliefs play an important role in determining contributions. Recall that participants could state a list of up to four (ordered) guesses on their co-players’ actions. On average, they state 2.1 guesses. Table 4 briefly summarises average beliefs in the three treatments. In order not to run out of observations for a number of cells, I group RPNE-set categories (ii)-(iv) under the broader category of ‘multiple RPNE’.

Average ...-stated belief	1st		2nd		3rd		4th	
NOINFO								
Full defection only	2.9	(36)	7.5	(21)	8.7	(11)	10.9	(7)
Multiple-equilibria groups	8.2	(12)	7.8	(8)	11.0	(4)	9.7	(3)
INFOTYPE								
Full defection only	3.6	(42)	5.9	(29)	7.4	(10)	9.3	(3)
Multiple-equilibria groups	8.9	(54)	9.6	(36)	10.0	(12)	9.1	(7)
INFOEQM								
Full defection only	3.0	(64)	6.8	(49)	5.6	(14)	11.3	(6)
Multiple-equilibria groups	7.1	(28)	8.5	(20)	7.8	(8)	6.0	(3)

Table 4: Average belief as a function of the belief’s position on the list of stated beliefs, by treatment and RPNE-set type. Numbers in parentheses indicate the respective numbers of belief-stating participants. Note that beliefs in NOINFO were elicited also displaying co-players’ conditional-contribution schedules. Hence, NOINFO-beliefs are informative only as a control that participants behave similarly in the treatments. They are not informative with respect to contribution decisions.

Table 4 shows a pronounced increase from the first- through to the fourth-stated belief in groups with only a single, full-defection RPNE. The size of the

[28]Note that it does not make sense to speak of ‘under-contributions’ in relation to any RPNE contribution (highest or otherwise) as long as there is a full-defection RPNE.

increase in conjunction with the numbers of observations shows that this effect can only partially be explained by 'equilibrium-belief types' who state full-defection as their only guess.[29] In other words, the further down a guess is in the list of beliefs, the further away it is from the RPNE-prediction for these groups. In contrast, in multiple-RPNE groups, average beliefs follow a rather flat pattern.

These observations set the stage for the analyses pertaining to our research questions, in Section 5.3. Note that in the remainder of the article, I subsume RPNE-set categories (ii)-(iv) under the broader category of 'multiple RPNE' (as in Table 4 above). This is done for ease of exposition only. Tables with detailed information for all categories can be found in Appendix A.

5.3 Equilibrium behaviour in the SIMPG-experiment

H 1: *Participants play a best-response to their stated beliefs.*

Table 5 displays the percentages of participants playing a best-response to their respective beliefs for the treatments with common knowledge of player types.[30] Recall that participants could state up to four guesses of what their co-players would do. Thus, the table is organised in analogy to the following reasoning: if a participant's action is not a best-response to the first-stated belief (which should be the contribution combination the participant considers most probable), at the same time being a best-response to the second- and third-stated beliefs, then this participant will enter the percentage in column "2". If the action is not a best-response to any of the stated beliefs, then the participant will add to the percentage in column "no belief".

Result 1. In both treatments with common knowledge of preferences, the average fraction of players not best-responding to any of their stated beliefs is only about 20%. This fraction is lower in groups with a single (full-defection) *revealed-preference Nash-equilibrium* (RPNE) than in groups with multiple RPNE.

Table 5 shows that in both INFOTYPE and INFOEQM, the fraction of participants not playing a best-response to any of their stated beliefs is lowest for groups in which there is only a full-defection RPNE. In this case, the above fraction amounts to 13–14%. On the other hand, when there is more than one equilibrium, the fraction of contributions that are not best-responses to any belief increases to 25–30%,

[29]To see this, suppose that all 43 single-RPNE-group participants who state a single guess only stated 0. Then, the 99 single-RPNE-group participants who state multiple guesses would have to state an average belief of 4.5 to get an average first-stated belief of 3.15. This, however, is clearly lower than the second-stated beliefs of 7.5 (NOINFO), 5.9 (INFOTYPE), and 6.8 (INFOEQM).

[30]In NOINFO, elicited beliefs are uninformative with respect to best-response behaviour, as the other participants' conditional-contribution vectors were shown during the belief-elicitation procedure also in this treatment.

Best-response to ...-stated belief	1st	2nd	3rd	4th	no belief	# obs.
infoType						
Full defection only	48	29	7	2	14	42
Multiple-equilibria groups	48	19	4	0	30	54
infoEqm						
Full defection only	39	38	11	0	13	64
Multiple-equilibria groups	57	18	0	0	25	28
Random benchmark						
Full defection only	46	5	1	0	48	
Multiple-equilibria groups	23	8	4	0	64	

Table 5: Percentage of participants for whom the belief in the column-title is the highest-ranked belief to which they play a best-response, by treatment and type of RPNE set. The expected frequencies in the random benchmark are calculated using the actual distribution of contributions from both INFO-treatments taken together.

yielding the overall average of about 20% in **Result 1**.[31] These fractions contrast with 48% and 64%, respectively, that would be expected if contributions and lists of stated beliefs were matched randomly.[32]

Having seen that the overwhelming majority of the participants do play a best-response to their beliefs, the natural next question is whether their contributions also coincide with an action from the RPNE set. This question is addressed next.

H 2: *Participants play* RPNE *actions.*

Table 6 provides a detailed picture of participants' absolute deviations from the nearest RPNE prediction, displaying the percentages of participants who deviate by a certain amount for each treatment and each type of RPNE set. The first aspect that calls our attention is that deviations are larger in groups with only a full-defection RPNE compared to groups with multiple RPNE. Note that this is not surprising given in multiple-RPNE groups, the average reference point (i.e., the RPNE action that is closest to a participant's chosen contribution) is closer to the centre of the interval of possible contributions. Also, the fact that the zero-deviation rate is higher in these groups is not very surprising: in a substantial fraction of them, many if not

[31] Two-sided Boschloo-tests on the difference of non-best-response frequencies between single- and multiple-RPNE groups yield p-values of 0.080 for INFOTYPE, 0.175 for INFOEQM, and 0.013 if I pool the data under the assumption that behaviour in both treatments is sufficiently similar.

[32] χ^2-tests yield $p \ll 0.001$ under the hypotheses that contributions and belief lists are matched randomly both for single- and for multiple-RPNE groups.

all possible actions were part of particular equilibria.[33]

Absolute deviation from (nearest) prediction	**0**	4	8	12	16	20	# obs.
NOINFO							
Full defection only	**56**	6	11	19	0	8	36
Multiple-equilibria groups	**67**	25	8	0	0	0	12
INFOTYPE							
Full defection only	**71**	7	12	5	2	2	42
Multiple-equilibria groups	**81**	14	4	0	0	0	54
INFOEQM							
Full defection only	**72**	8	11	3	2	5	64
Multiple-equilibria groups	**79**	14	7	0	0	0	28
Random benchmark							
Full defection only	**54**	9	14	9	4	11	
Multiple-equilibria groups	**74**	15	6	3	2	1	

Table 6: Percentage of participants deviating from the (nearest) prediction by the absolute amount in the column title, by treatment and type of equilibrium set. The expected frequencies in the random benchmark are calculated using the actual distribution of contributions from both INFO-treatments taken together.

Result 2.1. The percentage of RPNE actions tends to be higher in the treatments in which preferences are common knowledge.

To see this, simply compare the according zero-deviation percentages for NOINFO (56 and 67) to the corresponding figures for INFOTYPE (71 and 81) and INFOEQM (72 and 79).[34] Further, note that deviations also tend to be smaller (percentages for almost any absolute deviation are smaller in the common-knowledge-of-types treatments compared to NOINFO).

Result 2.2. About 75% of the participants in the treatments with common knowledge of player types play a pure-strategy RPNE action.

[33]The average numbers of equilibria in each type of equilibrium set are 1 (full defection only), 2.3 (full defection to moderate contributions), 3.4 (full defection to high contributions, and 3 (high contributions only), without notable treatment differences.

[34]A one-sided Boschloo-test for the comparison between NOINFO and INFOTYPE (INFOEQM/the pooled data from both common-knowledge treatments) yields p-values of 0.077 (0.051/0.043) for full-defection-only groups, and of 0.110 (0.188/0.114) for multiple-equilibria groups. Using data from all groups, the p-values are 0.011 (0.036/0.011). I use one-sided tests here because in NOINFO, it is not clear where equilibrium beliefs should come from, and therefore, actions need not correspond to predictions that are based on the actual preferences of the matched players; in contrast, there is no uncertainty about the game in INFOTYPE and INFOEQM, and hence, no inherent reason for a deviation from the (set) prediction.

As can be seen from the column printed in bold, in groups that face only a full-defection RPNE, this number is a little lower with 71–72%. This is clearly above the random-contributions benchmark of 54%.[35] What is noteworthy is that when deviations occur, they are not necessarily small: in single-RPNE groups, deviations of 8 or higher make up for roughly three fourths of all non-zero deviations.

In multiple-RPNE groups, 79–81% of the participants play a RPNE action, which is only slightly more than the 74% under a random-contributions benchmark. The large benchmark percentage stems from the fact that in these groups, many participants' sets of RPNE actions made up for a large fraction of the set of alternatives.[36]

Result 2.3. Under common knowledge of best-reply correspondences, providing players with the possibility to display all possible RPNE does not increase the percentage of RPNE-action choices.

Result 2.3 is evident from comparing the INFOTYPE and INFOEQM percentages of zero-deviations in Table 6 for the respective types of equilibrium sets (71 vs 72 for single- and 81 vs 79 for multiple-RPNE groups). Also, the non-zero deviations do not decrease.[37]

Returning to the broader picture, we have seen that most participants do play a best-response to their beliefs, and almost as many choose a RPNE action. Does this mean that participants are in equilibrium? A first necessary condition for this would be that participants' beliefs correspond to RPNE actions of the other group members. Whether this is indeed the case will be analysed in the following paragraphs.

H 3: *Participants' beliefs correspond to the actions their co-players' should take according to the* RPNE*-prediction.*

An overview of how participants' beliefs correspond to any of the equilibria they face can be found in Table 7. Table 7 should be read in analogy to the following reasoning: if a participant's first-stated belief does not correspond to the other players' actions from any of the pure-strategy RPNE but the second- and third-stated beliefs do, then this participant will enter the percentages in columns "2", "3", and "multiple RPNE beliefs". If none of the stated beliefs corresponds to a RPNE, then the participant will add to the percentage in column "no RPNE belief".

[35] This random benchmark is based on the concept that participants use a choice-generating process that need not lead to uniform randomisation and therefore uses the distribution of contributions in the INFO-treatments as a basis. For groups facing a single, full-defection RPNE, a χ^2-test yields a $p = 0.006$ under the random-choice hypothesis.

[36] For groups facing multiple-RPNE, the χ^2-test yields $p = 0.381$.

[37] Note that 90% of the participants in INFOEQM did look at the equilibria. Out of the remaining 9 participants, 6 played an RPNE-action, 2 deviated by 4, and 1 by 16 Euros. In 8 cases, the nearest RPNE prediction was 0, the remaining participant 'correctly' chose a contribution of 4 Euros.

Belief... is a RPNE belief	1	2	3	4	multiple RPNE beliefs	no RPNE belief	# obs.
INFOTYPE							
Full defection only	60	10	2	0	–	29	42
Multiple-equilibria groups	56	24	2	7	15	28	54
INFOEQM							
Full defection only	59	5	5	0	–	31	64
Multiple-equilibria groups	43	18	14	4	14	39	28
Random benchmark							
Full defection only	26	4	1	0	–	68	
Multiple-equilibria groups	27	8	2	1	0	64	

Table 7: Percentage of participants for whom the belief in the column-title is predicted by a RPNE, by treatment and type of RPNE set. Note that by construction, row-wise percentage sums may add up to more than 100%. The random benchmark assumes different belief-generating processes for first, second, third, and fourth beliefs, but that these processes are the same in single- and multiple-RPNE groups. This makes sense given the focus lies on whether RPNE determine beliefs.

Result 3. Under common knowledge of preferences, roughly one third of the participants do not state any belief that would correspond to others' actions in one of the existing *revealed-preference Nash-equilibria.*

As can be seen from the last data column of Table 7, 29–31% of the participants do not believe in the RPNE prediction even when there is only a full-defection equilibrium. The fact that this figure is not different between INFOTYPE and INFOEQM (29% vs 31%) indicates again that this is not due to participants not being able to calculate the existing RPNE. Interestingly, there is no indication that the fraction of players whose beliefs do not correspond to their co-players' RPNE actions is any different when we look at groups with multiple RPNE (for INFOTYPE this is obvious, with 29% vs 28%; for INFOEQM, a two-sided Boschloo-test yields $p = 0.283$).

Result 3 already suggests the equilibrium assumption about players' beliefs is too strong, despite the fact that participants' beliefs correspond to a RPNE far more often than chance would predict.[38] However, the equilibrium concept makes yet another assumption, namely that participants are able to predict others' choices. The following analysis reveals that this assumption is even less accurate.

H 4: *Participants' beliefs are empirically correct.*

Table 8 provides an overview of whether participants hold empirically accurate

[38]Binomial tests on the frequencies of observing that no stated belief is a RPNE belief under the hypothesis that beliefs are randomly matched to RPNE sets yield $p \ll 0.001$ for both types of groups.

beliefs, that is, whether they are able to predict correctly what other players—whose conditional-contribution vector is known—will contribute. The table is read in the same way as Table 5.

...-stated belief is empirically correct	1st	2nd	3rd	4th	no belief	# obs.
INFOTYPE						
Full defection only	40	7	0	0	52	42
Multiple-equilibria groups	11	7	0	2	80	54
INFOEQM						
Full defection only	31	8	5	0	56	64
Multiple-equilibria groups	7	4	0	0	89	28
Random benchmark						
Full defection only	16	3	1	0	79	
Multiple-equilibria groups	5	2	1	0	91	

Table 8: Percentage of participants for whom the belief in the column-title is the first belief that is correct, by treatment and type of RPNE set. The random benchmark is calculated using the same assumptions as before (cf. Tables 5, 6, and 7).

Result 4. Under common knowledge of preferences, roughly two thirds of the participants do not hold empirically correct beliefs: these participants do not include their co-players' contribution choices in a list that may contain up to four entries.

The last data column in Table 8 shows that in groups with only the full-defection equilibrium, slightly more than half of the participants do not specify a belief that correctly predicts the other group members' contributions. This is particularly surprising because participants may state up to four beliefs. The fraction of participants with empirically inaccurate beliefs is even higher in case of multiple equilibria, with more than 80% failing to include a correct belief in their list.[39] Yet, comparing the above percentages to those expected under random matching of beliefs and actions shows that revealing others' conditional-contribution vectors does help participants predict their co-players' actions to some degree.[40] As the analysis

[39]There are a whole number of possible alternative analyses on the correctness of beliefs. An alternative at one extreme would be to ask whether players ever state the correct contribution for at least one co-player somewhere on the list of stated beliefs, which is true for about 80% of the participants in both treatments and equilibrium-set types. Arguably, more informative criteria would be to look at the fraction of players who state both co-players' correct contributions in at least one stated belief (but not necessarily in the correct combination) or the players who are correct in all their stated beliefs about one specific other player (possibly being incorrect about the other in some or all stated beliefs). Both criteria are fulfilled for about 50% in case of the single, full-defection equilibrium, and 25% when there are multiple equilibria, with little difference between treatments. The message from these figures is the same as the one in the main text.

[40]χ^2-tests yield $p \ll 0.001$ for single- and $p = 0.022$ for multiple-equilibria groups.

in Section 5.5 will show, common knowledge of best-reply correspondences helps predominantly by revealing the *defectors* in a group, whose subsequent choices are easier to predict.

From what has been presented so far, the outcome of the test of **H 5** on the descriptive power of the RPNE is immediately obvious:

Result 5.1. Participants generally do not play a *revealed-preference Nash-equilibrium*: in only 10% of all groups, we observe a RPNE. In all of these groups, omnilateral defection is the only RPNE.

Figure 3 shows the distribution of groups playing a RPNE over treatments and groups, grouped by the type of RPNE sets.[41] As can be seen, only 3 single-equilibrium groups of each treatment coordinate on the (full-defection) RPNE. The low number of groups coordinating on the RPNE even when it is unique shows that this deviation from the theoretic prediction is due only very partially to the problem of equilibrium coordination. The following result makes this even clearer.

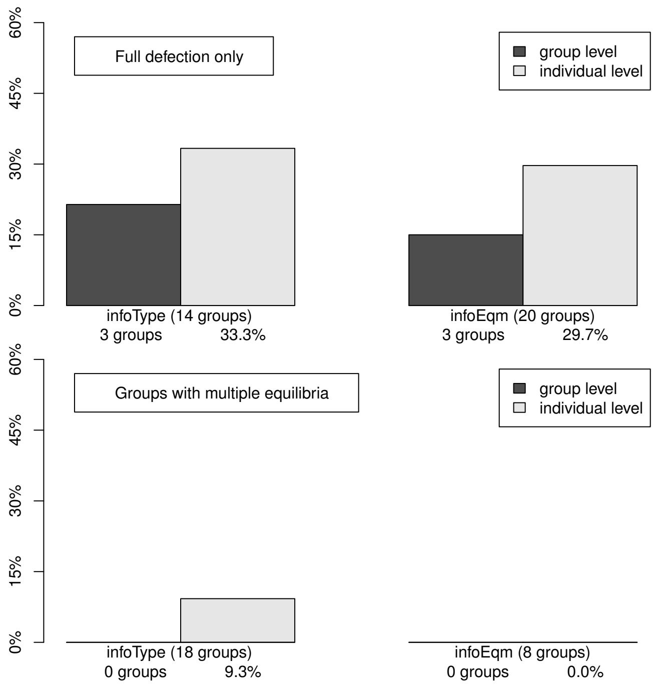

Figure 3: Number and percentage of groups playing a RPNE, percentage of participants playing a 'RPNE on the individual level', by treatment and type of RPNE set.

[41]Note that in INFOEQM, the number of groups sums up to 28 only, as groups with participants excluded from the analysis (cf. Section 4.4) had to be excluded here as well.

Result 5.2. Participants rarely play an 'RPNE on the individual level'. That is, they rarely choose a RPNE action that is a best-response to their belief which in turn is a belief that the co-players contribute according to one element from the set of RPNE.

In Figure 3, I juxtapose the fraction of participants playing a 'RPNE on the individual level' to the group-level data. As can be seen from the respective right-hand bars, the fraction of players conforming to the RPNE prediction at least individually is rather low, too. In groups with only a full-defection RPNE, they make up about one third of all participants, while in groups with multiple RPNE, they do not make up even 10%. Overall, they make up one fifth of all players.

Finally, looking at whether participants happen to play a best-response to others' contributions, I observe that this is the case for only 55% of the participants. In groups with a single, full-defection RPNE, this figure amounts to 71% in INFOTYPE (66% in INFOEQM), while in groups with multiple equilibria, the according figure is 37% (43%). In other words, the RPNE prediction does not describe accurately the behaviour of almost half of all participants even if we abstract from beliefs.

Summing up, if the assumptions underlying the presented research design hold, people cannot be expected to play a pure-strategy *revealed-preference Nash equilibrium* in a public-good game even if they know who they are playing with. While common knowledge of players' conditional-contribution vectors leads to a higher fraction of participants choosing a RPNE action, they rarely play a RPNE. From the data, this does not seem to be due to the complexity of calculating the existing RPNE or a failure to play a best-response to one's belief; rather, the most notable deviation from the theoretic prediction is in the missing empirical accuracy of beliefs even when the RPNE is unique. Therefore, the question arises of what may lead to participants' difficulty of correctly predicting others' contributions. In an attempt to get closer to this question, I perform an exploratory analysis of how beliefs depend on player types in the remainder of Section 5.

5.4 Classification of participants into types

For the exploratory analysis of how belief accuracy and player types go together, I split the data into the types described by Fischbacher, Gächter and Fehr (2001). For this purpose and to account for the modified setup, I group the other-player contribution combinations into three sets of seven combinations each, using the following characterisations:[42]

[42]For this grouping, I ordered the other-player contribution combinations by the respective empirical average response to them; using the combinations' means and variances lexicographically yields the same sets. The conditional contributions were averaged within the subsets in order to allow for minor inconsistencies that may arise due to the random-order one-by-one presentation of the possible contribution combinations of the other players.

Conditional cooperators. Participants were categorised as conditional cooperators if the following conditions held simultaneously: their conditional contributions for intermediate (high) other-player contributions were at least as high as for low (intermediate) contributions, the difference between conditional contributions for high and low other-player contributions was at least $\frac{20}{7}$, and their response to others' full contribution was not 0.[43]

Defectors. Participants were categorised as defectors if their average conditional contributions did not surpass a value of 2 for low, intermediate, and high other-player contributions.

Triangle contributors. Participants were categorised as triangle contributors if their average conditional contributions were strictly higher for intermediate other-player contributions than for low or high ones, or if their average conditional contributions increased monotonically in the other-player contributions but they would respond to others' full contributions by defecting.

Others. Participants were categorised as 'others' if they would not fit into any of the above three categories.

Figure 4 shows the conditional contributions, averaged separately over each type. From the upper left-hand panel it is obvious that the conditional cooperators are driving the general results reported above. However, the average schedule of participants grouped as triangle contributors also exhibits a clearly increasing trend up to an average of 18. This stems from the fact that we chose to classify as triangle contributors also those participants who are generally conditionally-cooperative but choose a contribution of 0 if both other players choose to cooperate fully. The right-hand panels on triangle contributors and 'others' show that for these participants, the variance in their co-players' contributions seems to play a much stronger role than for conditional cooperators.

5.5 Player types and participants' beliefs

Using the classification presented in Section 5.4, I now can ask three questions: (i) are participants of a certain type better at predicting others' contributions than those of other types, (ii) how does the type of the participant about whom a belief is stated influence the accuracy of the prediction, and (iii) does number of RPNE affect the accuracy of beliefs on top of the group composition?

Table 9 displays the percentage of participants who do not include the correct contribution combination anywhere in their stated list of beliefs, conditional on treatment and group composition. As becomes obvious from the table, the percentage of inaccurate beliefs is highest in groups composed of conditional cooperators

[43] The difference of $\frac{20}{7}$ was chosen to include players who would choose 20 in response to the full-contribution combination 20–20, and 0 for all other contribution combinations. This was the case for 1 (2/4) participant in noInfo (infoType/infoEqm), or 3% in total.

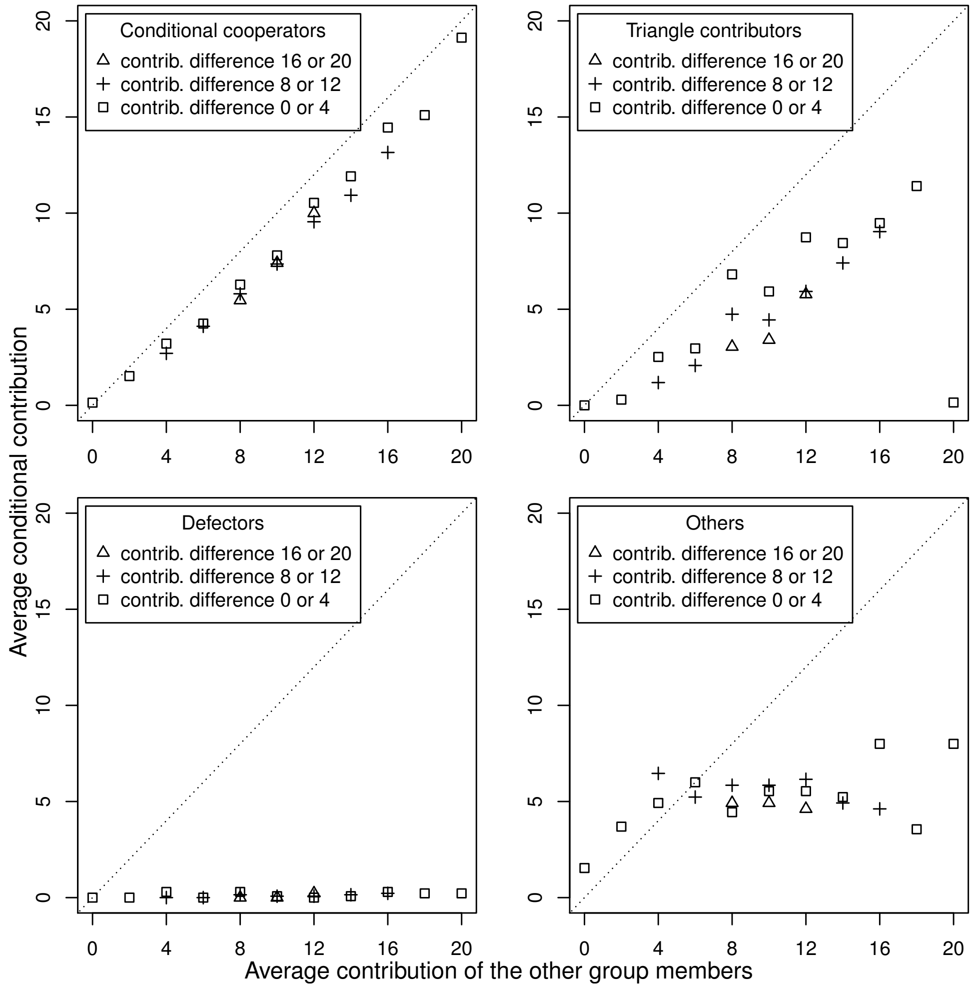

Figure 4: Average contribution conditional on others' average contribution, by player type. Again, note that the difference between others' contributions can be only 0, 8, or 16 if the others' average contribution is divisible by 4, and only 4, 12, or 20, if it is not.

only, while there seems to be little strategic uncertainty in groups that have at least two defectors. This begs the question of whether defectors have more accurate beliefs or whether their behaviour is easier to predict. Table 10 provides the data necessary to answer this question, pooling the data from both INFO-treatments.

Result 6. Player types do not differ in their ability to predict others' behaviour; however, types differ in how accurate their behaviour is predicted by others. Multiplicity of RPNE *per se* does not influence the accuracy of beliefs.

From part (a) of Table 10, it becomes immediately obvious that participants of

No belief accurate		No. obs		No. obs.
	INFOTYPE		INFOEQM	
2 or 3 defectors	**17**	6	**13**	8
1 defector, 1 conditional cooperator	**59**	27	**59**	29
2 conditional cooperators	**67**	27	**73**	41
3 conditional cooperators	**83**	36	**93**	14

Table 9: Percentage of participants for whom no stated belief is correct, by group composition and treatment.

	Full-defection only	No. obs.	**Multiple eq'ia**	No. obs.
(a) Inaccurate beliefs by...				
...defectors	**49**	35	**100**	6
...conditional cooperators	**60**	50	**83**	64
...triangle contributors	**50**	14	**71**	7
...others	**57**	7	**80**	5
(b) Inaccurate beliefs on...				
...defectors	**23**	70	**25**	12
...conditional cooperators	**39**	100	**50**	128
...triangle contributors	**33**	28	**46**	14
...others	**64**	14	**60**	10

Table 10: Percentage of participants (a) whose list of beliefs does not include the accurate contributions, and (b) whose contribution is not part of their co-player's belief list, by player type. Note that in part (b), observations are not independent, as each participant has a belief on *two* other participants. Furthermore, the numbers in parts (a) and (b) of the table are not comparable, as for (a), contribution pairs are evaluated, while in (b), only a single participant's contribution is analysed.

different player types do not differ in their ability to predict others' contributions.[44] Part (b) of Table 10 tells a more nuanced story: a comparison of the first row with the remaining data suggests that it is far easier to predict the behaviour of defectors compared to other player types. At the same time, it looks as if the type of RPNE set made little difference. A mixed-effects probit estimation on whether a participant states a correct belief about another participant underlines this impression: using participant random effects and defectors in single-RPNE groups as the baseline category, only the dummy variables for conditional cooperators and 'others' are significantly negative ($p = 0.031$ and $p = 0.003$, respectively) while the dummy variable for triangle contributors goes in the same direction but fails to be significantly different from 0 ($p = 0.308$). At the same time, neither the type of RPNE set nor any interaction effects of player type and RPNE-set type add anything

[44]The corresponding two-sided Boschloo-tests all yield $p > 0.3$.

to the explanatory power (all $p > 0.6$). In other words, the reason for the higher average percentage of inaccurate beliefs in multiple-RPNE groups (cf. Table 8 in Section 5.3) is the substantially lower fraction of easy-to-predict defectors, rather than the uncertainty stemming from the multiplicity of equilibrium actions.

Table 11 seems to suggest a reason for the apparent higher predictability of defectors. 90% of the unconditional defectors from the PREFS-experiment choose a contribution of 0 also in part SIMPG-experiment. However, in this case, no average belief of other players can deviate negatively from the true contribution, so that what for other types is split up into negative deviations and non-deviations will fall completely into the non-deviation category for defectors. Comparing the row-wise sums of the first two categories in Table 11 would seem to lend support to this contention.

Average belief deviation from true action	< -2	$\in [-2, 2]$	> 2	**No. obs.**
...if the target is a defector	9	61	30	82
...if the target is a conditional cooperator	31	36	33	230
...if the target is a triangle cooperator	18	43	40	40
...if the target is non-classifiable	33	29	38	24

Table 11: Percentage of belief deviations from true contributions, averaged over all stated beliefs, by target-player type.

The important question that remains is why it seems so hard to predict others' behaviour. One answer seems to be that participants—being well-aware of the fact that not everybody will conform to the equilibrium predictions—are not very good at assessing *which* (quarter of the) other players will deviate from the equilibrium prediction. However, this does not seem to be the full story. Figure 5 contrasts the fraction of average beliefs on each player type that deviates from the nearest RPNE by at most 2 with the corresponding percentages of actual RPNE actions.

Figure 5 shows that virtually all player types choose the RPNE action more frequently than is expected by the other group members (exceptions are the unclassifiable in case of the single RPNE, and the triangle contributors in case of multiple RPNE; yet, for neither of the two cases is there an adequate number of observations). In other words, under both types of RPNE sets, participants seem to underestimate the economic forces drawing others (and themselves) towards equilibrium. Notably, a third of all participants expect even defectors to contribute positive amounts when the latter observe others' schedules.

6 Summary and discussion

The over-arching question of this study is the question of why participants in public-good experiments contribute as much as they often do. In a recent contri-

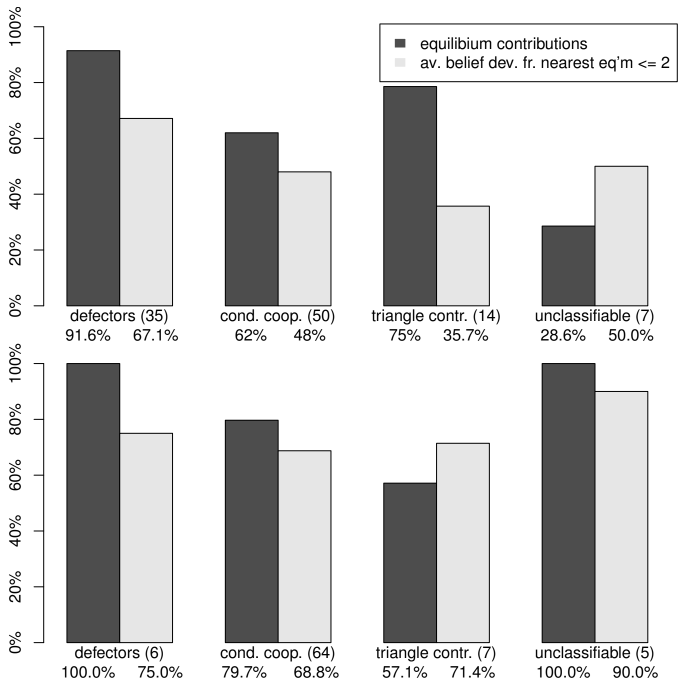

Figure 5: Percentages of RPNE contributions by the respective player types, compared to percentages of average beliefs them that deviate by at most 2 from the nearest RPNE action. Row 1 depicts groups with only a full-defection RPNE, row 2 those with multiple RPNE. The numbers of players of each type are given in parentheses; for the (partly dependent) observations concerning the beliefs, the given numbers are to be doubled.

bution, Fischbacher and Gächter (2010) single out overly optimistic beliefs to be the culprit; however, it remains unclear whether these false beliefs are the consequence of uncertainty over other players' preferences or whether beliefs would be incorrect even if other players' preferences were known.

The present study fills this gap. By making players' conditional-contribution vectors common knowledge in a one-shot public-good game combined with a detailed belief elicitation procedure, I show that uncertainty over other players' types is only part of the story (**Result 2.1**). In fact, the data show that the Nash-prediction based on participants' elicited conditional-contribution preferences (termed *revealed-preference Nash-Equilibrium*, or RPNE) holds for hardly any group in our study (**Result 5.1**), and even individually, participants' behaviour conforms rarely to the prediction (**Result 5.2**). While about 80% best-respond to one of their stated beliefs (**Result 1**) and about 75% of the participants do play a RPNE action (**Result 2.2**), one third of the participants do not include a single other-player-contributions combination from the pure-strategy RPNE set in their list of beliefs (**Result 3**), and roughly two thirds of them do not state empirically correct beliefs (**Result 4**). Importantly, these results hold for groups with multiple RPNE as well as for groups with a unique pure-strategy RPNE. Even abstracting from beliefs, roughly half of the participants do not play a best-response to their co-players' contributions. Notably, none of the above findings can be attributed to participants' potential inability to compute the RPNE set: providing players with the possibility to display the complete set of equilibria leads to an insignificantly lower percentage of RPNE choices (**Result 2.3**).

The above results specify Fischbacher and Gächter's 2010 finding in an important way: the main culprit for the apparent over-contributions seem to be false beliefs 'despite better knowledge', that is, despite players' conditional-cooperation preferences being commonly known. In an attempt to get a better understanding of the reasons for these false beliefs, I embarked on an exploratory analysis of beliefs using the player-type categories introduced in Fischbacher, Gächter and Fehr (2001). The analysis shows that the higher average percentage of inaccurate beliefs in multiple-RPNE groups can be explained by their respective group composition in conjunction with two simple facts: (i) different player types do not differ in their ability to predict others' behaviour, and (ii) the behaviour of defectors seems to be easier to predict compared to other types (**Result 6**). The second fact most likely is due to defectors' optimal choice being a corner solution. Notably, multiplicity of RPNE *per se* does not seem to play a role for the strategic uncertainty participants face as expressed in the fraction of inaccurate beliefs.

I can only speculate on why participants do so badly when predicting others' contributions. At least two reasons seem to be important: (i) they feel a substantial fraction will not play according to the RPNE prediction but mispredict who will be the deviating players, and (ii) they overestimate the frequency of such deviations. The latter is particularly surprising when considering that it holds even if the target player is a defector facing a full-defection RPNE only. Obtaining a better understanding of this issue promises to be an interesting point for future research.

Technical acknowledgements

The experiments were computerised using z-Tree (Fischbacher, 2007), participants were recruited using ORSEE (Greiner, 2004) with Mozilla Firefox. The equilibria of the game were calculated during the experiment using R (R Development Core Team, 2001, 2012, Ihaka, 1998), which was also used to analyse the data in combination with RKWard (Rödiger et al., 2012). For the statistical testing, R packages Exact (Calhoun, 2012; Boschloo test) and lme4 (Bates et al., 2012; mixed-effects probit regression) were of particular value. All this was done on a computer running on KDE-based (KDE e.V., 2012) Kubuntu, which required the use of wine for the programming of the experiment. The article was written using Kile.

References

Ambrus, Attila and Parag A. Pathak, "Cooperation over finite horizons: A theory and experiments," *Journal of Public Economics*, 2011, *95* (7-8), 500–512.

Bardsley, Nicholas, Robin Cubitt, Graham Loomes, Peter Moffatt, Chris Starmer, and Robert Sugden, *Experimental Economics: Rethinking the Rules*, Princeton University Press, 2009.

Bates, Douglas, Martin Maechler, and Ben Bolker, "Package 'lme4'," http://lme4.r-forge.r-project.org/ 2012. Last accessed on 31st January, 2013.

Bazerman, Max H. and William F. Samuelson, "I Won the Auction But Don't Want the Prize," *Journal of Conflict Resolution*, 1983, *27* (4), 618–634.

Binmore, Ken and Avner Shaked, "Experimental economics: Where next?," *Journal of Economic Behavior & Organization*, 2010, *73* (1), 87 – 100.

Blanco, Mariana, Dirk Engelmann, and Hans Theo Normann, "A within-subject analysis of other-regarding preferences," *Games and Economic Behavior*, 2011, *72* (2), 321 – 338.

Breuer, Wolfgang and Anselm Hüwe, "Explaining Individual Contributions in Public Goods Games using (only) Reciprocity and Overoptimism," 2013. Working paper.

Brosig, Jeannette, Thomas Riechmann, and Joachim Weimann, "Selfish in the End? An Investigation of Consistency and Stability of individual Behavior," February 2007. Working paper.

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

Calhoun, Peter, "Package 'Exact'," http://cran.r-project.org/web/packages/Exact/Exact.pdf 2012. Last accessed on 31st January, 2013.

Chaudhuri, Ananish, "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," *Experimental Economics*, 2011, *14*, 47–83.

Cheung, Stephen L., "New insights into conditional cooperation and punishment from a strategy method experiment," *Experimental Economics*, forthcoming.

Croson, Rachel T. A., "Theories of Commitment, Altruism and Reciprocity: Evidence from Linear Public Goods Games," *Economic Inquiry*, Apr 2007, *45* (2), 199 – 216.

Dijkstra, Jacob, "Explaining contributions to public goods: Formalizing the social-exchange heuristic," *Rationality and Society*, 2012, *24* (3), 324–342.

Fehr, Ernst and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 1999, *114* (3), 817–868.

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

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

_ , _ , and Ernst Fehr, "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment," *Economics Letters*, 2001, *71(3)*, 397–404.

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

Gächter, Simon, "Conditional cooperation: behavioral regularities from the lab and the field and their policy implications," in Bruno S. Frey and Alois Stutzer, eds., *Psychology and economics*, MIT Press Cambridge, MA 2007.

_ and Benedikt Herrmann, "Reciprocity, Culture and Human Cooperation: Previous Insights and a New Cross-Cultural Experiment," *Philosophical Transactions of the Royal Society B - Biological Sciences*, 2009, *364(1518)*, 791–806.

Greiner, Ben, "An Online Recruitment System for Economic Experiments," in Kurt Kremer and Volker Macho, eds., *Forschung und wissenschaftliches Rechnen 2003*, Vol. 63, Göttingen: Ges. für Wiss. Datenverarbeitung, 2004, pp. 79–93.

Healy, Paul J., "Epistemic Foundations for the Failure of Nash Equilibrium," 2011. Working paper.

Huck, Steffen, Hans-Theo Normann, and Jörg Oechssler, "Two are few and four are many: number effects in experimental oligopolies," *Journal of Economic Behavior & Organization*, 2004, *53* (4), 435 – 446.

Ihaka, Ross, "R: Past and Future History," in Sanford Weisberg, ed., *Proceedings of the 30th Symposium on the Interface*, 1998, pp. 392–396.

Kagel, John H. and Dan Levin, "The Winner's Curse and Public Information in Common Value Auctions," *American Economic Review*, Dec 1986, *76* (5), 894–920.

KDE e.V., "About KDE," http://www.kde.org/community/whatiskde/ 2012. Last accessed on 31st January, 2013.

Klumpp, Tilmann, "Finitely Repeated Voluntary Provision of a Public Good," *Journal of Public Economic Theory*, 2012, *14* (4), 547–572.

Ledyard, John O., "Public Goods: A Survey of Experimental Research," in John Kagel and Alvin Roth, eds., *Handbook of Experimental Economics*, Princeton University Press, 1995, pp. 111–194.

Nicklisch, Andreas and Irenaeus Wolff, "On the nature of reciprocity: Evidence from the ultimatum reciprocity measure," *Journal of Economic Behavior & Organization*, 2012, *84* (3), 892 – 905.

R Development Core Team, "What is R?," *R News*, 2001, *1*, 2–3.

_ , *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, 2012.

Rödiger, Stefan, Thomas Friedrichsmeier, Prasenjit Kapat, and Meik Michalke, "RKWard: A comprehensive graphical user interface and integrated development environment for Statistical Analysis with R," *Journal of Statistical Software*, 2012, *49* (9), 1–34.

Appendices A-C

For Appendices A to C, please refer to
<http://www.wiwi.uni-konstanz.de/fischbacher/home/staff/dr-irenaeus-wolff/>.

THURGAU INSTITUTE
OF ECONOMICS
at the University of Konstanz

Hauptstr. 90
CH-8280 Kreuzlingen 2

Telefon: +41 (0)71 677 05 10
Telefax: +41 (0)71 677 05 11

info@twi-kreuzlingen.ch
www.twi-kreuzlingen.ch