University of Heidelberg

Department of Economics

Discussion Paper Series | No. 586

The Conditional Contribution Mechanism for the Provision of Public Goods

Andreas Reischmann

March 2015

The Conditional Contribution Mechanism for the Provision of Public Goods*

Andreas Reischmann

Department of Economics
Heidelberg University
Germany

andreas.reischmann@awi.uni-heidelberg.de

March 14, 2015

Abstract

Many mechanisms have been designed to solve the free-rider problem in public good environments. The designers of those mechanisms focused on good static equilibrium properties. In this paper, I propose a new mechanism for the provision of public goods that has good dynamic properties instead. The mechanism gives all agents the possibility to condition their contribution on the total level of contribution provided by all agents. Under a reasonable variant of Better Response Dynamics all equilibrium outcomes are Pareto efficient. This makes the mechanism particularly suited for repeated public good environments. In contrast to many previously suggested mechanisms, it does further not require an institution that has the power to enforce participation and/or transfer payments. Neither does it use any knowledge of agents preferences.

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

JEL-Classification: D82, H41, C72

*I thank the participants of the Game Theory Festival 2013 at Stony Brook University, as well as the participants of the IWGTS at the University of Sao Paulo. I also thank Alvin Roth, Jörg Oechssler, Christoph Vanberg, Christoph Brunner, Jonas Hedlund, Thomas Tröger and Hannes Koppel as well as all participants of the seminars at Heidelberg University for helpful comments.

1 Introduction

Numerous mechanisms have been developed in an attempt to solve the free-rider problem in public good scenarios. All those mechanisms were developed with a static solution concept in mind. However, Healy (2006) shows that in repeated public good environments agents' actions can be well described by a dynamic better response behavior. This paper therefore presents a new mechanism that achieves efficient contribution levels under an adjusted better response dynamic. This mechanism is called the Conditional Contribution Mechanism (CCM).

In the CCM agents can free-ride and contribute unconditionally as in the Voluntary Contribution Mechanism. Moreover, agents have the possibility to conditionally contribute. In the most simple environment contribution is binary and agents' utility from the public good increases linearly with the level of the public good. In this environment an offer of conditional contribution has the form "I am willing to contribute, if at least k agents contribute in total". The mechanism then chooses the highest possible level of total contribution that satisfies all those conditions.

Under Better Response Dynamics agents switch only to messages with positive probability that make them weakly better off if nobody else switches as well. In the proposed mechanism all agents are indifferent between many of their messages. Thus, Better Response Dynamics are not sufficiently restrictive for the dynamic process to converge to any equilibria.

However, the conditional contribution structure of the mechanism makes some better responses more plausible than others in the long term. Once a certain level of contributions is reached, messages can be separated in two sets. Under the messages in one set other agents can change the outcome such that the first agent is worse off, but still has to contribute to the public good. The second set of messages makes sure that in all outcomes in which the first agent has to contribute he is at least as well off as he is in the current outcome.

The first kind of messages, which increase other agents incentives to free-ride, shall be called exploitable. The second kind shall be called unexploitable. The solution concept used in this paper, Unexploitable Better Response Dynamics, assumes that in the long run agents only choose strategies which are better responses and not exploitable.

The central result of the binary model is that an outcome is an equilibrium outcome of the proposed mechanism under Unexploitable Better Response Dynamics if and only if it is Pareto optimal and a strict Pareto improvement over the outcome with zero contribution.

The remaining parts of the paper generalize the environment. First, contributions can now be non-binary. Here the mechanism needs to be adjusted. However, the general idea of offering agents the options to free-ride, conditionally contribute, and unconditionally contribute remains unchanged. In this environment the equilibrium results mirror the results for binary contribution.

Second, the environment is generalized to cover weakly monotonic increasing instead of linear valuation functions. In this case Pareto optimality will not be enough to ensure that an outcome is part of a recurrent class. Since utility gained from the public good no longer increases linearly with the contribution towards the public good, there might now be coalitions of agents who benefit from reducing their own contributions even if all other agents then contribute nothing any more. In this environment an outcome is an outcome of a recurrent class of the mechanism under Unexploitable Better Response Dynamics if and only if it is in the core and any deviation of a coalition from this outcome makes at least one agent in that coalition strictly worse off. This holds if at least one such outcome exists. Existence can be guaranteed by adding only infinitesimal monetary incentives.

1.1 Related literature

This work relates in particular to three branches of the literature. The first one is given by work on mechanisms to increase contributions to public goods. The earliest work dates back to Lindahl (1919). However, his pricing system turned out to be not incentive compatible. The most prominent incentive compatible mechanisms were then designed by Clarke (1971) and Groves and Ledyard (1977). More recent advances are the Jackson-Moulin mechanism

(Jackson and Moulin, 1992) or the Falkinger mechanism (Falkinger et al., 2000). However, all those mechanisms have their own draw-backs. Some e.g. require participation to be enforceable, or a high level of information about other agents' preferences to reach the desired equilibrium.

Second, there are experimental studies on public good provision. For a general survey I refer to Ledyard (1994), or the more recent surveys of Chen (2008) and Chaudhuri (2011). Two smaller branches of this literature support the idea that the CCM should be successful.

First, the studies of Fischbacher et al. (2001) and Kocher et al. (2008) show that some agents have strict preferences for conditional cooperation. I do not use this fact in the equilibrium analysis. However, it is obvious that preferences for conditional cooperation make it more likely that agents choose to conditional contribution instead of free-riding.

Second, there are certain papers that compare the performance of the Voluntary Contribution Mechanism (VCM) experimentally to the performance of other simple public good mechanisms. Two mechanisms have been found to be able to increase contributions at least in some situations. The auction mechanism by Smith (1979, 1980) and the Provision Point Mechanism (PPM) studied e.g. in Rondeau et al. (1999, 2005). Those mechanisms have in common that they use a sharp discontinuity to prevent the incentives of free-riding. The CCM shares this property of a sharp discontinuity.

In the PPM and the auction mechanism the value of this discontinuity has to be chosen by the mechanism designer. With a lack of knowledge of agents' preferences this can lead to failure of the mechanism to provide any contributions. In the CCM the value of the discontinuity depends on agents' messages. Thus, it is no longer exogenously fixed but can dynamically adjust itself to the optimal value.

The third branch of the literature focuses on Better Response Dynamics in mechanisms. I already mentioned that Healy (2006) provides experimental evidence that agents' behavior in public good mechanisms can be well described by a better response model. The importance of Better Response Dynamics in mechanisms is further highlighted by the recent introduction of Better Response Dynamics into the implementation literature by Cabrales and Serrano (2011).

1.2 Plan of the paper

The remaining sections are structured as follows. In section 2, I introduce the Binary Conditional Contribution Mechanism (BCCM) in the simplest possible setting. Valuations are linear and contribution to the public good is binary. Section 3 introduces Unexploitable Better Response Dynamics and the outcomes of recurrent classes of the BCCM under UBRD are calculated. Section 4 removes the assumption that contributions are binary and introduces the (non-binary) Conditional Contribution Mechanism. In Section 5, the assumption of linear valuations is replaced with the weaker assumption of weakly increasing valuation functions. Section 6 provides a summary and discussion of the results. Proofs of all theorems can be found in the Appendix.

2 The Binary Conditional Contribution Mechanism

I consider a public good environment in the following form. All $n \in \mathbb{N}$ agents labeled i are considered to have one monetary unit available in each period, which they can either keep or invest in one unit of the public good. An outcome is then defined as $z = (z_1, ..., z_n)$ with $z_i \in \{0, 1\}$, $\forall i \in I := \{1, \ldots, n\}$, where $z_i = 1$ is interpreted as agent i investing his monetary unit into the public good and $z_i = 0$ represents agent i keeping his monetary unit for himself. For notational convenience define $\underline{z} = (0, \ldots, 0)$.

Further, all agents $i \in I$ have a valuation $\theta_i \in [0, 1)$ for the public good.[1] Utility of agent

[1] 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 of 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 this 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.

i is then given by a quasilinear utility function of the form

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

Valuations θ_i are further assumed to be such that some outcome z exists, which is a strict Pareto improvement over $\underline{z}$ for all agents i, who contribute in z. This assumption ensures that some strict improvement over $\underline{z}$ is possible.[2]

We do not make any specific assumption on whether agents are informed on the valuations of other agents or not. Nash equilibrium is considered only as a first predictor of possible dynamically stable outcomes and the main solution concept is a dynamic adjustment process. Therefore, the results of this paper apply whenever this adjustment process describes behavior reasonably well. This might be the case in environments with complete or incomplete information.

2.1 The mechanism

In the Binary Conditional Contribution Mechanism $G^{BCCM} := (M^{BCCM}, g^{BCCM})$ 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 $M_i^{BCCM} := \{1, 2, \ldots, n+1\}$, $\forall i \in I$. The chosen message is thereby 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.[3]

The outcome selected by the mechanism is the outcome with the highest possible level of

[2]If this were not the case, any Pareto improvement would rely on some agent's contribution, who is indifferent between this Pareto improvement and $\underline{z}$. No mechanism with the desired properties can be asked to provide strict incentives to contribute for this agent in such an environment. Thus, such cases are not considered in the equilibrium analysis.

[3]Since there are only n agents, there can never be $n+1$ contributing agents.

contributions such that all those statements are satisfied. Formally, define

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

The outcome of the mechanism is defined as $g^{BCCM}(m) = z$ with $z_i = 1$ if and only if $m_i \leq K(m)$.[4]

2.2 Nash equilibria of the BCCM

The BCCM has multiple Nash equilibria. An example shall demonstrate what properties an outcome must have to be a Nash equilibrium outcome.

Example 2.1 *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 incentive do deviate.*

The incentive structure in the example can be generalized. For any outcome there is a message profile that limits the options of agents to the following ones: Agents that currently do not contribute can only alter the outcome by unilaterally contributing themselves, which makes them worse off. Agents that currently contribute can only change the outcome to $\underline{z}$. This indicates that a certain outcome can be implemented as a Nash equilibrium if and only if there is no agent for which the deviation to $\underline{z}$ is profitable.

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

Theorem 2.2 *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 2.2 predicts equilibria which are Pareto efficient as well as equilibria which may not be Pareto efficient such as those equilibria with outcome $\underline{z}$. Thus, the Nash equilibrium concept does not make a clear prediction as to the equilibrium outcome of the mechanism. Nor does it predict the efficiency of equilibrium outcomes. Therefore, a suitable refinement of the Nash equilibrium concept is needed.

3 Unexploitable Better Response Dynamics

As mentioned in the introduction, Better Response Dynamics have been found to describe agents' behavior in repeated public good games rather well (Healy, 2006). Thus, the focus of this section is on Better Response Dynamics as a solution concept. In the following I demonstrate why simple Better Response Dynamics can not be used for the proposed mechanisms. And I motivate a variant of Better Response Dynamics that will be used instead. The intention of this concept is not to perfectly describe subject behavior. The purpose is rather to define a dynamic concept that captures all incentives relevant in the long term. The aim is to correctly predict the set of long term stable outcomes.

Better Response Dynamics assume that a mechanism is played repeatedly by the same agents over a finite or infinite number of periods t. In any period one or more agents are allowed to adjust their message. Agents deviate with positive probability from their current message m_i^t to any message m_i^{t+1} that is a better or best response to m^t. A recurrent class of such a dynamic concept 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.

However, when m_{-i}^t (the message profile containing the messages of all players but i) is fixed, all messages in the BCCM of agent i will lead to only two possible outcomes. This implies that agents will myopically be indifferent between most of their messages. A dynamic adjustment process that only considers myopic better or best response behavior will then have the entire strategy space as its only recurrent class. Thus, simple Better Response Dynamics are not restrictive enough as a solution concept.

I propose to combine the myopic better response condition with a second condition on behavior that is less myopic. Consider the following example.

Example 3.1 *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* $m_i^{t+1} \in \{1, 2, 3, 4\}$ *is a better response for agent* i *to the message profile* m^t. *None of those 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 making those agents worse and agent* 2 *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. The special structure of the mechanism makes it possible for agents to prevent this kind of incentives for exploitation without having to free-ride themselves.*

From a strategic perspective the exploitable messages in the example provide other agents with incentives to deviate to less cooperative messages. Thus, not choosing those messages can be interpreted like a second order better response behavior. Agents assume that other agents better respond to the message profile and choose of their own better responses the ones that are strategically optimal. There are more arguments that rationalize this behavior. It is easier, however, to provide those arguments once the term “exploitable“ and with it Unexploitable Better Response Dynamics are precisely defined.

Definition 3.2 *Given a message profile* m *and an outcome* $g(m) = z$*, a deviation from* m_i *to* m'_i *is called exploitable if there is* $m_{-i} \in M_{-i}$ *such that* $z'(m_{-i}) := g(m'_i, m_{-i}) \prec_i z$ *and* $z'_i(m_{-i}) > 0$*. A message* m'_i *is called unexploitable if it is not exploitable.*

In the following the assumptions of better responding and unexploitability are combined to one behavioral model.[5]

Definition 3.3 *In Unexploitable Better Response Dynamics (UBRD) all agents can adjust their message in every period. Agent* i *switches in period* t *to message* m_i^t *with strictly positive probability if and only if*

- m_i^t *is a (weak) better response to* m^{t-1} *and*
- m_i^t *is unexploitable with respect to* $z^{t-1} := g^{BCCM}(m^{t-1})$*.*

Revisit the example from above with this definition in mind.

Example 3.4 *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* $\underline{z}$*. Therefore, a message is exploitable in this case 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* $m_i = 3$*. Thus, when the current outcome is* $\underline{z}$ *a message is exploitable if and only if it is weakly dominated.*

Therefore, unexploitability can be summarized by two assumptions. First, if agents did not yet coordinate on any Pareto improvements, agents do not send weakly dominated messages. Second, once agents coordinated on a positive level of contributions, they do not choose messages that set incentives for other agents to free-ride on their contribution.

Furthermore, it is not necessary that all agents behave in an unexploitable way. If a large enough subgroup of agents behaves according to UBRD, while the rest of the agents is

[5] Such a model must further specify whether only one or all agents can change their message in a given period. The latter seems more reasonable for most applications (e.g. international environmental agreements). Thus, I assume in the analysis that all agents can adjust their message every period.

just better responding, the equilibrium outcomes will be as efficient as if all agents behaved according to UBRD. Consider again an example.

Example 3.5 *Assume there are 5 identical agents all with type* $\theta_i = 0.4$. *Let the current message profile be* $m = (5, 5, 5, 1, 1)$. *In this case only agents* 1 *through* 3 *send an unexploitable message. Nevertheless, neither of the agents can strictly benefit from any deviation. Although agent 4 and 5's messages are exploitable any attempt to exploit those 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 to support full cooperation.*

3.1 Equilibrium properties of the BCCM under UBRD

Under the stated assumptions agents will learn over time not to choose messages which make them worse off. And they will learn to choose messages that prevent other agents from exploiting their contribution offers. Under the combination of those two assumptions an outcome is stable if and only if it is Pareto optimal and no agent would be equally well or better off in $\underline{z}$. 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* $\theta_i \neq 0$. [6]

With this definition I can prove 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* $\underline{z}$.

[6] Agents who do not profit from the public good ($\theta_i = 0$) can never be strictly better off than in $\underline{z}$. If those agents are excluded this definition of *strict** is not necessary. However, the existence of agents with a valuation of $\theta_i = 0$ makes many mechanisms, which try to force agents to cooperate, to lead to outcomes that are not individually rational. It is thus important to include this case to demonstrate that the BCCM can handle it.

Let me again provide an example to improve the intuition for this result:

Example 3.8 *Consider a case 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. Those outcomes have in common that they are Pareto efficient in a non-transferable utility setting. Assume for example that the current message profile is* $m = (4,4,4,4,6)$. *Then agents 1 through 4 contribute to the public good, while agent 5 does not. Thus, the outcome is* $z = (1,1,1,1,0)$. *Any deviation of agent 5 would lead to* $z' = (1,1,1,1,1)$ *and would thus not be a better response. For agents 1 through 4 messages* $m_i \in \{5,6\}$ *would lead to the outcome* $\underline{z}$. *They are thus not better responses either. Messages* $m_i \in \{1,2,3\}$ *however make outcomes possible in which the agent has to contribute, but total contribution is less than* 4. *Thus, those messages are exploitable. Therefore, the given message profile is a steady state of UBRD.*[7]

4 Non-binary Conditional Contribution Mechanisms

The environment 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.[8]

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 be introduced afterwards.

[7] In this example the other steady states are given by $m' = (5,5,5,5,5)$ and $m'' = (3,3,3,6,6)$ (in any permutation)

[8] This 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.

4.1 The Natural Extension Mechanism

The natural extension of the BCCM will assign every agent i the message space $M_i^{NEM} := \{0, 0.01, \ldots, 0.99, 1\} \times \{0, 0.01, \ldots, n - 0.01, n\}$, where $m_i = (\alpha_i, \beta_i)$ is interpreted as "I am willing to contribute α_i to the public good if total contribution is at least β_i." For the analysis in this section I refer to this mechanism as the Natural Extension Mechanism (NEM). The outcome space is then given by $Z := \{0, 0.01, \ldots, 0.99, 1\}^n$, where z_i is the contribution of agent i to the public good in outcome z. $\underline{z} := (0, \ldots, 0)$ 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 $Z^{NEM}(m) \subset Z$ be the set of all outcomes that satisfy all conditions in m. This can be formalized by

$$z \in Z^{NEM}(m) \Leftrightarrow \left(z_i = 0 \text{ or } z_i = \alpha_i \text{ and } \sum_{j=1}^{n} z_j \geq \beta_i \right), \ \forall\, i \in I. \tag{3}$$

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

$$g^{NEM}(m) = \text{argmax}_{z \in Z^{NEM}(m)} \sum_{i=1}^{n} z_i. \tag{4}$$

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

Theorem 4.1 *An outcome z is an outcome of a Nash equilibrium of the NEM if and only if $z \succeq_i \underline{z}$, $\forall\, i \in I$.*

Revisit the example

Example 4.2 *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 NEM has undesirable equilibria under UBRD as well. The simplest way to show this is by considering an example.

Example 4.3 *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 $m_1' = (\alpha_1, \beta_1)$ 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. $\beta_1 = 0.3$ 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 (α_i, β_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.[9]

4.2 The Conditional Contribution Mechanism

I call this mechanism the Conditional Contribution Mechanism $G^{CCM} := (M^{CCM}, g^{CCM})$. Every agent can announce two tuples $\{(\alpha_i^1, \beta_i^1), (\alpha_i^2, \beta_i^2)\} \in M_i^{CCM} := M_i^{NEM} \times M_i^{NEM}$. The outcome $g^{CCM}(m)$ of the CCM is then defined as in the NEM 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:

$$z \in Z^{CCM}(m) \Leftrightarrow \; z_i = 0 \text{ or } \left\{ \exists l_i \in \{1,2\} : \; z_i = \alpha_i^{l_i} \text{ and } \sum_{j=1}^{n} z_j \geq \beta_i^{l_i} \right\}, \; \forall \, i \in I \tag{5}$$

The outcome of the CCM is then uniquely defined by

$$g^{CCM}(m) = \text{argmax}_{z \in Z^{CCM}(m)} \sum_{i=1}^{n} z_i.^{10} \tag{6}$$

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

[9] Depending 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.

[10] The 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 4.4 *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 4.5 *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.*[11]

5 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 quasi-linear utility functions $u_i(w_i, w_p) = w_i + f_i(w_p)$, where w_i is the private wealth of agent i and w_p is the total amount of wealth

[11] Agents 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.

invested into the public good by all agents. The functions f_i are only assumed to be weakly increasing in the level of the public good and may differ across agents.[12] Endowment and outcome space $Z := \{0, 0.01, \ldots, 1\}^n$ remain unchanged.[13]

In this setting Pareto optimality will not be enough to ensure that an outcome is part of a recurrent class. Utility gained from the public good increases no longer linearly with the contribution towards the public good. Therefore, there might now be groups of agents who benefit from reducing their own contributions even if all other agents would not contribute anything any more.

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.

Definition 5.1 *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).

Definition 5.2 *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 $z' \succ_i z$, $\forall \, i \in S$.*

Since I already demonstrated that Nash equilibrium does not even uniquely predict the outcome in the linear case I skip the static analysis and present only the result under UBRD. 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.

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

[13]A 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.

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.

Definition 5.3 *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 5.4 *Define the subset $S^C(z) \subset I$ via $i \in S^C(z)$ if and only if $f_i(\sum_{i=1}^n z_i) > 0$ as the set of agents that strictly benefit from the amount of public good in z.*

Theorem 5.5 *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. This is shown, by proving that the mechanism can be adjusted to guarantee existence at arbitrarily low expected costs.[14]

In the following theorem let Δ be a mapping from $Z \times I \to \mathbb{R}_+$. 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 payed if outcome z occurs. I write $G + \Delta$ to describe a mechanism G to which the additional payments Δ are added.

Theorem 5.6 *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.*

[14] Since costs are arbitrarily low I do not want to argue here who should pay those costs. Note though that in reality costs for setting any such incentives can never be arbitrarily low since the administration costs will be strictly positive. However, the theorem is not meant to fix the problem in applications, but rather to show that the problem is likely to have no effect in real applications at all. 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.

6 Summary and discussion

This paper introduces the class of Conditional Contribution Mechanisms for the provision of public goods. In these mechanisms agents can condition their contribution on the total contribution of all agents. There are efficient as well as inefficent Nash equilibrium outcomes. However, under Unexploitable Better Response Dynamics all equilibrium outcomes turn out to be Pareto efficient, in the non transferable utility sense.

A new concept, Unexploitable Better Response Dynamics, is introduced in the paper to predict the outcomes of the mechanisms. Although the concept is close to the standard concept of Better Response Dynamics and the new unexploitability condition can, besides other arguments, be related to eliminating weakly dominated strategies, there always remains some doubt as to the predictive power of a new solution concept. Therefore, experiments with these mechanisms have to be conducted. A first experiment with the binary environment shows that the BCCM significantly outperforms the VCM in terms of contribution rates and Unexploitable Better Response Dynamics is a good predictor for the stable equilibrium outcomes (Reischmann, 2015).

Good dynamic equilibrium properties combined with ambiguous Nash equilibrium properties indicate that the mechanism might only be suited for repeated public good problems. However, there are a lot of possibilities to adjust the mechanism for a one-shot game such that the dynamic properties are used. As one example the mechanism could be played five times with the highest contribution in the five trials being used as the outcome. This is close to the way in which the auction mechanism studied by Smith (1979, 1980) makes coordination possible. Further, agents could be allowed to communicate prior to the one shot game. This form of cheap talk communication was already used successfully to increase contributions in a standard VCM public goods game by Isaac et al. (1985). In the VCM agents have a myopic incentive to lie about the message they intend to send. In the CCM agents do not have such an incentive to lie, since failed coordination makes everyone worse off. Thus, communication should work even better with the CCM. Finding the best way to adjust the mechanism to one shot games is an interesting question for further research.

Everything considered, the class of Conditional Contribution Mechanisms is an important addition to the set of public good mechanisms. It satisfies individual rationality, incentive compatibility, and leads under UBRD to Pareto efficient outcomes in repeated public good environments. Furthermore, in the final analysis the only assumption on valuations is that they are weakly increasing in the level of the public good. Those weak assumptions make the mechanism applicable in a wide variety of public good settings.

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 2.2 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 I prove 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: In the discussion of the environment I assumed that there exists some Pareto improvement z over $\underline{z}$, which is strict for all $i \in I_1$. Such a Pareto improvement is further strict for all agents i with $\theta_i > 0$.

Let z be any such outcome 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 strategies 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 contribution 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 messages of agent j that made the change from z to z' possible was exploitable. Thus, j would not have chosen this message under UBRD. And 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. □

Proof of Theorem 4.1 Let $z := (z_1, \ldots, z_n) \in Z$ 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 $m_i = (z_i, \bar{\beta})$ 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 choses.

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.□

Proof of Theorem 4.4 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 $\bar{\beta} := \sum_{i=1}^{n} z_i$. Then $\alpha_i^1 = \alpha_i^2 = z_i$ and $\beta_i^1 = \beta_i^2 = \bar{\beta}$ 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 $\bar{\beta} := \sum_{i=1}^{n} z_i$ and $\bar{\beta}' := \sum_{i=1}^{n} z'_i$. Then for any agent i the message $\alpha_i^1 = z_i$, $\beta_i^1 = \bar{\beta}$, $\alpha_i^2 = z'_i$, $\beta_i^2 = \bar{\beta}'$ 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 $\alpha_i^1 = z'_i$, $\beta_i^1 = \bar{\beta}'$, $\alpha_i^2 = z'_i$, $\beta_i^2 = \bar{\beta}'$ 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 $\alpha_i^1 = 0$, $\beta_i^1 = 0$, $\alpha_i^2 = 0$, $\beta_i^2 = 0$ 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. □

Proof of Theorem 5.5 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 $\bar{\beta} := \sum_{i=1}^{n} z_i$. Then $\alpha_i^1 = \alpha_i^2 = z_i$ and $\beta_i^1 = \beta_i^2 = \bar{\beta}$ 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 $\bar{\beta} := \sum_{i=1}^n z_i$ and $\bar{\beta}' := \sum_{i=1}^n z_i'$. Then the message $(z_i, \bar{\beta}), (z_i', \bar{\beta}')$ 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 $\bar{\beta}' := \sum_{i=1}^n z_i' < \bar{\beta} := \sum_{i=1}^n z_i$ or this would be a Pareto improvement. Then in a first step the messages $(z_i, \bar{\beta}), (z_i', \bar{\beta}')$ are unexploitable better responses for every agent $i \in I_1'$. Once all agents $i \in I_1'$ switched to those messages, the messages $(z_i', \bar{\beta}'), (z_i', \bar{\beta}')$ and $(z_i, \bar{\beta}), (z_i, \bar{\beta})$ are both unexploitable better responses for those agents, since the current outcome is still z. But if now simultaneously one agent chooses $(z_i', \bar{\beta}'), (z_i', \bar{\beta}')$ and another one chooses $(z_j, \bar{\beta}), (z_j, \bar{\beta})$, 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. □

Proof of theorem 5.6 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 $r : Z \to \{1, ..., N_Z\}$ be any bijective mapping, which satisfies $\sum_{i=1}^n z_i > \sum_{i=1}^n z_i' \Rightarrow r(z) > r(z')$. Define the mapping $\Delta_{zi} = \frac{r(z)\min(\epsilon,\epsilon')}{2nN_Z}\ \forall\ i \in I$. Total cost of this mapping can be estimated in the following way:

$$\sum_i^n \Delta_{zi} = \sum_i^n \frac{r(z)\min(\epsilon,\epsilon')}{2nN_Z} \leq \sum_i^n \frac{N_Z \min(\epsilon,\epsilon')}{2nN_Z} \leq \frac{n\min(\epsilon,\epsilon')}{2n} \leq \frac{\epsilon}{2} \tag{7}$$

Thus the mapping has total cost of at most $\frac{\epsilon}{2}$. Assume now to the contrary that some agent i is indifferent between any two outcomes z and z' under the mechanism with the incentive scheme Δ. This indifference implies:

$$u_i(z) + \Delta_{zi} = u_i(z') + \Delta_{z'i} \Leftrightarrow u_i(z) - u_i(z') = \Delta_{z'i} - \Delta_{zi} \tag{8}$$

The absolute value of the left-hand side of this 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 incentive scheme Δ leads to a mechanism in which no agent is indifferent between any two outcomes.

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 $z''' := (\max\{z_1, z_1''\}, ..., \max\{z_k, z_k''\}, z_{k+1}'')$ 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.□

References

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.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14(1):47–83.

Chen, Y. (2008). Incentive-compatible mechanisms for pure public goods: A survey of experimental research. In Handbook of Experimental Economics Results, Vol. 1, ed. Charles R. Plott and Vernon L. Smith, pages 625 – 643. Elsevier.

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

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.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters*, 71(3):397 – 404.

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

Healy, P. J. (2006). Learning dynamics for mechanism design: An experimental comparison of public goods mechanisms. *Journal of Economic Theory*, 129(1):114 – 149.

Isaac, R. M., McCue, K. F., and Plott, C. R. (1985). Public goods provision in an experimental environment. *Journal of Public Economics*, 26(1):51 – 74.

Jackson, M. and Moulin, H. (1992). Implementing a public project and distributing its cost. *Journal of Economic Theory*, 57(1):125 – 140.

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.

Lindahl, E. (1919). Just taxation - A positive solution (E. Henderson, trans.). In R.A. Musgrave, A. P., editor, *Classics in the theory of public finance*, pages 98–123. Macmillan, London. 1958.

Owen, G. (1982). *Game Theory*, chapter XIII Games without side payments, pages 288 – 328. Academic Press.

Reischmann, A. (2015). Voluntary vs. Conditional Contribution Mechanism - An experimental study. *Heidelberg Discussionpaper*, (587). http://www.uni-heidelberg.de/fakultaeten/wiso/awi/forschung/discussionpaper.html.

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.

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.