GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 194

Providing Public Goods Without Strong Sanctioning Institutions

Anke Gerber*
Philipp C. Wichardt**

February 2007

*Anke Gerber, agerber@isb.unizh.ch
**Philipp C. Wichardt, University of Bonn, Department of Economics, Econ. Theory III, Adenauerallee 24-42, 53113 Bonn, Germany. philipp.wichardt@uni-bonn.de.

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.gesy.uni-mannheim.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Konrad Stahl, Ph.D. · Department of Economics · University of Mannheim · D-68131 Mannheim,
Phone: +49(0621)1812786 · Fax: +49(0621)1812785

Providing Public Goods Without Strong Sanctioning Institutions*

Anke Gerber†

Swiss Banking Institute, University of Zurich

Philipp C. Wichardt‡

Department of Economics, University of Bonn

This Version: February 7, 2007

Abstract This paper proposes a simple mechanism aimed to establish positive contributions to public goods in the absence of powerful institutions to sanction free-riders. The idea of the mechanism is to require players to commit to the public good by paying a deposit prior to the contribution stage. If all players commit in this way, those players who do not contribute their share to the public good forfeit their deposit. If there is no universal commitment, all deposits are refunded and the standard game is played. Given deposits are sufficiently high, prior commitment and full ex post contributions are part of a strict 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.

Key words: Public Goods, Cooperation, Institutions, Climate-Change Treaties
JEL code: C72, D61

*Acknowledgements: We are grateful to Sylvain Béal and Patrick Schmitz for helpful comments and discussions. Philipp Wichardt gratefully acknowledges financial support from the German Research Foundation (DFG) through GRK 629 and SFB/TR 15.

†e-mail: agerber@isb.unizh.ch

‡Postal address for correspondence: Philipp Wichardt, University of Bonn, Department of Economics, Econ. Theory III, Adenauerallee 24-42, 53113 Bonn, Germany.

1 Introduction

One of the most prominent examples for the failure of markets concerns the provision of public goods. 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). 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 is the introduction of sanctioning institutions (e.g. a reliable jurisdiction) that enforce contributions to public goods. 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). In fact, more recent studies even show that individuals, when facing the choice between a social environment with and without sanctioning possibilities, learn to choose the environment with sanctioning (Gürerk et al., 2006; see also Kosfeld et al., 2006). Moreover, upon being placed in the desired environment, most players indeed fully contribute to the public good and punish free-riders (Gürerk et al., 2006).

In order to be effective, however, sanctioning institutions have to be equipped with the power necessary to enforce the respective punishment. This, of course, is satisfied in case of local public goods, where governmental institutions exist to back up the enforcement. Powerful sanctioning institutions were also taken as given in the laboratory studies cited above. Yet, in a more and more globalised world, an increasing number of public goods that do not belong to this category have attracted considerable attention; most prominent among them are environmental issues related to global warming (cf. Dutta and Radner, 2004). Absent, e.g., a world government, effective sanctioning of free-riders is difficult to establish in these cases. Accordingly, the optimal provision of public goods is more awkward and we often witness considerable difficulties in implementing actions that were informally agreed upon (e.g. climate-change treaties like the Kyoto Protocol).

In the present paper we take up this issue and propose a simple mechanism aimed to establish full contribution to a public good in cases where effective ex post sanctioning is difficult, if not impossible, to enforce. The idea of the mechanism is to allow players to take an action, prior to the contribution stage, which renders full contribution to the public good a dominant strategy. More specifically, we consider a 2-stage variation 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 a 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. Obviously, universal commitment in the form of paying the deposit as well as full contributions to the public good now can be rationalised as a subgame perfect Nash equilibrium of the resulting game. Moreover, 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. A discussion of the mechanism is provided in Section 3.

2 The Mechanism

We consider a generalised standard public goods game PG^s with n players. In PG^s, each player i, $i = 1, ..., n$, chooses a level of contribution to a public good, c_i, from an idiosyncratic set of possible contributions, i.e. $c_i \in [0, \bar{c}_i]$, where $\bar{c}_i > 0$ can be thought of as the socially desired contribution of player i.[1] Payoffs for each player i are given by

$$\pi_i^s(c) = e_i - a_i c_i + b_i \sum_j c_j \ ,$$

[1] We comment on the determinants of $\bar{c}_i$ at the end of this section.

where $c := (c_1, ..., c_n)$, e_i is player i's initial endowment and a_i (b_i) denotes player i's marginal cost (benefit) of contributing one further unit to the public good.[2] Hence, we explicitly allow for heterogeneity among the players. In order to guarantee that the players' payoffs reflect the common public goods structure, we assume:

Assumption A: For all i, it holds that $a_i > b_i > a_i \dfrac{\bar{c}_i}{\sum_j \bar{c}_j}$.

Proposition 1 is well known and straightforward to prove.

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

Yet, under Assumption A, c^0 is Pareto dominated by $\bar{c} = (\bar{c}_1, \ldots, \bar{c}_n)$:

$$\pi_i^s(\bar{c}) = e_i - a_i \bar{c}_i + b_i \sum_j \bar{c}_j > e_i = \pi_i^s(c^0) \quad \text{for all } i.$$

To implement full contribution, i.e. $\bar{c}$, consider the following 2-stage variation of PG^s.

Stage 1 All players simultaneously choose to pay a deposit $d_i \in \{0, \bar{d}_i\}$ to a neutral institution, e.g. a fund. At the end of stage 1, the profile of all deposits paid, denoted by d, is revealed to the players. Thereafter, the game enters stage 2.

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 in stage 2 the public goods game PG^s is played. If, however, at the end of stage 1 we have $d_i = \bar{d}_i$ for all i, then in stage 2 a public goods game PG^* is played for which payoffs are as follows:

$$\pi_i^*(c) = \begin{cases} e_i - \bar{d}_i - a_i c_i + b_i \sum_j c_j, & \text{if } \ c_i < \bar{c}_i \\ e_i - a_i c_i + b_i \sum_j c_j, & \text{if } \ c_i = \bar{c}_i. \end{cases}$$

[2] We do not necessarily assume monetary contributions. For example, c_i can be some measure of reducing carbon dioxide emissions, which has a marginal cost of a_i and a marginal benefit of b_i for player i.

Thus, if all players have decided to pay their deposit, these deposits are only refunded to those players who contribute to the public good in the previously agreed way, i.e. for whom $c_i = \bar{c}_i$.

The 2-stage game defined above is denoted by $\widehat{PG}$. A strategy for player i in $\widehat{PG}$ is given by a tuple $(d_i, \hat{c}_i)$, where

$$\hat{c}_i : \{0, \bar{d}_1\} \times \ldots \times \{0, \bar{d}_n\} \rightarrow [0, \bar{c}_i].$$

The following result is readily established:

Proposition 2 *Assume that $\bar{d}_i > (a_i - b_i)\bar{c}_i$ for all i. Then, the strategy profile $(\bar{d}, \hat{c}^*)$ where $\bar{d} = (\bar{d}_1 \ldots, \bar{d}_n)$ and*

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

is a strict subgame perfect Nash equilibrium of $\widehat{PG}$. Moreover, for all i, $\hat{c}_i^*$ is a strictly dominant strategy at stage 2 and $\bar{d}_i$ is a dominant strategy at stage 1, given that play continues with the strategy profile $\hat{c}^*$ at stage 2.

Proof. Consider stage 2 first. If $d_j = 0$ for some j, then PG^s 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. Then,

$$\begin{array}{rrcll} & \pi_i^*(\bar{c}_i, c_{-i}) & > & \pi_i^*(c_i, c_{-i}) & \text{for all } c_i < \bar{c}_i \\ \Longleftrightarrow & (b_i - a_i)\bar{c}_i & > & (b_i - a_i)c_i - \bar{d}_i & \text{for all } c_i < \bar{c}_i \\ \Longleftrightarrow & \bar{d}_i & > & (a_i - b_i)\bar{c}_i & \end{array}$$

which is satisfied by assumption. 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 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,

$$\pi_i^s(\hat{c}^*(\bar{d}_i, d_{-i})) = e_i = \pi_i^s(\hat{c}^*(0, d_{-i})).$$

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

$$\pi_i^*(\hat{c}^*(\bar{d}_i, d_{-i})) = e_i - a_i\bar{c}_i + b_i \sum_j \bar{c}_j > e_i = \pi_i^s(\hat{c}^*(0, d_{-i})),$$

which is satisfied by Assumption A. ■

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 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 not a dominant strategy for player i, given that play continues with $\hat{c}^*$. In fact, $(\bar{d}, \hat{c}^*)$ is the unique subgame perfect Nash equilibrium in stage-wise 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 variation 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^s is played except that now player i's payoff function is given by:

$$\tilde{\pi}_i^s(c) = \begin{cases} e_i - \varepsilon\bar{d}_i - a_i c_i + b_i \sum_j c_j, & \text{if } \ d_i = \bar{d}_i \\ \\ e_i - a_i c_i + b_i \sum_j c_j, & \text{if } \ d_i = 0. \end{cases}$$

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:

$$\tilde{\pi}_i^*(c) = \begin{cases} e_i - \bar{d}_i - a_i c_i + b_i \sum_j c_j, & \text{if } \ c_i < \bar{c}_i \\ e_i - \varepsilon \bar{d}_i - a_i c_i + b_i \sum_j c_j, & \text{if } \ c_i = \bar{c}_i. \end{cases}$$

The next proposition shows that 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 $\widetilde{PG}$. The proof of Proposition 3 is immediate.

Proposition 3 *Let $\varepsilon > 0$ be such that $\bar{d}_i(1-\varepsilon) > (a_i - b_i)\bar{c}_i$ and $b_i \sum_j \bar{c}_j > a_i \bar{c}_i + \varepsilon \bar{d}_i$ for all i. Then there are exactly two subgame perfect Nash equilibria $(d^*, \hat{c}^*)$ of $\widetilde{PG}$. 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.*

To conclude our analysis, we finally discuss the determinants of the share $\bar{c}_i$ player i can be expected to contribute to the public good. In order to do so, let us interpret player i's endowment, e_i, as the maximum wealth player i is generally able/willing to invest into the provision of the public good, and let us assume that this is undisputed among players. Then, for feasibility reasons, it must be true that $\bar{d}_i + a_i \bar{c}_i \leq e_i$. As we have seen, the implementation of full contribution to the public good requires that $\bar{d}_i \geq (a_i - b_i)\bar{c}_i$.[3] Hence, $\bar{c}_i$ is bounded above by $e_i/(2a_i - b_i)$. Thus, ceteris paribus, the larger player i's initial endowment e_i, the larger her marginal benefit b_i derived from the public good, and the lower the marginal cost a_i, the larger is the contribution $\bar{c}_i$ player i can be asked to contribute to the public good. Yet, maximal contributions $\bar{c}_i = e_i/(2a_i - b_i)$ for all i are only feasible if Assumption A is satisfied, i.e. if

$$b_i \sum_j \frac{e_j}{2a_j - b_j} > a_i \frac{e_i}{2a_i - b_i} \text{ for all } i.$$

Accordingly, if plans are too ambitious regarding the targeted level of contributions to the public good, its provision may fail simply because of the players' budget constraints. This seems to be particularly relevant if we consider,

[3]For the sake of simplicity we neglect any costs for administering deposits.

for example, the desired effort levels to reduce the output of green-house gases. However, if the technology for producing the public good allows for a stepwise production process, then the mechanism proposed in this paper still can be used effectively since the target level of the public good can be achieved by a series of incremental increases. Contributions to the public good could, for example, be evaluated on an annual (monthly, ...) basis. If all contributions are found to be made as expected, existing deposits are paid back, so that again enough funds are available for the implementation of the next incremental increase in the public good.

3 Discussion

We have proposed a simple two-stage mechanism which implements full contributions to the public good in a subgame perfect Nash equilibrium. Similar to a smoker who publicly announces to refrain from smoking in order to make failure prohibitively costly (or a co-author who freely promises to deliver a revised version of the paper by the end of the week),[4] players in our mechanism can pay an ex ante deposit which, if paid by all, renders contributing to the public good a dominant strategy. This mechanism has favourable properties which we discuss in the following.

First, our mechanism implements the provision of public goods in a strict subgame perfect Nash equilibrium. If the cost for running the institution that administers deposits is negligible, the implementation can even be achieved in a stage-wise dominant strategy equilibrium. Hence, even though there exist zero-contribution equilibria as well, the coordination problem is less severe than in case of equilibria that are non-strict or not in dominant strategies.

Second, and most importantly, our mechanism does not require the presence or establishment of powerful institutions to implement full contributions to the public good (cf. Falkinger et al., 2000; Fehr and Gächter, 2000; Gürek et al., 2006; Kosfeld et al., 2006). The reason is that no ex post punishment of free-riders is required. All that is needed is an independent institution (e.g. the world bank)

[4]Whether announcements are strong enough to later enforce the desired action may be questionable. Their frequent use, however, indicates that the targeted commitment effect indeed is very similar to the one used in our mechanism.

that collects deposits, monitors the players' contributions and refunds deposits to those who have contributed to the public good. Hence, our mechanism is particularly suited to implement international agreements like the Kyoto Protocol, where there is no common authority that can enforce the punishment of free-riders.

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. Hence, societies who in principle are willing to provide a particular public good (e.g. clean air) can be expected to agree to the implementation of a self-sanctioning scheme (the deposit) which helps them to adhere to their intended contribution ex post. In fact, the mechanism proposed here is indeed quite similar to a self-sanctioning mechanism. Players themselves submit the deposit and, hence, can also easily be assigned the burden of proof that they contributed to the public good. Consequently, the institution in our case essentially has to administer the deposits, while its monitoring role is rather weak. This is different for a punishing institution, which — in line with conventional legal systems — has to prove that a free-rider did not contribute to the public good.

References

Dutta, P., and R. Radner (2004): "Self-enforcing Climate Change Treaties," *Proceedings of the National Academy of Sciences*, 101, 5174–5179.

Falkinger, J., E. Fehr, S. Gächter, and R. Winter-Ebmer (2000): "A Simple Mechanism for the Efficient Provision of Public Goods: Experimental Evidence," *American Economic Review*, 90, 247–264.

Fehr, E., and S. Gächter (2000): "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90, 980–994.

Gürek, O., B. Irlenbusch, and B. Rockenbach (2006): "The Competitive Advantage of Sanctioning Institutions," *Science*, 312, 108–111.

Kosfeld, M., A. Okada, and A. Riedl (2006): "Institution Formation in Public Goods Games," mimeo, Maastricht University.