The conditional contribution mechanism for repeated public goods - the general case*

Joerg Oechssler, Andreas Reischmann, and Andis Sofianos†
Department of Economics
University of Heidelberg

January 13, 2020

Abstract

We present a new and simple mechanism for repeated public good environments. In the Conditional Contribution Mechanism (CCM), agents send two messages of the form, "I am willing to contribute x units to the public good if in total y units are contributed." This mechanism offers agents risk-free strategies, which we call unexploitable. We prove that if agents choose unexploitable messages in a Better Response Dynamics model, all stable outcomes of the CCM are Pareto efficient. We conduct a laboratory experiment to investigate whether observed behavior is consistent with this prediction. In the complete information case we find that indeed almost 80% of outcomes are Pareto optimal. Furthermore, in comparison to the Voluntary Contribution Mechanism, the CCM leads to significantly higher contribution rates. Even under incomplete information, contributions are fairly high and do not deteriorate over time.

Keywords: Experiment; Public Goods; Mechanism Design; Better Response Dynamics.

JEL-Classification: C72; C92; D82; H41

*We thank Jobst Heitzig, Steffen Huck, Marco Lambrecht, Al Roth, Christoph Vanberg, Georg Weizsäcker and participants of seminars at WZB Berlin, PIK Potsdam, HeiKaMaX, the ESA Los Angeles, and the universities of Göttingen, Heidelberg, and Potsdam for helpful comments. The University of Heidelberg provided the funding for this research.

†Bergheimer Str. 58, 69115 Heidelberg, Germany, email: oechssler@uni-hd.de

1 Introduction

Many public goods have to be provided repeatedly. Currently probably the most prominent example would be climate policy where countries meet repeatedly at climate conferences and try to agree on limiting emissions of greenhouse gases. Another example would be contributions to international organizations like UNICEF or the fund-raising for global public goods like Wikipedia. But there are also numerous examples on a much smaller and local level like contributions to clubs, schools, and universities.

While many mechanisms have been proposed in order to deal with the free-rider problem, most are aimed at static public good problems. When public good problems are (finitely) repeated, this poses new challenges but also creates new opportunities.[1] Thus, we require a mechanism that takes into account the repeated nature of the problem. Furthermore, we desire that a suitable mechanism satisfies voluntary participation of all agents (i.e. individual rationality), even of those who may not benefit (enough) of the public good. For many practical applications a very important further requirement is that neither the mechanism designer nor the other agents are assumed to know the preferences of (other) agents. Finally we require that the mechanism is budget-balanced.

In this paper, we suggest a new and simple mechanism for such repeated public good environments that satisfies all of the above criteria. In the Conditional Contribution Mechanism (CCM), agents send two messages of the form, "I am willing to contribute x units to the public good if in total y units are contributed." The mechanism then simply selects the unique contribution profile that maximizes total contributions from all agents. The static game induced by the mechanisms has many Nash equilibria some of which are Pareto optimal and some are not. Zero-contribution is always a Nash equilibrium. However, if we use the dynamic nature of the game and assume that agents myopically choose better (or best) responses to the previous period's message profile, we may be able to select some Nash equilibria. Indeed, the CCM allows agents to use risk-free strategies that make sure that they are not taken advantage of. We call messages that make sure an agent is not worse off than in the previous period when he is a contributor, "unexploitable strategies". Our main theorem shows that if agents play unexploitable better responses, all dynamically stable outcomes of the CCM are Pareto optimal. Furthermore, for all agents who have any

[1] We do not refer here to the usual Folk theorem results for infinitely repeated public good games as in Pecorino (1999) and Marx and Matthews (2000) or to warm-glow preferences as in Romano and Yildirim (2001).

positive valuation for the public goods, this represents a strict Pareto improvement over zero-contribution.

The intuition why the CCM works well in theory and why it requires two messages from each agent is that one message is required to protect the status quo. By conditioning their contributions on the status quo, agents can prevent other agents from deviating. The second message is used to send a signal to other agents that one would be willing to contribute more if they did as well. This allows the mechanism to escape from getting stuck in a low-contribution outcome.

We conduct a laboratory experiment to investigate whether observed behavior is consistent with this prediction. In the complete information case we find that indeed almost 80% of outcomes are Pareto optimal. Furthermore, comparison treatments with the Voluntary Contribution Mechanism (VCM) show that the CCM leads to significantly higher contribution rates, which do not diminish over time. Under incomplete information, contributions are also fairly high and increasing over time. Surprisingly, the VCM works also quite well and better than under complete information.

The remainder of the paper is structured as follows. Section 2 provides a short survey of previously proposed public good mechanisms, with a focus on environments to which they may be applicable. Section 3 introduces the CCM and provides the equilibrium analysis. Section 4 describes the setting that is used in the laboratory experiment. In Section 5, we present and discuss the experimental results. Section 6 gives a short summary and discussion. Most proofs can be found in Appendix A. Finally, translations of written instructions and test questions can be found in Appendices B and C.

2 Related literature

The review of the whole literature on public good provision is far too extensive and beyond the scope of this paper. Thus we shall concentrate on mechanisms that are closely related. We can classify public good mechanisms by the environment for which they are suitable. As explained above, we require a mechanism that is suitable when the public good provision is repeated and when the mechanism designer does not know the agents' preferences. Furthermore, even the agents themselves need not know other agents' preferences. Finally, the complexity of the mechanism is a concern.

A mechanism that is applicable in all environments is the Voluntary Contribution Mechanism (VCM). As it is by far the most studied mechanism in the experimental literature, we shall also consider it as a benchmark. The VCM has the major disadvantage that it does not lead to Pareto efficient outcomes in equilibrium. In experiments, contribution rates decrease over time when the VCM is played repeatedly. See Ledyard (1995) for a survey on this literature. Therefore, when efficiency is important, the VCM may be more suitable for one-shot environments. Furthermore, the VCM may also be suitable for environments in which simplicity is more important than Pareto efficiency. This should apply, for example to most cases of charitable fund-raising.

Communication, in particular face-to-face communication, has been shown to substantially increase efficiency of the VCM in public good games (see e.g. Dawes et al., 1977 and Isaac and Walker, 1988). However, communication is less successful when it is anonymous and structured, for example when subjects can send only a number which indicates their suggested contribution (Bochet et al., 2006).

Another example of a simple mechanism is the Provision Point Mechanism (PPM), which shares with the CCM that the messages are conditional in nature. Under the PPM, public goods are only provided if total contributions exceed a predefined threshold. While the PPM is also applicable in all environments, the mechanism designer has to choose an optimal threshold. When the public good is binary, as for example in the "one streetlight problem" of Bagnoli and Lipman (1989), the optimal threshold is obvious. However, in other cases, like the "multiple streetlight problem" Bagnoli and Lipman (1989) or for continuous public goods, finding the optimal threshold may not be trivial.[2] The PPM may therefore be suited for environments in which there is an obvious optimal threshold. Instead, the CCM makes it possible for agents to find an optimal size for the public good via endogenously adjusting their individual conditions.[3]

Probably the first paper to explicitly model conditional contributions was Guttman (1978).[4] Contributions to a public good are modeled as a two-stage game in which agents announce in the first stage a "matching rate", which commits them to match the unconditional contributions of all other agents with a chosen rate. In the second stage, agents select the unconditional contributions. Under some assumptions the unique subgame perfect equilibrium implies Pareto efficient investments into the public good. However, two of

[2] Suppose a community wants to build a playground. The size (and costs) of the playground may vary widely and the city government would not know the optimal size without knowing the citizens' preferences.

[3] Another mechanism that has been found to sustain contribution levels in the case where an obvious threshold can be identified is the Hired Gun Mechanism, which incorporates punishment towards the lowest contributors (Andreoni and Gee, 2012, 2015).

[4] See also Guttman (1986).

the needed assumptions make it unsuitable for our purposes. First, agents need to know the marginal per capita return of all other agents. Second, utilities must be strictly concave in the public good ruling out the linear public good case. More recently, Heitzig (2019) shows for a very general framework of games including some public good games that a mechanism that enforces conditional commitment functions yields outcomes that are in the core and are therefore Pareto optimal. When strong equilibria in the sense of Aumann (1959) are selected, these are the only outcomes.

MacKay et al. (2015) and Schmidt and Ockenfels (2019) study mechanisms that are particularly geared towards reducing CO2 emissions through carbon pricing. Given the difficulty of international agreements Schmidt and Ockenfels (2019) suggest a clever mechanism that would implement for all participating countries the smallest commitment made by a country. This way, the pivotal country has an incentive to increase its commitment as this would mandate an increase from all participating countries. The whole game has multiple equilibria but in an experiment Schmidt and Ockenfels (2019) show that the mechanism performs quite well.

The approach most closely related to our work are the contractive mechanisms by Healy and Mathevet (2012). They design mechanisms that implement Pareto efficient outcomes, are budget-balanced, and individually rational. Furthermore, they are designed with a dynamic adjustment process in mind. The main challenge for these mechanisms might be their complexity. Especially when budget-balance is guaranteed off equilibrium, the outcome functions of the mechanisms become very complex and may be hard to comprehend for agents. Whether this complexity is a problem will be clearer once there is experimental evidence for these mechanisms.

One of the most well-known public good mechanisms is the Vickrey-Clarke-Groves (VCG) mechanism (Vickrey, 1961, Clarke, 1971, Groves and Ledyard, 1977). In general, the VCG mechanism requires that agents state their benefit of the public good, and then public good only provided if it is socially desirable. An agent's payment or contribution to the public good is determined by the externality he imposes on the other agents by choosing his message. The major drawbacks of the VCG mechanisms are two-fold. They are not budget balanced and they do not satisfy individual rationality. However, in cases where a central authority is present, participation can be seen as enforceable and an unbalanced budget, while not desirable, might at least be possible.

3 Theoretical analysis

We consider a set of agents $I := \{1, \ldots, n\}$ who interact over a given number of periods. There is one public and one private good. Each agent $i \in I$ has an endowment of one monetary unit in each period and decides how much of the endowment he wants to invest into the public good. We denote the investment of agent i into the public good by $z_i^t \in [0,1]$. Here $z_i^t = 1$ is interpreted as agent i investing all of his endowment in period t into the public good and $z_i^t = 0$ represents agent i investing all of his monetary unit into the private good. An outcome in a period t is defined as $z^t = (z_1^t, ..., z_n^t)$, $\forall i \in I$. Let $Z := [0,1]^n$ be the outcome space. For notational convenience define $\underline{z} := (0, \ldots, 0)$ as the zero-contributions outcome in which no agent invests anything into the public good.

We assume utility to be linear in both the private good and the public good. Agent i's valuation for the public good is given by $\theta_i \in [0,1)$.[5] The utility of agent i in period t is given by

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

Let $\succeq_i$ denote the corresponding preference relation.

Our setting allows for agents who have no use at all for the public good ($\theta_i = 0$). To talk meaningfully about strict Pareto improvements in this environment it is useful to define the following.

Definition 1 *An outcome z' is a strict* Pareto improvement over z if z' is a Pareto improvement over z that is strict for all agents with $\theta_i > 0$.*[6]

To make the problem interesting, we assume that preferences are such that there is scope for Pareto improvements by contributing to the public good.

Assumption 1 *There exists a strict* Pareto improvement over $\underline{z} := (0, 0, ..., 0)$.*

[5] Values $\theta_i < 0$ 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 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 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.

[6] Agents who do not benefit from the public good ($\theta_i = 0$) are indifferent between all outcomes in which they do not contribute. If these agents were excluded by assumption, this special definition of strict would not be necessary.

Lemma 1 *Assumption 1 is satisfied if and only if*

$$\sum_{j=1}^{n} \theta_j > 1.$$

Of course, the mechanism proposed below can also be applied in situations in which $\sum_{j=1}^{n} \theta_j < 1$. However, then in any profile with positive contribution levels there would be some agent who is worse off than in the no contribution outcome $\underline{z}$. Thus any mechanism that is based on voluntary participation can only implement $\underline{z}$.

Next we can characterize Pareto efficient outcomes.

Lemma 2 *An outcome profile $z \in Z$ is Pareto optimal if and only if $\sum_{i \in C} \theta_i \leq 1$, where $C := \{i : z_i < 1\}$.*

Corollary 1 *If z is Pareto optimal, then so is any $z' \geq z$.*

Proof. This follows directly from Lemma 2 by noting that $C' := \{i : z_i' < 1\} \subset C$. $\square$

3.1 The Conditional Contribution Mechanism

We define the Conditional Contribution Mechanism (CCM) as $G^{CCM} := (M^{CCM}, g^{CCM})$, where M^{CCM} describes the mechanism's message space and $g^{CCM} : M^{CCM} \mapsto Z$ describes the mechanism's outcome function. The CCM will allow agents to send *two* messages (α_i, β_i) of the form "I am willing to contribute α_i to the public good if total contributions are at least β_i." As a special case, this allows agents to unconditionally contribute nothing $\{(\alpha_i^1, \beta_i^1), (\alpha_i^2, \beta_i^2)\} = \{(0,0), (0,0)\}$ or to unconditionally contribute fully $\{(1,0), (1,0)\}$.[7] We assume for technical reasons that messages come from a finite message space, i.e. $(\alpha_i, \beta_i) \in \{0, \delta, 2\delta, ..., 1-\delta, 1\} \times \{0, \delta, 2\delta, ..., n-\delta, n\} =: M_i$.[8]

Since agents are allowed to send two messages, the message space in the CCM is given by $M^{CCM} := \prod_{i=1}^{n} M_i^{CCM}$, where $M_i^{CCM} := M_i \times M_i$.

[7] In case the two messages are such that $\alpha_i^1 \geq \alpha_i^2$ but $\beta_i^2 > \beta_i^1$, we consider the second message as redundant.

[8] However, δ can be arbitrarily small. Henceforth, we will assume that the continuous message space is sufficiently well approximated by the discrete one such that cumbersome case distinctions due to the discreteness can be avoided.

The outcome $g^{CCM}(m)$ of the CCM is then defined 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 $m \in M^{CCM}$:

$$z \in Z^{CCM}(m) \Leftrightarrow \; z_i = 0 \text{ or } \left\{ \exists l_i \in \{1,2\} : \; z_i = \alpha_i^{l_i} \text{ and } \sum_{j=1}^{n} z_j \geq \beta_i^{l_i} \right\}, \; \forall \, i \in I. \tag{2}$$

It is easy to see that $z \in Z^{CCM}(m)$ and $z' \in Z^{CCM}(m)$ imply together $z'' = (\max\{z_1, z_1'\}, \ldots, \max\{z_n, z_n'\}) \in Z^{CCM}(m)$. Thus, the outcome of the mechanism can be uniquely defined as

$$g^{CCM}(m) = \underset{z \in Z^{CCM}(m)}{\arg\max} \sum_{i=1}^{n} z_i. \tag{3}$$

Let $T(m) := \max_{z \in Z^{CCM}(m)} \sum_{i=1}^{n} z_i$ denote the corresponding total contribution level.

Note that agents' messages can be translated into a step-function, assuming WLOG that $\beta_i^1 \leq \beta_i^2$,

$$f_i(x) := \begin{cases} 0 & \text{if } x < \beta_i^1 \\ \alpha_i^1 & \text{if } \beta_i^1 \leq x < \beta_i^2 \\ \max\{\alpha_i^1, \alpha_i^2\} & \text{if } \beta_i^2 \leq x \end{cases}.$$

Total contributions $T(m)$ induced by the CCM mechanism can then easily be computed by adding up all step functions and taking the highest fixed point $x = \sum f_i(x)$. This makes sure that there is no computational problem in applications when n is large.

The definition of the outcome function requires that conditional contribution offers can be enforced by the institution that implements the outcome. In some cases, for example when contributions are monetary, this can be arranged without much difficulty. In other cases this is nearly impossible, without the use of additional commitment devices. The CCM is thus best suited for environments in which it is easy to make the conditional contribution offers binding.

3.2 Nash equilibria of the CCM

We will mainly be concerned with dynamic solution concepts to predict the outcomes of the CCM in a repeated setting. However, looking at the static Nash equilibria of the stage game provides some intuition on the incentive structure of the game. The following example demonstrates what properties an outcome must have to be a Nash equilibrium outcome.

Example 1 *Consider 5 identical agents with valuation* $\theta_i = 0.6,\ \forall\ i \in I$*. The trivial Nash equilibrium is given by* $m_i = \{(0,0),(0,0)\},\ \forall\ i \in I$*. Here no agent contributes to the public good. However, there are more equilibria as e.g. when agents* 1 *and* 2 *choose message* $m_i' = \{(0,0),(1,2)\}$ *and agents* 3 *through* 5 *choose* $m_i' = \{(0,0),(0,0)\}$*. In this case the first two agents will contribute to the public good and the outcome is* $z = (1,1,0,0,0)$*. The structure of the mechanism makes this an equilibrium. Agents 3 through 5 can only change the outcome to* $(1,1,\alpha_3,0,0)$*,* $(1,1,0,\alpha_4,0)$*, or* $(1,1,0,0,\alpha_5)$ *respectively, by unilateral deviation. Neither deviation is profitable for the respective agent. The first two agents can only change the outcome to* $\underline{z}$*, which is not profitable for them either. Thus, no agent has any incentive to deviate and the above message profile is a Nash equilibrium. Yet the outcome is not Pareto optimal as it is dominated e.g. by* $z' = (1,1,1,1,1)$*.*

In the second message profile in the example, m', all agents, who end up contributing to the public good, condition their contribution exactly on the aggregate level of contributions. All other agents choose to free-ride in any case. The latter agents can alter the outcome only by unilaterally contributing themselves. Since $\theta_i < 1$, this is not profitable. Agents that currently do contribute can either unilaterally increase their contribution (if $z_i < 1$), which is not profitable, or they can lower their contribution and change the outcome to $\underline{z}$. The latter will be a strict best reply if and only if $\underline{z} \succ_i z$. 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.

Lemma 3 *z is the outcome of a Nash-equilibrium of the CCM if and only if* $z \succeq_i \underline{z},\ \forall\ i \in I$*.*

Lemma 3 predicts equilibria which are Pareto efficient as well as equilibria which may not be Pareto efficient. In the next subsection, we will present arguments why only the Pareto efficient Nash equilibria will prevail as dynamically stable outcomes of a reasonable dynamic adjustment process.

3.3 Dynamic behavior in the CCM

In this subsection, we analyze the properties of the CCM under dynamic considerations. In our dynamic model, agents play the same game repeatedly over several periods in fixed groups. We follow the literature (in particular, Cabrales and Serrano, 2011) and assume that all agents can adjust their message from one period to the next with independent and fixed probability. We further assume that agents are myopic and take the behavior of the other agents from the previous period as given. There is evidence (see e.g. Healy, 2006) that agents' behavior in repeated public good mechanisms can be well described using best response dynamics. Whether one should assume that agents respond to behavior from the previous period, to the entire history, or, as done by Healy (2006), to an average over the last few periods, may depend on the application. In our case, given that the CCM mechanism reacts in a very discontinuous fashion,[9] we find it most plausible if agents react only to the most recent information from the previous period.

Accordingly, a message m_i^{t+1} is a best response if

$$u_i(g(m_i^{t+1}, m_{-i}^t)) \geq u_i(g(m_i', m_{-i}^t)), \text{ for all } m_i' \in M_i^{CCM}, \tag{4}$$

where g represents a general outcome function and m_{-i}^t is the message profile of all agents other than agent i in period t. If there are several best responses, we assume that all of them are chosen with strictly positive probability. However, given that there is ample evidence in public good games (see e.g. Ledyard (1995)) that agents do not fully exploit their strategic advantages, for example, by contributing in a voluntary contribution mechanism, we would also allow agents to simply choose a better response, although we do not require that all better responses are chosen with strictly positive probability.[10]

Formally, a message m_i^{t+1} is a better response (better than the current message m_i^t) if

$$u_i(g(m_i^{t+1}, m_{-i}^t)) \geq u_i(g(m_i^t, m_{-i}^t)). \tag{5}$$

Definition 2 *In Better Response Dynamics (BRD) all agents can adjust their message in every period with independent and fixed probability $\gamma \in (0, 1)$. If they have the opportunity, agent i switches in period $t + 1$ to a message m_i^{t+1} that is a better response. All messages that are best responses are chosen with strictly positive probability.*

As shown by Reischmann and Oechssler (2018) for the binary case, BRD by themselves do not guarantee that seemingly reasonable messages are actually dynamically stable. The

[9] Since the message space is finite, the word "discontinuous" is not meant literally here. Instead it describes the fact that a small change in one agent's message can change the outcome from full contribution to zero contribution or the other way round.

[10] In this sense our dynamics impose strictly weaker restrictions than best response dynamics.

following example illustrates this problem.

Example 2 *Assume there are 5 identical agents all with type $\theta_i = 0.4$. Assume that all 5 agents contribute to the public good under the current message profile $m_i^t = \{(1,5)(1,5)\}, \forall i$. Thus, all agents condition their contribution on all other agents contributing as well. This might seem to be a good candidate for a dynamically stable outcome. However, for any agent i any deviation to a message m_i^{t+1} with $\beta_i < 5$ keeps the outcome constant if all other agents repeat their message $m_{-i}^{t+1} = m_{-i}^t$. Therefore, these messages are all best (and therefore better) responses and the message profile under BRD in the next period might be, e.g. $m_i^{t+1} = \{(1,0)(1,0)\}, \forall i$. In this transition the outcome did not change. However, now the best response to m^{t+1} is to not contribute and the next profile could be $m_i^{t+2} = \{(0,0),(0,0)\}$, an outcome in which no agent contributes to the public good.*

With the motivation of the example in mind, we follow Reischmann and Oechssler (2018) and extend the myopic better response model using a second property, which is called unexploitability. Unexploitability captures non-myopic strategic incentives in the CCM and, therefore, complements the myopic better response condition with forward looking aspects. The formal definition of unexploitability is the following.

Definition 3 *Given an outcome $z^t = g(m_i^t, m_{-i}^t)$, a message m_i^{t+1} is called* exploitable *if there is any $m_{-i}^{t+1} \in M_{-i}$ such that $g(m_i^{t+1}, m_{-i}^{t+1}) = z^{t+1} \prec_i z^t$ and $z_i^{t+1} > 0$. A message m_i^{t+1} is called unexploitable if it is not exploitable.*

In words, a message of agent i is unexploitable if there is no chance that after any deviation of all other agents agent i has to contribute in the next period and is worse off. Definition 3 has two particular details that deserve discussion. Note first that all possible message profiles m_{-i} of other agents are considered. One could argue, since we assume a BRD model, that we should only consider profiles of better responses of other agents at this point. However, this dynamic model is designed with the most general case in mind, in which agents have no information on the preferences of other agents. In this case agents cannot tell whether a certain message of another agent is a better response. Therefore, from a player's perspective it seems rational to account for all possible choices. Note second that we only consider outcomes z' in which agent i contributes to the public good. This has to be the case since the choice of conditions in the CCM 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 i does not contribute, there is nothing to exploit.

The existence of cooperative but unexploitable messages is a special property of the CCM. As one comparison, in the VCM only contributing nothing is unexploitable. In other mechanisms there might be situations where the set of unexploitable best responses is empty. As the following case shows, this is not the case for the CCM.

Lemma 4 *Unexploitable best responses always exist in a CCM.*

When we combine the BRD with the requirement that messages be unexploitable, we get the following.

Definition 4 *An unexploitable better response dynamic (UBRD) is a BRD dynamic with the restriction that agents can only choose unexploitable messages.*

As mentioned before, an important feature of the dynamic process for the CCM mechanism is that it does not depend on complete information. In particular, neither the mechanism designer nor other agents need to be informed about (other) agents' valuations. To see this note that neither i's better responses (5) nor the definition of unexploitability depend on the other agents' $\theta_j, j \neq i$.

Remark 1 *Unexploitable better response dynamics in a CCM do not depend on any knowledge about other agents' valuations.*

Next, we define what we mean by dynamically stable outcomes.

Definition 5 *A recurrent class of UBRD 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.*

Our main theorem characterizes the long run stable outcomes of our mechanism.

Theorem 2 *An outcome $z \in Z$ is an outcome of a recurrent class of the CCM under UBRD if and only if it is a Pareto optimal outcome and a strict* Pareto improvement over $\underline{z}$.*

Note that the assumed environment does not include the possibility for transfer payments. Thus, a Pareto optimal outcome need not necessarily maximize the sum of utilities.

Example 3 *Consider the example with 5 agents. Each agent has type $\theta_i = 0.4$. It follows from Lemma 2 and Theorem 2 that 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.*

3.4 Why two messages?

An obvious question to ask is why we need two messages for each agent in the CCM. To show that one message is not sufficient to achieve Pareto optimal outcomes, we shall define the Simple Conditional Contribution Mechanism (SCCM), which is similar to the CCM except that the message space is just given by M_i. In the following we show that the SCCM can get stuck at very inefficient outcomes even if we use the same dynamic process.

Example 4 *Assume again five agents with type $\theta_i = 0.4, \forall i$. Assume further that in period t all agents sent message $m_i^t = (0.1, 0.5)$. Obviously, the resulting outcome $z^t = (0.1, 0.1, 0.1, 0.1, 0.1)$ is not Pareto optimal. Let us find all unexploitable better responses in period $t + 1$. Consider w.l.o.g agent 1. Any message $m'_1 = (\alpha_1, \beta_1)$ with $\alpha_1 < 0.1$ and $\beta_1 > \alpha_1$ will lead to $\underline{z}$ and is thus not a better response. Any message $m'_1 = (\alpha_1, \beta_1)$ with $\alpha_1 < 0.1$ and $\beta_1 \leq \alpha_1$ will lead to $z = (\alpha_1, 0, 0, 0, 0)$ and is thus not a better response, either. Any message $m'_1 = (\alpha_1, \beta_1)$ with $\alpha_1 > 0.1$ and $\beta_1 > 0.4 + \alpha_1$ will lead to $\underline{z}$ and is thus not a better response. Any message $m'_1 = (\alpha_1, \beta_1)$ with $\alpha_1 > 0.1$ and $\beta_1 \leq 0.4+\alpha_1$ will lead to $z = (\alpha_1, 0.1, 0.1, 0.1, 0.1)$ and is thus not a better response, either. This leaves only messages with $\alpha_1 = 0.1$. However of those messages the ones with $\beta_1 > 0.5$ lead to $\underline{z}$ and are not a better response and the ones with $\beta_1 < 0.5$ are exploitable. For example. $\beta_1 = 0.3$ 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.*

Agents can in this way get stuck on Pareto improvements over $\underline{z}$ which are far from Pareto optimal. Any deviation aiming to make further Pareto improvements possible would make the deviating agent worse off in the next period.

This problem can be solved by letting agents announce more than one tuple of the form (α_i, β_i).[11] 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.

4 Experimental design

The stage game is a standard linear public good game with five participants, which is repeated for 60 periods with fixed matchings. In every period each subject is endowed with 10 units. Contributions to the public good are possible only in integer values, participants can contribute anything from nothing to their whole endowment. Participants know that the public good for them has a value of $\theta_i \in \{0, 0.4\}$ in all periods. This implies that in each period, subjects are faced with the following payoff function

$$u_i(z) = 10 - z_i + \theta_i \sum_{j=1}^{5} z_j. \tag{6}$$

Table 1: Summary of treatments

	Complete info	Incomplete info
Conditional contribution mechanism	CCM (40)	CCMII (55)
Voluntary contribution mechanism	VCM (35)	VCMII (55)
VCM with communication	VCMC (40)	

Note: Number of subjects in parentheses.

We consider five treatments, which vary depending on the mechanism used and the information subjects have about the θ_j of other group members (see Table 1 for a summary).

[11] Allowing agents to send more than two messages would be possible but would not add anything in the current framework.

In the complete information treatments, all subjects know that everyone in their group has a θ_i of 0.4. In the incomplete information treatments, all subjects know that each group member can independently have a θ_i of 0 with a probability of 20% and a θ_i of 0.4 with a probability of 80%. Each participant knows their own θ_i which is kept constant for the entire experiment but never finds out the realized θ_i of the other group members.

In the VCM treatments, participants are simply asked how much they would like to contribute and consequently those contributions are implemented. In the CCM treatments, each participant has to send two messages of the form "I will contribute α if total contributions are at least β." The computer then determines which is the highest contribution level possible given the combination of conditional contributions within each group and implements it. For the incomplete information treatments the decisions and procedures are identical with the only difference being that participants have a note on their screen informing them of their own θ_i, while they do not know the θ_i of others in their group.

Finally, the VCM with communication (VCMC) is just like the VCM treatment except that in each round subjects can send two conditional messages just as in treatment CCM. However, these messages are non-binding cheap-talk.

After each round, the outcomes from the last round are listed in the bottom of the screen. In the VCM treatment, participants can see how much they and every other member of their group contributed to the public good. In the CCM and the VCMC treatments, they additionally see their own and all other group members' two messages.

To help participants understand how their payoff depends on their own and the others' decisions, a calculator is provided on the decision screen. In the VCM and the VCMC treatments, the calculator requires subjects to enter a (hypothetical) contribution level for each member of their group. In the CCM treatments, participants have to input two conditional contributions for each group member. The calculator then informs them how much they would contribute and what their payoff would be for the values they had entered.

From the second round onwards, the screens are filled in as a default with the numbers each participant had submitted in the previous round. This means that the outcomes of the previous round are also loaded in the calculator boxes. We chose this 'default' norm to assist participants in utilizing the calculator for best responding to their group members actions. Specifically for the CCM treatments, utilizing the calculator would be quite cumbersome as participants would need to input 10 values before they could obtain a calculated possible profit (two conditional contributions for each of the five members of the group).

Obviously, both the provision of the calculator and the default values are "nudges" to induce myopic best response behavior. Note however that in any real application of the CCM, the mechanism designer can and probably should also provide those nudges to participants, for example in form of a webpage or an app.[12]

After the completion of the public good game, the participants are asked to complete a short demographic questionnaire.

Implementation

We implemented 12 sessions with a total of 225 participants, of which 40 participants were in the CCM treatment (8 groups of five), 35 participants were in the VCM treatment (7 groups of five), and 40 were in the VCMC treatment (8 groups of five).[13] For the incomplete information treatments we were aiming for 12 groups since some groups without any $\theta_i = 0$-type subjects were to be expected. Due to no-shows we ended up with 11 groups in 3 sessions for both treatments (55 participants, each). All sessions were conducted at the AWI experimental lab at the University of Heidelberg using z-Tree (Fischbacher, 2007) and hroot (Bock et al., 2014) for recruitment from a student subject pool. At the beginning of each session, subjects were randomly placed in the different cubicles of the lab where they were asked to read the instructions. Once they were done with the instructions, a short comprehension quiz was administered. Once all subjects solved the quiz correctly and there were no more questions the experiment started. The sessions were conducted in German. English translations of the instructions and test questions can be found in the appendix.

The participants were informed that one randomly chosen round out of the 60 played would be chosen to determine their monetary payoff. The paying round was determined by rolling 10-sided die and a regular die by one of the participants in each session. The experimental units each corresponded to €1. The experiments lasted around 75 minutes and subjects earned on average €16.51.

[12] And participants should be happy to use such a webpage or app once they understand that it helps to achieve an efficient outcome.

[13] We have one group less in the VCM treatment due to a high number of 'no-shows' in one of our sessions. Since the VCM treatment consists of a common mechanism that has been studied in numerous experiments where qualitatively similar results are found we felt it was not appropriate to administer one session with only 1 group.

5 Experimental results

We start by considering the complete information treatments, then look at the incomplete information treatments, and finally consider the strategic behavior in all treatments.

5.1 Complete information

Of prime importance are the average contributions across treatments. Figure 1 shows the average contributions trend for the three complete information treatments. For the VCM we confirm the standard result of average contributions starting roughly at 50% but declining over time to less than 20% (comparable to Herrmann et al., 2008). The CCM, however, shows an increasing path with very high contribution rates after the first few periods. Table 2 lists the average contributions for all periods, the second half of the experiment and for periods 10-55 (excluding the end game effect). In all cases, contributions in the CCM are almost twice as high as in the VCM. According to MWU-tests (taking each group as one independent observation), we find significant differences for all periods ($p = 0.011$), the second half ($p = 0.015$) and periods 10-55 ($p = 0.018$).[14]

Table 2: Average contributions in complete information treatments

	CCM	VCM	VCMC
average contributions all periods	8.61	4.66	7.35
average contributions periods 30-60	8.77	4.40	7.23
average contributions periods 10-55	8.90	4.61	7.63

Average contributions in treatment VCMC are between those in VCM and in CCM. They are not significantly different to either one according to MWU-tests, although they are close to significance with respect to VCM ($p = .11$). Thus it seems that part of the success of CCM may be due to communication that is inherent in the mechanism. But this is not the whole story. Looking at individual groups (see Figure 2) one can see that 3 of the 8 groups in treatment VCMC completely collapse to no contribution while the remaining groups cooperate almost perfectly. This is different for the CCM. In no group does cooperation completely collapse but there are some volatile groups where contributions drop to zero in one round only to rebound back to full contribution in the next round. This

[14] All tests reported in this paper are two–sided tests.

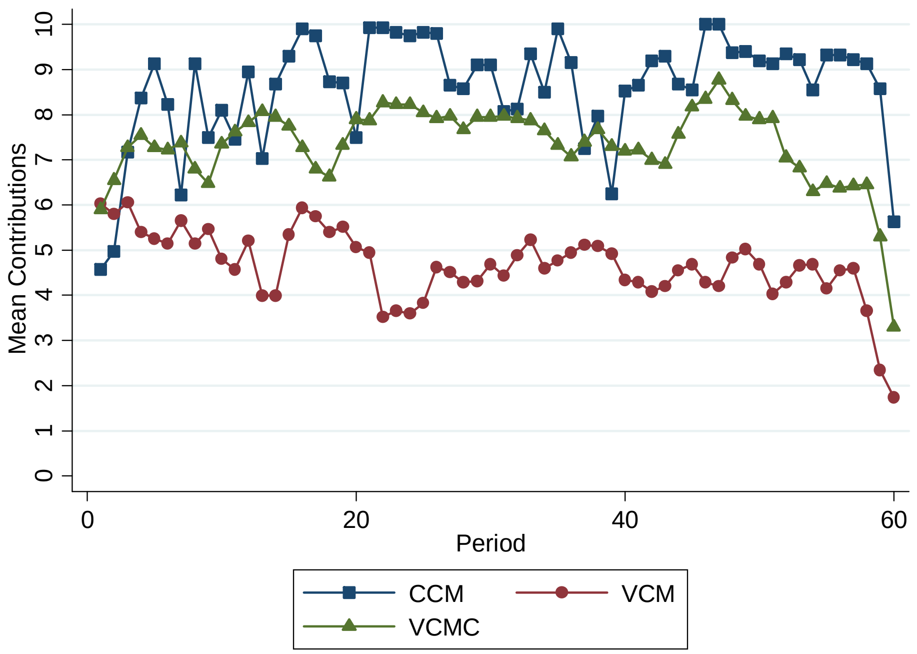

Figure 1: Time path of average contributions over all 60 periods, complete information treatments.

is typical for a situation when subjects condition on the current level of contribution and yet some subjects try to deviate. They quickly learn that such deviations are not profitable.

When we look at the end–game, in the CCM treatment half of the groups remain at maximal contribution (10) for each subject. This seems quite remarkable but should not be surprising. It is not rational for agents to deviate if enough subjects condition their contribution on the current contribution level. That is, if they were to play an unexploitable best response. What actually is remarkable is that there are 3 groups in which at least one subject did not understand this causing total contributions to drop to zero. We would expect that such mistakes become rarer if subjects have more experience by playing the repeated game repeatedly. On the other hand, in treatments VCM and VCMC there is an end–game effect in every group (in the groups where contributions had not already crashed earlier).

Next, we can check the prediction of Theorem 2. Accordingly, the CCM should converge to Pareto optimal outcomes. Table 3 lists the percentage of Pareto optimal outcomes for the three treatments. As predicted, the share of Pareto optimal outcomes is quite high in CCM at 78%. A MWU-test (at the group level) shows that the corresponding share is significantly lower at 24.3% for VCM ($p = 0.012$). Again, VCMC is intermediate between the other two treatments and is weakly significantly higher than VCM ($p = 0.080$) and not significantly different from CCM ($p = 0.293$).

Table 3: Percentage of Pareto optimal outcomes

	% of Pareto optimal outcomes
CCM	77.9%
	$>^{**}$
VCM	24.3%
	$<^{*}$
VCMC	59.0%

** sign. at 5%, MWU-test, * sign. at 10%.

5.2 Incomplete information

Figure 3 shows the time path of average contributions of the beneficiaries of the public good (i.e. subjects with $\theta_i = 0.4$) for the two incomplete information treatments. We

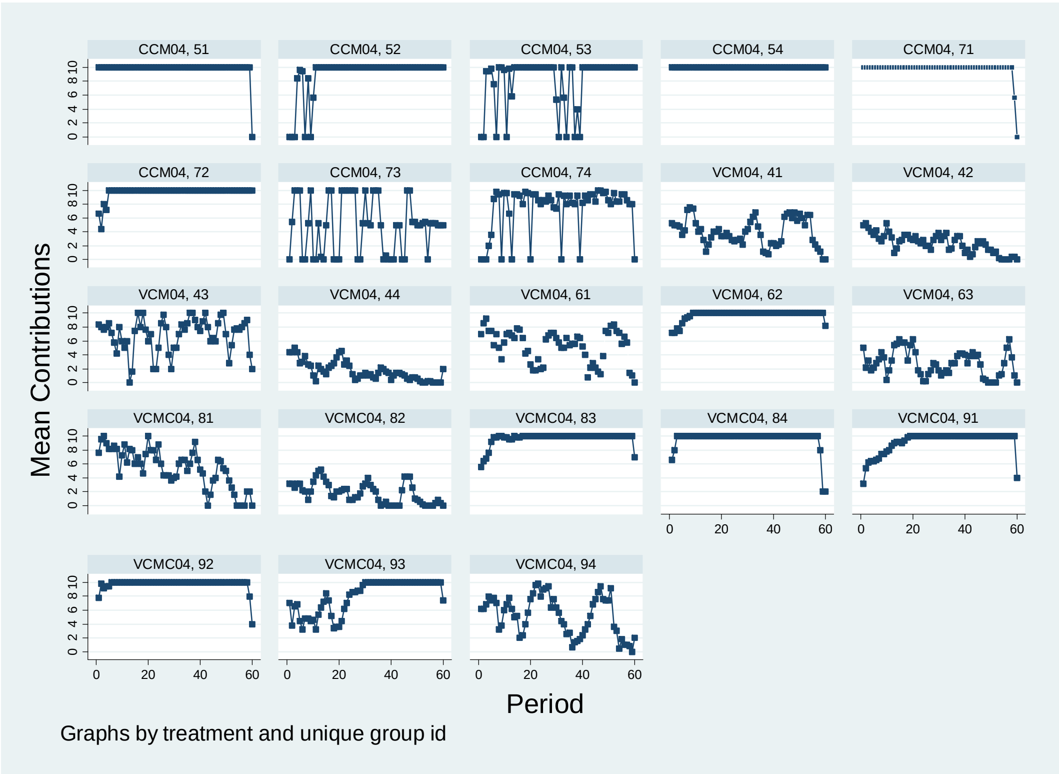

Figure 2: Mean contributions of individual groups over 60 periods, complete information treatments.

exclude the $\theta_i = 0$ types since they have a dominant strategy to contribute zero and do so in 94% of cases. The CCMII again shows an increasing path with fairly high contribution rates after the first few periods. Table 4 lists the average contributions for all periods, the second half of the experiment and for periods 10-55. As was to be expected, contributions in CCMII are lower than under complete information but they are not very much lower and there is no downward trend and no endgame effect.[15]

What is surprising, though, is that the VCM seems to do just as well as the CCM under incomplete information. According to MWU-tests (taking each group as one independent observation), we find no significant difference for all periods ($p = 0.87$), the second half ($p = 0.82$), and periods 10-55 ($p = 0.97$). In contrast to treatment CCMII, there is a strong endgame effect in VCMII.

The VCM seems to do better under incomplete information (counting only beneficiaries) than under complete information, which is a bit surprising (although the difference between VCM and VCMII is not significant for any of the considered time periods, $p > 0.30$). We are currently not aware of any prior experimental evidence showing this. So this is an interesting finding by itself and deserves more attention in future research.

The share of Pareto optimal outcomes is also fairly similar in the two incomplete information treatments at 49.9% for CCMII and 47.1% in VCMII, which is not significantly different (MWU-test, $p = 0.82$).

Table 4: Average contributions of beneficiaries in the incomplete information treatments

	CCMII	VCMII
average contributions all periods	6.29	6.39
average contributions periods 30-60	6.73	6.37
average contributions periods 10-55	6.69	6.62

Looking again at individual groups we can see some interesting patterns. Figure 4 shows the mean contributions at the group level including the type-0 subjects. Some groups almost perfectly coordinate, e.g. groups 101, 103, and 104 in CCMII. The average contribution is 8, which comes about because all beneficiaries contribute fully and the one type-0 subject not at all. Surprisingly, some of these cases exist even for the VCM, e.g. groups 131 and 152.

[15] Contributions are significantly different between CCM and CCMII according to MWU-tests with $p = 0.026$ for all periods, $p = 0.096$ for periods 30-60, and $p = 0.054$ for periods 10-55.

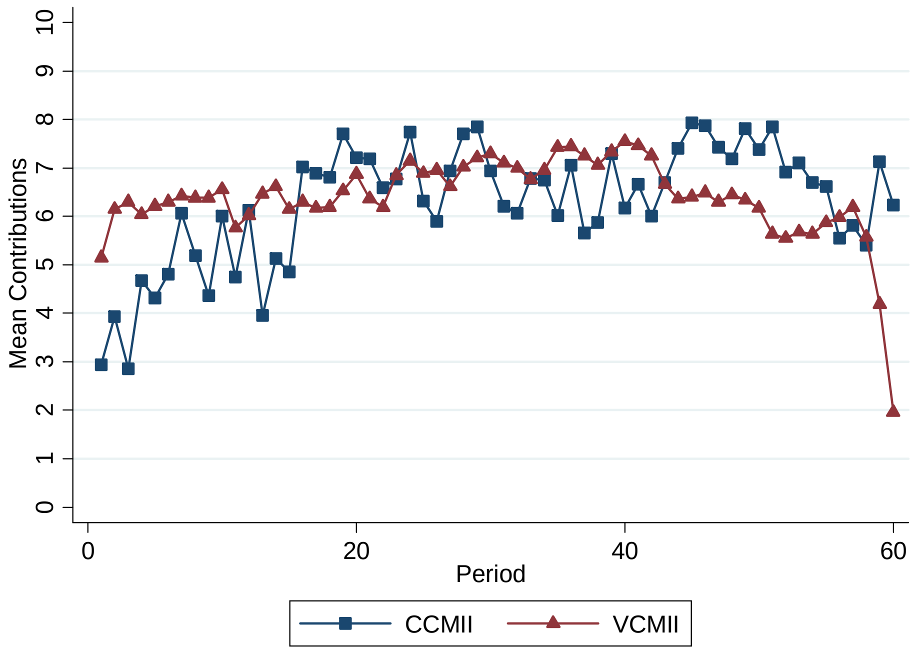

Figure 3: Time path of average contributions of beneficiaries over all 60 periods, incomplete information treatments.

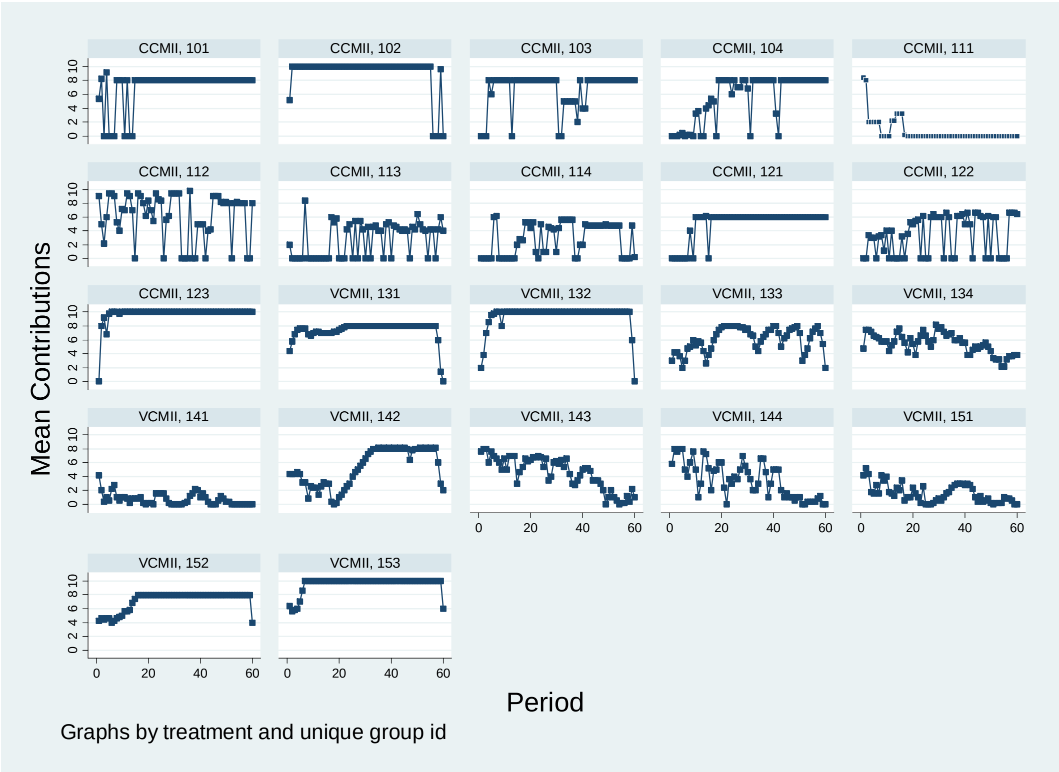

Figure 4: Mean contributions of individual groups over 60 periods (type-0 subjects included), incomplete information treatments

There are also two interesting groups that point out potential weaknesses of the CCM which might have to be addressed in future work. Group CCMII,111 is a group with 2 type-0 subjects and group contributions quickly converge to 0. The corresponding group in the VCM, group 141, shares almost the same fate. Yet, the reason why the CCM fails in this group is more interesting: Both type-0 subjects start by offering to contribute positive amounts and actually have to contribute 7 and 5, respectively, in the first period. Very likely this made the other subjects in their group believe that there are no type-0 subjects in the group and can account for the failure to coordinate.

Group CCMII,121 is also interesting but for a different reason. This group managed to coordinate on a Pareto efficient mean contribution of 6, which comes about when 3 subjects contribute fully and 2 not at all. However, this group had only one type-0 subject. It turns out that one subject perfectly mimicked a type-0 guy by never offering any contributions from period 1 to 60. This can happen even if the CCM works perfectly since the outcome is Pareto optimal as predicted.

5.3 Strategic behavior

To better understand the performance of the CCM it is useful to check which parts of the mechanisms and the dynamic process are supported by the data and which not. Table 5 lists a number of measures related to the strategic behavior of subjects in the CCM treatments. First, the assumption in the theory section that subjects play myopic better responses seems to be strongly supported by the data. In fact, more than 93% of choices are better responses to the message profile of the previous period in both CCM treatments. As discussed above, subjects do not always fully exploit their strategic position as only slightly more than 2/3 of decisions in CCM and 57% of decisions in CCMII are *best* responses.

Table 5: Strategic behavior of subjects in the CCM

	complete info	incomplete info
% of better responses	95.5%	93.4%
% of best responses	67.0%	57.0%
% of unexploitable messages	13.8%	44.8%
% high message unexploitable	63.8%	72.5%
% high message chosen by CCM	83.1%	61.7%
% of weakly unexploitable messages	97.9%	87.1%

The share of unexploitable messages is surprisingly low at 13.8% in CCM. Note however, that for a message to be counted as unexploitable both parts of the message must satisfy the condition for unexploitability. In many cases subjects will (correctly) assume that only one part of their message is relevant in the current situation and in 83.1% the high message (i.e. the message that promises the higher contribution) is the one chosen by the CCM mechanisms. As Table 5 shows, 63.8% of high messages are unexploitable. It seems that subjects make sure their – more likely relevant – high message is unexploitable and possibly give less attention to the low message. In CCMII the share of unexploitable messages is higher, which however, might be due to the higher share of no contribution messages (which are always unexploitable).

Finally, we find that subjects overwhelmingly make sure that their messages do not make them worse off than not contributing at all. This condition is satisfied if $\alpha_i \leq \theta_i\beta_i$ for both messages. We call such messages weakly unexploitable because the condition is implied by our unexploitability condition for period 2 through 60 once it is satisfied in the first period. As Table 5 shows, 97.9% of all messages are weakly unexploitable in CCM and 87.1% in CCMII. This shows that most subjects understand that they can use the CCM to make sure that they are not the suckers in their group: by using (weakly) unexploitable messages they would never contribute and be the victim of free-riders, which certainly is one of the dangers in the VCM.

Table 6 lists the share of outcomes where subjects receive a profit of less than 10, which is the no contribution profit. Under the CCM this almost never happens as only 0.13% of profits are below 10 (and only slightly higher in CCMII). In the VCM treatment, almost 7% of profits are below 10 (and more than 8% in VCMII). Even though this is not a very high share, it may discourage contributions. Communication alleviates the sucker problem somewhat but cannot solve it entirely as 3.92% of profits are below 10.

Table 6: The risk of becoming a sucker

	share of outcomes with profits < 10
CCM	0.13%
VCM	6.95%
VCMC	3.92%
CCMII	1.15%
VCMII	8.10%

6 Summary and discussion

In this paper we propose a new and simple mechanism that helps achieve dynamically stable Pareto efficient outcomes in a repeated public good environment. With many of the existing mechanisms that have been proposed in the literature a common limitation is typically that they are focusing on non-repeated cases of public good games. Situations that apply to the repeated case are widespread which makes it clearly necessary to study mechanisms that can achieve efficient outcomes with repetition.

We require that our mechanism satisfies three criteria: individual rationality, balanced budget, and that the mechanism designer and agents need *not* know the preferences of (other) agents. The Conditional Contribution Mechanism (CCM) asks agents to send two messages about their own level of contribution with a condition on total contributions. The mechanism then selects the contribution profile that maximizes the total contributions from all agents. Our main theorem says that if agents play unexploitable better responses all dynamically stable outcomes of the CCM are Pareto optimal.

We experimentally test the mechanism and find that under complete information almost 80% of outcomes are Pareto optimal. By comparing the CCM to a VCM treatment we find clear evidence that our mechanism outperforms the VCM with significantly higher contribution rates. Of course, given the existing literature on how communication helps foster high contribution rates, a potential concern might be that the CCM does better than the VCM simply due to the 'communication' aspect of it. We investigate this possibility in a cheap talk VCM treatment where participants offer non-binding messages of how much they would contribute for a given level of total contribution. We find that the VCMC treatment does better than the VCM in terms of contribution rates, but importantly it does not manage to do as good as the CCM. The CCM contribution rates are still slightly higher than the VCMC treatment as is the proportion of Pareto efficient outcomes.

One of the important characteristics of the CCM is that neither the mechanism designer nor other agents need to know the preferences of an agent. We also test how the CCM performs in an incomplete information situation experimentally and find that high contribution rates are sustained on average, but not as high as in the complete information treatments. Surprisingly we find that the VCM treatment with incomplete information also performs quite well in terms of contribution rates. This is an interesting finding by itself and deserves more attention in future research.

Overall, our theoretical predictions for the CCM appear to be corroborated in our experimental analysis. This is promising as the CCM is a very simple mechanism to implement and can be applied to various situations to help sustain long-term contributions. Of course, one caveat to the mechanism is that it assumes that the messages are binding and that this is enforceable. This assumption may not always be appropriate depending on the situation one is interested in. For example, when looking at international treaties it would be heroic to assume that any commitments involved in such treaties are enforceable even if under 'binding' agreements. Given this limitation it would seem reasonable to argue that the CCM is not exactly appropriate for such cases, but instead more appropriate for smaller scale situations. In cases where sports clubs or schools are trying to collect contributions for specific projects potentially the CCM could act as a helpful mechanism to assist socially optimal levels of contributions. Furthermore, in situations like with Wikipedia fund-raising, one could imagine a possibility where contributions come with conditions attached and are then triggered automatically thus making messages both binding and enforceable.

Computationally it should be easy to use the CCM even in very large populations. However, the feedback given to participants needs to be adjusted in this case since it is clearly impractical to provide participants with a list of all messages. To solve this problem one could provide participants with an app that is programmed similarly as the calculator used in the experiment except that it already takes the messages of other participants as given. Future research should test the CCM in the field.

Appendix A: Proofs

Proof of Lemma 1. Suppose $\sum_j \theta_j > 1$. Consider a contribution vector z such that $z_i = \theta_i$, $\forall i$. Then

$$u_i(z) = 1 - \theta_i + \theta_i \sum_j \theta_j \geq 1,$$

with the inequality being strict for all $\theta_i > 0$. Hence, z is a *strict** Pareto improvement over $\underline{z}$.

Conversely, suppose there exists a z which is a *strict** Pareto improvement over $\underline{z}$. Hence

$$1 - z_i + \theta_i \sum_j z_j \geq 1 \Leftrightarrow \theta_i \geq \frac{z_i}{\sum_j z_j}$$

with the inequality being strict for all $\theta_i > 0$. Summing over all i yields

$$\sum_i \theta_i > \frac{\sum_i z_i}{\sum_j z_j} = 1$$

as desired. $\square$

Proof of Lemma 2. ("only if") Suppose that $\sum_{i \in C} \theta_i > 1$. We will show that z is not Pareto optimal by constructing some z' which is a Pareto improvement over z. Let $\Delta_i := z_i' - z_i \geq 0, \forall i \in C$ and $\Delta_i = 0, \forall i \notin C$. We now specify that

$$\Delta_i = \theta_i M + \varepsilon, \forall i \in C$$

for some $\varepsilon > 0$ and $M > 0$. We choose an arbitrary M and ε small enough such that $z_i + \Delta_i < 1, \forall i \in C$. Next we show that the z' thus constructed is strictly better than z for all agents with $\theta_i > 0$ and weakly better for all $\theta_i = 0$.

$$\begin{aligned} u_i(z') &= 1 - z_i - \Delta_i + \theta_i \left(\sum_{j \in I} z_j + \sum_{j \in C} \Delta_j \right) \\ &= 1 - z_i - \theta_i M - \varepsilon + \theta_i \sum_{j \in I} z_j + \theta_i \sum_{j \in C} (\theta_i M + \varepsilon) > 1 - z_i + \theta_i \sum_{j \in I} z_j = u_i(z) \end{aligned}$$

$\Leftrightarrow$

$$-\theta_i M - \varepsilon + \theta_i \sum_{j \in C} \theta_i M + \theta_i |C| \varepsilon > 0, \tag{7}$$

where $|C|$ is the cardinality of C. Rearranging (7) yields

$$M\theta_i \left(\sum_{j \in C} \theta_i - 1 \right) + (\theta_i |C| - 1)\varepsilon > 0,$$

which is satisfied by assumption for all $\theta_i > 0$ and ε small enough.

("if") Suppose z is not Pareto optimal. We will show that $\sum_{i \in C} \theta_i > 1$. If z is not Pareto optimal, then some z' exists s.t.

$$\Delta_i \leq \theta_i \sum_{j \in C} \Delta_j, \forall i \in C,$$

with strict inequality for at least one i. Thus,

$$\sum_{j\in C}\Delta_j < \sum_{j\in C}\theta_j \sum_{j\in C}\Delta_j$$

and hence $\sum_{i\in C}\theta_i > 1$. $\square$

Proof of Lemma 3. 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 $\bar{\beta} := \sum_{i=1}^{n} z_i$. Then $m_i = \{(z_i, \bar{\beta}), (0, 0)\}, \forall i$, is a Nash-equilibrium of the mechanism with outcome z. To see this, note first that it is immaterial whether an agent deviates by changing his first or his second message since given the message profiles of the others, only one message is relevant. Thus, without loss of generality, we can denote any deviation by $\{(\alpha_i, \beta_i), (\alpha_i, \beta_i)\}$. There are four ways in which agent i can deviate. He can increase or decrease his proposed contribution. And he can increase or decrease his condition.

If he decreases his proposed contribution, $\alpha_i < z_i$, other agents' conditions are not satisfied and $\underline{z}$ will be chosen by the mechanism, which is not a profitable deviation, no matter what condition β_i he chooses.

If he increases his proposed contribution, $\alpha_i > z_i$, other agents' contributions will be unaffected, making agent i worse off if his condition β_i is satisfied. If his new condition β_i is not 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 $m_i'' = \{(0, 0), (0, 0)\}$. 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 Lemma 4. Let $br_i(m_{-i}^{t-1})$ denote a contribution level that is part of some best response of i to the previous period's message profile m_{-i}^{t-1}, which always exists since the set of contributions is finite. Let $T^{br}(m_{-i}^{t-1})$ denote the total contribution that would be induced by the CCM given m_{-i}^{t-1} and given agent i chooses to contribute $br_i(m_{-i}^{t-1})$. Define condition $\beta_i = T^{br}(m_{-i}^{t-1})$ if $T^{br}(m_{-i}^{t-1}) > 0$. If $T^{br}(m_{-i}^{t-1}) = 0$, let β_i be any condition that makes $(br_i(m_{-i}^{t-1}), \beta_i)$ unexploitable, i.e.

$$1 - br_i(m_{-i}^{t-1}) + \theta_i \beta_i \geq 1 - z_i^{t-1} + \theta_i \sum_{j \in I} z_j^{t-1} = u_i(z^{t-1}).$$

Thus, message $\left\{ \left(br_i(m_{-i}^{t-1}), \beta_i \right), (0,0) \right\}$ is an unexploitable best response to m_{-i}^{t-1}. Note that $(0,0)$ is always unexploitable and given that the first part of the message is a best response, the whole message is too. $\square$

Proof of Theorem 2. We prove this theorem in two steps. In step 1 we prove that outcomes that are Pareto optimal and *strict** Pareto improvements over $\underline{z}$ are indeed outcomes of recurrent classes of UBRD. And in step 2 we prove that from any other outcome the dynamics reach one of those recurrent classes with strictly positive probability.

Step1: Let $z \in Z$ be a Pareto optimal outcome, which is a *strict** Pareto improvement over $\underline{z}$. Define $\bar{\beta} := \sum_{i=1}^{n} z_i$. Then $\{(z_i, \bar{\beta}), (z_i, \bar{\beta})\}, \forall i$, 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.

Step 2: Assume now that the current message profile yields an outcome that is not Pareto optimal. We need to show that from there the dynamics eventually reach one of the recurrent classes. We shall do so by constructing a path that has strictly positive probability (although many other paths would work as well).

Step 2A: Let $br_i(m_{-i}^{t-1})$ and $T^{br}(m_{-i}^{t-1})$ be defined as in the proof of Lemma 4. Furthermore, as in the proof of Lemma 4 assume that agents in step 2A always choose $(0,0)$ as their second message.

Suppose in period t only one agent can adjust his strategy and for this agent there exists $br_i(m_{-i}^{t-1}) < \alpha_i^{t-1}$ for his first message. Agent i then switches to message $\left\{\left(br_i(m_{-i}^{t-1}), T^{br}(m_{-i}^{t-1})\right), (0,0)\right\}$ with positive probability. If there is no such agent, move to Step 2B. In period $t+1$, another agent j with $br_j(m_{-j}^t) < \alpha_j^t$ is selected and switches to $\left\{\left(br_j(m_{-j}^t), T^{br}(m_{-j}^t)\right), (0,0)\right\}$. Step 2A is repeated until there is no further agent who wants to decrease his contribution. Since the state space is finite, this happens in a finite number of steps. If the new message profile yields a Pareto optimal outcome, move to Step 2C. If not move to Step 2B.

Step 2B: The current message profile m^t yields an outcome z^t that is not Pareto optimal and $br_i(m_{-i}^t) \geq \alpha_i^t$, for all agents and all best replies. With positive probability all agents i for whom $br_i(m_{-i}^t) = \alpha_i^t$ can adjust their message in the next period. Let them all switch to $\left\{(br_i(m_{-i}^t), \sum_j z_j^t), (0,0)\right\}$. Note that in the new profile m^{t+1}, $br_i(m_{-i}^t) \leq \sum_j z_j^t$, for all i, because if there were an agent with $br_i(m_{-i}^t) > \sum_j z_j^t$, he would not have contributed and would therefore have had an alternative best response of $(0,0)$ contradicting the fact that $br_i(m_{-i}^t) \geq \alpha_i^t, \forall i$. We claim that m^{t+1} must be a Nash equilibrium. By construction, no agent wants to lower his contribution. But also no agent has a best response that would entail increasing his contribution since he would not trigger anyone else's condition to be satisfied.

If the outcome z^{t+1} is Pareto optimal, move to Step 2C. If not, then there exist Pareto improvements over z^{t+1}, some of which are Pareto optimal. Let z' denote some such Pareto optimal outcome with $z_i' \geq z_i^{t+1}, \forall i$.

Then for any agent i the message

$$m_i^{t+2} = \left\{ \left(z_i^{t+1}, \sum_j z_j^{t+1} \right), \left(z_i', \sum_j z_j' \right) \right\}$$

is an unexploitable best response to m_{-i}^t.

$\left(z_i^{t+1}, \sum_j z_j^{t+1}\right)$ is a best response because m^{t+1} is a Nash equilibrium. $\left(z_i', \sum_j z_j'\right)$ is always a best response since for unchanged m_{-i}^{t+1} this message cannot be satisfied since $\sum_j z_j' > \sum_j z_j^{t+1}$.

Message m_i^{t+2} is unexploitable because if i contributes with message m_i^{t+2}, he is weakly better off than in z^{t+1}. If the CCM implements the second part of the message, we know that $u_i(z'_i, z'_{-i}) \geq u_i(z^{t+1})$ by the fact that z' is a Pareto improvement over z^{t+1}. If all agents choose message m_i^{t+2}, the new outcome will be z'.

Step 2C: If z' is a *strict** Pareto improvement over $\underline{z}$,

$$m_i^{t+3} = \left\{ \left(z'_i, \sum_j z'_j \right), \ \left(z'_i, \sum_j z'_j \right) \right\}$$

is an unexploitable best response for all agents. Thus from any not Pareto optimal outcome a Pareto optimal outcome is reached with strictly positive probability.

If z' is not a *strict** Pareto improvement over $\underline{z}$, then there exists some agent i who is at least as well off in $\underline{z}$ as in z'. For this agent the message $\{(0,0),(0,0)\}$ in an unexploitable best response, while all other agents can choose m_i^{t+3}. In the next period, $\{(0,0),(0,0)\}$ is an unexploitable best response for all agents. And from there any Pareto optimal z that is a *strict** Pareto improvement over $\underline{z}$ can be reached by letting all agents choose message $\left\{(0,0), \left(z_i, \sum_j z_j\right)\right\}$. □

Appendix B: Instructions CCM

[These are translations of the original German instructions]

Welcome to our experiment! Please read these instructions carefully. Do not talk to your neighbor from now on. Please turn off your mobile phone and leave it turned off until the end of the experiment. If you have any questions, raise your hand. We will come to you. All participants have the same instructions.

In the experiment you will be divided in **groups of 5**. The experiment will last for **60 periods**. In each period you will interact with the same 4 other participants. The experiment runs completely anonymously. No participant will be informed about who he has interacted with or which payoff another participant received.

The currency for the experiment will be **Taler**. In each period you will start with 10 Taler and you can decide which fraction of those 10 Taler you want to invest into a common project. For each Taler one participant of your group invested into the common project, each member of the group will be paid **0.4 Taler**. You will keep the rest of the 10 Taler that you chose not to invest into the common project.

Example 1 *You invest 4 Taler into the project and additionally 2 other participants each invested 8 Taler, thus in total 20 (*$= 4+8+8$*) Taler are invested. You will receive for your investment and the investments of the two other participants in total* $20 \times 0.4 = 8$ *Taler. You keep the 6 Taler you did not invest. So you earn in total* $8 + 6 = 14$ *Taler in this period.*

Example 2 *All participants invest their 10 Taler into the common project. So you earn* $50 \times 0.4 = 20$ *Taler in this period.*

Example 3 *All participants invest 0 Taler into the common project. So you simply keep the 10 Taler you received in the beginning.*

Example 4 *All the other participants invest 10 Taler into the common project, you invest nothing. So you earn* $40 \times 0.4 + 10 = 26$ *Taler in this period.*

Example 5 *All the other participants invest nothing into the common project, you invest 10 Taler. So you earn* $10 \times 0.4 = 4$ *Taler in this period.*

Each participant can condition his contribution to the common project on how much all other participants contribute. All conditions have the following form:

"I will contribute X Taler, if in total at least Y Taler are contributed (by all members of the group including yourself)"

Example 6 *Suppose you choose the "I will contribute 5 Taler, if in total at least 20 Taler are contributed (by all members of the group including yourself)" Suppose the other participants contribute 18 Taler in total. Together with your 5 Taler it would make* $18 + 5 = 23$ *Taler. Your condition would be fulfilled and you would contribute 5 Taler.*

In fact, you should formulate **two different conditions of this kind**. For example, you could signal the other participants that you would be willing to contribute more, if all the others do so too. E.g. we could append the condition from example 6 by a higher condition:

First condition: **"I will contribute 5 Taler, if in total at least 20 Taler are contributed (by all members of the group including yourself)"**

Second condition: **"I will contribute 10 Taler, if in total at least 50 Taler are contributed (by all members of the group including yourself)"**

The computer then chooses the condition for each participant, such that the **highest total amount** for the project is reached. But the conditions make sure that you never pay more than you indicated.

For example, if you choose the two conditions from above, you will contribute 10 Taler if also all other participants contribute the maximal amount of 10 Taler, such that in total 50 Taler are contributed. If the group contributes in total an amount between 20 and 49 Taler you will contribute 5 Taler. And if the others in the group contribute less than 15 Taler such that even with your 5 Taler the condition of 20 Taler will not be reached, you will contribute nothing.

Payoff

The experiment lasts for 60 periods. A questionnaire will follow after the 60 periods. At the end of the experiment **one of the 60 periods will be randomly chosen** and the earnings of the chosen period will be paid with an exchange factor of 1 Taler = 1 €. The payment will be private and in cash. For example, if your earnings from the payment relevant round are 15 Taler, you will receive 15 €.

Structure of the screen

You have a printed example of the structure of the screen of the program you will make your decisions in each period. The screen is divided into three blocks.

The upper left block contains a calculator. Here you can test contributions and conditions for yourself and the other four participants (by default, the computer enters the numbers group members chose in the previous period. You can change them, of course). As soon as you have chosen contributions and conditions for all participants you can press the button “Calculate payoff”. Then the computer calculates the payoff you will receive in this case and the amount you would invest.

In the upper right block you enter your contributions and conditions that will be relevant for your later payoff. Again, by default, the computer enters the numbers group members chose in the previous period. If you want to change them, do it now. Beneath that, there is a red button. If you press it you submit your decision and leave the screen. Only when **all** participants have pressed the red button the experiment will continue. A time display on the right gives you a hint until when you should have made a decision. If the time runs out this has no consequences.

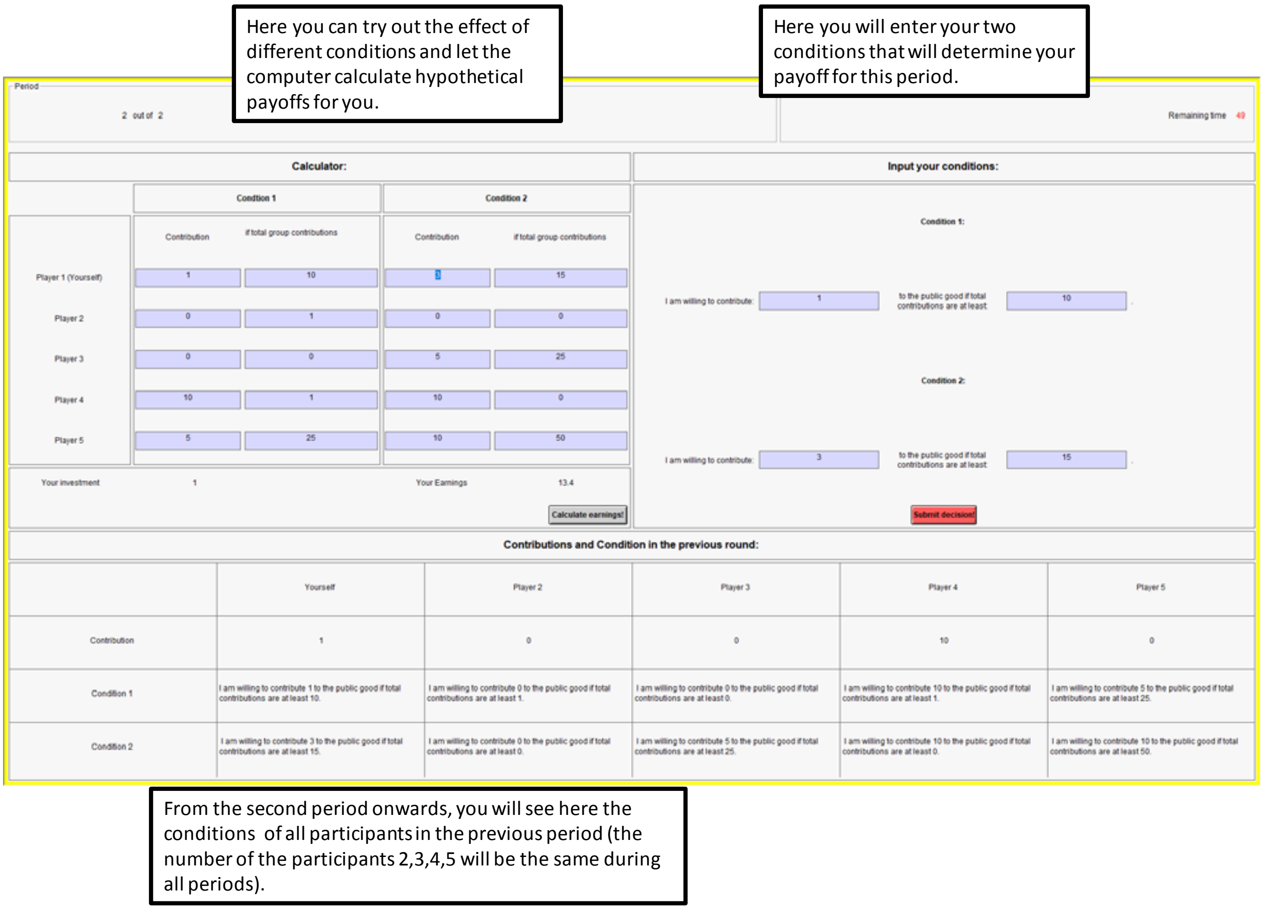

Figure 5: Screenshot CCM

In the bottom block you will see from the second period onwards which contributions and conditions all participants chose in the previous round. In the first period this block will be empty in the program.

Appendix C: Instructions VCM and VCMC

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

In the experiment you will be divided in **groups of 5**. The experiment will last for **60 periods**. You will be grouped with the same four participants in all periods. In each period you will interact with the same 4 other participants. The experiment runs completely anonymously. No participant will be informed about who he has interacted with or which payoff another participant received.

The currency for the experiment will be **Taler**. In each period you will start with 10 Taler and you can decide which fraction of those 10 Taler you want to invest into a common project. For each Taler one participant of your group invested into the common project, each member of the group will be paid **0.4 Taler**. You will keep the rest of the 10 Taler that you chose not to invest into the common project.

[`This part only in the incomplete information treatments:`

With a 20 % chance, you are type 1 and do not benefit from the collaborative project. In that case, you will only be paid the Talers you keep. All Talers that a participant in your group invests in the joint project will not affect your payout in this case.

With a 80 % chance, you are type 2 and benefit from the joint project. For every Taler that a participant in your group invests in the joint project, you (and every other type 2 member of the group) will receive 0.4 Talers

Example 1 *You invest 4 Taler into the project and additionally 2 other participants each invested 8 Taler, thus in total 20 (*$= 4+8+8$*) Taler are invested. You will receive for your investment and the investments of the two other participants in total* $20 \times 0.4 = 8$ *Taler. You keep the 6 Taler you did not invest. So you earn in total* $8 + 6 = 14$ *Taler in this period*

Example 2 *All participants invest their 10 Taler into the common project. So you earn* $50 \times 0.64 = 20$ *Taler in this period.*

Example 3 *All participants invest 0 Taler into the common project. So you simply keep the 10 Taler you received in the beginning.*

Example 4 *All the other participants invest 10 Taler into the common project, you invest nothing. So you earn* $40 \times 0.4 + 10 = 26$ *Taler in this period.*

Example 5 *All the other participants invest nothing into the common project, you invest 10 Taler. So you earn* $10 \times 0.4 = 4$ *Taler in this period.*

[`This part only in treatment VCMC:`

Each participant can send messages to the other participants. All conditions have the following form and will be shown to the other participants at the beginning of the next round:

"I will contribute X Taler, if in total at least Y Taler are contributed (by all members of the group including yourself)"

In fact, you can formulate **two different conditions of this kind**. For example, you could signal the other participants that you would be willing to contribute more, if all the others do so too. For example, you could send the following two messages:

First condition: **"I will contribute 5 Taler, if in total at least 20 Taler are contributed (by all members of the group including yourself)"**

Second condition: **"I will contribute 10 Taler, if in total at least 50 Taler are contributed (by all members of the group including yourself)"**

However, the messages are non-binding. That is, you can still choose your contributions freely. Also, you do not have to send messages if you do not want to.]

Payoff

The experiment lasts for 60 periods. A questionnaire will follow after the 60 periods. At the end of the experiment **one of the 60 periods will be randomly chosen** and the earnings of the chosen period will be paid with an exchange factor of 1 Taler = 1 €. The payment will be private and in cash. For example, if your earnings from the payment relevant round is 15 Taler, you will receive 15 €.

Structure of the Screen

You have a printed example for the structure of the screen of the program you will make your decisions with in each period. The screen is divided into three blocks.

The upper left block contains a calculator. Here you can test actions for you and the four other players (by default the computer will show here the decisions of the last period. Of course you can change them). Once you selected an action for every player you can press the button "Calculate payoff!" and the computer will calculate the payoff you would obtain in this case.

In the upper right block [`only in VCMC:` you enter the two messages which the other participants will see at the beginning of the next round. Furthermore...] you enter the

Figure 6: Screenshot VCM.

action that will be relevant for your payoff (here the computer will show as a default your contribution of the last period as well. If you want to change it, you have to do so now). Below there is a red button. When you press this button you submit your decision and leave the screen. Only when **all** players have pressed the red button the experiment continues. A clock on the upper right gives you a hint until when you should have made a decision. If the time runs out this has no consequences.

In the bottom block you will see from the second period onwards the actions of all players in the previous round. [`only in VCMC:` Furthermore you see the messages sent by all participants (if they sent any).] In the first period this block will be empty.

6.1 Test questions

CCM Quiz

1) Suppose you contribute 5 Talers to the joint project. The other participants contribute 15 Talers. What is your payout for this round?

2) Suppose you choose as a first condition: “I contribute 10 if a total of 30 is contributed.” As a second condition: “I contribute 9 if a total of 25 is contributed.” If the others contribute 18 in total, how much will you contribute in this round ?

VCM Quiz

1) Suppose you contribute 5 Talers to the joint project. The other participants contribute 15 Talers. What is your payout for this round?

2) Suppose you contribute 2 Talers to the joint project. The other participants contribute 38 Talers. What is your payout for this round?

VCMII Quiz

1) Suppose you contribute 5 Talers to the joint project and you are type 2. The other participants contribute 15 Talers. What is your payout for this round?

2) Suppose you contribute 2 Talers to the joint project and you are type 2. The other participants contribute 38 Talers. How much is your payout for this round?

CCMII Quiz

1) Suppose you contribute 5 Talers to the joint project and you are type 2. The other participants contribute 15 Talers. What is your payout for this round?

2) Suppose you choose as a first condition: “I contribute 10 if a total of 30 is contributed.” As a second condition: “I contribute 9 if a total of 25 is contributed.” If the others contribute 18 in total, how much will you contribute in this round?

References

Andreoni, J. and Gee, L. K. (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, 96(11-12):1036–1046.

Andreoni, J. and Gee, L. K. (2015). Gunning for efficiency with third party enforcement in threshold public goods. *Experimental Economics*, 18(1):154–171.

Aumann, R. J. (1959). Acceptable points in cooperative general n-person games. In Aumann, R. J., Luce, R., and Tucker, A., editors, *Contributions to the Theory of Games*, volume IV. Princeton University Press.

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.

Bochet, O., Page, T., and Putterman, L. (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization*, 60(1):11–26.

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.

Cabrales, A. and Serrano, R. (2011). Implementation in adaptive better-response dynamics: Towards a general theory of bounded rationality in mechanisms. *Games and Economic Behavior*, 73(2):360 – 374.

Clarke, E. (1971). Multipart pricing of public goods. *Public Choice*, 11:17–33.

Dawes, R. M., McTavish, J., and Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of personality and social psychology*, 35(1):1.

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

Groves, T. and Ledyard, J. (1977). Optimal allocation of public goods: A solution to the "free rider" problem. *Econometrica*, 45(4):783–809.

Guttman, J. M. (1978). Understanding collective action: matching behavior. *The American Economic Review*, 68(2):251–255.

Guttman, J. M. (1986). Matching behavior and collective action: Some experimental evidence. *Journal of Economic Behavior & Organization*, 7(2):171–198.

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.

Heitzig, J. (2019). Efficient non-cooperative provision of costly positive externalities via conditional commitments. *Available at SSRN.*

Herrmann, B., Thoni, C., and Gachter, S. (2008). Antisocial punishment across societies? *Science*, 319(5868):1362–1367.

Isaac, R. M. and Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, 103(1):179–199.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In Kagel John, Roth, A., editor, *Handbook of Experimental Economics*. Princeton University Press, Princeton.

MacKay, D. J. C., Cramton, P., Ockenfels, A., and Stoft, S. (2015). Price carbon - I will if you will. *Nature News*, 526(7573):315–316.

Marx, L. M. and Matthews, S. A. (2000). Dynamic voluntary contribution to a public project. *The Review of Economic Studies*, 67(2):327–358.

Pecorino, P. (1999). The effect of group size on public good provision in a repeated game setting. *Journal of Public Economics*, 72(1):121–134.

Reischmann, A. and Oechssler, J. (2018). The binary conditional contribution mechanism for public good provision in dynamic settings - theory and experimental evidence. *Journal of Public Economics*, 159:104–115.

Romano, R. and Yildirim, H. (2001). Why charities announce donations: a positive perspective. *Journal of Public Economics*, 81(3):423–447.

Schmidt, K. M. and Ockenfels, A. (2019). Negotiating climate cooperation. *mimeo*.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37.