

**Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2015/12**



**Public-Good Provision
in Large Economies**

Felix J. Bierbrauer
Martin F. Hellwig



MAX PLANCK SOCIETY



Public-Good Provision in Large Economies

Felix J. Bierbrauer / Martin F. Hellwig

October 6, 2015

Public-Good Provision in Large Economies*

Felix J. Bierbrauer[†] and Martin F. Hellwig[‡]

October 6, 2015

Abstract

In a large economy, a first-best provision rule for a public good is robustly implementable with budget balance because no one individual alone can affect the aggregate outcome. First-best outcomes can, however, be blocked by coalitions of agents acting in concert. With a requirement of immunity against robustly blocking coalitions, we find that, for a public good that comes as a single indivisible unit, a monotonic social choice function cannot condition on preference intensities but only on the population shares of people favoring one outcome over another. Any such social choice function can be implemented by a simple voting mechanism. With more public-good provision levels, more complicated mechanisms are required, but they still involve the counting of votes rather than an assessment of benefits. Monotonicity and immunity against robust blocking thus provide a foundation for the use of voting mechanisms.

Keywords: Public-good provision, Mechanism Design, Robust Incentive Compatibility, Monotonic Social Choice Functions, Immunity against Robustly Blocking Coalitions, Voting Mechanisms, Large Economy

JEL: D60, D70, D82, H41

*We benefitted from discussions with Alia Gizatulina, Mike Golosov, Kristoffel Grechenig, Christian Hellwig, David Martimort, Benny Moldovanu, and Nora Szech.

[†]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

Large Economies. We study public-good provision in an economy with many participants where each individual participant is too insignificant to have a noticeable influence on aggregate outcomes. In other parts of economic theory and political economy, such large economies play a central role. For private goods, the large-economy paradigm is deemed to provide the proper framework for studying what happens when there are millions of people and none of them has the power to affect market prices. In public economics, the large-economy paradigm is used to study people’s responses to taxation when each person individually is too insignificant to have a noticeable impact on the government’s budget. In political economy, the large-economy paradigm is used to study voting when no individual expects to be pivotal for the outcome.

By contrast, most analyses of public-good provision involve small-economy models in which each individual can have a noticeable impact on aggregate outcomes.¹ This approach reflects the theory’s focus on individual incentive compatibility. 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 requires a small-economy approach.

This small-economy approach may be appropriate for studying how people in a condominium might decide on how much to spend on maintenance and gardening, it is inappropriate for studying how a society with millions of people decides on how much to spend on national defense or on the judicial system. A large-economy model, in which no one person individually is able to affect the level of public-good provision precisely captures the notion that, as individuals, people do not expect to have any influence over such collective decisions.

The notion that individual agents are too insignificant to have a noticeable influence on aggregate outcomes is as relevant for certain problems of public-good provision as it is for competitive markets or for elections. Limiting the theory of public-good provision to models in which each agent has a noticeable influence on aggregate outcomes is akin to limiting the analysis of markets to models of bargaining and oligopoly without ever talking about perfect competition.

What are the issues? In a large economy, the problem of finding an individually incentive-compatible mechanism for public-good provision is trivial. Because no one person individually is able to affect the level of public-good provision, no one person 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 neither have an effect on the public-good

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

²For implementation in dominant strategies, see Clarke (1971), Groves (1973); Green and Laffont (1979 a), 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.

provision level nor on the payments that the person has to make, she may as well report her preferences truthfully.

If preferences are reported truthfully, there is no problem about implementing an efficient provision rule for the public good. For example, one can ask people what the public good is worth to them. If the cross-section average of the reports exceeds the per-capita cost of provision, the public good is provided, and the cost is evenly shared in the population. If the cross-section average of the reports falls short of the per-capita cost, the public good is not provided, and no payments are made. Participation in this scheme may not be voluntary, but there is no problem of incentive compatibility.³ Truthtelling is actually a (weakly) dominant strategy.

There is also no problem with budget balance. For finite economies, it is well known that dominant-strategy or robust Bayes-Nash implementation of efficient public-good provision rules are incompatible with budget balance.⁴ The payments needed to provide appropriate incentives do not match the costs in all states of the world. In large economies, this is not a problem because individual incentive provision is trivial.

However, we see another problem. In our view, the theory of public-good provision should be concerned about groups of people as well as individuals. In collective decision making, we usually observe interest groups and political parties that contribute to the coordination of individuals with similar interests and similar views. In relations with their members, these groups may be hampered by information and incentive problems of their own, but in their relations with the overall system, they often are able to affect aggregate outcomes.

To take account of group concerns, we impose a condition of coalition proofness. Under this condition, there must not be any group of participants that is able to affect the aggregate outcome by coordinating its members' reports so as to make a significant set of its members better off without making any other member worse off. We are not trying to analyse real-world organizations, such as political parties or industry lobbies, that influence public decisions. Our purpose is, rather, to investigate in abstract terms, based on the values that different people attach to a public good, what potential there is for coalitions to form and what constraints such coalition formation imposes on public-good provision.

In the following section, we use an example to explain why a requirement of coalition proofness is called for and why it makes a difference. We show that first-best implementation may need to rely on information from people who are actually harmed by the use that is made of their information. Because none of these people expects to affect the aggregate outcome, individually, there is no conflict with individual incentive compatibility, but as a group, these people can gain if they coordinate on falsifying their reports. A requirement of coalition proofness takes such group interests into account. When this is done, an efficient public-good provision rule may not be implementable.

We also use the example to show that the failure of coalition proofness in the large economy has a counterpart in small economies, i.e. in economies with finitely many participants. For

³We 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.

⁴See Green and Laffont (1979 a).

small economies, we actually identify two types of failure of coalition proofness. One type of failure, which has been observed in the literature, is related to incentive payments for individuals who are pivotal and to the associated lack of budget balance; this failure of coalition proofness is only relevant in small economies.⁵ A second type of failure involves the ability of potentially large groups with common interests to manipulate information so as to change the overall outcome. This failure is relevant in large economies and in small economies.

Our analysis. In the body of the paper, we develop these insights formally. We use a Bayesian approach to implementation but require robustness of the social choice function. Neither the public-good provision rule, nor the participants' payments are allowed to depend on the specification of beliefs that people have about each other. We also impose a such robustness condition on the design of deviating coalitions.

We focus on social choice functions that are monotonic in the sense that, if the distribution of public-good valuations in the population is shifted "upwards" in the sense of first-order stochastic dominance, then the level of public-good provision is not lowered.⁶ We show that, if a monotonic social choice function is immune to robust coalition formation, as well as anonymous and robustly implementable, then it can be implemented by a voting mechanism, i.e., a mechanism under which the level of the public-good that is provided depends on the number of people preferring the chosen level over the different alternatives, without regard to the intensities of people's preferences. Conversely, any anonymous and robustly implementable social choice function that is immune to robust coalition formation can be implemented by a voting mechanism.

For a public good that comes as a single indivisible unit, it suffices to ask people who is for and who is against the provision of the public good and to condition the outcome on the shares of votes for the two alternatives. This conclusion remains true if the public good comes in multiple units, and per-unit costs are constant. With multiple units and increasing per-unit costs, more complicated procedures can be used, but it remains true that, if the social choice function is to be immune against robust blocking, as well as monotonic, then it cannot take account of preference intensities and a generalized voting procedure can be used to implement it. For an important class of social choice functions, we find that it suffices to ask, for each potential level of public good how many people are for and how many people are against an additional one-unit increase in the provision level and to proceed with the increase if the population share of the proponents exceeds a specified threshold.

Our results thus provide a link between mechanism design and political economy approaches to public-goods provision. In practice, how much we spend on public-goods such as highways, national defense, or the judicial system is determined in the political system. In democratic societies, public decision making depends on voting with a "one-person-one-vote-principle". In most countries voting is about the people who will become members of parliaments and governments, but in some countries also have popular votes on substantive questions. Switzerland,

⁵See Bennett and Conn (1977), Green and Laffont (1979 b).

⁶The focus on monotonic provision rules is not restrictive if we think of an economy with a continuum of individuals as an idealization of a large, but finite economy. In an economy with finitely many individuals, monotonicity is an implication of individual incentive compatibility.

for example, has public referenda on such issues, from the installation of a municipal swimming pool to the new transalpine tunnels and even on whether to abolish the army.

Economists have traditionally been critical of voting because voting does not take account of preference intensities and therefore can lead to inefficient outcomes.⁷ 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 might be sub-optimal because the benefits for the proponents might be large while 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 robustly incentive-compatible and coalition-proof as well as anonymous. Mechanisms that take account of preference intensities necessarily violate one of these conditions.⁸

Related work. The large-economy approach to mechanism design has been pioneered by Hammond (1979, 1987), Mas-Colell and Vives (1993) and Guesnerie (1995). We extend their framework so as to allow for aggregate uncertainty, i.e. for uncertainty about the cross-section distribution of preferences in the economy as a whole.⁹

Coalition-proofness in public-good provision has been studied by Bennett and Conn (1977), Green and Laffont (1979 a), and Crémer (1996), who showed that Groves mechanisms, which implement efficient allocations in dominant strategies, are typically not coalition-proof. As we explain in the next section, their arguments involve failures of coalition proofness that arise when some agents are pivotal. By contrast, we focus on failures of coalition proofness that arise when no individual agent is pivotal but a significant set of agents can benefit by coordinating reports so that the overall outcome is changed.

Whereas Bennett and Conn (1977), Green and Laffont (1979 a), and Crémer (1996) relied on a concept of coalition proofness *ex post*, i.e., after the state of the world has been revealed, we consider coalition proofness *ex interim*, i.e., at a time when each agent knows his or her individual valuation but is uncertain about the overall state of the economy. Like Laffont and Martimort (1997, 2000), we impose incentive constraints on coalition formation that take account of the fact that each agent's valuation is known to that agent only and not to whoever organizes the coalition. Whereas Laffont and Martimort focus on deviations by the grand coalition of all agents, we allow for deviations by coalitions of arbitrary size. If a coalition comprises a subset of agents, the coalition organizer can use an incentive mechanism to learn the coalition members' valuations for the public good but he cannot learn the valuations of people outside the coalition. To take account of this lack of information, we require failures of coalition proofness to be *robust*,

⁷Thus, Buchanan and Tullock (1962) argue that vote-trading would be desirable because it provides a way to overcome this problem. Similarly, Casella (2005) argues that intensities could be taken into account if voters had an endowment of votes and could assign more votes to issues that are of greater importance to them. Goeree and Zhang (2013) propose to replace votes by monetary bids.

⁸When we refer to voting mechanisms, we do not necessarily mean majority rule. A mechanism involving majority voting can be but need not be optimal. 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 votes in favor. Majority voting is likely to be desirable if there is no such prior information about potential biases in voting.

⁹The combination of aggregate uncertainty with private values at the individual level gives rise to some measure-theoretic issues, for which we refer the reader to Qiao et al. (forthcoming).

i.e. the collective deviation that is implemented by a coalition must make the coalition members better off no matter what the situation of the rest of the economy may be.

In Bierbrauer and Hellwig (2015), we study the implications of coalition proofness for public-good provision in a setting with finitely many participants and show that any robustly implementable and coalition-proof social choice function must be implementable by a voting mechanism. However, in that paper we allow coalitions to condition their behaviours on the beliefs that participants have about each other and we require coalition proofness for each specification of the agents' belief systems. Coalition organizers are thus assumed to know more than the overall mechanism designer. For belief systems that are degenerate in the sense that all participants "know" what the state of the economy is, the problem of what a coalition organizer knows about about people outside the coalition is therefore moot. In Bierbrauer and Hellwig (2015), the proof that every robust and coalition-proof social choice function is a voting mechanism makes essential use of these "complete information belief systems." This proof strategy is not available here because we do not allow collective deviations to be conditioned on belief systems, let alone the situation of people outside coalition. We now require collective deviations by coalitions to be robustly designed and to be robustly advantageous to coalition members. With this robustness requirement, we also show that in a large economy, there is no room for side payments between coalition members; in contrast, in Bierbrauer and Hellwig (2015), absence of side payments was imposed as an assumption.

Bierbrauer (2009a,b, 2012, 2014); Bierbrauer and Sahn (2010) explore the interaction of optimal taxation and public-goods provision in large economies. These papers involve environments in which individuals differ not only in their preferences for public goods, but also in their productive abilities, as in the theory of optimal income taxation in the tradition of Mirrlees (1971). The resulting two-dimensional heterogeneity among agents introduces issues of multi-dimensional mechanism design, which are the focus of those papers. Bierbrauer (2009b), Bierbrauer (2012) and Bierbrauer (2014) also invoke a requirement of coalition proofness, which arises naturally because, under the assumption that there are just two "classes" in terms of productive abilities and only two possible valuations for a public good, it is easy to see which coalitions might benefit from manipulating a mechanism for income taxation and public-goods provision. In contrast, this paper looks at public-good provision in isolation, and develops a concept of coalition proofness without taking recourse to any natural "class distinctions".

In the following, Section 2 uses an example to explain in more detail why individual incentive compatibility conditions ought to be supplemented by conditions of coalition proofness and robustness. We also use the example to relate our analysis of public-good provision in a large economy to the traditional analyses of dominant-strategy implementation or Bayes-Nash implementation in finite economies. Subsequently, Section 3 presents our formal model and introduces the requirement of robust implementability. The requirement of coalition proofness is formally introduced in Section 4. Section 5 gives our main result, i.e., the characterization of robust and coalition-proof public-goods provision in a large economy. Section 6 discusses the welfare implications of our characterization, and Section 7 extends the analysis to allow for multiple provision levels of the public good. The last section contains concluding remarks.

2 Why Coalition Proofness? Why Robustness?

In this section, we explain why public-goods analysis should not only involve a requirement of individual incentive compatibility but also a requirement of coalition proofness. We use a simple example to illustrate our concerns.

In the example, the public good comes as a single indivisible unit. The per-capita cost of providing this unit is equal to 4. The benefit an agent draws from the public good if provided is either 0, or 3, or 10. The shares of agents with valuations 0, 3, and 10 in the population are denoted as s_0 , s_3 , and s_{10} , respectively. An efficient provision rule stipulates that the public good should be provided if $3s_3 + 10s_{10} > 4$ and that it should not be provided if $3s_3 + 10s_{10} < 4$. To implement this decision rule, one needs to know the values of s_3 and s_{10} . If agents' public-good valuations are their own private information, the needed information about s_3 and s_{10} can only be obtained if agents can be made to communicate their valuations. This is where individual incentive compatibility comes in.

To impose some more structure, we assume that there is some number α between zero and .7 such that the different agents' valuations are the realizations of independent and identically distributed random variables with probabilities $.7 - \alpha$ for the valuation 0, α for the valuation 3, and $.3$ for the valuation 10. For a large economy with a continuum of agents, this assumption implies that the shares s_0, s_3, s_{10} of agents with valuations 0, 3, and 10 in the population are nonrandom and equal to $.7 - \alpha, \alpha$, and $.3$.¹⁰ The public good should then be provided if $\alpha > \frac{1}{3}$ and should not be provided if $\alpha < \frac{1}{3}$. The requisite resources can be obtained by a payment rule under which everybody pays 4 if the public good is provided and 0 if it is not provided. Because no one agent is ever pivotal for the provision of the public good, a mechanism implementing this rule for efficient public-good provision with equal cost sharing is incentive-compatible.

If α is common knowledge, this reasoning is unproblematic. By contrast, if α is the realization of a nondegenerate random variable $\tilde{\alpha}$, there is a problem because it is not *a priori* clear whether the public good should be provided or not. 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 infers that $\tilde{\alpha} > \frac{1}{3}$ and that the public good should be provided.

At this point, however, 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

¹⁰For a formal treatment of the law of large numbers in a model with a continuum of agents, see Sun (2006) and Qiao et al. (forthcoming).

mechanism, the answer to this question is that nobody minds reporting his or her valuation honestly because nobody feels that his or her report will make a difference to anything anyway. We find this answer unconvincing.

What precisely is amiss? Two objections come naturally. First, there is something arbitrary about the assumption that, if agents are indifferent between the different messages at their disposal, they resolve this indifference in favour of truthtelling. 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, yet in the probability-zero event where she might make a difference, reporting the valuation 0 would lead to a preferred outcome, with non-provision of the public good and no payment. Such considerations underlie the assumption of sincere voting in political economy. Under this assumption, people vote their preferences even though, as individuals, they do not expect their votes to have an effect on aggregate outcomes.¹¹ Conceptually, this involves what we may call a super-dominance criterion, i.e. a criterion of robustness to the elimination of strategies that are undominated in the usual sense, but dominated in fictitious probability-zero events. In our example, the efficient provision rule with equal cost sharing does not satisfy this criterion.

Second, in our example, a coalition of people who value the public good at either 0 or 3 could prevent the implementation of the efficient provision rule with equal cost sharing by coordinating their reports so that the fraction of people reporting 3 is always less than $\frac{1}{3}$. The public good would not be provided at all, and all coalition members would be better off. The reports that such a coalition would recommend to its members would all be individually incentive-compatible. Since no one person can affect the aggregate outcome, no coalition member has anything to gain by deviating from the manipulation. In contrast to a cartel that tries to eliminate competition among its members, such a coalition would not have an incentive compatibility problem of its own.

Which of these observations provides the “right” way for dealing with the issue? In addressing this question, it is important to keep in mind that the notion of a large economy with a continuum of agents is an idealization, which is not to be taken literally. No economy actually has a continuum of agents but the continuum model is useful because it puts the focus on certain features of strategic interaction that are essential in large finite economies as well as the continuum model, but are more difficult to analyse in the finite models where the analyst must keep track of exactly how many agents there are and how insignificant each agent is relative to the aggregate.¹²

Given this interpretation of the continuum economy as an idealization that captures essential features of large finite economies, any condition that is imposed in the continuum model should have a natural analogue in the large finite model. In the following, we therefore reconsider our example with finitely many agents. We will see that, in the finite setting, a super-dominance criterion has no bite, but coalition proofness does.

¹¹See, for example, Austen-Smith and Banks (1996).

¹²In this context, it is worth recalling that, in experimental studies of double oral auctions with, say, twenty participants, outcomes tend to be indistinguishable from the Walrasian, competitive equilibrium outcome, even though the market power of any one participant is, strictly speaking, not zero.

Dominant-Strategy Implementation and Coalition-Proofness in Finite Economies.

Consider an economy with n agents. As before, we assume that the public good comes as a single indivisible unit, with a per-capita provision cost equal to 4, so with n agents the costs are equal to $4n$. The agents' valuations for the public good are again 0, 3, or 10. Thus, if S_3 and S_{10} are the numbers of agents with valuations 3 and 10, a first-best provision rule requires that the public good be provided if $3S_3 + 10S_{10} > 4n$ and not be provided if $3S_3 + 10S_{10} < 4n$. For specificity, we assume that the public good is also provided if $3S_3 + 10S_{10} = 4n$.

With finitely many participants, strategic interdependence of the participants is important. In the literature on public-good provision, this is usually taken into account by considering implementation through dominant-strategy equilibria or implementation through Bayes-Nash equilibria. We first consider dominant-strategy implementation.

Dominant-strategy implementation of first-best provision rules is obtained by Groves mechanisms, which induce agents to take account of the externalities that their choices may impose on others.¹³ In our example, a Groves mechanism calls for public-good provision according to the first-best provision rule and specifies the payment of any agent i as a function of the agent's reported valuation \hat{v}_i and the numbers \hat{S}_3 and \hat{S}_{10} of other agents reporting 3 and 10, so that

$$p_i(\hat{v}_i, \hat{S}_3, \hat{S}_{10}) = h_i(\hat{S}_3, \hat{S}_{10}) + 4n - 3\hat{S}_3 - 10\hat{S}_{10} \quad \text{if } 4n \leq \hat{v}_i + 3\hat{S}_3 + 10\hat{S}_{10}, \quad (1)$$

and

$$p_i(\hat{v}_i, \hat{S}_3, \hat{S}_{10}) = h_i(\hat{S}_3, \hat{S}_{10}) \quad \text{if } 4n > \hat{v}_i + 3\hat{S}_3 + 10\hat{S}_{10}, \quad (2)$$

for some arbitrary function h_i . Given this form of payment rule and given the first-best provision rule, one easily verifies that truth-telling is a dominant strategy for each agent.

We further specialize the analysis by assuming that, for any i , the function h_i takes the form

$$h_i(\hat{S}_3, \hat{S}_{10}) = \max[0, 3\hat{S}_3 + 10\hat{S}_{10} - 4(n - 1)]. \quad (3)$$

We thus obtain the special case of a Clarke-Groves mechanism.¹⁴ From (1) - (3), one easily sees that this mechanism never runs a deficit. Table 1 below describes a Clarke-Groves mechanism for an arbitrary but finite number n of individuals. By way of normalization, we replace the aggregate valuation \bar{V} by the average valuation $\bar{v} = \frac{\bar{V}}{n}$. Thus, the public good is provided if $\bar{v} \geq 4$ and is not provided if $\bar{v} < 4$.

Table 1.

	$v_i = 0$	$v_i = 3$	$v_i = 10$
$\bar{v} \leq 4(1 - \frac{1}{n})$	$p_i = 0$	$p_i = 0$	$p_i = 0$
$4(1 - \frac{1}{n}) < \bar{v} < 4$	$p_i = n\bar{v} - 4(n - 1)$	$p_i = 0$	$p_i = 0$
$4 \leq \bar{v} < 4 + \frac{6}{n}$	$p_i = 4$	$p_i = 4$	$p_i = 4n - (\bar{v}n - 10)$
$\bar{v} \geq 4 + \frac{6}{n}$	$p_i = 4$	$p_i = 4$	$p_i = 4$

¹³See Groves (1973).

¹⁴See Clarke (1971).

The table exhibits four cases, no payments and no provision if \bar{v} is very low; payments by agents with valuation zero and no provision if \bar{v} is close to but less than 4; provision with equal cost sharing if \bar{v} is very high; finally, provision with extra payments by agents with valuation 10 if \bar{v} is close to but not less than 4.

Truth-telling is a dominant strategy, and moreover, whenever an individual is pivotal for public-goods provision, truth-telling is the unique best response. Thus, truth-telling is perfectly in line with a super-dominance criterion.

There is, however, a problem with coalition proofness. Coalition-proofness can fail in two ways. First, if $4(1 - \frac{1}{n}) < \bar{v} < 4$ and for $4 \leq \bar{v} < 4 + \frac{6}{n}$, the grand coalition of all agents might reduce payments by reporting that all have the valuation zero (if $4(1 - \frac{1}{n}) < \bar{v} < 4$) or that all have the valuation 10 (if $4 \leq \bar{v} < 4 + \frac{6}{n}$). Second, if $\bar{v} > 4 + \frac{6}{n}$ and $S_{10} < 4n$, a coalition of agents with valuation 3 would wish to forestall the provision of the public good by reporting that their valuation is actually 0. The first type of failure of coalition proofness has previously been noted by Bennett and Conn (1977) and Green and Laffont (1979 b). The second type of failure is an exact analogue of the failure of coalition proofness in the continuum model.

If the number of participants is large, the first type of failure of coalition proofness is not very important. To see this, recall the assumption that, for some $\alpha \in (0, .7)$, the different agents' valuations are independent and identically distributed with probabilities $.7 - \alpha$ for the valuation 0, α for the valuation 3, and $.3$ for the valuation 10. If n is large, then, by the law of large numbers, with a probability close to one, the shares of agents with valuations 0, 3, and 10 will be close to $.7 - \alpha$, α , and $.3$. If $\alpha > \frac{1}{3}$, it follows that, with a probability close to one, the per-capita aggregate valuation \bar{v} will be greater than $4 + \frac{6}{n}$, and, similarly, if $\alpha < \frac{1}{3}$, \bar{v} will be less than $4(1 - \frac{1}{n})$ with a probability close to one. In both cases, if $\alpha > \frac{1}{3}$ and if $\alpha < \frac{1}{3}$, the events $4(1 - \frac{1}{n}) < \bar{v} < 4$ and $4 \leq \bar{v} < 4 + \frac{6}{n}$ become unlikely if n becomes large. So does the coalition proofness problem (and the budget balance problem) associated with these events.

If we think about α itself as the realization of a random variable which has probability zero of taking the value $\frac{1}{3}$, it follows that for large n , the events $4(1 - \frac{1}{n}) < \bar{v} < 4$ and $4 \leq \bar{v} < 4 + \frac{6}{n}$ are unlikely. The continuum model in which no one agent is ever pivotal reflects this finding. With large n , the question whether the public good should be provided or not depends much more on whether α is greater or less than $\frac{1}{3}$ than on whether the deviations of the individual v_i s from their means are positive or negative. However, for *any* α , both with finitely many individuals and with a continuum of individuals, first-best implementation is incompatible with the condition that a coalition of agents with valuations of 3 should not find it more attractive to jointly report valuations of 0 instead.

Bayes-Nash Implementation, Coalition-Proofness, and Robustness. For Bayes-Nash implementation, as usual, it makes a big difference whether agents' valuations are treated as mutually independent or as correlated. In our example, we have independent private values if the parameter α is taken to be fixed and commonly known; we have correlated values if α is taken to be the realization of a nondegenerate random variable $\tilde{\alpha}$. In the latter case, each agent's valuation is correlated with $\tilde{\alpha}$, and therefore the different agents' valuations are correlated with each other.

The literature on public-good provision in Bayesian models with finitely many agents usually assumes independent, rather than correlated, private values.¹⁵ From the perspective of a large economy, this case is not very interesting because, if α is known, it is also clear whether it is efficient for the public good to be provided or not. In a large economy with independent private values, the implementation of an efficient rule for public-good provision does not require any information from participants. If the question whether the public good should be provided or not is to be nontrivial, α must be the realization of a nondegenerate random variable $\tilde{\alpha}$ that is not directly observed.¹⁶

If $\tilde{\alpha}$ is a nondegenerate random variable, each agent's beliefs about $\tilde{\alpha}$ will vary with his own valuation for the public good. A high value of $\tilde{\alpha}$ makes the valuation 0 less likely and a valuation of 3 more likely. Observation of the valuation 0 therefore induces a downward adjustment and of the valuation 3 an upward adjustment in the probability that an agent assigns to a high value of $\tilde{\alpha}$. Because the probability of the valuation 10 is independent of α , agents' beliefs after observing the valuation 10 are the same as their prior beliefs.

As was first pointed out by Crémer and McLean (1985,1988), differences in beliefs induce differences in preferences over lotteries whose outcomes depend on the value of $\tilde{\alpha}$, and these differences in preferences over lotteries can be exploited for mechanism design. Such valuation-and-outcome-dependent payment schemes can actually provide for a coalition-proof, Bayesian-incentive-compatible implementation of first-best provision rules.

As an example, consider the payment scheme in Table 2 where, as in Table 1, \bar{v} is the aggregate per-capita valuation of the public good. In the continuum model, of course, $\bar{v} = 3\alpha + 3$, where α is the realization of $\tilde{\alpha}$. As before, assume that the public good is provided if and only if $\bar{v} \geq 4$.

Table 2.

	$v_i = 0$	$v_i = 3$	$v_i = 10$
$\bar{v} < 4$	$p_i = -2.1$	$p_i = 8.4$	$p_i = -2.1$
$4 \leq \bar{v}$	$p_i = 10$	$p_i = 0$	$p_i = 10$

For $v_i \in \{0, 3, 10\}$, let $\beta_L(v_i)$ and $\beta_H(v_i) = 1 - \beta_L(v_i)$ be the probabilities that an agent with public-good valuation v_i assigns to the events $\{\tilde{\alpha} < \frac{1}{3}\}$ and $\{\tilde{\alpha} \geq \frac{1}{3}\}$ or, equivalently, the events $\{\bar{v} < 4\}$ and $\{\bar{v} \geq 4\}$. Given that individuals cannot influence the level of public-good provision, one easily verifies that the payment scheme in Table 2 is strictly Bayesian incentive-compatible if

$$\beta_L(0) = \frac{5}{6}, \beta_H(0) = \frac{1}{6}, \beta_L(3) = \frac{1}{4}, \beta_H(3) = \frac{3}{4}, \beta_L(10) = \frac{1}{2}, \beta_H(10) = \frac{1}{2}. \quad (4)$$

These values of $\beta_L(v_i)$ and $\beta_H(v_i)$ for $v_i \in \{0, 3, 10\}$ can actually be derived from a common prior that assigns a probability of one half each to the two possible values .2 and .6 of the random variable $\tilde{\alpha}$. For this specification of $\tilde{\alpha}$, the given mechanism is also feasible in the large economy:

¹⁵The seminal paper is d'Aspremont and Gérard-Varet (1979). Subsequent papers include Güth and Hellwig (1986), Mailath and Postlewaite (1990), Hellwig (2003) and Norman (2004).

¹⁶For a general discussion of this point, see Bierbrauer and Hellwig (in preparation).

If $\tilde{\alpha} = .2$, the aggregate per-capita payment is equal to $.2 \cdot 8.4 - .8 \cdot 2.1 = 0$; if $\tilde{\alpha} = .6$, the aggregate per-capita payment is equal to $.6 \cdot 0 + .4 \cdot 10 = 4$, which is just the per-capita cost of providing the public good, as stipulated for this event.¹⁷

The given incentive mechanism is also coalition-proof. Under the payment scheme in Table 3, as opposed to equal cost sharing, 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-good provision. Without their cooperation, however, a coalition that would reduce the incidence of public-good provision cannot form. By a similar 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.

Straightforward continuity considerations imply that coalition-proof Bayesian-Nash implementation of first-best provision rules is also obtained for the n -agent version of the example, provided that n is large. To see this, observe that the type-dependent posterior probabilities (4) for the events $\tilde{\alpha} = .2$ and $\tilde{\alpha} = .6$ have nothing to do with the number of agents in the economy. Observe also that, if n is large, then, by the law of large numbers, the type-dependent posterior probabilities of the events $\{\bar{v} < 4\}$ and $\{\bar{v} \geq 4\}$ are close to the type-dependent posterior probabilities of the events $\tilde{\alpha} = .2$ and $\tilde{\alpha} = .6$ in (4). For the posterior probabilities given by (4), the conditions for Bayesian incentive-compatibility hold with strict inequality in the large economy. Therefore, the mechanism given by a first-best provision rule and the payment scheme in Table 3 is also Bayesian-incentive-compatibility in an n -agent version of the example with very large n , and so is a mechanism that results from taking a small perturbation of the payment scheme in Table 3 as may be required to ensure that expected payments are equal to expected costs.

With a continuum as well as a finite number of participants, Crémer-McLean-type mechanisms can be used for coalition-proof, Bayesian incentive-compatible implementation of first-best public-good provision rules. However, these mechanisms are not robust. The payment scheme given in Table 3 works for a prior that assigns probabilities $\frac{1}{2}$ each to the outcomes .2 and .6 but does not work for some other priors. For example, if the prior assigns probability $\frac{1}{3}$ to the outcome .2 and $\frac{2}{3}$ to the outcome .6, a person with the valuation $v_i = 10$ has posterior probabilities $\beta_L(10) = \frac{1}{3}, \beta_H(10) = \frac{2}{3}$; with these probabilities, this person expects to pay $\frac{1}{3} \cdot 8.4 = 2.8$ if he or she claims to have the valuation 3 and to pay $\frac{1}{3} \cdot (-2.1) + \frac{2}{3} \cdot 10 = 5.9$ if he or she is honest. Bayesian incentive compatibility is violated.

Like Ledyard (1978) and Bergemann and Morris (2005), we consider it unreasonable to suppose that a mechanism designer has the information about the type dependence of participants' beliefs that he would need to implement the appropriate Crémer-McLean payment scheme. We will therefore impose a requirement of robustness, i.e., the mechanism that is used to determine the level of public-good provision and the different agents' payments 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.¹⁸ This require-

¹⁷The given mechanism is also individually rational: Interim expected payoffs are $\frac{5}{6} \cdot 2.1 - \frac{1}{6} \cdot 10 = \frac{5}{6}$ for agents with $v_i = 0$, $-\frac{1}{4} \cdot 8.4 + \frac{3}{4} \cdot 3 = \frac{6}{4}$ for agents with $v_i = 3$, and $\frac{1}{2} \cdot 2.1$ for agents with $v_i = 10$.

¹⁸See Bergemann and Morris (2005). Börgers and Smith (2014) suggest that robustness may be too strong a

ment eliminates the possibility of using Crémer-McLean-type mechanisms to achieve coalition proofness as well as individual incentive compatibility.

For models with private information, it is well known that robust Bayes-Nash and dominant-strategy implementability requirements are equivalent in the sense that they impose the same restrictions on the design of incentive mechanisms. Coalition-proofness therefore restricts the scope for robust Bayes-Nash implementation just as it restricts the scope for dominant-strategy implementation. Our previous discussion carries right over.

In the following, we study the implications of coalition proofness for robust Bayes-Nash implementation. By comparison to the dominant-strategy approach, the Bayes-Nash approach has the advantage that it is explicit about the decision problems participants face, including a full specification of their beliefs about other agents' types and their expectations about other agents' behaviors. It also allows for a richer discussion of coalition proofness. As far as we know, the notion of robustly blocking coalitions that we propose has no counterpart in a dominant-strategy framework.¹⁹

3 Robust Implementation in a Large Economy

Payoffs and Social Choice Functions. We consider an economy with a continuum of agents of measure 1. There is one private good and one public good. The public good comes as a single indivisible unit.²⁰ 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 set V of possible valuations, which is the same for all i . We assume that V is a compact interval, $V = [v_{\min}, v_{\max}] \subset \mathbb{R}_+$, and that k is in the interior of V .

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 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

condition. They give examples in which outcomes depend on the participants' beliefs about each other, but this dependence is due to people's behaviors rather than the incentive mechanism. Unfortunately, we do not as yet have a simple characterization of the requirements that their approach would impose on mechanism design.

¹⁹As developed by Bennett and Conn (1977), Green and Laffont (1979 b), and Moulin (1999), the requirement of coalition proofness in a dominant-strategy setting requires that, regardless of what the reports of the other agents may be, a coalition of agents cannot benefit from a coordinated false communication of types. In the terminology of Bergemann and Morris (2005), this is a condition of *ex post* coalition proofness. This condition presumes that coalition behaviour can be conditioned on the valuations of all participants, including the non-members of the coalition.

²⁰In Section 7 we show that our results extend to a setting with a compact set of possible public-good provision levels and a linear cost function.

probability measures on 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, 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.²¹

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

$$\int_V P_F(s, v) ds(v) \geq kQ_F(s) \tag{5}$$

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

A social choice function is said to be monotonic if $Q_F(s) \geq Q_F(s')$ whenever s dominates s' in the sense of first-order stochastic dominance, i.e. whenever $\int g(v) ds(v) \geq \int g(v) ds'(v)$ for every nondecreasing function g . Monotonicity reflects the notion that, if public-good valuations go up, the level of public-good provision should not go down.

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 distributions 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 $X \subset \mathcal{M}(T)$, $\beta(X | t_i)$ 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 $\beta : T \rightarrow \mathcal{M}(\mathcal{M}(T))$ as the *belief system* of the economy.²²

²¹Anonymity is a substantive constraint. Using the idea of sampling, that has been developed by Green and Laffont (1979 a), Bierbrauer and Sahn (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.

²²We do not assume that the belief system is compatible with a common prior. As shown in Bierbrauer and

A typical element of $\mathcal{M}(T)$ will be denoted by δ . The cross-section distribution δ of abstract types induces the cross-section distribution $s(\delta) = \delta \circ \tau^{-1}$ of payoff types, or public-good valuations. For any subset V' of V we write $s(V' | \delta)$ for the mass of individuals that the distribution $s(\delta)$ assigns to payoff types in V' . A generic element of $\mathcal{M}(V)$ will be denoted by s .

We will mostly consider payoff type distributions that do not have mass points and beliefs that assign probability zero to payoff type distributions with mass points. We say that a belief system $\beta : T \rightarrow (\mathcal{M}(T))$ is *admissible* if and only if for all $t \in T$, the induced belief function $\beta(s^{-1}(\mathcal{M}^{na}(V)) | t) = 1$, where $\mathcal{M}^{na}(V)$ is the set of measures on V that do not have atoms. This restriction will allow us to neglect agents who are indifferent between different alternatives because the set of such agents has measure zero.

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 $[(V, \mathcal{V}), \beta_s]$, where V is the set of possible valuations and \mathcal{V} is the Borel σ -algebra on V . 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.

Robust Incentive Compatibility. We focus on social choice functions that can be implemented as the truth-telling equilibrium of a direct mechanism. Such social choice functions will be called incentive-compatible. Formally, a social choice function $F = (Q_F, P_F)$ is said to be incentive-compatible on a given type space $[(T, \mathcal{T}), \tau, \beta]$, if, for all $t, t' \in T$

$$U(t | t) \geq U(t | t'), \tag{6}$$

where

$$U(t | t') := \int_{\mathcal{M}(T)} \{\tau(t)Q_F(s(\delta)) - P_F(\tau(t'), s(\delta))\} d\beta(\delta | t)$$

is the interim expected utility of an individual with type t that reports type t' under a direct mechanism for the given social choice function F .

An anonymous social choice function F is said to be *robustly incentive-compatible* or *robustly implementable* if, for every (T, \mathcal{T}) , and $\tau : T \rightarrow V$, the inequalities in (6) hold for every admissible belief system β .

Equivalence of Robust and Ex post Incentive Compatibility. The following Proposition asserts that robust incentive compatibility is equivalent to *ex post incentive compatibility*: once s has become known, no individual regrets having revealed his type to the mechanism. Ex post incentive compatibility of course is in turn equivalent to dominant-strategy incentive compatibility. The Proposition 1 adapts an argument of Bergemann and Morris (2005) to the given setup.²³

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).

²³A proof can be found in Bierbrauer and Hellwig (2010).

Proposition 1 *An anonymous social choice function $F = (Q_F, P_F)$ is robustly incentive-compatible if and only if it is ex post incentive compatible in the sense that, for all v and v' in V and all $s \in \mathcal{M}(V)$,*

$$vQ_F(s) - P_F(v, s) \geq vQ_F(s) - P_F(v', s). \quad (7)$$

By inspection of (7), in our setting, *ex post* incentive compatibility 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* incentive compatibility. This observation yields the following corollary to Proposition 1.

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

Given Corollary 1, we will represent a robustly incentive-compatible social choice function as a pair (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 $s \in \mathcal{M}(V)$.

Robust Incentive Compatibility of a First-best Provision Rule with Equal Cost Sharing. An anonymous social choice function $F = (Q_F, P_F)$ yields 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 (5). By standard arguments, this requires that the public good should be provided if the aggregate valuation $\bar{v}(s) := \int_V v ds(v)$ exceeds the cost k and should not be provided if $\bar{v}(s)$ 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:

Theorem 1 *A first-best anonymous social choice function $F = (Q_F, P_F)$ is robustly incentive-compatible if and only if, for all $s \in \mathcal{M}(V)$,*

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

Theorem 1 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 budget balance, as well as robust incentive compatibility.

This conclusion stands in marked contrast to the findings of the literature for economies with finitely many participants. For example, Green and Laffont (1979 a) showed that dominant-strategy implementation of first-best social choice rules is not compatible with budget balance. The reason is apparent from Table 1: Incentive provision to agents who may be pivotal requires payments that have nothing to do with the real costs that must be covered, whether the public good is provided or not.

In the large economy these concerns are moot because no agent is ever pivotal. Thus, in the absence of participation constraints, Theorem 1 suggests that in large economies robust incentive compatibility of first-best public-good provision rules does not involve any fundamental difficulties.

Nevertheless, we do not regard Theorem 1 as a satisfactory basis for the normative theory of public-good provision in a large economy. As we discussed in the introductory sections we consider the requirements of robust incentive compatibility 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.

4 Coalition Proofness

To implement a first-best outcome, one must know the aggregate public-good valuation $\bar{v}(s)$. This information is derived from people's reports about their individual valuations. With the social choice function in Theorem 1, 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, truth-telling is an interim Bayes-Nash equilibrium.

However, people with similar valuations have similar interests. Collectively, they might upset the truth-telling equilibrium, for example by manipulating the social outcome by coordinating on exaggerated reports. If all agents with payoff types $\tau(t) > k$ were to report the type t' so that $\tau(t') = v_{\max}$, they would raise the mechanism's assessment of the aggregate public-good valuation from its true level $\bar{v}(s)$ to the level $\bar{v}(s) + \int_{\{\tau(t) > k\}} [v_{\max} - \tau(t)] d\delta(t)$. This collective exaggeration of enthusiasm for the public good might cause the public good to be provided even though the aggregate valuation $\bar{v}(s)$ lies below the per-capita cost k . From the perspective of types with $\tau(t) > k$, this outcome would dominate the truth-telling equilibrium. It is also worth observing that, because people are indifferent as to what they report, this collective deviation by agents with payoff types above k , in combination with truth-telling by everybody else, is also an interim Bayes-Nash equilibrium of the game induced by the overall mechanism.

To eliminate the possibility of such a manipulation, we impose a requirement of coalition proofness in addition to anonymity and robust incentive compatibility.

Modelling coalition proofness. There are many ways to formalize a notion of coalition proofness.²⁴ One can study an extensive form game of coalition formation, or one can use a normal-form specification. One can assume that coalitions are able to use side-payments between coalition members, or one can disallow side payments. One can allow for indirect mechanisms, or focus on the truth-telling equilibria of direct mechanisms.²⁵ Finally, one may or may not impose additional requirements such as the condition that any use of a participant’s private information by a coalition must respect incentive-compatibility constraints or the condition that any deviating coalition must itself be robust to the formation of subcoalitions, which in turn might have to be robust to the formation of subsubcoalitions, and so on, see Bernheim et al. (1986).

In a previous version of this paper, we modelled coalition formation by an extensive-form game as in Laffont and Martimort (1997, 2000).²⁶ The extensive-form approach has the advantage of showing precisely how a deviating coalition gets the information that it uses to block the implementation of the social choice function. However, it is cumbersome to work with, so we use a short-cut that yields essentially the same results in a much simpler manner.

The shortcut that we use was introduced by Bennett and Conn (1977) and Green and Laffont (1979 b) and has been used, among others, by Moulin (1999) and Mehta et al. (2009). Given our focus on the implementation of a social choice function as the truth-telling equilibrium of a direct mechanism,²⁷ we represent coalition proofness by the requirement that there must not exist a group of agents who would all benefit from a false communication of preferences. Whereas Bennett and Conn (1977) and Green and Laffont (1979 b) applied the concept *ex post*, we apply it *ex interim*, with agents evaluating coalition outcomes in terms of their type-dependent beliefs. Moreover, we impose a robustness requirement on the collective deviations that block the implementation of a social choice function.

Apart from this robustness requirement, our notion of coalition proofness is very rudimentary. No account is given of extensive-form considerations, individual incentive compatibility in coalitions, or the potential susceptibility of coalitions to further deviations by sub-coalitions. In principle, it would be desirable to take all these restrictions into account. However, in a continuum economy and with a requirement of robustness in coalition design, some of these

²⁴Important references include Green and Laffont (1979 b), Bernheim et al. (1986), Crémer (1996), Laffont and Martimort (1997, 2000), Moulin (1999), Che and Kim (2006).

²⁵With a requirement of coalition proofness, the validity of the revelation principle cannot be taken for granted, see Boylan (1998) or Bierbrauer (2014).

²⁶See Bierbrauer and Hellwig (2010). In that game, an overall mechanism is announced at stage 0. At stage 1, a coalition organizer may propose a manipulative side-mechanism. At stage 2, individuals would decide whether or not to participate in this coalition. At stage 3, agents who have chosen to participate would send messages to the coalition organizer. At stage 4, the coalition organizer would use the information provided by these messages in order to choose recommendations to coalition members about the messages ("lies") they should send to the overall mechanism. Finally, at stage 5, individuals would send messages to the overall mechanism. On the basis of these messages, the overall mechanism would determine the level of public-good provision and the different agents’ payments. Sequential equilibrium conditions ensure that individual incentive constraints are satisfied and that no one relies on information that has not previously been provided in the course of play in the game.

²⁷This approach simplifies the exposition considerably. However, it raises the question whether the set of implementable social choice functions can be enlarged by allowing for non-direct mechanisms. In a previous discussion paper we have shown that the answer is “no”, see Bierbrauer and Hellwig (2011).

concerns are moot. As mentioned, extensive-form considerations do not make a difference at all. Nor do concerns for individual incentive-compatibility in coalitions.²⁸ Concerns about further deviations by sub-coalitions are also moot because the coalitions that matter for our analysis are all homogeneous and there is no point in forming sub-coalitions.²⁹ We will discuss some of these concerns as we go along.

Robustly Blocking Coalitions. We continue to impose anonymity and robust incentive compatibility of the social function. Thus, by Corollary 1, payments as well as public-good provision levels are independent of individual announcements. We think of a coalition as being run by a coalition organizer, who proposes the collective deviation, collects reports from coalition members and then chooses a profile of reports that coalition members are to make to the overall mechanism. Because individuals expect their announcements to have no effects on outcomes, any announcement is individually a best response. Individual incentive compatibility of the reports suggested by the coalition organizer is thus automatically given.

Consider an abstract type space (T, \mathcal{T}) , a payoff type function $\tau : T \rightarrow V$ and a robustly implementable social choice function $F = (Q_F, \bar{P}_F)$. For the moment, we neglect the possibility of side payments in a coalition. A collective deviation for a set $T' \subset T$ is specified as a strategy (a “lie”) $\ell_{T'} : \hat{\mathcal{M}}(T') \rightarrow \mathcal{M}(T)$, where $\hat{\mathcal{M}}(T')$ is the set of measures on T' that have total measure less than or equal to one. We think of $\ell_{T'}(\delta_{T'})$ as a probability distribution over reports or “lies” for individuals with types in T' . One may think of $\ell_{T'}(\delta_{T'})$ as a lottery determining the report that any one individual with a type in T' sends to the overall mechanism when the distribution of types in T' is $\delta_{T'}$. For any measurable set $\tilde{T} \subset T$, $\ell_{T'}(\tilde{T} \mid \delta_{T'})$ is then the probability that the report will be in the set \tilde{T} .³⁰ In this interpretation, a law of large numbers would imply that $\ell_{T'}(\tilde{T} \mid \delta_{T'})$ is also the share of reports from coalition members that lie in \tilde{T} . The lies are a function of $\delta_{T'}$, as opposed to $\delta = (\delta_{T'}, \delta_{T \setminus T'}) \in \hat{\mathcal{M}}(T') \times \hat{\mathcal{M}}(T \setminus T')$, because the coalition organizer observes only the types in the set of participating individuals and remains ignorant about the distribution of types in the complimentary set of agents, if he attracts agents with types in T' , he knows $\delta_{T'}$ but not $\delta_{T \setminus T'}$.

We write $\hat{\delta}(\ell_{T'}, \delta)$ for the overall cross-section distribution of reports that is generated by $\ell_{T'}$ if the true cross-section distribution of types is δ . Thus we have

$$\hat{\delta}(\ell_{T'}, \delta) = \delta(T') \cdot \ell_{T'}(\delta_{T'}) + \delta_{T \setminus T'}, \quad (8)$$

where $\delta_{T \setminus T'} \in \hat{\mathcal{M}}(T \setminus T')$ is the (non-normalized) cross-section distribution of types in $T \setminus T'$. The implied payoff type distribution is $s(\hat{\delta}(\ell_{T'}, \delta))$. If (T, \mathcal{T}) is actually the naive type space (V, \mathcal{V}) ,

²⁸As discussed in Bierbrauer and Hellwig (2015), with finitely many participants, individual incentive compatibility concerns do impose restrictions on coalition formation, but for robust coalition proofness, these restrictions do not affect the analysis.

²⁹In Bierbrauer and Hellwig (2015), we do impose the requirement that deviating coalitions have to be subcoalition-proof, subcoalitions have to be subsubcoalition-proof, and so on. With a requirement of robust coalition proofness, in Bierbrauer and Hellwig (2015), this requirement plays a role in proving the analogue of Theorem 3 below because it eliminates coalitions involving agents with heterogeneous interests.

³⁰The report $\ell_{T'}(\cdot)$ of any one member of the coalition might also be made to depend on the person’s type. This would not make a difference, however, because individual incentive compatibility holds anyway and only the aggregate “lie” matters.

we write $\hat{s}(\ell_{V'}, s)$ for the cross-section distribution of reported (payoff) types that is implied by the collective deviation $(V', \ell_{V'})$ when the true cross-section distribution is s .

Given a type space $[(T, \mathcal{T}), \tau, \beta]$, we say that a collective deviation $(T', \ell_{T'})$ *blocks* the anonymous and robustly implementable social choice function F on $[(T, \mathcal{T}), \tau, \beta]$ if

$$\begin{aligned} & \int_{\mathcal{M}(T)} \{\tau(t)Q_F(s(\hat{\delta}(\ell_{T'}, \delta))) - \bar{P}_F(s(\hat{\delta}(\ell_{T'}, \delta)))\} d\beta(\delta | t) \\ & \geq \int_{\mathcal{M}(T)} \{\tau(t)Q_F(s(\delta)) - \bar{P}_F(s(\delta))\} d\beta(\delta | t) \end{aligned} \quad (9)$$

for all $t' \in T'$, with a strict inequality for a non-negligible set of $t' \in T'$. Given an abstract type space (T, \mathcal{T}) and a payoff type function $\tau : T \rightarrow V$, we say that a collective deviation $(T', \ell_{T'})$ blocks the anonymous and robustly implementable social choice function F *robustly* on $[(T, \mathcal{T}), \tau]$ if it blocks F on $[(T, \mathcal{T}), \tau, \beta]$ for every admissible belief system β such that for $t \in T'$, the belief $\beta(t)$ assigns positive probability to the event that the deviation $(T', \ell_{T'})$ departs from truth-telling. The social choice function F is *immune against robust blocking*, if there is no abstract type space $[(T, \mathcal{T}), \tau]$ and no collective deviation $(T', \ell_{T'})$ that blocks F robustly on $[(T, \mathcal{T}), \tau]$.

Immunity against robust blocking is weaker than the condition of robust coalition proofness that we used in Bierbrauer and Hellwig (2015). A social choice function F is robustly coalition proof if there is no type space $[(T, \mathcal{T}), \tau, \beta]$ and no collective deviation $(T', \ell_{T'})$ that blocks F on $[(T, \mathcal{T}), \tau, \beta]$. This condition allows the collective deviation $(T', \ell_{T'})$ to be adapted to the particular type space on which it is used. Coalition organizers are thus presumed to have more information than the overall mechanism designer who wants to implement the social choice function. Immunity against robust blocking involves no such presumption.³¹

Irrelevance of Side Payments in Robust Coalition Design. What would happen if we allowed for side payments between coalition members? In principle, side payments enlarge the scope for coalition formation, so one might expect the set of coalition-proof social choice functions to become smaller if side payments are allowed.

If coalitions can condition their behaviors on belief systems, allowing for side payments may imply that no social choice function is robustly coalition-proof. The reason is that side payments introduce a possibility for agents to trade on differences in their beliefs. With quasi-linear utility functions, the gains from such trades can be arbitrarily large and can outweigh any concerns about public-good provision. However, such gambles depend on the specification of beliefs. They

³¹The requirement of robust coalition proofness also makes it possible to fudge the issue of incomplete information about the distribution of types among non-members of a coalition. Robust coalition proofness implies, in particular, that there are no blocking coalitions under complete information. Thus, if F is robustly coalition-proof, there is no type space $[(T, \mathcal{T}), \tau, \beta_\delta]$ with a belief system β_δ under which all agents “know” the state of the economy to be $\delta \in \mathcal{M}^{na}(T)$ such that some collective deviation $(T', \ell_{T'})$ blocks F on $[(T, \mathcal{T}), \tau, \beta_\delta]$. For every $\delta \in \mathcal{M}^{na}(T)$ and every collective deviation $(T', \ell_{T'})$, there exists a non-negligible set of types $t \in T'$ such that

$$\tau(t)Q_F(s(\hat{\delta}(\ell_{T'}, \delta))) - \bar{P}_F(s(\hat{\delta}(\ell_{T'}, \delta))) < \tau(t)Q_F(s(\delta)) - \bar{P}_F(s(\delta)),$$

which is a condition of coalition proofness *ex post*, when δ is common knowledge among the participants. In Bierbrauer and Hellwig (2015), we use these constraints to show that every robust and coalition-proof social choice function is a voting mechanism. Here, this proof strategy is not available because collective deviations have to be attractive to the participants for every admissible belief systems.

disappear if we require the collective deviations themselves to be robust. We briefly show that there is no role for side payments in robust coalition design.

To see this, consider an extended notion $(T', \ell_{T'}, z_{T'})$ of a collective deviation, where T' is the set of deviating types, $\ell_{T'}$ is the coalition's reporting strategy, and $z_{T'}$ is a function that indicates the side payments to coalition members. Reports and side payments depend on the set T' and the cross-section distribution $\delta_{T'}$ or types in T' but not on the belief system β or on the distribution of types in $T \setminus T'$. Given a belief system β , an individual of type $t' \in T'$ who participates in this coalition receives the expected side payment $\int_{\mathcal{M}(T')} z_{T'}(t', \delta_{T'}) d\beta(\delta_{T'} | t')$ and achieves the expected utility

$$\int_{\mathcal{M}(T)} \{\tau(t)Q_F(\hat{s}(\ell_{T'}, \delta)) - \bar{P}_F(\hat{s}(\ell_{T'}, \delta))\} d\beta(\delta | t) + \int_{\mathcal{M}(T')} z_{T'}(t', \delta_{T'}) d\beta(\delta_{T'} | t').$$

Individuals realize that their behavior has no effect on public-good provision but might affect the side payments they receive. Side payment functions must therefore satisfy the following incentive, participation and feasibility constraints:

Incentive compatibility. Because individuals are free to misreport their types, the side payment function must satisfy,

$$\int_{\mathcal{M}(T')} z_{T'}(t', \delta_{T'}) d\beta(\delta_{T'} | t') \geq \int_{\mathcal{M}(T')} z_{T'}(\hat{t}', \delta_{T'}) d\beta(\delta_{T'} | t'), \quad (10)$$

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

Participation constraint. Because individuals are free not to join the coalition, their expected receipts must be nonnegative, i.e., it must be the case that

$$\int_{\mathcal{M}(T')} z_{T'}(t', \delta_{T'}) d\beta(\delta_{T'} | t') \geq 0, \quad (11)$$

for all $t' \in T'$.

Feasibility Constraint. The coalition organizers must not expect to lose money.³²

We require incentive and participation constraints to hold robustly, i.e. for every admissible belief system.³³ Robust incentive compatibility again implies that the side payment an agent receives must be independent of the agent's announcement. Thus there is a function $\bar{z}_{T'} : \mathcal{M}(T') \rightarrow \mathbb{R}$ so that, for all $t' \in T'$, $z_{T'}(t', \delta_{T'}) = \bar{z}_{T'}(\delta_{T'})$, and the participation constraint implies that

$$\int_{\mathcal{M}(T')} \bar{z}_{T'}(\delta_{T'}) d\beta(\delta_{T'} | t') \geq 0.$$

³²Our formulation of this constraint is deliberately vague because we have not said anything about the coalition organizer's beliefs. One way to make the feasibility constraint precise is to impose it ex post, i.e. to postulate it separately for every $\delta_{T'}$. A weaker version would allow for averaging on the basis of some specific beliefs that the coalition organizer might have. As we explain below, if the incentive and participation constraints are robustly satisfied, this modelling choice is inconsequential for the conclusion that there is no role for side payments.

³³The arguments of Crémer and McLean (1988) imply that, for any *fixed*, belief function β , generically, there is a nontrivial side payment function that satisfies incentive and participation constraints and in addition, for any δ , the inequality

$$\int_{T'} z_{T'}(t', \delta_{T'}) d\delta(t') \leq 0,$$

so that the coalition organizer does not have to contribute any money of his own.

For this inequality to hold regardless of the belief system β , it must be the case that $\bar{z}_{T'}(\delta_{T'}) \geq 0$ and hence

$$\int_{T'} z_{T'}(t', \delta_{T'}) d\delta(t') = \bar{z}_{T'}(\delta_{T'}) \geq 0,$$

for all $\delta_{T'}$. At any $\delta_{T'}$ with $\bar{z}_{T'}(\delta_{T'}) > 0$, the coalition organizer then makes a loss. Thus, either $\bar{z}_{T'}(\delta_{T'}) = 0$ for all $\delta_{T'}$, or the payment scheme $z_{T'}$ violates the feasibility constraint.

Thus, contrary to what one might have expected, in our large-economy model with robust incentive, participation, and feasibility constraints, side payments do not actually enlarge the scope for coalition formation.

5 Monotonic Social Choice Functions and Voting Mechanisms

In this section, we show that social choice function that is robustly implementable and immune against robust blocking can be implemented by a voting mechanism.

Voting mechanisms. A simple 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, or abstain. The message set is therefore the triple $\{\text{alternative 0, alternative 1, } \textit{abstain}\}$.
- Alternative 1 stipulates that the public good should be provided and that each participant should make a payment $P_{\Phi}^1 \geq k$. Alternative 0 stipulates that the public good should not be provided and that each participant should make a payment $P_{\Phi}^0 \geq 0$.
- The outcome that is implemented depends monotonically on the shares of people voting for the different alternatives. For example, there might be a threshold $m_{\Phi}^1 \in [0, 1]$ such that alternative 1 is implemented if and only if the share of people voting for it exceeds m_{Φ}^1 .

The classical efficiency conditions for public goods due to ? and ? emphasize preference intensities, i.e., the average of the individuals' willingness to pay for public goods is the major determinant of the desirability of public-goods provision. From this perspective, the use of voting mechanisms cannot be justified. The neglect of preference intensities in voting appears as a major source of distortions.

In the following, we will however show that such a neglect of preference intensities is actually implied by monotonicity, robust implementability and immunity to robust blocking. If a monotonic, anonymous and robustly implementable social choice function were to condition the provision of the public good on preference intensities, it would be robustly blocked by some coalition. With robust incentive-compatibility 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.

Outcome Dependent Payment Schemes. We begin by showing that, under robust implementability and immunity to robust blocking, payments are a function of the public-goods provision level: If two states s and s' are such that $Q_F(s) = Q_F(s')$, then it must also be true that $\bar{P}_F(s) = \bar{P}_F(s')$. The logic is very simple: If 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 of all participants could use a collective deviation to induce the outcome with low payments when the actual state would call for the outcome with high payments.

Proposition 2 *If an anonymous and robustly implementable social choice function F is immune to robust blocking, then there exist numbers P_F^0 and P_F^1 so that, for all $v \in V$ and all $s \in \mathcal{M}^{na}(V)$,*

$$P_F(v, s) = \begin{cases} P_F^0, & \text{if } Q_F(s) = 0, \\ P_F^1, & \text{if } Q_F(s) = 1. \end{cases} \quad (12)$$

Proof. Let $F = (Q_F, \bar{P}_F)$ be an anonymous and robustly implementable social choice function. For $q \in \{0, 1\}$, let \hat{P}_F^q be the infimum of $\bar{P}_F(s)$ over the set $\{s \mid Q_F(s) = q\}$. Given an abstract type space (T, \mathcal{T}) , and a payoff type function $\tau : T \rightarrow V$, for any $\varepsilon > 0$, let δ_ε^q be such that

$$Q_F(s(\delta_\varepsilon^q)) = q \quad \text{and} \quad \bar{P}_F(s(\delta_\varepsilon^q)) \in [\hat{P}_F^q, \hat{P}_F^q + \varepsilon).$$

Consider a collective deviation ℓ_T such that, for all $q \in \{0, 1\}$,

$$\begin{aligned} \ell_T(\delta) &= \delta_\varepsilon^q \quad \text{if } Q_F(s(\delta)) = q \quad \text{and} \quad \bar{P}_F(s(\delta)) > \hat{P}_F^q + \varepsilon, \quad \text{and} \\ \ell_T(\delta) &= \delta \quad \text{otherwise.} \end{aligned}$$

One easily verifies that the collective deviation ℓ_T blocks F on all type spaces $[(T, \mathcal{T}), \tau, \beta]$ with belief systems that assign positive probability to events $\{Q_F(s(\delta)) = q, \bar{P}_F(s(\delta)) > \hat{P}_F^q + \varepsilon\}$, for $q \in \{0, 1\}$. If the social choice function F is immune to robust coalitions, it follows that all belief systems must assign zero probability to the events $\{s \mid Q_F(s(\delta)) = q, \bar{P}_F(s(\delta)) > \hat{P}_F^q + \varepsilon\}$, for $q \in \{0, 1\}$. Since ε can be arbitrarily small, it follows that all belief systems must assign zero probability to the events $\{s \mid Q_F(s(\delta)) = q, \bar{P}_F(s(\delta)) > \hat{P}_F^q\}$, for $q \in \{0, 1\}$. The proposition follows immediately. \blacksquare

5.1 Robust Implementability, Monotonicity, and Immunity to Robust Blocking Imply Voting

Given Proposition 2, 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 $F = (Q_F, \bar{P}_F)$. Given such a social choice function, we denote by

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

the sets of payoff types of net gainers and net losers from public-good provision, respectively. These two sets will play a key role in our analysis. For ease of exposition, we will drop the argument $P_F^1 - P_F^0$ whenever this can be done without creating confusion.

If $s \in \mathcal{M}^{na}(V)$, then almost all people have payoff types in V_1 or V_0 , i.e.,

$$s(V_1) + s(V_0) = 1. \quad (13)$$

Because the distribution s has no mass points, the set of people who are indifferent between the two alternatives can be neglected because it has measure zero. These people have the public-good valuation $v = P_F^1 - P_F^0$, and $s \in \mathcal{M}^{na}(V)$ implies $s(\{P_F^1 - P_F^0\}) = 0$.

Theorem 2 *If a monotonic, anonymous and robustly implementable social choice function F is immune to robust blocking, then for all s and s' in $\mathcal{M}^{na}(V)$,*

$$s(V_1) \geq s'(V_1) \quad \text{implies} \quad Q_F(s) \geq Q_F(s'). \quad (14)$$

Proof. Let F satisfy the conditions of the theorem. For any $\sigma \in [0, 1]$, define

$$Q_F^*(\sigma) := Q_F(\sigma s_{v_{\max}} + (1 - \sigma) s_{v_{\min}}), \quad (15)$$

where $s_{v_{\max}}$ and $s_{v_{\min}}$ are the measures on V that assign all probability mass to the extreme points v_{\max} and v_{\min} . We claim that, for any $s \in \mathcal{M}^{na}(V)$, it must be the case that

$$Q_F(s) = Q_F^*(s(V_1)). \quad (16)$$

By (15) and the monotonicity of Q_F , $Q_F^*(\cdot)$ is a nondecreasing function, so the conclusion of the theorem follows immediately from (16).

It remains to prove (16). Suppose that $Q_F(s) < Q_F^*(s(V_1))$ for some $s \in \mathcal{M}^{na}(V)$. Then $Q_F(s) = 0$ and $Q_F^*(s(V_1)) = 1$. We claim that, in this case, F can be robustly blocked on the naive type space (V, \mathcal{V}) . To see this, consider the collective deviation (V_1, ℓ_{V_1}) such that, for any $s' \in \mathcal{M}(V)$ with $s'(V_1) > 0$,

$$\ell_{V_1}(s'_{V_1}) = \frac{1}{s'(V_1)} \cdot s'_{V_1} \quad \text{if } Q_F^*(s'(V_1)) = 0, \quad (17)$$

and

$$\ell_{V_1}(s'_{V_1}) = s_{v_{\max}} \quad \text{if } Q_F^*(s'(V_1)) = 1, \quad (18)$$

where s'_{V_1} is the restriction of s' to V_1 . The report $\ell_{V_1}(s'_{V_1})$ is in fact the truth if $Q_F^*(s'(V_1)) = 0$, but, if $Q_F^*(s'(V_1)) = 1$, all coalition members lie by saying that their public-good valuation is v_{\max} .

By the definition of Q_F^* and the monotonicity of Q_F , we have

$$Q_F^*(s'(V_1)) \leq Q_F(\hat{s}(\ell_{V_1}, s')) \quad (19)$$

for all s' . Hence $Q_F^*(s'(V_1)) = 1$ implies $Q_F(\hat{s}(\ell_{V_1}, s')) = 1$ for all s' . Therefore, the collective deviation (V_1, ℓ_{V_1}) blocks F on the type space $[(V, \mathcal{V}), \beta]$ for any β such that, for any $t \in V_1$, $\beta(t) \in \mathcal{M}(\mathcal{M}^{na}(V))$ assigns positive probability to the set

$$\{s' \in \mathcal{M}^{na}(V) \mid Q_F(s') = 0 \quad \text{and} \quad Q_F^*(s'(V_1)) = 1\}. \quad (20)$$

If $Q_F(s) < Q_F^*(s(V_1))$ for some $s \in \mathcal{M}^{na}(V)$, the set (20) is nonempty, and the collective deviation (V_1, ℓ_{V_1}) blocks F robustly on (V, \mathcal{V}) . Conversely, immunity to robust blocking implies that the set (20) is empty and hence that there is no $s \in \mathcal{M}^{na}(V)$ such that $Q_F(s) < Q_F^*(s(V_1))$.

Alternatively, if $Q_F(s) > Q_F^*(s(V_1))$ for some $s \in \mathcal{M}^{na}(V)$, a similar argument shows that F is robustly blocked by the collective deviation (V_0, ℓ_{V_0}) such that, for any $s' \in \mathcal{M}^{na}(V)$ with $s'(V_0) > 0$,

$$\ell_{V_0}(s'_{V_0}) = \frac{1}{s'(V_0)} \cdot s'_{V_0} \quad \text{if } Q_F^*(1 - s'(V_0)) = 1, \quad (21)$$

and

$$\ell_{V_0}(s'_{V_0}) = s_{v_{\min}} \quad \text{if } Q_F^*(1 - s'(V_0)) = 0. \quad (22)$$

For any $s' \in \mathcal{M}^{na}(V)$, (13) implies

$$s'(V_0) + s'(V_1) = 1,$$

so conditioning on $1 - s'(V_0)$ in (21) and (22) is equivalent to conditioning on $s'(V_1)$. By the same argument as before, the collective deviation (V_0, ℓ_{V_0}) blocks F on $[(V, \mathcal{V}), \beta]$ for any β such that, for any $t \in V_1$, $\beta(t) \in \mathcal{M}(\mathcal{M}^{na}(V))$ assigns positive probability to the set

$$\{s' \in \mathcal{M}^{na}(V) \mid Q_F(s') = 1 \quad \text{and} \quad Q_F^*(s'(V_1)) = 0\}. \quad (23)$$

If $Q_F(s) > Q_F^*(s(V_1))$ for some $s \in \mathcal{M}^{na}(V)$, the set (23) is nonempty, and the collective deviation (V_0, ℓ_{V_0}) blocks F robustly on (V, \mathcal{V}) .

Immunity against robust blocking thus yields (16) for any $s \in \mathcal{M}^{na}(V)$. ■

The underlying argument is very simple: If the social choice function is monotonic, people with payoff types in V_1 cannot lose by claiming that their valuations are at the very top, at v_{\max} , and people with payoff types cannot lose by claiming that their valuations are at the very bottom, at v_{\min} . Moreover, coalitions of such people can win by such claims, unless the social choice function abstracts from intensities of preferences altogether. Indeed, they can do so robustly, i.e. the strategies of claiming extreme valuations can be used for blocking no matter what the specification of beliefs may be. To be immune against robust blocking, the social choice function must therefore take the form (16).

Monotonicity of the social choice function is not a very restrictive condition. The first-best public-good provision rule in Theorem 1 is monotonic. Moreover, in models with finitely many participants, monotonicity is implied by (robust) incentive compatibility. Monotonicity therefore is a natural condition to impose if we think of the continuum model as an idealization of models with finitely many participants.

The proof of Theorem 2 does not actually use the full strength of the assumption that Q_F is monotonic. All that is used is the assumption that, for any σ , the function

$$(s_0, s_1) \rightarrow Q_F(\sigma s_1 + (1 - \sigma)s_0) \quad (24)$$

from $\mathcal{M}(V) \times \mathcal{M}(V)$ into $\{0, 1\}$ has a saddle-point, i.e. that there exists a pair $(s_0^*, s_1^*) \in \mathcal{M}(V)^2$ such that, for all $(s_0, s_1) \in \mathcal{M}(V)^2$,

$$Q_F(\sigma s_1^* + (1 - \sigma)s_0) \geq Q_F(\sigma s_1^* + (1 - \sigma)s_0^*) \geq Q_F(\sigma s_1 + (1 - \sigma)s_0^*).$$

Monotonicity of Q_F ensures that the pair $(s_{v_{\min}}, s_{v_{\max}})$ is a saddle-point of (24). This property is crucial for the proof of Theorem 2.³⁴

Monotonicity, or the existence of a saddle-point for (24), would not be needed for Theorem 2 if the social choice function were required to satisfy robust coalition proofness rather than immunity to robust blocking.³⁵ If collective deviations can be conditioned on belief systems, then for complete-information belief systems, they can actually be conditioned on the distribution of types among people who are not coalition members. This conditioning allows a coalition to avoid the risk that, for some constellation of reports of the non-members, the collective deviation might move the level of public-good provision in the wrong direction. The analysis here shows that this risk is irrelevant if the social choice function is monotonic. In this case, collective deviations in the direction of exaggerating preferences, upwards for people with payoff types in V_1 , downwards for people with payoff types in V_0 , will never mover the level of public-good provision in a direction that coalition members dislike.

Theorem 2 implies that, for any monotonic, anonymous and robustly implementable social choice function F that is immune to robust blocking, there exists a threshold level \bar{s}_1 such that

$$Q_F(s) = 1 \quad \text{implies} \quad s(V_1) > \bar{s}_1 \tag{25}$$

and

$$Q_F(s) = 0 \quad \text{implies} \quad s(V_1) < \bar{s}_1. \tag{26}$$

To implement such F , it suffices to ask who is in favour of having the public good provided and to condition the outcome on whether or not the population share of the people in favour exceeds the given threshold \bar{s}_1 . As an immediate corollary to Theorem 2, we therefore obtain:

Corollary 2 *If an anonymous and robustly implementable monotonic social choice function F is immune to robust blocking, then it can be implemented by a simple voting mechanism.*

In view of (13), the implication (14) in Theorem 2 is equivalent to

$$s(V_0) \geq s'(V_0) \quad \text{implies} \quad Q_F(s) \leq Q_F(s'), \tag{27}$$

and the public-good provision rule in (25) and (26) is equivalent to the rule

$$Q_F(s) = 1 \quad \text{implies} \quad s(V_0) < 1 - \bar{s}_1, \tag{28}$$

³⁴The saddle-point condition can be interpreted as a Nash equilibrium condition for a strictly competitive game between the coalition organizers for people with payoff types in V_0 and for people with payoff types in V_1 , respectively, when the population shares of the two coalitions are $1 - \sigma$ and σ .

³⁵For economies with finitely many agents, this is the main result in Bierbrauer and Hellwig (2015). The argument given there applies in the continuum model as well.

$$Q_F(s) = 0 \text{ implies } s(V_0) > 1 - \bar{s}_1. \quad (29)$$

Thus the voting mechanism might also ask who opposes the provision of the public good, stipulating provision if and only if the vote share of the opponents is sufficiently small.

What about people who are indifferent? The preceding analysis has relied on the assumption that almost nobody is indifferent. If this assumption is violated, i.e. if we allow for cross-section payoff type distributions with mass points at $v = P_F^1 - P_F^0$, we would actually get a nonexistence result. In this case, suitable adaptations of the arguments in the proof of Theorem 2 would lead to the conclusion that the social choice function must satisfy both implications, (14) and (27).³⁶ Then there exist thresholds \bar{s}_1 and \bar{s}_0 (not necessarily equal to $1 - \bar{s}_1$) such that the public-good provision rule satisfies (25), (26) as well as (28), (29) with $1 - \bar{s}_1$ replaced by \bar{s}_0 . However, this is only possible if $\bar{s}_1 = 0$ or $\bar{s}_0 = 1$, i.e. if $Q_F(s) = 1$ for all s or $Q_F(s) = 0$ for all s . If we allow for payoff type distributions with mass points at $v = P_F^1 - P_F^0$, a monotonic, anonymous, and robustly incentive-compatible social choice function that is immune to robust blocking cannot have a nontrivial rule for public-good provision.³⁷

In Theorem 2, the set of individuals who are indifferent between the two alternatives has measure zero because the analysis is restricted to admissible belief systems, i.e. belief systems that assign probability zero to payoff type distributions with mass points. Any other restriction that serves the same purpose will yield the same conclusion. Thus, we obtain:

Remark 1 *The conclusion of Theorem 2 holds for any payoff type space V and any social choice function F with outcome-dependent payments P_F^0 and P_F^1 such that $P_F^1 - P_F^0 \notin V$.*

This extension of Theorem 2 is important for model specifications with discrete payoff types. With discrete payoff types, the assumption that the payoff type distribution has no mass points does not make much sense. Remark 1 shows that the conclusion of Theorem 2 holds anyway unless the difference $P_F^1 - P_F^0$ is an element of V . In the example of Section 2, with $V = \{0, 3, 10\}$, this requirement is satisfied if the social choice function stipulates equal cost sharing, i.e., $P_F^0 = 0$, and $P_F^1 = k = 4$.

5.2 Voting Mechanisms are Immune to Robust Blocking

The following result provides a converse to Theorem 2 and its corollary.

Theorem 3 *If an anonymous, robustly implementable social choice function $F = (Q_F, P_F^0, P_F^1)$ satisfies condition (14) for all s and s' in $\mathcal{M}^{na}(V)$, then F is immune to robust blocking.*

³⁶In the argument used to establish (14), group people who are indifferent with people who have payoff types in $V_0(P_F^1 - P_F^0)$. To establish (27), use a symmetric argument, where people who are indifferent are grouped with people who have payoff types in $V_1(P_F^1 - P_F^0)$.

³⁷This destructive conclusion is avoided if, somewhat arbitrarily, we assume that people with payoff type $v = P_F^1 - P_F^0$ never join a coalition with people whose payoff types are in V_0 (or never join a coalition with people whose payoff types are in V_1). Equivalently, in the definition of blocking, we might have a weak Pareto criterion for coalitions that intend to raise the level of public-good provision and a strict Pareto criterion for coalitions that intend to lower the level of public-good provision.

Proof. Let $F = (Q_F, P_F^1, P_F^0)$ be a robustly incentive compatible anonymous social choice function that satisfies condition (14). We will show that F is also immune to robust blocking. For suppose that the social choice function F is blocked on some type space $[(T, \mathcal{T}), \tau, \beta]$ with an admissible belief system β . Let \hat{T} be the blocking coalition, and let $\hat{V} = \tau(\hat{T})$ be the set of associated payoff types. Depending on how \hat{V} relates to the sets V_0 and V_1 , we distinguish three cases.

Case 1: $\hat{V} \cap V_1 = \emptyset$.

In this case, obviously, $\hat{V} \cap V_0 \neq \emptyset$, and there is no loss of generality in assuming that $\hat{V} = V_0$ and $\hat{T} = T_0 := \tau^{-1}(V_0)$. For a collective deviation ℓ_{T_0} by people with payoff types in V_0 to block F , there must exist some cross-section type distribution δ such that

$$Q_F(s(\delta)) = 1$$

and

$$Q_F(s(\hat{\delta}(\ell_{T_0}, \delta))) = 0.$$

By condition (14), it follows that the set of agents whose reports suggest that they have valuations in V_1 must be strictly greater at δ than at $\hat{\delta}(\ell_{T_0}, \delta)$, i.e.

$$\delta(T_1) > \hat{\delta}(T_1 | \ell_{T_0}, \delta). \quad (30)$$

However, the definition (8) of $\hat{\delta}(\ell_{T_0}, \delta)$ implies

$$\hat{\delta}(T_1 | \ell_{\hat{T}}, \delta) \geq \delta(T_1), \quad (31)$$

which is incompatible with (30). The assumption that there is a blocking coalition \hat{T} with $\tau(\hat{T}) \cap V_1 = \emptyset$ has thus led to a contradiction and must be false.

Case 2: $\hat{V} \cap V_0 = \emptyset$.

The argument is precisely symmetric to the one given for Case 1 and is left to the reader.

Case 3: $\hat{V} \cap V_0 \neq \emptyset$ and $\hat{V} \cap V_1 \neq \emptyset$.

For a collective deviation $\ell_{\hat{T}}$ by a coalition of people with payoff types in V_0 and people with payoff types in V_1 to block F at $[(T, \mathcal{T}), \tau, \beta]$, there must exist nonempty sets D_{10} and D_{01} of type distributions in $\mathcal{M}(T)$ such that, for $\delta_0 \in D_{10}$,

$$Q_F(s(\delta_0)) = 1 \quad \text{and} \quad Q_F(s(\hat{\delta}(\ell_{\hat{T}}, \delta_0))) = 0,$$

and for $\delta_1 \in D_{01}$,

$$Q_F(s(\delta_1)) = 0 \quad \text{and} \quad Q_F(s(\hat{\delta}(\ell_{\hat{T}}, \delta_1))) = 1.$$

Moreover, the different participants' beliefs must be such that the different participants attach more weight to the gains from the changes they like than to the losses from the changes they dislike. Thus, for an agent with type $t_0 \in T_0$, it must be the case that

$$\frac{\beta(D_{10} | t_0)}{\beta(D_{01} | t_0)} \geq 1,$$

and for an agent with type $t_1 \in T_1$,

$$\frac{\beta(D_{10} | t_1)}{\beta(D_{01} | t_1)} \leq 1,$$

and at least one of the inequalities must be strict for a nonnegligible set of agents in \hat{T} . However, not all belief systems that assign positive probability to δ_0 and δ_1 satisfy these inequalities. Blocking of F by the coalition \hat{T} therefore cannot be robust. ■

Without any additional conditions, Theorem 3 would not be true if we replaced the requirement of immunity to robust coalitions by a requirement of coalition-proofness which allows for belief-dependent coalition formation: Collective deviations by individuals with payoff types in both V_0 and V_1 would block F if beliefs were such that the “trade” involved in raising public-good provision in some states while lowering it in other states would seem mutually beneficial. In the proof of Theorem 3 the requirement that blocking by a collective deviation should be robust is used to eliminate this possibility. If belief-dependent coalitions are allowed for, such “trades” can also be eliminated by requiring coalitions themselves to be immune to the formation of subcoalitions. A coalition of people with payoff types in both V_0 and V_1 that tries to implement such a “trade” is itself vulnerable to the possibility that one of the subgroups might use an additional collective deviation of its own in order to manipulate reports to the primary coalition organizer so that those states in which the original coalition would provide for a manipulation that people in the subgroup dislike would never be “observed”. In Bierbrauer and Hellwig (2015), we use a version of this argument to prove an analogue of Theorem 3 for robust coalition-proofness in finite economies.³⁸

6 Welfare Implications

Limits to First-Best Implementation. From Theorems 1 and 2, we immediately obtain:

Corollary 3 *If there exist s and s' in $\mathcal{M}^{na}(V)$, such that $s(V_0(k)) \leq s'(V_0(k))$ and $\bar{v}(s) < k < \bar{v}(s')$, then there is no anonymous social choice function that yields first best outcomes and that is robustly implementable and immune to robust blocking.*

Whereas the literature on public-goods provision in small economies has mainly developed impossibility theorems on the basis of participation constraints and budget balance, Theorem 1 shows that, in the large economy, those concerns are moot, but first-best implementation may be vulnerable to manipulation by groups with common interest. We saw this already in the example in Section 2, where the size of the group of net beneficiaries was always 0.3 and the first-best provision rule depended on the population shares of people with valuation 0 and people with valuation 3, all of whom would be harmed by public-good provision. In that example,

³⁸In Bierbrauer and Hellwig (2015), we actually use the Bernheim et al. (1986) approach in which subcoalitions themselves must again be immune to blocking by admissible subsubcoalitions, etc. In the continuum model, this concept involves the difficulty that there is no minimal coalition on which to build a hierarchy of admissible coalitions of increasing size.

obviously the conditions of Corollary 3 are satisfied, for example, with $s_0 = 0.7$, $s_3 = 0$, $s_{10} = 0.3$ and $s'_0 = 0$, $s'_3 = 0.7$, $s'_{10} = 0.3$.

The following table provides another example, which will be useful for discussing second-best considerations. 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.³⁹

Table 3.

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

(32)

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

For the example in Table 3, 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 3, one immediately sees that the set of net beneficiaries has a population share of 0.7 in state 1 and of 0.6 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 Theorem 1. In more concrete terms, any mechanism that would implement a social choice function with first-best outcomes would be robustly blocked by a coalition of people with valuations 5 and 10. The coalition would use a collective deviation that involves truthtelling if the coalition's population share is 0.6. If the population share of the coalition is 0.7, the collective deviation involves reporting the valuation 10 with probability 5/7, and the valuations 0 and 5 with probability 1/7 each. The overall mechanism is thus made to believe that the state is 2, even when the true state is 1.

Second-Best Considerations. If first-best is out of reach, the overall mechanism designer faces 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 the requirement of immunity against robust blocking. For instance, with the example in Table 3, 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

³⁹The payoff type distributions in Table 3 obviously have mass points. However, with equal cost sharing the assumption of Remark 1 is satisfied so the conclusion of Theorem 2 holds.

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 immunity against robust blocking. This is equivalent to the problem of choosing P_F^0 , P_F^1 and $Q_F : \mathcal{M}(V) \rightarrow \{0, 1\}$ so as to maximize the expected aggregate surplus

$$E^M[(\bar{v}(s) - P_F^1)Q_F(s) - P_F^0(1 - Q_F(s))] \quad (33)$$

subject to the feasibility constraints that $P_F^0 \geq 0$, $P_F^1 \geq k$, and the condition that for every pair s and s' , $s(V_1) \geq s'(V_1)$ implies $Q_F(s) \geq Q_F(s')$. The expectations operator E^M in (33) indicates that expectations over s are taken with respect to the mechanism designer's subjective beliefs.

With Table 3, 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 an excessive payment 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 immunity to robust coalitions.⁴⁰

⁴⁰Similarly, in the example of Section 2, the mechanism designer might want to impose a payment of 1 whenever the public good is not provided and a payment of 4 whenever the public good is provided.

7 Multiple Provision Levels

To what extent does our analysis depend on the assumption that the public goods comes in a single indivisible unit so that there are only two alternatives, provision and non-provision? Extending Theorem 1 to allow for multiple provision levels is obviously unproblematic. The finding that, in a large economy, first-best public-good provision with equal cost sharing is robustly implementable holds quite generally.

The question is whether Theorems 2 and 3 also hold for the general case of multiple provision levels. In the following, we provide an answer to this question.

Multiple provision levels with a linear cost function. Extending our previous analysis is easy if the provision cost function is linear (up to some upper bound \bar{Q}) and the payment rule provides for equal cost sharing. In this case, if the provision cost per capita is kQ , there are again two sets of payoff types, $V_0 = \{v|v < k\}$ and $V_1 = \{v|v > k\}$ so that, for any two provision levels Q_0 and Q_1 with $Q_1 > Q_0$ agents with payoff types in V_0 prefer Q_0 and agents with payoff types in V_1 prefer Q_1 . An easy adaptation of the proof of Theorem 2 then shows that, if a monotonic, anonymous and robustly implementable social choice function is immune to robust coalitions, the public-good provision rule must take the form

$$Q_F(s) = Q^*(s(V_1)),$$

where Q^* is a nondecreasing function. Such a social choice function can again be implemented by a voting mechanism, though this time, the votes would not be about the alternatives that are to be chosen. The vote might be about whether to have zero public-good provision or not, and the level would depend on the population share of the people in favour of some public-good provision. (With utility taking the form vQ , of course, efficiency would mandate that Q^* should take the values 0 and \bar{Q} so the linear-cost case is similar to the binary case anyway.)

Multiple provision levels with increasing marginal costs. Extending our analysis to multiple provision levels is nontrivial, however, if the cost of an additional unit of the public-good is increasing or decreasing in Q . The case of increasing marginal costs is particularly interesting. In the following, we show how our approach works when there are $n + 1$ provision levels $0, 1, \dots, n$ and the provision cost is k_1 for $Q = 1$, $k_1 + k_2$ for $Q = 2$, and, more generally, $\sum_{i=1}^j k_i$ for $K = j$, where

$$v_{\min} < k_1 < k_2 < \dots < k_n < v_{\max}. \quad (34)$$

Depending on the distribution of payoff types, any one of the provision levels $0, 1, \dots, n$ can be efficient. Specifically, a first-best provision rule requires that

$$Q_F^{FB}(s) = \begin{cases} 0, & \text{if } \bar{v}(s) < k_1, \\ j, & \text{if } \bar{v}(s) \in (k_{j-1}, k_j), j = 1, \dots, n - 1, \\ n, & \text{if } \bar{v}(s) > k_n. \end{cases}$$

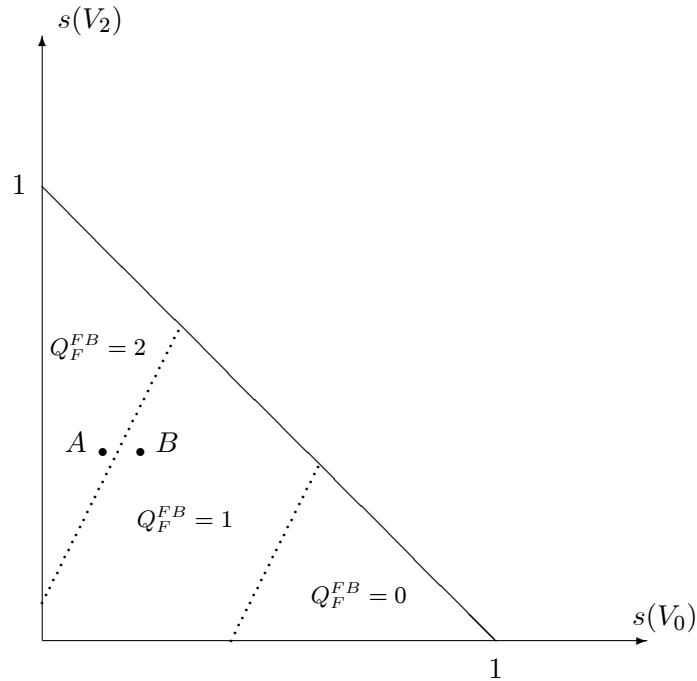
Figure 1 below illustrates the first-best provision rule for the case $n = 2$, when there are three provision levels. On the axes, we have the population shares $s(V_0)$ and $s(V_2)$ of the sets $V_0 =$

$\{v|v < k_1\}$ and $V_2 = \{v|v > k_2\}$. For $s \in \mathcal{M}^{na}(V)$, the population share $s(V_1)$ of the set $V_1 = \{v|k_1 < v < k_2\}$ of people who prefer the intermediate provision level $Q = 1$ is $1 - s(V_0) - s(V_2)$. The figure represents the set

$$\{(s(V_0), s(V_2)) \mid 0 \leq s(V_0) \leq 1, 0 \leq s(V_2) \leq 1 \text{ and } s(V_0) + s(V_2) \leq 1\}. \quad (35)$$

of possible constellations of $s(V_0)$ and $s(V_2)$ with the understanding that $s(V_1) = 1 - s(V_0) - s(V_2)$. Under the assumption that the conditional means of v on the sets V_0, V_1, V_2 are fixed and do not change as $s(V_1)$ changes,⁴¹ the figure shows how the first-best provision rule depends on the population shares of the three groups.

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



However, the first-best provision rule is again not immune to robust blocking. To see this, consider the points $A = (s(V_0)^A, s(V_2)^A)$ and $B = (s(V_0)^B, s(V_2)^B)$ in Figure 1. These points have been chosen so that $s(V_2)^A = s(V_2)^B$, $Q_F^{FB}(s^A) = 2$, and $Q_F^{FB}(s^B) = 1$. Consider the naive type space (V, \mathcal{V}) and a collective deviation $(V_0 \cup V_1, \ell_{V_0 \cup V_1})$ by agents with payoff types in $V_0 \cup V_1$ such that, whenever, the population share of this coalition is $1 - s(V_2)^A$ and whenever the population share of agents with payoff types in V_0 is $s(V_0)^A$, the reporting strategy $\ell_{V_0 \cup V_1}$ has a share $\frac{s(V_0)^B}{1 - s(V_2)^A}$ of coalition members report that their payoff types are in V_0 and a share $1 - \frac{s(V_0)^B}{1 - s(V_2)^A}$ of coalition members report that their payoff types are in V_1 . As a result of this manipulation, the overall mechanism realizes $Q_F^{FB}(s^B)$ even though the actual state of the economy is s^A . All coalition members benefit from this manipulation and, from an interim perspective, all coalition

⁴¹This assumption is only made in Figure 1. Our subsequent analysis does without it.

members expect to benefit from the collective deviation $(V_0 \cup V_1, \ell_{V_0 \cup V_1})$ whenever they have beliefs that assign positive probability to state s^A . Thus the collective deviation $(V_0 \cup V_1, \ell_{V_0 \cup V_1})$ blocks the first-best provision rule robustly on (V, \mathcal{V}) .

What can we say in general about the implications of requiring social choice functions to be immune to robust blocking, as well as anonymous and robustly incentive-compatible? The following result answers this question for a particular class of social choice functions, whose range includes all possible provision levels.

Theorem 4 *Consider a monotonic, anonymous and robustly implementable social choice function F for the model with $n + 1$ public-good provision levels with increasing marginal provision costs and suppose that the payment rule provides for equal cost sharing. Assume that there exist numbers $\sigma_i, i = 0, \dots, n$, in $[0, 1]$ such that, for any i ,*

$$Q_F(\sigma_i s_{v_{\max}} + (1 - \sigma_i) s_{v_{\min}}) = i, \quad (36)$$

where $s_{v_{\min}}$ and $s_{v_{\max}}$ are the degenerate measures that assign probability one to the minimum v_{\min} and the maximum v_{\max} of V . If F is immune to robust blocking, then there exist numbers $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_n$ such that

$$0 < \bar{s}_1 < \dots < \bar{s}_n < 1 \quad (37)$$

and, moreover, for any $s \in \mathcal{M}^{na}(V)$ and $j = 0, \dots, n$,

$$Q_F(s) = i, \text{ if } s(W_{i-1}^+) > \bar{s}_i \text{ and } s(W_i^+) < \bar{s}_{i+1}, \quad (38)$$

where, for notational convenience, we write $\bar{s}_0 = 0$ and $\bar{s}_{n+1} = 1$, and where, for $i = 0, 1, \dots, n$, $W_i^+ = (k_{i+1}, v_{\max}]$ is the set of payoff types for whom an increase in the provision level from $Q = i$ to $Q = i + 1$ is desirable. Moreover, for $j = 1, \dots, n$, $Q_F(s)$ is constant on each of the sets $\{s \mid \sum_{i=j}^n s(V_i) = \bar{s}_j\}$, taking either the value $Q_F(s) = j$ or the value $Q_F(s) = j - 1$.

Corollary 4 *If a social choice function satisfies the assumptions of Theorem 4, then it can be implemented by a sequence of votes on the question whether people want to raise the public good by one unit, first from zero to one, then from one to two, ... finally from $n-1$ to n . Thresholds for proceeding to the next stage are given by $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_n$. If stage j is reached but the threshold for stage $j + 1$ is not met, the public-good provision level is $Q_F(s) = j$.*

For the case $n = 2$, Theorem 4 is illustrated in Figure 2, which again shows the set (35). The set V_2 of payoff types that prefer the provision level $Q = 2$ is equal to the set W_1^+ of payoff types for whom an increase from $Q = 1$ to $Q = 2$ is desirable. If s is atomless, the set $V_1 \cup V_2$ of payoff types that prefer the provision level $Q = 1$ or the provision level $Q = 2$ has the same measure as the set W_0^+ of payoff types for whom an increase from $Q=0$ to $Q=1$ is desirable, i.e. $s(W_0^+) = s(V_1 \cup V_2) = s(V_1) + s(V_2)$. According to the theorem, the public good is provided at

the maximum level $Q = 2$ if $s(W_1^+) > \bar{s}_2$, which corresponds to the case $s(V_2) > \bar{s}_2$ in the figure. It is provided at the level $Q = 1$ if $s(W_1^+) < \bar{s}_2$ and $s(W_0^+) > \bar{s}_1$, which corresponds to the case $s(V_2) < \bar{s}_2$ and $s(V_0) = 1 - s(V_1) - s(V_2) < 1 - \bar{s}_1$ in the figure. It is not provided at all if $s(W_0^+) < \bar{s}_1$, which corresponds to the case $s(V_0) > 1 - \bar{s}_1$ in the figure.

Figure 2: Immunity to robust blocking: The case of three provision levels

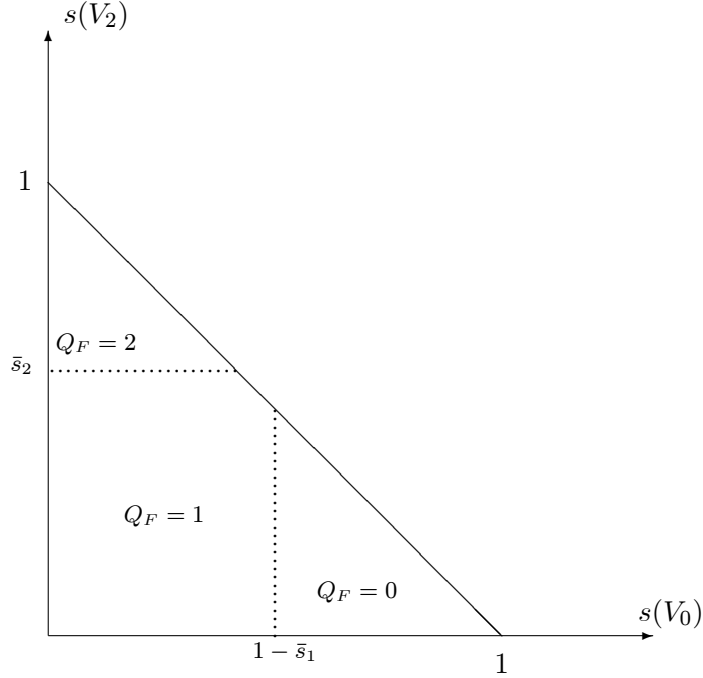


Figure 2 also illustrates the condition that there exist numbers $\sigma_i, i = 0, \dots, n$, in $[0, 1]$ such that, for any i , the public-good provision level at $\sigma_i s_{v_{\max}} + (1 - \sigma_i) s_{v_{\min}}$ satisfies (36). For any $\sigma \in [0, 1]$, the payoff type distribution $\sigma s_{v_{\max}} + (1 - \sigma) s_{v_{\min}}$ corresponds to a point on the boundary of the set (35). The condition requires that, as one moves along this boundary from the point $(1, 0)$ to the point $(0, 1)$, Q_F takes all possible values, from 0 to n , in the figure the values 0, 1, 2; without ever jumping, e.g. from 0 to 2. Below, we discuss this assumption further.

7.1 Proof of Theorem 4

To prove Theorem 4, we first specify the thresholds $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_n$.

Lemma 1 *Under the assumptions of Theorem 4, there exist thresholds $\bar{s}_1, \bar{s}_2, \dots, \bar{s}_n$ satisfying (37) such that*

$$Q_F(\sigma s_{v_{\max}} + (1 - \sigma) s_{v_{\min}}) = j, \quad \text{if } \sigma \in (\bar{s}_j, \bar{s}_{j+1}), \quad (39)$$

where again $\bar{s}_0 = 0$ and $\bar{s}_{n+1} = 1$.

Proof. By the monotonicity of Q_F , the function

$$\sigma \rightarrow \hat{Q}_F(\sigma) := Q_F(\sigma s_{v_{\max}} + (1 - \sigma)s_{v_{\min}}) \quad (40)$$

is nondecreasing. Under the assumptions of Theorem 4, for each j , the set $\{\sigma \in [0, 1] \mid \hat{Q}_F(\sigma) = j\}$ is nonempty. For $j \geq 1$, let \bar{s}_j be the infimum of this set. (39) and (37) follow by the monotonicity of the function (40). \blacksquare

We next proceed by induction. We first show that, for any $s \in \mathcal{M}^{na}(V)$, the claims of the Theorem are true if $s(W_{n-1}^+) > \bar{s}_n$. Then we show that, if they are true for s satisfying $s(W_j^+) > \bar{s}_{j+1}$, then they must also be true for any s satisfying $s(W_{j-1}^+) > \bar{s}_j$. Throughout the analysis, we make repeated use of the fact that, if $s \in \mathcal{M}^{na}(V)$, then for any $j < n$,

$$s(W_j^+) = \sum_{i=j+1}^n s(V_i). \quad (41)$$

We also introduce the notation W_j^- for the set of payoff types for whom a decrease of the provision level from $Q = j$ to $Q = j - 1$ would be desirable, and we note that, if $s \in \mathcal{M}^{na}(V)$, then for any $j > 0$,

$$s(W_j^-) = \sum_{i=0}^{j-1} s(V_i). \quad (42)$$

Lemma 2 *Under the assumptions of Theorem 4,*

$$Q_F(s) = n$$

for any $s \in \mathcal{M}^{na}(V)$ such that $s(V_n) > \bar{s}_n$.

Proof. We show that if the lemma is false, then F can be robustly blocked on the naive type space (V, \mathcal{V}) . If the lemma is false, there exists $s \in \mathcal{M}^{na}(V)$ such that $s(V_n) > \bar{s}_n$ and $Q_F(s) < n$. Consider the collective deviation (V_n, ℓ_{V_n}) where ℓ_{V_n} takes the constant value $s_{v_{\max}}$. By monotonicity, we have

$$Q_F(\hat{s}(\ell_{V_n}, s')) \geq Q_F(s') \quad (43)$$

for all $s' \in \mathcal{M}^{na}(V)$. By monotonicity, we also have

$$Q_F(\hat{s}(\ell_{V_n}, s')) \geq \hat{Q}_F(s'(V_n)). \quad (44)$$

By Lemma 1 and monotonicity, it follows that

$$Q_F(\hat{s}(\ell_{V_n}, s')) = n \quad (45)$$

whenever $s'(V_n) > \bar{s}_n$. For any $s \in \mathcal{M}^{na}(V)$, therefore, $s(V_n) > \bar{s}_n$ and $Q_F(s) < n$ together imply that $Q_F(\hat{s}(\ell_{V_n}, s)) > Q_F(s)$. The collective deviation (V_n, ℓ_{V_n}) blocks F on the naive type space

$[(V, \mathcal{V}), \beta]$ for any belief system that has people with payoff types in V_n give positive probability to the set of states s with $s(V_n) > \bar{s}_n$ and $Q_F(s) < n$. \blacksquare

In the induction argument, we repeatedly refer to Figure 3. This figure is similar to Figure 2, except that now, for some fixed j , the axes show the population shares $s(W_j^-)$ and $s(W_j^+)$ of the set of people who want the public-good provision level to be lower and the set of people who want the public-good provision level to be higher than j . We assume that (38) is true for all s such that $s(W_j^+) > \bar{s}_{j+1}$ and proceed to show that (38) is also true for all s such that $s(W_j^+) < \bar{s}_{j+1}$ and $s(W_{j-1}^+) > \bar{s}_j$ or, equivalently, $s(W_j^-) < 1 - \bar{s}_j$. The induction assumption corresponds to the statement that (38) is true and $Q_F(s) > j$ for all s corresponding to points above the horizontal line $s(W_j^+) = \bar{s}_{j+1}$ in the figure. The induction argument will show that (38) is also true and $Q_F(s) = j$ for all s corresponding to points in the intermediate area between the horizontal line $s(W_j^+) = \bar{s}_{j+1}$ and the vertical line $s(W_j^-) = 1 - \bar{s}_j$ in the figure. (The theorem also implies that $Q_F(s) < j$ for all s corresponding to points to the right of the vertical line $s(W_j^-) = 1 - \bar{s}_j$ in the figure, but this is not part of the induction argument.)

The different parts of the proof correspond to different sets of points in the figure. In the first step, we verify our claim for distributions $s \in \mathcal{M}^{na}(V)$ for which $s(W_j^+) < \bar{s}_{j+1}$ and $s(W_{j-1}^+) > \bar{s}_j$ and, in addition, $s(V_j) = 0$, or, equivalently, $s(W_j^-) + s(W_j^+) = 1$. This is the set of payoff type distributions that correspond to points on the outer boundary of the triangle in the figure, between the vertical lines $s(W_j^-) = 1 - \bar{s}_{j+1}$ and $s(W_j^-) = 1 - \bar{s}_j$. In the second step, we verify our claim for all remaining distributions $s \in \mathcal{M}^{na}(V)$ for which $s(W_j^-) \in (1 - \bar{s}_{j+1}, 1 - \bar{s}_j)$; these distributions correspond to points on the vertical strip underneath the boundary points considered in the first step. In the third step, finally we establish our claim for all distributions $s \in \mathcal{M}^{na}(V)$ for which $s(W_j^-) \leq 1 - \bar{s}_{j+1}$ and $s(W_j^+) < \bar{s}_{j+1}$. This corresponds to the rectangle between the axes, the vertical line $s(W_j^-) = 1 - \bar{s}_{j+1}$ and the horizontal line $s(W_j^+) = \bar{s}_{j+1}$.

The following lemma establishes our claim for distributions that correspond to points on the outer boundary of the triangle in Figure 3.

Lemma 3 *Fix $j < n$. Under the assumptions of Theorem 4, if (38) holds for all $s \in \mathcal{M}^{na}(V)$ with $s(W_j^+) > \bar{s}_{j+1}$, then (38) also holds for any $s \in \mathcal{M}^{na}(V)$ with $s(V_j) = 0$, and $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$.*

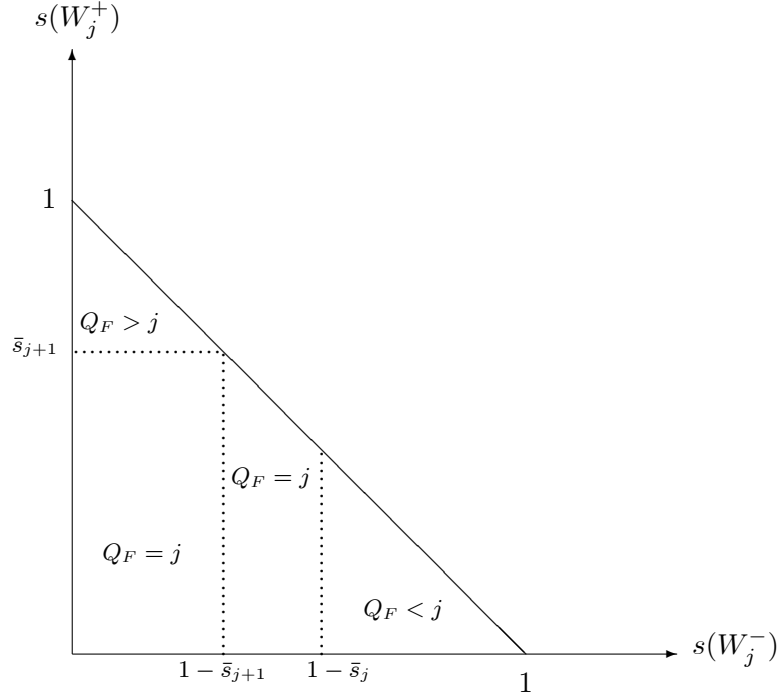
Proof. The proof proceeds in three steps. In the first step, we show that, under the induction hypothesis, for $s \in \mathcal{M}^{na}(V)$ satisfying $s(V_j) = 0$, and $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, $Q_F(s)$ cannot exceed $j + 1$. To see this, consider the payoff type distribution

$$s' = \alpha s_{W_j^+} + (1 - \alpha)s_{W_j^-},$$

where $s_{W_j^+}$ and $s_{W_j^-}$ are the restrictions of s to the sets $W_j^+ = (k_j, v_{\max}]$ and $W_j^- = [v_{\min}, k_j)$ and

$$\alpha := \frac{\bar{s}_{j+1} + \bar{s}_{j+2}}{2} \cdot \frac{1}{s(W_j^+)}.$$

Figure 3: Immunity to robust blocking: The intermediate provision level



(This is well defined because s satisfies $s(W_j^+) > \bar{s}_j$.) The distribution s' satisfies

$$s(W_j^+) = s'(W_j^+) = \alpha s(W_j^+) = \frac{\bar{s}_{j+1} + \bar{s}_{j+2}}{2}$$

and hence $s'(W_j^+) \in (\bar{s}_{j+1}, \bar{s}_{j+2})$. By the induction hypothesis, it follows that $Q_F(s') = j + 1$. Since, by construction, s' dominates s in the sense of first-order stochastic dominance, we also have $Q_F(s) \leq Q_F(s')$ and therefore $Q_F(s) \leq j + 1$.

In the second step, we show that, for $s \in \mathcal{M}^{na}(V)$ satisfying $s(V_j) = 0$, and $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, $Q_F(s)$ cannot be less than j . If this claim were false, then F would be robustly blocked on the naive type space (V, \mathcal{V}) . To see this, consider the collective deviation $(W_j^+ \cup V_j, \ell_{W_j^+ \cup V_j})$ where $\ell_{W_j^+ \cup V_j}$ is a reporting strategy that asks coalition members to report $s_{v_{\max}}$ whenever $s(V_j) = 0$, and $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, and to report the truth in all other circumstances. For any $s' \in \mathcal{M}^{na}(V)$ such that $s'(V_j) = 0$, and $s'(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, this reporting strategy generates a cross-section distribution of reports $\hat{s}(\ell_{W_j^+ \cup V_j}, s')$ that weakly dominates the distribution $\sigma s_{v_{\max}} + (1 - \sigma)s_{v_{\min}}$ for $\sigma = s'(W_j^+)$. By the monotonicity of F , it follows that, for such s' ,

$$Q_F(\hat{s}(\ell_{W_j^+ \cup V_j}, s')) \geq \hat{Q}_F(s(W_j^+)) = j.$$

By the first step of the argument, we also have $Q_F(\hat{s}(\ell_{W_j^+ \cup V_j}, s')) \leq j + 1$. If $s'(V_j) = 0$, it follows that almost all members of the coalition $W_j^+ \cup V_j$ prefer the public-good provision level $Q_F(\hat{s}(\ell_{W_j^+ \cup V_j}, s'))$ to any level below j . Thus, if $Q_F(s)$ were less than j for some $s \in \mathcal{M}^{na}(V)$

satisfying $s(V_j) = 0$, and $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, the collective deviation $(W_j^+ \cup V_j, \ell_{W_j^+ \cup V_j})$ would robustly block F on the naive type space (V, \mathcal{V}) .

In the third step, we show that, for $s \in \mathcal{M}^{na}(V)$ satisfying $s(V_j) = 0$, and $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$ we must have $Q_F(s) = j$. If this claim were false, then F would be robustly blocked on (V, \mathcal{V}) . To see this, consider the collective deviation $(W_j^- \cup V_j, \ell_{W_j^- \cup V_j})$ where $\ell_{W_j^- \cup V_j}$ is a reporting strategy that asks coalition members to report $s_{v_{\min}}$ whenever $s(V_j) = 0$ and $s(W_j^-) \in (1 - \bar{s}_{j+1}, 1 - \bar{s}_j)$ and to report the truth in all other circumstances. For any $s' \in \mathcal{M}^{na}(V)$ such that $s'(V_j) = 0$ and $s'(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, this reporting strategy generates a cross-section distribution of reports $\hat{s}(\ell_{W_j^- \cup V_j}, s')$ that is weakly dominated by the distribution $\sigma s_{v_{\max}} + (1 - \sigma)s_{v_{\min}}$ for $\sigma = 1 - s(W_j^-) = s(W_j^+)$. By the monotonicity of F , it follows that, for such s' ,

$$Q_F\left(\hat{s}(\ell_{W_j^- \cup V_j}, s')\right) \leq \hat{Q}_F\left(s(W_j^+)\right) = j.$$

By the second step of the argument, therefore, $Q_F(\hat{s}(\ell_{W_j^- \cup V_j}, s')) = j$, which all coalition members prefer to any higher level. Thus if $Q_F(s)$ were equal to $j+1$ for some $s \in \mathcal{M}^{na}(V)$ satisfying $s(V_j) = 0$ and $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, the collective deviation $(W_j^- \cup V_j, \ell_{W_j^- \cup V_j})$ would robustly block F on the naive type space (V, \mathcal{V}) . The lemma follows immediately. \blacksquare

The following lemma establishes our claim for distributions corresponding to points in Figure 3 that lie on the vertical strip below the boundary points considered in Lemma 3.

Lemma 4 *Fix $j < n$. Under the assumptions of Theorem 4, if (38) holds for all $s \in \mathcal{M}^{na}(V)$ with $s(W_j^+) > \bar{s}_{j+1}$, then (38) also holds for all $s \in \mathcal{M}^{na}(V)$ with $s(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$.*

Proof. We first note that any $s \in \mathcal{M}^{na}(V)$ with $s(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$ is dominated by the distribution

$$s' = s + s(V_j)s_{v_{\max}} - s_{V_i} \tag{46}$$

which is obtained from s by shifting all mass from the set V_j to the singleton $\{v_{\max}\}$. By construction, $s'(W_{j-1}^+) = s(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$ and $s'(V_j) = 0$. Hence, by Lemma 3, $Q_F(s') = j$. By the monotonicity of F , therefore, $Q_F(s) \leq j$.

To prove the lemma, it therefore suffices to show that, for any $s \in \mathcal{M}^{na}(V)$ such that $s(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$, we cannot have $Q_F(s) < j$. For this purpose, we show that, if we had $Q_F(s) < j$ for some such s , then F could be robustly blocked on the naive type space (V, \mathcal{V}) . To see this, consider the collective deviation $(W_{j-1}^+, \ell_{W_{j-1}^+})$ where $\ell_{W_{j-1}^+}$ is a reporting strategy that asks coalition members to report

$$\ell_{W_{j-1}^+}(s_{W_{j-1}^+}) = \frac{1}{s(W_{j-1}^+)} \cdot s_{W_{j-1}^+} + \frac{s(V_j)}{s(W_{j-1}^+)} s_{v_{\max}} \tag{47}$$

whenever $s(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$ and to report the truth in all other circumstances. For any $s \in \mathcal{M}^{na}(V)$ such that $s(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$, the distribution of reports $\hat{s}(\ell_{W_{j-1}^+}, s)$ that is generated by this reporting strategy satisfies the conditions of Lemma 3 so we have $Q_F(\hat{s}(\ell_{W_{j-1}^+}, s)) = j$.

Since all coalition members prefer the provision level j to any lower level, it follows that, if there were a distribution $s \in \mathcal{M}^{na}(V)$ such that $s(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$ and $Q_F(s) < j$, then the collective deviation $(W_{j-1}^+, \ell_{W_{j-1}^+})$ would block F robustly on (V, \mathcal{V}) . The lemma follows immediately. \blacksquare

Finally, we consider distributions corresponding to points in Figure 3 that lie to the left of the vertical line $s(W_j^-) = 1 - \bar{s}_{j+1}$ and below the horizontal line $s(W_j^+) = \bar{s}_{j+1}$.

Lemma 5 *Fix $j < n$. Under the assumptions of Theorem 4, if (38) holds for all $s \in \mathcal{M}^{na}(V)$ with $s(W_j^+) > \bar{s}_{j+1}$, then (38) also holds for all $s \in \mathcal{M}^{na}(V)$ with $s(W_{j-1}^+) > \bar{s}_j$.*

Proof. Since all other cases are covered by Lemma 4 or by the induction hypothesis, it suffices to prove that (38) holds for all $s \in \mathcal{M}^{na}(V)$ with $s(W_j^+) < \bar{s}_{j+1}$ and $s(W_{j-1}^+) > \bar{s}_j$. For any such s , obviously, there exists $s' \in \mathcal{M}^{na}(V)$ such that s dominates s' and, moreover, $s'(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$. (To obtain s' from s , shift an appropriate amount of mass from V_j to the singleton $\{v_{\min}\}$.) For such s' , Lemma 4 implies $Q_F(s') = j$. For any $s \in \mathcal{M}^{na}(V)$ such that $s(W_j^+) < \bar{s}_{j+1}$ and $s(W_{j-1}^+) > \bar{s}_j$, we therefore have $Q_F(s) \geq j$.

To complete the proof it now suffices to show that, for any $s \in \mathcal{M}^{na}(V)$ such that $s(W_j^+) < \bar{s}_{j+1}$ and $s(W_{j-1}^+) > \bar{s}_j$, we cannot have $Q_F(s) > j$. We give separate arguments for the case $s(W_j^+) > \bar{s}_j$ and for the case $s(W_j^+) \leq \bar{s}_j$.

Case 1. For the case $s(W_j^+) > \bar{s}_j$, we show that, if we had $Q_F(s) > j$ for some such s , then F could be robustly blocked on the naive type space (V, \mathcal{V}) . To see this, consider the collective deviation $(W_{j+1}^-, \ell_{W_{j+1}^-})$ where $\ell_{W_{j+1}^-}$ is a reporting strategy that asks coalition members to report

$$\ell_{W_{j+1}^-}(s_{W_{j+1}^-}) = \frac{1}{s(W_{j+1}^-)} \cdot s_{W_{j+1}^-} + \frac{s(V_j)}{s(W_{j+1}^-)} s_{v_{\min}} - \frac{s_{V_j}}{s(W_{j+1}^-)} \quad (48)$$

if $s(W_{j+1}^-) \in (1 - \bar{s}_{j+1}, 1 - \bar{s}_j)$, or equivalently, if $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, and to report the truth in all other circumstances. Heuristically, the strategy $\ell_{W_{j+1}^-}$ asks all people with payoff types in V_j to report v_{\min} , leaving everything else unchanged. For any $s \in \mathcal{M}^{na}(V)$ such that $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$, the distribution of reports $\hat{s}(\ell_{W_{j+1}^-}, s)$ that is generated by this reporting strategy satisfies the conditions of Lemma 3, so we have $Q_F(\hat{s}(\ell_{W_{j+1}^-}, s)) = j$. Since all coalition members prefer the provision level j to any higher level, it follows that, if there were a distribution $s \in \mathcal{M}^{na}(V)$ such that $s(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$ and $Q_F(s) > j$, then the collective deviation $(W_{j+1}^-, \ell_{W_{j+1}^-})$ would block F robustly on (V, \mathcal{V}) . Hence we must have $Q_F(s) = j$ for any such j .

Case 2. For the case $s(W_j^+) \leq \bar{s}_j$ and $s(W_{j-1}^+) > \bar{s}_j$, it suffices to note that any such s is dominated by a distribution s' satisfying $s'(W_j^+) \in (\bar{s}_j, \bar{s}_{j+1})$ and itself dominates a distribution s'' satisfying $s''(W_{j-1}^+) \in (\bar{s}_j, \bar{s}_{j+1})$. By Case 1, $Q_F(s') = j$. By Lemma 4, $Q_F(s'') = j$. By the monotonicity of F , $Q_F(s') \geq Q_F(s) \geq Q_F(s'')$ and hence $Q_F(s) = j$. This completes the proof of the lemma. \blacksquare

Lemma 6 *Under the assumptions of Theorem 4, (38) holds for all $s \in \mathcal{M}^{na}(V)$.*

Proof. Given Lemmas 2 and 5, the lemma follows by induction on j . ■

Lemma 7 *Under the assumptions of Theorem 4, $Q_F(s) = \hat{Q}_F(\bar{s}_j)$ for any s with $s(W_{j-1}^+) = \bar{s}_j$.*

Proof. By monotonicity, $\hat{Q}_F(\bar{s}_j) = j$ or $\hat{Q}_F(\bar{s}_j) = j - 1$. Whenever there exists a distribution s with $s(W_{j-1}^+) = \bar{s}_j$ and $Q_F(s) \neq \hat{Q}_F(\bar{s}_j)$, one can use the arguments in the proofs of Lemmas 3, 4 and 5 to construct a robustly blocking coalition. Thus, if $\hat{Q}_F(\bar{s}_j) = j$, the coalitions considered in the proofs of Lemmas 3 and 4 could be used to construct a collective deviation that would block F robustly on (V, V) if there existed any s with $s(W_{j-1}^+) = \bar{s}_j$ and $Q_F(s) \neq j$. Alternatively, if $\hat{Q}_F(\bar{s}_j) = j - 1$, the arguments in the proofs of Lemmas 3 and 5 could be used to construct a collective deviation that would block F robustly on (V, V) if there existed any s with $s(W_{j-1}^+) = \bar{s}_j$ and $Q_F(s) > j - 1$. We leave the details to the reader. ■

7.2 Discussion

Theorem 4 and Corollary 4 show that the basic conclusions of Theorem 2 and Corollary 2 remain valid when there are multiple public-good provision levels with increasing marginal provision costs and if all provision levels appear in the range of the social choice function when there are only people with extreme payoff types. If the social choice function is required to be immune to robust blocking as well as monotonic, anonymous and robustly implementable, the provision decision can only be conditioned on the population shares of the different interest groups. Implementation can use a set of simple voting mechanisms where, in any one vote, participants are asked whether or not they want the public-good provision level to be increased from j to $j + 1$.

The additional assumption that all provision levels appear in the range of the social choice function when there are only people with extreme payoff types is special but not unreasonable. Social choice functions that condition on a comparison of the aggregate (expected) benefits with the costs of an additional unit of the public good will satisfy the assumption because, as population mass is shifted from the lower end of the payoff type set to the upper end, the aggregate (expected) benefits of an additional unit of the public good move continuously from v_{\min} to v_{\max} .

If the additional assumption about the social choice function is not satisfied, we lose the simple characterization of social choice functions as being implementable by a collection of simple votes on unit increases in the level of public-good provision. However, we do *not* lose the property that, with a requirement of immunity to robust blocking, as well as monotonicity, public-good provision can only be conditioned on the population shares of the different groups.

Rather than stating and proving this claim formally, we illustrate the argument by means of Figure 4, which deals with the case $n = 2$. We now assume that the function

$$\sigma \rightarrow \hat{Q}_F(\sigma) := Q_F(\sigma s_{v_{\max}} + (1 - \sigma)s_{v_{\min}})$$

does not take the value one, and that there exists $\bar{s}_2 \in (0, 1)$ such that $\hat{Q}_F(\sigma) = 0$ if $\sigma < \bar{s}_2$ and $\hat{Q}_F(\sigma) = 2$ if $\sigma > \bar{s}_2$. For specificity, we also assume that $\hat{Q}_F(\bar{s}_2) = 2$.⁴² By arguments similar to those given in the proof of Lemmas 2 and 6 it follows that, if the social choice function is immune to robust blocking as well as monotonic, we must have $Q_F(s) = 2$ whenever $s(W_1^+) = s(V_2) \geq \bar{s}_2$ and $Q_F(s) = 0$ whenever $s(W_1^-) = s(V_0) > \bar{s}_0$, where, with an abuse of notation, we now define $\bar{s}_0 := 1 - \bar{s}_2$. By arguments similar to those given in the proofs of Lemmas 3 - 5, we also find that, if $Q_F(s) = 1$ for some $s \in \mathcal{M}^{na}(V)$, then we must also have $Q_F(s')$ for all $s' \in \mathcal{M}^{na}(V)$ such that $s'(V_2) \leq s(V_2)$ and $s'(V_0) \leq s(V_0)$. Hence, the (closure of the) set of $s \in \mathcal{M}^{na}(V)$ such that $Q_F(s) = 1$ must correspond to a rectangle in Figure 4. If no such point exists, we are back in the world of Theorem 2 and Corollary 2. If such a point exists and the “northeast” vertex of the rectangle corresponds to the point (\bar{s}_0, \bar{s}_2) in the figure, we are back in the world of Theorem 4 and Corollary 4. However, we cannot rule out the possibility that, as shown in the figure, the “northeast” vertex (\hat{s}_0, \hat{s}_2) of the rectangle might lie in the interior of the set (35). In this case, the monotonicity of F implies that $Q_F(s) = 2$ if $s(V_2) \in (\hat{s}_2, \bar{s}_2]$ and $s(V_0) < \hat{s}_0$ and that $Q_F(s) = 0$ if $s(V_2) < \hat{s}_2$ and $s(V_0) \in (\hat{s}_0, \bar{s}_0]$.⁴³

This leaves the rectangle with “southwest” vertex (\hat{s}_0, \hat{s}_2) and “northeast” vertex (\bar{s}_0, \bar{s}_2) . Immunity to robust blocking implies that, in this rectangle, the choice between the provision levels zero and two can only depend on whether or not the population share $s(V_2 \cup V_{12})$ of the set of people who prefer $Q = 2$ over $Q = 0$ meets (or exceeds) a certain threshold or not, or equivalently, on whether the population share $s(V_0 \cup V_{10})$ of the set of people who prefer $Q = 0$ over $Q = 2$ meets (or exceeds) a certain threshold or not. Any reliance on intensities of preferences is ruled out by the requirement of immunity to robust blocking because, with a monotonic social choice function, people with payoff types in V_0 and V_2 can never lose and possibly gain by reporting extreme valuations, v_{\min} or v_{\max} , whereas people with payoff types in V_{12} who prefer $Q = 1$ over $Q = 2$ and $Q = 2$ over $Q = 0$ cannot lose but possibly gain by reporting the highest possible valuation in V_1 .⁴⁴

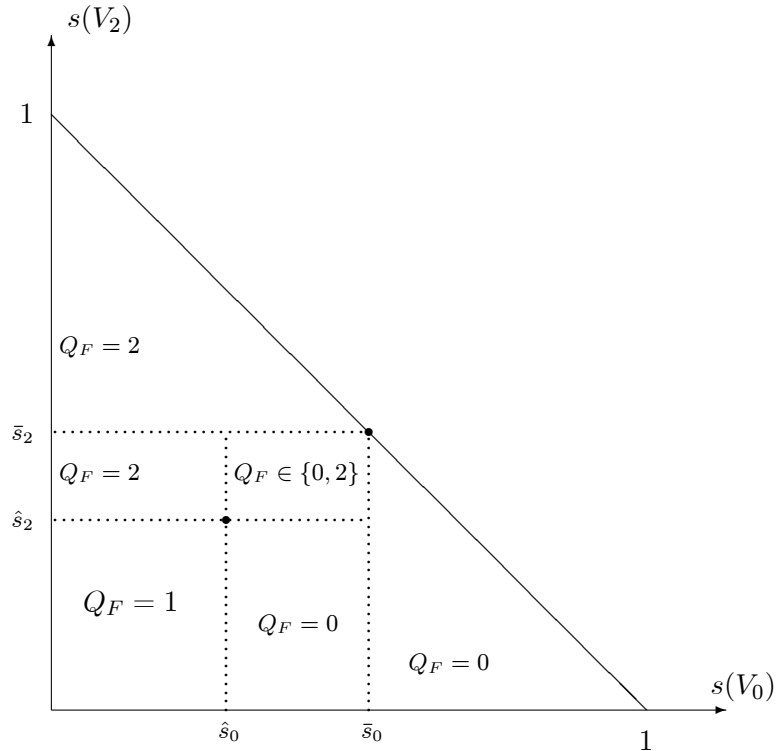
To conclude this section, we note that, if a monotonic and anonymous social choice function with equal cost sharing can be implemented by a voting mechanism, then it is also immune to

⁴²If $\hat{Q}_F(\bar{s}_2)$ were equal to one, we would be back in the setting of Theorem 4. If $\hat{Q}_F(\bar{s}_2)$ were equal to zero, the analysis would be similar to the one we give.

⁴³Figure 4 illustrates the difference between the requirements of immunity to robust blocking, and *ex post* coalition proofness (or robust coalition proofness). The outcomes for $s(V_2) \in (\hat{s}_2, \bar{s}_2]$ and $s(V_0) < \hat{s}_0$, on the one hand, and for $s(V_2) < \hat{s}_2$ and $s(V_0) \in (\hat{s}_0, \bar{s}_0]$, on the other hand, violate *ex post* coalition-proofness. For example, if $Q_F(s) = 2$ for $s(V_2) \in (\hat{s}_2, \bar{s}_2]$ and $s(V_0) < \hat{s}_0$ and the set of people with payoff types in V_{10} , i.e. the set of people who prefer $Q = 1$ over $Q = 0$ and $Q = 0$ over $Q = 2$ is sufficiently large, a coalition of people with payoff types in $V_0 \cup V_{10}$ can induce $Q = 0$ rather than $Q = 2$ by reporting that all coalition members have payoff types in V_0 . Thus, if all coalition members know that $s(V_2) \in (\hat{s}_2, \bar{s}_2]$ and $s(V_0) < \hat{s}_0$, as they do *ex post*, or in the case of a complete-information belief system with degenerate beliefs concentrated at s , this coalition can block the social choice function. Our concept of robust blocking, however, requires coalitions to do without such information. If people with payoff types in V_{10} do not know $s(V_2)$ (or are unable to infer $s(V_2)$ from the information available to the coalition $V_0 \cup V_{10}$), they must be afraid that $s(V_2) < \hat{s}_2$, in which case the coalition members’ all reporting payoff types in V_0 would move the provision level from $Q = 1$ to $Q = 0$ and make the agents with payoff types in V_{10} worse off.

⁴⁴The words “highest” and “lowest” in this sentence are not to be taken literally but to be read as shorthand for suitable approximations of suprema and infima.

Figure 4: Three provision levels: $\hat{Q}_F(\sigma) \neq 1$, for all σ .



robust blocking. The proof is similar to the proof of Theorem 3 and is left to the reader.

8 Concluding Remarks

Our paper has four messages. First, it is important to study public-good provision in large economies where any one individual in isolation is too insignificant to affect the level of public-good provision. In societies with millions of participants, important public-good provision problems are best understood from a large-economy perspective. The large-economy paradigm provides for important simplifications. In particular, individuals do not ever see themselves as being pivotal, and therefore considerations of individual incentive compatibility are trivial. In a society with a million people, the view that individuals are never pivotal is not quite correct, but the probability of being pivotal is on the order of 10^{-4} , an order of magnitude that is unlikely to make much of a difference to people's behaviors.⁴⁵

In a large economy, most of the issues that have been studied in the context of finite economies are moot. Efficient provision of a public good is obtained by asking people how much the public good is worth to them, providing it if the aggregate benefits exceed the costs and sharing the costs equally among the participants. In the absence of participation constraints, this mechanism

⁴⁵As is well known, in a system with n participants, the probability of being pivotal is on the order of $n^{-\frac{1}{2}}$. See the discussion in Hellwig (2003) and the references given there.

implements first-best outcomes robustly. It also satisfies budget balance, i.e. the mechanism designer does not run a surplus or a deficit.

Second, whereas most of the theoretical literature on public-good provision has focused on individual incentives, we argue that coalition proofness should also be a major concern. The scheme that is used for first-best implementation may rely on information from people who are hurt by the use that is made of it. These people provide this information because, as individuals, they feel that their reports do not make a difference to the outcome. A requirement of coalition proofness takes account of their collective interest (without violating individual incentive compatibility).

Coalition proofness matters in large economies as well as small. In the literature, issues of coalition proofness in public-good provision have mainly been tied to the failure of budget balance in dominant-strategy or robust Bayesian implementation. In large economies, these concerns are moot, like the concerns about individual incentive compatibility and about budget balance. However, concerns about collective misrepresentations of preference intensities matter all along and impose significant constraints on mechanism design. In large economies, these concerns matter even if the coalitions themselves must satisfy a robustness condition, which prevents them from conditioning on people's beliefs about each other.

Third, it is desirable to impose a robustness requirement on blocking coalitions as well as the social choice functions. Organizers of deviating coalitions cannot be presumed to know the participants's beliefs. Nor can they be presumed to know the characteristics of non-members of their coalitions. Requirements of robust coalition proofness or ex post coalition proofness are therefore too strong. We replace them with a requirement of immunity to robust blocking.

Fourth, if social choice functions are monotonic and if they must be immune to robust blocking, then mechanism design is confined to the use of voting mechanisms, i.e. public-good provision can only be conditioned on the population shares of people favouring the different alternatives. Economists tend to criticize the use of voting mechanisms because they fail to take account of the intensities of preferences. Our analysis shows that such neglect of preference intensities is unavoidable if coalitions of participants can collectively misrepresent their preferences. For example, a coalition of people for whom the benefits from the public good are smaller than the costs per person could prevent the implementation of first-best outcomes by having all coalition members report that the public good is worth nothing to them, perhaps even that they are harmed by it. Or a coalition of people for whom the benefits exceed the costs might coordinate on exaggerating the benefits they report. Haven't we all heard or seen such cheap-talk exaggerations in media discussion? Given the scope for collective misrepresentations of preference, the only source of information about preferences that can be reliably used is in fact in the numbers of votes for and against the installation of the public good.

The focus on numbers of votes for the different alternatives entails a loss of information and a loss of efficiency. These losses are an unavoidable consequence of the fact that first-best implementation is vulnerable to collective misreporting. The challenge now is to obtain a better understanding of the tradeoffs involved in designing second-best mechanisms.

References

- Austen-Smith, D. and Banks, J. (1996). Information Aggregation, Rationality and the Condorcet Jury Theorem. *American Political Science Review*, 90:34–45.
- Bennett, E. and Conn, D. (2010). The Group Incentive Properties of Mechanisms for the Provision of Public Goods. *Public Choice*, 29: 95-102.
- Bergemann, D. and Morris, S. (2005). Robust Mechanism Design. *Econometrica*, 73:1771–1813.
- Bernheim, B., Peleg, B., and Whinston, M. (1986). Coalition-proof Nash equilibria I. Concepts. *Journal of Economic Theory*, 42:1–12.
- Bierbrauer, F. (2009a). A note on Optimal Income Taxation, Public-Goods provision and Robust Mechanism Design. *Journal of Public Economics*, 93:667–670.
- Bierbrauer, F. (2009b). Optimal Income Taxation and Public-Good Provision with Endogenous Interest Groups. *Journal of Public Economic Theory*, 11:311–342.
- 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 Hellwig, M. (2011). Mechanism Design and Voting for Public-Good Provision. Preprint 2011/31, Max Planck Institute for Research on Collective Goods.
- Bierbrauer, F. and Hellwig, M. (2015). Public Goods, Mechanism Design and Voting. Working Paper, Max Planck Institute for Research on Collective Goods.
- Bierbrauer, F. and Hellwig, M. (in preparation). Mechanism Design and the Difference between Public and Private Goods. 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.
- Bierbrauer, F. (2012). Distortionary Taxation and the Free-Rider Problem. *International Tax and Public Finance*, 19:732–752.
- Bierbrauer, F. (2014). Optimal Tax and Expenditure Policy with Aggregate Uncertainty. *American Economic Journal: Microeconomics*, 6:205–257.
- Buchanan, J. and Tullock, G. (1962). *The Calculus of Consent*. University of Michigan Press, Ann Arbor.
- Boylan, R. (1998). Coalition-Proof Implementation. *Journal of Economic Theory*, 82:132–143.
- Börger, T. and Smith, D. (2014). Robust Mechanism Design and Dominant Strategy Voting Rules. *Theoretical Economics*, 9: 339–360.
- Casella, A. (2005). Storable Votes. *Games and Economic Behavior*, 51: 391–419.

- Che, Y. and Kim, J. (2006). Robustly Collusion-Proof Implementation. *Econometrica*, 74:1063–1107.
- 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.
- Crémer, J. (1996). Manipulation by Coalition Under Asymmetric Information: The Case of Groves Mechanisms. *Games and Economic Behavior*, 13:39–73.
- 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.
- Goeree, J. and Zhang, J. (1979). *Electoral Engineering: One Man, One Bid*. Discussion Paper, University of Zurich.
- Green, J. and Laffont, J.-J. (1979 a). *Incentives in Public Decision-Making*. North-Holland Publishing Company.
- Green, J. and Laffont, J.-J. (1979 b). On Coalition Incentive Compatibility. *Review of Economic Studies*, 46: 243-254.
- Groves, T. (1973). Incentives in Teams. *Econometrica*, 41:617–663.
- Guesnerie, R. (1995). *A Contribution to the Pure Theory of Taxation*. Cambridge University Press.
- Hammond, P. (1979). Straightforward Individual Incentive Compatibility in Large Economies. *Review of Economic Studies*, 46:263–282.
- Hammond, P. (1987). Markets as Constraints: Multilateral Incentive Compatibility in Continuum Economies. *Review of Economic Studies*, 54:399–412.
- Güth, W. and Hellwig, M. (1986). The Private Supply of a Public Good. *Journal of Economics*, Supplement 5:121–159.
- Hellwig, M. (2003). Public-good Provision with Many Participants. *Review of Economic Studies*, 70:589–614.
- Hellwig, M. (2011). Incomplete-Information Models of Large Economies with Anonymity: Existence and Uniqueness of Common Priors. Preprint 2011/08, Max Planck Institute for Research on Collective Goods.

- Hindriks, J., and Myles, G. (2006). *Intermediate Public Economics*. MIT Press, Cambridge, MA.
- 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.
- 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.
- Mas-Colell, A. and Vives, X. (1993). Implementation in Economies with a Continuum of Agents *Review of Economic Studies*, 60:613–629.
- Mirrlees, J. (1971). An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies*, 38:175–208.
- Moulin, H. (1999). Incremental Cost Sharing: Characterization by Coalition Strategy-Proofness. *Social Choice and Welfare*, 16:175–208.
- Mehta, A. and Roughgarden, T. and Sundarajan, M. (2009). Beyond Moulin Mechanisms. *Games and Economic Behavior*, 67:125–155.
- Norman, P. (2004). Efficient Mechanisms for Public Goods with Use Exclusion. *Review of Economic Studies*, 71:1163–1188.
- Qiao, L., Sun, Y., and Zhang, Z. (forthcoming). Conditional exact law of large numbers and asymmetric information economies with aggregate uncertainty. *Economic Theory*, DOI 10.1007/s00199-014-0855-6.
- Sun, Y. (2006). The Exact Law of Large Numbers via Fubini extension and Characterization of Insurable Risks. *Journal of Economic Theory*, 126:31–69.