Journal of Public Economics 93 (2009) 429–439

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/econbase

Providing public goods in the absence of strong institutions

Anke Gerber[a,*], Philipp C. Wichardt[b]

[a] Department of Economics, University of Hamburg, Germany
[b] Department of Economics, University of Bonn, Germany

ARTICLE INFO

Article history:
Received 15 January 2008
Received in revised form 15 October 2008
Accepted 16 October 2008
Available online 30 October 2008

JEL classification:
C72
D61

Keywords:
Public goods
Cooperation
Institutions
Climate-change treaties

ABSTRACT

This paper proposes a simple two-stage mechanism to establish positive contributions to public goods in the absence of powerful institutions to provide the public good and to sanction free-riders. In this mechanism players commit to the public good by paying a deposit prior to the contribution stage. If there is universal commitment, deposits are immediately refunded whenever a player contributes her specified share to the public good. If there is no universal commitment, all deposits are refunded and the standard game is played. For suitable deposits, prior commitment and full ex post contributions are supported as a subgame-perfect Nash equilibrium for the resulting game. As the mechanism obviates the need for any ex post prosecution of free-riders, it is particularly suited for situations where players do not submit to a common authority as in the case of international agreements.

1. Introduction

One of the most prominent examples for the failure of markets concerns the provision of public goods. In this case, the conflict of interest between the socially desirable and individually optimal contribution to the public good commonly prevents the implementation of Pareto optimal solutions — not only theoretically but also empirically (e.g. Fehr and Gächter, 2000). Moreover, due to the often immense welfare gains at stake (e.g. clean water/air), the question of how to establish high contribution rates in public goods games is a key issue in economic policy.

An "easy" way out of such social dilemmas are sanctioning institutions that can enforce the desired contributions to public goods (e.g. a reliable jurisdiction). In fact, already casual evidence suggests that, once individual deviations from previously agreed contribution rates can be appropriately punished, many public goods can be — and indeed are — established at a level close to the social optimum (e.g. public transport, health care, quiet sleeping hours at night). Evidence from laboratory experiments further supports this observation (e.g. Falkinger et al., 2000; Fehr and Gächter, 2000); more recent studies even show that individuals, when facing the choice between a social environment with and without sanctioning possibilities, choose the environment with sanctioning (Gürek et al., 2006; see also Kosfeld et al., forthcoming). Moreover, upon being placed in the desired environment, most players indeed fully contribute to the public good and punish free-riders (Gürek et al., 2006). And, of course, inverting the role of the institution, an effect similar to that of sanctions can be achieved through the implementation of some rewarding scheme, given that the institution is provided with the required funds; rewards may, for example, take the form of matching grants as in Boadway et al., (1989). In either case, however, it is necessary that the players submit to some common authority which has the right/power to enforce some kind of punishment or rewarding scheme.

* Corresponding author. Department of Economics, University of Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Germany.
E-mail addresses: anke.gerber@wiso.uni-hamburg.de (A. Gerber), philipp.wichardt@uni-bonn.de (P.C. Wichardt).

0047-2727/$ – see front matter
doi:10.1016/j.jpubeco.2008.10.006

Alternatively, the successful provision of a public good is also possible if the good can be provided by some central institution. This of course requires that the institution is in a position not only to produce the public good but also to obtain the necessary resources from the agents. Thus, the institution either needs the power to tax, or else it must be possible to condition the provision of the public good on the contributions of the agents; in the latter case, considered by Bagnoli and Lipman (1989), the public good is provided if the sum of the contributions is sufficiently large.[1]

In all these cases, however, the provision of the public good eventually hinges on the existence of an institution which is strong in at least one of the following senses: (i) the institution has the power to implement transfers between the agents, i.e. it has to be able to impose (monetary) punishments and/or rewards; or (ii) it has the ability/technology to produce the public good itself and to raise the necessary funds, e.g. because it also has the power to tax. Clearly, such conditions are likely to be satisfied in case of local public goods, where governmental institutions exist to back up the enforcement of the desired contributions and/or where the public good is directly provided by the government (education, street-light, national security etc.). Yet, not least due to the effects of globalisation, an increasing number of public goods do not belong to this category.

Prominent examples for public goods which have to be provided in the absence of strong institutions can be found in connection with environmental issues such as global warming. In fact, questions about the implementation of such public goods have attracted considerable scientific attention in recent years (e.g. Carraro, 1999; Bosello et al., 2003; Dutta and Radner, 2004). In the case of global warming, for example, the public good "clean air" has to be provided in a decentralised way so that each country has to take measures in order to reduce greenhouse gas emissions. At the same time effective sanctioning of free-riders is difficult to establish since there is no institution like a world government which can punish sovereign countries. Accordingly, it is rather unsurprising that we observe a major compliance problem with climate-change treaties like the Kyoto Protocol.[2]

In the present paper, we take up this issue and propose a simple mechanism aimed to establish the provision of a public good in an environment without strong institutions; i.e. the agents cannot be forced to contribute to the public good (e.g. by taxes), effective ex post sanctioning is difficult, if not impossible to enforce, and the public good has to be provided in a decentralised way by the agents themselves. For our analysis, we take as given some agreement on individual contributions to the public good; i.e. we are not concerned with the agents' incentives to actually reach such an agreement but only with its eventual implementation. The considered agreement can, for example, be thought of as the result of some unspecified cooperative bargaining process among the agents. It may be characterised by a number of desirable properties like efficiency, individual rationality, equity etc. Like the Kyoto protocol, however, the agreement is not enforceable due to the absence of sufficiently strong institutions. Hence, it will be implemented only if all agents find it in their self-interest to do so.

The idea of our mechanism is to allow players to take an action, prior to the contribution stage, which renders it a dominant strategy to comply with the agreement. More specifically, we consider a 2-stage modification of a general public goods game. In stage 1, players can choose to pay a deposit to a neutral institution. If at the end of stage 1 everyone has paid the required deposit, then in stage 2 the public goods game is played and deposits are refunded to those who contribute to the public good. If some player has not paid the deposit, all deposits are refunded (potentially deducting a small fee to sustain the institution administering deposits) and the standard public goods game is played in stage 2. With these modifications, universal commitment in the form of paying the deposit and full contributions to the public good now can be rationalised as a subgame-perfect Nash equilibrium of the resulting game. Moreover, the corresponding equilibrium even is in stage-wise weakly dominant strategies. Thus, the mechanism does not require that players know each other's preferences but only that the social planner who designs the mechanism has perfect information about the agents' preferences so as to determine the adequate contributions and deposits.[3] Furthermore, as players essentially execute their own punishment (pay the deposit), the neutral institution considered here only has to resist demands to repay forfeited deposits. This, however, appears far easier to enforce than collecting fines from free-riders ex post.

The rest of the paper is structured as follows. In Section 2, we introduce and analyse our mechanism. We start with an introductory example, then present the general analysis and finally discuss extensions to multi-period public goods problems and issues concerning renegotiation-proofness. Section 3 concludes.

2. Model and results

This section is divided into three parts: an introductory example (Section 2.1), the general analysis (Section 2.2), and a discussion of possible extensions (Section 2.3).

[1] Yet another way many public goods problems are "solved" in practice is via social norms which prescribe the desired behaviour and which are followed by the players in order to avoid mental distress such as feelings of guilt (e.g. Charness and Dufwenberg, 2006) or conflicts with ones identity (e.g. Wichardt, 2008; see also Akerlof and Kranton, 2000, 2005). In such cases, however, there is no conflict between individual incentives and aggregate welfare which is the topic of our paper.

[2] The Kyoto Protocol, in fact, nicely demonstrates the problems that arise in the absence of strong institutions. The current enforcement mechanism of the Kyoto Protocol simply "punishes" a party who has exceeded the assigned amount of emissions by imposing an additional emission deduction of 30% in the next period and by suspending the Party's eligibility to make transfers under emission trading until the Party is reinstated (cf. "http://unfccc.int/kyoto_protocol/compliance/items/2875.php"). The form of punishment used here, however, obviously suffers from a similar compliance problem as the original agreement and hence cannot be considered an effective threat.

[3] Clearly, if the agents themselves negotiate the individual contributions to the public good and determine deposits afterwards, they must have perfect information about each others' preferences.

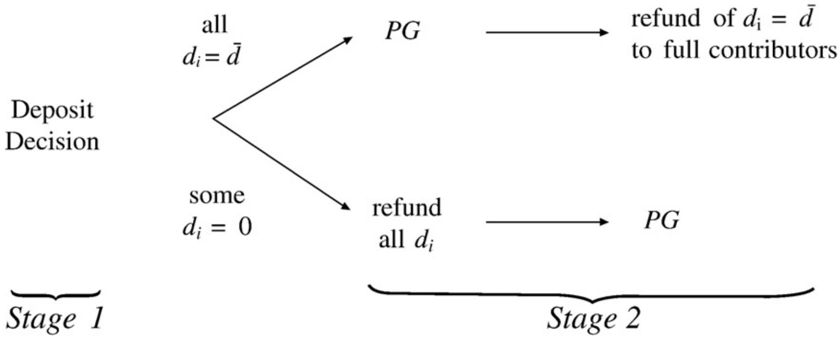

Fig. 1. Graphical illustration of the 2-stage game $\widehat{PG}$. If in stage 1 all players pay the required deposit, PG is played and deposits are refunded to those who contributed $\bar{c}$; otherwise all deposits are refunded first and then PG is played.

2.1. Introductory example

The purpose of the subsequent example is to illustrate the main features of the proposed mechanism for the simple case of a symmetric linear public goods problem. We concentrate on the main aspects of the argument and hence defer any discussion of our assumptions to the next subsection where the general case is analysed.

Consider the following linear, symmetric public goods game denoted by PG. There are two players. Each of them is equipped with an initial endowment e and has to choose a contribution $c_i \in [0,e]$, $i=1,2$, to some public good. The players' preferences over contribution profiles (c_1,c_2) are represented by the utility function $\pi_i(c_1,c_2)=e-c_i+a(c_1+c_2)$, $i=1,2$, with $\frac{1}{2}<a<1$.

Given the specification of PG, it is a strictly dominant strategy for both players to choose $c_i=0$. Accordingly, the unique Nash-equilibrium of PG is given by $c_1^0=c_2^0=0$. However, the outcome corresponding to $c_1^0=c_2^0=0$ is Pareto dominated by the full contribution outcome, i.e. $c_1=c_2=e=:\bar{c}$, because $\pi_i(\bar{c},\bar{c})=e-\bar{c}+2a\bar{c}=2ae>e=\pi_i(0,0)$.[4] The question, thus, is whether and how full contribution can be implemented.

In the sequel, we show that full contributions, $\bar{c}$, by both players can be implemented as a subgame-perfect equilibrium in a 2-stage variation of PG, denoted by $\widehat{PG}$, which is structured as follows: In stage 1, both players simultaneously decide whether to transfer a refundable deposit $\bar{d}>0$ to a neutral institution, e.g. a trustee. The size of the deposit $\bar{d}$ will be determined later. At the end of stage 1, after all players i have chosen $d_i \in \{0, \bar{d}\}$, the profile of deposits (d_1,d_2) is revealed to both players. If both players have deposited $\bar{d}$, then stage 2 directly continues with PG and deposits are refunded to those players who contributed the desired amount, i.e. $\bar{c}$, to the public good. If $d_i=0$ for at least one player i, then, in stage 2, all deposits are refunded immediately and PG is played after that; see Fig. 1 for an illustration.

The stage game of $\widehat{PG}$ that results in case some player i has deposited $d_i=0$ is denoted by PG^0; the stage game that results in case $d_1=d_2=\bar{d}$ is denoted by PG^*. Because PG^0 starts by (automatically) refunding all deposits, the payoff structure of PG^0 is identical to that of PG. Hence, zero contribution is the dominant strategy for both players so that the unique Nash equilibrium of PG^0 again is given by $c_1=c_2=0$. By contrast, the players' payoffs in PG^*, where refunds are conditional on full contribution to the public good, are given by:

$$\pi_i^*(c_1,c_2) = \begin{cases} e-\bar{d}-c_i+a(c_1+c_2), & \text{if } c_i \neq \bar{c} \\ e-c_i+a(c_1+c_2), & \text{if } c_i=\bar{c}. \end{cases}$$

From this it follows immediately that player i's dominant strategy in PG^* is $c_i=\bar{c}$, if deposits are sufficiently large, i.e. if $\bar{d}>(1-a)\bar{c}$.[5] Accordingly, for any profile (d_1,d_2) of deposits paid in stage 1, there is a unique Nash equilibrium in strictly dominant strategies for the resulting subgame at stage 2 and player i's strictly dominant strategy is given by

$$\hat{c}_i^*(d_1,d_2) = \begin{cases} \bar{c}, & \text{if } d_1=d_2=\bar{d} \\ 0, & \text{else.} \end{cases}$$

Intuitively, by paying the deposit $\bar{d}$ in stage 1, players effectively commit to full contributions in the public goods game by implementing a deterrent punishment for themselves. The commitment is conditional, though, as the threat of punishment is only activated in case all players committed to the public good in this way. However, once it is activated, no further steps to execute the punishment are necessary as deposits are already paid and hence the credibility of the (conditional) commitment is not an issue.

Finally, if play continues with the strategy profile $(\hat{c}_1^*,\hat{c}_2^*)$ at stage 2, then $d_i=\bar{d}$ is a weakly dominant strategy for player $i,i=1,2$, at stage 1. To see this, consider the case of player 1. If player 2 deposits $\bar{d}$, then $d_1=\bar{d}$ is strictly better than $d_1=0$. This follows because $\pi_i(\bar{c},\bar{c})>\pi_i(0,0)$ and $(\hat{c}_1^*,\hat{c}_2^*)$ implements $c_1=c_2=\bar{c}$ if $d_1=d_2=\bar{d}$ but only $c_1=c_2=0$ if $d_1=0$. If instead player 2 chooses $d_2=0$,

[4] As maximal contributions need not be Pareto optimal in the general case analysed in Section 2.2, we already use the more general notation, i.e. $\bar{c}$ instead of e, here. Later $\bar{c}$ will refer to the desired contribution.

[5] Observe that $\bar{d}>(1-a)\bar{c}$ for $\bar{d}$ sufficiently close to e.

then the outcome that is implemented by $(\hat{c}_1^*, \hat{c}_2^*)$ in stage 2 is independent of d_1, namely $c_1 = c_2 = 0$, so that player 1 is indifferent between $d_1 = 0$ and $d_1 = \bar{d}$. By symmetry the same argument holds for player 2. Accordingly, $(d_i, \hat{c}_i)_{i=1,2}$, with $d_i = \bar{d}$ and $\hat{c}_i = \hat{c}_i^*$ for $i = 1,2$, is the unique subgame-perfect Nash equilibrium in stage-wise (weakly) dominant strategies.[6] Moreover, in this equilibrium both players contribute the desired amount $\bar{c}$ to the public good.

2.2. General analysis

We now consider a standard public goods problem with n players referred to as consumers. Each consumer i, $i = 1,\ldots, n$, has a preference relation over consumption of a private and a public good. These preferences are represented by a utility function $U_i : \mathbb{R}_+^2 \to \mathbb{R}$, where $U_i(x_i, z)$ is i's utility from consuming x_i units of the private good and z units of the public good. We assume that U_i is strictly increasing:

Assumption A. For all i, it holds that

$$\frac{\partial U_i(x_i, z)}{\partial x_i} > 0 \quad \text{and} \quad \frac{\partial U_i(x_i, z)}{\partial z} > 0.$$

Each consumer i has an initial endowment $e_i > 0$ of the private good while the initial level of the public good is zero. The public good is produced in a decentralised way, i.e. for all i there exists a differentiable production function f_i: $[0, e_i] \to \mathbb{R}_+$ with $f_i(0) = 0, f_i' > 0$, and $f_i'' < 0$. If consumer i contributes $c_i \in [0, e_i]$, $i = 1,\ldots, n$, the realised amount of the public good is

$$z = F\left(\sum_{i=1}^{n} f_i(c_i)\right)$$

for some differentiable function F: $\mathbb{R}_+ \to \mathbb{R}_+$ with $F' > 0$, and consumer i's utility is given by:

$$\pi_i(c_{1,\ldots,}c_n) := U_i\left(e_i - c_i, F\left(\sum_{j=1}^{n} f_i(c_j)\right)\right)$$

Remark 1. From a formal point of view the differentiation between the global and the individual production functions, i.e. between F and the f_i, is not necessary. Yet, in view of applications, it is convenient to separate between these functions. For example, in the context of global warming, one may want to differentiate between technological progress, that reduces the costs for the abatement of greenhouse gas emissions in some country, and natural changes in the decomposition of greenhouse gases in the atmosphere. In this case, f_i can be thought of as the country's technology for reducing greenhouse gas emissions, while $F(\Sigma_i q_i)$ can be interpreted as the "quality" of the atmosphere if each country i has reduced its greenhouse gas emissions by q_i.[7]

Finally, we assume that positive contributions are always strictly dominated by zero contributions. As we will discuss later (see Section 2.3) this assumption can be relaxed. Nevertheless, at the present stage it is useful as it best captures the essence of the public goods problem at issue.

Assumption B. For all $(c_1,\ldots, c_n) \in \times_{j=1}^{n} [0, e_j]$, it holds that

$$\frac{\frac{\partial U_i(x_i, z)}{\partial x_i}}{\frac{\partial U_i(x_i, z)}{\partial z}} > f_i'(c_i)F'(z),$$

for all x_i and for all $i = 1,\ldots, n$, where $z = F(\Sigma_j f_j(c_j))$.

Hence, we assume that for each consumer i the marginal rate of substitution of the private for the public good is larger than i's marginal rate of transformation in the public good's production. Under Assumption B it follows that $a\pi_i(c_1,\ldots, c_n)/\partial ci < 0$ for all i and all contribution profiles $(c_1,\ldots, c_n)$.

Remark 2. The linear public goods problem presented as an introductory example satisfies Assumptions A and B. In this case, $U_i(x_i, z) = x_i + az$ and $f_i(c_i) = c_i$ for some positive constant a with $a < 1$, and F is given by $F(q) = q$.

The resulting public goods problem is modelled as an n-player game PG, where each player i has a strategy set $S_i = [0, e_i]$ and payoff function $\pi_i: \times_{j=1}^{n} S_j \to \mathbb{R}$. Under Assumption B the following Proposition is immediate:

Proposition 1. *The strategy profile c^0 with $c_i^0 = 0$ for all i is the unique Nash equilibrium for PG. Moreover, $c_i^0 = 0$ is a strictly dominant strategy for all i.*

As, in general, the no contribution equilibrium of Proposition 1 does not lead to an efficient outcome, there commonly is a conflict between individual incentives and aggregate welfare. The mechanism proposed below remedies this deficiency. More

---

6 Of course, there also exists another subgame-perfect equilibrium, where both players $i = 1,2$, choose $d_i = 0$ in stage 1 and then play $\hat{c}_i^*$ in stage 2. However, since $d_i = 0$ is weakly dominated by $d_i = \bar{d}$ if play continues with $(\hat{c}_1^*, \hat{c}_2^*)$ at stage 2, this equilibrium seems less plausible than the full contribution equilibrium characterised first.

7 Observe that the quality of the atmosphere need not be a linear function of the aggregate reduction in greenhouse gas emissions.

specifically, it is designed to implement any contribution profile $\bar{c}=(\bar{c}_1,\ldots,\bar{c}_n)\in\times_{j=1}^{n}(0,e_j]$ that strictly Pareto dominates the no contribution profile, i.e. any $\bar{c}$ that satisfies the following condition:

PDOM. For all $i=1,\ldots,n$, $\pi_i(\bar{c}_1,\ldots,\bar{c}_n)>\pi_i(0,\ldots,0)$.

Beyond Condition PDOM, no further assumptions on $\bar{c}$ are necessary. In particular, it is not required that $\bar{c}$ yields a Pareto efficient allocation.

Remark 3. As mentioned before, the vector of contributions $\bar{c}$ can, for example, be thought of as the result of some cooperative bargaining process among the players which cannot be enforced, e.g. due to a lack of strong institutions. In case of the linear public goods problem discussed in the introductory example Condition PDOM is satisfied for $a>1/2$.

In order to implement $\bar{c}$, consider the following 2-stage modification of PG:

Stage 1. All players simultaneously choose to pay a deposit $d_i\in\{0,\bar{d}_i\}$ with $\bar{d}_i>0$ to a neutral institution, e.g. a trustee; later, $\bar{d}_i$ will be chosen such that contributing $\bar{c}_i$ in the public goods game PG is a dominant strategy for player i. At the end of stage 1, the profile of all deposits paid, $d=(d_1,\ldots,d_n)$, is revealed to the players.

Stage 2. The interaction in this stage depends on d in the following way. If at the end of stage 1 we have $d_i=0$ for some i, then all deposits are refunded and the public goods game PG is played. If, however, at the end of stage 1 we have $d_i=\bar{d}_i$ for all i, then a public goods game PG* is played for which payoffs are as follows:

$$\pi_i^*(c_1,\ldots,c_n)=\begin{cases}U_i\left(e_i-c_i-d_i,F\left(\Sigma_j f_j(c_j)\right)\right), & \text{if } c_i\neq\bar{c}_i\\ U_i\left(e_i-c_i,F\left(\Sigma_j f_j(c_j)\right)\right), & \text{if } c_i=\bar{c}_i\end{cases}$$

Thus, if each player i has decided to pay the deposit $\bar{d}_i$, these deposits are only refunded to those players who contribute to the public good in the desired way, i.e. for whom $c_i=\bar{c}_i$. For the moment we simply assume that the neutral institution keeps any forfeited deposits; a discussion of alternative ways to deal with non-refunded deposits is provided in Section 2.3.1.

The 2-stage game defined above is denoted by $\widehat{\mathrm{PG}}$. Player i's payoff in $\widehat{\mathrm{PG}}$ is given by:

$$\Pi_i[(d_1,c_1),\ldots,(d_n,c_n)]=\begin{cases}\pi_i^*(c_1,\ldots,c_n), & \text{if } d_j=\bar{d}_j \text{ for all } j\\ \pi_i(c_1,\ldots,c_n), & \text{else}\end{cases}$$

A strategy for player i in $\widehat{\mathrm{PG}}$ is given by a tuple $(d_i,\hat{c}_i)$, where $d_i\in\{0,\bar{d}_i\}$ and $\hat{c}_i$: $\{0,\bar{d}_1\}\times\ldots\times\{0,\bar{d}_n\}\to[0,e_i]$.

Let $\bar{d}_i\leq e_i$ be such that

$$U_i\left(e_i-\bar{c}_i,F\left(\sum_{j\neq i}f_j(c_j)+f_i(\bar{c}_i)\right)\right)>U_i\left(e_i-\bar{d}_i,F\left(\sum_{j\neq i}f_j(c_j)\right)\right) \tag{1}$$

for all $c_j\in[0,e_j]$ $j\neq i$.[8] Since U_i,f_i and F are strictly increasing and $\bar{c}_i>0$, it follows that inequality (1) is satisfied for $\bar{d}_i=\bar{c}_i$. Hence, by continuity it is also satisfied for any $\bar{d}_i$ in a sufficiently small neighbourhood of $\bar{c}_i$. The following result, then, is readily established:

Proposition 2. *Let Assumptions A and B be satisfied. Then, for any profile $\bar{c}$ of contributions to the public goods game PG that satisfy Condition PDOM, the strategy profile $(\bar{d}_i,\hat{c}_i^*)_{i=1,\ldots,n}$ with*

$$\hat{c}_i^*(d)=\begin{cases}\bar{c}_i, & \text{if } d_j=\bar{d}_j \text{ for all } j\\ 0, & \text{if } d_j=0 \text{ for some } j\end{cases}$$

is a subgame-perfect Nash equilibrium of $\widehat{\mathrm{PG}}$ if $\bar{d}_i$ satisfies inequality (1) for all i. Moreover, for all i, $\hat{c}_i^*$ is a strictly dominant strategy at stage 2 and $\bar{d}_i$ is a weakly dominant strategy at stage 1, if play continues with the strategy profile $\hat{c}^*$ at stage 2.[9]

Proof. Consider stage 2 first. If $d_j=0$ for some j, then PG is played in stage 2. In this case, $\hat{c}_i^*(d)=0$ for all i is the unique Nash equilibrium and it is in strictly dominant strategies (cf. Proposition 1). If $d_j=\bar{d}_j$ for all j, then $c_i=\bar{c}_i$ is a strictly dominant strategy for player i in PG*. To see this, let c_{-i} be an arbitrary profile of contributions for all players except i. By Eq. (1) we have

$$\begin{aligned}&U_i\left(e_i-\bar{c}_i,F\left(\sum_{j\neq i}f_j(c_j)+f_i(\bar{c}_i)\right)\right)>U_i\left(e_i-\bar{d}_i,F\left(\sum_{j\neq i}f_j(c_j)\right)\right)\\ &\Leftrightarrow\pi_i^*(\bar{c}_i,c_{-i})>\pi_i^*(0,c_{-i})\\ &\Leftrightarrow\pi_i^*(\bar{c}_i,c_{-i})>\pi_i^*(c_i,c_{-i})\quad\text{for all } c_i\neq\bar{c}_i\end{aligned}$$

where the last equivalence follows from Assumption B. Hence, $\hat{c}_i^*(d)$ as given in the statement of the proposition is the unique Nash equilibrium in stage 2 and it is in strictly dominant strategies. What remains to be shown is that $\bar{d}_i$ is a weakly dominant strategy for player i at stage 1 given that play continues with $\hat{c}^*$. Consider first the case where $d_j=0$ for some $j\neq i$. Then,

[8] Choosing $\bar{d}_i$ so that inequality (1) is satisfied, of course, requires knowledge of the players' utility functions.

[9] $\bar{d}_i$ is only weakly dominant since player i is indifferent between $d_i=\bar{d}_i$ and $d_i=0$ whenever $dj=0$ for some $j\neq i$, in which case all deposits are refunded.

independent of player i's choice of d_i all deposits are refunded and the game continues with the play of PG at stage 2. Hence,

$$\begin{aligned}\Pi_i\left[\left(\bar{d}_i,\hat{c}_i^*(\bar{d}_i,d_{-i})\right),\left(d_{-i},\hat{c}_{-i}^*(\bar{d}_i,d_{-i})\right)\right] &= \pi_i\left(\hat{c}^*(\bar{d}_i,d_{-i})\right)\\ &= U_i(e_i,0)\\ &= \pi_i(\hat{c}^*(0,d_{-i}))\\ &= \Pi_i\left[\left(0,\hat{c}_i^*(0,d_{-i})\right),\left(d_{-i},c_{-i}^*(0,d_{-i})\right)\right].\end{aligned}$$

Next, consider the case where $d_j=\bar{d}_j$, for all $j\neq i$. Then, by Condition PDOM

$$\begin{aligned}\Pi_i\left[\left(\bar{d}_i,\hat{c}_i^*(\bar{d}_i,\bar{d}_{-i})\right),\left(\bar{d}_{-i},\hat{c}_{-i}^*(\bar{d}_i,\bar{d}_{-i})\right)\right] &= \pi_i^*\left(\hat{c}^*(\bar{d}_i,\bar{d}_{-i})\right)\\ &= \pi_i^*(\bar{c}_1,\dots,\bar{c}_n)\\ &= \pi_i(\bar{c}_1,\dots\bar{c}_n)\\ &> \pi_i(0,\dots,0)\\ &= \pi_i\left(\hat{c}^*(0,\bar{d}_{-i})\right)\\ &= \Pi_i\left[\left(0,\hat{c}_i^*(0,\bar{d}_{-i})\right),\left(\bar{d}_{-i},\hat{c}_{-i}^*(0,\bar{d}_{-i})\right)\right]\end{aligned}$$

□

As we have seen in the proof of Proposition 2, for any deposit profile d there is a unique equilibrium $\hat{c}^*(d)$ at stage 2 of $\widehat{PG}$. This is not true for stage 1 since $\bar{d}_i$ is only weakly dominant for player i. Any tuple $(d,\hat{c}^*)$, with $d_i=0$ for at least two i, also constitutes a subgame-perfect Nash equilibrium of $\widehat{PG}$. In such an equilibrium, there is no contribution to the public good at stage 2. However, $d_i=0$ is weakly dominated by $d_i=\bar{d}_i$ if play continues with $\hat{c}^*$. In fact, $(\bar{d},\hat{c}^*)$ is the unique subgame-perfect Nash equilibrium in stage-wise weakly dominant strategies.

As a last step, we show that the above result remains to hold if the institution collecting and administering the deposits is assumed to be costly, and if players, when paying their deposits, have to bear a small share of this cost. To see this, consider following modification of $\widehat{PG}$, denoted by $\widetilde{PG}$. Different from the previous case, assume now that a fraction $\varepsilon>0$ of any deposit made is kept for the maintenance of the respective institution. Thus, if at the end of stage 1 we have $d_i=0$ for some i, then in stage 2 the game PG is played except that now player i's payoff function is given by:

$$\tilde{\pi}_i(c_1,\dots,c_n)=\begin{cases}U_i\left(e_i-\varepsilon\bar{d}_i-c_i,F\left(\Sigma_j f_j(c_j)\right)\right), & \text{if } d_i=\bar{d}_i\\ U_i\left(e_i-c_i,F\left(\Sigma_j f_j(c_j)\right)\right), & \text{if } d_i=0\end{cases}$$

If at the end of stage 1 we have $d_i=\bar{d}_i$ for all i, then again PG* is played with the following modification of the players' payoff functions:

$$\tilde{\pi}_i^*(c_1,\dots,c_n)=\begin{cases}U_i\left(e_i-\bar{d}_i-c_i,F\left(\Sigma_j f_j(c_j)\right)\right), & \text{if } c_i\neq\bar{c}_i\\ U_i\left(e_i-\varepsilon\bar{d}_i-c_i,F\left(\Sigma_j f_j(c_j)\right)\right), & \text{if } c_i=\bar{c}_i\end{cases}$$

Hence, player i's payoff in $\widetilde{PG}$ is

$$\tilde{\Pi}_i[(d_1,c_1),\dots,(d_n,c_n)]=\begin{cases}\tilde{\pi}_i^*(c_1,\dots,c_n), & \text{if } d_j=\bar{d}_j \text{ for all } j\\ \tilde{\pi}_i(c_1,\dots c_n), & \text{else.}\end{cases}$$

Proposition 3 below states that also under these modifications everyone paying the deposit in stage 1 and full contributions to the public good in stage 2 still can be implemented as a subgame-perfect Nash equilibrium of $\widetilde{PG}$.

Proposition 3. *Let* Assumptions A and B *be satisfied and let $\bar{c}$ be any contribution profile that satisfies Condition PDOM. Furthermore, assume that $\bar{d}_i$ satisfies Eq.* (1) *for all i. Then, there exists $\bar{\varepsilon}>0$ such that for $0<\varepsilon<\bar{\varepsilon}$, $\widehat{PG}$ has exactly two subgame-perfect Nash equilibria $(d^*,\hat{c}^*)$. In one equilibrium, $d_i^*=\bar{d}_i$ for all i in stage 1; in the other equilibrium $d_i^*=0$ for all i in stage 1. In both equilibria $\hat{c}^*$ is as given in* Proposition 2. *Both equilibria are strict.*

The proof immediately follows from the fact that by continuity Condition PDOM implies that

$$U_i\left(e_i-\varepsilon\bar{d}_i-\bar{c}_i,F\left(\sum_j f_j(\bar{c}_j)\right)\right) > U_i(e_i,0)$$

and that inequality (1) implies that

$$U_i\left(e_i - \varepsilon\,\bar{d}_i - \bar{c}_i, F\left(\sum_{j\neq i} f_j(c_j) + f_i(\bar{c}_i)\right)\right) > U_i\left(e_i - \bar{d}_i, F\left(\sum_{j\neq i} f_j(c_j)\right)\right)$$

for all i if ε is sufficiently small.

2.3. Discussion and extensions

In the following, we provide a discussion of several extensions of the preceding model. To begin with, we propose some alternatives to the use of forfeited deposits off the equilibrium path. We then show how the proposed mechanism can be applied to public goods games that do not have a unique equilibrium in dominant strategies. Afterwards we demonstrate how the mechanism applies in cases where the provision of the public good necessitates contributions over more than one period so that intertemporal aspects have to be taken into account. Finally, we address issues concerning incentives to renegotiate and argue that our mechanism indeed is renegotiation-proof in the sense of Farrell and Maskin (1989).

2.3.1. Forfeited deposits

In our game, whenever all players have paid the required deposits in stage 1, these deposits are only refunded to those players who contribute the desired amount to the public good in stage 2. As we have seen, in equilibrium all players contribute the desired amount and hence all deposits are refunded. However, off the equilibrium path some deposits are not refunded and the question is what happens to these forfeited deposits. So far, we have assumed that these deposits are kept by the neutral institution without specifying their further use.

In our view there are at least three alternatives, which all support the equilibrium outcome we have derived before. The first is an inefficient "burning" of the forfeited deposits. The second is to add forfeited deposits to the contributors' refunds, e.g. according to the relative size of their contribution $\bar{c}_i$ to the public good compared to sum of all full contributions $\bar{c}_j$ that were actually made and to simply refund all deposits in case of insufficient contributions by all players ($c_i \neq \bar{c}_i$ for all i).[10] The third is to keep the forfeited deposits within the respective fund and to save them for (efficient) use on issues that also are of common interest to all players but that lie outside the public goods problem considered here, e.g. in the context of international agreements, to finance first aid in case of catastrophes such as droughts. The advantage of the second alternative, of course, is that it technically remains within the boundaries of the strategic problem at hand. A social planner, who has to solve multiple allocation problems, may find the third option slightly more appealing, though.

2.3.2. Public goods game without dominant strategies

So far, we have assumed that for each player i zero contribution, i.e. $c_i = 0$, is a strictly dominant strategy in PG (Assumption B). However, as mentioned before, this assumption was made for expositional purposes only. In fact, the desired contributions can still be implemented as a subgame-perfect Nash equilibrium of our extended game if we abandon Assumption B and, instead, apply a minor adjustment to the corresponding condition PDOM on the Pareto dominating strategy profile. To this end let ξ_i denote a mixed strategy of player i. We then require that the targeted contribution profile $\bar{c} = (\bar{c}_1, \ldots, \bar{c}_n)$, Pareto dominates all Nash equilibria $(\xi_1^0, \ldots, \xi_n^0)$ of *PG* including mixed equilibria.[11] Hence, we replace condition PDOM by the following condition PDOM′, where, slightly abusing notation, π_i now refers to expected payoffs:

PDOM′: For all $i = 1, \ldots, n$ and any Nash equilibrium $\xi^0 = (\xi_1^0, \ldots, \xi_n^0)$ of *PG*, it holds that $\pi_i(\bar{\xi}_1, \ldots, \bar{\xi}_n) > \pi_i(\xi_1^0, \ldots, \xi_n^0)$, where $\bar{\xi}_i$ denotes the mixed strategy of player i that places probability one on $\bar{c}_i, i = 1, \ldots, n$; i.e. the contribution profile $\bar{c}$ Pareto dominates the Nash equilibrium ξ^0.

Proposition 2 then becomes:

Proposition 2′. *Let* Assumption A *be satisfied. Then, for any contribution profile* $(\bar{c}_1, \ldots, \bar{c}_n)$ *that satisfies PDOM′ and any Nash equilibrium* $(\xi_1^0, \ldots, \xi_n^0)$ *of PG, the strategy profile* $(\bar{d}_i, \hat{\xi}_i^*)_{i=1,\ldots,n}$ *with*

$$\hat{\xi}_i^*(d) = \begin{cases} \xi_i^0, & \text{if } d_j = 0 \text{ for some } j \\ \bar{\xi}_i, & \text{if } d_j = \bar{d}_j \text{ for all } j \end{cases}$$

is a subgame-perfect Nash equilibrium of $\widehat{\mathrm{PG}}$ *if* $\bar{d}_i$ *satisfies inequality* (1) *for all i. Moreover, for all i,* $\bar{d}_i$ *is a weakly dominant strategy at stage 1, if play continues with the strategy profile* $\hat{\xi}^*$ *at stage 2.*

The proof is analogous to Proposition 2 except that in case $d_j = 0$ for some j playing $\hat{\xi}_i^*(d) = \xi_i^0$ is not a dominant strategy for player i at stage 2. Instead, in this case, $\hat{\xi}^*(d)$ with $\hat{\xi}_i^*(d) = \xi_i^0$ for all i constitutes "only" a Nash equilibrium of the subgame in stage 2.

2.3.3. Intertemporal aspects and renegotiation-proofness

The subsequent discussion is divided into two parts. To begin with we outline how our mechanism can be applied in case the underlying public goods game extends over more than one period. In the second part, we consider the possibility of renegotiations and argue that the players' strategies implemented by the proposed mechanism are renegotiation-proof in the sense of Farrell and Maskin (1989).

---

[10] The strategy profile specified in Proposition 2 remains a subgame-perfect equilibrium under the described refunding scheme because the best any player i can do in stage 2 still is to play $\bar{c}_i$, given $d_j = \bar{d}_j$ for all j in stage 1 and given $c_j = \bar{c}_j$ for all $j \neq i$; otherwise player i would lose $\bar{d}_i$ which under this scheme is redistributed to players $j \neq i$.

[11] Observe that PG has at least one Nash equilibrium in mixed strategies since the strategy sets are compact and the payoff functions are continuous.

A repeated public goods problem. For the preceding argument, we have assumed that the public good to be provided requires contributions only in one period. Yet, many public goods, in particular non-durable ones, necessitate repeated contributions over a longer horizon. A valid question to be asked, therefore, is whether the mechanism proposed in Section 2.2 extends to the case of repeated contributions to a public good. In the sequel, we address this issue and demonstrate by means of a simplified example how our mechanism can indeed be adapted to public goods that require repeated contributions by all players.[12]

Intuitively, the main idea again is to first collect deposits and to then refund in each period a share of these deposits to those who contributed to the public good in that period. Collecting deposits once at the beginning is preferable to a repeated collection at the start of each period as it saves transaction costs[13]; also, it implies a credible long run commitment in case everyone has paid the deposit.[14] If the deposits collected at the beginning are sufficiently high to cover the necessary refunds in each period, initial commitment, i.e. full deposit payments, and provision of the desired per period contributions to the public good again can be implemented as a subgame-perfect Nash equilibrium.

Consider an n-player version of the linear public goods problem PG discussed in Section 2.1, where for all $i=1,\ldots,n$, $U_i(x_i,z)=x_i+az$, with $\frac{1}{n}<a<1$, $f_i(c_i)=c_i$ and $F(q)=q$. Suppose that there are $T\geq 1$ periods and that players discount future payoffs with a per period discount factor $\delta=\frac{1}{1+r}$, where r is the one-period riskless interest rate. In all periods $t=1,\ldots,T$, each player i receives an endowment $e>0$ of the private good that she can either consume, invest into a riskless asset or spend on the production of the public good. For simplicity, we assume that the public good is non-durable, i.e. there is no stock of the public good that can be transferred across periods; the generalisation to the case of durable public goods, i.e. cases with "stock-effects," is straightforward.

As players' utility functions are linear in the private good and because $\delta=\frac{1}{1+r}$, players are indifferent between immediate consumption and investment into the riskless asset. To further simplify the exposition, we therefore assume that all players immediately consume any fraction of period-t-endowments that is not spent on the public good. Hence, we obtain a repeated game, PG^T, where in each period t player i chooses a contribution $c_i^t\in[0,e]$ to the public good and player i's payoff in period t is

$$\pi_i^t(c_1^t,\ldots,c_n^t)=e-c_i^t+a\sum_{j=1}^{n}c_j^t.$$

Given any sequence of contribution profiles $(c^1,\ldots,c^T)$, player i's overall payoff then is given by

$$\sum_{t=1}^{T}\delta^{t-1}\pi_i^t(c_1^t,\ldots,c_n^t).$$

It is an immediate consequence of the above specification of PG^T, that the T-period public goods game has a unique subgame-perfect equilibrium in dominant strategies, namely zero contributions in all periods by all players, i.e. $c_i^t=0$ for all t and all i. Because $an>1$ by assumption, this equilibrium is Pareto dominated by any sequence of contribution profiles $(c^1,\ldots,c^T)$ with $c_i^t=\bar{c}>0$ for all t and all i. The purpose of the subsequent discussion is to show how this contribution profile can be supported as a subgame-perfect equilibrium in a T-period version of our mechanism, denoted by $\widehat{PG}^T$.

As indicated above, the basic idea of the T-period variation of our mechanism again is to collect deposits whose conditional refund renders full desired contribution to the public good, $\bar{c}$, a dominant strategy for all players. Hence, similar to the one-period case, the mechanism starts with an additional deposit stage preceding the contribution stage of period 1. As before, if all players have paid the required deposit, part of the deposit is refunded each period and the remainder is invested into the riskless asset; the details are explicated below. Compared to the previous 1-period case, however, two modifications are necessary. First of all, deposits now have to provide incentives for T periods instead of 1. And secondly, in case deposit payments are incomplete and all deposits are refunded immediately, we assume that the next period again starts with a new deposit stage. This is done in order to allow for an implementation of the desired outcome with a one period delay.

Following this procedure and again assuming $\bar{d}>\bar{c}(1-a)$, the deposits required at the beginning of a period t are given by

$$\bar{D}^t=\bar{d}\sum_{\tau=t}^{T}\left(\frac{1}{1+r}\right)^{\tau-t}. \tag{2}$$

[12] It has come to our attention that Gersbach and Winkler (2007) in a recent working paper propose a related approach to an intertemporal public goods problem albeit in the specific context of global warming. Moreover, although their analysis uses a refunding mechanism which is similar to ours, it does not address some issues which we believe are essential to this problem. Most importantly, their paper limits itself to a 2-period game with homogeneous players and proposes a mechanism that relies on taxes and their redistribution through a refunding scheme. Yet, it does not discuss the consequences of this redistribution for the players' initial incentives to participate in the refunding scheme, although this aspect is crucial in the more realistic case of heterogeneous players; in our model the participation decision corresponds to the decision to pay the required deposit in stage 1 of the game.

[13] International treaties, e.g. on greenhouse gas reductions, commonly have to be approved by national parliaments. And treaties that require repeated payments would usually imply repeated voting, repeated control of the deposits made etc., all of which is time-consuming and costly.

[14] Recall that there is also an inefficient subgame-perfect Nash equilibrium of the two-stage deposit-contribution game where at least two players do not pay the required deposit and there is no contribution to the public good at stage 2. Therefore, if we simply repeat the two-stage game, there is always the risk that players play the inefficient equilibrium at some point in the future even if they have played the efficient full-contribution equilibrium in the past. If, however, deposits are collected once at the beginning, then a successful completion of the deposit stage implies that full contributions to the public good are a strictly dominant strategy for all players in all future periods.

Given this, consider an arbitrary period t. We distinguish two cases: (a) there is a deposit stage in period t and (b) there is none. Case (a) occurs in period 1 and in any period $t>1$, unless the deposit stage was successfully completed in some period $\tau<t$; below we will give a definition of the successful completion of the deposit stage. To begin with, consider the case where there is a deposit stage in period t. In this case, each player first chooses a deposit $D_i^t \in \{0,\bar{D}^t\}$.[15] Once all deposit choices are made, the resulting profile of deposits is revealed to all players and the game enters the contribution stage. In the contribution stage, players have to decide on their contribution to the public good. If $D_i^t=0$ for some i, then all deposits are refunded immediately and afterwards the game PG is played, where each player i chooses $c_i^t \in [0,e]$. Hence, player i's period t payoff is given by

$$\pi_i^t(c_1^t,\ldots,c_n^t) = e - c_i^t + a\sum_{j=1}^{n} c_j^t.$$

If instead, $D_i^t=\bar{D}^t$ for all i, we say that *the deposit stage was successfully completed in period t*. In this case, the game PG is played and afterwards, each player i who contributes $c_i^t=\bar{c}$ to the public good in period t obtains a partial refund $\bar{d}$ out of her deposit $\bar{D}^t$. Accordingly, player i's period t payoff is given by

$$\pi_i^{*t}(c_1^t,\ldots,c_n^t) = \begin{cases} e - \bar{D}^t - c_i^t + a\sum_{j=1}^{n} c_j^t, & \text{if } c_i^t \neq \bar{c} \\ e - \bar{D}^t + \bar{d} - c_i^t + a\sum_{j=1}^{n} c_j^t, & \text{if } c_i^t = \bar{c} \end{cases}$$

Moreover, whatever remains of player i's deposit $\bar{D}^t$ (which is $\bar{D}^t$, if $c_i^t\neq\bar{c}$, and $\bar{D}^t-\bar{d}$, if $c_i^t=\bar{c}$) is transferred to an account that bears the riskless interest rate. Thus, by construction, the value of player i's account in period $t+1$ is at least $\bar{D}^{t+1}$. Accordingly, once the deposit stage has been successfully completed in some period t, no further deposit stages in later periods are necessary.

Finally, if there is no deposit stage in period $t>1$ (recall that there is always is a deposit stage in period 1), then the game immediately enters the contribution stage for that period and payoffs are

$$\pi_i^{*t}(c_1^t,\ldots,c_n^t) = \begin{cases} e - c_i^t + a\sum_{j=1}^{n} c_j^t, & \text{if } c_i^t \neq \bar{c} \\ e + \bar{d} - c_i^t + a\sum_{j=1}^{n} c_j^t, & \text{if } c_i^t = \bar{c} \end{cases}$$

Proposition 4 below generalises our previous results for the one period case to the present T-period setting. It shows that all players paying the required deposit in period 1 and full contributions to the public good in all later periods again is the outcome of a subgame-perfect equilibrium of $\widehat{\text{PG}}^T$.

Proposition 4. *Let PG be the linear public goods game specified above and let* $\bar{d}>(1-a)\bar{c}$. *Then the following strategies support the contribution* $\bar{c}$ *by all players in all periods as a subgame-perfect Nash equilibrium of* $\widehat{\text{PG}}^T$. *For any period t,* $1\le t\le T$:

(i) If there is a deposit stage in period t, each player i pays the full deposit, i.e. $D_i^{*t}=\bar{D}^t$, *for all* $i=1,\ldots,n$; *then,*
   1. *if* $D_i^{*t}\neq\bar{D}^t$ *for some i, each player i,* $i=1,\ldots,n$, *chooses* $c_i^{*t}=0$ *in the contribution stage of period t.*
   2. *if* $D_i^{*t}=\bar{D}^t$ *for all i, each player i,* $i=1,\ldots,n$, *chooses* $c_i^{*t}=\bar{c}$ *in the contribution stage of period t.*

(ii) If there is no deposit stage in period t, i.e. the deposit stage of some period $\tau<t$ *was successfully completed, then each player i,* $i=1,\ldots,n$, *chooses* $c_i^{*t}=\bar{c}$ *in the contribution stage of period t.*

For all i and all t, c_i^{*t} *is a strictly dominant strategy in the contribution stage of period t. Moreover, for all i, if there is a deposit stage in period t, then* D_i^{*t} *is a weakly dominant strategy for player i, if play continues with the strategy profile* $(c_1^{*\tau},\ldots,c_n^{*\tau})$ *in the contribution stages of all subsequent periods* $\tau\ge t$.

The proof of Proposition 4 is a straightforward generalisation of the proof of Proposition 2, so we only give a sketch of the argument.

Sketch of the Proof. As before, we first note that once all players have paid the required deposits in some period t^*, i.e. there was a deposit stage in t^* and $D_i^{t^*}=\bar{D}^{t^*}$ for all i, then full contribution to the public good in all subsequent periods, i.e. $c_i^t=\bar{c}$ for all $t\ge t^*$, is a strictly dominant strategy for all players; again this essentially follows from the fact that $\bar{d}>(1-a)\bar{c}$. Similarly, $c_i^t=0$ again is the optimal choice for all contribution stages that have not been preceded by a successfully completed deposit stage. Thus, it remains to be shown that paying the full deposit whenever there is a deposit stage in some period t^* is a weakly dominant strategy, if play continues with $(c_1^{*t},\ldots,c_n^{*t})$ in the contribution stages of all subsequent periods $t\ge t^*$. To see this, consider period T first. If there is a deposit stage in period T, then $\bar{D}^T=\bar{d}$ is a weakly dominant strategy for any player, if play continues with $(c_1^{*T},\ldots,c_n^{*T})$; this follows directly from Proposition 2. The argument, then, can be iterated back in order to cover all periods $t<T$, i.e. if there is a

[15] Regarding the payment of the deposit, we assume that players can borrow $\bar{D}^t$ at a per period interest rate of r.

deposit stage in period $t<T$, then $\bar{D}^t$ is a weakly dominant strategy for any player if subsequent play follows the equilibrium strategies as specified in the proposition. This is due to the fact that paying the deposit has the same consequences as not paying it if some other player does not pay the required deposit; in this case, all deposits are refunded immediately. If, however, all other players pay the required deposits, then paying the deposit is optimal as in that case deposits can be used to induce the desired contributions to the public good already in period t and, by construction,

$$\sum_{\tau=t}^{T}\delta^{\tau-t}\left[e-\bar{c}+\bar{d}+an\bar{c}\right]-\bar{D}^{t}>e+\sum_{\tau=t+1}^{T}\delta^{\tau-t}\left[e-\bar{c}+\bar{d}+an\bar{c}\right]-\delta\bar{D}^{t+1}.$$

□

Remark 4. Regarding the use of forfeited deposits, the alternatives discussed in Section 2.3.1 are also feasible here. As before, though, the exact details of the solution are of minor importance and can be adapted to the problem at hand.

Renegotiation-proofness. Subgame-perfection ensures that the equilibrium strategies are individually credible, i.e. no player has an incentive to change her strategy in any subgame on or off the equilibrium path, it may nevertheless be that the punishment is not optimal for the group as a whole, since there exists another continuation equilibrium that strictly Pareto-dominates the "punishment equilibrium." For example, if equilibrium strategies prescribe the punishment of deviators in a way that hurts all players, no punishment and simply proceeding with the earlier equilibrium path strictly Pareto-dominates the execution of the punishment.[16] Accordingly, in this case there is an incentive for the group of all players to renegotiate so that the initial threat to punish is incredible. A subgame-perfect equilibrium that is robust to such renegotiations is called *renegotiation-proof* (cf. Farrell and Maskin, 1989).

Naturally, renegotiation-proofness plays an important role in the context of international environmental agreements and their enforcement (Barrett, 1994; see also Finus and Rundshagen, 1998; Froyn and Hovi, 2008). In the context of our mechanism, however, it is not an issue as all players effectively take care of their own punishment, namely by paying the required deposit to a central agency; in case of later insufficient contributions by some player, the punishment then is carried out by the agency simply by keeping the respective deposit. Thus, there is no costly punishment on the part of the non-deviators, which usually gives rise to potential Pareto-improvements through renegotiations. Moreover, in any subgame the equilibrium strategies, where full deposits are paid at the earliest opportunity and full contributions are paid in all contribution stages from then on, strictly Pareto-dominate any other subgame-perfect equilibrium. Hence, there is no scope for improvements, neither on nor off the equilibrium path.[17]

3. Concluding remarks

In many situations, like in the case of global warming, there is a general consensus among the affected parties that the provision of a particular public good is desirable. Yet, at the same time everyone knows that, ex post, there is a strong incentive to free ride on the contributions of others if there is no institution which can enforce the desired contributions. Hence, low contribution rates are what commonly results.

The mechanism proposed in this paper dissolves the tension between the socially desirable and the individually rational through the introduction of an initial commitment stage. In this stage, players can voluntarily commit to making the desired contribution to the public good by paying an adequate deposit. If all players commit to the public good, deposits are only refunded to those who later make the full desired contribution; otherwise all deposits are refunded immediately. Thus, in effect, commitment to the public good implements a self-sanctioning scheme which, if activated, renders full contributions to the public good a strictly dominant strategy.

The main advantage of the initial commitment stage is that, similar to the public goods decision, general commitment again is socially desirable. Yet, different from the contribution decision in the public goods game, commitment now is also individually rational. This is because the respective self-sanctioning scheme is activated only if all players choose to commit and commitment has no consequences otherwise; contributions to the public good itself, by contrast, are irreversible once made. Accordingly, free-riding is not an issue in the commitment stage.

Moreover, the proposed mechanism does not require the presence or establishment of powerful institutions to implement the desired contributions to the public good. The reason is that no ex post punishment of free-riders is required because of the conditional ex ante self-sanctioning involved in the procedure. All that is needed is an independent institution that collects deposits, monitors the players' contributions and refunds deposits to those who have made the desired contributions to the public good. Hence, the mechanism is particularly suited to implement agreements in situations where there is no common authority for all players that can enforce the punishment of free-riders, e.g. in the case of international agreements such as the Kyoto protocol.

[16] For a specific example consider an infinitely repeated Prisoner's Dilemma game and assume that both players play grim-trigger (i.e. cooperate as long as the opponent does, defect until infinity in case of a deviation of the opponent). In this case, given a single deviation, players are better off by coordinating again on cooperation rather than executing the punishment.

[17] Of course, an inefficient use of forfeited deposits may induce a desire to renegotiate the mechanism itself. This, however, does not concern the players' strategy choices and, hence, is not at issue in the context of renegotiation-proofness in the sense defined above. Nevertheless, requiring an efficient use of forfeited deposits certainly not only renders the mechanism more appealing from a theoretical perspective but also is more plausible in view of applications (cf. Section 2.3.1).

Finally, regarding practical applications, we believe that these are both possible and promising. In fact, whenever international agreements are negotiated, a possible way to complement other incentives to comply with the agreement, e.g. due to reputation effects, would be to add some monetary deposit, which then is used in the way described above; the deposit could, e.g., be payable to the World Bank. Moreover, given the complementary character of the deposits, we would speculate that the size of the deposits necessary to alleviate the compliance problem is considerably smaller than what the preceding discussion — which neglects external aspects such as reputation — might suggest. Yet, to ultimately judge in how far the implementation of such a mechanism is possible, given the diversity of national legal systems and the different national interest groups, is beyond the scope of this paper. Nevertheless, our results suggest that from an economic perspective it might be worthwhile to explore what is politically feasible in this respect.

Acknowledgements

We are grateful to Sylvain Béal, Patrick Schmitz, Avner Shaked and three anonymous referees for helpful comments. Anke Gerber acknowledges support from the center of excellence “Integrated Climate System Analysis and Prediction” (CliSAP) at the University of Hamburg. Philipp Wichardt gratefully acknowledges financial support from the German Research Foundation (DFG) through GRK 629 and SFB/TR 15 at the University of Bonn. This paper was originally circulated under the title "Providing Public Goods without Strong Sanctioning Institutions" (IEW-Working Paper No. 303, University of Zurich, 2006).

References

Akerlof, G., Kranton, R., 2000. Economics and identity. Quarterly Journal of Economics 115, 715–753.
Akerlof, G., Kranton, R., 2005. Identity and the economics of organizations. Journal of Economic Perspectives 19, 9–32.
Bagnoli, M., Lipman, B.L., 1989. Provision of public goods: fully implementing the core through private contributions. The Review of Economic Studies 56, 583–601.
Barrett, S., 1994. The biodiversity supergame. Environmental and Resource Economics 4, 111–122.
Boadway, R., Pestieau, P., Wildasin, D., 1989. Tax-transfer policies and the voluntary provision of public goods. Journal of Public Economics 39, 157–176.
Bosello, F., Buchner, B., Carraro, C., 2003. Equity, development, and climate change control. Journal of the European Economic Association 1, 601–611.
Carraro, C., 1999. International Environmental Agreements on Climate Change. Kluwer Academic Publishers, Dordrecht. The Netherlands.
Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74, 1579–1601.
Dutta, P., Radner, R., 2004. Self-enforcing climate change treaties. Proceedings of the National Academy of Sciences 101, 5174–5179.
Falkinger, J., Fehr, E., Gächter, S., Winter-Ebmer, R., 2000. A simple mechanism for the efficient provision of public goods: experimental evidence. The American Economic Review 90, 247–264.
Farrell, J., Maskin, E., 1989. Renegotiation in repeated games. Games and Economic Behavior 1, 327–360.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. The American Economic Review 90, 980–994.
Finus, M., Rundshagen, B., 1998. Renegotiation-proof equilibria in a global emission game when players are impatient. Environmental and Resource Economics 12, 275–306.
Froyn, C., Hovi, J., 2008. A climate agreement with full participation. Economics Letters 99, 317–319.
Gersbach, H., Winkler, R., 2007. On the Design of Global Refunding and Climate Change. CER-ETH - Center of Economic Research at ETH Zurich, Woking Paper No. 07/69.
Gürek, O., Irlenbusch, B., Rockenbach, B., 2006. The competitive advantage of sanctioning institutions. Science 312, 108–111.
Kosfeld, M., Okada, A., Riedl, A. forthcoming. Institution Formation in Public Goods Games. The American Economic Review.
Wichardt, P., 2008. Why and How Identity Should Influence Utility. SSRN eLibrary.