

Reischmann, Andreas

Conference Paper

Conditional Contribution Mechanisms for the Provision of Public Goods in Dynamic Settings - Theory and Experimental Evidence

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Public Goods, No. B10-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Reischmann, Andreas (2016) : Conditional Contribution Mechanisms for the Provision of Public Goods in Dynamic Settings - Theory and Experimental Evidence, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Public Goods, No. B10-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/145613>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Conditional Contribution Mechanisms for the Provision of Public Goods in Dynamic Settings - Theory and Experimental Evidence

XXXXXXXXXXXXXXXXXX

February 26, 2016

Abstract

Many mechanisms have been developed to solve the free-rider problem in private public good provision. Most of these mechanisms were designed to have good static (Nash) equilibrium properties. I present in this paper a new class of mechanisms, the Conditional Contribution Mechanisms (CCM), which are designed to have good dynamic equilibrium properties instead. The CCMs give all agents the possibility to condition their contribution on the total level of contribution provided by all agents. Their dynamic incentive structure makes the CCMs particularly suited for repeated public goods. I prove for a very general class of environments that all equilibria of the CCMs under a new variant of Better Response Dynamics, called Unexploitable Better Response Dynamics (UBRD), are Pareto efficient. I further present a first experimental study of one mechanism's performance compared to the performance of the Voluntary Contribution Mechanism. In an environment with binary contribution and linear valuations, agents play the mechanisms in a repeated setting. I test one case of complete information and homogeneous valuations and a second case with incomplete information and heterogeneous valuations. In both cases a significantly higher contribution rate can be observed when the CCM is used. Furthermore, all stable outcomes of the CCM, which are observed in the experiment, are in line with the prediction of UBRD.

Keywords: Experimental Economics, Public Goods, Mechanism Design, Better Response Dynamics.

JEL-Classification: C9, D82, H41, C72

1 Introduction

Public goods are characterized by non-excludability and non-rivalry in consumption. If contribution to public goods is voluntary, egoistic utility maximizing agents choose their contribution to a public good on basis of their individual price and individual benefit. Therefore, the positive external effects of individual contributions are neglected, leading to an inefficient level of public good provision. This is the well known free-rider problem. Numerous attempts have been made to find a mechanism that provides efficient contributions in public good environments. I give a short survey of previously proposed mechanisms in section 2. In this paper, I propose a new class of mechanisms, which I call the Conditional Contribution Mechanisms (CCM). In these mechanisms agents can choose conditional contribution offers as messages.

An optimal mechanism for public good provision should implement a Pareto efficient outcome in dominant strategies, or at least in a unique Nash equilibrium. The mechanism should be individually rational, i.e. all agents should be ex ante better off by participating in the mechanism. It should be budget balanced, i.e. the sum of transfers payed to and received from an eventual central authority should be zero. The mechanism should not make use of any information, e.g. on individual preferences, that is unlikely to be available to the mechanism designer in applications. And the mechanism should not be too complicated, such that agents can be trusted to play the mechanism as intended. Furthermore, the optimal mechanism should not have any environmental restrictions. I.e. it should not make use of specific tools, as e.g. communication, which are not available to agents in every environment.¹ It is apparent from this extensive list of properties, that such a mechanism does not exist. In fact, already Pareto optimality, budget balance and dominant strategy implementation are incompatible (Groves and Ledyard, 1977b). Therefore, there is a need to make some restrictive assumptions on public good environments and search for optimal mechanisms in

¹I do not claim that this list of properties is complete. There are always further desirable features that can be asked of a mechanism. Fairness considerations, which might be captured by a switch from Pareto optimal outcomes to core outcomes, could be one such further desirable property.

those environments.

The main difference between this attempt and previous attempts is that to this point most mechanisms aimed at an implementation of Pareto efficient outcomes in Nash equilibria. In this sense, mechanisms proposed so far were designed to have good static equilibrium properties. The CCMs on the other hand are specifically designed with a focus on their dynamic equilibrium properties and are therefore particularly suited for repeated public goods. There are many public goods that have to be supplied repeatedly. Financing of online public goods like Wikipedia, renovations of infrastructure, like roads or children playgrounds, and the cleanness of the kitchen in a shared flat are only three examples of very different scale. One reason, why repeated public good environments have not gained more attention so far, might be that dynamic behavior in repeated mechanisms was hard to predict.

However, Healy (2006) provides solid indication that agents behavior in repeated interaction in public good mechanisms can be well described by a model of better response dynamics. Therefore, I assume in this paper that there is a repeated public good environment in which agents' behavior can be described by a behavioral model that is based on better response dynamics. For such environments, I propose a new class of mechanisms, the Conditional Contribution Mechanisms. I give a formal analysis of the CCMs' dynamic equilibrium properties and I demonstrate by experimental evidence that one version of the CCMs outperforms the Voluntary Contribution Mechanism (VCM) significantly in two different environments. Furthermore, I show that the behavioral model, which is used in the theory sections, makes correct predictions according to the experimental evidence presented.

In the Conditional Contribution Mechanisms, agents can free-ride and contribute unconditionally, just as in the VCM. Moreover, agents have the possibility to conditionally contribute. In the first part of the paper, I consider an environment in which contribution is binary. In such an environment, a conditional contribution offer is especially simple. Agents may announce: "I am willing to contribute to the public good if at least k agents contribute in total." The mechanism then selects the highest possible level of total contribution that

satisfies all those conditions. I call this simple version of the CCMs the Binary Conditional Contribution Mechanism (BCCM). I claim in the following that this kind of mechanism has very good dynamic behavioral properties. I.e. in my opinion this mechanism is well suited to allow agents to coordinate over time on Pareto efficient equilibrium outcomes in repeated public good environments.²

I model a repeated public good environment in the following way. There is a finite number of agents, one private and one public good. All agents interact via the same mechanism over a finite or infinite number of periods. Every agent has the same endowment in every period and has to spend his entire endowment on the two goods.³ The basic behavioral model that I apply to this environment is Better Response Dynamics (BRD). In this concept, all agents can change their message from one period to the next. And a message is a better response to the message profile of the previous period if this message leads to a (weak) increase in utility, when everybody else sticks to the message from the previous period. Under BRD agents switch with strictly positive probability from their message in the previous period to a certain message if and only if this message is a better response to the message profile of the previous period. A set of message profiles that, once reached, can never be left by BRD, and that does not contain a smaller set with the same property is called a recurrent class. If such a set is a singleton, i.e. it only contains a single message profile, it is called an absorbing state. An outcome is considered an equilibrium outcome under such a dynamic model if it is the outcome of a recurrent class or of an absorbing state.

If agents behave according to BRD, then conditional contribution offers have two distinct properties that make them suitable for repeated interaction in public goods. First, imagine that in the current period nobody contributes to the public good. By offering a contribution conditional on some (higher) total level of contribution, agent i can cooperate without being

²The CCMs are supposed to be simple decentralized mechanisms and applicable in environments without a central authority. As such the mechanisms do not make use of transfer payments between agents. Therefore, I also consider the notion of Pareto efficiency in this paper without the possibility of transfers. It may, therefore, be possible that Pareto improvements on efficient outcomes can be found, when transfer payments are allowed.

³I.e. there is no consideration of consumption or saving between periods.

worse off in the short term. If others behave similarly, contributions can be increased without anyone suffering short term losses. Thus, conditional contribution offers have the potential to leave inefficient outcomes under BRD.

Second, imagine that contributions are on a Pareto efficient level. Assume further that all agents, who contribute positive amounts, condition that contribution on the same level, the aggregate level of contributions. Such a message profile, in which all conditions equal the sum of contributions, shall be called perfectly coordinated. Whenever the message profile is perfectly coordinated, every single agent is pivotal. I.e. no agent can reduce his or her contribution, without total contributions going to zero. This indicates that conditional contribution offers have the potential to stabilize efficient outcomes under BRD.

In fact, there is only one kind of a better response that makes agents leave a Pareto optimal outcome, once a perfectly coordinated message profile is reached. Agents can lower their condition. Such a deviation never alters the outcome immediately (contribution offers did not change and all conditions are still satisfied). However, it can introduce incentives for other agents to free ride on the deviator's contributions. I claim that lowering one's condition in a coordinated state is not rational, and I call such a deviation exploitable. Accordingly, deviations which are not exploitable will be called unexploitable. I add unexploitability as a second condition to the model of Better Response Dynamics. Therefore, I assume that agents only deviate to messages which are unexploitable and a better response with respect to the current outcome. I prove that under this behavioral model, Unexploitable Better Response Dynamics (UBRD), the CCMs lead only to Pareto efficient equilibrium outcomes.

In the second part of the paper, I test this theory in a lab experiment in two environments. In both environments, contribution is binary and the benefit from the public good is linear. The environments differ in terms of preferences and information structure. For the first environment, I choose homogeneous valuations and complete information, because this is a standard environment. In this way the results are better comparable to previous experiments. For the second environment, I choose heterogeneous valuations and incomplete information,

since this is more realistic and more challenging. In both environments, I compare the BCCM with the VCM, repeated over 20 periods. I find that the BCCM significantly outperforms the VCM in terms of contributions and efficiency in both environments. Furthermore, all stable equilibrium outcomes, which can be observed in the BCCM treatments, are in line with the prediction of the behavioral model UBRD.

Finally, in the third part of the paper, I generalize the environment. Contributions are no longer assumed to be binary, and the utility functions are much more general. In an environment with non-binary contribution, a conditional contribution offer has the form: "I am willing to contribute x to the public good if total contribution is at least y ." I demonstrate that the simple generalization of the BCCM, in which all agents announce one such offer, has inefficient equilibria under UBRD. I call this mechanism the Single Conditional Contribution Mechanism (SCCM). However, the slightly more complex version, in which all agents are allowed to make two such offers in each period, solves this issue. Since this is the most general form of the CCMs, I simply call it the Conditional Contribution Mechanism (CCM).

The rest of the paper is structured as follows. Section 2 provides a short survey of previously proposed public good mechanisms and their environmental restrictions, as well as a discussion of environmental restrictions for the mechanisms I propose in this paper. Section 3 introduces the Binary Conditional Contribution Mechanism and provides the equilibrium analysis under Unexploitable Better Response Dynamics. Section 4 describes the experimental setting that is used to test the BCCM in the lab. In Section 5, I present and discuss the experimental results. In Section 6, I extend the theoretical analysis to environments in which contributions are not binary and preferences over the public good are very general. Finally, Section 7 gives a short summary and discussion of the paper. Proofs to all theorems can be found in Appendix A. Translations of written instructions and test questions handed out to agents in the experiment can be found in Appendix B and Appendix C.

2 Related literature.

Since I propose new mechanisms for public good provision, my work relates to all previously proposed mechanisms in this area. However, a complete and extensive survey of all those mechanisms is beyond the scope of this paper. Therefore, I give a short discussion of the most prominent work in this area. One way of presenting these mechanisms is by listing all desirable properties which they do and do not satisfy. However, the final goal of the design of public good mechanisms must be the application to real problems. Therefore, I try to go one step further and give for all mechanisms a characterization of all environments to which I think they could be reasonably applied. At the very end of this section, I also give the same characterization for the CCMs.

The first mayor point of distinction between environments is whether or not there is a central authority. The most prominent examples of mechanisms for environments with a central authority are the well known Vickrey-Clarke-Groves (VCG) mechanisms (Vickrey, 1961, Clarke, 1971, Groves and Ledyard, 1977a). The mayor disadvantages of the VCG mechanisms are that they are not budget balanced and that participation does not have to be rational for all agents. However, in environments with a central authority, participation is usually considered to be enforceable and an unbalanced budget, while not desirable, might at least be possible. Therefore, the VCG mechanisms are suitable for environments with a central authority. The tax and subsidy scheme of Falkinger (1996) falls into this category as well. In this mechanism, agents receive a subsidy or have to pay a tax, depending on the deviation of their contribution to the public good from the average contribution of everyone else. While this mechanism has an advantage over the VCG mechanisms by having a balanced budget, the Falkinger mechanism only has an efficient interior equilibrium when the induced prices for the public good happen to be the Lindahl prices (Kirchsteiger and Puppe, 1997). This turns out to be almost only satisfied, when preferences of all agents are perfectly homogeneous. In environments with homogeneous preferences, there is also experimental support for this mechanism (Falkinger et al., 2000, Bracht et al., 2008). Thus, the Falkinger mechanism is

suitable for environments with a central authority, in which preferences are homogeneous.

The second category of mechanisms covers mechanisms that are very simple. The Voluntary Contribution Mechanism is the most prominent example. It is applicable in all environments. However, it has the mayor disadvantage that it does not lead to Pareto efficient outcomes in equilibrium. Especially when interaction is repeated, contribution rates under the VCM get very low in later periods. See Ledyard (1994) for a survey on this extensive branch of the literature. A second example of a simple mechanism is the Provision Point Mechanism (PPM). In the PPM there exists a certain threshold. When total contributions do not meet at least this threshold no public good is provided and the money is usually refunded. While the PPM is also applicable in all environments, the difficulty of choosing an optimal threshold varies across environments. When the public good is binary, as e.g. in the one streetlight problem (Bagnoli and Lipman, 1989), the optimal threshold is obvious. However, in other cases, like the multiple streetlight problem or continuous public goods, finding the optimal threshold may not be trivial. There are multiple studies that demonstrate that in suitable environments the PPM leads to higher contributions to the public good than the VCM (see e.g. Rondeau et al. (1999, 2005), Rose et al. (2002)). Because of its simplicity the PPM is regularly applied in practice, as e.g. on crowd funding platforms. Bagnoli and Lipman (1989) attempted to extend the idea of the PPM to continuous public goods. Their generalization produced efficient outcomes in theory under a special refinement of the Nash equilibrium, the Successively Undominated Strictly Perfect Equilibrium (SUSPE). However, experimental evidence suggests that agents do not play the game according to this refinement (Bagnoli et al., 1992).

The third category covers mechanisms that offer potential rewards for contributions. Such mechanisms are e.g. auction and lottery mechanisms. Schram and Onderstal (2009) compare a first-price winner-pay auction, a first-price all-pay auction and a lottery. They find that out of those three mechanisms the all-pay auction leads to significantly higher contributions. Morgan and Sefton (2000) present an experiment in which a lottery leads to higher contri-

butions to a public good than the VCM. They further find that higher prize money leads to a more effective mechanism. Contrary to the findings of Schram and Onderstal (2009), Corazzini et al. (2010) show an experiment in which a lottery outperforms an all-pay auction. Still in their experiment both mechanisms fare better than the VCM. The main disadvantage of such mechanisms is the risk of losing the prize money. When preferences for the public good are rather low, total contributions might be lower than the prize money, leading to an unbalanced budget. One further environmental constraint is that the use of monetary incentives limits the applicability to environments in which contributions are also monetary.⁴ However, in environments where there is good indication that contributions will exceed the prize money such mechanism may not be a bad choice.

In contrast to the use of rewards, mechanisms of the fourth category aim to increase contributions by allowing punishments. This literature started with Ostrom et al. (1992) and Hirshleifer and Rasmusen (1989), who showed that punishment or ostracism can theoretically and experimentally increase cooperation under repeated interaction. This idea became even more popular when Fehr and Gächter (2000) demonstrated that agents even punish in one-shot games out of anger, although punishment in their experiments is costly and there is no potential gain in future periods. This implies that the threat of punishment is also credible in one-shot games. Punishment can therefore increase cooperation in both cases. However, the use of punishment has one severe disadvantage. While punishment usually increases contribution rates, it does not increase efficiency by the same margin. This is due to the fact that the punishment itself is socially costly.

The fifth category covers mechanisms that use an extensive form. One example is the auction mechanism of Smith (Smith, 1979, 1980). This auction mechanism lets all agents announce a marginal contribution, which they are willing to pay for one unit of the public good, and a desired quantity. The mechanism then proposes to supply the average of these desired quantities if the sum of all marginal contributions is at least as high as the marginal

⁴Otherwise the budget is definitely negative.

price. In the next stage agents vote on the proposed outcome. Only unanimous acceptance implements the proposed quantity. In case the vote fails, the entire procedure is repeated up to five times. One further extensive form mechanism is the subsidy mechanism proposed by Danziger and Schnytzer (1991). They propose that agents announce a price in the first stage, by which they will subsidize all other contributions in the second stage. In the second stage agents choose their contributions based on the announced subsidies. Varian (1994) proposes a similar mechanism for two agents in more general environments. Furthermore, there is the Jackson-Moulin mechanism (Jackson and Moulin, 1992), which was later generalized by Bag (1997). In this mechanism all agents announce the total benefit of a public project in the first stage. If the highest announced number is lower than the cost of the project the project is not undertaken. In the second stage, all agents announce their individual benefit. If the sum of those benefits is higher than the highest total benefit announced in period one, the project is implemented and costs are shared according to some predefined rule. All those mechanisms have in common, that they use sub-game perfection as a refinement. This leads to high requirements of information on part of the participating agents to play the equilibrium. Smith (1979) incorporated this problem into his mechanism by giving agents five trials. In this way, agents have some time to learn the information by trial and error. However, this comes with the disadvantage that it might take a long time to play the mechanism, which comes at a cost as well.⁵ Therefore, these mechanisms are best suited for environments in which agents have much information on other agents' preferences and/or environments in which simultaneous play can be implemented easily and at low costs.

Finally, the sixth category covers other mechanisms that focus on dynamic equilibrium behavior. There is first of all the canonical mechanism by Maskin (1999). This mechanism is designed to implement general social choice functions. Therefore, it is, in theory, also applicable to public good environments. The main challenge when applying the canon-

⁵Think about larger groups that are not in one room as in lab experiments and that can not be forced to sit at their computer at the same time. In such groups simultaneous play will have to be implemented by waiting for the slowest responding agent's input in each round. With up to ten rounds (five times sending messages and five times voting) this can take up to some weeks.

ical mechanism is that it uses the entire type profile of all agents in every agent's message space. One can choose relatively simple type spaces to test the canonical mechanism in the lab. However, it is entirely unclear how those type spaces should be modeled in the field, without message dimensionality getting completely out of hand. Probably most related to my work are the contractive mechanisms by Healy and Mathevet (2012). They manage to design a mechanism that satisfies nearly all desired properties. It Nash implements Pareto efficient outcomes, it is budget-balanced and it is individually rational. The main challenge in this mechanism might be its complexity. Especially when budget-balance is guaranteed off equilibrium, the outcome function of the mechanism becomes very mathematical and hard to comprehend for agents. The main difference to the approach in this paper is that their mechanism is continuous, while the mechanism in this paper works exactly because of its discontinuities.

Of course, the mechanisms that I propose in this study, the Conditional Contribution Mechanisms, have environmental constraints as well. First of all, the CCMs are designed for repeated interaction. While I demonstrate in this paper that all outcomes of the mechanisms, which are stable in the long run, are Pareto efficient, this does not tell us anything about efficiency in one shot public good games. Therefore, the CCMs are probably only suited for repeated public good environments. Second, the CCMs use conditional contribution offers and I assume that those offers can be enforced. This is not a problem in environments in which contributions are monetary, however when contributions are not monetary enforceability of such offers must always be considered.⁶ And third, the behavioral model of Unexploitable Better Response Dynamics must give an accurate description of long term stable outcomes in the respective environment. This is an empirical question. I provide experimental evidence in section 5 that the model gives an accurate description of long term stable outcomes in binary contribution environments with linear valuations for the public good. Whether this extends to other environments, is an interesting question for further research.

⁶When contributions are monetary they can be submitted with each offer and be refunded if the corresponding condition is not met, just as in the PPM when the threshold is not met.

3 The Binary Conditional Contribution Mechanism

This section introduces one Conditional Contribution Mechanism in a binary contribution environment with linear valuations for the public good. In environments with binary contribution the mechanism has a particularly simple form.⁷ I call it the Binary Conditional Contribution Mechanism (BCCM). In section 6, I show how the mechanism can be generalized to environments with non-binary contributions. And I prove that similar results to the ones in this section hold for very general preferences over the public good.

Consider the following environment. There are $n \in \mathbb{N}$ agents labeled i that interact over a certain number of periods t . There is one public and one private good. Each agent has an endowment of one monetary unit in each period, which he or she can either invest into one unit of the private or one unit of the public good. An outcome in a certain period t is then defined as $z^t = (z_1^t, \dots, z_n^t)$ with $z_i^t \in \{0, 1\}$, $\forall i \in I := \{1, \dots, n\}$. Here $z_i^t = 1$ is interpreted as agent i investing his monetary unit in period t into the public good. And $z_i^t = 0$ represents agent i investing his monetary unit in period t into the private good. Let $Z := \{0, 1\}^n$ be the outcome space that contains all possible outcomes in a certain period. For notational convenience define $\underline{z} = (0, \dots, 0)$ as the outcome in which no agent invests into the public good.

Since I intend to apply a dynamic solution concept, utility is defined for each period separately. In this section, I assume utility to be linear in the both the private good and the public good.⁸ Further, all agents $i \in I$ have a certain valuation $\theta_i \in [0, 1)$ for the public

⁷Whether contributions are binary or not, is either given by the environment and observable to everyone. Or it is a design choice by the mechanism designer. Therefore, it is not problematic to propose different forms of a mechanism, depending on whether contribution is binary or not.

⁸This assumption is made in this section for two reasons. First to get some simple intuition on how the CCMs work. And second to cover the theory for the experiment in sections 4 and 5. In later sections this assumption will be relaxed considerably.

good.⁹ Utility of agent i in period t is therefore given by

$$u_i^t = 1 - z_i^t + \theta_i \sum_{j=1}^n z_j^t. \quad (1)$$

The main solution concept in this paper is a dynamic adjustment process that will be formally introduced in subsection 3.2. Therefore, the results of this paper apply whenever this adjustment process describes equilibrium outcomes reasonably well. This might be the case in environments with complete or incomplete information. Therefore, I do not make any specific assumption on whether agents are informed on other agents' valuations or not.

In this environment, I define the Binary Conditional Contribution Mechanism as $G^{BCCM} := (M^{BCCM}, g^{BCCM})$, where M^{BCCM} describes the mechanism's message space and $g^{BCCM} : M^{BCCM} \mapsto Z$ describes the mechanism's outcome function. In the BCCM every agent can choose a natural number between 1 and $n + 1$. Thus, the message space is defined as $M^{BCCM} = \prod_{i=1}^n M_i^{BCCM}$, with $M_i^{BCCM} := \{1, 2, \dots, n+1\}$, $\forall i \in I$.¹⁰ The chosen message is interpreted in the following way: Choosing message $m_i = k$ is like saying "I'm willing to contribute to the public good if at least k agents (including myself) contribute in total." Note that with the messages $m_i = 1$ and $m_i = n + 1$ players can decide to contribute in any or no case, respectively.¹¹

Given a chosen message profile, the outcome selected by the mechanism is the outcome with the highest possible level of contributions such that all message statements are satisfied.

⁹Values $\theta_i < 0$ are excluded, since then the public good would be a bad for those agents. If this were the case a mechanism that does not use transfers can never guarantee Pareto improvements. Thus, the mechanism proposed in this paper should only be applied if valuations for the public good of all agents are weakly positive. Values $\theta_i \geq 1$ are excluded for simplicity of notation. Any agent with $\theta_i \geq 1$ has a weakly dominant strategy to contribute the entire endowment to the public good. Thus, there is no need to provide additional incentives to these kind of agents. Therefore, including the possibility of $\theta_i \geq 1$ would not lead to a significant change in any results of the paper, but would complicate notation at several points.

¹⁰If the number of participating agents is unknown to either the mechanism designer or the participating agents themselves, the message space can be chosen as $M_i^{BCCM} := \mathbb{N}$ instead. This is less elegant, since agents now have some redundant options. However, all relevant results are unaffected.

¹¹Since there are only n agents, there can never be $n + 1$ contributing agents.

Formally, define

$$K(m) := \max \left\{ k \in \{0, 1, \dots, n\} \mid \sum_{i=1}^n \mathbb{1}_{(m_i \leq k)} \geq k \right\}.^{12} \quad (2)$$

The outcome of the mechanism is defined as $g^{BCCM}(m) = z$ with $z_i = 1$ if and only if $m_i \leq K(m)$.

3.1 Nash equilibria of the BCCM

I will later use a dynamic solution concept to predict the outcomes of the BCCM. However, looking at the Nash equilibria of the stage game provides a lot of intuition on what the incentive structure looks like. Therefore, I will start with a discussion of Nash equilibria. The following example demonstrates what properties an outcome must have to be a Nash equilibrium outcome.

Example 3.1 *Consider 5 identical agents with valuation $\theta_i = 0.4 \forall i \in I$. The trivial Nash equilibrium is given by $m_i = 6, \forall i \in I$, where no agent contributes to the public good. However, there are more equilibria as e.g. when agents 1, 2 and 3 choose message $m_i = 3$ and agents 4 and 5 choose $m_i = 6$. In this case the first three agents will contribute to the public good: $z = (1, 1, 1, 0, 0)$. The structure of the mechanism makes this an equilibrium. Agents 4 or 5 can only change the outcome to $z' = (1, 1, 1, 1, 0)$ or $z'' = (1, 1, 1, 0, 1)$ respectively by unilateral deviation. Neither deviation is beneficial. And the first three agents can only change the outcome to \underline{z} , which is not beneficial either. Thus, no agent has any (strict) incentive to deviate.*

In the second message profile in the example ($m = (3, 3, 3, 6, 6)$) all agents, who end up contributing to the public good, condition their contribution exactly on the aggregate level of contributions. And all other agents choose to free-ride in any case. I call such a message profile perfectly coordinated. In a perfectly coordinated message profile any agent can

¹²In equation (2) $\mathbb{1}_{(m_i \leq k)}$ denotes the indicator function, which is 1 if $m_i \leq k$ and 0 otherwise.

only alter the outcome in one way by a unilateral deviation. Agents that currently do not contribute can only alter the outcome by unilaterally contributing themselves. Since $\theta_i < 1$ this is not profitable. Agents that currently do contribute can only change the outcome to \underline{z} . This may or may not be profitable depending on the current outcome z . This reasoning demonstrates that a certain outcome z can be implemented as a Nash equilibrium if and only if there is no agent for which the deviation from z to \underline{z} is profitable.¹³

Theorem 3.2 *z is the outcome of a Nash-equilibrium of the BCCM if and only if $z \succeq_i \underline{z}, \forall i \in I$.*

Theorem 3.2 predicts equilibria which are Pareto efficient as well as equilibria which may not be Pareto efficient. In the next subsection I will present my considerations on why only the Pareto efficient Nash equilibria will prevail as stable outcomes of a reasonable dynamic adjustment process.

3.2 Unexploitable Better Response Dynamics

In this subsection, I construct and apply a dynamic behavioral model for the Conditional Contribution Mechanisms.¹⁴

Since Better Response Dynamics (BRD) have been found to give a good description of agents' behavior in repeated public good mechanisms (Healy, 2006), a BRD model will be the basis for my behavioral model. In such a dynamic model agents play the same game repeatedly over several periods. And in our case all agents will be allowed to adjust their message from one period to the next.¹⁵ A message m_i^{t+1} in period $t + 1$ is called a better

¹³It is easy to see that an outcome that can be realized as a Nash equilibrium outcome by some message profile can also be realized as a Nash equilibrium by a perfectly coordinated message profile.

¹⁴Constructing a new behavioral model for a new mechanism might be unusual and this step can certainly be criticized. However, it is my believe that we should put more weight on what we believe is rational play in a given mechanism, rather than simply put our trust into one of the standard solution concepts. I firmly believe that the model of Unexploitable Better Response Dynamics, which I introduce in this paper, combines all relevant rational incentives to correctly predict the outcomes of dynamic stable states in the CCMs. However, since I can not simply expect the reader to share my believes, I present experimental evidence that supports the presented model in chapter 5.

¹⁵Any dynamic model must specify whether only one or all agents can change their message in a given period. I choose the latter, since it seems more reasonable for most applications.

response for agent i to the message profile m^t in period t if m_i^{t+1} had provided a least as much utility to i as m_i^t if he had chosen m_i^{t+1} already in period t . Mathematically, m_i^{t+1} is a better response if

$$u_i(g(m_i^{t+1}, m_{-i}^t)) \geq u_i(g(m_i^t, m_{-i}^t)) \quad (3)$$

In this equation g represents a general outcome function and m_{-i}^t is the message profile of all agents other than agent i in period t . A recurrent class of BRD is a set of message profiles, which if ever reached by the dynamics, is never left and which contains no smaller set with the same property. If such a recurrent class consists of a single message profile, it is called an absorbing state. The equilibrium outcomes of Better Response Dynamics are defined as all outcomes of their recurrent classes or absorbing states. I will also call such outcomes the long term stable outcomes of the dynamic model.

For some mechanisms all better responses might be equally rational and the standard BRD model might be a good predictor of dynamic behavior. However, I argue that for the CCMs some better responses are more rational than others, because of strategic implications. Consider the following example.

Example 3.3 *Assume there are 5 identical agents all with type $\theta_i = 0.4$. Assume that currently 4 agents contribute to the public good. The message profile could e.g. be $m^t = (4, 4, 3, 3, 6)$. In this case agents 1 through 4 contribute to the public good. Consider now agent 1. Any message $m_1^{t+1} \in \{1, 2, 3, 4\}$ is a better response for agent i to the message profile m^t . None of these messages would change the outcome if no other agent changes his message at the same time. However, the message $m_1^{t+1} = 3$ gives agent 2 an incentive to deviate to $m_2^{t+2} = 6$ in the following period. Under the new message profile $m^{t+2} = (3, 6, 3, 3, 6)$ only agents 1, 3 and 4 would contribute to the public good. Thus, agents 1, 3, and 4 are worse off and agent 2 is better off. The same would be true for the messages $m_1^{t+1} = 2$ and $m_1^{t+1} = 1$. Messages $m_1^{t+1} \in \{1, 2, 3\}$ can thus be exploited by agent 2 in a later period, making agent 2 better off and agent 1 worse off. Therefore, I call these messages exploitable. The special*

structure of the CCMs makes it possible for agents to prevent this kind of incentives for exploitation without having to free-ride themselves.

With the motivation of the example in mind, I propose to combine the myopic better response condition with a second condition on behavior. I call this condition unexploitability. As demonstrated by the example, unexploitability captures non-myopic strategic incentives in the CCMs and, therefore, complements the myopic better response condition to form a balanced dynamic behavioral model.¹⁶ Before I present more intuition on this model, I formally define Unexploitable Better Response Dynamics (UBRD).

Definition 3.4 *Given a message profile m^t and an outcome $g(m^t) = z^t$, a message m_i^{t+1} is called exploitable if there is any $m_{-i}^{t+1} \in M_{-i}$ such that $z^{t+1} = g(m_i^{t+1}, m_{-i}^{t+1}) \prec_i z^t$ with $z_i^{t+1} > 0$. A message m_i^{t+1} is called unexploitable if it is not exploitable.*

This definition has two particular details that deserve a discussion. Note first that all possible message profiles of other agents m_{-i} are considered. One could argue, since I assume a BRD model, that I should only consider profiles of better responses of other agents at this point. However, the UBRD model is designed with the most realistic case in mind, in which agents have no information on the preferences on other agents. In this case agents can not tell whether a certain message of another agent is a better response. Therefore, from such a player's perspective it seems rational to account for all possible choices. Note second that I only consider outcomes z^{t+1} in which agent i contributes a strictly positive amount. This has to be the case, since the choice of conditions in the BCCM only allows agents to exclude some outcomes in which they have to contribute. Agents never have any influence over outcomes in which they do not contribute. Furthermore, the term exploitable suggests that agent i is exploited by another agent. If he does not contribute, there is nothing to exploit.

¹⁶There is one further indication that choosing exploitable messages is not rational. Being exploitable is in some cases identical to being weakly dominated. This can best be demonstrated by the following example. Assume there are 5 identical agents all with type $\theta_i = 0.4$. Let the current message profile be $m = (6, 6, 6, 6, 6)$. In this case no agent contributes and the outcome is \underline{z} . In this situation a message is exploitable if it makes outcomes possible in which an agent is worse off than in \underline{z} . Those messages are only $m_i = 1$ and $m_i = 2$. Both messages are weakly dominated by $m_i = 3$. In general, when the current outcome is \underline{z} a message is exploitable if and only if it is weakly dominated.

In the next step I combine unexploitability with the BRD model into one behavioral model.

Definition 3.5 *In Unexploitable Better Response Dynamics (UBRD) all agents can adjust their message in every period. Agent i switches in period $t + 1$ to message m_i^{t+1} with strictly positive probability if and only if*

- (1) m_i^{t+1} is a (weak) better response to m^t and
- (2) m_i^{t+1} is unexploitable.

This model is well defined for the CCMs. In other words, unexploitable better responses always exist.¹⁷ If agent i currently contributes to the public good, conditioning his contribution on the current total level of contribution is always an unexploitable better response. And if agent i currently does not contribute to the public good, unconditional free-riding is always an unexploitable better response. The rest of the paper uses the following definition to simplify notation.

Definition 3.6 *z' is a strict* Pareto improvement over z if z' is a Pareto improvement over z , that is strict for all agents with type $\theta_i \neq 0$.*¹⁸

With this definition I can formulate the central result for the binary model.

Theorem 3.7 *An outcome $z \in Z$ is an outcome of some recurrent class of the BCCM under UBRD if and only if it is a Pareto optimal outcome and a strict* Pareto improvement over z , if at least one such an outcome exists. If no such outcome exists, the dynamics cycle over all outcomes, which are weak Pareto improvements over z .*¹⁹

¹⁷This statement holds only for the CCMs. In most other mechanisms, there might be situations where the set of unexploitable better responses is empty.

¹⁸Agents, who do not benefit from the public good ($\theta_i = 0$), can never be strictly better off than in z . If those agents are excluded by assumption, this special definition of *strict** is not necessary. However, when there exist agents with a valuation of $\theta_i = 0$, many mechanisms, which try to force agents to cooperate, violate individual rationality. In my opinion it is worth the additional notation to demonstrate that the BCCM can handle this case.

¹⁹Note again at this point that I consider a concept of Pareto optimality without the possibility of transfer payments.

Note first that the assumption of existence is very weak. It will be true in all cases, in which there are significant efficiency gains possible. With the other assumptions in this section, we can formulate this assumption mathematically in the following way. A Pareto optimal outcome, which is a *strict** Pareto improvement over \underline{z} exists, if and only if

$$\exists k \leq n \text{ and } i_1, \dots, i_k \in I : \theta_i > \frac{1}{k}, \forall i \in \{i_1, \dots, i_k\} \quad (4)$$

Let me provide an example to further improve the intuition for the main result:

Example 3.8 *Revisit the example with 5 identical agents all with type $\theta_i = 0.4$. The theorem predicts that all outcomes in which 3, 4, or 5 agents contribute to the public good are outcomes of recurrent classes of the BCCM. Assume for example that the current message profile is $m^t = (4, 4, 4, 4, 6)$. Then agents 1 through 4 contribute to the public good, while agent 5 does not. Thus, the outcome is $z^t = (1, 1, 1, 1, 0)$. Any unilateral deviation of agent 5 would lead to $z^{t+1} = (1, 1, 1, 1, 1)$ and would thus not be a better response. For agents 1 through 4 unilateral deviations to $m_i^{t+1} \in \{5, 6\}$ would lead to the outcome $z^{t+1} = \underline{z}$. They are, thus, not better responses either. And for agents 1 through 4 messages $m_i^{t+1} \in \{1, 2, 3\}$ make outcomes possible in which the agent has to contribute, and total contribution is less than 4. Thus, those messages are exploitable. Therefore, the message profile $m^t = (4, 4, 4, 4, 6)$ is an absorbing state of UBRD.²⁰*

As demonstrated by example 3.8, Pareto optimal outcomes can only be left under Better Response Dynamics if agents choose exploitable messages. Despite all described strategic incentives I do not expect that no exploitable messages at all will be chosen in experiments. Instead my trust in the predictive power of the model comes from the following observation.

When agents play the BCCM and they better respond, they will reach a Pareto optimal outcome eventually, even by random play, since there is no absorbing state under BRD. Once agents reach such a Pareto optimal outcome, one of three cases has to happen. In the first

²⁰In this example the other absorbing states are given by $m' = (5, 5, 5, 5, 5)$ and $m'' = (3, 3, 3, 6, 6)$ (in any permutation)

case, which is described by UBRD, enough agents play unexploitable to stabilize the Pareto optimal outcome. In the second case, some agents play exploitable such that incentives for exploitation exist. In this case, assume that the remaining agents do not exploit them (e.g. because they have preferences for conditional cooperation, as described in Fischbacher et al. (2001), Fischbacher and Gächter (2010)). If the exploitable messages are not exploited, the outcome is again not left under UBRD. It is, therefore, stable as predicted by UBRD, even if agents do not actually play unexploitable. In the third case, enough agents play exploitable such that they can be exploited. And then they are exploited. In this case, the outcome, which should be stable according to UBRD, is left. This is the case which is most critical for the theory. In this last case, however, choosing exploitable messages makes agents strictly worse off once they are exploited. And no matter how often such an outcome is left, agents will at some point return to the same or another absorbing state of UBRD, because there are no other stable outcomes under BRD. Therefore, I claim that in this case, after some time, enough agents will realize that choosing exploitable messages makes them worse off. And they will stop to choose them. Once this happens we return to case one and the outcome is once again stable.

In fact, not even all agents have to play unexploitable to remove incentives for exploitation. This strengthens the argument that in the third case eventually enough agents will learn to do so. Consider again an example to demonstrate this.

Example 3.9 *Assume there are 5 identical agents all with type $\theta_i = 0.4$. Let the current message profile be $m^t = (5, 5, 5, 1, 1)$. In this case only agents 1 through 3 send an unexploitable message. Nevertheless, neither of the agents can strictly benefit from any unilateral deviation. Although agent 4 and 5's messages are exploitable, any attempt to exploit these agents would leave only agents 4 and 5 contributing. Thus, total contribution to the public good would go down by 3. This makes all agents worse off. Thus, in this example it is sufficient if 60% of agents behave according to UBRD and 40% behave according to BRD to support full cooperation.*

4 Experiment on the BCCM

The main prediction of last section’s theoretical model is that only Pareto efficient outcomes are stable outcomes of the BCCM under UBRD. The focus of this section is the design of an experiment to test this prediction. Since the theory should hold for cases of complete and incomplete information about agents’ preferences, I test the BCCM in two different environments. The first environment covers complete information and homogeneous valuations. I include this case, since it is very simple and it provides basic intuition on whether the prediction of the UBRD model holds. The second environment covers incomplete information and heterogeneous valuations. I include this case, since it is particularly challenging and it provides good intuition on whether or not the results are stable across different environments.²¹

Besides validating the prediction of UBRD, I want to evaluate the level of contributions that the BCCM can provide. For this reason, I need to compare the contribution rates of the BCCM to some other mechanism that can be used as a benchmark. The mechanism that I use as a benchmark is a standard binary Voluntary Contribution Mechanism (VCM). In this mechanism agents have only two options in any given period. They can contribute in any case or free-ride in any case. I use the VCM as a benchmark for three reasons. First, the VCM is still used in many applications to provide important public goods (e.g. Wikipedia). Therefore, improvements on the VCM in such environments are still important. Second, most other experimental studies so far chose the VCM as a comparison as well (e.g. Smith (1979, 1980), Morgan and Sefton (2000), Corazzini et al. (2010)). Thus, this choice provides some intuition on how the BCCM compares to those other mechanisms as well. Third, the VCM is the most standard of all public good mechanisms and therefore best understood. And although the VCM may not lead to the most efficient levels of contribution, it still enables a better evaluation of the contribution rates of another mechanism than any other choice.²²

²¹Comparisons between the two environments, with the same mechanism, are not the focus of this experiment. Therefore, it is not a problem that more than one variable changes between the two environments.

²²Because of the similarities between the CCMs and the PPM, the PPM might seem to be a good comparison as well. However, this is not the case. A closer examination reveals that the PPM and the CCMs are suited for more or less disjoint sets of environments. The success of the PPM depends on the mechanism

In each treatment groups of five agents play one of the two mechanisms as a stage game repeated over 20 periods. In every period each agent is endowed with 10 points.²³ Each agent can either invest those points in a group project (the public good) or keep them in his private account. As in the theory part in the previous section, contributions are binary in this experiment. Therefore, the endowment can not be divided between the two accounts. An outcome is again described by $z = (z_1, z_2, z_3, z_4, z_5)$. And $z_i = 0$ still denotes that agent i does not invest his points into the project. Since every agent has an endowment of 10 points, $z_i = 10$ implies in the experimental part that agent i does invest his points into the project. All agents are provided with a linear valuation θ_i for the public good. Given their valuation agents have the following payoff function in each period:

$$\Pi_i = 10 - z_i + \theta_i \sum_{j=1}^5 z_j \quad (5)$$

In the complete information case every agent knows all players' valuations for the public good and valuations are homogeneous with $\theta_i = 0.6, \forall i$. In the incomplete information case agents only know their own valuation and that every agent has a valuation of $\theta_i = 0$ with a probability of 20% and a valuation of $\theta_i = 0.6$ with a probability of 80%. I call those agents who do not benefit from the public good type 1 agents. And I call agents who do benefit type 2 agents. The valuation for every agent is constant over all periods and all draws of valuations are independent. Therefore, it is possible that all or none of the agents are of type 1.²⁴ In expectation, there is one agent per group, who does not benefit from the public good. Every random draw is used for one group with each mechanism to ensure comparability.

designer's choice of the threshold. The PPM is, therefore, best suited for environments in which the socially optimal choice of the threshold is known to the mechanism designer. Instead, the main point of the CCMs is that agents can find a socially optimal threshold, which is not known to the mechanism designer, by an endogenous dynamic adjustment process.

²³10 points are chosen to ensure that the number of points earned in each period is a natural number in all cases.

²⁴In the end, the random draws resulted in three groups with only type 2 agents, five groups with one type 1 agent and four groups with two type 1 agents.

4.1 Conducting the experiment

The experiments were conducted at the Alfred-Weber Institute (AWI) of Heidelberg University. agents were recruited via ORSEE (Greiner, 2004) from the AWI agent pool, which consists mainly of students of differing agents. In total 195 agents took part in the experiment. Seven groups played the VCM with complete information and eight groups played the BCCM with complete information. In the incomplete information treatments, each mechanism was played by 12 groups. Sessions lasted between 45 minutes and one hour.

When the agents entered the lab, they were randomly allocated to their seats by drawing numbered cards. All seats are separated by screens to prevent any interaction of agents outside of the program. Every agent was handed one set of instructions and test questions in German. English translations of the instructions and test questions can be found in Appendix A and B. Once all agents answered the test questions correctly and there were no more questions, a computer program written in z-Tree (Fischbacher, 2007) was started. The program randomly matched agents into groups of five. Groups stayed the same over all 20 periods. Every group played only one mechanism and only one information setting. In the incomplete information treatments, the random draw of types was performed by the program at the beginning of period one.

After the last period, there was a short questionnaire asking for personal characteristics such as gender and previous knowledge of game theory. Afterwards agents were called by seat number to receive their payoff in cash and in private. In every period, type 2 agents and agents in the complete information treatment could earn between 6 and 34 points. Type 1 agents earned 10 points per period if they followed their dominant strategy not to invest into the public good in any case. Points of all periods were added up. agents were payed 1€ for every 40 points. Type 1 agents in the incomplete information treatment received an additional 5€ lump sum to compensate them for the lower earning possibilities. Average earnings per agent were 11.55€.

4.2 Expectation on behavior

The Voluntary Contribution Mechanism gives every agent the choice whether or not he contributes to the public good. However, no agent has any influence over any other agent's contribution. If agents care only for their own profit, it is easy to see, and well known in the literature, that free-riding is a dominant strategy. The straight forward Nash prediction is, therefore, that all agents will free-ride. Since free-riding is a strictly dominant strategy any refinement of Better Response Dynamics will also predict this outcome as a unique absorbing state.

However, it is equally well known that this theoretical prediction is seldom to never observed in experiments. The general observation is a contribution rate of about 40-60% of the efficient level in the first period. If the public good game is played repeatedly, as it is in this study, the typical experimental finding is that contribution rates decline over time. Social preferences (e.g. inequality aversion as in Fehr and Schmidt (1999)) as well as preferences for conditional cooperation (Fischbacher et al., 2001, Fischbacher and Gächter, 2010, Kocher et al., 2008) in combination with a better responding behavior (Healy, 2006) explain these findings well. Since those findings are very persistent (see Ledyard (1994) for a survey of the early findings and Burger and Kolstad (2009) for a recent example with the binary VCM), this is also what I expect to find in this experiment.

The theoretical prediction for the BCCM, as derived in the last chapter, is that under Unexploitable Better Response Dynamics all stable outcomes are Pareto optimal. Therefore, the theory does not predict anything for the first few periods. Instead, whenever a group, after arbitrary play in some periods, sticks to one specific outcome until the last period, my behavioral model predicts that this outcome is Pareto efficient in all cases.

5 Experimental results

I run this experiment with the intention to answer two questions. First, is the model of UBRD suited to predict long term stable outcomes of Conditional Contribution Mechanisms? Second, can the BCCM improve contribution rates to public goods compared to the VCM?

5.1 Contribution rates

In this subsection, I discuss whether or not the BCCM can increase contributions to the public good significantly, compared to the VCM. Therefore, I compare total contributions in groups under the BCCM to the contributions under the VCM, using the Wilcoxon-Rank-Sum Test. In each group I take the average of total contributions over a certain number of periods. First, I consider all periods to get an impression of contributions overall. Second, I only consider the last 10 periods, to get an impression of the long run contributions, once a certain level of convergence has taken place. Since the last 10 periods include the well known endgame effect, I further consider periods 9 to 18 as third choice. This makes it possible to look at long run contributions without the end game effect.

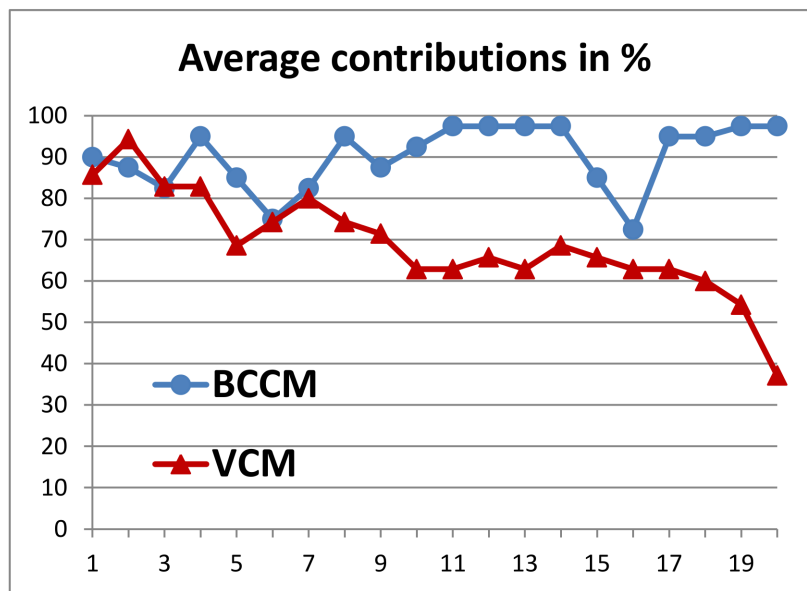


Figure 1: Comparison of average contributions over all groups in the complete information treatment.

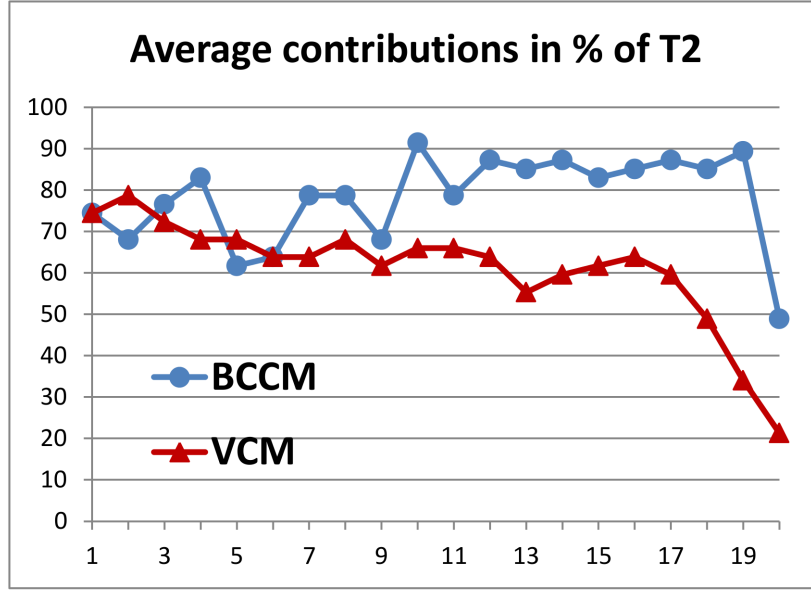


Figure 2: Comparison of average contributions over all groups in the incomplete information treatment.

Average contributions per period over all groups with complete information are displayed in Figure 1. The figure makes the following immediate observations possible. First, the average contribution rate in the first period under the VCM is surprisingly high. The reason for this is probably the binary contribution environment in combination with the rather small group size of 5. Second, contributions in the VCM decline over time as expected. Third, contribution rates in the BCCM are similar to the VCM in early periods but much higher in later periods. Fourth, the BCCM does, under complete information, not suffer from any endgame effect. All these observations support the theoretical prediction that the BCCM has better dynamic properties than the VCM. In fact the BCCM leads already to significantly higher contributions when all periods are taken into account ($p = 0.0425$). When only the last 10 periods are considered, the effect is highly significant ($p = 0.0080$). And when I exclude the endgame effect (periods 9 to 18) the results are still significant ($p = 0.0388$).

In the incomplete information treatments, the following type distributions were drawn. Three groups contained only type 2 agents. Five groups contained one type 1 and four type 2 agents. And four groups contained two type 1 and three type 2 agents. The type 1 agents

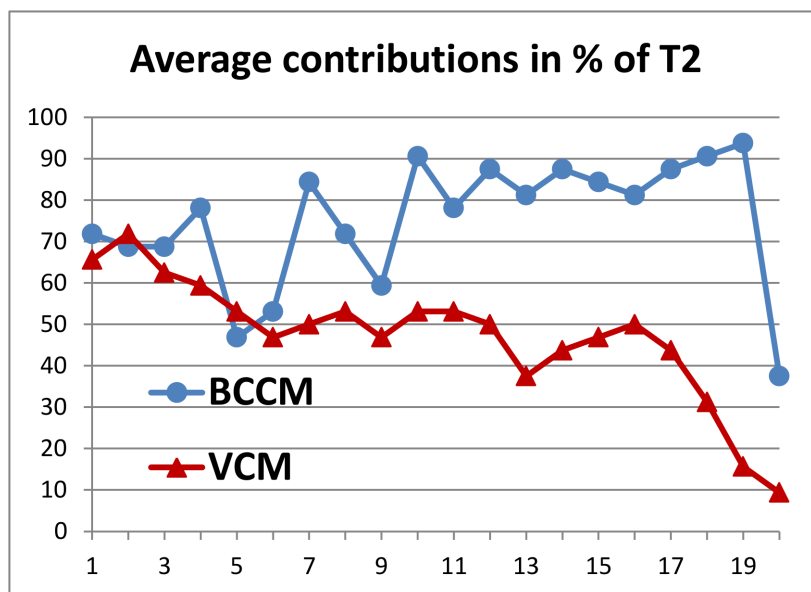


Figure 3: Comparison of average contributions over all groups with at least one type 1 agent.

have a dominant strategy to free-ride. Besides a few mistakes in period 1 and one mistake in period 2, all type 1 agents also chose this strategy. Therefore, contribution rates in the incomplete information treatments are always compared in terms of average contributions of type 2 agents. Average contributions of type 2 agents per period over all groups with incomplete information are displayed in Figure 2. Observations from this figure differ from the complete information case in only one way. Under incomplete information the BCCM suffers from a severe endgame effect.²⁵ There are two reasons for this. First, many groups play a stable Pareto efficient equilibrium over some periods ahead of period 20. Nevertheless, some agents in these groups harm themselves by deviating in period 20, which results in a complete breakdown of contributions. One explanation for this behavior is that some agents make mistakes because of the somewhat higher complexity of the incomplete information treatment. Another explanation might be that those agents are familiar with the successful VCM behavior, to contribute most of the periods and free-ride at the very end. And they apply this behavior without much thought. The second reason for the endgame effect is that

²⁵In fact this endgame effect only exists in groups with exactly one type 1 agent. And it occurs in all of those groups.

some groups reach Pareto efficient outcomes, but not enough agents play unexploitable to stabilize the equilibrium in period 19. This creates incentives for individual agents to deviate in period 20. However, more than one deviation usually leads again to a complete breakdown. These complete breakdowns make the endgame effect in the BCCM even bigger than in the VCM. This second effect might vanish when more periods are played. This gives agents more time to converge to stable equilibria. Besides the endgame effect, the results are very similar. Average contributions are higher under the BCCM than under the VCM. When all periods are considered, this effect is only weakly significant ($p = 0.0602$). However, for the last 10 periods results are again significant at the 1% level with a p-value of $p = 0.0078$. And when the last two periods are excluded the increase is significant at the 5% level ($p = 0.0199$).

Figure 3 shows average contribution rates in the incomplete information treatments when only the more challenging groups are considered. These are the groups that contain at least one type 1 agent. This leads to lower contribution rates under both mechanisms in the first half of the experiment. However, groups playing the BCCM again manage to achieve high contribution rates in periods 10 to 19. Groups playing the VCM, on the other hand, can not stop the decline of contributions. This observation strengthens the impression that the BCCM robustly reaches high long term contribution rates, even in settings in which coordination in early periods is difficult.

Result 5.1 *Under complete as well as under incomplete information the Binary Conditional Contribution Mechanism leads to more efficient contribution rates than the Voluntary Contribution Mechanism.*

One typical goal of the implementation problem is that the designed mechanism should lead to Pareto efficient outcomes. Whether the BCCM leads to Pareto efficient outcomes can only be answered qualitatively. In the case of complete information, when all periods are considered, 91.88% of outcomes are Pareto efficient. When only the last 10 periods are taken into account 96.25% of outcomes are efficient. And in the last 4 periods every single outcome is Pareto efficient. Note again that Pareto efficiency is considered without

the possibility of transfer payments. Thus, an outcome is Pareto efficient in the complete information treatment if four or five agents contribute to the public good.

While the theoretical prediction of Pareto efficient outcomes fits the data very well in the complete information case, the situation is not as clear under incomplete information. In those treatments 75.42% of all outcomes under the BCCM are Pareto efficient. This number increases slightly to 80.83% in the last 10 periods, but decreases again to 75% in the last 4 periods, because of the endgame effect under incomplete information.

Result 5.2 *Under complete information the Binary Conditional Contribution Mechanism converges to Pareto efficient outcomes. Under incomplete information about three out of four outcomes are Pareto efficient.*

5.2 Unexploitable Better Response Dynamics

Finally, I am interested in the model of Unexploitable Better Response Dynamics itself. How well does the model fit the data for the Binary Conditional Contribution Mechanism?

The more general model of better responding agents fits the data very well. In the complete information treatment about 93% of chosen messages are better responses. In the incomplete information treatment the value is even higher at 96%. This is in both cases high enough to claim that BRD is a good description of agents' behavior.²⁶ However, only around half of all messages are also unexploitable better responses in the two treatments (41% under complete and 53% under incomplete information).

There is no support for a theory that agents learn to choose unexploitable messages over time under incomplete information (52% of messages are unexploitable better responses in the last ten and 53% in the last 5 periods). And only weak support for a learning towards unexploitability under complete information (35% in the last ten and 47% in the last 5 periods).

²⁶Note that this observation is not in conflict with the dips in earlier periods. Since high contribution rates can be obtained in exploitable strategies a better response behavior can in some periods lead to a decline in contributions.

However, the main intention of Unexploitable Better Response Dynamics is not the prediction of individual behavior, but the prediction of the long term stable outcomes. Here UBRD predicts that the long term stable outcomes are the Pareto efficient outcomes. If the dynamics are considered to have converged to a stable outcome if at least four out of the last five outcomes are identical, then 14 out of 20 groups converge to an outcome in the BCCM treatments.²⁷ Of those 14 outcomes all 14 are Pareto efficient. This supports the conclusion that UBRD predicts the dynamically stable outcomes of the BCCM correctly. In comparison, under the definition of convergence from above, 8 out of 19 groups under the VCM reach a stable outcome. Of those 8 outcomes 4 are Pareto efficient and 4 are not Pareto efficient.

5.2.1 Individual group behavior

In the last part of the results section, I will not provide any more statistics. Instead, I provide some examples of behavior in selected groups. This provides a better intuition of whether or not UBRD is a useful behavioral model for the CCMs.

Figure 4 shows the messages of all agents over all periods in one specific group. This example is taken from experiments with incomplete information and agents 1 and 5 do not benefit from the public good. In the first ten periods, contribution rates oscillate between 0% and 100%. Agents might be trying to figure out, who benefits from the public good and who does not, as well as trying to coordinate their messages. From period ten onwards, the group is in an absorbing state of Unexploitable Better Response Dynamics and the outcome is Pareto efficient. Thus, in this group UBRD perfectly predicts behavior, as well as the long term stable outcome.

The second case in Figure 5 demonstrates the predictive power of the UBRD concept when no exploitation takes place. From the first to the last period agents choose exploitable messages. Thus, agents do not at all behave according to UBRD. Nevertheless, since no

²⁷This definition of convergence is rather loose. It allows to include those groups that reach a stable outcome, and in which some agent deviates in period 20. A loose definition of convergence makes this test more challenging for the theory, since it considers more cases in which the theory has to apply.

Group 1										
Periods:	1	2	3	4	5	6	7	8	9	10
S1 (T1)	6	6	6	6	6	6	6	6	6	6
S2	3	6	3	3	3	3	3	3	3	3
S3	4	3	6	2	6	3	3	3	3	3
S4	2	2	2	2	2	6	3	3	4	3
S5 (T1)	6	6	6	6	6	6	6	6	6	6

Periods:	11	12	13	14	15	16	17	18	19	20
S1 (T1)	6	6	6	6	6	6	6	6	6	6
S2	3	3	3	3	3	3	3	3	3	3
S3	3	3	3	3	3	3	3	3	3	3
S4	3	3	3	3	3	3	3	3	3	3
S5 (T1)	6	6	6	6	6	6	6	6	6	6

Figure 4: Example for a group with perfect convergence under UBRD.

exploitation takes place, the outcomes are Pareto efficient in all periods. Thus, in this group the outcome, but not individual behavior, is correctly predicted by UBRD.

The third case in Figure 6 demonstrates that it is not necessary that all agents behave unexploitable, such that an outcome is stable. In this group only agents 3 and 4 play unexploitable. Nevertheless, no agent has any incentive to free-ride. Thus, UBRD correctly predicts the outcome in all periods and partially predicts agents' individual behavior.

In the fourth case in Figure 7, coordination is not achieved (or only achieved in the last two periods if one wants to count that as convergence). In this specific group, agent 2 is of type 1 and makes two errors in periods 1 and 2. This might induce other agents to believe he is type 2. Therefore, some agents try repeatedly to make him contribute by conditioning their contribution on a total level that includes agent 2's contribution. This kind of fight, where one agent tries to free-ride while one other agent conditions his contribution on a level that includes the first agent's contribution is the main force that keeps contributions low in some groups. I claim that this behavior will rarely be persistent over time since in general both agents have a short term incentive to give in. Therefore, these kind of fights will be settled at some point. This is the moment when the stable outcomes of UBRD are reached.

In summary, UBRD does not perfectly capture agent behavior in all groups. However,

Group 6										
Period:	1	2	3	4	5	6	7	8	9	10
S1	2	2	2	2	2	2	2	2	2	2
S2	3	2	2	2	2	2	2	2	2	2
S3	1	1	1	1	1	1	1	1	1	1
S4	1	1	1	1	1	1	1	1	1	1
S5	2	2	1	1	2	2	2	2	2	2

Period:	11	12	13	14	15	16	17	18	19	20
S1	2	2	2	2	2	2	2	2	2	2
S2	2	2	2	2	2	2	2	2	2	2
S3	1	1	1	1	1	1	1	1	1	1
S4	1	1	1	1	1	1	1	1	1	1
S5	2	2	2	2	2	2	2	2	2	5

Figure 5: Example for a group with no exploitation.

Group 4										
Period:	1	2	3	4	5	6	7	8	9	10
S1	2	2	2	2	2	5	4	4	4	4
S2	2	2	2	2	2	2	2	2	2	2
S3	5	5	5	5	5	5	5	5	5	5
S4	5	5	5	5	5	5	5	5	5	5
S5	4	3	4	4	4	4	4	4	4	4

Period:	11	12	13	14	15	16	17	18	19	20
S1	4	4	4	4	4	4	4	4	4	4
S2	2	2	2	2	2	2	2	2	2	2
S3	5	5	5	5	5	5	5	5	5	5
S4	5	5	5	5	5	5	5	5	5	5
S5	4	4	4	4	4	4	4	4	4	4

Figure 6: Example for a group in which two unexploitable agents stabilize the perfectly efficient outcome.

Group 7										
Periods:	1	2	3	4	5	6	7	8	9	10
S1 (T1)	6	6	6	6	6	6	6	6	6	6
S2 (T1)	3	2	6	6	6	6	6	6	6	6
S3	5	3	3	4	4	6	5	5	5	3
S4	2	2	3	4	3	3	3	3	3	3
S5	1	1	1	1	6	1	1	1	1	2

Periods:	11	12	13	14	15	16	17	18	19	20
S1 (T1)	6	6	6	6	6	6	6	6	6	6
S2 (T1)	6	6	6	6	6	6	6	6	6	6
S3	3	4	2	3	2	1	6	4	3	3
S4	3	4	4	4	4	4	4	3	3	3
S5	2	2	2	2	2	6	4	4	3	3

Figure 7: Example for a group with late or no convergence under UBRD

in all groups in which an outcome is stable over the last periods, this outcome is one of the outcomes predicted by UBRD. Thus, the data partially supports UBRD as a predictor for actual behavior, and it completely supports UBRD as a predictor of long term stable outcomes. And the second point is what the model is intended to predict correctly.

6 Non-binary Conditional Contribution Mechanisms

The environment of section 3 can be generalized to a setting in which contribution is not binary, while keeping the mechanism similar. Assume that every agent can invest any amount between 0 and 1 into the public good. Because it is closer to reality and it keeps the dynamic analysis simpler, I assume a smallest indivisible monetary unit of 0.01.²⁸

The BCCM can be adjusted to this environment in a very natural way. However, this natural extension turns out to have equilibria under dynamic considerations, which are not Pareto optimal. Nevertheless, this failure of the natural extension is an important motivation for the more complex message space of the Conditional Contribution Mechanism, which will

²⁸This discretization resembles the money structure in most countries. All results in the paper hold with any other finite discretization as well as with different levels of income.

be introduced afterwards.

6.1 The Single Conditional Contribution Mechanism

The natural extension of the BCCM, the Single Conditional Contribution Mechanism (SCCM), will assign every agent i the message space $M_i^{SCCM} := \{0, 0.01, \dots, 0.99, 1\} \times \{0, 0.01, \dots, n - 0.01, n\}$, where $m_i = (x_i, y_i)$ is interpreted as “I am willing to contribute x_i to the public good if total contribution is at least y_i .” The outcome space is then given by $Z := \{0, 0.01, \dots, 0.99, 1\}^n$, where z_i is the contribution of agent i to the public good in outcome z . $\underline{z} := (0, \dots, 0)$ is used as before as the outcome with no contribution to the public good by anyone. The level of contribution selected by the mechanism is again the highest level of total contribution such that all conditions are satisfied. Formally, let $Z^{SCCM}(m) \subset Z$ be the set of all outcomes that satisfy all conditions in m . This can be formalized by

$$z \in Z^{SCCM}(m) \Leftrightarrow \left(z_i = 0 \text{ or } z_i = x_i \text{ and } \sum_{j=1}^n z_j \geq y_i \right), \forall i \in I. \quad (6)$$

It is easy to see that $z \in Z^{SCCM}(m)$ and $z' \in Z^{SCCM}(m)$ imply together $z'' = (\max\{z_1, z'_1\}, \dots, \max\{z_n, z'_n\}) \in Z^{SCCM}(m)$. Thus, the outcome of the mechanism is uniquely defined by

$$g^{SCCM}(m) = \operatorname{argmax}_{z \in Z^{SCCM}(m)} \sum_{i=1}^n z_i. \quad (7)$$

The structure of Nash equilibria is similar to the binary case:

Theorem 6.1 *An outcome z is an outcome of a Nash equilibrium of the NEM if and only if $z \succeq_i z, \forall i \in I$.*

Revisit the example

Example 6.2 *Each of five agents has type $\theta_i = 0.4$. Assume $z = (0.5, 0.4, 0.3, 0.2, 0.1)$. Then $z \succ_i \underline{z} \forall i \in I$. This outcome is the outcome of the Nash equilibrium given by $m_i = (z_i, 1.5)$.*

This is a Nash equilibrium since no agent can reduce his contribution without the outcome becoming \underline{z} . And neither can any agent by changing his message increase any other agent's contribution. Thus, the options for unilateral deviations can be reduced to the same cases as in the binary model.

Unfortunately, the SCCM has undesirable equilibria under UBRD as well. The simplest way to show this is by considering an example.

Example 6.3 *Assume again each of five agents has type $\theta_i = 0.4$. Assume further that in period t all agents sent message $m_i^t = (0.1, 0.5)$ and $z^t = (0.1, 0.1, 0.1, 0.1, 0.1)$. Let us find all unexploitable better responses in period $t + 1$. Consider w.l.o.g agent 1. Any message $m'_1 = (x_1, y_1)$ with $x_1 < 0.1$ and $y_1 > x_1$ will lead to \underline{z} and is thus not a better response. Any message $m'_1 = (x_1, y_1)$ with $x_1 < 0.1$ and $y_1 \leq x_1$ will lead to $z = (x_1, 0, 0, 0, 0)$ and is thus not a better response, either. Any message $m'_1 = (x_1, y_1)$ with $x_1 > 0.1$ and $y_1 > 0.4 + x_1$ will lead to \underline{z} and is thus not a better response. Any message $m'_1 = (x_1, y_1)$ with $x_1 > 0.1$ and $y_1 \leq 0.4 + x_1$ will lead to $z = (x_1, 0.1, 0.1, 0.1, 0.1)$ and is thus not a better response, either. This leaves only messages with $x_1 = 0.1$. However of those messages the ones with $y_1 > 0.5$ lead to \underline{z} and are not a better response and the ones with $y_1 < 0.5$ are exploitable. $y_1 = 0.3$ e.g. could lead after deviations of the other agents to $m'_j = (0.05, 0.3)$, $\forall j \in \{2, 3, 4, 5\}$ to $z' = (0.1, 0.05, 0.05, 0.05, 0.05)$. In this outcome agent 1 is worse off than in z^t but contributes a strictly positive amount. Thus, his message was exploitable. The only unexploitable better response is thus $m'_1 = (0.1, 0.5)$. This implies that message profile m^t is an absorbing state of UBRD. However, $z^t = (0.1, 0.1, 0.1, 0.1, 0.1)$ is not Pareto optimal.*

Agents can in this way get stuck on Pareto improvements over \underline{z} which are not Pareto optimal. Any deviation aiming to make further Pareto improvements possible would make the deviating agent worse off in the next period. And such a deviation is infeasible under a better response behavior.

This problem can be solved by letting agents announce more than one tuple of the form (x_i, y_i) . This grants agents a higher flexibility in their strategy giving them the opportunity

to explore Pareto improvements with some tuples, while securing the current level of cooperation with one other tuple. As it turns out a message of two such tuples is already enough to solve the issue. Simplicity is a further desirable feature of mechanisms once practical implementations are considered. Thus, the mechanism I propose in the following paragraph lets agents announce exactly two tuples.²⁹ I call this mechanism the Conditional Contribution Mechanism (CCM).

6.2 The Conditional Contribution Mechanism

In the CCM every agent can announce two tuples $\{(x_i^1, y_i^1), (x_i^2, y_i^2)\} \in M_i^{CCM} := M_i^{SCCM} \times M_i^{SCCM}$. The outcome $g^{CCM}(m)$ of the CCM is then defined as in the SCCM as the outcome with the highest level of contribution consistent with the messages chosen. Let $Z^{CCM}(m) \subset Z$ be the set of feasible outcomes for a message profile m :

$$z \in Z^{CCM}(m) \Leftrightarrow z_i = 0 \text{ or } \left(\exists l_i \in \{1, 2\} : z_i = x_i^{l_i} \text{ and } \sum_{j=1}^n z_j \geq y_i^{l_i} \right), \forall i \in I \quad (8)$$

The outcome of the CCM is then uniquely defined by

$$g^{CCM}(m) = \operatorname{argmax}_{z \in Z^{CCM}(m)} \sum_{i=1}^n z_i. \quad (9)$$

The additional tuple in the message has no effect on Nash equilibrium outcomes, since only one of the two announced tuples per agent is responsible for the outcome. Such a mechanism

²⁹Depending on the application different versions of the mechanism are possible. The more tuples agents can send, the more flexible they are. Thus, more tuples could lead to faster convergence. However, more tuples also make the mechanism more complicated. Therefore, a reasonable version for applications might be to let agents announce any amount of tuples they choose between one and some upper bound. This gives agents the simple option of choosing one tuple, while also giving them the option to choose very detailed messages. This mechanism is from the theoretical perspective identical to the version in the paper. The paper version is chosen since it simplifies notation, especially in proofs.

³⁰The outcome can easily be computed by translating the messages of all agents into step-functions, adding them up and taking the highest fixed point of the resulting function. This makes sure that there is no problem in computation, when n is large.

can thus only be found and argued for, when dynamic properties are taken into consideration. The CCM has indeed the desired positive dynamic properties:

Theorem 6.4 *An outcome $z \in Z$ is an outcome of some recurrent class of the CCM under UBRD if and only if it is a Pareto optimal allocation and a strict* Pareto improvement over \underline{z} .*

An example shall provide some intuition for this result.

Example 6.5 *Consider the example with 5 agents. Each agent has type $\theta_i = 0.4$. Then in all outcomes of recurrent classes 3 agents contribute their entire endowment. The two other agents can contribute any amount. Take for example the outcome $z = (1, 1, 1, 0.5, 0.5)$. This outcome is supported by the messages $m_i = ((1, 4), (1, 4))$ for $i = 1, 2, 3$ and $m_i = ((0.5, 4), (0.5, 4))$ for $i = 4, 5$. The combination of unexploitability and better responding behavior makes sure that the outcome cannot be left to another outcome with lower contributions and the unexploitability condition implies further that the outcome cannot be left to any outcome with higher contributions since either agent 4 or 5 would be worse off than in z . Consider for example the message $m'_4 = ((0.5, 4), (1, 5))$. This deviation in itself does not change the outcome, thus it is a better response. However if agent 5 also switches to $m'_5 = ((0.5, 4), (1, 5))$, the outcome would change to $z' = (1, 1, 1, 1, 1)$. Since $u_{4/5}(z) = 2.1 > 2.0 = u_{4/5}(z')$ the messages m'_4 and m'_5 are exploitable.³¹*

6.3 Non-linear valuation functions

In this section I drop the assumption that valuations are linear and replace it by a weaker assumption. Consider a finite number n of agents with utility functions $u_i(w_i, w_p)$. Here w_i is the private wealth of agent i and w_p is the total amount of wealth invested into the public good by all agents. The functions u_i are only assumed to be strictly increasing and continuous

³¹Agents 1 through 3 did not actively exploit the messages of agents 4 and 5 in this example. In some sense these agents exploited each other. However, the important point is that the deviation from z to z' is not desirable for agents 4 and 5.

in w_i and weakly increasing (and possibly not continuous) in w_p .³² Further, utility functions may differ across agents. Endowment and outcome space $Z := \{0, 0.01, \dots, 1\}^n$ remain unchanged.³³

The fact, that utility gained from the public good increases no longer linearly with the contribution towards the public good, changes the properties of the outcomes of recurrent classes under UBRD. Coalitions of agents might in some instances benefit from collectively reducing their own contributions, even if all other agents would not contribute anything any more. And under the behavioral model, UBRD, deviations by coalitions can occur. Therefore, Pareto optimality will not be enough to ensure that an outcome is part of a recurrent class in this setting.

In the proofs, I use that the options for deviations of coalitions can be limited to outcomes in which no agent outside the coalition contributes. I call such outcomes enforceable, since coalitions cannot force other agents to contribute. When coalitions' options for deviations are limited to their enforceable outcomes, the equilibrium outcomes of the CCM under UBRD can be captured by the core.

Note that the core captures, besides Pareto efficiency, also some fairness considerations. Therefore, implementing the core is at least as desirable as implementing Pareto efficient outcomes. The main disadvantage of the core is the fact that it might be empty. I discuss this concern at the very end of this section.

Definition 6.6 *An outcome $z \in Z$ is enforceable for a coalition $S \subset I$ if $z_i = 0 \forall i \notin S$. The set of all enforceable outcomes for coalition S shall be denoted Z_S*

As in the case of Pareto efficiency I use a standard definition of the core for games without transferable utility as e.g. in (Owen, 1982, p. 293).

³²Note that this includes the cases of agents not profiting at all from the public good, or who get satiated at some level.

³³A further generalization to different endowments for different agents only complicates notation. The mechanism can easily be adjusted by enhancing the message space and all main results would be unaffected.

Definition 6.7 *An outcome $z \in Z$ is in the core if there is no $S \subset I$, $S \neq \emptyset$, and $z' \in Z_S$, such that $z' \succ_i z$, $\forall i \in S$.*

As in the previous results there needs to be a strict disincentive for agents to deviate. Since the outcome space is finite the usual core definition does not guarantee this.

I therefore need a definition, which is somewhat stronger than the usual core definition to describe the equilibrium outcomes. Possibilities for deviations under indifference need to be excluded. The following two technical definitions allow me to formulate the theorem in a simple way.

Definition 6.8 *A core allocation z is strict* for a subset $S \subset I$ of agents if for any enforceable outcome z' of a coalition S' with $S' \cap S \neq \emptyset$ there exists some agent $i \in S'$ with $z \succ_i z'$.*

Definition 6.9 *Define the subset $S^C(z) \subset I$ via $i \in S^C(z)$ if and only if $u_i(1 - z_i, \sum_{j=1}^n z_j) > u_i(1 - z_i, 0)$ as the set of agents that strictly benefit from the amount of public good in z .*

Theorem 6.10 *Assume there exists at least one outcome z that is a core allocation and strict* for $S^C(z)$. Then an outcome z' is an outcome of a recurrent class of the CCM under UBRD if and only if it is a core allocation that is strict* for $S^C(z')$.*

If no such outcome exists the result would be a cycling behavior of the dynamics. It is not obvious that the assumption of existence of such an outcome is satisfied in all relevant cases. However, the existence problem only exists on an infinitesimal level. The mechanism can be adjusted to guarantee existence at arbitrarily low expected costs.³⁴

In the following theorem let Δ be a mapping from $Z \times I \rightarrow \mathbb{R}_+$. The interpretation is that the mapping defines for any agent and any outcome some expected payment $\Delta(z, i) := \delta_{zi}$ that agent i gets paid if outcome z occurs. I write $G + \Delta$ to describe a mechanism G to which the additional payments Δ are added.

³⁴Note first, that since costs are arbitrarily low, there is no need to argue here who should pay those costs. Note further, that only expected costs can be arbitrarily low as the assumption of a smallest monetary unit makes arbitrarily low payments only possible as lotteries.

Theorem 6.11 *For any environment with weakly increasing valuation functions and for any $\epsilon > 0$ there exists a mapping Δ such that in the game $CCM + \Delta$ there exists a core allocation z , which is strict* for the subset $S^C(z)$. Further, the expected cost of Δ is less than ϵ .*³⁵

7 Summary and discussion

This paper introduces the class of Conditional Contribution Mechanisms for the provision of public goods in repeated environments. In these mechanisms agents can condition their contribution on the total level of contribution provided by all agents. Based on the observation by Healy (2006), that Better Response Dynamics describe agent behavior in public good mechanisms well, I motivate a new behavioral model. This model, Unexploitable Better Response Dynamics, predicts that all outcomes of absorbing states of the mechanisms are Pareto efficient.

An experiment was conducted with the aim to test the performance of the Binary Conditional Contribution Mechanism (BCCM) for public good provision. Since this is the first test a simple binary contribution environment with linear valuations is chosen. In the experiment the BCCM is compared to the standard Voluntary Contribution Mechanism in one setting with complete and one with incomplete information.

In all settings the BCCM leads to significantly higher contribution rates than the VCM. This effect is especially large if only the second half of the experiment is considered. In those periods convergence in many groups of the BCCM is complete and average contribution rates are rather stable at 93% (complete information) or 81% (incomplete information). By comparison, average contribution rates over the same periods under the VCM are 60% (complete information) and 53% (incomplete information). Another important difference between the mechanisms is that in groups playing the BCCM no decline of contributions

³⁵Note that in the proof of this theorem I make use of the specific utility functions of all agents, to design the payments Δ . And these utility functions will in general not be known. The theorem can therefore not be interpreted as a fix of the problem in applications. It should rather be seen as an indication that the problem of existence only exists on an infinitesimal level and is therefore likely to be no problem in applications at all.

over time can be observed.

Furthermore, the experimental results support the dynamic model Unexploitable Better Response Dynamics. The model gives an accurate prediction of the long term stable outcomes of the BCCM in the test environment. And all those outcomes are Pareto efficient.

Everything considered, the class of Conditional Contribution Mechanisms is an important addition to the set of public good mechanisms. It satisfies individual rationality and leads under UBRD to Pareto efficient outcomes in repeated public good environments. This makes the CCMs particularly suited for repeated public good environments.

Appendix

General notation: In many proofs I have to show that some outcome z is some sort of equilibrium. In those proofs I need to distinguish between two subsets of agents. The subset of agents who contribute to the public good in z , shall be called $I_1 \subset I$. And the subset of agents who do not contribute to the public good in z shall be called $I_0 \subset I$. If I need a second outcome z' in the proof, those sets will be called I'_1 and I'_0 , respectively.

Proof of Theorem 3.2 Let z be an allocation such that no agent strictly prefers \underline{z} to z and define $k := \sum_{i=1}^n z_i$. Then the message profile $m_i = k \forall i \in I_1, m_i = n + 1 \forall i \in I_0$ is a Nash equilibrium with the desired outcome. It is obvious that $g^{BCCM}(m) = z$. In the following I show that m is a Nash equilibrium.

If some agent i in I_1 deviates to a message $m'_i < k$, the outcome does not change. If he changes his message to some $m'_i > k$, the new outcome will be \underline{z} . Since no agent strictly prefers \underline{z} to z , this can not make agent i strictly better off. Thus agents in I_1 have no strict incentive to deviate.

If some agent j in I_0 deviates to $m'_j > k + 1$, the outcome does not change. If he changes his message to $m'_j \leq k + 1$ he will contribute and total contribution will be $k + 1$. Since $\theta_j \in [0, 1)$ this will make him worse off. Thus also the agents in I_0 have no incentive to deviate and m is indeed a Nash equilibrium.

Let on the other hand z be an outcome such that any agent i strictly prefers \underline{z} to z . Let then m be any message profile leading to the outcome z . By choosing the message $m'_i = n + 1$ any outcome that might occur is at least as good for agent i as \underline{z} . Thus i has an incentive to deviate. Thus m can not be a Nash equilibrium. \square

Proof of Theorem 3.7 Assume first that there exists an outcome, which is Pareto optimal and a *strict** Pareto improvement over \underline{z} . I prove this part of the theorem in two steps. In step 1, I show that any outcome with the described properties is an outcome of a recurrent class of the dynamics. In step 2, I show that from any other outcome the dynamics reach

such a recurrent class with strictly positive probability.

Step1: Let z be any outcome, which is Pareto optimal and a *strict** Pareto improvement over \underline{z} , and let $k = \sum_{i=1}^n z_i$. Then $m_i = k$ if and only if $i \in I_1$ and $m_i = n + 1$ if and only if $i \in I_0$ is part of a recurrent class of UBRD with outcome z . I prove this by checking that no deviation to a different outcome is compatible with UBRD.

For any agent $i \in I_1$ deviations to any $m_i = k' > k$ will lead to the outcome \underline{z} . Since z is a strict Pareto improvement over \underline{z} for those agents, this is not a better response. Deviations to any $m_i = k' < k$ make outcomes possible in which i contributes but total contribution is less than k . Thus, those messages are exploitable. Thus, no agent in I_1 will change their message according to UBRD. If only agents in I_0 change their messages, total contributions can only increase. No agent $i \in I_0$ will choose any $m_i = k' < k + 2$ since then this agent i would contribute. Since $\theta_i \in [0, 1)$ agent i would be worse off. Thus, this is not a better response for agent i .

Assume now that after some deviations of agents $i \in I_0$ under UBRD, the outcome nevertheless changes from z to z' . Since z was Pareto optimal at least one agent, call him j , is worse off in z' than in z . Since we already noted that no agent in I_1 has any incentive to deviate, total contributions are higher in z' than in z . Thus, $j \in I'_1$ or agent j could not be worse off in z' . This implies that the message of agent j that made the change from z to z' possible was exploitable. Thus, j would not have chosen this message under UBRD. Therefore, z is indeed the outcome of a recurrent class of the UBRD process.

Step2: Take now any outcome $z \in Z$ which is not Pareto optimal or not a strict Pareto improvement over \underline{z} for all i with $\theta_i > 0$. Then I distinguish two cases. In case 1 z is Pareto optimal but not a strict Pareto improvement over \underline{z} for all i with $\theta_i > 0$. Then there exists some agent i , who contributes, but would be better off by or indifferent to not contributing even if this will lead to \underline{z} . Thus for this agent $m_i = n + 1$ is a (weak) better response. Further $m_i = n + 1$ can never be exploitable. If all other contributing agents chose unexploitable messages the switch to $m_i = n + 1$ will lead to the outcome \underline{z} . From \underline{z} the dynamics reach

any recurrent class with Pareto optimal outcome z , which is a *strict** Pareto improvement over \underline{z} , with positive probability. All messages in any such recurrent class are unexploitable better responses, whenever the current outcome is \underline{z} .

In case 2 z is not Pareto optimal. Then there exists a Pareto optimal outcome z' , which is a Pareto improvement over z . Assume that in z' , k' agents will contribute. Then for those agents who contribute in z' but not in z , $m_i = k'$ is an unexploitable better response. Once all those agents play $m_i = k'$, the outcome switches to z' . Thus the dynamics reach z' with positive probability. Now z' is either a Pareto optimum which is a *strict** Pareto improvement over \underline{z} , or we are in case 1.

For the second part assume that there does not exist any outcome, which is Pareto optimal and a *strict** Pareto improvement over \underline{z} . I show in three steps that in this case there is only one recurrent class and it contains all outcomes, which are weak Pareto improvements over \underline{z} .

Step 1: In this step I show that from all outcomes which are not part of the recurrent class the dynamics reach the recurrent class with strictly positive probability. In all outcomes, which are not weak Pareto improvements over \underline{z} , some agents must be worse off than in \underline{z} . Assume now that all those agents choose to free-ride in the next period, which is an Unexploitable Better Response. Assume that all other agents that currently contribute condition their contribution in the next period on the current total level of contributions. Then the outcome in the next period will be \underline{z} .

Step 2: In this step I show that the recurrent class is never left under UBRD. Assume that the current outcome is some weak Pareto improvement over \underline{z} . Assume further that after some deviations according to UBRD the outcome in the next period is no weak Pareto improvement over \underline{z} . Then at least one agent must now be worse off than in \underline{z} . Since only contributing agents can be worse off than in \underline{z} , this agent's deviation to the message in the current period was exploitable, which is a contradiction.

Step 3: In this step I show that there exists no smaller recurrent class inside of the one

described above. Assume that the current outcome in period t is $z^t = \underline{z}$. And let z' be any weak Pareto improvement over z different from \underline{z} . Let $k' = \sum_{i=1}^n z'_i$. If all agents $i \in I'_1$ choose $m_i^{t+1} = k'$ and all other agents choose $m_j^{t+1} = n + 1$, then $z^{t+1} = z'$. It is easy to see that all these messages are Unexploitable Better Responses. Assume next that the current outcome z is any weak Pareto improvement over \underline{z} . Since this Pareto improvement over \underline{z} can not be *strict** by assumption, there is at least one contributing agent that is indifferent between z and \underline{z} . If those agents and all agents in I_0 choose $m_i^{t+1} = n + 1$ in the next period and all other agents in I_1 choose to condition on the current total level of contributions, the outcome in the next period will be $z^{t+1} = \underline{z}$. It is again easy to verify that all these messages are Unexploitable Better Responses. This demonstrates that any outcome in the recurrent class is reached from any other such outcome with strictly positive probability. \square

Proof of Theorem 6.1 Let $z := (z_1, \dots, z_n) \in Z$ be an outcome, such that $z \succeq_i \underline{z} \forall i \in I$, and define $\bar{y} := \sum_{i=1}^n z_i$. Then $m_i = (z_i, \bar{y})$ is a Nash-equilibrium of the mechanism with outcome z . There are four ways in which any agent i can deviate from this message. He can increase or decrease his proposed contribution. And he can increase or decrease his condition.

Any decrease in the offered contribution will fail to satisfy all other agents conditions and can thus only lead to outcomes, which are worse for agent i , no matter what condition he chooses.

Any (weak) increase in the offered contribution will not lead to an increase of other agents' contributions. Thus, such an increase combined with a condition that can be satisfied will only lead to a (weakly) higher contribution by agent i . If the increase in the offered condition is combined with a condition that can not be satisfied the outcome will be \underline{z} . In both cases agent i is (weakly) worse off. Thus, no agent has any incentive to deviate and m is a Nash equilibrium.

Let now $z \in Z$ be an outcome such that some agent i strictly prefers \underline{z} to z . Given any message profile m' leading to the outcome z agent i can profitably deviate to $m''_i = (0, 0)$. This gives him an outcome which is at least as good as \underline{z} and thus strictly better than z .

Therefore, there is no message profile that makes z a Nash equilibrium outcome. \square

Proof of Theorem 6.4 I prove this theorem in two steps. In step 1 I prove that the described outcomes are indeed outcomes of recurrent classes of UBRD. And in step 2 I prove that from any other outcome the dynamics reach one of those recurrent classes with strictly positive probability.

Step1: In the discussion of the environment I assumed that there exists some Pareto improvement z over \underline{z} , which is strict for all $i \in I_1$. Take then any Pareto optimal outcome z' , which is a Pareto improvement over z . Then z' is a Pareto optimal outcome, which is strict for all $i \in I'_1$. Assume to the contrary that some $i \in I'_1$ were indifferent between z' and \underline{z} , then his valuation θ_i must be positive. But then i was either better off in z than in z' if $i \in I_0$, or he was worse off in z than in \underline{z} if $i \in I_1$. Both possibilities lead to a contradiction. Note further that any Pareto improvement z over \underline{z} , which is strict for all $i \in I_1$ is further strict for all agents i with $\theta_i > 0$.

Thus, there exists a Pareto optimal outcome $z \in Z$, which is a strict Pareto improvement over \underline{z} for all agents i with $\theta_i > 0$. Let z be such an outcome and define $\bar{y} := \sum_{i=1}^n z_i$. Then $x_i^1 = x_i^2 = z_i$ and $y_i^1 = y_i^2 = \bar{y}$ is part of a recurrent class of UBRD with outcome z . Assume to the contrary that after deviations of some agents consistent with UBRD the outcome changes from z to some $z' \neq z$. Note that $z' \neq z$ implies in this environment that not all agents are equally well off in z' as in z . Then at least one agent is worse off in z' than in z (otherwise this would be a Pareto improvement over z). If one of the agents who is worse off contributes in z' a strictly positive amount then his message that led to the outcome z' was either exploitable or no better response and he would not have chosen it in UBRD. Thus, all agents, who are worse off in z' than in z , need to contribute zero in z' . Assume to the contrary that in the group of the other agents who are equally well or better off in z' than in z there are some agents who contribute more in z' than in z . Then it would be a Pareto improvement over z if those agents made the contributions as in z' , while all other agents made contributions as in z . This cannot be the case since z was Pareto optimal. Thus, all

agents contribute weakly less in z' than in z . This implies that total contributions are lower in z' than in z . Then there is one agent in this group whose contribution sank relatively to the contributions in z by the lowest percentage. If this agent is better off in z' than in z he would still be better off in \underline{z} since the valuation of the public good is linear. This contradicts that z was a *strict** Pareto improvement over \underline{z} . This yields a contradiction and thus it is not possible that the outcome changes under UBRD once the described message profile is reached.

Step2: Assume now that the current outcome z is not Pareto optimal. Then there exists a Pareto improvement z' over z such that z' is Pareto optimal. Define again $\bar{y} := \sum_{i=1}^n z_i$ and $\bar{y}' := \sum_{i=1}^n z'_i$. Then for any agent i the message $x_i^1 = z_i, y_i^1 = \bar{y}, x_i^2 = z'_i, y_i^2 = \bar{y}'$ is an unexploitable better response to their current message. If all agents choose this message the outcome will be z' . Thus the dynamics reach this message profile with strictly positive probability. Once it is reached the new outcome is z' and now $x_i^1 = z'_i, y_i^1 = \bar{y}', x_i^2 = z'_i, y_i^2 = \bar{y}'$ is an unexploitable better response for all agents. Thus from any not Pareto optimal outcome a message profile, like the one in the first part of this proof, is reached with strictly positive probability.

If z' is a strict Pareto improvement over \underline{z} for all agents i with $\theta_i > 0$ the proof is complete. If it is not, then there exists some agent $i \in I'_1$ who is at least as well off in \underline{z} as in z' . For this agent the message $x_i^1 = 0, y_i^1 = 0, x_i^2 = 0, y_i^2 = 0$ in an unexploitable better response. Thus the dynamics move from any Pareto optimum like z' to \underline{z} with positive probability. From \underline{z} any Pareto optimal allocation, which is a *strict** Pareto improvement over \underline{z} , is reached with positive probability in the way described above. \square

Proof of Theorem 6.10 In the first part of the proof I show that any core outcome z , which is *strict** for $S^C(z)$, is an outcome of recurrent classes of the dynamics.

Let z be an outcome of the mechanism and let z be a core allocation, which is *strict** for $S^C(z)$. Define $\bar{y} := \sum_{i=1}^n z_i$. Then $x_i^1 = x_i^2 = z_i$ and $y_i^1 = y_i^2 = \bar{y}$ is part of a recurrent class of UBRD with outcome z . Assume to the contrary that after deviations of some agents

consistent with UBRD the outcome changes to some $z' \neq z$. Then at least one agent $i \in I'_1$ is worse off in z' than in z (otherwise this would be a coalition improvement over z). Agent i 's message, which led to the outcome z' , was thus either exploitable or no better response and he would not have chosen it in UBRD.

In the second part of the proof I show that from all other allocations the dynamics move with strictly positive probability to a core allocation, which is *strict** for $S^C(z)$.

Assume that the dynamics are in a state with some outcome z , which is not Pareto optimal and let z' be any Pareto optimal allocation, which is a Pareto improvement over z . Define $\bar{y} := \sum_{i=1}^n z_i$ and $\bar{y}' := \sum_{i=1}^n z'_i$. Then the message $(z_i, \bar{y}), (z'_i, \bar{y}')$ is an unexploitable better response for any agent i . Thus the dynamics move with strictly positive probability from z to any such z' .

I can thus assume that the dynamics are in a state with some outcome z , which is Pareto optimal, but not a core outcome that is strict for $S^C(z)$. Then there exists a coalition S and an outcome $z' \in Z_S$ such that all agents $i \in I'_1$ are at least as well off in z' than in z . This implies that $\bar{y}' := \sum_{i=1}^n z'_i < \bar{y} := \sum_{i=1}^n z_i$ or this would be a Pareto improvement. Then in a first step the messages $(z_i, \bar{y}), (z'_i, \bar{y}')$ are unexploitable better responses for every agent $i \in I'_1$. Once all agents $i \in I'_1$ switched to those messages, the messages $(z'_i, \bar{y}'), (z'_i, \bar{y}')$ and $(z_i, \bar{y}), (z_i, \bar{y})$ are both unexploitable better responses for those agents, since the current outcome is still z . But if now simultaneously one agent chooses $(z'_i, \bar{y}'), (z'_i, \bar{y}')$ and another one chooses $(z_j, \bar{y}), (z_j, \bar{y})$, then contribution breaks down entirely and the outcome will be \underline{z} . From \underline{z} any core allocation, which is a Pareto improvement over \underline{z} and *strict** for $S^C(z)$ will be reached with strictly positive probability in the way described above. \square

Proof of theorem 6.11 I prove this theorem in two steps. In step 1, I show that it is possible to design arbitrarily cheap incentive schemes, such that no agent is indifferent between any two outcomes. In step 2, I show that this leads to the existence of a core outcome in the given environment. Finally, when every agent has a strict preference between any two outcomes then any core outcome is *strict** for all subsets of agents. Thus, there exists a core

outcome z , which is strict for $S^C(z)$.

Step 1: Let $\epsilon > 0$. Define $\epsilon' := \min_{i \in I} \min_{z, z' \in Z: u_i(z) \neq u_i(z')} |u_i(z) - u_i(z')|$ as the smallest positive difference in utility between any two outcomes for any agent. Let $N_Z := \#Z$ be the number of possible outcomes and let $r : Z \rightarrow \{1, \dots, N_Z\}$ be any bijective mapping, which satisfies $\sum_{i=1}^n z_i > \sum_{i=1}^n z'_i \Rightarrow r(z) > r(z')$. Let $m \in \mathbb{N}$. Define the mapping $u_{zim}^+ = \frac{r(z)\epsilon'}{2nmN_Z} \forall i \in I$. Interpret u_{zim}^+ as the utility increment we want to add to agent i 's utility in outcome z . Since u_i is strictly increasing in w_i and since $u_{zim}^+ < \epsilon' \forall z \in Z, i \in I$ there exists a payment $\Delta_{zim} < 0.01$ that adds exactly the utility u_{zim}^+ . If the payments Δ_m are defined in this way, no agent is indifferent between any two outcomes, since this would imply.

$$u_i(z) + u_{zim}^+ = u_i(z') + u_{z'im}^+ \Leftrightarrow u_i(z) - u_i(z') = u_{zim}^+ - u_{z'im}^+ \quad (10)$$

The absolute value of the left-hand side of the second equation is either equal to zero or weakly bigger than ϵ' . However, since $r(z) \neq r(z')$ the absolute value of the right-hand side is strictly bigger than zero and strictly smaller than ϵ' . This leads to a contradiction. Therefore, adding the payments Δ_m leads to a mechanism in which no agent is indifferent between any two outcomes. Further, out of continuity of u_i in w_i there exists an $m \in \mathbb{N}$ such that total expected costs of Δ_m are less than ϵ .

Step 2: I prove this step by induction over the number of agents in the economy. For the beginning assume there are $n = 1$ agents. Then existence of a core outcome is equivalent to the existence of an outcome which gives the agent maximal utility. Since our state space is finite this is trivial. Thus, one may assume that for an economy with $n = k$ agents there exists a core outcome. Let's now look at an economy with $n = k + 1$ agents. Call the coalition of agents 1 through k in this economy C . Then by assumption there is an outcome z , with $z_{k+1} = 0$, from which no subcoalition of C can improve. I call this a core outcome in the coalition C . Let z' be the Pareto optimal Pareto improvement over z , in which agent $k + 1$ gets the highest utility. Then no subcoalition of C can improve on z' . Otherwise z could not have been a core outcome in coalition C . Assume to the contrary

a coalition C' including agent $k + 1$ can improve from z' to an outcome z'' . Then total contributions are less in z'' than in z' or this would be a further Pareto improvement. Then $z''' := (\max\{z_1, z_1''\}, \dots, \max\{z_k, z_k''\}, z_{k+1}'')$ is a Pareto improvement over z in which agent $k + 1$ is better off than in z'' (since $\sum_{i=1}^n z_i > \sum_{i=1}^n z_i' \Rightarrow r(z) > r(z')$) and thus better off than in z' . This contradicts the assumptions on z' . Thus, no coalition can improve on z' and therefore z' is in the core. \square

Appendix B

This Appendix covers the experiment instructions. They are translations from the German original. The German version can be obtained on request from the author. The different instructions for the four treatments are given in the following order: 1.) VCM, complete information, 2.) CCM, complete information, 3.) VCM, incomplete information 4.) CCM, incomplete information.

7.1 Instructions for VCM with complete information

Instructions

Welcome to our experiment! Please read the instructions carefully. Do not talk to your neighbor from now on. Shut down your mobile phone and keep it turned off until the experiment ends. If you have any questions, raise your hands. We will come to you. All participants have got the same instructions.

In the experiment you will be divided in **groups of 5**. The experiment will last for **20 periods**. You will be grouped with the same four players in all periods. The experiment is entirely anonymous. No player will be informed whom he was grouped up with or what payoff any other player obtains.

Points are the currency in the experiment. In every period you start with 10 points. These points will by the end of the period either be added entirely to your private account,

or will be invested entirely into a common project. For every player who invests his **10 points** into the project all players obtain **6 points**.

Example 1: You invest you 10 points into the project and 2 other players invested into the project additionally. You will get for your investment and for the investment of the other 2 players 6 points each. Thus you will get $3 \times 6 = 18$ points in total added to your account.

Example 2: You do not invest your 10 points into the project and 2 players invested into the project in total. Your will keep your 10 points and get additionally 6 points each for the investment of the other 2 players. Thus you get $10 + 2 \times 6 = 22$ points in total added to your account.

Every player can choose in every period between two actions:

- You can invest your 10 points into the project.
- Or you can keep your 10 points for yourself.

All players decide simultaneously.

Payoff of all periods

After it was determined who contributes to the project in a given period, all players get the corresponding points added to their account.

Then a new period starts. After 20 periods there will be a questionnaire. After the experiment you will be called to receive your money. You will receive your **earnings for all periods** at a rate of 40 points=1€. The payment will be private and in cash.

Program structure

You obtained a printed example for the structure of the program, which you will use to submit your decision in every period. The screen is divided into three blocks.

The block on the upper left side contains a calculator. Here you can test actions for you and the four other players. Once you select an action for every player the computer will calculate the payoff you would obtain in this case.

In the upper right block you enter the action that will be relevant for your payoff. Below there is a red button. When you push this button you submit your decision and leave the screen. Only when **all** players pushed the button the experiment continues. A clock on the upper right hints at the time in which your decision should be made. **If the time runs out this has no effect.**

From period two on the actions of all players of all previous periods and your payoff in those periods will be displayed in the big block below. In the first period this block will be empty.

The texts in the green frames on the printed example of the program are comments that explain the print. They will not be displayed in the actual program.

7.2 Instructions for CCM with complete information

Instructions

Welcome to our experiment! Please read the instructions carefully. Do not talk to your neighbor from now on. Shut down your mobile phone and keep it turned off until the experiment ends. If you have any questions, raise your hands. We will come to you. All participants have got the same instructions.

In the experiment you will be divided in **groups of 5**. The experiment will last for **20 periods**. You will be grouped with the same four players in all periods. The experiment is entirely anonymous. No player will be informed whom he was grouped up with or what payoff any other player obtains.

Points are the currency in the experiment. In every period you start with 10 points. These points will by the end of the period either be added entirely to your private account, or will be invested entirely into a common project. For every player who invests his **10 points** into the project all players obtain **6 points**.

Example 1: You invest you 10 points into the project and 2 other players invested into the project additionally. You will get for your investment and for the investment of the other

2 players 6 points each. Thus you will get $3 \times 6 = 18$ points in total added to your account.

Example 2: You do not invest your 10 points into the project and 2 players invested into the project in total. You will keep your 10 points and get additionally 6 points each for the investment of the other 2 players. Thus you get $10 + 2 \times 6 = 22$ points in total added to your account.

Every player can **choose** in every period **between six different conditions**:

- 0=Contribute in any case.
- 1=Contribute only if at least one other player contributes, too.
- 2=Contribute only if at least two other players contribute, too.
- 3=Contribute only if at least three other players contribute, too.
- 4=Contribute only if all four other players contribute, too.
- 5=Contribute in no case.

The computer selects the highest amount of players, which can contribute to the project, without violation the condition of any player. These players will then automatically contribute to the project. The other players will not contribute.

Example 1: 3 players choose condition "1" and the other two players choose condition "5". Then those 3 players, who chose condition "1" will contribute to the project.

Example 2: 3 players choose condition "3" and the other two players choose condition "5". Then no player will contribute to the project.

Payoff of all periods

After it was determined who contributes to the project in a given period, all players get the corresponding points added to their account.

Then a new period starts. After 20 periods there will be a questionnaire. After the experiment you will be called to receive your money. You will receive your **earnings for all periods** at a rate of 40 points=1€. The payment will be private and in cash.

Program structure

You obtained a printed example for the structure of the program, which you will use to submit your decision in every period. The screen is divided into three blocks.

The block on the upper left side contains a calculator. Here you can test conditions for you and the four other players. Once you select a condition for every player the computer will calculate the payoff you would obtain in this case.

In the upper right block you enter the condition that will be relevant for your payoff. Below there is a red button. When you push this button you submit your decision and leave the screen. Only when **all** players pushed the button the experiment continues. A clock on the upper right hints at the time in which your decision should be made. **If the time runs out this has no effect.**

From period two on the conditions of all players of all previous periods and your payoff in those periods will be displayed in the big block below. In the first period this block will be empty.

The texts in the green frames on the printed example of the program are comments that explain the print. They will not be displayed in the actual program.

7.3 Instructions for VCM with incomplete information

Instructions

Welcome to our experiment! Please read the instructions carefully. Do not talk to your neighbor from now on. Shut down your mobile phone and keep it turned off until the experiment ends. If you have any questions, raise your hands. We will come to you. All participants have got the same instructions.

In the experiment you will be divided in **groups of 5**. The experiment will last for **20 periods**. You will be grouped with the same four players in all periods. The experiment is entirely anonymous. No player will be informed whom he was grouped up with or what payoff any other player obtains.

Points are the currency in the experiment. In every period you start with 10 points. These points will by the end of the period either be added entirely to your private account, or will be invested entirely into a common project.

At the beginning of the first period every player will be assigned one **type**, which he will keep **for the entire game**.

With a chance of 20% you are type 1 and you do not benefit from the common project. In this case in each period your 10 points will be added to your private account if you do not invest them into the project. And 0 points will be added to your private account if you invest into the project. If any other players invest into the project does not influence your payoff in this case.

With a chance of 80% you are type 2 and you benefit from the common project. In this case in each period your 10 points will be added to your private account as well if you do not invest them into the project, but 6 points will be added to your private account if you invest into the project. Additionally you receive 6 points for every other player, who also invests into the project.

The types are drawn independently, especially different players may thus have different types. Every player gets displayed his type in every period. He does not get to know the types of the other players.

Example 1: You are type 2 and you invest you 10 points into the project and 2 other players invested into the project additionally. You will get for your investment and for the investment of the other 2 players 6 points each. Thus you will get $3 \times 6 = 18$ points in total added to your account.

Example 2: You are type 2 and you do not invest your 10 points into the project and 2 players invested into the project in total. Your will keep your 10 points and get additionally 6 points each for the investment of the other 2 players. Thus you get $10 + 2 \times 6 = 22$ points in total added to your account.

Example 3: You are type 1 and you do not invest your 10 points into the project and 2 players invested into the project in total. You will keep your 10 points and get no additional points for the investment of the other 2 players. Thus you get 10 points in total added to your account.

Every player can choose in every period between two actions:

- You can invest your 10 points into the project.
- Or you can keep your 10 points for yourself.

All players decide simultaneously.

Payoff of all periods

After it was determined who contributes to the project in a given period, all players get the corresponding points added to their account.

Then a new period starts. After 20 periods there will be a questionnaire. After the experiment you will be called to receive your money. You will receive your **earnings for all periods** at a rate of 40 points=1€. The payment will be private and in cash. If you are type 1 you will receive 5€ additionally to compensate for your lower earning possibilities.

Program structure

You obtained a printed example for the structure of the program, which you will use to submit your decision in every period. The screen is divided into three blocks.

The block on the upper left side contains a calculator. Here you can test actions for you and the four other players. Once you select an action for every player the computer will calculate the payoff you would obtain in this case.

In the upper right block you enter the action that will be relevant for your payoff. Additionally in this block your type is displayed and whether you benefit from the project. Below there is a red button. When you push this button you submit your decision and leave the

screen. Only when **all** players pushed the button the experiment continues. A clock on the upper right hints at the time in which your decision should be made. **If the time runs out this has no effect.**

From period two on the actions of all players of all previous periods and your payoff in those periods will be displayed in the big block below. In the first period this block will be empty.

The texts in the green frames on the printed example of the program are comments that explain the print. They will not be displayed in the actual program.

7.4 Instructions for CCM with incomplete information

Instructions

Welcome to our experiment! Please read the instructions carefully. Do not talk to your neighbor from now on. Shut down your mobile phone and keep it turned off until the experiment ends. If you have any questions, raise your hands. We will come to you. All participants have got the same instructions.

In the experiment you will be divided in **groups of 5**. The experiment will last for **20 periods**. You will be grouped with the same four players in all periods. The experiment is entirely anonymous. No player will be informed whom he was grouped up with or what payoff any other player obtains.

Points are the currency in the experiment. In every period you start with 10 points. These points will by the end of the period either be added entirely to your private account, or will be invested entirely into a common project.

At the beginning of the first period every player will be assigned one **type**, which he will keep **for the entire game**.

With a chance of 20% you are type 1 and you do not benefit from the common project. In this case in each period your 10 points will be added to your private account if you do not invest them into the project. And 0 points will be added to your private account if you

invest into the project. If any other players invest into the project does not influence your payoff in this case.

With a chance of 80% you are type 2 and you benefit from the common project. In this case in each period your 10 points will be added to your private account as well if you do not invest them into the project, but 6 points will be added to your private account if you invest into the project. Additionally you receive 6 points for every other player, who also invests into the project.

The types are drawn independently, especially different players may thus have different types. Every player gets displayed his type in every period. He does not get to know the types of the other players.

Example 1: You are type 2 and you invest you 10 points into the project and 2 other players invested into the project additionally. You will get for your investment and for the investment of the other 2 players 6 points each. Thus you will get $3 \times 6 = 18$ points in total added to your account.

Example 2: You are type 2 and you do not invest your 10 points into the project and 2 players invested into the project in total. Your will keep your 10 points and get additionally 6 points each for the investment of the other 2 players. Thus you get $10 + 2 \times 6 = 22$ points in total added to your account.

Example 3: You are type 1 and you do not invest your 10 points into the project and 2 players invested into the project in total. Your will keep your 10 points and get no additional points for the investment of the other 2 players. Thus you get 10 points in total added to your account.

Every player can **choose** in every period **between six different conditions**:

- 0=Contribute in any case.
- 1=Contribute only if at least one other player contributes, too.
- 2=Contribute only if at least two other players contribute, too.

- 3=Contribute only if at least three other players contribute, too.
- 4=Contribute only if all four other players contribute, too.
- 5=Contribute in no case.

The computer selects the highest amount of players, which can contribute to the project, without violation the condition of any player. These players will then automatically contribute to the project. The other players will not contribute.

Example 1: 3 players choose condition "1" and the other two players choose condition "5". Then those 3 players, who chose condition "1" will contribute to the project.

Example 2: 3 players choose condition "3" and the other two players choose condition "5". Then no player will contribute to the project.

Payoff of all periods

After it was determined who contributes to the project in a given period, all players get the corresponding points added to their account.

Then a new period starts. After 20 periods there will be a questionnaire. After the experiment you will be called to receive your money. You will receive your **earnings for all periods** at a rate of 40 points=1€. The payment will be private and in cash. If you are type 1 you will receive 5€ additionally to compensate for your lower earning possibilities.

Program structure

You obtained a printed example for the structure of the program, which you will use to submit your decision in every period. The screen is divided into three blocks.

The block on the upper left side contains a calculator. Here you can test conditions for you and the four other players. Once you select a condition for every player the computer will calculate the payoff you would obtain in this case.

In the upper right block you enter the condition that will be relevant for your payoff. Additionally in this block your type is displayed and whether you benefit from the project.

Below there is a red button. When you push this button you submit your decision and leave the screen. Only when **all** players pushed the button the experiment continues. A clock on the upper right hints at the time in which your decision should be made. **If the time runs out this has no effect.**

From period two on the conditions of all players of all previous periods and your payoff in those periods will be displayed in the big block below. In the first period this block will be empty.

The texts in the green frames on the printed example of the program are comments that explain the print. They will not be displayed in the actual program.

Appendix C

In addition to instructions agents had to fill out a slide of comprehension questions. A translation of the German original is given exemplary for the case of the CCM and incomplete information:

Comprehension questions - Experiment PGCCM

You are asked to complete two test questions to check whether you understood the instructions completely.

Choose in the following test question 1 a condition for each player. Choose at least **three different** conditions:

Your condition (player 1): ____

Condition player 2: ____

Condition player 3: ____

Condition player 4: ____

Condition player 5: ____

Underline those players, who would contribute to the project in this case:

Player 1 Player 2 Player 3 Player 4 Player 5

What payoff would you obtain in this period if you are of type 2? _____

Choose also in the following test question 2 a condition for each player. Choose at least **three different** conditions, such that the number of players, who contribute to the project,

differs in test question 1 and 2:

Your condition (player 1): ____

Condition player 2: ____

Condition player 3: ____

Condition player 4: ____

Condition player 5: ____

Underline those players, who would contribute to the project in this case:

Player 1 Player 2 Player 3 Player 4 Player 5

What payoff would you obtain in this period if you are of type 1? _____

References

- Bag, P. K. (1997). Public goods provision: Applying jackson-moulin mechanism for restricted agent characteristics. *Journal of Economic Theory*, 73(2):460 – 472.
- Bagnoli, M., Ben-David, S., and McKee, M. (1992). Voluntary provision of public goods. *Journal of Public Economics*, 47(1):85 – 106.
- Bagnoli, M. and Lipman, B. L. (1989). Provision of public goods: Fully implementing the core through private contributions. *The Review of Economic Studies*, 56(4):583–601.
- Bracht, J., Figuières, C., and Ratto, M. (2008). Relative performance of two simple incentive mechanisms in a public goods experiment. *Journal of Public Economics*, 92(1-2):54 – 90.
- Burger, N. E. and Kolstad, C. D. (2009). Voluntary public goods provision, coalition formation, and uncertainty. Working Paper 15543, National Bureau of Economic Research.
- Clarke, E. (1971). Multipart pricing of public goods. *Public Choice*, 11:17–33.
- Corazzini, L., Faravelli, M., and Stanca, L. (2010). A prize to give for: An experiment on public good funding mechanisms*. *The Economic Journal*, 120(547):944–967.
- Danziger, L. and Schnytzer, A. (1991). Implementing the lindahl voluntary-exchange mechanism. *European Journal of Political Economy*, 7(1):55 – 64.
- Falkinger, J. (1996). Efficient private provision of public goods by rewarding deviations from average. *Journal of Public Economics*, 62(3):413–422.
- Falkinger, J., Fehr, E., Gächter, S., and Winter-Ebmer, R. (2000). A simple mechanism for the efficient provision of public goods: Experimental evidence. *The American Economic Review*, 90(1):247–264.
- Fehr, E. and Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, 90(4):980–994.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):pp. 817–868.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters*, 71(3):397 – 404.
- Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *The American Economic Review*, 100(1):pp. 541–556.
- Greiner, B. (2004). The online recruitment system orsee 2.0-a guide for the organization of experiments in economics. *University of Cologne, Working paper series in economics*, 10(23):63–104.

- Groves, T. and Ledyard, J. (1977a). Optimal allocation of public goods: A solution to the "free rider" problem. *Econometrica*, 45(4):783–809.
- Groves, T. and Ledyard, J. (1977b). Some limitations of demand revealing processes. *Public Choice*, 29(2):107–124.
- Healy, P. J. (2006). Learning dynamics for mechanism design: An experimental comparison of public goods mechanisms. *Journal of Economic Theory*, 129(1):114 – 149.
- Healy, P. J. and Mathevet, L. (2012). Designing stable mechanisms for economic environments. *Theoretical Economics*, 7(3):609–661.
- Hirshleifer, D. and Rasmusen, E. (1989). Cooperation in a repeated prisoners' dilemma with ostracism. *Journal of Economic Behavior & Organization*, 12(1):87 – 106.
- Jackson, M. and Moulin, H. (1992). Implementing a public project and distributing its cost. *Journal of Economic Theory*, 57(1):125 – 140.
- Kirchsteiger, G. and Puppe, C. (1997). On the possibility of efficient private provision of public goods through government subsidies. *Journal of Public Economics*, 66(3):489 – 504.
- Kocher, M. G., Cherry, T., Kroll, S., Netzer, R. J., and Sutter, M. (2008). Conditional cooperation on three continents. *Economics Letters*, 101(3):175 – 178.
- Ledyard, J. O. (1994). Public goods: A survey of experimental research. Public Economics 9405003, EconWPA.
- Maskin, E. (1999). Nash equilibrium and welfare optimality*. *The Review of Economic Studies*, 66(1):23–38.
- Morgan, J. and Sefton, M. (2000). Funding public goods with lotteries: Experimental evidence. *The Review of Economic Studies*, 67(4):785–810.
- Ostrom, E., Walker, J., and Gardner, R. (1992). Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86:404–417.
- Owen, G. (1982). *Game Theory*, chapter XIII Games without side payments, pages 288 – 328. Academic Press.
- Rondeau, D., Poe, G. L., and Schulze, W. D. (2005). VCM or PPM? a comparison of the performance of two voluntary public goods mechanisms. *Journal of Public Economics*, 89(8):1581 – 1592.
- Rondeau, D., Schulze, W. D., and Poe, G. L. (1999). Voluntary revelation of the demand for public goods using a provision point mechanism. *Journal of Public Economics*, 72(3):455 – 470.
- Rose, S. K., Clark, J., Poe, G. L., Rondeau, D., and Schulze, W. D. (2002). The private provision of public goods: tests of a provision point mechanism for funding green power programs. *Resource and Energy Economics*, 24:131 – 155. Economist and editor George Tolley - A special issue in his honour.

- Schram, A. J. and Onderstal, S. (2009). Bidding to give: An experimental comparison of auctions for charity*. *International Economic Review*, 50(2):431–457.
- Smith, V. L. (1979). An experimental comparison of three public good decision mechanisms. *The Scandinavian Journal of Economics*, 81(2):pp. 198–215.
- Smith, V. L. (1980). Experiments with a decentralized mechanism for public good decisions. *The American Economic Review*, 70(4):pp. 584–599.
- Varian, H. R. (1994). A solution to the problem of externalities when agents are well-informed. *The American Economic Review*, 84(5):pp. 1278–1293.
- Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37.