Public Good Provision in a Large Economy

Felix J. Bierbrauer and Martin F. Hellwig

Max Planck Institute for Research on Collective Goods, Bonn

February 2009

Abstract

We propose a new approach to the normative analysis of public-good provision in an economy that is large so that any one individual is too insignificant to have a noticeable effect on the provision levels of public goods. In such an economy, the standard mechanism design problem of calibrating people's payments to the influence they have on public-good provision is moot. In the absence of participation constraints, the first-best provision rule of providing the public good if and only if the average per capita valuation exceeds the per capita cost can be implemented if the costs are shared equally among individuals. Equal cost sharing is actually necessary if the mechanism is to be robust in the sense of Bergemann and Morris (2005). However, the first-best provision rule with equal cost sharing is vulnerable to collective deviations in the sense of Laffont and Martimort (2000). Thus, people with valuations below the per capita provision cost would all benefit from a collective deviation inducing a downward bias into the assessment of the average per capita valuation. We develop a concept of coalition-proofness and show that a coalition-proof and robust mechanism cannot condition on the average per capita valuation, but only on the population shares of people with valuations above and below the per capita provision costs. The result suggests an intriguing link between mechanism design theory for large economies and voting.

Keywords: Public Good Provision, Mechanism Design, Large Economy

JEL: D60, D70, D82, H41

1 Introduction

In this paper, we propose a new approach to the normative analysis of public-good provision in large economies. By a large economy, we understand an economy with many people in which each individual is too insignificant to have a noticeable effect on variables such as the prices of private goods or the provision levels of public goods. We consider the large-economy paradigm to be appropriate for studying how a society with millions of people can best determine the appropriate levels of resources that are to be devoted to matters such as national defense or the court system, which concern the entire population. We also believe that, when applied to a large economy, the standard mechanism design approach to public-good provision provides unsatisfactory results.

The standard mechanism design approach to public-good provision focusses on issues of *individual incentive compatibility*. Under asymmetric information about individual preferences, the question is whether individuals have proper incentives to provide the system as whole with the information about preferences that it needs for efficient public-good provision. In a "small" economy, in which each individual has a distinct chance of being "pivotal", i.e. of having a noticeable effect on the provision of a public good, this requires that people's financial contributions must be precisely calibrated to their expressions of preferences. The calibration must be such that people have neither an incentive to overstate their preferences on the assumption that the increase in public-good provision is paid for by somebody else nor an incentive to understate their preferences on the assumption that the money which they can thereby save is worth more than the reduction in public-good provision. The implications of this requirement have been thoroughly explored in the literature. It is well known that an efficient provision rule can be implemented if the calibration of payments to expressed preferences is such that, in terms of expectations, at least, people are induced to take account of the external effects that they impose on others whenever they are "pivotal" for the provision or non-provision of the public good.[1]

In a large economy, these concerns are moot. In such an economy, any notion that a person's payments should be calibrated to the effects that this person's communication about her preferences have on the provision of the public good leads to the simple conclusion that their payments should be independent of what they communicate. If what they say affects neither the collective decision on public-good provision nor the payments they have to make, individual incentive compatibility is trivial. If what they say is deemed to have no effect whatsoever, they may as well tell the truth. According to the standard mechanism design approach, the aggregate of the information that is thus communicated can serve as a basis for implementing an efficient provision rule for the public good. Participation in the system may not be voluntary, but there is no problem of incentive compatibility.[2]

[1] For implementation in dominant strategies, see Clarke (1971), Groves (1973), Green and Laffont (1979), for (interim) Bayes-Nash implementation, see d'Aspremont and Gérard-Varet (1979). More recently, Bergemann and Morris (2005) have studied interim implementation with a requirement of robustness with respect to the specification of agents' beliefs about the other participants.

[2] By contrast to the literature on public-good provision through multilateral bargaining, we do not insist on voluntary participation. Participation constraints are irrelevant if the state has powers of coercion and these powers can be used to make people contribute to financing a public good even when it does not benefit them.

We want to take issue with this view. We explain our concerns by a simple example. Suppose that the public good in question comes as a single indivisible unit. The provision cost *per capita* of the population is 4. A fraction $\frac{3}{10}$ of the population assigns a value of 10 to the public good, a fraction s a value of 3, and a fraction $\frac{7}{10} - s$ a value of 0. An efficient provision rule stipulates that the public good should be provided if the average *per capita* valuation exceeds 4, and that it should not be provided if the average *per capita* valuation is less than 4. In other words, the public good should be provided if $s > \frac{1}{3}$ and should not be provided if $s < \frac{1}{3}$. The requisite resources can be obtained by imposing a payment rule under which everybody pays 4 if the public good is provided and 0 if it is not provided. If people believe that, individually, they are too insignificant to affect the provision of the public good, a mechanism involving this provision and payment rule is incentive compatible.

If s is common knowledge, this reasoning is unproblematic. This is the case, for instance, if we think of the large-economy model as a limit of finite-economy models with independent private values in which the number of participants becomes large. However, if s is common knowledge, the implementation of an efficient provision rule does not require any information from participants because, even before any such information is provided, it is commonly known whether the public good should be provided or not.[3]

By contrast, if s is the realization of a nondegenerate random variable $\tilde{s}$, the problem of whether the public good should be provided or not involves a genuine information problem. In this case, the information whether the public good should be provided or not must be inferred from the participants' reports about their preferences. If the fraction of people reporting a valuation of 3 exceeds $\frac{1}{3}$, one may infer that $s > \frac{1}{3}$ and that the public good should be provided.

At this point, we have a problem with the notion that efficient provision can be implemented with a payment rule under which everybody pays 4 if the public good is provided and 0 if it is not provided. Why should people with a valuation of 3 report this valuation honestly? Reporting a valuation of 3 contributes to making provision of the public good more likely, if only infinitesimally. If the public good is provided, these people enjoy a benefit of 3 and have to pay 4, for a net payoff equal to -1. Each one of them would be better off if the public good was not provided. Moreover, the public good would indeed not be provided if each one of these people reported a valuation of 0. Why then should they report honestly, rather than claiming that the public good is worth nothing to them?

If individual incentive compatibility is the only requirement for the public-good provision mechanism, the answer to this question is that nobody minds reporting his or her valuation honestly because nobody feels that his or her report will make a difference to anything anyway. We consider this answer to be unconvincing. Therefore, we propose a new approach to the analysis of public-good provision in a large economy.

This new approach involves a requirement of *coalition proofness* as well as individual incentive compatibility. In the given example, the people who value the public good at either 0 or 3 have an incentive to sabotage the efficient provision rule with equal cost sharing by forming a coalition

[3]Thus, in models with independent private values, the problem of whether to provide the public good or not becomes moot if one takes limits as the number of participants becomes large and the law of large numbers sets in.

to coordinate reports in such a way that the fraction of people reporting 3 is always below $\frac{1}{3}$. By contrast to cartel formation in industrial economics, the distorted reports that this sabotage action requires would all be individually incentive-compatible.

Given this requirement of coalition proofness, in the given example, with equal sharing of public-good provision costs, it is impossible to condition public-good provision on s. More generally, we shall find that public-good provision can be conditioned on the sizes of the set of people who are net beneficiaries of public-good provision and of its complement, the set of people who are harmed by public-good provision, but not on any additional information, e.g., information about the intensities of people's likes and dislikes. In the example, the two sets of people have sizes $\frac{7}{10}$ and $\frac{3}{10}$, regardless of s, and the mechanism designer is reduced to a rule that stipulates public-good provision or not, depending on whether the *ex ante* expectation of people's valuation of the public good is greater or less than the *per capita* cost 4, or, equivalently, whether he considers the *ex ante* expectation of $\tilde{s}$ to be greater or less than $\frac{1}{3}$.

A requirement of coalition proofness has previously been introduced by Laffont and Martimort (1997, 2000). Our approach differs from theirs in that we focus on coalitions consisting of subsets of the entire population, with coalition membership depending on people's types. Thus, in the above example, we considered a coalition of all people who value the public good at 0 or 3. By contrast, Laffont and Martimort focussed on coalitions of all people, regardless of their types. This focus was appropriate for their purpose, which was to eliminate the possibility, established by Crémer and McLean (1985, 1988), that the mechanism designer might exploit the slightest correlations in individual preferences in order to appropriate the entire surplus that is generated.

In our analysis, coalition proofness is not used to prevent the mechanism designer from appropriating rents, but as a device to articulate the inherent conflict between people who benefit from public-good provision and people who are harmed by it. Therefore, we focus on coalitions of subsets of people with common interests. If common interests are the basis of coalition membership, it is natural to have coalition membership depend on people's types. Thus, in the given example, the common interests of people who are harmed by public-good provision are put into focus by a concept of coalition proofness that allows for collective manipulations of individual reports by the coalition of people who value the public good at 0 or at 3.

However, we do follow Laffont and Martimort in requiring that coalition formation and the behaviour of coalition members satisfy the same information and incentive constraints as the underlying incentive mechanism itself. In particular, we require that the decision to join a coalition and the behaviour as a coalition member must be individually incentive-compatible. The information problems of coalition formation and behaviour are actually more complex in our setting than in Laffont and Martimort (1997, 2000) because, apart from problems of individual incentive compatibility of stipulated behaviours of coalition members, a coalition that consists of a subset of the population also must deal with the problem that its information about people outside the coalition is incomplete.

In addition to imposing coalition proofness, we require that individual incentive compatibility be *robust* in the sense of Bergemann and Morris (2005). Outcomes are allowed to depend *only* on those aspects of the participants' types that are relevant for their payoffs, i.e., their public-goods

preferences. They are *not* allowed to depend on other aspects of the participants' types such as the beliefs that they have about other people's payoffs or other people's beliefs. Moreover, the outcome function must be individually incentive-compatible regardless of how the non-payoff-relevant aspects of people's types are specified.

In the context of a large economy, robustness implies that people's payments cannot be made to depend on their types. This is in line with the notion that payments should be calibrated to the effects that this person's communication about her preferences have on the provision of the public good, which in a large economy are zero. Deviations from this principle could be incentive-compatible, if, conditional on their types, people have different beliefs about the state of the economy and the prospects for public-good provision and a type dependence of payments allows them to bet on the differences in beliefs. However, the incentive-compatibility compatibility of such deviations is not robust to changes in the specification of beliefs.

To explain the issue, we return to the above example and consider a type-dependent payment rule that requires people who value the public good at 3 to pay 0 if the public good is provided and to pay 8 if the public good is not provided. People who value the public good at 0 or 10 pay 10 if the public good is provided and receive 2 if the public good is not provided. Under this rule, people who value the public good at 3 are no longer averse to having revealed that s is greater and not less than $\frac{1}{3}$. When the public good is provided, their net payoff is equal to 3, when the public good is not provided, their net payoff is equal to -8.

If the random variable $\tilde{s}$ can only take the values $\frac{2}{10}$ and $\frac{6}{10}$, the combination of this type-dependent payment rule with an efficient provision rule, providing for non-provision if $\tilde{s} = \frac{2}{10}$ and for provision if $\tilde{s} = \frac{6}{10}$, is also compatible with budget balance.

The resulting mechanism is incentive-compatible if type-dependent beliefs are derived from a common prior that assigns probability one half to each of the two possible realizations of $\tilde{s}$ and that assigns values $\frac{7}{10} - s, s$, and $\frac{3}{10}$ to any one person's conditional probabilities, given the event $\tilde{s} = s$, of having valuations 0, 3 and 10. Given this common prior, the probability of public-good provision, i.e., of the event $\tilde{s} = \frac{6}{10}$, is assessed at $\frac{1}{6}$ by a person with valuation 0, at $\frac{3}{4}$by a person with valuation 3, and at $\frac{1}{2}$by a person with valuation 10. These differences in beliefs allow for an incentive-compatible dependence of payments on types. The resulting payments scheme can be interpreted as a combination of sharing of the cost of efficient public-good provision and a system of bets on the state of the economy.[4]

However, if the common prior were to assign probabilities one third to the event $\tilde{s} = \frac{2}{10}$ and two thirds to the event $\tilde{s} = \frac{6}{10}$, the given scheme would no longer be incentive compatible. With beliefs determined by the prior $(\frac{1}{3}, \frac{2}{3})$, the people who value the public good at 10 would consider the payment scheme that is meant for people with valuation 3 to be more attractive than the payment scheme that is meant for themselves. The incentive compatibility of the given type-dependent payment scheme is thus not robust to changes in the specification of beliefs.

[4]The introduction of the system of bets also has the effect of shifting the expected payoff that a person with valuation 0 receives from public-good provision from $\frac{1}{6}(-4) = -\frac{2}{3}$ to $\frac{1}{6}(-10) + \frac{5}{6}2 = 0$. For a person with valuation 3, the expected payoff is shifted from $-\frac{3}{4}$ to $\frac{1}{4}$, for a person with valuation 10, from 3 to 1. By the type-contingent system of bets, the people who value the public good are made to "share" the benefits with the result that the other people's expected payoffs from the system become nonnegative.

In combination with a requirement of anonymity, robustness implies that the costs of public-good provision must be shared equally. This provides for a clear distinction between between people whose net payoffs are increased by the provision of the public good and people whose net payoffs are lowered by the provision of the public goods. Both groups provide natural candidates for assessing coalition proofness of a provision rule for the public good.

The main result of our analysis shows that, if one imposes coalition proofness, as well as robust individual incentive compatibility and anonymity, then the sizes of the two groups, the group of people who are harmed by public-good provision and its mirror image, the group of people who benefit from public-good provision, represent the *only* information that can be used in determining whether the public good is to be provided or not. By contrast, information concerning the intensity of likes and dislikes cannot be used. Apart from exceptional circumstances, therefore, it is impossible to implement a first-best provision rule by a coalition proof, robustly incentive-compatible anonymous mechanism. By contrast to previous impossibility results, this finding does not involve any participation constraints or a multi-dimensional information problem. Instead, it follows from the observation that coalition proofness and robust incentive compatibility together destroy the possibility of conditioning on intensities of preferences.

Mechanisms that condition the provision of the public good on the numbers of its adherents and its opponents are reminiscent of voting mechanisms. In our analysis, optimal coalition proof, robustly incentive-compatible, anonymous mechanisms differ from traditional voting mechanisms in that the decision to provide the public good or not is based on an assessment of expected benefits and costs conditional on numbers of adherents and opponents, rather than any (qualified) majority rule. Even so, we find it intriguing that, once coalition proofness is imposed in addition to robust incentive compatibility and anonymity, the mechanisms that we are concerned with involve numbers of votes, for and against, rather than any attempt to measure willingness to pay. Implementation can be done by a show of hands, rather than any more complicated procedure.

The remainder is organized as follows. Section 2 contains the description of the environment and the characterization of robust public good mechanisms in a large economy. In Section 3 we define the notion of a collective manipulation mechanism and of a coalition-proof rule for public good provision. Section 4 characterizes the provision rules that are both robust and coalition-proof. Finally, in Section 5 we solve for the optimal provision rule for public goods. All proofs are in the Appendix.

2 Robust Implementation in a Large Economy

2.1 Payoffs, Allocations, and Social Choice Functions

We consider an economy with an atomless measure space of agents $(A, \mathcal{A}, \lambda)$, where λ is normalized so that $\lambda(A) = 1$. There are one private good and one public good. The public good comes as a single indivisible unit. Its installation requires aggregate resources equal to k units of the private good.

Given a public-good provision level $Q \in \{0, 1\}$ and a required contribution of p units of the

private good, an agent with valuation v obtains the payoff $vQ - p$, where v belongs to the set V of possible valuations of the public good.

A profile of public goods preferences is a function $\mathbf{v} : A \to V$. We require the map $\mathbf{v}$ to be measurable, so that the cross-section distribution of valuations, $\lambda \circ \mathbf{v}^{-1}$, is well defined. It belongs to the set $\mathcal{M}(V)$ of probability measures on V.

An *allocation* for this economy consists of a public-good provision level $Q \in \{0, 1\}$ and a measurable function $\mathbf{p}$ from A into $\mathbb{R}$ such that, for each $a \in A$, $\mathbf{p}(a)$ is the number of units of the private good that agent a has to contribute.

A *social choice function* is a map F from preference profiles to feasible allocations. It consists of a provision rule $Q_F : \mathbf{v} \mapsto Q_F(\mathbf{v})$, that specifies for each preference profile whether or not the public good is provided and a payment rule $p_F : (a, \mathbf{v}) \mapsto p_F(a, \mathbf{v})$ that specifies a payment for each agent a and each preference profile $\mathbf{v}$.

Following Guesnerie (1995), we impose a requirement of *anonymity*. A social choice function $F = (Q_F, p_F)$ is said to be *anonymous* if the following two requirements are met: (i) *Recipient anonymity*: For every preference profile $\mathbf{v}$ and every pair of agents a and a' with $\mathbf{v}(a) = \mathbf{v}(a')$, $p_F(a, \mathbf{v}) = p_F(a', \mathbf{v})$. (ii) *Anonymity in influence*: For any pair of preference profiles $\mathbf{v}$ and $\mathbf{v}'$, $\lambda \circ \mathbf{v}^{-1} = \lambda \circ \mathbf{v}'^{-1}$ implies $Q_F(\mathbf{v}) = Q_F(\mathbf{v}')$ and, $p_F(a, \mathbf{v}) = p_F(a', \mathbf{v}')$, for every every pair of agents a and a' with $\mathbf{v}(a) = \mathbf{v}'(a')$. By standard measurability theory, e.g., Result (8), p. 43, in Hildenbrand (1974), one obtains:

Lemma 1 *A social choice function $F = (Q_F, p_F)$ is anonymous if and only if there exist functions $\hat{Q}_F : \mathcal{M}(V) \to \{0, 1\}$, and $\hat{p}_F : V \times \mathcal{M}(V) \to \mathbb{R}$ such that for all $\mathbf{v}$ with $\lambda \circ \mathbf{v}^{-1} = s$, $Q_F(\mathbf{v}) = \hat{Q}_F(s)$ and $p_F(a, \mathbf{v}) = \hat{p}_F(v, s)$, for all a with $\mathbf{v}(a) = a$.*

With a slight abuse of notation, for anonymous social choice functions, we also write $F = (\hat{Q}_F, \hat{p}_F)$, where $(\hat{Q}_F, \hat{p}_F)$ is the pair given by Lemma 1. Under an anonymous social choice function, the public-good provision level Q depends only on the cross-section distribution $\lambda \circ \mathbf{v}^{-1}$ of preference parameters, and the payment of agent a depends only on the distribution $\lambda \circ \mathbf{v}^{-1}$ and on $\mathbf{v}(a)$. The preference of agent a does not affect the public-good provision level or the payment of any agent a' other than a.

2.2 Types and Beliefs

Following Bergemann and Morris (2005), we model information by means of an abstract type space $\mathfrak{T} = [(T, \mathcal{T}), \mathbf{t}, \tau, \beta]$, where $(T, \mathcal{T})$ is a measurable space, $\mathbf{t}$ is a measurable map from A into T, τ is a measurable map from T into V, and β is a measurable map from T into the space $\mathcal{M}(\mathcal{M}(T))$ of probability distributions over measures on T. We interpret $t(a)$ as the abstract "type" of agent a, $\tau(t)$ as the preference parameter of an agent with abstract type t, and $\beta(t)$ as the "belief" of an agent with abstract type t. Given the mappings $\mathbf{t}$ and τ, we refer to $\mathbf{v}(a) = \tau(\mathbf{t}(a))$ as the *payoff type* and to $\mathbf{b}(a) = \beta(\mathbf{t}(a))$ as the *belief type* of agent a.

The belief type $b(a)$ indicates the agent's beliefs about the other agents. We specify these beliefs in terms of cross-section distribution of types in the economy. Thus, $\mathbf{b}(a)$ is a probability

measure on the space $\mathcal{M}(T)$ of these cross-section distributions.

We think of the cross-section distribution of types as the realization δ of a random variable $\tilde{\delta}$ and of the type $\mathbf{t}(a)$ of any agent a as the realization of a random variable $\tilde{t}$.

Type space $\mathfrak{T}$ is a *common prior type space* if the beliefs of any agent a can be identified with the marginal distribution of $\tilde{\delta}$ that is obtained after conditioning on the agent's type. Formally, $\mathfrak{T}$ is a *common prior type space* if there is a measure P on $T \times \mathcal{M}(T)$ such that for every $X \subset \mathcal{M}(\mathcal{M}(T))$ and every $t \in T$:

$$\beta(t)[X] = P(\tilde{\delta} \in X \mid t) \; .$$

The law of large numbers holds, if conditional on the event $\tilde{\delta} = \delta$, $\tilde{t}$ has the probability distribution δ. Formally, there is a measure P such that for very $B \subset T$ and every $\delta \in \mathcal{M}(T)$:

$$\mathrm{pr}(\tilde{t} \in B \mid \tilde{\delta} = \delta) = \delta(B) \; .$$

Hence, we can identify the conditional probability distribution of $\tilde{t}$ given the event $\tilde{\delta} = \delta$ with the cross-section distribution δ itself.

2.3 Incentive Compatibility and Robust Implementation

Information about types is assumed to be private. A social choice function is interim implementable on a given type space if, for this type space, there exists a mechanism, specifying a message set for each agent and a function from message profiles to allocations, and there exists an equilibrium of the strategic game induced by the mechanism such that the equilibrium allocation is equal to allocation stipulated by the social choice function.

Given the restriction to anonymous social choice functions, we also restrict attention to anonymous mechanisms. An anonymous mechanism consists of a set of reports R and a function f_R that maps distributions of reports $\rho \in \mathcal{M}(R)$ into allocations. It can be represented by a provision rule $\hat{Q}_{f_R} : \rho \mapsto \hat{Q}_{f_R}(\rho)$ that specifies for each cross-section distribution of reports whether the public good is provided and a payment rule $\hat{p}_{f_R} : (t, \delta) \mapsto \hat{p}_f(t, \delta)$ that specifies an agent's payment as a function of his report $r \in R$ and the distribution of reports ρ.

A strategy $\sigma^* : T \to R$ is said to be an interim Nash equilibrium for the game induced by f_R on a given type space $\mathfrak{T}$, if for all t and all r,

$$\int_{\mathcal{M}(T)} [\tau(t)\hat{Q}_{f_R}(\delta \circ \sigma^{*-1}) - \hat{p}_{f_R}(\sigma^*(t), \delta \circ \sigma^{*-1})] d\beta(t) \geq$$
$$\int_{\mathcal{M}(T)} [\tau(t)\hat{Q}_{f_R}(\delta \circ \sigma^{*-1}) - \hat{p}_{f_R}(r, \delta \circ \sigma^{*-1})] d\beta(t) \; .$$

An anonymous mechanism f_R is said to *achieve the social choice function* F if

$$f_R(\delta \circ \sigma^{*-1}) = F(\delta \circ \tau^{-1}) \tag{1}$$

for all $\delta \in \mathcal{M}(T)$. The allocation that is induced by the mechanism f_R for a given cross-section distribution δ of abstract types must coincide with the allocation that the social function F stipulates for the corresponding distribution of payoff types $\delta \circ \tau^{-1}$.

The mechanism f_R is said to implement *the social choice function F on the type space $\mathfrak{T}$* if it has an equilibrium σ^* that achieves F.

An anonymous social choice function F is said to be *robustly implementable* if, for every type space $\mathfrak{T}$, there exists an anonymous mechanism f_R that implements F on $\mathfrak{T}$.

Proposition 1 *An anonymous social choice function $F = (\hat{Q}_F, \hat{p}_F)$ is robustly implementable if and only if it satisfies the following ex post incentive compatibility constraints: For all v and and v' in V and all $s \in \mathcal{M}(V)$,*

$$v\hat{Q}_F(s) - \hat{p}_F(v,s) \geq v\hat{Q}_F(s) - \hat{p}_F(v',s) \; . \tag{2}$$

By inspection of (2), in our setting, *ex post* implementability is equivalent to the requirement that $\hat{p}_F(v,s) = \hat{p}_F(v',s)$ for all v, v' and s. If the payment of some agent was, for some s, smaller than the payment of some other agent, the latter would like to imitate the agent with the small payment. This would contradict ex post implementability. This observation yields the following corollary to Proposition 1.

Corollary 1 *An anonymous social choice function $F = (\hat{Q}_F, \hat{p}_F)$ is robustly implementable if and only if payments are independent of individual payoff types, i.e., there is a function $\bar{p}_F : \mathcal{M}(V) \to \mathbb{R}$ such that $\hat{p}_F$ takes the form $\hat{p}_F(v,s) = \bar{p}_F(s)$ for all $v \in \Theta$ and all $s \in \mathcal{M}(V)$.*

Given Corollary 1, we will represent a robustly implementable social choice function in the following as a pair of functions $(\hat{Q}_F, \bar{p}_F)$, where $\bar{p}_F(s)$ is the lump sum contribution to the cost of public good provision if the cross-section distribution of payoff types equals $s \in \mathcal{M}(V)$.

2.4 Robust Implementation of First-Best Allocations

Given an economy with agent space $(A, \mathcal{A}, \lambda)$ and preference profile $\mathbf{v}$, we say that an allocation is *feasible* if aggregate payments cover the costs of public-good provision, i.e., if $\int_A p(a)d\lambda(a) \geq kQ$. An allocation is said to be *first-best* if it maximizes the aggregate surplus

$$\int_A [\mathbf{v}(a)Q - p(a)]d\lambda(a)$$

over the set of feasible allocations. A social choice function is said to *yield first-best outcomes* if, for every preference profile $\mathbf{v}$ the allocation $(Q_F(\mathbf{v}), \{p_F(a,\mathbf{v})\}_{a\in A})$ is first-best. Trivially, one obtains:

Lemma 2 *A social choice function $F = (Q_F, p_F)$ yields first-best outcomes if and only if, for any $\mathbf{v}$,*

$$\textstyle\int_A p(a,\mathbf{v})d\lambda = kQ_F(\mathbf{v}) \; , \quad \textit{where} \quad Q_F(\mathbf{v}) = \begin{cases} 0, & \textit{if} \quad \int_A \mathbf{v}(a)d\lambda < k, \\ 1, & \textit{if} \quad \int_A \mathbf{v}(a)d\lambda > k. \end{cases}$$

An anonymous social choice function $F = (\hat{Q}_F, \hat{p}_F)$ yields first-best outcomes if and only if, for any $s \in \mathcal{M}(V)$,

$$\int_V \hat{p}_F(v,s)ds = k\hat{Q}_F(s)\ , \quad \textit{where} \quad \hat{Q}_F(s) \begin{cases} 0, & \textit{if} \quad \bar{\mathbf{v}}(s) < k, \\ 1, & \textit{if} \quad \bar{\mathbf{v}}(s) > k, \end{cases}$$

where $\bar{\mathbf{v}}(s) := \int_V v\ ds$.

The public good should be provided if the aggregate valuation $\bar{\mathbf{v}}(s)$ exceeds the cost k and should not be provided if the aggregate valuation is less than k. Aggregate payments should exactly cover the cost of public-good provision, i.e., there should be no slack in the feasibility constraint. Upon combining these conditions with Corollary 1, we obtain:

Proposition 2 *An anonymous social choice function $F = (\hat{Q}_F, \hat{p}_F)$ yields first-best outcomes and is robustly implementable if and only if, for all $s \in \mathcal{M}(V)$ and all $v \in V$, $\hat{p}_F(v,s) = k\hat{Q}_F(s)$ where $\hat{Q}_F(s)$ is zero or one depending on whether $\bar{\mathbf{v}}(s)$ is less than k or greater than k.*

Proposition 2 provides a general possibility result for first-best implementation in a large economy. Required contributions so that the cost of public-good provision are equally shared; this ensures feasibility (budget balance), as well as robust implementability. Because people never see themselves as having any influence on public-good provision and because people's payments do not depend on their types, each individual is indifferent as to what message he or she sends to "the system". Given this indifference, one may as well tell the truth.

Robust implementation of first-best allocations is *not* compatible with the imposition of interim participation constraints. Under equal cost sharing, anybody with a payoff type below k would wish to veto the the social choice function if he could: If the public good is provided, his payoff is negative because he has to pay more than the public good is worth to him; if the public good is not provided, his payoff is zero. On average, therefore, he loses from this regime.

This observation is in line with the findings of the literature on Bayesian mechanisms with independent private values.[5] It stands in contrast to the literature on Bayesian mechanisms with correlated values, according to which, in generic models with finitely many types, differences in beliefs can be used to provide for interim individually rational, incentive-compatible implementation of first-best allocations.[6] The contrast is due to the robustness requirement, which, in particular, requires that payments schemes to be independent of the beliefs that agents may have about the other agents in the economy.[7]

However, in this paper, we are not concerned about the violation of participation constraints. Participation constraints matter *only* if one adheres to the contractarian view of government

[5] See Güth and Hellwig (1986), Rob (1989), Mailath and Postlewaite (1990), Schmitz (1997), Hellwig (2003) and Norman (2004).

[6] Johnson, Pratt, and Zeckhauser (1990), d'Aspremont, Crémer, and Gérard-Varet (1990, 2004).

[7] For an example see the introduction. The payment scheme considered there is not incentive compatible if the underlying prior is modified. The fact that different belief specifications can negate the possibility theorems of Johnson, Pratt, Zeckhauser (1990) and d'Aspremont, Crémer, Gérard-Varet (1990, 2004) has previously been observed by Neeman (2004), as well as Bergemann and Morris (2005).

and the state that underlay Lindahl's (1919) original treatment of public goods. If one takes the state's power of coercion for granted and has no qualms about the state's using this power to implement first-best allocations, Proposition 2 suggests that the implementation of first-best allocations in large economies faces no fundamental difficulties.

As was explained in the introduction to this paper, we do not share this rather sanguine view. Even when abstracting from the problem of coercion, we believe that Proposition 2 does *not* provide a satisfactory basis for the normative theory of public-good provision in a large economy. As we have explained in the introduction, we consider the requirements of robust implementation to be too weak to do full justice to the information and incentive problems of public-good provision in a large economy. Therefore we now turn to a discussion and analysis of coalition proofness as an additional restriction on social choice functions and incentive mechanisms.

3 Coalition-Proof Implementation

For first-best outcomes to be obtained, "the system" must be able to ascertain the aggregate public-good valuation $\bar{\mathbf{v}}(s)$. The example in the introduction shows that some of the people who are providing this information may be effectively hurt by the use to which the information is put. In such a case, incentive compatibility holds *only* because any person alone is unable to affect the social outcome and is therefore indifferent about the message that he or she transmits to the mechanism that is to implement the social choice function.

However, people with similar valuations have similar interests. Collectively, they could upset the functioning of the mechanism. Therefore, they would seem to have an incentive to form a coalition in order to collectively manipulate the social outcome. In the following, we formalize this idea and study its implications. Of course, we require that, whatever collective manipulation they might be part of, people's behaviours must be individually incentive-compatible. Our treatment is inspired by Laffont and Martimort (1997, 2000), but there are also some important differences.

3.1 Collective Manipulation Mechanisms

Let $\mathfrak{T}$ be a given type space, and let $f_R = (\hat{Q}_{f_R}, \hat{p}_{f_R})$ be an anonymous mechanism. We consider the possibility that this mechanism is manipulated by a coalition of people with specified types, who collectively deviate from truth-telling. We think of this coalition as being operated by a coalition manager who announces a collective manipulation mechanism and asks people to join in order to manipulate messages to the overall mechanism. Conditional on a profile of messages that he receives from coalition members, the coalition manger will choose a profile of lies that coalition members should transmit to the overall mechanism.

We formalize a manipulation mechanism such that the coalition manager chooses the messages sent by the individuals who agree to join the manipulation mechanism.[8] We consider the following sequential structure:

[8] It will become clear below that the assumption that the coalition manager makes choices for individuals can be replaced by the weaker assumption that he makes recommendations and that individuals choose whether or not to follow them; see footnote 13 below.

1. The overall mechanism f_R is announced.

2. A coalition organizer may propose a manipulation mechanism. This is publicly observable.

3. Individuals choose whether to report directly to the mechanism designer, or let their report be chosen by the coalition organizer. The coalition organizer does not observe the reports that are sent directly to the overall mechanism. Likewise, the overall mechanism designer does not observe the communication between individuals and the coalition organizer.

4. The overall mechanism receives a profile of reports, one for each individual, and the corresponding allocation is implemented.

We assume that the coalition manger can not use side payments to facilitate coalition formation. In a large economy, this is without loss of generality. The reason is that individuals will participate only if this involves no personal cost. This implies, in particular, that they are not willing to make payments to the coalition organizer.[9]

The coalition manager uses a direct mechanism, which works as follows: he asks all individuals to announce a message from the set $T^e = T \cup \{\emptyset\}$. If an agent sends $\emptyset$ this indicates that he does not participate in the manipulation mechanism and that the coalition manager does not choose a report to the overall mechanism for this agent. Agents who participate announce some type t to the coalition manager.[10]

We limit attention to anonymous manipulation mechanisms. The manipulated message that an individual sends to the overall mechanism is a random variable with a distribution $h : T \times \mathcal{M}(T^e) \to \mathcal{M}(R)$. Hence, given that the coalition organizer receives a distribution of messages $\chi \in \mathcal{M}(T^e)$, the probability that an individual who has sent a message $t' \neq \emptyset$ to the coalition organizer will send a message in a subset R' of R to the overall mechanism equals $h(t', \chi)[R']$.

The distribution of reports g to the overall mechanism that is generated by the coalition organizer is a function of the distribution of messages that the coalition organizer receives. Formally, for $\chi \in \mathcal{M}(T^e)$, we denote by $g(\chi)[R']$ the mass of agents for whom the coalition organizer chooses a report that belongs to a subset R' of R. The function $g(\chi)$ belongs to a set of feasible distributions of reports that we denote by $G(\chi)$.[11]

It is convenient to assume that the probability space that is used by the coalition organizer satisfies a law of large numbers for large economies. Consequently, $h(t', \chi)[R']$ can be interpreted not only as the probability that the manipulated message of an individual who has communicated t' to the coalition organizer lies in R', but also as the fraction of individuals (among those who have sent t' to the coalition organizer) whose manipulated message lies in R'.[12]

[9] An earlier version of this paper contained a proof of this claim. It is available upon request.

[10] In the Appendix we prove a revelation principle for manipulation mechanisms. This implies that we may assume without loss of generality that the message set for individuals who join the manipulation mechanism is equal to the set of types T, and that, in equilibrium, these individuals report their types truthfully to the coalition organizer.

[11] Formally, $G(\chi)$ is the set of non-negative measurable functions from T^e to R so that $z \in G(\chi)$ if and only if $\int_R dz(r) = \chi(T)$.

[12] A law of large numbers for large economies has been formalized in different ways. For approaches that are

In the following we illustrate these formal definitions by means of the example of a collective manipulation that was discussed in the Introduction.

Example 1 *Consider a type space with the following properties: there are three possible types* $T = \{t^1, t^2, t^3\}$*, and two possible cross-section distributions of payoff types* $\delta^0 = (0.6, 0.1, 0.3)$ *and* $\delta^1 = (0.1, 0.6, 0.3)$*. All individuals belief that these two type distributions are equally likely; for all* $t \in T$*,* $\beta(t)[\delta^1] = \beta(t)[\delta^2] = \frac{1}{2}$*. Individuals differ however in their payoff types,* $\tau(t^1) = 0$*,* $\tau(t^2) = 3$*, and* $\tau(t^3) = 10$*.*

Suppose the per capita cost of public goods provision is equal to $k = 4$*, that the overall mechanism is a direct mechanism with a truth-telling equilibrium that achieves efficient public good provision based on equal cost sharing. Consequently, for* $\delta = \delta^1$ *the public good is provided and the payoff of a type* t *individual equals* $\tau(t) - k$*. For* $\delta = \delta^0$ *the public good is not provided and each individual realizes a payoff of* 0*.*

We now describe a manipulation by individuals with types t^1 *and* t^2 *which blocks public goods provision whenever* $\delta = \delta^1$*. Note that, given that all types except one participate, the coalition organizer learns whether* $\delta = \delta^0$*, or* $\delta = \delta^1$*.*

If $\delta = \delta^0$*, the outcome is not manipulated so that* $h(t^1, \delta^0)[t^1] = 1$*,* $h(t^2, \delta^0)[t^2] = 1$*, and* $g(\delta^0)[t^1] = 0.6$ *and* $g(\delta^0)[t^2] = 0.1$*; i.e., for* $\delta = \delta^0$ *the coalition organizer reports truthfully both at the individual level and at the aggregate level.*

By contrast, if $\delta = \delta^1$*,* $h(t^1, \delta^1)[t^1] = 1$*,* $h(t^2, \delta^1)[t^1] = \frac{5}{6}$*,* $h(t^2, \delta^1)[t^2] = \frac{1}{6}$*, and* $g(\delta^1)[t^1] = 0.6$ *and* $g(\delta^1)[t^2] = 0.1$*; hence, if* $\delta = \delta^1$*, the coalition organizer manipulates the messages sent by individuals with types* t^2 *so that each of these individuals sends message* t^1 *to the overall mechanism with probability* $\frac{5}{6}$ *and message* t^2 *with probability* $\frac{1}{6}$*.*

Strategies and payoffs

A strategy for the game induced by the overall mechanism and the manipulation mechanism is a pair of functions μ and ν. The function $\mu : T \to T^e$ determines whether or not an individual participates, and the message sent to the coalition organizer in case of participation. Given the revelation principle for manipulation mechanisms that we prove in the Appendix, it entails no loss of generality to assume that $\mu(t) \neq \emptyset$ implies that $\mu(t) = t$. The function $\nu : \mu^{-1}(\emptyset) \to R$ specifies the message that a non-participant sends to the overall mechanism.

Given that all other individuals behave according to μ and ν, the expected payoff $U(\mu, \nu, t, r)$ of a type t individual from sending report r immediately to the overall mechanism equals

$$U(\mu, \nu, t, r) = \int_{\mathcal{M}(T)} u(\mu, \nu, t, r, \delta) d\beta(t) \ ,$$

with

$$u(\mu, \nu, t, r, \delta) \ := \ \tau(t)\hat{Q}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) - \hat{p}_{f_R}(r, g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) \ .$$

based on independent and identically distributed random variables, see Judd (1985) and Al-Najjar (2004). Al Najjar's work also contains an extension to environments in which types are identically distributed but correlated. An alternative approach for identically distributed and correlated random variables is provided by Alòs-Ferrer (2002).

The expected payoff $V(\mu, \nu, t, r)$ of a type t individual from announcing type t' to the coalition organizer equals

$$V(\mu, \nu, t, t') = \int_{\mathcal{M}(T)} v(\mu, \nu, t, t', \delta) d\beta(t)$$

with

$$\begin{aligned} v(\mu, \nu, t, r, \delta) \quad := \quad & \tau(t)\hat{Q}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) \\ & - \int_R \hat{p}_{f_R}(r, g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) dh(t', \delta \circ \mu^{-1}) \ . \end{aligned}$$

Coalition-Proof Interim Nash Equilibrium

Given a type space $\mathfrak{T}$ and a mechanism f_R. The strategy σ^* is a coalition-proof interim Nash equilibrium if it is an interim Nash equilibrium and there exists no manipulation mechanism (g, h) and no strategy-pair (μ, ν) such that

1. The *participation constraints for a manipulation mechanism* are satisfied: for all t, with $\mu(t) \neq \emptyset$,

$$V(\mu, \nu, t, \mu(t)) > \int_{\mathcal{M}(T)} [\tau(t)\hat{Q}_{f_R}(\delta \circ \sigma^{*-1}) - \hat{p}_{f_R}(\sigma^*(t), \delta \circ \sigma^{*-1})] d\beta(t) \ ,$$

and

2. The *equilibrium conditions for a manipulation mechanism* are satisfied:

(i) Participating individuals prefer to communicate with the coalition organizer according to μ, over any alternative communication strategy. Formally, for all t with $\mu(t) \neq \emptyset$ and for all $t' \in T$, $V(\mu, \nu, t, \mu(t)) \geq V(\mu, \nu, t, t')$.

(ii) Participating individuals prefer to communicate with the coalition organizer according to μ, over directly reporting to the overall mechanism. Formally, for all t with $\mu(t) \neq \emptyset$ and for all $r \in R$, $V(\mu, \nu, t, \mu(t)) \geq U(\mu, \nu, t, r)$.

(iii) Non-Participants prefer to communicate with the overall mechanism designer according to ν, over any alternative communication strategy. Formally, for all t with $\mu(t) = \emptyset$ and for all $r \in R$, $U(\mu, \nu, t, \nu(t)) \geq U(\mu, \nu, t, r)$.

(iv) Non-Participants prefer to communicate with the overall mechanism designer according to ν, over communicating with the coalition organizer. Formally, for all t with $\mu(t) = \emptyset$ and for all $t' \in T$, $U(\mu, \nu, t, \nu(t)) \geq V(\mu, \nu, t, t')$.

3.2 Robust Implementability as a Coalition-Proof Equilibrium

The following proposition characterizes the social choice functions which are robustly implementable as a coalition-proof interim Nash equilibrium; i.e., the social choice functions with the property that, for every type space $\mathfrak{T}$, there exists a mechanism that implements this social function as a coalition-proof interim Nash equilibrium.

Proposition 3 *A social choice function with equal payments $F = (\hat{Q}_F, \bar{p}_F)$ is robustly implementable as a coalition-proof interim Nash equilibrium if and only if there is no type space $\mathfrak{T} = [(T, \mathcal{T}), \mathbf{t}, \tau, \beta]$ and no subset T' of T with a manipulation of type announcements g so that*

$$\begin{aligned}&\int_{\mathcal{M}(T)} \Big\{\tau(t)\hat{Q}_F\Big((g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1}\Big) \\ &\qquad -\bar{p}_F\Big((g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1}\Big)\Big\}\, d\beta(t) \\ &> \int_{\mathcal{M}(T)} \{\tau(t)\hat{Q}_F(\delta \circ \tau^{-1}) - \bar{p}_F(\delta \circ \tau^{-1})\} d\beta(t)\ ,\end{aligned}$$

where μ is such that $\mu(t) = t$ for all $t \in T'$ and ν is such that $\nu(t) = t$ for all $t \in T \backslash T'$.

This Proposition characterizes the social choice functions that are robustly implementable as a coalition-proof equilibrium. By Proposition 1 a necessary condition is that the social choice function has equal payments. The additional constraint that is implied by the requirement of coalition-proofness is that there must not exist a subset of types who can manipulate the decision on public goods provision in such a way that they are all better off.

The proof of the proposition is in the Appendix. The logic is as follows: The restriction to social choice functions with equal payments implies that it is without loss of generality to limit attention to mechanisms with equal payments, i.e., to mechanisms with the property that, for a given distribution of reports $\rho \in \mathcal{M}(R)$, individuals who send different reports have to make the same payment. For this class of mechanisms however, any manipulation mechanism satisfies the equilibrium conditions. If all reports are treated equally, then any profile of reports is an equilibrium, so that the equilibrium conditions are trivially satisfied.[13] Consequently, a manipulation mechanism only has to satisfy the participation constraint. This observation makes it possible to prove a revelation principle. In particular, we show that if there is no direct mechanism with equal payments that achieves a given social choice function in a truthful equilibrium without being manipulated, then there is also no indirect mechanism with this property. Finally, upon replacing the equilibrium allocation of the direct mechanism by the social choice function we obtain the non-manipulability condition in Proposition 3.

An interesting observation is that the requirement of robustness is actually necessary for the possibility to prove the revelation principle; i.e., for a fixed type space it is not possible to show that a social choice function is implementable as a coalition-proof interim Nash equilibrium only if it is implementable by a direct mechanism. The reason is akin to the well-known result that the implementation of a social choice function as the unique equilibrium of some mechanism may require the use of non-direct mechanisms.[14] To see the similarity, note that, for a given type space, our notion of coalition-proofness requires that the possibility to reach a certain equilibrium must not be endangered by the existence of a second equilibrium which is attractive for a subset of types. Since the use of non-direct mechanism generally makes it possible to

[13] This observation implies in particular that we may replace the assumption that the coalition organizer chooses reports to the overall mechanism by the weaker assumption that he recommends reports to individuals, who then have a choice whether or not to follow this recommendation. Given that all individuals have to make the same payment to the overall mechanism, individuals are also willing to make any report that is recommended to them.

[14] See Bassetto and Phelan (2008), Jackson (2001), or Moore (1992).

get rid of undesired equilibria that a direct mechanism possesses, we are unable to prove a revelation principle based on our notion of a coalition-proof equilibrium. However, once we insist on robust implementability as a coalition-proof equilibrium, we can restrict attention to social choice functions with equal payments. For this class of social choice functions, we can establish that the revelation principle holds.

4 Characterization of social choice functions that are robustly implementable as a coalition-proof equilibrium

In this section, we characterize the set of social choice functions that are implementable as a coalition-proof interim Nash equilibrium equilibrium on every type space; i.e., we characterize the social choice functions with equal payments that satisfy the non-manipulability condition in Proposition 3.

We proceed in two steps. First, we fix a common prior type space $\mathfrak{T}$ and study the social choice functions with equal payments that are non-manipulable for this given type space. We then move on to a characterization of social choice functions that fulfill this condition on every type space.

Proposition 4 *Let $F = (\hat{Q}_F, \bar{p}_F)$ be a social choice function that is implementable as a coalition-proof equilibrium on a common prior type space $\mathfrak{T}$. Then there exist numbers p_F^0 and p_F^1 such that $\bar{p}_F(s) = p_F^0$, whenever $\hat{Q}_F(s) = 0$ and $\bar{p}_F(s) = p_F^1$, whenever $\hat{Q}_F(s) = 1$.*

The Proposition follows from the observation that whenever a decision on public good provision can be induced either in a way that is cheap for all individuals or in a way that is expensive for all individuals, then there is an incentive to form a grand coalition of all agents who manipulate their announcements in such a way that the expensive state will never be revealed.

Given Proposition 4, a social choice function is in the following written as $F = (\hat{Q}_F, p_F^0, p_F^1)$. Denote by $V^1 = \{v \mid v > p_F^1 - p_F^0\}$ the set of payoff types who prefer public good provision and by $V^0 = \{v \mid v \leq p_F^1 - p_F^0\}$ the set of payoff types who oppose public good provision. Denote by $C^1 = \tau^{-1}(V^1)$ and $C^0 = \tau^{-1}(V^0)$ the corresponding sets of types. If we want to emphasize that the sets V^1 and V^0 depend on the payments p_F^1 and p_F^0 we write $V^0(p_F^0, p_F^1)$ and $V^1(p_F^0, p_F^1)$. We omit this reference if this creates no confusion.

It will prove convenient to introduce a shorthand notation that for truth-telling behavior by a manipulation by all opponents and by all proponents of the public good, respectively.

Consider first a manipulation by all types in C^1. Let μ_1 be their communication with the coalition organizer, i.e., $\mu_1(t) = t$ if $t \in C^1$ and $\mu_1(t) = \emptyset$, otherwise. Let ν_1 be the complementary strategy of truthful communication with the overall mechanism for individuals with types in C^0, i.e., $\nu_1(t) = t$ for $t \in C^0$. Truth-telling g_1^* is a manipulation strategy with the property that, for every δ, and for every subset C of C^1, $g_1^*(\delta \circ \mu_1^{-1})[C] = \delta \circ \mu_1^{-1}(C)$.

In a similar way we characterize truth-telling by all types in C^0 as a manipulation strategy g_0^* which is applied under the assumption messages are sent according to functions μ_0 and ν_0,

that are defined analogous to μ_1 and ν_1.

Proposition 5 *Consider a social choice function $F = (\hat{Q}_F, p_F^0, p_F^1)$. Let $\mathfrak{T}$ be a common prior type space.*

i) There is no manipulation mechanism that is beneficial for individuals with types in C, for any $C \subset C^0$, if and only if these types can not reduce the probability of public good provision. Formally, for every δ,

$$g_0^*(\delta \circ \mu_0^{-1}) \in argmin_{g_0 \in G(\delta \circ \mu_0^{-1})} E[\hat{Q}_F((g_0 + \delta \circ \nu_0^{-1}) \circ \tau^{-1}) \mid \delta \circ \mu_0^{-1}] \,. \tag{3}$$

ii) There is no manipulation mechanism that is beneficial for individuals with types in C, for any $C \subset C^1$, if and only if these types can not increase the probability of public good provision. Formally, for every δ,

$$g_1^*(\delta \circ \mu_1^{-1}) \in argmax_{g_1 \in G(\delta \circ \mu_1^{-1})} E[\hat{Q}_F((g_1 + \delta \circ \nu_1^{-1}) \circ \tau^{-1}) \mid \delta \circ \mu_1^{-1}] \,. \tag{4}$$

Proposition 5 establishes that there are increasing returns to the size of a coalition. If more opponents of public good provision agree on a manipulation this is beneficial for each subcoalition of opponents. Moreover, if there is a manipulation by all opponents, then there is no scope for manipulations by a subcoalition because the gains from coalition formation are already exhausted.

The conditions stated in Proposition 5 would be sufficient to establish that $F = (\hat{Q}_F, p_F^0, p_F^1)$ is coalition-proof if there was no possibility for individuals in C^0 and in C^1 to agree on a manipulation. The next Proposition establishes that this is indeed the case if attention is limited to provision rules which are efficient. Provision rule $\hat{Q}_F$ is said to be *efficient* given p_F^0 and p_F^1 if there is no provision rule $\hat{Q}' : \mathcal{M}(V) \to \{0, 1\}$ such that, for all $t \in T$

$$E[\hat{Q}'(s)(\tau(t) - p_F^1) - (1 - \hat{Q}'(s))p_F^0 \mid t] \geq E[\hat{Q}_F(s)(\tau(t) - p_F^1) - (1 - \hat{Q}_F(s))p_F^0 \mid t] \,,$$

with a strict inequality for at least one t.

Proposition 6 *Given a common prior type space $\mathfrak{T}$. Given p_F^0 and p_F^1, $\hat{Q}_F$ is efficient among the set of coalition-proof provision rules if and only if it is efficient among the provision rules satisfying (3) and (4).*

With a common prior type space, it is not possible to increase the probability of public good provision from the perspective of individuals who benefit from public good provision and simultaneously decrease it for those who are harmed if the public good is provided. Consequently, efficiency among the provision rules satisfying (3) and (4) implies efficiency among all provision rules.

Proposition 7 *Given a common prior type space $\mathfrak{T}$. Given p_F^0 and p_F^1, a provision rule $\hat{Q}_F$ satisfies (3) and (4) if and only if it satisfies the following two properties:*

i) For every pair δ and δ' such that $\delta(C_F^1) = \delta'(C_F^1)$,

$$\hat{Q}_F(\delta \circ \tau^{-1}) = \hat{Q}_F(\delta' \circ \tau^{-1}). \tag{5}$$

ii) For every pair δ and δ' such that $\delta(C_F^1) \geq \delta'(C_F^1)$,

$$\hat{Q}_F(\delta \circ \tau^{-1}) \geq \hat{Q}_F(\delta' \circ \tau^{-1}). \tag{6}$$

Proposition 7 provides a simple characterization of coalition-proof provision rule for a given type space. The decision on public good provision is a non-decreasing function of the share of individuals who benefit from public good provision. In particular, public good provision cannot respond to a change in preference intensities that leaves the share of individuals in favor of public good provision unaffected. Coalition-proofness implies that a provision rule has to be implementable by a very simple mechanism. Everybody who wants to have the public good is asked to raise his hand and the public good is provided if the number of hands exceeds some threshold.

The proof of Proposition 7 uses results from the analysis of two person zero sum games. Such games do not have pure strategy equilibria unless one party has a dominant or a dominated action. In our setting, coalition-proofness requires essentially that truth-telling is an equilibrium in pure strategies in a zero sum game between the opponents and the proponents of the public good. As a consequence, for every δ, one of these parties must have a dominant action, i.e., it must be able to choose announcements such that the preferred outcome results. Since all states where population shares are the same generate the same set of feasible actions for the coalition organizer, this implies that in all of these states the outcome must be the same.

So far the analysis was based on a given common prior type space $\mathfrak{T}$. However, the analysis can be easily generalized. The following proposition states that if $\hat{Q}_F$ is coalition-proof on some common prior type space, then it is coalition-proof on every type space.

Proposition 8 *Given p_F^0 and p_F^1. Let $\hat{Q}_F$ be such that (5) and (6) hold. Then $\hat{Q}_F$ is implementable as a coalition-proof interim Nash equilibrium on every type space.*

The proposition follows immediately from the fact that, whatever the type space is, the relevant coalitions are the set of individuals with payoff types in V^0 and the set of individuals with payoff types in V^1. To see this, note that conditions (5) and (6) can be equivalently stated as follows: whenever $s, s' \in \mathcal{M}(V)$ are such that $s(V^1) = s'(V^1)$, then $\hat{Q}_F(s) = \hat{Q}_F(s')$ and whenever $s(V^1) \geq s'(V^1)$, then $\hat{Q}_F(s) \geq \hat{Q}_F(s')$.

Given this property, whatever the types pace – i.e. whatever the functions τ, that assigns payoff types in V to types in T, and β, that assigns beliefs in $\mathcal{M}(\mathcal{M}(T))$ to types in T, are – individuals with payoff types in V^1 have no way to increase the probability of public good provision. If they lie about their types such that there are more announcements of payoff types in V^0, the probability of public good provision goes down and this not beneficial. If they lie about their types such that the announcements of payoff types in V^0 does not go up, then the

decision on public good provision is not affected and hence this manipulation is not beneficial either.

5 Implications for efficiency

In this section, we discuss the welfare implications of coalition-proofness. We first show that the requirement of coalition-proofness implies that first best allocations can not be reached and then move on to a characterization of second-best allocations which satisfy the requirement of coalition-proofness.

Proposition 9 *Suppose there is a set of states s and s' such that $s(V^1) = s'(V^1)$ and $\bar{v}(s) < k < \bar{v}(s')$. Then, there is no social choice function that yields first best outcomes and is robustly implementable as a coalition-proof equilibrium.*

Proposition 9 is a direct consequence of Propositions 2 and 8. According to Proposition 2, a necessary condition for the possibility to achieve, simultaneously, first best outcomes and robust implementability is that the public good is provided whenever the average valuation of the public good exceeds the per capita cost of providing it. According to Proposition 8 the decision on public goods provision has to be the same for every pair of states which give rise to the same population shares of individuals who benefit from public goods provision. In general, these two requirements are incompatible.

To see this, recall the example that we discussed already in the introductory section. Per capita cost of public goods provision are equal to $k = 4$. The set V^0 consists of payoff types 0 and 3 and the set V^1 contains individuals with $v = 10$. Let state s be such that the population share of individuals with valuation 0 equals $\frac{7}{10}$, the share of individuals with valuation 3 equals 0 and the share of individuals with valuation 10 equals $\frac{3}{10}$. In state s', by contrast, the share of individuals with valuation 0 equals 0, the share of individuals with valuation 3 equals $\frac{7}{10}$ and the share of individuals with valuation 10 equals $\frac{3}{10}$. First best requires that $\hat{Q}_F(s) = 0$ and $\hat{Q}_F(s') = 1$. Coalition-proofness requires that $\hat{Q}_F(s) = \hat{Q}_F(s')$. Obviously, this is incompatible.

The assumption that there exist two states s and s' such that $s(V^1) = s'(V^1)$ and $\bar{v}(s) < k < \bar{v}(s')$ is very weak. For instance, if we let $k \in (0, 1)$ and $V = [0, 1]$ this condition will be satisfied.

Given the impossibility to achieve efficient outcomes for every state s, we now turn to a characterization of second best allocations. We assume that there is a mechanism designer who has specific beliefs about s and who chooses a coalition-proof social choice function in order to maximize expected welfare, i.e., he chooses p_F^0, p_F^1 and $\hat{Q}_F : \mathcal{M}(V) \to \{0, 1\}$ in order to maximize

$$E[(\bar{\mathbf{v}}(s) - p_F^1)\hat{Q}_F(s) - p_F^0(1 - Q_F(s))]$$

subject to the monotonicity constraint, $\hat{Q}_F(s) \geq \hat{Q}_F(s')$, whenever $s(V^1) > s'(V^1)$, the constraint that $\hat{Q}_F(s) = \hat{Q}_F(s')$, whenever $s(V^1) = s'(V^1)$, and the budget constraints $p_F^1 \geq k$ and $p_F^0 \geq 0$. Obviously, which social choice function is optimal depends on the mechanism designer's

beliefs. Our results below clarify how alternative assumptions about the mechanism designer's beliefs shape the second-best mechanism. In particular, we are interested in identifying conditions such that the monotonicity constraints and the feasibility constraints are not binding, so that deviations from first best are entirely due to the constraint that the decision on public good provision can only be based on the information about the population share of individuals whose valuation exceeds the per capita cost of public goods provision.

The following assumption will make it possible to show that the monotonicity constraint will bot be binding.

Assumption 1 *For every pair p_F^0 and p_F^1, and every pair of numbers x and $x' \in [0,1]$, $x > x'$ implies*

$$E[\bar{\mathbf{v}}(s) \mid s(V^1(p_F^0, p_F^1)) = x] > E[\bar{\mathbf{v}}(s) \mid s(V^1(p_F^0, p_F^1)) = x'] \, .$$

Assumption 1 states that the expected value of the average valuation of the public good is an increasing function of the number of individuals who benefit from public good provision. This assumption is illustrated by the following example.

Example 2 *Suppose there are five possible payoff types $V = \{0, 3, 4, 5, 10\}$ and four possible distributions of payoff types, $\{s^j\}_{j=1}^4$, where, for each j, s^j is a vector $s^j = (s_0^j, s_3^j, s_4^j, s_5^j, s_{10}^j)$ that lists the population shares of the different payoff types. Assume that the per capita cost of public good provision k equals 4.5. Hence, with $p_F^0 = 0$ and $p_F^1 = k$ coalition V^0 consists of 0, 3 and 4 whereas V^1 consists of types 5 and 10. The population shares of the different payoff types are listed in the following table.*

j	s_0^j	s_3^j	s_4^j	s_5^j	s_{10}^j	$\bar{\mathbf{v}}(s^j)$
1	*0.5*	*0.3*	*0*	*0.1*	*0.1*	*2.4*
2	*0*	*0*	*0.8*	*0.1*	*0.1*	*4.7*
3	*0.2*	*0.1*	*0*	*0.6*	*0.1*	*4.3*
4	*0.2*	*0.1*	*0*	*0.1*	*0.6*	*6.8*

Let the beliefs of the mechanism designer on the cross-section distribution of payoff types be given by $(\alpha^j)_{j=1}^4$, where $\alpha^j := prob(\tilde{s} = s^j)$, and note that

$$E[\bar{\mathbf{v}}(s) \mid s(V^1) = 0.2] = \alpha_1 2.4 + \alpha_2 4.7 \quad \textit{and} \quad E[\bar{\mathbf{v}}(s) \mid s(V^1) = 0.7] = \alpha_3 4.3 + \alpha_4 6.8 \, .$$

Hence, if α^3 is high relative to α^4, and α^2 is high relative to α^1, then Assumption 1 is violated because the expected value of $\bar{v}(s)$ given that 20 per cent of the population benefit from public good provision exceeds the expected value of $\bar{v}(s)$ given that 70 per cent of the population benefit from public good provision.

Assumption 1 implies that the monotonicity constraint – $\hat{Q}_F(s) \geq \hat{Q}_F(s')$, whenever $s(V^1) > s'(V^1)$ – is not binding. To illustrate this, fix p_F^1 and p_F^0 and consider a relaxed problem

of maximizing $E[(\bar{\mathbf{v}}(s) - p_F^1)\hat{Q}_F(s) - p_F^0(1 - Q_F(s))]$ subject to the constraint that $s(V^1) = s'(V^1)$ implies $\hat{Q}_F(s) = \hat{Q}_F(s')$. Obviously, it is optimal to choose $\hat{Q}_F(s) = 1$ if and only if $E[\bar{\mathbf{v}}(s) - (p_F^1 - p_F^0) \mid s(V^1)] > 0$. Assumption 1 implies that the resulting provision rule satisfies the monotonicity constraint $\hat{Q}_F(s) \geq \hat{Q}_F(s')$ whenever $s(V^1) > s'(V^1)$. Hence, under Assumption 1 the monotonicity constraint will never be binding.

This can also be illustrated by means of Example 2. If α^3 is high relative to α^4, and α^2 is high relative to α^1, then, with $p_F^0 = 0$ and $p_F^1 = k$, the solution to the relaxed problem is such that $\hat{Q}_F(s) = 0$ whenever $s(V^1) = 0.7$, and $\hat{Q}_F(s) = 1$ whenever $s(V^1) = 0.2$. Hence, the monotonicity constraint is violated.

These observations prove the following Proposition.

Proposition 10 *The monotonicity constraint $\hat{Q}_F(s) \geq \hat{Q}_F(s')$, whenever $s(V^1) > s'(V^1)$ is not binding at a solution of the second best problem if and only if Assumption 1 holds.*

Our second assumption ensures that the budget constraint is binding at a solution to the second-best problem.

Assumption 2

i) Let $p_F^0 = 0$ and $p_F^1 > k$. Define $\hat{x}(p_F^1)$ implicitly by the equation

$$E[\bar{\mathbf{v}}(s) - p_F^1 \mid s(V^1(0, p_F^1)) = \hat{x}(p_F^1)] = 0 \, .$$

$\hat{x}(p_F^1)$ is an increasing function of p_F^1.

ii) Let $p_F^0 > 0$ and $p_F^1 = k$. Define $\bar{x}(p_F^0)$ implicitly by the equation

$$E[\bar{\mathbf{v}}(s) - (k - p_F^0) \mid s(V^1(p_F^0, k)) = \bar{x}(p_F^0)] = 0 \, .$$

$\bar{x}(p_F^0)$ is a decreasing function of p_F^0.

Suppose that $p_F^0 = 0$ and $p_F^1 > k$. Assumption 2 says that a decrease of p_F^1 does not only imply that the set of individuals who prefer public good provision over non-provision, V^1, contains more payoff types but also that the set of states in which public good provision is desirable expands. In particular, in all states where public good provision was desirable with high p_F^1, public good provision remains desirable if p_F^1 is decreased. Analogously, suppose that $p_F^0 > 0$ and $p_F^1 = k$. A decrease of p_F^0 implies that the set V^0 becomes larger. In addition the set of states in which non-provision is desirable expands. The following example shows that this assumption is not trivially satisfied.

Example 3 *Let the set of states be as in Example 2. Suppose that the feasibility constraints are binding, so that $p_F^0 = 0$ and $p_F^1 = k$. The constraints, $\hat{Q}_F(s) \geq \hat{Q}_F(s')$, whenever $s(V^1) > s'(V^1)$, and $\hat{Q}_F(s) = \hat{Q}_F(s')$, whenever $s(V^1) = s'(V^1)$ imply that the following provision rules*

are admissible: (i) $\hat{Q}_F(s) = 0$*, for all* s*, (ii)* $\hat{Q}_F(s) = 1$*, for all* s*, and (iii)* $\hat{Q}_F(s^1) = \hat{Q}_F(s^2) = 0$*,* $\hat{Q}_F(s^3) = \hat{Q}_F(s^4) = 1$*. Let the beliefs of the mechanism designer be given by* $\alpha^1 = 0.1$*,* $\alpha^2 = 0.25$*,* $\alpha^3 = 0.6$ *and* $\alpha^4 = 0.05$*. It is easy to check that, under these assumptions, the optimal provision rule is* $\hat{Q}_F(s) = 0$*, for all* s*, and the realized level of expected utilitarian welfare equals* 0*. Now choose* $p_F^0 = 0$*, and* $p_F^1 = 5 + \epsilon$*, where* ϵ *is positive and arbitrarily small. This changes the set of individuals who benefit from public goods provision,* $V^1(0, 5 + \epsilon)$ *consists only of the individuals with a valuation* 10*. As a consequence the following provision rule becomes admissible,* $\hat{Q}_F(s^1) = \hat{Q}_F(s^2) = \hat{Q}_F(s^3) = 0$*, and* $\hat{Q}_F(s^4) = 1$*. Moreover, this provision rule yields a strictly positive level of expected welfare. This violates Assumption 2:* p_F^1 *is decreased from 5 to 4.5 and the set of states in which public good provision is desirable shrinks.*

Proposition 11 *Suppose that Assumption 1 holds. A second best allocation is such that the feasibility constraints hold as an equality if and only if, in addition, Assumption 2 holds.*

Imposing Assumptions 1 and 2 has the following implication. Consider an initial situation with $p_F^0 = 0$ and $p_F^1 > k$. If we lower the payment p_F^1 then, public goods provision remains desirable in all states where it was desirable before. Moreover, if it public goods provision is undesirable with a reduced level of p_F^1, then it is also undesirable with the high level p_F^1. In addition, there is a set of states where the desirability of public goods provision changes due to the reduction in the payment. Given that it would still be possible not to provide the public good in these states, welfare is increased due to reduction of p_F^1. However, these arguments rely on Assumption 1. Otherwise a binding monotonicity constraint may imply that these potential improvements cannot be realized. A formal proof of these assertions can be found in the Appendix.

Given Assumptions 1 and 2, we may safely ignore the monotonicity constraint, $\hat{Q}_F(s) \geq \hat{Q}_F(s')$, whenever $s(V^1) > s'(V^1)$, and we may assume that the feasibility constraints $p_F^1 \geq k$ and $p_F^0 \geq 0$ are both binding. Consequently, an optimal provision rule for the public good maximizes $E[(\bar{\mathbf{v}}(s) - k)\hat{Q}_F(s)]$ subject to the constraint that $\hat{Q}_F(s) = \hat{Q}_F(s')$, whenever $s(V^1) = s'(V^1)$.

Depending on the mechanism designer's beliefs it may even by the case that this last constraint is not binding. This is established by the following Proposition, which we state without proof.

Proposition 12 *For a given set of beliefs, there is a social choice function that is robustly implementable as a coalition-proof equilibrium and achieves a first best allocation almost surely, if and only if Assumptions 1 and 2 hold, and, moreover, there is no pair of subsets* S' *and* S'' *of* S *that occur with positive probability each and satisfy the following two properties:*

i) For all $s \in S' \cup S''$*,* $s(V^1) = x$*, for some* $x \in [0, 1]$*.*

ii) $E[\bar{\mathbf{v}}(s) \mid s \in S'] < k < E[\bar{\mathbf{v}}(s) \mid s \in S'']$*.*

The Proposition shows that first best outcomes are possibly achieved almost surely from the perspective of a mechanism designer who has specific beliefs. However, in the light of Proposition

9 this possibility is not robust with respect to the specification of the mechanism designer's beliefs; i.e., whenever there is a set of beliefs such that efficiency can be achieved almost surely there is another set of beliefs so that the probability of achieving an efficient outcome is strictly less than one.

To illustrate this consider once more Example 2, and suppose, for the sake of the argument, that the mechanism designer's beliefs satisfy Assumptions 1 and 2. If the mechanism designer's beliefs assign probability zero to states s_1 and s_3, $\alpha_1 = \alpha_3 = 0$, then efficiency is achieved almost surely be choosing $Q = 0$ in states s_1 and s_2 and $Q = 1$ in states s_3 and s_4. The fact that states s_1 and s_3 occur with probability zero imply that the inefficiencies associated with these states are not felt by the mechanism designer. Obviously, this is no longer true for beliefs such that α_1 or α_3 are bounded away from zero.

6 Concluding Remarks

We studied the problem of public good provision in a large economy with uncertainty about the distribution of preferences. In a large economy, the standard notion of individual incentive compatibility becomes trivially fulfilled under any anonymous allocation mechanism. Since a single individual has no chance of affecting the outcome, there is no reason to misrepresent private information on preferences. Under the assumption that individuals can coordinate their behavior in order to induce outcomes that are more attractive to them, coalition-proofness, by contrast, becomes a binding constraint.

Our notion of coalition-proofness can be defined in an equivalent way for finite economies. In a finite economy, however, both individual and collective incentive compatibility are relevant constraints on the set of incentive-feasible outcomes and this complicates the analysis. Besides its plausibility for allocation problems involving millions of individuals, the focus on a large economy has the advantage that coalition-proofness can be characterized in an easy and intuitive manner: Only ordinal information on preferences can be used as a basis for collective decision making, because all individuals who agree on the ranking of alternatives will coordinate in such a way that their most preferred outcome has a high chance of being implemented.

In our model with a yes-or-no-decision on public good provision this implies, that an optimal decision can be implemented with a simple voting rule: Ask people to raise their hand if they want the public good to be provided and provide the public good if the number of hands is high enough.

References

Al-Najjar, N. (2004). Aggregation and the law of large numbers in large economies. *Games and Economic Behavior*, 47:1–35.

Alòs-Ferrer, C. (2002). Individual randomness in economic models with a continuum of agents. Working Paper 9807, Dpt. of Economics, University of Vienna.

Bassetto, M. and Phelan, C. (2008). Tax riots. *Review of Economic Studies*, 75:649–669.

Bergemann, D. and Morris, S. (2005). Robust mechanism design. *Econometrica*, 73:1771–1813.

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

Crémer, J. and McLean, R. (1985). Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica*, 53:345–361.

Crémer, J. and McLean, R. (1988). Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica*, 56:1247–1257.

d'Aspremont, C., Crémer, J., and Gérard-Varet, L. (1990). Incentives and the existence of pareto-optimal revelation mechanisms. *Journal of Economic Theory*, 51:233–254.

d'Aspremont, C., Crémer, J., and Gérard-Varet, L. (2004). Balanced Bayesian mechanisms. *Journal of Economic Theory*, 115:385–396.

d'Aspremont, C. and Gérard-Varet, L. (1979). Incentives and incomplete information. *Journal of Public Economics*, 11:25–45.

Green, J. and Laffont, J. (1979). *Incentives in Public Decision-Making*. North-Holland Publishing Company.

Groves, T. (1973). Incentives in teams. *Econometrica*, 41:617–663.

Guesnerie, R. (1995). *A Contribution to the Pure Theory of Taxation*. Cambridge University Press.

Güth, W. and Hellwig, M. (1986). The private supply of a public good. *Journal of Economics*, Supplement 5:121–159.

Hellwig, M. (2003). Public-good provision with many participants. *Review of Economic Studies*, 70:589–614.

Hildenbrand, W. (1974). *Core and Equilibria of a Large Economy*. Princeton University Press.

Jackson, M. (2001). A crash course in implementation theory. *Social Choice and Welfare*, 18:655–708.

Johnson, S., Pratt, J., and Zeckhauser, R. (1990). Efficiency despite mutually payoff-relevant private information: The finite type case. *Econometrica*, 58:873–900.

Judd, K. (1985). The law of large numbers with a continuum of i.i.d. random variables. *Journal of Economic Theory*, 35:19–25.

Laffont, J. and Martimort, D. (1997). Collusion under asymmetric information. *Econometrica*, 65:875–911.

Laffont, J. and Martimort, D. (2000). Mechanism design with collusion and correlation. *Econometrica*, 68:309–342.

Mailath, G. and Postlewaite, A. (1990). Asymmetric bargaining procedures with many agents. *Review of Economic Studies*, 57:351–367.

Moore, J. (1992). Implementation, contracts, and renegotiation in environments with complete information. In Laffont, J.-J., editor, *Advances in Economic Theory: Sixth World Congress, vol. I.* Cambridge, UK, Cambridge University Press.

Neeman, Z. (2004). The relevance of private information in mechanism design. *Journal of Economic Theory*, 117:55–77.

Norman, P. (2004). Efficient mechanisms for public goods with use exclusion. *Review of Economic Studies*, 71:1163–1188.

Osborne, M. and Rubinstein, A. (1994). *A course in Game Theory.* MIT Press, Cambridge, MA.

Rob, J. (1989). Pollution claim settlements under private information. *Journal of Economic Theory*, 47:307–333.

Schmitz, P. (1997). Monopolistic provision of excludable public goods under private information. *Public Finance/ Finance Publiques*, 52:89–101.

A Proofs

A.1 Proof of Proposition 1

By the revelation principle, we may without loss of generality restrict attention to direct mechanisms (mechanisms with $R = T$), and truth-telling equilibria ($\sigma^*(t) = t$, for all t). Accordingly, a social choice function is robustly implementable if and only if, a direct mechanism implements F as a truthful equilibrium on every type space.

For a given type space, truth-telling is an interim Nash equilibrium for the game induced by direct mechanism f if the following equilibrium condition holds:

$$\int_{\mathcal{M}(T)} [\tau(t)\hat{Q}_f(\tilde{\delta}) - \hat{p}_f(t,\tilde{\delta})]d\beta(t) \geq \int_{\mathcal{M}(T)} [\tau(t)\hat{Q}_f(\tilde{\delta}) - \hat{p}_f(t',\tilde{\delta})]d\beta(t) \ , \tag{7}$$

for all t and all t' in T.

"$\Longleftarrow$": Given an ex post incentive compatible social choice function $F = (\hat{Q}_F, \hat{p}_F)$, construct a direct mechanism $f^* = (\hat{Q}^*_f, \hat{p}^*_f)$ as follows: for every $\delta \in \mathcal{M}(T)$, let $\hat{Q}^*_f(\delta) = \hat{Q}_F(\delta \circ \tau^{-1})$ and let $\hat{p}^*_f(t,\delta) = \hat{p}_F(\tau(t), \delta \circ \tau^{-1})$, for all $t \in T$. We show that truth-telling is an equilibrium for every type space; i.e., for any type space $\mathfrak{T}$ and any type t,

$$\begin{aligned} t \ &\in \ \text{argmax}_{t' \in T} \ \int_{\mathcal{M}(T)} [\tau(t)\hat{Q}^*_f(\tilde{\delta}) - \hat{p}^*_f(t',\tilde{\delta})d\beta(t) \\ &= \ \text{argmax}_{t' \in T} \ \int_{\mathcal{M}(V)} \Big(\int_{\{\delta | \delta \circ \tau^{-1} = s\}} d\beta(t) \Big) [\tau(t)\hat{Q}_F(s) - \hat{p}_F(\tau(t'), s)] \ , \end{aligned}$$

or, equivalently, for every $v \in V$,

$$v \in \text{argmax}_{v' \in V} \int_{\mathcal{M}(V)} \Big(\int_{\{\delta | \delta \circ \tau^{-1} = s\}} d\beta(t) \Big) [v\hat{Q}_F(s) - \hat{p}_F(v', s)] \ .$$

This latter statement follows from the fact that $F = (\hat{Q}_F, \hat{\rho}_F)$ is ex post implementable.

"$\Longrightarrow$": Fix some arbitrary $s' \in \mathcal{M}(\Theta)$ and consider the corresponding complete information type space, i.e. suppose that

$$\int_{\{\delta | \delta \circ \tau^{-1} = s'\}} d\beta(t) = 1 \tag{8}$$

for all $t \in T$. If $f = (\hat{Q}_f, \hat{p}_f)$ implements F on this type space, then it has to be true that (i) truth-telling is an equilibrium and (ii) that f achieves F. After rewriting the equilibrium conditions in (7) using that (8) holds, and replacing f by F, the equilibrium conditions read as,

$$\tau(t)\hat{Q}_F(s') - \hat{p}_F(\tau(t), s') \geq \tau(t)\hat{Q}_F(s') - \hat{\rho}_F(\tau(t'), s') \ ,$$

for all t and all t'. Equivalently, this can be written as

$$v\hat{Q}_F(s') - \hat{p}_F(v, s') \geq v\hat{Q}_F(s') - \hat{p}_F(v, s') \ ,$$

for all v and all $v' \in V$. Since s' was arbitrary, this condition has to be satisfied for any $s \in \mathcal{M}(V)$. Hence, $F = (\hat{Q}_F, \hat{p}_F)$ is ex post incentive compatible. ■

A.2 A Revelation Principle for Manipulation Mechanisms

Fix a type space $\mathfrak{T}$, a mechanism f_R, and an interim Nash equilibrium σ^*. We show in the following that whenever there is a non-direct manipulation mechanism that satisfies both the participation constraints and the equilibrium conditions for a manipulation mechanism, then there exists also a direct manipulation mechanism that is based on truth-telling of participating types that satisfies these conditions.

Consider a manipulation mechanism which is such that individuals send a message from a set $X^e = X \cup \emptyset$ to the coalition organizer. Again, agents who choose to send a message $x \neq \emptyset$ to the coalition organizer, are those who participate. Agents who send $x = \emptyset$ are those who choose their report to the overall mechanism without help of the coalition organizer.

The message that a participating individual sends to the overall mechanism is now a random variabel with a distribution $h_X : X \times \mathcal{M}(X^e) \to \mathcal{M}(R)$. At an aggregate level, the reports generated by the coalition organizer are represented by a function g_X which assigns to each $\chi_X \in \mathcal{M}(X^e)$ a measure $g_X(\chi_X) \in \mathcal{M}(R, \lambda^+(\chi_X))$.

Let $\mu_X : T \to X^e$ determine whether or not an individual participates, and the message sent to the coalition organizer in case of participation. The function $\nu_X : \mu^{-1}(\emptyset) \to R$ specifies the message that a non-participant sends to the overall mechanism.

Suppose σ^* is not coalition-proof, i.e., suppose there exists a manipulation mechanism represented by h_X and g_X so that the strategy (μ_X, ν_X) satisfies both the participation constraints and the equilibrium conditions for a manipulation mechanism.

With reference to μ_X, define the function $\mu : T \to T \cup \{\emptyset\}$ as follows: $\mu(t) = t$ if $\mu_X(t) \neq \emptyset$ and $\mu(t) = \emptyset$ if $\mu_X(t) = \emptyset$.

We show that there exists a direct manipulation mechanism with $X^e = T^e$ that can be represented by functions h and g which is such that the strategy (μ, ν_X) satisfies the participation constraints and the equilibrium conditions for a manipulation mechanism.

The function $h : T \times \mathcal{M}(T^e) \to \mathcal{M}(R)$ is defined with reference to h_X in the following way: for all t with $\mu(t) \neq \emptyset$, $h(\mu(t), \delta \circ \mu^{-1}) = h(\mu_X(t), \delta \circ \mu_X^{-1})$.

The function g specifies for each $\chi \in \mathcal{M}(T^e)$ a measure over reports $g(\chi) \in \mathcal{M}(R, \lambda^+(\chi))$. It is constructed such that, for every δ, $g(\delta \circ \mu^{-1}) = g_X(\delta \circ \mu_X^{-1})$.

By construction, we have that, for all $t \in T$, sending a message $t' \neq \emptyset$ to the coalition organizer yields an expected payoff of $V(\mu, \nu_X, t, t') = V(\mu, \nu_X, t, x)$, for some $x \in X$. Also, for a type t with $\mu(t) \neq \emptyset$, $V(\mu, \nu, t, t) = V(\mu_X, \nu_X, t, \mu_X(t))$, i.e., the expected payoff rom truth-telling under the direct manipulation mechanism is the same as the expected payoff from following μ_X under the non-direct mechanism. Likewise, for all $t \in T$, sending a report r directly to the overall mechanism yields an expected payoff of $U(\mu, \nu_X, t, r) = U(\mu_X, \nu_X, t, r)$, for all $r \in R$.

These observations imply that the "new" manipulation mechanism satisfies the participation constraints and the equilibrium conditions.

■

A.3 Proof of Proposition 3

The social choice functions that are robustly implementable as a coalition-proof interim Nash equilibrium are a subset of those that are robustly implementable as a Nash-equilibrium. As we have shown in Proposition 1, the latter can be represented by a provision rule $\hat{Q}_F : \mathcal{M}(V) \to \{0, 1\}$ and a payment rule $\bar{p}_F : \mathcal{M}(V) \to \mathbb{R}$.

In the following we take the type space $\mathfrak{T}$ and a social choice function with equal payments $F = (\hat{Q}_F, \bar{p}_F)$ as given. We show that, for this restricted class of social choice functions, the revelation principle holds.

This social choice function is implementable as a coalition-proof interim Nash equilibrium on the given type space if there exists a mechanism f_R and a coalition-proof equilibrium strategy $\sigma^* : T \to R$ such that the equilibrium allocation coincides, for every δ, with the allocation stipulated by the social function, i.e., for every δ,

$$\hat{Q}_F(\delta \circ \tau^{-1}) = \hat{Q}_{f_R}(\delta \circ \sigma^{*-1})$$

and

$$\bar{p}_F(\delta \circ \tau^{-1}) = \hat{p}_{f_R}(\sigma^*(t), \delta \circ \sigma^{*-1}) \,, \tag{9}$$

for all $t \in T$.

Suppose without loss of generality that the set of reports R contains no superfluous element, i.e., for every $r \in R$, there exists t with $\sigma^*(t) = r$. Equation (9) has the following implication:

the mechanism f_R must be a mechanism with equal payments, i.e., for every $\rho \in \mathcal{M}(R)$ there exists a number $\bar{p}_{f_R}(\rho)$ such that

$$\hat{p}_{f_R}(r, \rho) = \bar{p}_{f_R}(\rho) \ , \tag{10}$$

for all $r \in R$.

Given that we may, without loss of generality, restrict attentions to mechanisms with equal payments, the definition of coalition-proofness is greatly simplified. Any manipulation mechanism and any pair of functions μ and ν that determine the communication of participating individuals with the coalition organizer, and non-participating individuals with the overall mechanism, respectively, trivially satisfy the equilibrium conditions for a manipulation mechanism. If the mechanism responds to each report equally, then every profile of reports is an equilibrium. In particular, the profile of reports induced by the manipulation mechanism is an equilibrium.

This considerations imply that the definition of implementability as a coalition-proof equilibrium can be simplified considerably: A social choice function with equal payments $F = (\hat{Q}_F, \bar{p}_F)$ is implementable as a coalition-proof interim Nash equilibrium if and only if there is a mechanism with equal payments $f_R = (\hat{Q}_{f_R}, \bar{p}_{f_R})$ with equilibrium σ^* that achieves the social choice function and has the following property: There is no manipulation mechanism that satisfies the participation constraint. Formally, there is no manipulation mechanism (g, h) with a strategy (μ, ν) such that for all t with $\mu(t) \neq \emptyset$,

$$\int_{\mathcal{M}(T)} \{\tau(t)\hat{Q}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) - \bar{p}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1})\} d\beta(t) \quad > \\ \int_{\mathcal{M}(T)} \{\tau(t)\hat{Q}_{f_R}(\delta \circ \sigma^{*-1}) - \bar{p}_{f_R}(\delta \circ \sigma^{*-1})\} d\beta(t) \qquad .$$

Moreover, by the revelation principle for manipulation mechanisms (see Section A.2 of the Appendix) we may without loss of generality assume that $\mu(t) = t$, for all t who participate. Also, since the mechanism has equal payments and every profile of reports leads to an equilibrium, we may assume that $\nu(t) = \sigma^*(t)$, for all types who do not participate.

The key step in the proof of Proposition 3 is to establish that the revelation principle holds for social choice function with equal payments. This is stated formally in the following Lemma.

Lemma 3 *A social choice function with equal payments $F = (\hat{Q}_F, \bar{p}_F)$ is implementable as a coalition-proof interim Nash equilibrium if and only if it is implementable as the truthful equilibrium of a direct mechanism.*

Proof Let $f_R = (\hat{Q}_{f_R}, \bar{p}_{f_R})$ be a non-direct mechanism with an equilibrium strategy σ^* that implements $F = (\hat{Q}_F, \bar{p}_F)$. With reference to f_R construct a direct mechanism $f = (\hat{Q}_f, \bar{p}_f)$ which is such that, for all δ,

$$\hat{Q}_f(\delta) = \hat{Q}_{f_R}(\delta \circ \sigma^{*-1}) \tag{11}$$

and

$$\bar{p}_f(\delta) = \bar{p}_{f_R}(\delta \circ \sigma^{*-1}) \ . \tag{12}$$

Now suppose that for this direct mechanism coalition-proofness fails; i.e., there is a manipulation (g, h) and a subset T' of T such that for all $t \in T'$,

$$\int_{\mathcal{M}(T)} \{\tau(t)\hat{Q}_f(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) - \bar{p}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1})\} d\beta(t) \tag{13}$$
$$> \int_{\mathcal{M}(T)} \{\tau(t)\hat{Q}_f(\delta) - \bar{p}_f(\delta)\} d\beta(t) \ .$$

where μ is such that $\mu(t) = t$ for all $t \in T'$ and ν is such that $\nu(t) = t$ for all $t \in T \backslash T'$.

With reference to this manipulation mechanism we now construct a new manipulation mechanism (g', h') for the game induced by the non-direct mechanism f_R. In particular, we choose g' such that, for every δ and every subset T'' of T,

$$g'(\delta \circ \mu^{-1})[\sigma^*(T'')] = g(\delta \circ \mu^{-1})[T''] \ .$$

This ensures that, for every δ,

$$\hat{Q}_f(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) = \hat{Q}_{f_R}(g'(\delta \circ \mu^{-1}) + \delta \circ \nu'^{-1}) \tag{14}$$

and

$$\bar{p}_f(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) = \bar{p}_{f_R}(g'(\delta \circ \mu^{-1}) + \delta \circ \nu'^{-1}) \ , \tag{15}$$

where the function ν' is such that $\nu'(t) = \sigma^*(t)$ for all t with $\mu(t) = \emptyset$.

Substituting equations (14), (15), (11) and (12) into (13) implies that for all $t \in T'$,

$$\int_{\mathcal{M}(T)} \left\{\tau(t)\hat{Q}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu'^{-1}) - \bar{p}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu'^{-1})\right\} d\beta(t) \tag{16}$$
$$> \int_{\mathcal{M}(T)} \{\tau(t)\hat{Q}_{f_R}(\delta \circ \sigma^{*-1}) - \bar{p}_{f_R}(\delta \circ \sigma^{*-1})\} d\beta(t) \ ,$$

hence a contradiction to the assumption that $f_R = (\hat{Q}_{f_R}, \bar{p}_{f_R})$ implements $F = (\hat{Q}_F, \bar{p}_F)$ as a coalition-proof equilibrium. ■

The Proposition is an immediate corollary of this Lemma. To see this, note that the assumption that f_R achieves F in conjunction with equations (11) and (12) implies that we can substitute $\hat{Q}_F\Big((g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1}\Big)$ for $\hat{Q}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1})$ and $\bar{p}_F\Big((g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1}\Big)$ for $\bar{p}_{f_R}(g(\delta \circ \mu^{-1}) + \delta \circ \nu^{-1})$ in (13).

A.4 Proof of Proposition 4

Let P be a common prior on $\mathcal{M}(T)$, and P_V be the corresponding marginal probability distribution over distributions of payoff-types. Suppose there exists $\bar{s}$ such that $\hat{Q}_F(\bar{s}) = 1$ and a set $S \in \mathcal{M}(V)$ with $P_V(S) > 0$ such that $s' \in S$ implies $\hat{Q}_F(s') = 1$ and $\bar{p}_F(s') > \bar{p}_F(\bar{s})$.

Consider a manipulation mechanism with $\mu(t) = t$, for all t, i.e. all agents participate, Hence the coalition organizer learns the true distribution of payoff types, s, from the reports of individuals. Whenever $s \in S$ he chooses the announcements for individuals such that the cross section distribution of announced payoff types equals $\bar{s}$ with probability 1. Otherwise he reports the types of individuals truthfully to the overall mechanism. Obviously under the manipulation mechanism all individuals are strictly better off. Hence, a contradiction to the assumption that F is implementable as a coalition-proof equilibrium. Likewise we show that there cannot exist $\bar{s}'$ such that $\hat{Q}_F(\bar{s}') = 1$ and a set S'' with $P_V(S'') > 0$ such that $s'' \in S''$ implies $\hat{Q}_F(s'') = 1$ and $\pi_F(s'') > \pi_F(\bar{s}')$. ■

A.5 Proof of Proposition 5

The proof proceeds in two steps. We first show (Lemma 4) that every type t with $\tau(t) \in C^0$ (C^1) benefits from a manipulation with support C^0 (C^1) that attempts to minimize (maximize) the probability of public good provision. We then show in a second step (Lemma 5), that there is no beneficial manipulation for some set of types C contained in C^0 (C^1) if the probability of public good provision is minimized (maximized) conditional on observing $\delta \circ \mu_0^{-1}$ ($\delta \circ \mu_1^{-1}$).

Lemma 4 *Consider a manipulation for individuals with types in C, i.e., $\mu^{-1}(T) = C$. Suppose that C can be partitioned into the subsets C' and C''. For any manipulation of messages g, so that for any δ,*

$$E[\hat{Q}_F((g(\delta \circ \mu^{-1}), \delta \circ \nu^{-1}) \circ \tau^{-1}) \mid \delta \circ \mu^{-1}] \leq E[\hat{Q}_F((\delta_C, \delta_{-C}) \circ \tau^{-1}) \mid \delta \circ \mu^{-1}] \,, \tag{17}$$

we have that, for every δ,

$$E[\hat{Q}_F((g(\delta \circ \mu^{-1}), \delta \circ \nu^{-1}) \circ \tau^{-1}) \mid \delta \circ \mu_{C'}^{-1}] \leq E[\hat{Q}_F((\delta_C, \delta_{-C}) \circ \tau^{-1}) \mid \delta \circ \mu_{C'}^{-1}] \,, \tag{18}$$

where $\mu_{C'}$ is a function so that $\mu_{C'}(t) = t$ if $t \in C'$ and $\mu_{C'}(t) = \emptyset$ otherwise.

Proof The claim follows from a straightforward application of the law of iterated expectations. To see this, note that the inequalities in (18) can be equivalently written as

$$E\Big[E[\hat{Q}_F((g(\delta \circ \mu^{-1}), \delta \circ \nu^{-1}) \circ \tau^{-1}) \mid \delta \circ \mu^{-1}] \mid \delta \circ \mu_{C'}^{-1}\Big]$$

$$\leq E\Big[E[\hat{Q}_F((\delta_C, \delta_{-C}) \circ \tau^{-1}) \mid \delta \circ \mu^{-1}] \mid \delta \circ \mu_{C'}^{-1}]\Big] \,.$$

Hence, (18) is implied by (17). ■

Lemma 4 implies that whenever there exists some manipulation mechanism g for types in C^0 achieves a reduction in the probability of public good provision – relative to the outcome under the truthful announcement strategy g_0^*, then this manipulation makes all individuals with types in C^0 strictly better off.

The following Lemma shows that if a manipulation by all types in C^0 is unable to achieve a reduction in the probability of public good provision, then the same holds true for any manipulation by types in a subset C' of C^0.

Lemma 5 *If there is no manipulation that is beneficial for individuals with types in C^0, then there is no manipulation that is beneficial for types in C', where C' is a strict subset of C^0. Likewise, if there is no manipulation that is beneficial for individuals with types in C^1.then there is no manipulation that is beneficial for types in C'', where C'' is a strict subset of C^1.*

Proof We show that if there is a beneficial manipulation for types in C', then there is also a beneficial manipulation for types in C^0.

Suppose there is a beneficial manipulation for types in $C' \subset C^0$; i.e., there is δ and a strategy $g_{C'}$ such that

$$E[\hat{Q}_F((g_{C'}(\delta \circ \mu_{C'}^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1}) \mid \delta \circ \mu_{C'}^{-1}] < E[\hat{Q}_F(\delta \circ \tau^{-1}) \mid \delta \circ \mu_{C'}^{-1}] \,,$$

where $\mu_{C'}$ is a function so that $\mu_{C'}(t) = t$ if $t \in C'$ and $\mu_{C'}(t) = \emptyset$ otherwise and $\nu(t) = t$, for all $t \in T \setminus C'$. By the law of iterated expectations, this inequality can be equivalently written as

$$E\Big[E[\hat{Q}_F((g_{C'}(\delta \circ \mu_{C'}^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1}) \mid \delta \circ \mu_{C^0}^{-1}] \mid \delta \circ \mu_{C'}^{-1}\Big]$$

$$< E\Big[E[\hat{Q}_F(\delta \circ \tau^{-1}) \mid \delta \circ \mu_{C^0}^{-1}] \mid \delta \circ \mu_{C'}^{-1}\Big] \,.$$

This implies that there exist values of $\delta \circ \mu^{-1}$ for which this manipulation is beneficial for all types in C^0. Moreover, truthful reporting to the overall mechanism for types in $C^0 \setminus C'$ and following $g_{C'}$ for individuals in C' is a possible manipulation mechanism for all individuals with types in C^0. ■

A.6 Proof of Proposition 6

The "only if"-part is trivial. Hence, it remains to be shown that efficiency in the set of provision rules satisfying (3) and(4) implies efficiency in the set of coalition-proof provision rules, given p_F^0 and p_F^1.
Obviously, if provision rule $\hat{Q}_F$ is efficient in the set of provision rules satisfying (3) and(4), then there is no beneficial manipulation of $\hat{Q}_F$ by the grand coalition of all types $C = C^0 \cup C^1$.

We now establish that there is also no beneficial manipulation of $\hat{Q}_F$ with support $C = C^{0'} \cup C_1$ where $C^{0'}$ is a strict subset of C^0. (A symmetric argument implies that is also no manipulation with support $C = C^0 \cup C^{1'}$ for $C^{1'} \subset C^1$.)

The fact that there is no manipulation by the grand coalition of all types implies that truth-telling solves the following optimization problem for individuals with types in C^0: For every $\delta \circ \mu_{C^0}^{-1}$, choose $g(\delta \circ \mu_{C^0}^{-1})$ in order to minimize $E[\hat{Q}_F((g(\delta \circ \mu_{C^0}^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1}) \mid \delta \circ \mu_{C^0}^{-1}]$

subject to the constraints that

$$E[\hat{Q}_F((\delta \circ \mu_{C^{0'}}^{-1}, \delta \circ \nu^{-1}) \circ \tau^{-1}) \mid t] \geq E[\hat{Q}_F(\delta \circ \tau^{-1}) \mid t] \,,$$

for all types, $t \in C_1$. Using the arguments in Lemma 5 once more implies that no subset of C^0 can improve on this outcome.
We finally have to show that there is no beneficial manipulation for types in $C = C^{0'} \cup C^{1'}$ where $C^{0'}$ is a strict subset of C^0 and $C^{1'}$ is a strict subset of C^1.

We already know that there is no beneficial manipulation for types in $C = C^{0'} \cup C_1$. This implies in particular, that truth-telling solves the following problem: for every $\delta \circ \mu_C^{-1}$, choose $g(\delta \circ \mu_C^{-1})$ in order to maximize $E[\hat{Q}_F((g(\delta \circ \mu_C^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1} \mid \delta \circ \mu_C^{-1}]$ subject to the constraints that

$$E[\hat{Q}_F((g(\delta \circ \mu_C^{-1}) + \delta \circ \nu^{-1}) \circ \tau^{-1} \mid t] \leq E[\hat{Q}_F(\delta \circ \tau^{-1}) \mid t] \,,$$

for all types $t \in C^{0'}$. By Lemma 5, no subset of C_1 can improve on this outcome. ■

A.7 Proof of Proposition 7

Proof of the "only if" part.

In Appendix B we study strictly competitive games under incomplete information. It is shown that such games do not possess Bayes-Nash equilibria in which, ex interim, both players choose a pure strategy unless one player has an action that is strictly dominant or an action that is strictly dominated. These observations are established in Proposition 13.
Coalition-proofness requires that truth-telling solves the optimization problems in (3) and (3), respectively. As a consequence, for the present setting, Proposition 13 has the following implication.

Corollary 2 *Let S_α be the set of states with $\delta(C^1) = \alpha$. We can write $S_\alpha = S_{1-\alpha}^0 \times S_\alpha^1$, where $S_{1-\alpha}^0$ is a set that contains all possible realizations of $\delta \circ \mu_0^{-1}$ with the property that $\delta \circ \mu_0^{-1}(C^0) = 1 - \alpha$. We define $S_{1-\alpha}^0$ in the analogous way. For any α, coalition-proofness implies that one of the following statements has to be true:*

i) There is $x \in S_{1-\alpha}^0$ such that either $\hat{Q}_F(x, y) = 0$ or $\hat{Q}_F(x, y) = 1$, for every $y \in S_\alpha^1$.

ii) There is $y \in S_\alpha^1$ such that either $\hat{Q}_F(x, y) = 0$ or $\hat{Q}_F(x, y) = 1$, for every $x \in S_{1-\alpha}^0$.

Lemma 6 *A provision rule $\hat{Q}_F$ with the property that there exist δ and δ' such that $\delta(C^1) = \delta'(C^1)$ and $\hat{Q}_F(\delta \circ \tau^{-1}) \neq \hat{Q}_F(\delta' \circ \tau^{-1})$ is not coalition-proof.*

Proof Assume there exist δ and δ' such that $\delta(C^1) = \delta'(C^1) = \alpha$, $\hat{Q}_F(\delta \circ \tau^{-1}) = 0$ and $\hat{Q}_F(\delta' \circ \tau^{-1}) = 1$.

Suppose that the first statement in the corollary is true. Moreover, assume that there is $x \in S_{1-\alpha}^0$ such that $\hat{Q}_F(x, y) = 0$, for every $y \in S_\alpha^1$. Then, a manipulation mechanism for C^0

such that whenever $\delta \in S_\alpha$, $g(\delta \circ \mu_0^{-1}) = x$ with probability 1 (and truth-telling otherwise) is beneficial. Hence, a contradiction to coalition-proofness.
Now suppose that there exists $x \in S^0_{1-\alpha}$ such that $\hat{Q}_F(x, y) = 1$, for every $y \in S^1_\alpha$. Then, whenever $\delta \circ \mu_0^{-1} = x$ announcing instead $g(\delta \circ \mu_0^{-1}) = \delta' \circ \mu_0^{-1}$ (and truth-telling, otherwise), is beneficial for C^0. Hence, a contradiction to coalition-proofness.
Likewise one arrives at a contradiction to coalition-proofness if the second statement in the corollary is true. ■

Lemma 7 *A provision rule $\hat{Q}_F$ with the property that for some α_1 and α_2 with $\alpha_2 > \alpha_1$ there exist $\delta \in S_{\alpha_1}$ and $\delta' \in S_{\alpha_2}$ such that $\hat{Q}_F(\delta \circ \tau^{-1}) > \hat{Q}_F(\delta' \circ \tau^{-1})$ is not coalition-proof.*

Proof We know from Lemma 6 that, under coalition-proofness, the provision level is constant across states with the same population shares. If $\hat{Q}_F(\delta \circ \tau^{-1}) > \hat{Q}_F(\delta' \circ \tau^{-1})$, then C^0 will report a population share of α_1 if the true population share is α_2 or C^1 will report a population share of $1 - \alpha_2$ if the true population share is $1 - \alpha_1$. ■

Proof of the "if" - part.

We show that there does not exist a manipulation that is beneficial for a subset of types C' included in the set C^0. (A symmetric argument shows that there does not exist a manipulation that is supported only by a subset of C^1.)
Note first that a manipulation with support $C' \subset C^0$, has no impact on public good provision if it is such that for all δ, $\int_C dg(\delta \circ \mu^{-1})[t] = \delta \circ \mu^{-1}(C)$, i.e., a reshuffling population shares among types who oppose public good provision does not affect the decision on provision. Hence, such a manipulation is not beneficial. Moreover, a manipulation which is such that mass is shifted to types who are in favor of public good provision, $\int_C dg(\delta \circ \mu^{-1})[t] < \delta \circ \mu^{-1}(C)$, increases the probability of public good provision and is hence not beneficial either. ■

A.8 Proof of Proposition 8

See the proof of the "if" - part of Proposition 7. ■

A.9 Proof of Proposition 11

The arguments in the body of the text establish that Assumption 1 is both necessary and sufficient for the possibility to characterize the optimal social choice function as the solution of a relaxed problem that ignores the monotonicity constraint.

We show in the following that if Assumption 1 holds, then Assumption 2 implies that a welfare maximizing choice of p^0_F and p^1_F is such that $p^0_F = 0$ and $p^1_F = k$, i.e., there is no slack in the feasibility constraints.

Obviously, an optimal choice of p^1_F and p^0_F requires that $p^1_F = k$, or that $p^0_F = 0$. Otherwise it would be possible to decrease both p^1_F and p^0_F without violating the feasibility constraint and

without changing the sets V^0 and V^1. Hence, the provision rule for the public good would not need to be adjusted. Obviously the resulting social choice function would lead to higher welfare.

Now suppose that a social choice function has $p_F^0 = 0$ and $p_F^1 > k$. We show in the following that it is possible to decrease p_F^1 and to achieve a higher level of welfare if Assumptions 1 and 2 are satisfied.

Let $k \leq \bar{p}_F^1 < p_F^1$. A decrease of p_F^1 implies that more individuals prefer provision over non-provision, $V^1(0, \bar{p}_F^1)$ contains more payoff types than $V^1(0, p_F^1)$, and that $\hat{x}(\bar{p}_F^1) \leq \hat{x}(p_F^1)$ decreases. This has the following implications:

(a) Whenever it was optimal to provide the public good with p_F^1 it is also optimal to provide the public good with $\bar{p}_F^1$. Hence, in all states s such that $\hat{Q}_F(s) = 1$ is optimal under p_F^1, moving from p_F^1 to $\bar{p}_F^1$ leads ex post to higher welfare.

(b) In the states where $\hat{Q}_F(s) = 0$ is optimal given p_F^1 and given $\bar{p}_F^1$, welfare is unaffected.

(c) Consider the set of states where $\hat{Q}_F(s) = 0$ is optimal given p_F^1 and $\hat{Q}_F(s) = 1$ is optimal given $\bar{p}_F^1$. Given p_F^1 expected welfare equals zero in these states. By a revealed preferences argument, with $\bar{p}_F^1$, expected welfare is non-negative. To see this, recall that by Assumption 1 monotonicity constraints may be ignored. Consequently, $\hat{Q}_F(s) = 0$ remains an admissible choice for this set of states.

Consequently, a decrease of p_F^1 leads to an increase of welfare. The same arguments can be used to establish that $p_F^1 = k$ and $p_F^0 > 0$ cannot be optimal.

Finally, Example 3 in the body of the text establishes the "only if"-part of the Proposition.

■

B Strictly competitive games under incomplete information

This section of the appendix contains a result on strictly competitive games under incomplete information. We show that the characterization of Nash-equilibria under complete information for this class of games can be extended to games of incomplete information.

In the following we discuss games that differ with respect to their information structure but have the following properties in common. There are two players denoted by C_0 and C_1. The action set for player C_0 is given by $\{1, \ldots, m\}$. A typical action of C_0 is denoted by i. The action set for C_1 is $\{1, \ldots, n\}$ with a typical action denoted by k.

If C_0 chooses action i and C_1 chooses action k, then the outcome of the game is denoted by $Q_{ik} \in \{0, 1\}$. The strictly competitive structure arises because of the following property: Whenever C_0 strictly prefers outcome Q_{ik} over outcome Q_{jl}, then C_1 strictly prefers outcome Q_{jl}. Without loss of generality (see Osborne and Rubinstein (1994)) we represent these preferences by the following utility functions: C_0 evaluates outcomes according to u_0. This function is such that $u_0(Q_{ik}) = 0$ if $Q_{ik} = 0$, and $u_0(Q_{ik}) = -1$ if $Q_{ik} = 1$. C_1 has utility function u_1 which is such that $u_0(Q_{ik}) = -u_1(Q_{ik})$ for all possible outcomes .

We impose the assumption that neither player has an action that ensures his preferred outcome irrespective of the opponent's behavior:

Assumption 3

i) for each i, there exist k and l such that $Q_{ik} = 0$ and $Q_{il} = 1$.

ii) for each k, there exist i and j such that $Q_{ik} = 0$ and $Q_{jk} = 1$.

The complete information game

In the complete information game, a (mixed) strategy for player C_0 is a list $\alpha = (\alpha_1, \ldots, \alpha_m)$ that assigns a probability weight to each action in $\{1, \ldots, m\}$. The set of strategies for C_0 is hence given by the $m-1$ dimensional unit simplex, denoted by Δ^{m-1}. Likewise a strategy for C_1 is denoted by $\beta = (\beta_1, \ldots, \beta_n)$ and belongs to Δ^{n-1}.

A strategy pair (α, β) induces the following expected payoffs: For C_0 the expected payoff equals

$$U_0(\alpha, \beta) := \sum_{i=1}^{m} \sum_{k=1}^{n} \alpha_i \beta_k u_0(Q_{ik}) \; .$$

Likewise, the payoff for C_1 is,

$$U_1(\alpha, \beta) := \sum_{i=1}^{m} \sum_{k=1}^{n} \alpha_i \beta_k u_1(Q_{ik}) \; .$$

Obviously, for any (α, β), $U_0(\alpha, \beta) = -U_1(\alpha, \beta)$. Hence, the complete information game is strictly competitive.

In the subsequent derivations we will make repeated use of the following properties of this game.

Lemma 8

i) If (α^*, β^*) constitutes Nash-equilibrium of the complete information game, then

$$\alpha^* \in argmax_{\alpha \in \Delta^{m-1}} [min_{\beta \in \Delta^{n-1}} U_0(\alpha, \beta)] \quad and$$

$$\beta^* \in argmax_{\beta \in \Delta^{n-1}} [min_{\alpha \in \Delta^{m-1}} U_1(\alpha, \beta)] \quad .$$

ii) If (α^*, β^*) constitutes Nash-equilibrium of the complete information game, then

$$max_\alpha min_\beta \, U_0(\alpha, \beta) \;=\; min_\beta max_\alpha \, U_0(\alpha, \beta) \;=\; U_0(\alpha^*, \beta^*) \; .$$

In particular, all Nash equilibria of the complete information game yield the same payoffs.

For a proof see Proposition 22.2 in Osborne and Rubinstein (1994).

The incomplete information version of this game

We consider the following incomplete information extension of the complete information game: For each player, there is a set of possible types. The set of types for C_0 is given by $\{t_0^1, \ldots, t_0^q\}$ and the set of types for C_1 is $\{t_1^1, \ldots, t_1^r\}$. Consequently, the set of states is given by $\{t_0^1, \ldots, t_0^q\} \times \{t_1^1, \ldots, t_1^r\}$. Denote by p_{jl} the common prior probability that player C_0 has type t_0^j and player C_1 has type t_1^l.
For any type of player C_0, his action set is equal to $\{1, \ldots, m\}$. Hence, in the incomplete information game, a strategy for player C_0 is a list $s_0 = (\alpha^1, \ldots, \alpha^q)$ from $(\Delta^{m-1})^q$, where α^j is the mixed strategy that C_0 chooses with type t_0^j. Likewise a strategy for C_1 is denoted by $s_1 = (\beta^1, \ldots, \beta^r)$ and belongs to $(\Delta^{n-1})^r$.
Ex ante expected payoffs in the incomplete information game induced by a pair of strategies s_0, s_1 are given by

$$EU_0(s_0, s_1) := \sum_{j=1}^{q} \sum_{l=1}^{r} p_{jl} U_0(\alpha^j, \beta^l) \; .$$

for C_0 and by

$$EU_1(s_0, s_1) := \sum_{j=1}^{q} \sum_{l=1}^{r} p_{jl} U_1(\alpha^j, \beta^l) \; .$$

for C_1. Note that the incomplete information game is strictly competitive.

Lemma 9 *Under Assumption 1, the complete information game has no Nash equilibrium in pure strategies.*

Lemma 10 (α^*, β^*) *is a Nash equilibrium of the complete information game if and only if* (s_0^*, s_1^*) *with* $s_0^* = (\alpha^*, \ldots, \alpha^*)$ *and* $s_1^* = (\beta^*, \ldots, \beta^*)$ *is a Nash equilibrium of the incomplete information game.*

Proof
"$\Longrightarrow$": Let α^* satisfy $U_0(\alpha^*, \beta^*) \geq U_0(\alpha, \beta^*)$ for all $\alpha \in \Delta^{m-1}$. Then, obviously, for all $(\alpha^1, \ldots, \alpha^m) \in (\Delta^{m-1})^q$

$$\sum_{j=1}^{q} \sum_{l=1}^{r} p_{jl} U_0(\alpha^*, \beta^*) \geq \sum_{j=1}^{q} \sum_{l=1}^{r} p_{jl} U_0(\alpha^j, \beta^*) \; .$$

Equivalently, for all s_0, $EU_0(s_0^*, s_1^*) \geq EU_0(s_0, s_1^*)$.
"$\Longleftarrow$": Suppose that, for all $(\alpha^1, \ldots, \alpha^m) \in (\Delta^{m-1})^q$

$$\sum_{j=1}^{q} \sum_{l=1}^{r} p_{jl} U_0(\alpha^*, \beta^*) \geq \sum_{j=1}^{q} \sum_{l=1}^{r} p_{jl} U_0(\alpha^j, \beta^*) \; .$$

This implies that it has to be true that for all j, and for all $\alpha^j \in \Delta^{m-1}$,

$$U_0(\alpha^*, \beta^*) \sum_{l=1}^{r} p_{jl} \geq U_0(\alpha^j, \beta^*) \sum_{l=1}^{r} p_{jl} \ .$$

Equivalently, for all j, and for all $\alpha^j \in \Delta^{m-1}$, $U_0(\alpha^*, \beta^*) \geq U_0(\alpha^j, \beta^*)$. ■

Under $s_0^* = (\alpha^*, \ldots, \alpha^*)$ and $s_1^* = (\beta^*, \ldots, \beta^*)$ equilibrium expected payoffs are given by

$$EU_0(s_0^*, s_1^*) := U_0(\alpha^*, \beta^*) \quad \text{and} \quad EU_1(s_0^*, s_1^*) := U_1(\alpha^*, \beta^*) \ .$$

Lemma 11 *It follows from Observation 10 and from part ii) of Lemma 8 that all Nash equilibria of the incomplete information game generate these expected payoffs.*

For any strategy pair (s_0, s_1), denote ex interim expected payoffs by

$$EU_0(s_0, s_1 \mid t_0^j) = \frac{1}{\sum_{l=1}^{r} p_{jl}} \sum_{l=1}^{r} p_{jl} U_0(\alpha^j, \beta^l) \ ,$$

and

$$EU_1(s_0, s_1 \mid t_1^l) = \frac{1}{\sum_{j=1}^{l} p_{jl}} \sum_{j=1}^{q} p_{jl} U_1(\alpha^j, \beta^l) \ .$$

Under $s_0^* = (\alpha^*, \ldots, \alpha^*)$ and $s_1^* = (\beta^*, \ldots, \beta^*)$ equilibrium expected payoffs ex interim satisfy

$$EU_0(s_0^*, s_1^* \mid t_0^j) := U_0(\alpha^*, \beta^*) \quad \text{and} \quad EU_1(s_0^*, s_1^* \mid t_1^l) := U_1(\alpha^*, \beta^*) \ .$$

for all j and all l, respectively.
The following observation establishes that any Nash-equilibrium of the incomplete information game generates these expected payoff levels ex interim.

Lemma 12 *Let $(\hat{s}_0, \hat{s}_1)$ be a Nash equilibrium of the incomplete information game. Then*

$$EU_0(\hat{s}_0, \hat{s}_1 \mid t_0^j) := U_0(\alpha^*, \beta^*) \quad \textit{and} \quad EU_1(\hat{s}_0, \hat{s}_1 \mid t_1^l) := U_1(\alpha^*, \beta^*) \ .$$

for all j and all l, respectively.

Proof Denote $V_0(\alpha) := \min_{\beta \in \Delta^{n-1}} U_0(\alpha, \beta)$. It follows from Observation 1 and part i) of Lemma 8 that

$$\alpha^* = \text{argmax}_{\alpha \in \Delta^{m-1}} V_0(\alpha) \ ,$$

Step 1. In the complete information game, for any β, by choosing α^*, C_0 can ensure a payoff of at least $V_0(\alpha^*)$. Formally, $\forall \beta : U_0(\alpha^*, \beta) \geq V_0(\alpha^*)$. Suppose otherwise, then there exists $\hat{\beta}$ such that

$$U_0(\alpha^*, \hat{\beta}) < V_0(\alpha^*) = \min_{\beta} U_0(\alpha^*, \beta) \ .$$

Hence, a contradiction.

Step 2. One has $U_0(\alpha^*, \beta^*) \leq V_0(\alpha^*)$. To see this, suppose to the contrary that $U_0(\alpha^*, \beta^*) > V_0(\alpha^*)$. Consequently, there exists β' such that

$$U_0(\alpha^*, \beta^*) > U_0(\alpha^*, \beta') \, .$$

Using that the game is strictly competitive this implies that

$$U_1(\alpha^*, \beta') > U_1(\alpha^*, \beta^*) \, .$$

But this contradicts that (α^*, β^*) is a Nash equilibrium of the complete information game.

Step 3. Now suppose that there exists j such that

$$EU_0(\hat{s}_0, \hat{s}_1 \mid t_0^j) \neq U_0(\alpha^*, \beta^*)$$

By Observation 11 ex ante expected payoffs are

$$EU_0(\hat{s}_0, \hat{s}_1) := U_0(\alpha^*, \beta^*) \quad \text{and} \quad EU_1(\hat{s}_0, \hat{s}_1) := U_1(\alpha^*, \beta^*) \, .$$

Hence there must exist k such that

$$EU_0(\hat{s}_0, \hat{s}_1 \mid t_0^k) < U_0(\alpha^*, \beta^*)$$

Now suppose that instead of playing according to $\hat{s}_0$, C_0 with type t_0^k chooses α^*. By *Step 1* this yields an interim expected payoff of at least $V_0(\alpha^*)$. By *Step 2* this payoff exceeds $U_0(\alpha^*, \beta^*)$; i.e. C_0 with type t_0^k can do better than playing according to $\hat{s}_0$. This implies that $(\hat{s}_0, \hat{s}_1)$ cannot be a Nash equilibrium of the incomplete information game. Hence, a contradiction. ■

Lemma 13 *In the incomplete information game, in each state, at least one player randomizes ex interim.*

Proof By observation 12 and part i) of Lemma 8 it follows that, in every state, the strategy chosen by C_0 solves $\max_\alpha \min_\beta U_0(\alpha, \beta)$ and the strategy chosen by C_1 solves $\max_\beta \min_\alpha U_1(\alpha, \beta)$. If these strategies were both pure strategies, then the complete information game would possess a Nash-equilibrium in pure strategies. Hence, a contradiction to Observation 9. ■

Lemmas 9 to 13 establish the following proposition.

Proposition 13 *Consider the strictly competitive game of incomplete information. Under Assumption 3, any Bayes-Nash equilibrium of this game has the property that each type of each player randomizes.*