Bierbrauer, Felix; Hellwig, Martin

Working Paper

Mechanism design and voting for public-good provision

Preprints of the Max Planck Institute for Research on Collective Goods, No. 2011,31

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Suggested Citation: Bierbrauer, Felix; Hellwig, Martin (2011) : Mechanism design and voting for public-good provision, Preprints of the Max Planck Institute for Research on Collective Goods, No. 2011,31, Max Planck Institute for Research on Collective Goods, Bonn

This Version is available at:
https://hdl.handle.net/10419/57474

**Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2011/31**

Mechanism Design and Voting for Public-Good Provision

Felix Bierbrauer
Martin Hellwig

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2011/31

Mechanism Design and Voting for Public-Good Provision

Felix Bierbrauer / Martin Hellwig

December 2011

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

Mechanism Design and Voting for Public-Good Provision*

Felix J. Bierbrauer† and Martin F. Hellwig‡

December 2, 2011

Abstract

We propose a new approach to the normative analysis of public-good provision. In addition to individual incentive compatibility, we impose conditions of robust implementability and coalition proofness. Under these additional conditions, participants' contributions can only depend on the level of public-good provision. For a public good that comes as a single indivisible unit, provision can only depend on the population share of people in favour of provision. Robust implementability and coalition proofness thus provide a foundation for the use of voting mechanisms. The analysis is also extended to a specification with more than two public-good provision levels.

Keywords: Public-good provision, Mechanism Design, Voting Mechanisms, Large Economy

JEL: D60, D70, D82, H41

*We are grateful for discussions with and comments from Alia Gizatulina, Mike Golosov, Kristoffel Grechenig, Christian Hellwig, David Martimort, Benny Moldovanu, and Nora Szech, as well as an editor and three referees of this journal.

†University of Cologne, Center for Macroeconomic Research, Albertus-Magnus-Platz, 50923 Köln, Germany. Email: bierbrauer@wiso.uni-koeln.de

‡Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, 53113 Bonn, Germany. Email: hellwig@coll.mpg.de

1 Introduction

In allocation theory and public economics, the theory of public-good provision stands out as a monolith that is hardly related to anything else. It is an essential part of the curriculum, but the relation between the allocation mechanisms that are provided by the theory and the allocation mechanisms that are used in the real world is hardly ever discussed. The relation of public-good provision theory to other parts of allocation theory, the theory of taxation, and political economy is also hardly ever considered.

This state of affairs reflects two particular features of the theory as it appears in our curriculum.[1] First, the analysis focuses largely on questions of individual incentive compatibility. Second, the analysis focuses on "small" rather than "large" economies. The key question is how to calibrate people's payments to their expressions of preferences so that they have no wish either to understate their preferences for the public good (so as to reduce their payments) or to overstate their preferences (so as to get a greater provision level at other people's expense).[2] For this question to be nontrivial, each person must have a distinct chance of being "pivotal", i.e., of having a noticeable effect on the level of public-good provision through the expression of her preferences.

This small-economy approach to public-good provision stands in marked contrast to the way we deal with problems involving private goods. For private goods, the large-economy paradigm, where no one person is able to affect market prices, is deemed to provide the proper framework for studying what happens when there are millions of people and none of them individually has market power. This paradigm serves as a conceptual idealization and as a normative standard for assessing real-world markets.

In our view, the large-economy paradigm should also have a central place in public-good provision theory. The formulation of the public-good provision problem that we have in our curriculum may be appropriate for studying how people in a condominium may decide on how much to spend on maintenance and gardening. It is not appropriate, however, for studying how a society with millions of people can decide on how much to spend on national defense or on the judicial system. For this latter question, we need a large-economy model, where no one person individually is able to affect the level of aggregate per-capita spending on the public good in question.

In a large economy, the problem of mechanism design for individual incentive compatibility is trivial. Because no one person individually is able to affect the level of public-good provision, no one is ever "pivotal". For individual incentive compatibility, it therefore suffices to have payments that are independent of what people say. If the preferences that a person expresses have no effect on either public-good provision or the payments that the person has to make, she may as well report her preferences truthfully. Given that preferences are reported truthfully, there is no problem about implementing an efficient provision rule for the public good. Participation in

[1] See, for instance, Fudenberg and Tirole (1991), Mas-Colell et al. (1995), or Hindriks and Myles (2006).

[2] 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.

the system may not be voluntary, but there is no problem of incentive compatibility.[3]

In our opinion, the analysis must not stop here. We believe that, in addition to individual incentive compatibility, we must be concerned with issues of coalition proofness and of robustness of allocation mechanisms.

We illustrate our concerns by means of an 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 α a value of 3, and a fraction $\frac{7}{10} - \alpha$ 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 $\alpha > \frac{1}{3}$ and should not be provided if $\alpha < \frac{1}{3}$. 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 α is common knowledge, this reasoning is unproblematic.[4] By contrast, if α is the realization of a nondegenerate random variable $\tilde{\alpha}$, 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 $\tilde{\alpha} > \frac{1}{3}$ and that the public good should be provided.

At this point, we are bothered by 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.[5] We find this answer unconvincing. We are therefore led to the conclusion that individual incentive compatibility should not be the only requirement for public-good provision mechanisms.

We propose to impose requirements of *robustness* and *coalition proofness* in addition to in-

[3] 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.

[4] 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. However, if α 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.

[5] As we show in Section 9, the same problem arises in the finite economy.

dividual incentive compatibility. *Robustness* requires that the public-good provision mechanism should not depend on the beliefs that participants have about the other participants. *Coalition proofness* requires that there should be no scope for a group of participants to form a coalition in order to co-ordinate their reports so as to influence the aggregate outcome in a way that makes some coalition members better off and no coalition member worse off. In imposing robustness, we follow Ledyard (1978) and Bergemann and Morris (2005), in imposing coalition proofness, Laffont and Martimort (1997, 2000). In Section 2 below, we explain in more detail why we believe that these are the right conditions to impose.

In the above example, the mechanism that we used for first-best implementation is robust but not coalition-proof. Robustness is implicit in the observation that neither the rule for determining public-good provision nor the rule for determining people's payments depend on the specification of beliefs that people have about each other (for a mechanism that is non-robust, see Section 2.2 below). Coalition proofness, however, fails because a coalition of people who value the public good at either 0 or 3 could co-ordinate their reports so that the fraction of people reporting 3 would always be less than $\frac{1}{3}$ and the public good would not be provided, making all of them better off. Moreover, in contrast to a cartel trying to eliminate competition among its members, such a coalition would not have an incentive compatibility problem of its own; the reports that it recommends to its members would all be individually incentive-compatible.

In Sections 3 and 4 of the paper, we will specify these conditions formally and explore their implications. For a public good that comes as a single indivisible unit, we will show that any anonymous public-good provision mechanism that satisfies these additional requirements must take the form of a voting mechanism, i.e., a mechanism that asks people to vote for or against the provision of the public good, and that makes the provision rule condition on the shares of votes for the two alternatives.

Economists have traditionally been critical of voting mechanisms because they fail to take account of preference intensities. If there are many people opposing the provision of the public good and few people promoting it, a voting mechanism will stipulate non-provision, which is sub-optimal if the proponents could draw very large benefits from the public good, and the opponents do not feel very strongly about the matter. Our analysis shows that this criticism is irrelevant if public-good provision mechanisms must be coalition-proof and robust as well as anonymous. Mechanisms that take account of preference intensities necessarily violate one of these conditions.

When we refer to voting mechanisms, we are not assuming that voting must be governed by the majority rule. If, at the stage of mechanism design, there is prior information that beneficiaries of the public good feel strongly about it and opponents do not, it may be desirable to have a rule by which the public good is already provided if a sufficiently large minority is in favour. Majority voting is likely to be desirable if there is no such prior information which would permit a discrimination of alternatives at the mechanism design stage.

The reasoning underlying our analysis is fairly straightforward. In a large economy, robustness implies that payments must be independent of individual types. In any situation, therefore, payments must be the same for all agents (Proposition 1). Coalition proofness then implies that payments depend only on the level of public-good provision; if they were conditioned on some-

thing else, then, in some circumstances, there would be room for the grand coalition of all agents to co-ordinate and manipulate reports so as to lower the required payments without changing the level of public-good provision.

Given that payments depend only on public-good provision, there is a natural specification of winners and losers from public-good provision, those participants for whom the benefits from the enjoyment of the public good exceed the difference between payment levels in the events of provision and non-provision and those participants for whom the benefits from the enjoyment of the public good fall short of this difference. These two groups define the key coalitions to consider in assessing whether a provision rule for the public good is coalition-proof. For coalition proofness, the level of public-good provision must be a non-decreasing function of the population share of the group of net beneficiaries. If the decision on public-goods provision was made dependent on preference intensities, then the decision would be vulnerable to manipulations by a deviating coalition.

Our results also imply that, apart from exceptional circumstances, it is not possible to implement a first-best provision rule if coalition proofness and robustness requirements are imposed. By contrast to previous impossibility results, this finding does not involve participation constraints or a multi-dimensional information problem. It follows directly from the observation that coalition-proofness and robust incentive compatibility together destroy the possibility of conditioning on intensities of preferences.

In the following, Section 2 provides some additional motivation for the requirements of coalition proofness and robustness that we impose. Section 3 presents our formal model and introduces the notion of robust implementation. Section 4 introduces the notion of coalition proofness. Section 5 gives our main result, i.e., the characterization of robust and coalition-proof public-goods provision in a large economy. Section 7 discusses the welfare implications of this characterization. Section 9 establishes that our main result extends to an economy with finitely many individuals. In Section 8, we extend our analysis and allow for more than two possible provision levels of the public good. The last section contains concluding remarks. All proofs are in the Appendix.

2 Why Coalition-Proofness and Robustness?

In this section, we explain why the requirements of coalition proofness and robustness are germane to the example that we have presented above and why an approach that relies on these requirements is to be preferred to alternative approaches that might also deal with the problem raised by the example.

2.1 Coalition Proofness

As mentioned, coalition proofness makes a difference in the example because a coalition of people who value the public good at either 0 or 3 can manipulate the fraction of reports with valuation 3 that are received by the mechanism. If they co-ordinate their reports so that the fraction of people reporting 3 is always less than $\frac{1}{3}$, they can prevent the public good from being provided,

which makes all of them better off. By imposing coalition proofness, we thus address the concern that participants who value the public good at 3 are willing to report their valuation truthfully only because, individually, they feel that their reports do not matter anyway, when, collectively, truthtelling may make all of them worse off.

This concern could also be addressed at the level of individual decision making, by imposing additional conditions on the way in which people resolve their indifference between different reports. In our example, an agent who values the public good at 3 might consider that, even though, with probability one, her report does not make a difference, in the probability zero event where she might be pivotal, it would better to have reported the valuation 0 because this would contribute to not having the public good provided. There is thus a sense in which truthtelling is weakly dominated. The efficient provision rule with equal cost sharing is not robust to the elimination of strategies that are dominated in this sense.

Thus, it seems that a requirement of robustness to the elimination of weakly dominated strategies would achieve the same purpose as a requirement of coalition proofness.[6] By comparison to coalition proofness, this approach has the advantage that it does not transcend the level of individual decision making. It is therefore much simpler.

In the present context, however, a criterion of robustness to the elimination of weakly dominated strategies has the disadvantage that its application seems limited to economies that are large, in the strict sense that no one individual alone can influence the level of public-good provision. In a large finite economy, under a Clarke-Groves mechanism for public-good provision, truthtelling is a dominant strategy for every individual; truthtelling equilibria thus are robust to the elimination of dominated strategies. In the transition from large finite economies to the large-economy limit with a continuum of agents, however, the property of robustness of truthtelling equillibria to the elimination of weakly dominated strategies exhibits a discontinuity. Whereas this property holds in all finite economies, it fails in the large-economy limit where outcome functions are completely insensitive to the behavior of individuals. We consider this discontinuity in the transition from large finite economies to the large-economy limit to be problematic because, as a matter of principle, we think of large-economy models with a continuum of agents as a mathematical idealization (and simplification) of large finite economies. There should thus be some assurance that properties derived from the analysis of a large-economy model should approximately hold in a large finite economy as well.

By contrast to the dominance criterion, a requirement of coalition proofness meets this test. Whereas most of the analysis of this paper is carried out in a large-economy setting, in part B of the Appendix we show that our analysis extends to finite economies, i.e., in finite as well as large economies, the only social choice functions that satisfy robust implementability and coalition proofness are those that can be implemented by voting mechanisms.

A second reason for studying the implications of coalition proofness involves the comparison between private and public goods. For private goods, it is well known that, in a large economy in which every participant is insignificant relative to the aggregates, the set of competitive-

[6] In political economy, this assumption is often referred to as *sincere voting*, i.e., it is assumed that people vote their preferences even though, as individuals, they do not expect their votes to have an effect on aggregate outcomes. For a clarification of this terminology, see, for instance, Austen-Smith and Banks (1996).

equilibrium allocations coincides with the set of core allocations, i.e., the set of allocations that cannot be blocked by any coalitions, i.e., coalition proofness imposes no restrictions on implementability. It is therefore of interest to observe that, for public goods, a requirement of coalition proofness substantially restricts the set of attainable allocations.

The requirement of coalition proofness that we impose will be formulated so as to take account of incentive constraints due to information asymmetries between the different participants of a coalition that might form. In this respect, we follow Laffont and Martimort (1997, 2000), who treated the problem of organizing a coalition whose members would co-ordinate their reports as an mechanism design problem of its own, with distinct incentive and participation constraints for all participants.[7] In the example above, it is easy to see that these constraints are satisfied, i.e., if a coalition of people who value the public good at either 0 or 3 proposes to co-ordinate reports so that the fraction of people reporting 3 is less than $\frac{1}{3}$, then this proposal is individually rational and incentive-compatible for all intended participants.

2.2 Robustness

In addition to coalition proofness, we impose a requirement of robustness along the lines of Ledyard (1978) and Bergemann and Morris (2005). The mechanism that is used to determine the level of public-good provision must not depend on the details of the stochastic specification of the model. Individual incentive compatibility must be robust to changes in the specification of individuals' probabilistic beliefs.

This robustness requirement eliminates the possibility of using type-dependent differences in beliefs to support type-dependent payment rules. This possibility has been extensively discussed in the context of auctions with correlated values. In our setting, correlated values arise naturally if the question whether the public good should be provided or not is to involve a genuine information problem.[8] For there to be a genuine information problem, there must be some prior uncertainty about the aggregate valuation of the public good. If the aggregate valuation is derived from the cross-section distribution of individual valuations, there must be some correlation of individual and aggregate valuations. Given this correlation, individuals' beliefs will vary with their valuations. Robustness precludes the exploitation of this type dependence of beliefs for purposes of mechanism design.

To understand the issue, take another look at the example in the introduction. As before, a fraction $\frac{3}{10}$ of the population assigns a value of 10 to the public good, a fraction α a value of 3, and a fraction $\frac{7}{10} - \alpha$ a value of 0. We now assume that all agents regard α as the realization of a random variable with possible values $\alpha_H = \frac{6}{10}$ and $\alpha_L = \frac{2}{10}$. With a per-capita provision cost equal to 4, it is thus efficient to provide the public good if $\alpha = \alpha_H$ and not to provide it if $\alpha = \alpha_L$.

Whereas, before, we had assumed that all agents make the same payments, we now consider a payment rule which has agents' payments depend on their public-good valuations. For $v \in$

[7] Laffont and Martimort, however, focussed on deviations by the grand coalition of all agents. In contrast, we shall mainly be concerned with smaller coalitions involving specified subsets of agents, in particular, coalitions of agents who favour or oppose proposals for public-good provision.

[8] For a general discussion of this point, see Bierbrauer and Hellwig (2011).

$\{0, 3, 10\}$, let $P_H(v)$ and $P_L(v)$ be the payments that an agent with public-good valuation v is required to make if $\alpha = \alpha_H$ and if $\alpha = \alpha_L$. Feasibility requires that, on aggregate, these payments add up to the cost of public-good provision, i.e., that

$$\left(\frac{7}{10} - \alpha_H\right) P_H(0) + \alpha_H P_H(3) + \frac{3}{10} P_H(10) = 4$$

and

$$\left(\frac{7}{10} - \alpha_L\right) P_L(0) + \alpha_L P_L(3) + \frac{3}{10} P_L(10) = 0.$$

With $\alpha_H = \frac{6}{10}$ and $\alpha_L = \frac{2}{10}$, these conditions are obviously satisfied if we set $P_H(0) = P_H(10) = 10, P_H(3) = 0$, $P_L(0) = P_L(10) = -2$, and $P_L(3) = 8$. For $\alpha = \alpha_H$, when the public good is provided, its cost is entirely borne by people who value the public good at 0 or 10. People who value it at 3 contribute nothing. However, these people are required to make payments - and the other participants receive payments - when $\alpha = \alpha_L$ and the public good is not provided.

Can such a payment rule be incentive-compatible? The answer depends on agents' beliefs. Because, individually, agents do not affect the level of public-good provision, each agent will try to minimize his expected payment. Let $\beta_H(v)$ and $\beta_L(v)$ denote the probabilities of the events $\alpha = \alpha_H$ and $\alpha = \alpha_L$ as assessed by an agent with public-good valuation v. The payment rule $(P_H(\cdot), P_L(\cdot))$ is incentive-compatible if

$$\beta_H(v) P_H(v) + \beta_L(v) P_L(v) \leq \beta_H(v) P_H(\hat{v}) + \beta_L(v) P_L(\hat{v}) \tag{1}$$

for all v and $\hat{v}$. For the payment rule specified above, one easily verifies that this incentive compatibility condition is fulfilled if $\beta_H(3) \geq \beta_L(3)$, $\beta_H(0) \leq \beta_L(0)$, and $\beta_H(10) \leq \beta_L(10)$. Beliefs satisfying these inequalities can be generated by Bayesian updating on a common prior if each agent's information about this own public-good valuation contains suitable information about α. For instance, if the prior assigns the probability $\frac{1}{2}$ to both α_H and α_L and if, conditional on α, the public-good valuation of any given individual i is equal to 0 with probability $\frac{7}{10} - \alpha$, 3 with probability α, and 10 with probability $\frac{3}{10}$, Bayesian updating yields $\beta_H(0) = \frac{1}{6}$, $\beta_H(3) = \frac{3}{4}$, $\beta_H(10) = \frac{1}{2}$, and, for each v, $\beta_L(v) = 1 - \beta_H(v)$. Individual incentive compatibility is satisfied.

With this payment rule, first-best public-good provision is also coalition-proof because people who value the public good at 3 are no longer averse to having the public good provided. They get a net payoff of 3 when the public good is provided and a net payoff of -8 when it is not provided. They are therefore unwilling to join any coalition that would reduce the incidence of public-goood provision. Without their co-operation, however, a coalition that would reduce the incidence of public-good provision cannot form. By the same argument, people who value the public good at 0 would not join any coalition that would increase the incidence of public-good provision, and, therefore, such a coalition cannot form.

In this example, incentive compatibility of a payment rule with type-dependent payments is supported by differences in beliefs. The logic of the argument is the same as in the analysis of Crémer and McLean (1985, 1988) showing that, generically, full surplus extraction can be obtained in models of auctions with correlated values. In our example, individual valuations

are correlated with α. This correlation induces a type dependence of preferences over outcome-contingent lotteries. This type dependence of preferences over lotteries provide a basis for having incentive mechanisms that are coalition-proof as well as feasible and incentive-compatible.

However, these properties are not robust. The specification of outcome-contingent lotteries in the payment rules must be precisely calibrated to the valuation-contingent beliefs $\beta_H(v)$ and $\beta_L(v)$ that people have. Unless payments are independent of public-good valuations, there always exist beliefs that violate (1) for some v and $\hat{v}$. Suppose for instance, that in the above example, the prior probabilities of α_H and α_L are $\frac{2}{3}$ and $\frac{1}{3}$, rather than $\frac{1}{2}$ and $\frac{1}{2}$. Then a person with valuation 10 will have beliefs $\beta_H(10) = \frac{2}{3}$, $\beta_L(10) = \frac{1}{3}$, the the specified payment rule will violate (1) for this person.

More generally, the incentive compatibility of the payment rules that are used in the Bayesian approach may be sensitive to details in the specification of priors and beliefs. Following Ledyard (1978) and Bergemann and Morris (2005), we consider it unreasonable to suppose that the mechanism designer has the information about participants' beliefs that he would need in order to calibrate precisely an incentive mechanism to these details. Imposing a requirement of robustness, we therefore require that incentive mechanisms under consideration should be incentive-compatible and feasible regardless of the specification of priors and beliefs.[9]

3 Robust Implementation in a Large Economy

3.1 Payoffs and Social Choice Functions

We consider an economy with a continuum of agents of measure 1.[10] There is one private good and one public good. The public good comes as a single indivisible unit. Its installation requires aggregate resources (per-capita) equal to k units of the private good.

Given a public-good provision level $Q \in \{0, 1\}$, the utility of any agent i is given as $v_i Q - P_i$, where v_i is the agent's valuation of the public good and P_i is his contribution to the cost of public-good provision. The valuation v_i belongs to a measurable space $(V, \mathcal{V})$ of possible valuations, which is the same for all i.

A social choice function determines under what conditions the public good is to be provided and what contributions are to be made by the different individuals. Following Guesnerie (1995), we impose an anonymity requirement by which the level of public-good provision as well as the payments of individuals with a given valuation v are unchanged under any permutation of individual characteristics that leaves the cross-section distribution of preferences unaffected. Thus, an anonymous social function determines how public-good provision levels and payment

[9] In Bierbrauer and Hellwig (2011), we also show that robustness eliminates the dichotomy between models with independent and with correlated values. Without robustness, for models with participation constraints, the Bayesian approach yields impossibility theorems for first-best implementation with independent values and possibility theorems for first-best implementation with correlated values. With robustness, we obtain impossibility theorems for first-best implementation with correlated as well as independent values. In either case, with participation constraints, if there are many participants, approximately first-best implementation is possible with private goods and hardly anything is possible with public goods.

[10] In Section 9, we extend the analysis to economies with finitely many agents.

rules depend on the cross-section distribution of preferences. We refer to the latter as the state of the economy. Formally, the state of the economy is an element s of the set $\mathcal{M}(V)$ of probability measures on $(V, \mathcal{V})$. An anonymous social choice function is a pair $F = (Q_F, P_F)$ of functions $Q_F : s \mapsto Q_F(s)$ and $P_F : (s, v) \mapsto P_F(s, v)$ such that, for any state of the economy s, $Q_F(s) \in \{0, 1\}$ is the level of public-good provision in the state s, and $P_F(s, \cdot)$ is a function indicating how, in state s, an agent's payment depends on the agent's valuation.

Anonymity is a requirement of equal treatment. Two individuals with the same characteristics have to make the same contribution to the cost of public-goods provision. In addition, the decision whether to provide the public good does not depend on the identity of the agents with certain preferences, but only on the cross-section distribution of those preferences in the economy as a whole.[11]

For any $s \in \mathcal{M}(V)$, the payment rule $P_F(s, \cdot)$ is taken to be integrable with respect to v. The integral $\int P_F(s, v)ds(v)$ corresponds to the aggregate revenue that is collected in state s. We say that the anonymous social choice function $F = (Q_F, P_F)$ yields feasible outcomes if and only if, in any state of the economy, the aggregate revenue is sufficient to cover the public-good provision cost $kQ_F(s)$, i.e., if and only if the inequality

$$\int_V P_F(s, v)ds(v) \geq kQ_F(s) \tag{2}$$

is satisfied for all $s \in \mathcal{M}(V)$.

3.2 Types and Beliefs

Information about types is assumed to be private. As usual, we model information by means of an abstract type space. Let $(T, \mathcal{T})$ be a measurable space, τ a measurable map from T into V, and β a measurable map from T into the space $\mathcal{M}(\mathcal{M}(T))$ of probability distributions over measures on T. We interpret $t_i \in T$ as the abstract "type" of agent i, $v_i = \tau(t_i)$ as the *payoff type*, i.e., the public-good valuation of agent i and $\beta(t_i)$ as the *belief type* of agent i.

The belief type $\beta(t_i)$ 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, $\beta(t_i)$ is a probability measure on the space $\mathcal{M}(T)$ of these cross-section distributions. For any event $X \subset \mathcal{M}(T)$, $\beta(X \mid t_i)$ is the probability that type t_i of agent i assigns to the event that the cross-section distribution of types δ belongs to the set X. We refer to the map $\beta : T \to \mathcal{M}(\mathcal{M}(T))$ as the *belief system* of the economy.[12]

[11] Anonymity is a substantive constraint. Using the idea of sampling, that has been developed by Green and Laffont (1979), Bierbrauer and Sahm (2010) show that first-best outcomes can be implemented by a procedure where public-good preferences are elicited from a representative sample of the population only. If payment rules differ between the members of the sample and the rest of the population, the payment rule for the sample can be used to provide proper incentives and the payment rule for the rest can be used to finance public-good provision. By contrast, first-best is out of reach if all individuals have the same influence on public-good provision and the payment rule is the same for all.

[12] We do not assume that the belief system is compatible with a common prior. Our robustness requirement is therefore stronger than it would be under such an assumption. As shown in Bierbrauer and Hellwig (2010), however, our analysis would be unchanged if we restricted ourselves to belief systems that are compatible with common priors. The existence and uniqueness of common priors for the given setup is discussed in Hellwig (2011)

A typical element of $\mathcal{M}(T)$ will be denoted by δ. We denote by $\theta(\delta) = \delta \circ \tau^{-1}$ the cross-section distribution of valuations associated with δ. For any subset V' of V we write $\theta(V' \mid \delta)$ for the mass of individuals that the distribution $\theta(\delta)$ assigns to payoff types in V'.

The belief system β is said to be *degenerate* if, for some $\delta \in \mathcal{M}(T)$ and all $t \in T$, the measure $\beta(t)$ assigns all probability mass to the singleton $\{\delta\}$ i.e., if all agents "know" the cross-section distribution of types to be δ. The degenerate belief system which assigns all probability mass to δ will be denoted as β_δ.

In addition to the general notion of an abstract type space $[(T, \mathcal{T}), \tau, \beta]$, we shall also make use of the special notion of a *naive* type space $[(V, \mathcal{V}), \beta_s]$. This is the special case of an abstract type space in which agents' types are given by their public-good valuations so that $(T, \mathcal{T}) = (V, \mathcal{V})$ and τ is the identity mapping. A generic distribution of payoff types will in the following be denoted by $s \in \mathcal{M}(V)$ and $\beta_s(v)$ denotes the probabilistic beliefs of an individual with payoff type v regarding the cross-section distribution of payoff types.

3.3 Implementability on a given type space

For implementation of social choice functions, we consider anonymous incentive mechanisms of the form $f = (R, q, p)$, where R is a set of possible reports to the mechanism, $q : \mathcal{M}(R) \to \{0, 1\}$ specifies a public-goods provision level as a function of the cross-section distribution of messages, and $p : R \times \mathcal{M}(R) \to \mathbb{R}$ specifies an individual's payment as a function of the message sent by this individual and, again, of the cross-section distribution of messages. A social choice function F is *interim implementable on a given type space* if, for this type space, there exists an anonymous mechanism $f = (R, q, p)$ and there exists a symmetric interim Nash equilibrium of the strategic game induced by this mechanism such that, for any cross-section distribution of types δ, the equilibrium outcome of the game induced by the mechanism is equal to the outcome that the social choice function F stipulates for the payoff-type distribution $\theta(\delta)$.

For a formal statement of this condition, we need some more notation. Given an anonymous mechanism $f = (R, q, p)$, a (mixed) strategy profile is a measurable function $\sigma : T \to \mathcal{M}(R)$ that specifies, for each type $t \in T$, a lottery $\sigma(t)$ over reports $r \in R$ that an agent might make. The strategy profile σ corresponds to an *interim Nash equilibrium* if, for each $t \in T$, the lottery $\sigma(t)$ assigns probability one to the set of reports that an agent of type t considers optimal when he or she anticipates that the other agents' reports are given by σ. In specifying the agent's expectations, we assume that he or she relies on a law of large numbers for large economies and equates the cross-section distribution of reports with the probability distribution of the report that is submitted by a randomly drawn agent. More precisely, he or she anticipates that, if all agents choose the strategy σ and if the cross-section distribution of types is δ, then the cross-section distribution of reports received by the mechanism is $\Delta(\sigma, \delta, f) \in \mathcal{M}(R)$, where, for any subset $R' \subset R$,

$$\Delta(R' | \sigma, \delta, f) := \int_T \sigma(R' \mid t') \, d\delta(t'); \tag{3}$$

here $\sigma(R' \mid t')$ is the probability, under the lottery $\sigma(t')$, that a report belongs to the set R'. Given the anticipation that other agents play according to the strategy σ, the expected payoff

of an agent of type t sending the report r is therefore given as

$$U(r \mid t, \sigma, f) := \int_{\mathcal{M}(T)} \{\tau(t) q(\Delta(\sigma, \delta, f)) - p(r, \Delta(\sigma, \delta, f))\} \, d\beta(\delta \mid t). \tag{4}$$

A strategy σ^* is an interim Nash equilibrium on the given type space if

$$\int_R U(r' \mid t, \sigma^*, f) d\sigma^*(r'|t) \geq U(r \mid t, \sigma^*, f) \tag{5}$$

for all $t \in T$ and all $r \in R$. If the cross-section distribution of types is δ, then the outcome corresponding to this equilibrium is given by the functions $q(\Delta(\sigma^*, \delta, f))$ for the public-good provision rule and $p(\cdot, \Delta(\sigma^*, \delta, f))$ for the payment rule. The mechanism $f = (R, q, p)$ and the interim equilibrium σ^* *implement* the social choice function$F = (Q_F, P_F)$ if

$$q(\Delta(\sigma^*, \delta, f)) = Q_F(\delta \circ \tau^{-1}) \tag{6}$$

for all $\delta \in \mathcal{M}(T)$, and

$$p(r, \Delta(\sigma^*, \delta, f)) = P_F(\tau(t), \delta \circ \tau^{-1}) \ , \tag{7}$$

for all $\delta \in \mathcal{M}(T)$, all $t \in T$, and $\sigma^*(t)$-almost all $r \in R$.

3.4 Robust Implementability and $MLRP$-Robust Implementability

An anonymous social choice function F is said to be *robustly implementable* if, for every $(T, \mathcal{T})$, and $\tau : T \to V$, there exists an anonymous mechanism f and an interim Nash equilibrium σ^* that implement F on the type space $[(T, \mathcal{T}), \tau, \beta]$, for every belief system β.[13] Hence, F is robustly implementable if there exists a mechanism f and a strategy σ^* so that (5)–(7) hold for every β.

For some of our analysis, this robustness requirement is too strong. Therefore, we also use a weaker concept of P-robust implementability. Given some property P that belief systems may have, an anonymous social choice function F is said to be *P-robustly implementable* if, for every $(T, \mathcal{T})$, and $\tau : T \to V$, there exists an anonymous mechanism f and an interim Nash equilibrium σ^* that implement F on the type space $[(T, \mathcal{T}), \tau, \beta]$, for every belief system β that has property P.

We shall be particularly interested in P-robust implementability when P is the monotone-likelihood-ratio property requiring that, for any x, y, z in $\mathbb{R}$ and any t and $t' \in T$, $\tau(t') \geq \tau(t)$ implies

$$\frac{\beta(\{\delta \mid \theta((x, \infty) \cap V \mid \delta) \geq z\} \mid t')}{\beta(\{\delta \mid \theta((-\infty, x) \cap V|\delta) \geq y\} \mid t')} \geq \frac{\beta(\{\delta \mid \theta((x, \infty) \cap V \mid \delta) \geq z\} \mid t)}{\beta(\{\delta \mid \theta((-\infty, x) \cap V|\delta) \geq y\} \mid t)}. \tag{8}$$

[13] Our notion of robustness is slightly stronger than that of Bergemann and Morris (2005). Like Bergemann and Morris, we require implementability on every type space, but, following Ledyard (1978), we go further than they do and require that the mechanism that is used for implementation should be the same regardless of what the belief system is. In contrast, Bergemann and Morris allow the mechanism to depend on β. This difference is irrelevant if one is only concerned with the characterization of robustly implementable social choice functions. It matters, however, for the characterization of robust and coalition-proof social choice functions; see the discussion in Section 4.2 below.

Thus, an anonymous social choice function F is said to be *MLRP-robustly implementable* if, for every $(T, \mathcal{T})$ and $\tau : T \to V$ there exists an anonymous mechanism f and an interim Nash equilibrium σ^* that implement F on the type space $[(T, \mathcal{T}), \tau, \beta]$, for every belief system β that has the monotone-likelihood-ratio property. Under the monotone-likelihood-ratio property, there is a sense in which beliefs about the cross-section distribution of payoff-types, $\theta(\delta)$, depend monotonically on the agent's own payoff type. If the agent's payoff type is higher, he considers the likelihood of a cross-section distribution assigning more than a certain weight to payoff types above a certain threshold to be higher relative to the likelihood of a cross-section distribution assigning more than a certain weight to payoff types below the threshold.

The monotone-likelihood-ratio property is obviously satisfied by any belief system with beliefs that are independent of types. In particular, it is satisfied by any degenerate belief system. The following result is therefore relevant for $MLRP$-robustly implementability.

Proposition 1 *For any property P that is satisfied by every degenerate belief system, an anonymous social choice function $F = (Q_F, P_F)$ is P-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 $s \in \mathcal{M}(V)$,*

$$vQ_F(s) - P_F(v, s) \geq vQ_F(s) - P_F(v', s) \; . \tag{9}$$

Proposition 1 adapts an argument of Bergemann and Morris (2005) to the given setup.[14] For any property P that is satisfied by every degenerate belief system, P-robust implementability is equivalent to *ex post incentive compatibility*: suppose that a direct mechanism is used to implement a social choice function in a truthtelling equilibrium. Then, once s has become known, no individual regrets having revealed his type to the mechanism.

By inspection of (9), in our setting, *ex post* implementability is equivalent to the requirement that $P_F(v, s) = P_F(v', s)$ 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 *For any property P that is satisfied by every degenerate belief system, an anonymous social choice function $F = (Q_F, P_F)$ is P-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 P_F takes the form $P_F(v, s) = \bar{P}_F(s)$ for all $v \in V$ and all $s \in \mathcal{M}(V)$. Robust implementation is provided by the direct mechanism $f_F = (T, q_F, p_F)$ and the "honest" strategy profile h, i.e., $h(t)$ is the degenerate lottery that assigns probability one to the payoff type $\tau(t)$. Consequently, for any $t \in T$ and any $\delta \in \mathcal{M}(V)$,*

$$q_F(\delta) = Q_F(\delta \circ \tau^{-1}) \quad \textit{and} \quad p_F(t, \delta) = \bar{P}_F(\delta \circ \tau^{-1}) \; . \tag{10}$$

[14] A proof can be found in Bierbrauer and Hellwig (2010).

Given Corollary 1, we will represent a robustly implementable social choice function in the following as a pair of functions $(Q_F, \bar{P}_F)$, 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 $s \in \mathcal{M}(V)$.

3.5 Robust Implementation of First-Best Allocations

An anonymous social choice function $F = (Q_F, P_F)$ is said to yield first-best outcomes if, for all $s \in \mathcal{M}(V)$ the pair $(Q_F(s), P_F(s, \cdot))$ maximizes the aggregate surplus

$$\int_V \{vQ_F(s) - P_F(s, v)\} ds(v)$$

subject to the feasibility condition (2). By standard arguments, this requires that the public good should be provided if the aggregate valuation $\bar{\mathbf{v}}(s) := \int_V v ds(v)$ exceeds the cost k and should not be provided if $\bar{\mathbf{v}}(s)$ is less than k. Moreover, there should be no slack in the feasibility constraint, i.e., aggregate payments should exactly cover the cost of public-good provision. Upon combining these observations with Corollary 1, we obtain:

Proposition 2 *A first-best anonymous social choice function $F = (Q_F, P_F)$ is robustly implementable if and only if, for all $s \in \mathcal{M}(V)$,*

$$Q_F(s) = \begin{cases} 0, & \text{if} \quad \bar{\mathbf{v}}(s) < k, \\ 1, & \text{if} \quad \bar{\mathbf{v}}(s) > k, \end{cases} \qquad \text{and} \quad P_F(v, s) = kQ_F(s), \quad \text{for all } v \in V \ .$$

Proposition 2 provides a general possibility result for robust first-best implementation in a large economy. People are asked for their payoff types. The public good is provided if and only if the reported average per-capita valuation exceeds k. Required contributions are set so that the costs of public-good provision are equally shared; this ensures feasibility (budget balance), as well as robust implementability. Thus, in the absence of participation constraints, Proposition 2 suggests that the implementation of first-best allocations in large economies faces no fundamental difficulties.[15]

However, we do not regard Proposition 2 as a satisfactory basis for the normative theory of public-good provision in a large economy. As we discussed 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 such an economy. In the following section, we therefore introduce the analysis of coalition proofness as an additional restriction on social choice functions and incentive mechanisms.

[15] 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 has a net payoff below zero if the public good is provided and a payoff of zero if the public good is not provided. Such a person would not like to participate. In this paper, however, we are not concerned with participation constraints. Participation constraints play no role if the government has powers of coercion.

4 Robust Implementability and Coalition-Proofness

To implement a first-best outcome, one must know the aggregate public-good valuation $\bar{\mathbf{v}}(s)$. This information is derived from people's reports about their individual valuations. Under the mechanism $f_F = (V, Q_F, P_F)$ in Proposition 2, people are willing to provide this information because they see themselves as being unable to influence the outcome at all. Being unable to influence anything, they are indifferent as to what they report. Given this indifference, truthtelling is an interim Nash equilibrium of the game induced by f_F.

Because of this very indifference, however, this game has many equilibria. Moreover, for certain subsets of the population, some of these equilibria are more attractive than the truthtelling equilibrium. Consider the set of types with $\tau(t) > k$ who benefit if the public good is provided. If all of them were to submit a report $v^* = \max V$, they would raise the mechanisms assessment of the aggregate public-good valuation from its true level $\bar{\mathbf{v}}(s)$ to the level $\bar{\mathbf{v}}(s) + \int_{\{\tau(t)>k\}}[v^* - \tau(t)]d\delta(t)$, which is higher if the set $\{t \in T|\tau(t) \in (k, v^*)\}$ has positive mass under the cross-section type distribution δ. For some δ, this collective exaggeration of enthusiasm for the public good would change the provision decision, inducing the public good to be provided even though the aggregate valuation $\bar{\mathbf{v}}(s)$ is below the per-capita cost k. From the perspective of types with $\tau(t) > k$, the equilibrium with the strategy $\hat{\sigma}$ that stipulates reports v^* for them and truthtelling for the other types will therefore dominate the truthtelling equilibrium.

For a single individual, there is no reason to deviate from truthtelling to the exaggeration strategy. However, people with similar valuations have similar interests. Collectively, they might upset the truthtelling equilibrium. Therefore, they would seem to have an incentive to form a coalition in order to manipulate the social outcome collectively. To eliminate the possibility of such a manipulation, we impose a requirement of coalition proofness in addition to robust implementability.

4.1 Coalition Proofness

With this requirement, we follow Laffont and Martimort (1997, 2000). However, whereas Laffont and Martimort focussed on collective deviations by the grand coalition of all agents, we focus on deviations by coalitions of subsets of agents, where coalition membership depends on types.

Proceeding formally, let σ^* be an interim Nash equilibrium for the game that is induced by the mechanism f on a given type space $[(T, \mathcal{T}), \tau, \beta]$. A subset T' of the type set T is said to *block* the equilibrium σ^* if there exists a function $\sigma'_{T'} : T' \to \mathcal{M}(R)$ so that:

(a) The strategy profile $(\sigma'_{T'}, \sigma^*_{T\setminus T'})$, which is obtained by people having types in T' play according to $\sigma'_{T'}$ and types in $T\setminus T'$ play according to σ^*, is also an interim Nash equilibrium for the game that is induced by the mechanism f on $[(T, \mathcal{T}), \tau, \beta]$.

(b) For all $t' \in T'$,

$$\int_R U(r' \mid t', (\sigma'_{T'}, \sigma^*_{T\setminus T'}), f)d\sigma'_{T'}(r'|t) \geq \int_R U(r' \mid t', \sigma^*, f)d\sigma^*(r'|t), \tag{11}$$

and, for some $t' \in T'$, the inequality in (11) is strict.

In this definition, condition (a) is an incentive compatibility requirement. Like Laffont and Martimort (1997, 2000), we insist that the collective deviation that a coalition induces should itself be individually incentive-compatible so that no coalition member has an incentive to deviate from the deviation associated with the coalition[16]. In addition, condition (a) requires that there should also be no incentives for non-members of the coalition to deviate from the choice $\sigma^*(t)$ that is stipulated by the original strategy profile σ^*. Condition (b) is the usual requirement that the a blocking coalition provides a Pareto-superior outcome to its members.[17]

An interim Nash equilibrium σ^* for the game that is induced by the mechanism f on $[(T, \mathcal{T}), \tau, \beta]$ is said to be *coalition-proof* if there is no set $T' \subset T$ that blocks it. A social choice function F is said to be robustly implementable and coalition-proof if, for every $(T, \mathcal{T})$, and $\tau : T \to V$, there exists a mechanism f and a strategy profile σ^* such that, for every belief system β, σ^* is a coalition-proof interim Nash equilibrium for the game induced by f on the type space $[(T, \mathcal{T}), \tau, \beta]$, and, moreover, f and σ^* implement F.

Given a property P that a belief system may have, F is *P-robustly implementable and coalition-proof* if, for every $(T, \mathcal{T})$, and $\tau : T \to V$, there exists a mechanism f and a strategy profile σ^* such that, for every belief system β with the property P, σ^* is a coalition-proof interim Nash equilibrium for the game induced by f on the type space $[(T, \mathcal{T}), \tau, \beta]$, and, moreover, f and σ^* implement F.

Our definition of blocking does not allow the deviating agents to use incentive-compatible side-payments to enlarge the set of agents who are willing to participate in a blocking manipulation. In a model with a continuum of agents, however, this restriction is without loss of generality. Since deviating idividuals do not perceive themselves as being pivotal for the outcomes induced by a deviation, their participation constraint would be violated as soon as they were asked to make a payment. If such a payment was asked for, they would be better off free-riding; that is, they would benefit from the outcome induced by deviation, without making a personal contribution.

The requirement that a social choice function should be robustly implementable and coalition-proof bears some relation to the literature on "full" implementability of a social choice function.[18] This literature aims at characterizing those social choice functions that can be implemented by

[16] In contrast to Laffont and Martimort, however, we do not explicitly study an extensive-form game of coalition formation and merely look at coalition-proof equlibria in a normal-form formulation. An extensive-form formulation is considered in Bierbrauer and Hellwig (2010); this form leads to the same charcterization of robustly implementable and coalition-proof social choice function as the simpler formulation that is provided here.

[17] For ease of exposition, we do not yet invoke the requirement that a deviation must itself be coalition-proof, i.e., that there must not be an incentive for a subset T'' of T' to deviate to a strategy $\sigma''_{T''}$ given that all other types behave according to $(\sigma'_{T'}, \sigma^*_{T \setminus T'})$. This corresponds to the notion of Bernheim et al. (1986) that collective manipulations themselves must be coalition-proof. As we have shown in Bierbrauer and Hellwig (2010), our main result gets stronger if we require that deviating coalitions are subcoalition-proof. In this case, we can establish the equivalence of robust and coalition-proof mechanisms, on the one hand, and voting mechanisms on the other for all type spaces and not just for those satisfying the $MLRP$.

[18] See Jackson (2001), or Moore (1992) for an overview. Recently, Bergemann and Morris (2009) have discussed full implementability in connection with the requirement of robustness with respect to the specification of the belief system.

a mechanism so that *every* equilibrium of the mechanism achieves a desired social outcome.[19] Like this literature, we are concerned with alternative equilibria of the game induced by the mechanism f on the type space $[(T, \mathcal{T}), \tau, \beta]$. However, whereas the literature on full implementation is concerned about equilibrium multiplicity per se, we are only concerned with equilibrium multiplicity to the extent that a group of agents might benefit from a collective deviation which induces an alternative equilibrium.

4.2 Preliminary Results on Robust Implementability and Coalition Proofness

For robustly implementable social choice functions, Corollary 1 shows that implementation can rely on direct mechanisms and truthtelling equilibria. The follow proposition asserts the same for robustly implementable and coalition-proof social choice functions.

Proposition 3 *For any property P that is satisfied by every degenerate belief system, a P-robustly implementable anonymous social choice function $F = (Q_F, \bar{P}_F)$ is coalition-proof if and only if, on every type space $[(T, \mathcal{T}), \tau, \beta]$, the "honest" strategy profile h is a coalition-proof interim Nash equilibrium for the game induced by the direct mechanism $f_F = (T, q_F, p_F)$, where $f_F = (T, q_F, p_F)$ and h are as specified in Corollary 1. In this case, in particular, truthtelling is a coalition-proof interim Nash equilibrium for the game induced by the mechanism $f_F = (V, Q_F, \bar{P}_F)$ on the naive type space $[(V, \mathcal{V}), \beta]$, for every β.*

For any property P that is satisfied by every degenerate belief system, Proposition 1 implies that, if an anonymous social choice function is P-robustly implementable, then it is ex post implementable. The following proposition provides an analogous result for coalition proofness. For a formal statement, we need some more notation. Given the type space $[(T, \mathcal{T}), \tau, \beta]$ and the direct mechanism $f_F = (T, q_F, p_F)$, a *collective deviation* is defined as a pair $\pi = (T', \ell'_{T'})$, such that $T' \subset T$ is the set of deviating types and $\ell'_{T'} : T' \to \mathcal{M}(T)$ is a function that specifies a "lie" $\ell'_{T'}(t')$, i.e., a lottery over (typically false) type announcements, for each $t' \in T'$. Along the same lines as before, we write $(\ell'_{T'}, h_{T \setminus T'})$ for the strategy profile that is obtained by having types in T' report payoff types according to $\ell'_{T'}$ and types in $T \setminus T'$ report payoff types according to h. Given this strategy profile, if the cross-section distribution of types is δ, the cross-section distribution of reports received by the mechanism f_F, is equal to $\Delta((\ell'_{T'}, h_{T \setminus T'}), \delta, f_F)$ as specified by equation (3) above.

Proposition 4 *Let P be a property that is satisfied be every degenerate belief system. If a P-robustly implementable anonymous social choice function $F = (Q_F, \bar{P}_F)$ is coalition-proof, then, there is no $s \in \mathcal{M}(V)$ and there is no collective deviation $\pi = (V', \ell'_{V'})$ such that π blocks the honest strategy profile h in the game induced by the mechanism $(V, Q_F, \bar{P}_F)$ on a naive type*

[19] By contrast, a large part of the mechanism design literature requires only that there is *some* mechanism with *some* equilibrium that achieves the outcomes stipulated by a given social choice function.

space $[(V,\mathcal{V}),\beta_s]$ *with a degenerate belief system, i.e., there are no* s *and* $\pi=(V',\ell'_{V'})$ *such*

$$\begin{aligned} &v'Q_F(\Delta((\ell'_{T'},h_{T\setminus T'}),s,f_F))-\bar{P}_F(\Delta((\ell'_{T'},h_{T\setminus T'}),s,f_F))\\ &\geq v'Q_F(s)-\bar{P}_F(s), \end{aligned} \tag{12}$$

for all $v'\in V'$*, where, for some* $v'\in V'$*, the inequality in (12) is strict.*

Condition (12) concerns the attractiveness of collective deviations *ex post*: if (12) holds for all $v'\in V'$, with a strict inequality for some v', and if the cross-section distribution s is known to the participants, then the coalition of agents with types in V' wants to block the truthtelling equlibrium by deviating to the "lie" $\ell'_{V'}$. Proposition 4 asserts that, if the social choice function is coalition-proof, then no such coalition exists. The argument is derived from the observation that, for the degenerate belief system β_s, coalition proofness of a strategy profile in the game induced by the mechanism f_F on the naive type space $[(V,\mathcal{V}),\beta_s]$ is equivalent to coalition proofness ex post for the cross-section type distribution s.

Proposition 4 asserts the *necessity* of coalition proofness ex post. This raises the question whether coalition proofness ex post is also *sufficient* for a robustly implementable anonymous social choice function to be coalition-proof. Without any additional restrictions on belief systems, the answer to this question is negative. If we know that the truthtelling equilibrium for the direct mechanism $f_F=(V,Q_F,P_F)$ cannot be blocked ex post, we still cannot rule out the possibility that, for nondegenerate probabilistic belief systems, the truthtelling equilibrium for this mechanism might be blocked by a coalition whose members all expect the collective deviation to improve the outcome in some aggregate states and to worsen it in others and beliefs are such that, taking expectations over aggregate states, they all expect to benefit from the deviation.

5 The Main Result: Implementability by a voting mechanism

This section contains our main result, namely that any social choice functions that is $MLRP$-robust and coalition-proof can be reached by a voting mechanism. We proceed as follows: we first define what we mean by a voting mechanism and then provide a characterization of $MLRP$-robust and coalition-proof social choice functions. We will then show that, under a weak assumption, any such social choice function can be implemented by a voting mechanism.

A voting mechanism Φ is defined as a mechanism with the following properties:

- People are presented with two alternatives and can vote for one or the other. The message set R_Φ is therefore a binary set, $R_\Phi=\{\text{alternative } 0,\text{alternative } 1\}$.
- Alternative 1 stipulates that the public good should be provided and that each participant should make a payment $P^1_\Phi\geq k$. Alternative 0 stipulates that the public good should not be provided and that each participant should make a payment $P^0_\Phi\geq 0$.
- Alternative 1 is implemented if the share of people voting for it exceeds a given threshold $m_\Phi\in[0,1]$. If the share of people voting for alternative 1 is less than m_Φ, alternative 0 is implemented. If the share of people voting for alternative 1 is equal to m_Φ, then either alternative 0 or alternative 1 is implemented.

Economists have long been critical of the prominent role of voting in political decision making, arguing that the neglect of preference intensities in voting was a major source of distortions. The following result shows that this property of voting mechanisms is actually implied by robust implementability and coalition proofness. If an anonymous social choice function is to be robustly implementable and coalition-proof, it must not condition the provision of the public good on preference intensities. For robust implementabibility and coalition proofness, the provision of the public good can only be conditioned on the size of the set of people who benefit from public-good provision and the size of the set of people who are hurt by public-good provision. Moreover, the provision rule must be monotonic in the sense that it is not possible to have the public good provided when the set of net beneficiaries is smaller than in some other instance where it is not provided.

The following two propositions, which are proven in Section 6, provides a characterization of $MLRP$-robustly implementable and coalition-proof social choice functions.

Proposition 5 *If an anonymous social choice function F is $MLRP$ -robustly implementable and coalition-proof then there exist numbers P_F^0 and P_F^1 so that, for all $v \in V$ and all $s \in \mathcal{M}(V)$,*

$$P_F(v,s) = \begin{cases} P_F^0, & \text{if} \quad Q_F(s) = 0\,, \\ P_F^1, & \text{if} \quad Q_F(s) = 1\,. \end{cases} \tag{13}$$

If the participants' payments were high in one state and low in another when both states involve the same level of public-good provision, then the grand coalition coalition of all participants together could use a collective deviation to induce the outcome with low payments when the actual state would call for high payments.

Given this lemma, we restrict our attention to social choice functions with payments that depend only on whether the public good is provided or not. For such social choice functions, we find it convenient to write $F = (Q_F, P_F^0, P_F^1)$ rather than $F = (Q_F, P_F)$.

Given such a social choice function, we denote by

$$V_1(P_F^1 - P_F^0) := \{v \in V \mid v > P_F^1 - P_F^0\} \text{ and } V_0(P_F^1 - P_F^0) := \{v \in V \mid v < P_F^1 - P_F^0\}$$

the sets of payoff types of net gainers and net losers from public-good provision, respectively.

In the following we fix a social choice function $F = (Q_F, P_F^0, P_F^1)$ and ask under what condition it is robustly implementable as a coalition-proof equilibrium. For ease of exposition we assume that there are no indifferent types i.e., that $P_F^1 - P_F^0 \cap V = \emptyset$.

Proposition 6 *Consider a social choice function $F = (Q_F, P_F^0, P_F^1)$ and suppose that $P_F^1 - P_F^0 \cap V = \emptyset$. This social choice function is $MLRP$-robustly implementable and coalition-proof if and only if for all s and s' in $\mathcal{M}(V)$,*

$$s(V_1(P_F^1 - P_F^0)) \geq s'(V_1(P_F^1 - P_F^0)) \quad \text{implies} \quad Q_F(s) \geq Q_F(s'). \tag{14}$$

Corollary 2 *Consider a social choice function $F = (Q_F, P_F^0, P_F^1)$ and suppose that $P_F^1 - P_F^0 \cap V = \emptyset$. This social choice function is $MLRP$-robustly implementable and coalition-proof if and only if it is implementable by a voting mechanism.*

The conditions in Propositions 5 and 6 are equivalent to the ex post coalition proofness conditions in Proposition 4. They immediately imply that the social choice function must be implementable by a voting mechanism. Because Proposition 4 is not restricted to $MLRP$-robustly implementable and coalition-proof social choice functions, this implication is actually more general than the Proposition make it appear. For any property P that is satisfied by all degenerate belief systems, an anonymous social choice function that is P-robustly implementable and coalition-proof must be implementable by a voting mechanism.

The restriction to $MLRP$-robust implementability is needed for the converse, i.e., the statement that implementability by a voting mechanism is *sufficient* for $MLRP$-robust implementability and coalition proofness of an anonymous social choice function. The monotone-likelihood-ratio property of belief systems eliminates the possibility that, for nondegenerate probabilistic belief systems, the truthtelling equilibrium for the given mechanism might be blocked by a coalition whose members all expect the collective deviation to improve the outcome in some aggregate states and to worsen it in others. Given the monotonicity of the public-good provision rule, the monotone-likelihood-ratio property ensures that, when the participants take expectations over the different aggregate states, they cannot all expect to benefit from the deviation.

Remarks on Robustness and Weak Coalition Proofness

In the preceding analysis, the requirement of $MLRP$-robust implementability cannot be replaced by the stronger requirement of robust implementability. If there are no restrictions on belief systems, one can always find belief systems that give rise to the possibility of blocking the truthtelling equilibrium for a social choice function that satisfies the conditions of Proposition 6 simply because the different groups assign different probability weights to the different aggregate states.

The requirement of $MLRP$-robust implementability can, however, be strengthened to robust implementability if, at the same time, the requirement of coalition proofness is replaced by a notion of *weak coalition proofness*. This weaker concept only considers coalitions that are themselves immune to the formation of sub-coalitions.[20] More formally, it can be shown that a social choice function F is robustly implementable and weakly coalition-proof if and only if there exist numbers P_F^0 and P_F^1 so that the conditions of Propositions 5 and 6 are satisfied.

The argument involves three steps. The first step is to show that the requirement of weak coalition-proofness does also imply that the payment scheme can be reduced to two payments P_F^1 and P_F^0. The second step is the observation that manipulations that are supported solely

[20] As explained in Bierbrauer and Hellwig (2010), for technical reasons, the analysis of weak coalition-proofness is interesting only if attention is restricted to social choice functions with the property that the problem of forming a manipulation that achieves a certain outcome $Q \in \{0, 1\}$ with minimal payments is well-defined.

by gainers or losers are weakly coalition-proof because all coalition members have aligned interests. Steps 1 and 2 imply that the conditions in Propositions 5 and 6 remain necessary if we replace coalition-proofness by weak-coalition-proofness. The proof that these conditions are also sufficient is then based on the following insight: whenever a manipulation stipulates that some gainers claim to be losers and some losers claim to be gainers of public-goods provision, the manipulation itself fails to be coalition-proof. Given that the losers claim to be gainers, the gainers possess a deviation that makes all of them better off, namely to communicate truthfully that they are gainers, so as to make the provision of the public good more likely.

6 Proof of the main result

Proof of Proposition 5. Suppose that the lemma is false, and let $F = (Q_F, \bar{P}_F)$ be a $MLRP$-robustly implementable and coalition-proof social choice function such that, for some s and $\bar{s}$, $Q_F(s) = Q_F(\bar{s})$ and $\bar{P}_F(s) > \bar{P}_F(\bar{s})$. Consider the naive type space $[(V, \mathcal{V}), \beta_s]$ with the degenerate belief system β_s that assigns all probability mass to the aggregate state s, and let $f_F = (V, Q_F, \bar{P}_F)$ be the direct mechanism that implements F in truthtelling strategies on this type space. Let ℓ_V be such that $\Delta((\ell'_V, h_{\emptyset'}), s, f_F) = \bar{s}$. Then the collective deviation $\pi = (V, \ell_V)$ by the grand coalition of all agents induces the outcome

$$(Q_F(\Delta((\ell'_V, h_{\emptyset'}), s, f_F)), \bar{P}_F(\Delta((\ell'_V, h_{\emptyset'}), s, f_F))) = (Q_F(\bar{s}), \bar{P}_F(\bar{s})),$$

which provides a payoff

$$vQ_F(\bar{s}) - \bar{P}_F(\bar{s}) > vQ_F(s) - \bar{P}_F(s)$$

to a participant with payoff type v. By Proposition 4, it follows that F cannot be coalition-proof. The assumption that the lemma is false has thus led to a contradiction. ■

Necessity of the Conditions of Proposition 6. We seek to show the following: given a social choice function $F = (Q_F, P^1_F, P^0_F)$ and given that $P^1_F - P^0_F \cap V = \emptyset$, F is robustly implementable and coalition-proof only if, for all s and s' in $\mathcal{M}(V)$, $s(V_1(P^1_F - P^0_F)) \geq s'(V_1(P^1_F - P^0_F))$ implies $Q_F(s) \geq Q_F(s')$.

Suppose that this condition fails to hold, i.e., that $Q_F(s) = 0$ and $Q_F(s') = 1$ for some s and s' such that $s(V_1(p^1_F - p^0_F)) \geq s'(V_1(p^1_F - p^0_F))$. To focus on the essentials of the argument, suppose first that, in fact, $s(V_1(P^1_F - P^0_F)) = s'(V_1(P^1_F - P^0_F))$ and hence $s(V_0(P^1_F - P^0_F)) = s'(V_0(P^1_F - P^0_F))$. We claim that, for some belief system that has the monotone-likelihood-ratio property, the "honest" strategy profile h cannot be coalition-proof.

Consider the coalition of people with payoff types in $V_0(P^1_F - P^0_F)$. All these people prefer the outcome for the aggregate state s to the outcome for the aggregate state s'. Thus, if the belief system is degenerate and all beliefs assign all probability mass to the payoff type distribution s', a coalition of people with payoff types in $V_0(P^1_F - P^0_F)$ would block the honest strategy profile if they could find a reporting strategy ℓ_{T_0} that would induce a reports distribution $\Delta((\ell_{V_0}, h_{V \setminus V_0}), s', f_F)$ such that $Q_F(\Delta((\ell_{V_0}, h_{V \setminus V_0}), s', f_F)) = 0$.

For instance, since $Q_F(s) = 0$, this coalition would block the honest strategy profile if it could induce the reports distribution $\Delta((\ell_{V_0}, h_{V\setminus V_0}), s', f_F) = s$. Because the reports distribution $\Delta((\ell_{V_0}, h_{V\setminus V_0}), s', f_F)$ also depends on the (honest) reports of people with payoff types in $V_1(P_F^1 - P_F^0)$, it may not be able to do so. However, because its size is the same under both s and s', there does exist a reporting strategy ℓ_{V_0} for this coalition such that the induced reports distribution takes the form

$$\Delta((\ell_{V_0}, h_{V\setminus V_0}), s', f_F) = s'',$$

where s'' satisfies

$$s''(\hat{V}) = s(\hat{V}) \quad \text{if} \quad \hat{V} \subset V_0(P_F^1 - P_F^0)\ .$$

and

$$s''(\hat{V}) = s'(\hat{V}) \quad \text{if} \quad \hat{V} \subset V_1(P_F^1 - P_F^0)\ .$$

If the coalition of people with payoff types in $V_0(P_F^1 - P_F^0)$ is to be prevented from blocking the honest strategy profile h when beliefs assign all probability mass to the type distribution s', it must be the case that $Q_F(s'') = 1$.

But now consider the coalition of people with types in $V_1(p_F^1 - p_F^0)$. All these people prefer the outcome for the aggregate state s' and, therefore, also the outcome for the aggregate state s'' to the outcome for the aggregate state s. Thus, if the belief system is degenerate and beliefs assign all probability mass to the type distribution s, a coalition of people with types in $V_1(p_F^1 - p_F^0)$ will block the honest strategy profile if they can find a reporting strategy ℓ_{V_1} that induces the reports distribution

$$\Delta_{f_F}((\ell_{V_1}, h_{V\setminus V_1}), s) = s'',$$

which yields the outcome $Q_F(s'') = 1$. Because, by the definition of s'', the coalition of people with payoff types in $V_1(p_F^1 - p_F^0)$ has the same size under s'' as under s and s' and, moreover, the restrictions of s'' and s to the set $V_0(P_F^1 - P_F^0)$ are the same, such a reporting strategy ℓ_{V_1} is in fact available.

Thus, if there are two aggregate states satisfying $s(V_1(p_F^1 - p_F^0)) = s'(V_1(p_F^1 - p_F^0))$, the implementation of a social choice function prescribing $Q_F(s) = 0$ and $Q_F(s') = 1$ can necessarily be blocked. Either it can be blocked by a coalition of people with payoff types in $V_0(P_F^1 - P_F^0)$ when beliefs put all probability mass on s, or it can be blocked by a coalition of people with payoff types in $V_1(P_F^1 - P_F^0)$ when beliefs put all probability mass on s'.

In the preceding argument, the assumption that $s(V_1(p_F^1 - p_F^0)) = s'(V_1(p_F^1 - p_F^0))$, is not really needed. A little reflection shows that, if $s(V_1(p_F^1 - p_F^0)) \geq s'(V_1(p_F^1 - p_F^0))$, the blocking coalitions in the preceding argument have even more scope for finding collective deviations so as to generate reports distributions equal to s''. ■

Sufficiency of the Conditions of Proposition 6. We consider a social choice function $F = (Q_F, P_F^1, P_F^0)$ and assume that $P_F^1 - P_F^0 \cap V = \emptyset$. Moreover, we consider a belief system with the $MLRP$. We seek to show the following: if F is such that for all s and s' in $\mathcal{M}(V)$,

$s(V_1(P_F^1 - P_F^0)) \geq s'(V_1(P_F^1 - P_F^0))$ implies $Q_F(s) \geq Q_F(s')$, then, with a direct mechanism f_F, the strategy profile h is coalition-proof.

The argument proceeds in two steps. In the first step, we show that, under the given conditions, h cannot be blocked by a coalition of people with homogeneous interests. In the second step, we show that h can neither be blocked by a coalition of people with conflicting interests.

Step 1. Consider a coalition of people with payoff types in $V_0(P_F^1 - P_F^0)$ or a coalition of people with payoff types in $V_1(P_F^1 - P_F^0)$.

For a collective deviation by people with payoff types in $V_0(P_F^1 - P_F^0)$ to block the implementation of F by the mechanism f_F, this deviation must induce the outcome $Q = 0$ for some type distribution δ for which the mechanism $f_F = (T, q_F, p_F)$ stipulates $q_F(\delta) = 1$.

Given that $s(V_1(P_F^1 - P_F^0)) \geq s'(V_1(P_F^1 - P_F^0))$ implies $Q_F(s) \geq Q_F(s')$, this would only be possible if the coalition could make the set of people with payoff types in $V_1(P_F^1 - P_F^0)$ appear to be smaller than it actually is. This, however, is not possible if people with payoff types in $V_1(P_F^1 - P_F^0)$, who do not belong to the presumed blocking coalition, report their types honestly. By a precisely symmetric argument, there also is no collective deviation by people with payoff types in $V_1(P_F^1 - P_F^0)$ that can block the implementation of F.

Step 2. It remains to be shown that there is no collective deviation that is attractive for individuals with types in $V_0(P_F^1 - P_F^0)$ and for individuals with types in $V_1(P_F^1 - P_F^0)$.

For a collective deviation by people with conflicting interests to block the implementation of F by the mechanism f_F, this deviation must induce the outcome $Q = 0$ for some type distributions in $D_{01} \subset \mathcal{M}(T)$ for which F stipulates $Q = 1$ and the outcome $Q = 1$ for type distributions in $D_{10} \subset \mathcal{M}(T)$ for which F stipulates the outcome $Q = 0$. Moreover, the different participants' beliefs must be such that each participant attaches more weight to the gains from changes that he or she likes than to the losses from changes that he or she dislikes. Hence, if $t'' \in T'$ and $\tau(t'') \in V_0(P_F^1 - P_F^0)$, then

$$\frac{\beta(D_{01} \mid t'')}{\beta(D_{10} \mid t'')} \geq 1, \tag{15}$$

and if $t' \in T'$ and $\tau(t') \in V_1(P_F^1 - P_F^0)$, then

$$\frac{\beta(D_{01} \mid t')}{\beta(D_{10} \mid t')} \leq 1, \tag{16}$$

with a least one type in T' for which the inequality is strict.

Given that the social choice function satisfies (14), there exists a critical value c so that $Q_F(s) = 1$ if and only if $s(V_1(P_F^1 - P_F^0)) \geq c$. Consequently, for any pair δ and δ' with $\delta \in D_{01}$ and $\delta' \in D_{10}$,

$$\theta(V_1(P_F^1 - P_F^0) \mid \delta) \geq \theta(V_1(P_F^1 - P_F^0) \mid \delta') \ .$$

But then the $MLRP$ implies that for any pair t'' and t' where $\tau(t'') \in V_0(P_F^1 - P_F^0)$ and $\tau(t') \in V_1(P_F^1 - P_F^0)$ that

$$\frac{\beta(D_{01} \mid t')}{\beta(D_{10} \mid t'')} \geq \frac{\beta(D_{01} \mid t')}{\beta(D_{10} \mid t'')}$$

which contradicts the assumption that the inequalities in (15) and (16) hold, and that one holds as a strict inequality. ■

7 Welfare Implications

7.1 Limits to First-Best Implementation

We now turn to the welfare implications of imposing coalition-proofness, as well as robust implementability. We begin with an example that illustrates some of the issues that arise.

Example 1 *In this example, there are three possible payoff types* $V = \{0, 5, 10\}$. *The per-capita cost of public-good provision is* $k = 4.5$. *There are two possible cross-section distributions* s^j, $j = 1, 2$ *of payoff types. The population shares* s^j_v *of the different payoff types under these two cross-section distributions are given in the following table.*

j	s^j_0	s^j_5	s^j_{10}	$\bar{\mathbf{v}}(s^j)$
1	*0.3*	*0.7*	*0*	*3.5*
2	*0.4*	*0.1*	*0.5*	*5.5*

(17)

The last column in the table indicates the cross-section average valuation $\bar{\mathbf{v}}(s^j)$ *of the public good for each distribution.*

In this example, first-best implementation requires that the public good should not be provided in state 1 and that the public good should be provided in state 2. With equal cost sharing, the associated payment outcomes would be $P^0_F = 0$ and $P^1_F = 4.5$. Given these payments, the set of opponents of public-good provision consists of all types with valuations 0 and the set of net beneficiaries of public-good provision consists of all types with valuations 5 and 10. From Table 1, one immediately sees that the set of net beneficiaries has a population share of 0.8 in state 1 and of 0.7 in state 2. Because the population share of the set of net beneficiaries is larger in state 1 than in state 2, first-best implementation runs afoul of the monotonicity requirement in Theorems 1. In more concrete terms, any mechanism that would implement a social choice function with first-best outcomes would be vulnerable to a deviation by individuals with valuations 5 and 10. If these individuals believe that state 1 is the true state of the economy, they benefit if all of them report a valuation of 10 with probability 5/7, and valuations 0 and 5 with probability 1/7 each, thereby giving the impression that the true state is 2, rather than 1.

The possibility that robust first-best implementation may run afoul of coalition-proofness is also illustrated by the example in the introduction, with possible valuations 0, 3, and 10, and a per-capita provision cost equal to 4. In that example, all cross-section distributions of types involved population shares 0.3 of net beneficiaries and 0.7 of opponents of public-good provision. A robustly implementable and coalition-proof social choice function would have to be insensitive to whatever people report, which is incompatible with the efficiency requirement that the public

good be provided if and only if the population share of individuals with valuation 3 is sufficiently large. By contrast to this earlier example, the example here shows that coalition-proofness has bite even if the population share of net beneficiaries differs from state to state.

More generally, we obtain:

Corollary 3 *If there is a pair of states s and s', such that $s(V_0(k)) \leq s'(V_0(k))$ and $\bar{\mathbf{v}}(s) < k < \bar{\mathbf{v}}(s')$, then there is no social choice function that yields first best outcomes and is robust and coalition-proof.*

7.2 Second-Best Considerations

If first-best is out of reach, the overall mechanism designer is faced with a second-best problem. Given the impossibility of achieving efficient outcomes in every state s, he must choose between different deviations from efficiency that are compatible with robustness and coalition-proofness. For instance, in Example 1, he can decide whether it is better to forego the net benefits from public-good provision in state 2 or to incur the net losses from public-good provision in state 1. He might also want to change the boundary between yes-sayers and no-sayers by imposing a payment scheme that raises more funds than he needs, e.g., by asking for a payment $P_F^1 = 5.1$ if the public good is provided, rather than $P_F^1 = k = 4.5$, in order to turn people with valuations 5 from net beneficiaries into opponents of public-good provision. This would allow him to implement a first-best public-good provision rule, but there would be a waste of resources in state 2, when the public good is provided.

Any assessment of tradeoffs between the different kinds of inefficiency must rely on a system of weights that the mechanism designer assigns to the different states. For specificity, we assume that the mechanism designer has his own prior beliefs and chooses a social choice function in order to maximize expected aggregate surplus according to these beliefs, subject to the requirements of feasibility, robust implementability and coalition-proofness. Given our characterization robust implementability and coalition-proofness, this is equivalent to the problem of choosing P_F^0, P_F^1 and $Q_F : \mathcal{M}(V) \to \{0, 1\}$ so as to maximize the expected aggregate surplus

$$E^M[(\bar{\mathbf{v}}(s) - P_F^1)Q_F(s) - P_F^0(1 - Q_F(s))] \tag{18}$$

subject to the feasibility constraints that $P_F^0 \geq 0$, $P_F^1 \geq k$, and the coalition-proofness condition that for every pair s and s', $s(V_1(P_F^1 - P_F^0)) \geq s'(V_1(P_F^1 - P_F^0))$ implies $Q_F(s) \geq Q_F(s')$. The expectations operator E^M in (18) indicates that expectations over s are taken with respect to the mechanism designer's subjective beliefs.

In Example 1, the solution to this second-best problem depend on the probabilities ρ_1^M and ρ_2^M that the mechanism designer assigns to the different states. If the benefits of public-good provision are foregone in state 2, then, relative to first-best, there is a net per capita welfare loss of $5.5 - 4.5 = 1.0$ in this state. If the public good is provided in state 1, when it should not be, the per-capita welfare loss is $4.5 - 3.5 = 1.0$. If the mechanism designer deems the two states to be equiprobable, he will be indifferent between excessive provision in state 1 and non-provision

in state 2. If he deems state 2 to be more likely than state 1, he will prefer excessive provision in state 1 to non-provision in state 2; the reverse is true if he deems state 1 to be more likely.

In any case, though, non-provision in state 2 is dominated by a scheme involving non-provision in state 1 and provision with a payment $P_F^1 = 5.1 > k$ in state 2. This scheme involves a per-capita welfare loss, relative to first-best, that is equal to $5.1 - 4.5 = 0.6$ in state 2. If the mechanism designer deems the two states to be equiprobable, he will prefer this scheme even to an arrangement involving excessive provision of the public good in state 1. Excessive provision of the public good in state 1, i.e., provision of the public good in both states, with non-wasteful payments $P_F^0 = 0$ and $P_F^1 = k = 4.5$ is only preferred if the probability assigned to state 1 is less than 3/8. If the probability assigned to state 1 exceeds 3/8, the second-best social welfare function stipulates (the efficient) non-provision of the public good in state 1 and provision with a wasteful payment requirement in state 2. A wilful waste of resources may thus be part of a second-best solution when first-best solutions are ruled out by robustness and coalition-proofness.

8 An example with three provision levels

We now return to a setup with a continuum of individuals and study an extension that allows for more than two possible provision levels of the public good. More specifically, we assume that there are three possible provision levels, $Q \in \{0, 1, 2\}$. The resource requirement is given by a cost function K with $K(0) = 0$, $K(1) = k_1$ and $K(2) = k_2$, with $0 < k_1 < k_2$. We assume that the cost function is convex, $k_2 > 2k_1$.

For simplicity, as in Section 2, we focus on a model with three possible payoff types so that $V = \{0, x, 1\}$, where $0 < x < 1$. A state s of the economy is hence a triple $s = (s^0, s^x, s^1) \in \mathcal{M}(V)$. We also assume that, under equal cost sharing, individuals with payoff type $v = 1$ prefer the large provision level of 2 over the intermediate provision level of 1, which is in turn preferred over no public goods-provision at all; hence, $2 - k_2 > 1 - k_1 > 0$. The ranking is reversed for individuals with a payoff type of $v = 0$. Individuals with an intermediate payoff type prefer the intermediate provision level over non-provision and over the large provision level, $x - k_1 > 2x - k_2$ and $x - k_1 > 0$. For the sake of concreteness we assume finally that $2x - k_2 < 0$, so that the intermediate types prefer non-provision over the large provision level.

Figure 1 below represents the set of states by depicting the set

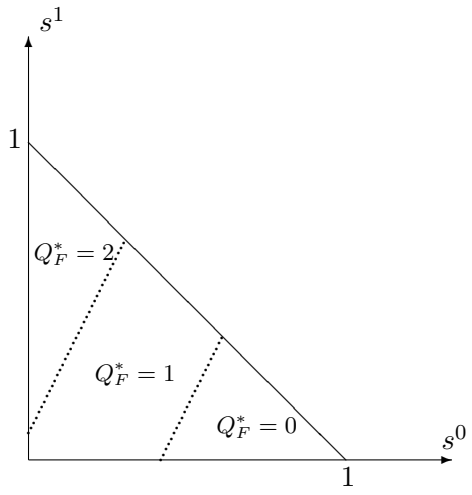
$$\{(s^1, s^0) \mid 0 \leq s^0 \leq 1; 0 \leq s^1 \leq 1 \text{ and } s^0 + s^1 \leq 1\} .$$

For any point in this set it is understood that $s^x = 1 - s^0 - s^1$. The figure also illustrates the first-best provision rule,

$$Q_F^*(s) = \begin{cases} 0, & \text{if } s^1 \leq -\frac{x-k_1}{1-x} + \frac{x}{1-x} s^0 , \\ 1, & \text{otherwise} , \\ 2, & \text{if } s^1 \geq \frac{k_2-k_1-x}{1-x} + \frac{x}{1-x} s^0 . \end{cases}$$

Corollary 1 and Proposition 4 generalize in a straightforward way to the given setup, since the proofs did not rely on there being only two possible provision levels of the public good. By

Figure 1: First-best in a setup with three provision levels



Corollary 1, robust implementability implies that, for every state s, there is a payment $\bar{P}_F(s)$ that every individual has to make, irrespective of the individual's payoff type. Proposition 4 states a necessary condition for robustness and coalition-proofness: From an ex post perspective – that is, after the state s has been revealed to individuals – there is no group of individuals who could have induced a preferred outcome by a collective lie about their payoff types.

For simplicity, we assume in the following that the individuals' payments are set in such a way that the feasibility constraint holds as an equality, $\bar{P}_F(s) = K(Q_F(s))$, for every s.[21] Consequently, under a direct mechanism with a truthtelling equilibrium, the payoff of an individual with valuation v in state s is given by the indirect utility function

$$\tilde{U}(s \mid v) := vQ_F(s) - K(Q_F(s)) \ .$$

The following Corollary is an implication of Proposition 4: There must not be a state s so that – under the assumption that individuals knew the state to be s – individuals with the same payoff type would benefit from a collective lie about their preferences.

Corollary 4 *If a social choice function is robust and coalition-proof, then the following statements have to be true:*

i) Given s^1, $\tilde{U}(s^0, 1-s^0-s^1, s^1 \mid 0)$ is a non-decreasing function of s^0. Given s^x, $\tilde{U}(s^0, s^x, 1-s^x-s^1 \mid 0)$ is a non-decreasing function of s^0.

ii) Given s^0, $\tilde{U}(s^0, s^x, 1-s^0-s^x \mid x)$ is a non-decreasing function of s^x. Given s^1, $\tilde{U}(1-s^0-s^x, s^x, s^1 \mid x)$ is a non-decreasing function of s^x.

[21] This is an innocent assumption as long as we are interested in the question whether first-best outcomes can be implemented. However, as we have seen in Section 7, second-best considerations may render slack in the resource constraint desirable.

iii) Given s^0, $\tilde{U}(s^0, 1 - s^0 - s^1, s^1 \mid 1)$ *is a non-decreasing function of* s^1. *Given* s^x, $\tilde{U}(1 - s^x - s^1, s^x, s^1 \mid 1)$ *is a non-decreasing function of* s^1.

Part i) of the Corollary addresses deviations by individuals who all have a payoff type of 0. These individuals must not benefit from collectively announcing a payoff type of x so that s^0 goes down and $s^x = 1 - s^0 - s^1$ goes up. Likewise, these individuals must not benefit from increasing s^1 at the expense of s^0. Parts ii) and iii) of the Corollary state the analogous conditions for deviations by individuals who have a payoff type of x and for deviations by individuals, who all have a payoff type of 1, respectively.

The following Proposition is an adaptation of our main result in Theorem 1 to the current setup. For brevity, it focusses only on the implications of robustness and coalition-proofness.[22]

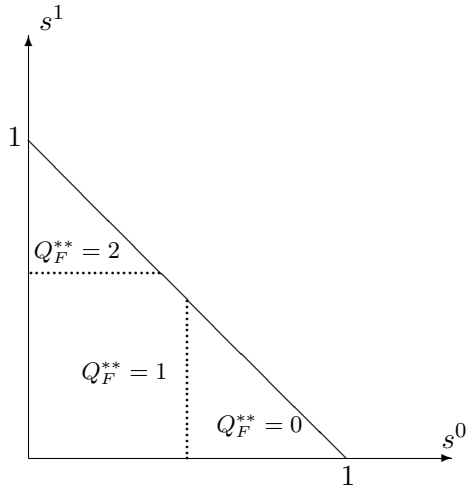
Proposition 7 *If a social choice function is robust and coalition-proof, then:*

i) Suppose that there is some $\bar{s}$ *so that* $Q_F(\bar{s}) = 2$. *Then,* $Q_F(s) = 2$, *whenever* $s^1 \geq \bar{s}^1$.

ii) Suppose that there is some $\bar{s}$ *so that* $Q_F(\bar{s}) = 2$, *and some* $\hat{s}$ *so that* $Q_F(\hat{s}) = 0$. *Then,* $Q_F(s) = 0$, *whenever* $s^0 \geq \hat{s}^0$.

A provision rule satisfying the necessary conditions in Proposition 7 is illustrated in Figure 5.

Figure 2: Second-best in a setup with three provision levels



A comparison of Figures 1 and 2 reveals that it is impossible to implement a first-best social choice function. For instance, under a first-best provision rule, the set of states in which the

[22] One way to show that the conditions in Proposition 7 are not only necessary but also sufficient would, again, be to employ a notion of weak coalition-proofness which requires that the outcome of a collective deviation must not be undermined by a further deviation of a subcoalition.

outcome is $Q = 2$ is bounded from below by an upward-sloping line. Now, whenever the true state s is on or above this line, increasing s^0 at the expense of s^x (and leaving s^1 constant) implies that we enter the region in which the outcome is $Q = 1$. Since intermediate types prefer the outcome $Q = 1$ over the outcome $Q = 2$, this implies that – whenever the belief system is such that intermediate types put enough probability mass on states so that $Q_F^*(s) = 1$ – they can jointly benefit from understating their preferences, i.e., from falsely declaring a payoff type of 0. Analogously, if these individuals believe that the most likely outcome is $Q_F^*(s) = 0$, they can jointly benefit from exaggerating; that is, from falsely declaring a payoff type of 1.

With the second-best provision rule in Figure 2 these problems are eliminated. If s^0 is increased at the expense of s^x, then either the public-goods provision level remains constant at a level of 0 or, the outcome changes from $Q = 1$ into $Q = 0$. Hence, increasing s^0 at the expense of s^x, is either inconsequential, or makes the intermediate types worse off, implying that there is no longer a rationale for an understatement of preferences by the intermediate types. Likewise, increasing s^1 at the expense of s^x changes the outcome from $Q = 1$ into $Q = 2$, if anything, so that there is neither an incentive to exaggerate public-goods preferences.

A second-best provision rule can be implemented by the following voting mechanism: ask each individual for his preferred alternative. The outcome is $Q = 2$ if and only if the population share of people in favor of $Q = 2$ exceeds a threshold level. Likewise, the outcome is $Q = 0$ if and only if the population share of people in favor of $Q = 0$ exceeds some other threshold level. In all other cases, the intermediate provision level $Q = 1$ is implemented. This voting mechanism differs from the one in the model with two provision levels because an action set with three alternatives is needed. The common feature is that only ordinal information on preferences can be used for a decision on public-goods provision: whether the outcome is $Q = 2$ or not depends only on how many individuals like this outcome best, and not on the preference intensities of the average person who does not like this outcome best. Likewise, whether or not the outcome $Q = 0$ is obtained depends only on s^0. It does not matter how big s^x is and how big s^1 is as long as $s^x + s^1 = 1 - s^0$.

9 Concluding Remarks

Our main subject in this paper has been the problem of mechanism design for public-good provision in a large economy with prior uncertainty as to whether it is efficient for the public good to be provided or not. In this economy, conditions for individual incentive compatibility are simple because no one individual can affect the aggregate outcome. If there are no participation constraints, therefore, a social choice function that yields first-best outcomes can be implemented simply by asking people about their preferences and having them share the costs evenly if the public good is provided. In some instances, however, such schemes are implausible because they rely on information that (collectively) hurts the people who provide it; and people's willingness to provide this information is based solely on the consideration that, as individuals, they are unable to affect the outcome anyway. We impose a requirement of coalition-proofness to eliminate this possibility.

When coalition-proofness is imposed along with robustness, the implementability of a social

choice function that yields first-best outcomes can no longer be taken for granted. Social choice functions are robustly implementable and coalition-proof if and only if the provision can be characterized by a threshold such that the public good is provided if the population share of the net beneficiaries exceeds the threshold and is not provided if the population share of the net beneficiaries falls short of the threshold. Preference intensities cannot play a role. Net beneficiaries are the people for whom the benefits of the public good exceed the costs of the contribution they have to make; contributions are the same for all people and depend only on whether the public good is provided or not. Generally, such threshold rules cannot be used to implement first-best outcomes, because they are not responsive to the preference intensities of those who benefit and those who are harmed by public-good provision.

The main part of our analysis has been based on a model with a continuum of agents even though our main result extends to a finite economy. The reason for our emphasis of the continuum economy is twofold. First, in the continuum economy, one does not have to keep track of the small probability events in which a single individual may be decisive for the decision on public-goods provision. This makes it much easier to get an intuitive understanding of the main result. Second, the large economy framework is the standard paradigm for the normative and positive analysis of allocation problems involving private goods. The contribution of this paper is to provide, within this paradigm, a treatment of the information and incentive problems that are specific to problems of public-good provision.

References

Austen-Smith, D. and Banks, J. (1996). Information Aggregation, Rationality and the Condorcet Jury Theorem. *American Political Science Review*, 90:34–45.

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

Bierbrauer, F. (2010). Optimal income taxation and public-goods provision with preference and productivity shocks. Preprint 2010/18, Max Planck Institute for Research on Collective Goods.

Bierbrauer, F. and Hellwig, M. (2011). The Samuelson critique of a voluntary exchange theory of public finance: an incentive-theoretic perspective. Working Paper, Max Planck Institute for Research on Collective Goods.

Bierbrauer, F. and Hellwig, M. (2010). Public-good Provision in a large economy. Preprint 2010/02, Max Planck Institute for Research on Collective Goods.

Bierbrauer, F. and Sahm, M. (2010). Optimal democratic mechanisms for income taxation and public-goods provision. *Journal of Public Economics*, 94:453-466.

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

Bergemann, D. and Morris, S. (2009). Robust implementation in direct mechanisms. *Review of Economic Studies*, 76:1175–1204.

Bernheim, B., Peleg, B., and Whinston, M. (1986). Coalition-proof Nash equilibria I. concepts. *Journal of Economic Theory*, 42:1–12.

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

Fudenberg, D., and Tirole, J. (1991). *Game Theory*. MIT Press, Cambridge, MA.

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.

Hellwig, M. (2011). Incomplete-Information Models of Large Economies with Anonymity: Existence and Uniqueness of Common Priors in. Preprint 2011/08, Max Planck Institute for Research on Collective Goods.

Hindriks, J., and Myles, G. (2006). *Intermediate Public Economics*. MIT Press, Cambridge, MA.

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

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.

Ledyard, J. (1978). Incentive compatibility and incomplete information. *Journal of Economic Theory*, 18:171–189.

Lindahl, E. (1919). *Die Gerechtigkeit der Besteuerung*. Lund.

Mailath, G. and Postlewaite, A. (1990). Asymmetric Information Bargaining Problems with Many Agents. *Review of Economic Studies*, 57:351–367.

Mas-Colell, A., Whinston, M., and Green, J. (1995). *Microeconomic Theory*. Oxford University Press, New York.

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.

Sun, Y. (2006). The exact law of large numbers via Fubini extension and characterization of insurable risks. *Journal of Economic Theory*, 126:31–69.

A Appendix

Proof of Proposition 3

The "if" part of the proposition is trivial. To prove the "only if" part, suppose that the social choice function F is robustly implementable and coalition-proof, i.e., that there exists a mechanism $f = (R, q, p)$ which reaches F by means of a strategy profile σ^* so that (i) σ^* is a Nash equilibrium on every type space with a belief system that satisfies Property P, and, in particular, on every type space with a degenerate belief system, and (ii) that on every such type space σ^* is coalition-proof.

The proof is by contradiction. Hence, suppose that, in the game induced by the direct mechanism $f_F = (T, q_F, p_F)$ on the type space $[(T, \mathcal{T}), \tau, \beta]$, the "honest" strategy profile h is blocked by the set $T' \subset T$. Let $\ell_{T'} : T' \to \mathcal{M}(T)$ be the associated "lying" strategy for types in T' and let $(\ell_{T'}, h_{T\backslash T'})$ be the strategy profile that results from having types in T' report payoff types according to $\ell_{T'}$ and types in $T\backslash T'$ report payoff types according to h. If the true cross-section distribution of types is δ, then by (3), the strategy profile $(\ell_{T'}, h_{T\backslash T'})$ gives rise to a cross-section distribution of reports $\Delta((\ell_{T'}, h_{T\backslash T'}), \delta, f_F)$ such that

$$\Delta(\hat{T}|(\ell_{T'}, h_{T\backslash T'}), \delta, f_F) = \int_{T'} \ell_{T'}(\hat{T}|t')d\delta(t') + \delta(\hat{T}\backslash T') \tag{19}$$

for any $\hat{T} \subset T$. Under the strategy profile h, the cross-section distribution of reports is just $\Delta(h, \delta, f_F) = \delta$. Because T' blocks the strategy profile h, we have

$$\int_T U(\hat{t} \mid t', (\ell_{T'}, h_{T\backslash T'}), f_F)d\ell_{T'}(\hat{t}|t') \geq U(t' \mid t', h, f_F), \tag{20}$$

for all $t' \in T'$, with a strict inequality for some $t' \in T'$. By (4) and by the definition of q_F and p_F, we have

$$\begin{aligned} U(\hat{t} \mid t', (\ell_{T'}, h_{T\backslash T'}), f_F) \;=\; & \int_{\mathcal{M}(T)} \{\tau(t')Q_F(\Delta((\ell_{T'}, h_{T\backslash T'}), \delta, f_F) \circ \tau^{-1}) \\ & -\bar{P}_F(\Delta((\ell_{T'}, h_{T\backslash T'}), \delta, f_F) \circ \tau^{-1})\} d\beta(\delta|t') \,, \end{aligned} \tag{21}$$

and

$$U(t' \mid t', h, f_F) \;=\; \int_{\mathcal{M}(T)} \{\tau(t')Q_F(\delta \circ \tau^{-1}) - \bar{P}_F(\delta \circ \tau^{-1})\} d\beta(\delta|t') \,, \tag{22}$$

for all $\hat{t} \in T$ and all $t' \in T'$.

In following we will show that an "equivalent" version of this manipulation exists under any other mechanism that implements the given social choice function in a robust and coalition-proof way. Hence, let $f = (R, q, p)$ and σ^* be any other mechanism and interim Nash equilibrium that implement F on $[(T, \mathcal{T}), \tau, \beta]$. Let $\sigma'_{T'} : T' \to \mathcal{M}(R)$ be given by setting

$$\sigma'_{T'}(R'|t') = \int_T \sigma^*(R'|\hat{t}) d\ell_{T'}(\hat{t}|t')$$

for any $R' \subset R$ and $t' \in T'$, and consider the strategy profile $(\sigma'_{T'}, \sigma^*_{T\setminus T'})$ that results from having types in T' reporting according to $\sigma'_{T'}$ and types in $T\setminus T'$ according to σ^*. If the cross-section distribution of types is δ, then by (3), the strategy profile $(\sigma'_{T'}, \sigma^*_{T\setminus T'})$ gives rise to a cross-section distribution of reports $\Delta((\sigma'_{T'}, \sigma^*_{T\setminus T'}), \delta, f)$ such that

$$\begin{aligned}\Delta(R'|(\sigma'_{T'}, \sigma^*_{T\setminus T'}), \delta, f) &= \int_{T'} \sigma'_{T'}(R'|t') d\delta(t') + \int_{T\setminus T'} \sigma^*(R'|t) d\delta(t) \\ &= \int_{T'} \int_T \sigma^*(R'|\hat{t}) d\ell_{T'}(\hat{t}|t')\, d\delta(t') + \int_{T\setminus T'} \sigma^*(R'|\hat{t}) d\delta(\hat{t}) \qquad (23)\end{aligned}$$

for all $R' \subset R$. Upon reversing the order of integration in the double integral on the right-hand side and using (19), we find that (23) can be rewritten in as

$$\Delta(R'|(\sigma'_{T'}, \sigma^*_{T\setminus T'}), \delta, f) = \int_T \sigma^*(R'|t) d\Delta(t|(\ell_{T'}, h_{T\setminus T'}), \delta, f_F)$$

and, therefore, that

$$\Delta((\sigma'_{T'}, \sigma^*_{T\setminus T'}), \delta, f) = \Delta(\sigma^*, \Delta((\ell_{T'}, h_{T\setminus T'}, f_F), \delta), f). \qquad (24)$$

Because the mechanism $f = (R, q, p)$ and strategy profile σ^* implement F, we also have

$$q(\Delta(\sigma^*, \delta, f)) = Q_F(\delta \circ \tau^{-1})$$

and

$$\sigma^*(\{r \in R | p(r, \Delta(\sigma^*, \delta, f)) = \bar{P}_F(\delta \circ \tau^{-1})\}|t) = 1$$

for all $t \in T$ and all $\delta \in \mathcal{M}(T)$. By the definition of $\sigma'_{T'}$ and by (24), it follows that

$$q(\Delta((\sigma'_{T'}, \sigma^*_{T\setminus T'}), \delta, f)) = Q_F(\Delta(\sigma^*, \Delta((\ell_{T'}, h_{T\setminus T'}), \delta, f_F)) \circ \tau^{-1})$$

and

$$\sigma'_{T'}(\{r \in R | p(r, \Delta((\sigma'_{T'}, \sigma^*_{T\setminus T'}), \delta, f)) = \bar{P}_F(\Delta_f(\sigma^*, \Delta((\ell_{T'}, h_{T\setminus T'}), \delta, f_F)) \circ \tau^{-1})\}|t') = 1$$

for all $t' \in T'$ and all $\delta \in \mathcal{M}(T)$. By (4) and (21), therefore,

$$U(r \mid t', (\sigma'_{T'}, \sigma^*_{T\setminus T'}), f) = U(\hat{t} \mid t', (\ell_{T'}, h_{T\setminus T'}), f_F)$$

for $\sigma'_{T'}(t')$-almost all $r \in R$, $\ell_{T'}(t')$-almost all $\hat{t} \in T$, and all $t' \in T'$. By (4) and (22), we also have

$$U(r \mid t', \sigma^*, f) = U(t' \mid t', h, f_F)$$

for σ^*-almost all $r \in R$ and all $t' \in T'$. Thus,

$$\int_R U(r \mid t', (\sigma'_{T'}, \sigma^*_{T\setminus T'}), f) d\sigma'_{T'}(r|t') = \int_T U(\hat{t} \mid t', (\ell_{T'}, h_{T\setminus T'}), f_F) d\ell_{T'}(\hat{t}|t')$$

and

$$\int_R U(r \mid t', \sigma^*, f) d\sigma^*(r) = U(t' \mid t', h, f_F)$$

for all $t' \in T'$. By (20), it follows that

$$\int_R U(r \mid t', (\sigma'_{T'}, \sigma^*_{T\setminus T'}), f) d\sigma'_{T'}(r|t') \geq \int_R U(r \mid t', \sigma^*, f) d\sigma^*(r)$$

for all $t' \in T'$, with a strict inequality for some $t' \in T'$.

The set of types T' that possess a lie that blocks the truthtelling equilibrium for the direct mechanism f_F thus also possesses a deviation $\sigma_{T'}$ that makes them better off relative to the the equilibrium σ^* for the mechanism f. To establish that $(T', \sigma_{T'})$ blocks the equilibrium σ^* it therefore remains to be shown that $(\sigma_{T'}, \sigma^*_{T\setminus T'})$ is a Nash equilibrium on the given type space. This follows from the following observations: All individuals, the deviating ones as well as the non-deviating ones, choose, almost surely, reports that belong to the support of $\sigma^*(t)$, for some $t \in T$. The fact that σ^* is a Nash equilibrium on every type space with a degenerate belief system implies that all individuals are giving a best response if they behave according to $(\sigma_{T'}, \sigma^*_{T\setminus T'})$ on the given type space. ■

Proof of Proposition 4

If the robustly implementable anonymous social choice function $F = (Q_F, \bar{P}_F)$ is coalition-proof, then, by Proposition 3, for every $s \in \mathcal{M}(V)$, the "honest" strategy profile h is a coalition-proof interim Nash equilibrium for the game induced by the direct mechanism $f_F = (V, Q_F, \bar{P}_F)$ on the naive type space $[(V, \mathcal{V}), \beta_s]$. Thus, there exists no collective deviation $\pi = (V', \ell'_{V'})$ such that $(\ell'_{V'}, h_{V\setminus V'})$ is an interim Nash equilibrium for the game induced by $f_F = (V, Q_F, \bar{P}_F)$ on the type space $[(V, \mathcal{V}), \beta_s]$ and, moreover, the inequality (12) holds for all $v' \in V'$, where, for some $v' \in V'$, the inequality is strict. Because individual announcements do not affect outcomes, any strategy profile $(\ell'_{V'}, h_{V\setminus V'})$ that is induced by a collective deviation $\pi = (V', \ell'_{V'})$ is an interim Nash equilibrium for the game induced by $f_F = (V, Q_F, \bar{P}_F)$. The proposition follows immediately. ■

Proof of Proposition 7

Part A. We first show that statement i) in Proposition 7 is true. Suppose there is $\bar{s} = (\bar{s}^0, \bar{s}^x, \bar{s}^1)$ so that $Q_F(\bar{s}^0, \bar{s}^x, \bar{s}^1) = 2$.

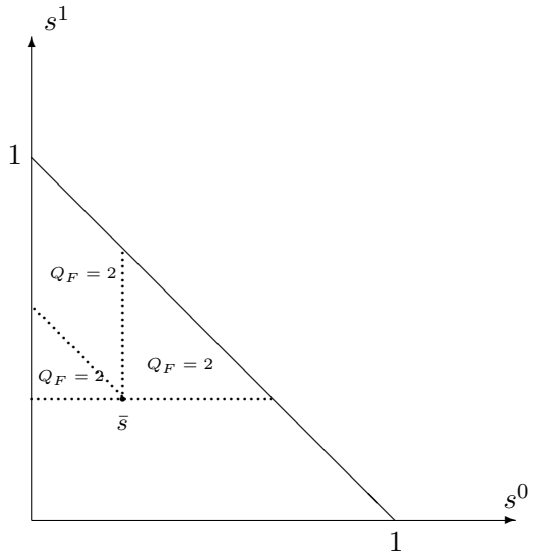
Step 1. It has to be true that for all $s = (s^0, s^x, s^1)$ with $s^0 = \bar{s}^0$ and $s^1 \geq \bar{s}^1$, $Q_F(s) = 2$. Otherwise there would be a state $\hat{s}$ so that $\hat{s}^0 = \bar{s}^0$, and $\hat{s}^1 > \bar{s}^1$, but $\tilde{U}(\hat{s} \mid 1) < \tilde{U}(\bar{s} \mid 1)$, a contradiction to Corollary 4. Analogously, it has to be true that for all $s = (s^0, s^x, s^1)$ with $s^x = \bar{s}^x$ and $s^1 \geq \bar{s}^1$, $Q_F(s) = 2$. Repeating this argument implies that $Q_F(s) = 2$, for all s so that $s^0 \leq \bar{s}^0$ and $s^0 + s^1 \geq \bar{s}^0 + \bar{s}^1$. The argument is illustrated graphically in Figure 3. The

vertical line starting at $\bar{s}$ is the locus of points so that $s^0 = \bar{s}^0$ and $s^1 \geq \bar{s}^1$. The downward-sloping line is the locus of points so that $s^1 \geq \bar{s}^1$ and $s^x = \bar{s}^x$. If $Q_F(\bar{s}) = 2$, then we need to have $Q_F(s) = 2$, for all s in the trapezoid in the northwest of $\bar{s}$.

Step 2. It has to be true that for all $s = (s^0, s^x, s^1)$ with $s^0 > \bar{s}^0$ and $s^1 = \bar{s}^1$, $Q_F(s) = 2$. Otherwise there would be a state $\hat{s}$ so $\hat{s}^1 = \bar{s}^1$, $\hat{s}^x < \bar{s}^x$, and $\tilde{U}(\hat{s} \mid x) > \tilde{U}(\bar{s} \mid x)$, a contradiction to Corollary 4. Analogously, it has to be true that for all $s = (s^0, s^x, s^1)$ with $s^0 = \bar{s}^0$ and $s^1 > \bar{s}^1$, $Q_F(s) = 2$. Repeating this argument implies that $Q_F(s) = 2$, for all s so that $s^0 \geq \bar{s}^0$ and $s^1 \geq \bar{s}^1$. The argument is illustrated graphically in Figure 3. If $Q_F(\bar{s}) = 2$, then we need to have $Q_F(s) = 2$, for all s in the triangle in the northeast of $\bar{s}$.

Step 3. It has to be true that for all $s = (s^0, s^x, s^1)$ with $s^0 < \bar{s}^0$ and $s^1 = \bar{s}^1$, $Q_F(s) = 2$. Otherwise there would be a state $\hat{s}$ so $\hat{s}^1 = \bar{s}^1$, $\hat{s}^0 < \bar{s}^0$, and $\tilde{U}(\hat{s} \mid 0) > \tilde{U}(\bar{s} \mid 0)$, a contradiction to Corollary 4. Analogously, it has to be true that for all $s = (s^0, s^x, s^1)$ with $s^x = \bar{s}^x$ and $s^0 < \bar{s}^0$, $Q_F(s) = 2$. Repeating this argument implies that $Q_F(s) = 2$, for all s so that $s^0 \leq \bar{s}^0$ and $s^0 + s^1 \leq \bar{s}^0 + \bar{s}^1$. The argument is illustrated graphically in Figure 3. If $Q_F(\bar{s}) = 2$, then we need to have $Q_F(s) = 2$, for all s in the triangle in the northwest of $\bar{s}$.

Figure 3: Implications of $Q_F(\bar{s}) = 2$



Part B. We now show that statement ii) in Proposition 7 is true. Suppose that there is some $\bar{s}$ so that $Q_F(\bar{s}) = 2$, and some $\hat{s}$ so that $Q_F(\hat{s}) = 0$.

By part i) there exists a number $\tilde{s}^1 > 0$ so that $Q_F(s) = 2$ if and only if $s^1 \geq \tilde{s}^1$. This situation is illustrated in Figure 4. In the region below the $s^1 = \tilde{s}^0$-line the provision level has to be either 0 or 1.

We can now adapt the argument from part *A* to show that, if there is a state $\hat{s}$ with $Q_F(\hat{s}) = 0$, then it has to be the case that $Q_F(s) = 0$, for all s with $s^0 \geq \hat{s}^0$. The details of this exercise are left to the reader.

■

B The finite economy

We now show that our main results in Propositions 5 and 6 extend to an economy with finitely many individuals. In the finite economy, the analysis is more complicated because an individual's action not only affects the own payment, but possibly also the decision on public-goods provision. Consequently, a group deviation does not automatically satisfy the condition that each deviating individual's action is a best response to the behavior of all other individuals. However, as will be shown below in Lemma 2, individuals do generally have multiple best responses. As we will see, this leaves sufficient scope for the formation of coalitions that are individually incentive-compatible so that our main result goes through. All proofs for the economy with finitely many individuals are in Section B.3.

As in the body of the text, we stick to the assumption that blocking manipulations cannot make use of on side payments. In an economy with finitely many individuals, this assumption involves a loss of generality. We conjecture, however, that one could prove a limit result so that, as the number of individuals goes to infinity, the effectiveness of side-payments as a means of facilitating coalition formation goes to zero. In fact, such limit results have already been provided by Mailath and Postlewaite (1990) for an independent private values model and by Neeman (2004) for a model which allows for a much larger set of type spaces.

B.1 The model

This section provides an adaptation of our model to an economy with finitely many individuals.

Individuals. The set of individuals is given by $I = \{1, \ldots, n\}$. Individual i has utility a function $u_i = v_i Q - P_i$, where $v_i \in V$ is the individual's valuation of a public good (i's payoff type), $Q \in \{0, 1\}$ and P_i is a monetary payment. We assume that V is a bounded subset of $\mathbb{R}_+$, that contains 0 as its smallest element and $\bar{v}$ as its largest element. We write $v = (v_1, \ldots, v_n)$ for a typical vector that lists the payoff types of all individuals.

Individual i's type t_i belongs to a set of possible types T, which is taken to be the same for all individuals. Each individual privately observes the own type. An individual's payoff type is determined by the function $\tau : T \to V$. We assume throughout that τ is surjective. We will occasionally use the shorthand notation $\tau(t) = (\tau(t_i), \ldots \tau(t_n))$ for the vector of payoff types that is induced by a vector of types $t = (t_1, \ldots, t_n)$.

An individual's belief type is determined by the function $\beta : T \to \mathcal{M}(T_{-i})$, where $\mathcal{M}(T_{-i})$ is the set of probability measures over the possible types of all individuals, except individual i, $T_{-i} = T^{n-1}$. The collection $[T, \tau, \beta]$ is referred to as a type space.

Mechanisms. A mechanism $f = (R_1, \ldots, R_n, q, p_1, \ldots, p_n)$ consists of: (i) for each individual i, a set of feasible reports R_i; we write $r = (r_1, \ldots, r_n)$ for a typical vector that lists the reports of all individuals; (ii) a public-goods provision rule $q : \prod_i R_i \to \{0, 1\}$, which determines, for each vector of reports, whether or not the public good is provided; and (iii) for each individual i, a function $p_i : \prod_i R_i \to \mathbb{R}$, which determines i's payment as a function of all reports. We will occasionally write $R_{-i} = \prod_{j \neq i} R_j$ for the set of possible reports by all individuals except i.

Given a mechanism, a (mixed) strategy for player i is a function $\sigma_i : T \to \mathcal{M}(R_i)$. Hence, we think of the report that is sent by type t_i of individual i as a random variable r_i, which is distributed according to a probability distribution $\sigma_i(t_i)$. In the following, we denote by $\sigma(R_i' \mid t_i)$ the probability according to which type t_i of player i chooses a message in a subset R_i' of R_i. We denote by $R_i^+(\sigma, t_i)$ the smallest subset R_i' of R_i so that $\sigma(R_i' \mid t_i) = 1$.

A strategy profile is in the following written as $\sigma = (\sigma_1, \ldots, \sigma_n)$. Let $t = (t_i, \ldots, t_n)$ be a given type profile. Conditional on this type profile, the profile of reports received by the mechanism is a random variable $r = (r_1, \ldots, r_n)$. We denote by $\sigma(t) = \prod_i \sigma_i(t_i)$ the probability distribution of r conditional on t.

Given a mechanism f and a type space, a strategy profile σ^* is called an interim Nash equilibrium provided that for all i, and all t_i,

$$R_i^+(\sigma_i^*, t_i) \in \operatorname{argmax}_{r_i \in R_i} \int_{T_{-i}} \int_{R_{-i}} u(f(r_{-i}, r_i), \tau(t_i))\, d\sigma_{-i}^*(r_{-i} \mid t_{-i})\, d\beta(t_{-i} \mid t_i)\ , \tag{25}$$

where

$$u(f(r_{-i}, r_i), \tau(t_i)) := \tau(t_i) q(r_{-i}, r_i) - p_i(r_{-i}, r_i)\ .$$

In the following, we denote the expected payoff of type t_i of individual i under a strategy profile σ by

$$U_i(\sigma, t_i, f) := \int_{T_{-i}} \int_{R} u(f(r), \tau(t_i))\, d\sigma(r \mid t_{-i})\, d\beta(t_{-i} \mid t_i)\ .$$

Social Choice Functions. A social choice function $F = (Q_F, P_{F1}, \ldots, P_{Fn})$ consists of (i) a public-goods provision rule $Q_F : V^n \to \{0, 1\}$, which determines, for each vector of payoff types, whether or not the public good is provided, and (ii) for each individual i, a function $P_{Fi} : V^n \to \mathbb{R}$, which determines i's payment.

A social choice function F is said to be implementable on a given type space if there exists a mechanism f with an interim Nash equilibrium σ^* so that for all $t = (t_1, \ldots, t_n)$,

$$q(r) = Q_F(\tau(t)) \quad \text{and} \quad p_i(r) = P_{Fi}(\tau(t)), \quad \text{for all } i\ , \tag{26}$$

$\sigma^*(t)$-almost surely.

Robust Implementability. Given a set T of types and a function τ, a social choice function F is said to be robustly implementable if there is a mechanism f with an equilibrium σ^* so that (25) and (26) hold on every type space $[T, \tau, \beta]$.

The following lemma, that we state without proof, is due to Ledyard (1978) and Bergemann and Morris (2005): Robust implementability is equivalent to implementability by means of a mechanism such that individuals announce payoff types, and such that truthtelling is a dominant strategy equilibrium. The lemma is the finite economy analogue to Proposition 1.

Lemma 1 *A social choice function F is robustly implementable if and only if it is payoff-type dominant strategy incentive compatible: for all i, all v_i, all v_i', and all v_{-i},*

$$v_i Q_F(v_{-i}, v_i) - P_{Fi}(v_{-i}, v_i) \geq v_i Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i')\ . \tag{27}$$

The next Lemma provides a complete characterization of robustly implementable social choice functions.

Lemma 2 *A social choice function F is robustly implementable if and only if it has the following properties: For every i, and every v_{-i}, there exists a cutoff type $c_i(v_{-i}) \geq 0$ so that*

$$Q_F(v_{-i}, v_i) = \begin{cases} 0, & \text{if} \quad v_i < c_i(v_{-i}) \ , \\ 1, & \text{if} \quad v_i \geq c_i(v_{-i}) \ . \end{cases} \tag{28}$$

If $0 < c_i(v_{-i}) \leq \bar{v}$, then there exist numbers $P^0_{Fi}(v_{-i})$ and $P^1_{Fi}(v_{-i})$ so that

$$P_{Fi}(v_{-i}, v_i) = \begin{cases} P^0_{Fi}(v_{-i}), & \text{if} \quad v_i < c_i(v_{-i}) \ , \\ P^1_{Fi}(v_{-i}), & \text{if} \quad v_i \geq c_i(v_{-i}) \ , \end{cases} \tag{29}$$

where

$$\inf\{v_i \in V \mid v_i > c_i(v_{-i})\} \geq P^1_{Fi}(v_{-i}) - P^0_{Fi}(v_{-i}) \geq \sup\{v_i \in V \mid v_i \leq c_i(v_{-i})\} \ . \tag{30}$$

If $c_i(v_{-i}) > \bar{v}$, or $c_i(v_{-i}) = 0$ then there exists a number $\bar{P}_{Fi}(v_{-i})$ so that

$$P_{Fi}(v_{-i}, v_i) = \bar{P}_{Fi}(v_{-i}) \ , \tag{31}$$

for all v_i.

Consider a direct mechanism that implements a social choice function with the properties in Lemma 2 in a truth-telling equilibrium. Observe that on every complete information type space, i.e., a type space where all individuals assign probability 1 to a specific payoff type profile v, individuals have multiple best responses. If $Q_F(v) = 0$, then every individual is willing to understate his preferences since this has neither an impact on the provision level, nor on the individual's payment. Likewise, if $Q_F(v) = 1$, then every individual is willing to exaggerate his preferences. In the following, we will show that these multiple best responses generate a degree of freedom for incentive-compatible coalition formation. As a preliminary step, however, we need to adapt our notion of a coalition-proof equilibrium to our model of a finite economy.

Coalition-proof equilibrium. Fix a mechanism f and a type space. An interim Nash equilibrium σ^* is said to be coalition-proof if the following does not exist: A deviation by a set of individuals $I' \subset I$ to a strategy profile $\sigma'_{I'} = (\sigma'_i)_{i \in I'}$ being such that

i) The strategy $(\sigma'^*_{I \setminus I'}, \sigma'_{I'})$ is an interim Nash equilibrium.

ii) For all $i \in I'$ and all t_i so that $\sigma'_i(t_i) \neq \sigma^*(t_i)$, $U_i((\sigma'^*_{I \setminus I'}, \sigma'_{I'}), t_i, f) \geq U_i(\sigma^*, t_i, f)$. Moreover, there is at least one $i \in I'$ with a type t_i so that $U_i((\sigma'^*_{I \setminus I'}, \sigma'_{I'}), t_i, f) > U_i(\sigma^*, t_i, f)$.

Robust and coalition-proof social choice functions. Given a set T of types and a function τ, a social choice function F is said to be robustly and coalition-proof if there is a mechanism f and strategy σ^* such that (i) σ^* is a coalition-proof interim Nash equilibrium on every type space $[T, \tau, \beta]$, and (ii) for every t, the equilibrium allocation coincides with the allocation stipulated by the social choice function, i.e., (26) holds.

The following Lemma states a necessary condition for robustness and coalition-proofness. It is the finite-economy-version of Proposition 4. In the finite economy, the condition that each deviators is, individually, giving a best response is not trivially fulfilled. Hence, there are participation and an incentive constraints that need to be satisfied. Proposition 4, by contrast, stated only a participation constraint.

Lemma 3 *If a social choice function is robust and coalition-proof, then the following does not exist: a profile of payoff types v, a subset I' of I with a deviation $v'_{I'} \neq v_{I'}$ so that*

i) The deviators benefit. For all $i \in I'$,

$$v_i Q_F(v_{I\setminus I'}, v'_{I'}) - P_{Fi}(v_{I\setminus I'}, v'_{I'}) \geq v_i Q_F(v_{I\setminus I'}, v_{I'}) - P_{Fi}(v_{I\setminus I'}, v_{I'}) \ ,$$

with a strict inequality for at least some $i \in I'$.

ii) Each deviator gives a best response. For all $i \in I'$,

$$v_i Q_F(v_{I\setminus I'}, v'_{I'-i}, v'_i) - P_{Fi}(v_{I\setminus I'}, v'_{I'-i}, v'_i) \geq v_i Q_F(v_{I\setminus I'}, v'_{I'-i}, v_i) - P_{Fi}(v_{I\setminus I'}, v'_{I'-i}, v_i) \ .$$

The *inequalities* in Lemma 3 state properties of a social choice function that are necessary for robustness and coalition-proofness. The *words* in Lemma 3 are based on the interpretation of such a social choice function as a direct mechanism in which individuals communicate their payoff types and in which truthtelling is a dominant strategy equilibrium. A truthtelling equilibrium is coalition-proof only if there is no profile of payoff types so that some individuals benefit from lying (property i)) and choose an action that is as good as the truth (property ii)).

We could add a third requirement, namely that *each non-deviator is giving a best response:* For all $j \in I \setminus I'$, and all v'_j,

$$v_i Q_F(v_{I\setminus I'}, v'_{I'-j}, v_j) - P_{Fi}(v_{I\setminus I'}, v'_{I'-j}, v_j) \geq v_i Q_F(v_{I\setminus I'}, v'_{I'-j}, v'_j) - P_{Fi}(v_{I\setminus I'}, v'_{I'-j}, v'_j) \ .$$

However, since robust implementability of a social choice function is equivalent to dominant strategy incentive compatibility, this requirement will be trivially fulfilled by any social choice function that is of interest to us.

B.2 The main result in the finite economy

In the finite economy we replace the condition of anonymity by the weaker condition of symmetry.[23] This condition says that individuals with the same payoff types make the same payment

[23] Symmetry in the finite economy is weaker in that a single individual may be pivotal for whether or not the public good is provided. In a continuum economy and under anonymity this is impossible.

and, moreover, that a permutation of the individuals' types does not affect the decision on public-goods provision. Formally, we say that a social choice function is *symmetric*, if, for every v, and every pair of individuals $(i,j) \in I^2$, $P_{Fi}(v_{-i-j}, v_i, v_j) = P_{Fj}(v_{-i-j}, v_j, v_i)$ and $Q_F(v_{-i-j}, v_i, v_j) = Q_F(v_{-i-j}, v_j, v_i)$.

The following Proposition, which is proven in the Appendix, is the finite economy analogue to Proposition 5.

Proposition 8 *A symmetric social choice function F is robustly implementable and coalition-proof only if there exist numbers P_F^0 and P_F^1, so that for all v and all i,*

$$P_{Fi}(v) = \begin{cases} P_F^0 & \text{if} \quad Q_F(v) = 0\,, \\ P_F^1 & \text{if} \quad Q_F(v) = 1\,. \end{cases} \tag{32}$$

Since robust implementability implies, in particular, $MLRP$-robust implementability, the necessary condition in Proposition 8 applies if we limit attention to belief systems satisfying $MLRP$. If we do this, we can show that the implementability by a voting mechanism is both necessary and sufficient for the robustness and coalition-proofness of a social choice function. This follows from Proposition 9 which is the finite economy analogue of Proposition 6. The proof is omitted because it parallels the one of Proposition 6.

Proposition 9 *Consider a social choice function $F = (Q_F, P_F^0, P_F^1)$ and suppose that $P_F^1 - P_F^0 \cap V = \emptyset$. Denote by $s_1(v) := \#\{i \mid v_i > P_F^1 - P_F^0\}$ the number of individuals who are net gainers from public-good provision, given a payoff profile v. This social choice function is $MLRP$-robustly implementable and coalition-proof if and only if for all v and v',*

$$s_1(v) \geq s_1(v') \quad \text{implies} \quad Q_F(v) \geq Q_F(v'). \tag{33}$$

The basic insight is the one previously obtained for an economy with a continuum of individuals: Under robustness, coalition-proofness and symmetry there is a payment P_F^1 that everybody has to deliver if the public good is provided and another payment P_F^0 that is relevant if the public good is not provided. Given these payments we can define the set of individuals who prefer non-provision over provision. Under robustness and coalition-proofness, the decision on public-goods provision can only reflect the number of individuals within and outside this set. Information on preference intensities cannot be used.

B.3 Proofs

Proof of Lemma 2

We first show that a social choice function satisfying equations (28) - (31) also satisfies (27) and hence is robustly implementable. Fix i, v_{-i}, and v_i. (i) If $c_i(v_{-i}) > \bar{v}$, then $v_i Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i') = -\bar{P}_i(v_{-i})$ for any $v_i' \in V_i$. Likewise, if $c_i(v_{-i}) = 0$, then $v_i Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i') = v_i - \bar{P}_i(v_{-i})$, for any $v_i' \in V_i$. Since these expressions do net depend on v_i', dominant strategy

incentive compatibility is trivially fulfilled. (ii) Now Suppose that $0 < c_i(v_{-i}) \leq \bar{v}$. Assume first that $v_i < c_i(v_{-i})$. Then, $v_i' < c_i(v_{-i})$, implies that $v_i Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i') = -P_i^0(v_{-i})$; and $v_i' \geq c_i(v_{-i})$ implies that $v_i Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i') = -P_i^0(v_{-i})$. Hence, (27) holds if and only if

$$P_{Fi}^1(v_{-i}) - P_{Fi}^0(v_{-i}) \geq v_i \ .$$

By (30) this is fulfilled for all $v_i < c_i(v_{-i})$. A symmetric argument establishes that (27) holds under the assumption that $v_i \geq c_i(v_{-i})$.

We now show that if a social choice function satisfies (27) then it does also satisfy (28) - (31). Fix i, and v_{-i}. (i) Upon adding the two incentive constraints $v_i Q_F(v_{-i}, v_i) - P_{Fi}(v_{-i}, v_i) \geq v_i Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i')$ and $v_i' Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i') \geq v_i' Q_F(v_{-i}, v_i) - P_{Fi}(v_{-i}, v_i)$, we find that $(v_i - v_i')(Q_F(v_{-i}, v_i) - Q_F(v_{-i}, v_i')) \geq 0$. Hence, $v_i \geq v_i'$ implies $Q_F(v_{-i}, v_i) \geq Q_F(v_{-i}, v_i')$, and hence the existence of a cutoff level $c_i(v_{-i})$ so that (28) holds. (ii) Let $Q_F(v_{-i}, v_i) = Q_F(v_{-i}, v_i')$, then the two incentive constraints imply that $P_{Fi}(v_{-i}, v_i) = P_{Fi}(v_{-i}, v_i')$, and hence that (29) and (31) hold true. (iii) Finally, if $v_i < c_i(v_{-i}) \leq v_i'$, then the incentive constraint $v_i Q_F(v_{-i}, v_i) - P_{Fi}(v_{-i}, v_i) \geq v_i Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i')$ simplifies to $-P_{Fi}^0(v_{-i}) \geq v_i - P_{Fi}^1(v_{-i})$. If this holds for all $v_i < c_i(v_{-i})$, then it must be true that $P_{Fi}^1(v_{-i}) - P_{Fi}^0(v_{-i}) \geq \sup\{v_i \in V \mid v_i \leq c_i(v_{-i})\}$. Analogously, for $v_i' < c_i(v_{-i}) \leq v_i$, the incentive constraint $v_i Q_F(v_{-i}, v_i) - P_{Fi}(v_{-i}, v_i) \geq v_i Q_F(v_{-i}, v_i') - P_{Fi}(v_{-i}, v_i')$ simplifies to $v_i - P_{Fi}^1(v_{-i}) \geq -P_{Fi}^0(v_{-i})$. If this holds for all $v_i \geq c_i(v_{-i})$, then it must be true that $\inf\{v_i \in V \mid v_i > c_i(v_{-i})\} \geq P_{Fi}^1(v_{-i}) - P_{Fi}^0(v_{-i})$. This shows that (30) holds. ■

Proof of Lemma 3

The proof is by contradiction. We consider a robustly implementable social choice function and suppose that there is a profile of payoff types v, a subset I' of I with a deviation $v_{I'}' \neq v_{I'}$ so that properties i) and ii) hold and show that this yields a contradiction to the assumption that F is not only robustly implementable but also coalition-proof.

Let f be a mechanism, so that σ^* is an interim Nash equilibrium on every type space and moreover suppose that (26) holds, so that f implements the social choice function.

Consider a type space so that it is commonly known among individuals that the type profile t is such that $\tau(t) = v$. More formally, the belief system β is such that $\beta(t_{-i} \mid t_i) = 1$ for all i. We show in the following that, on this type space, σ^* is not coalition-proof.

Consider a strategy profile $\bar{\sigma} := (\sigma_{I \setminus I'}^{\prime *}, \sigma_{I'}')$, where, for any $i \in I'$, $\sigma_i'(t_i) = \sigma^*(\tau^{-1}(v_i'))$. Note that σ' is constructed in such a way that each individual in I' behaves as if he still followed σ^*, but had a payoff type of $v_i' \neq v_i$

Step 1. We first show that, on the given type space, $U_i(\sigma', t_i, f) \geq U_i(\sigma^*, t_i, f)$, for all $i \in I'$ and all t_i so that $\sigma_i'(t_i) \neq \sigma_i^*(t_i)$; with a strict inequality for at least one type of at least one $i \in I'$. Given that $\beta(t_{-i} \mid t_i) = 1$ for all i, and given that (26) holds,

$$U_i(\bar{\sigma}, t_i, f) = v_i Q_F(v_{I \setminus I'}, v_{I'}) - P_{Fi}(v_{I \setminus I'}, v_{I'}) \ .$$

The construction of $\bar{\sigma}$ and (26) imply that

$$U_i(\sigma', t_i, f) = v_i Q_F(v_{I \setminus I'}, v_{I'}') - P_{Fi}(v_{I \setminus I'}, v_{I'}')$$

Hence, i) implies that $U_i(\sigma', t_i) \geq U_i(\sigma^*, t_i)$, for all $i \in I'$ and all t_i so that $\sigma'_i(t_i) \neq \sigma^*(t_i)$; with a strict inequality for at least one type of at least one $i \in I'$.

Step 2. It remains to be shown that $\bar{\sigma}$ is an interim Nash equilibrium on the given type space.

(a) We show that, for every $i \in I'$, there is no message $r_i \in R_i$ that yields a higher payoff than behaving according to $\sigma'_i(t_i)$. Suppose otherwise, then there is i and $r_i \in R_i$ so that

$$\int_{R_{-i}} u(f(r_{-i}, r_i), \tau(t_i))\, d\bar{\sigma}_{-i}(r_{-i} \mid t_{-i}) > U_i(\bar{\sigma}, t_i, f) = v_i Q_F(v_{I\setminus I'}, v'_{I'}) - P_{Fi}(v_{I\setminus I'}, v'_{I'})\ . \quad (34)$$

Property ii) in Lemma 3 and inequality (34) imply

$$\int_{R_{-i}} u(f(r_{-i}, r_i), \tau(t_i))\, d\sigma'_{-i}(r_{-i} \mid t_{-i}) > v_i Q_F(v_{I\setminus I'}, v'_{I'-i}, v_i) - P_{Fi}(v_{I\setminus I'}, v'_{I'-i}, v_i)\ . \quad (35)$$

Now consider a type space so that it is commonly known among individuals that the type profile equals t' where $t'_i = t_i$ and $\tau(t') = (v_{I\setminus I'}, v'_{I'-i}, v_i)$. Note that $\bar{\sigma}_{-i}(t_{-i}) = \sigma^*_{-i}(t'_{-i})$, so that inequality (35) can be rewritten as

$$\int_{R_{-i}} u(f(r_{-i}, r_i), \tau(t_i))\, d\sigma^*_{-i}(r_{-i} \mid t'_{-i}) > v_i Q_F(v_{I\setminus I'}, v'_{I'-i}, v_i) - P_{Fi}(v_{I\setminus I'}, v'_{I'-i}, v_i)\ . \quad (36)$$

Given that (26) holds the right-hand side of (36) equals

$$\int_{R_{-i}} u(f(r_{-i}, m_i^{**}), \tau(t_i))\, d\sigma^*_{-i}(r_{-i} \mid t'_{-i})$$

for some $m_i^{**} \in M_i^+(\sigma_i^*, t_i)$. Hence, inequality (36) implies that there exists $r_i \in R_i$ and $m_i^{**} \in M_i^+(\sigma_i^*, t_i)$ so that

$$\int_{R_{-i}} u(f(r_{-i}, r_i), \tau(t_i))\, d\sigma^*_{-i}(r_{-i} \mid t'_{-i}) > \int_{R_{-i}} u(f(r_{-i}, m_i^{**}), \tau(t_i))\, d\sigma^*_{-i}(r_{-i} \mid t'_{-i})\ .$$

But this implies that σ^* is not an interim Nash equilibrium on a type space where it is commonly known among individuals that the type profile equals t'. Hence, a contradiction to the assumption that σ^* is an interim Nash equilibrium on every type space.

(b) We now show that for every $i \in I \setminus I'$, there is no message $r_i \in R_i$ that yields a higher payoff than behaving according to $\sigma_i^*(t_i)$. Otherwise, there is $r_i \in R_i$ and so that

$$\int_{R_{-i}} u(f(r_{-i}, r_i), \tau(t_i))\, d\bar{\sigma}_{-i}(r_{-i} \mid t_{-i}) > U_i(\bar{\sigma}, t_i, f)\ ,$$

where $U_i(\bar{\sigma}, t_i, f) = v_i Q_F(v_{I\setminus I'}, v'_{I'-i}, v_i) - P_{Fi}(v_{I\setminus I'}, v'_{I'-i}, v_i)$ Hence,

$$\int_{R_{-i}} u(f(r_{-i}, r_i), \tau(t_i))\, d\bar{\sigma}_{-i}(r_{-i} \mid t_{-i}) > v_i Q_F(v_{I\setminus I'}, v'_{I'-i}, v_i) - P_{Fi}(v_{I\setminus I'}, v'_{I'-i}, v_i)\ ,$$

which is equivalent to inequality (35) above. We can now use the same arguments as in (a) to arrive at a contradiction to the assumption that σ^* is an interim Nash equilibrium on every type space. ■

Proof of Proposition 8

We say that a social choice function is *neutral* if it has the following two properties:

a) For any pair of individuals i and j, any v_{-i-j}, and any pair v_i and v_i': $Q_F(v_{-i-j}, v_i, 0) = Q_F(v_{-i-j}, v_i', 0) = 0$ and $P_{Fi}(v_{-i-j}, v_i, 0) = P_{Fi}(v_{-i-j}, v_i', 0)$ imply that $P_{Fj}(v_{-i-j}, v_i, 0) = P_{Fj}(v_{-i-j}, v_i', 0)$.

b) For any pair of individuals i and j, any v_{-i-j}, and any pair v_i and v_i': $Q_F(v_{-i-j}, v_i, \bar{v}) = Q_F(v_{-i-j}, v_i', \bar{v}) = 1$ and $P_{Fi}(v_{-i-j}, v_i, \bar{v}) = P_{Fi}(v_{-i-j}, v_i', \bar{v})$ imply that $P_{Fj}(v_{-i-j}, v_i, \bar{v}) = P_{Fj}(v_{-i-j}, v_i', \bar{v})$.

Neutrality requires that a change in individual i's payoff type, which is inconsequential both for the decision on public-goods provision and for i's payments, does not affect the payment of individual j if (a) j has the minimal valuation of the public good and the public good is not provided, or (b) j has the maximal valuation of the public good and the public good is provided.

Lemma 4 *If a symmetric social choice function is robust and coalition-proof, then it is neutral.*

Proof We only prove that a symmetric, robust and coalition-proof satisfies property a) in the definition of neutrality above. The proof of part b) is analogous.

Suppose that condition a) in the definition of neutrality is violated, i.e., suppose that there are individuals i and j, any v_{-i-j}, and a pair v_i and v_i' so that $Q_F(v_{-i-j}, v_i, 0) = Q_F(v_{-i-j}, v_i', 0) = 0$, $P_{Fi}(v_{-i-j}, v_i, 0) = P_{Fi}(v_{-i-j}, v_i', 0)$, but $P_{Fj}(v_{-i-j}, v_i, 0) \neq P_{Fj}(v_{-i-j}, v_i', 0)$. We show that this implies that coalition-proofness fails.

Suppose the true payoff type profile equals $(v_{-i-j}, v_i, 0)$. Consider a deviation from truthtelling by individuals i and j, and suppose they report instead $(v_i', 0)$. By dominant strategy incentive compatibility, individual j, who still reports truthfully, is giving a best response. Individual i's deviation neither affects the decision on public-goods provision, nor his payment. Hence, the deviation yields the same payoff as truthtelling, and therefore is a best response. Since individual i's payoff is unaffected he is willing to participate in the deviation. Coalition-proofness, therefore requires that individual j is not made strictly better off by the deviation, which requires that $P_{Fj}(v_{-i-j}, v_i, 0) \geq P_{Fj}(v_{-i-j}, v_i', 0)$. Since we hypothesized that $P_{Fj}(v_{-i-j}, v_i, 0) \neq P_{Fj}(v_{-i-j}, v_i', 0)$, it must be the case that $P_{Fj}(v_{-i-j}, v_i, 0) > P_{Fj}(v_{-i-j}, v_i', 0)$. Then, if the true payoff type profile equals $(v_{-i-j}, v_i', 0)$, a deviation by i and j to $(v_i, 0)$ is such that both are giving a best response, individual i is willing to participate and j is made strictly better off. Hence, a contradiction to coalition-proofness. ■

Lemma 5 *Let F by symmetric, robust and coalition-proof. Then it has the following properties:*

i) For every pair of individuals i and j, and for every v with $Q_F(v) = 0$,

$$Q_F(v_{-i-j}, 0, 0) = 0 \quad \textit{and} \quad P_{Fi}(v_{-i-j}, 0, 0) = P_{Fi}(v_{-i-j}, v_i, v_j)\ . \tag{37}$$

ii) For every pair of individuals i and j, and 1 for every v with $Q_F(v) = 1$,

$$Q_F(v_{-i-j}, \bar{v}, \bar{v}) = 1 \quad \textit{and} \quad P_{Fi}(v_{-i-j}, \bar{v}, \bar{v}) = P_{Fi}(v_{-i-j}, v_i, v_j). \tag{38}$$

iii) For every pair of individuals i and j, and for every v, $P_{Fi}(v_{-i-j}, v_i, v_j) = P_{Fj}(v_{-i-j}, v_i, v_j)$.

Proof We only proof part i). The proof of part ii) is analogous. Exploiting symmetry part iii) follows immediately from i) and ii).

Let F by symmetric, robust and coalition-proof. Fix some v so that $Q_F(v) = 0$ and suppose that two individuals $I' = \{i, j\}$ jointly announce $(0, 0)$ instead of (v_i, v_j).

Step 1. We first verify that this deviation satisfies property ii) in Lemma 3, i.e., both deviators are individually giving a best response: by (28) we have that for all k, and all v_{-k}, that $v'_k < v_k$ implies that $Q_F(v_{-k}, v'_k) \leq Q_F(v_{-k}, v_k)$. Hence, we have that

$$Q_F(v_{-i-j}, v_i, 0) = Q_F(v_{-i-j}, 0, v_j) = Q_F(v_{-i-j}, 0, 0) = 0\ . \tag{39}$$

Equations (28) and (29) then imply that

$$P_{Fi}(v_{-i-j}, 0, 0) = P_{Fi}(v_{-i-j}, v_i, 0) \quad \text{and} \quad P_{Fj}(v_{-i-j}, 0, 0) = P_{Fj}(v_{-i-j}, 0, v_j)\ . \tag{40}$$

Consequently,

$$v_i Q_F(v_{-i-j}, 0, 0) - P_{Fi}(v_{-i-j}, 0, 0) = v_i Q_F(v_{-i-j}, v_i, 0) - P_{Fi}(v_{-i-j}, v_i, 0)$$

and

$$v_j Q_F(v_{-i-j}, 0, 0) - P_{Fj}(v_{-i-j}, 0, 0) = v_j Q_F(v_{-i-j}, 0, v_j) - P_{Fi}(v_{-i-j}, 0, v_j)\ .$$

Hence, property ii) in Lemma 3 holds.

Step 2. Coalition-proofness therefore requires that either at least one individual does not benefit from this deviation,

$$P_{Fi}(v_{-i-j}, 0, 0) > P_{Fi}(v_{-i-j}, v_i, v_j) \quad \text{or} \quad P_{Fj}(v_{-i-j}, 0, 0) > P_{Fj}(v_{-i-j}, v_i, v_j)\ , \tag{41}$$

or, that both individuals are indifferent, which requires that

$$P_{Fi}(v_{-i-j}, 0, 0) = P_{Fi}(v_{-i-j}, v_i, v_j) \quad \text{and} \quad P_{Fj}(v_{-i-j}, 0, 0) = P_{Fj}(v_{-i-j}, v_i, v_j)\ . \tag{42}$$

We show in the following that (41) implies a contradiction to coalition-proofness, so that (42) has to be true. This will complete the proof, since by symmetry we have that $P_{Fi}(v_{-i-j}, 0, 0) = P_{Fj}(v_{-i-j}, 0, 0) =: P_{ij}(v_{-i-j}, 0, 0)$, so that (42) implies in particular that $P_{Fi}(v_{-i-j}, v_i, v_j) = P_{Fj}(v_{-i-j}, v_i, v_j)$.

The proof that (41) cannot be true proceeds by contradiction. Hence, suppose that these inequalities hold true. This gives rise to the following possibilities:

Case 1: $P_{ij}(v_{-i-j}, v_i, v_j) > P_{Fi}(v_{-i-j}, 0, 0)$ and $P_{ij}(v_{-i-j}, 0, 0) > P_{Fj}(v_{-i-j}, v_i, v_j)$. Then,

if the profile of payoff types equals $(v_{-i-j}, 0, 0)$, a deviation so that i and j jointly announce (v_i, v_j) instead of $(0, 0)$ makes both of them better off, and, by (40), is such that both are giving a best response. Hence, a contradiction to the assumption that F is coalition-proof.

Case 2: $P_{Fj}(v_{-i-j}, v_i, v_j) > P_{ij}(v_{-i-j}, 0, 0) > P_{Fi}(v_{-i-j}, v_i, v_j)$. It follows from equations (28) and (29) and from the symmetry of F that

$$P_{Fj}(v_{-i-j}, v_i, v_j) = P_{Fj}(v_{-i-j}, v_i, 0)$$

and

$$P_{Fi}(v_{-i-j}, v_i, v_j) = P_{Fi}(v_{-i-j}, 0, v_j) = P_{Fj}(v_{-i-j}, v_j, 0) \ .$$

Hence, condition $P_{Fj}(v_{-i-j}, v_i, v_j) > P_{ij}(v_{-i-j}, 0, 0) > P_{Fi}(v_{-i-j}, v_i, v_j)$ can be equivalently written as

$$P_{Fj}(v_{-i-j}, v_i, 0) > P_{Fi}(v_{-i-j}, 0, 0) > P_{Fj}(v_{-i-j}, v_j, 0) \ . \tag{43}$$

This is a contradiction to neutrality provided that

$$Q_F(v_{-i-j}, v_i, 0) = Q_F(v_{-i-j}, v_j, 0) \quad \text{and} \quad P_{Fi}(v_{-i-j}, v_i, 0) = P_{Fi}(v_{-i-j}, v_j, 0) \ . \tag{44}$$

To see that (44) holds, note that (39) and symmetry imply that $Q_F(v_{-i-j}, v_i, 0) = Q_F(v_{-i-j}, v_j, 0)$. Given that this is true, it follows from (28) and (29), that $P_{Fi}(v_{-i-j}, v_i, 0) = P_{Fi}(v_{-i-j}, v_j, 0)$.

Case 3: $P_{Fi}(v_{-i-j}, v_i, v_j) > P_{ij}(v_{-i-j}, 0, 0) > P_{Fj}(v_{-i-j}, v_i, v_j)$. A similar reasoning as in Case 2 can be used to arrive at a contradiction.

■

Corollary 5 *If a symmetric social choice function is robust and coalition-proof, then there exist numbers P_F^0 and P_F^1 so that, for all v, and all i,*

$$P_{Fi}(v) = \begin{cases} P_F^0 & \text{if} \quad Q_F(v) = 0 \ , \\ P_F^1 & \text{if} \quad Q_F(v) = 1 \ . \end{cases} \tag{45}$$

Proof We only show that there is a number P^0 so that, for all i, $P_{Fi}(v) = P^0$, whenever $Q_F(v) = 0$. It follows from Lemma 5 that, for all v, all individuals pay the same, i.e., $P_{Fi}(v) = P_{Fj}(v) := \bar{P}(v)$, for any pair (i, j). From part i) it follows that, starting from an arbitrary v, with $Q_F(v) = 0$, if we successively replace valuations v_i, that are possibly different from 0, by 0, the decision on provision and the payment $\bar{P}$ remain unaffected. Hence, if v is such that $Q_F(v) = Q_F(0, \ldots, 0) = 0$, then also $\bar{P}(v) = \bar{P}(0, \ldots, 0)$. ■