Public-Good Provision in Large Economies 1: Linear Payment and Provision Cost Functions*

Felix J. Bierbrauer† and Martin F. Hellwig‡

March 24, 2016

Abstract

According to the Lindahl-Samuelson rule, a public good should be provided if and only if the aggregate per-capita valuation exceeds the per-capita cost. In a large economy with private information and aggregate uncertainty about valuations, the Lindahl-Samuelson rule with equal cost sharing is shown to be robustly incentive-compatible. However, it is vulnerable to collective deviations. We introduce a concept of robust blocking by collective deviations and show that monotonic, anonymous and robustly incentive-compatible social choice functions are immune against robust blocking if and only if they can be implemented by voting mechanisms.

Keywords: Public-goods, Robust Mechanism Design, Coalition-proofness, Voting mechanisms, Large Economies

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

Incentive problems in public-good provision are usually analysed in models with finitely many participants, in which each individual can have a noticeable impact on aggregate outcomes.[1] With a focus on achieving individual incentive compatibility, the problem is 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 problem 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.[3]

The notion that any one individual can have a noticeable effect on the level of public-good provision makes sense if we think about people in a condominium deciding on how much to spend on gardening and maintenance. However, this notion is not very relevant for studying how a society with millions of people decides on how much to spend on defense or on the judicial system.

In other areas of economics, analyses of decision making and resource allocation in societies with millions of people are usually based on a large-economy paradigm where each individual is too insiginificant to have a noticeable influence on aggregate outcomes. For private goods, this paradigm underlies the theory of competitive equilibrium, where no one agent has the power to affect prices. In public economics, the large-economy paradigm is used to study the impact of taxation when no one person individually has 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. In the area

[1]For typical textbook treatments, see Fudenberg and Tirole (1991), Mas-Colell et al. (1995), or Hindriks and Myles (2006).

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

[3]Some work in public economics and macroeconomics involves public-good provision in large economies; see e.g. Barro (1990), Boadway and Keen (1993), Battaglini and Coate (2008). However, in this work, the benefits from public-provision are assumed to be known, so there is no need to deal with incentive problems in ascertaining people's tastes.

of public economics, the discrepancy between the small-economy approach to public-good provision and the large-economy approach to taxation is particularly vexing because it stands in the way of an integrated welfare analysis of public spending and taxation.

The notion that individual agents are too insignificant to have a noticeable influence on aggregate outcomes is as relevant for public-good provision as it is for the allocation of private goods through 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.

In this paper and in Bierbrauer and Hellwig (2016b), we study public-good provision in a large economy in which no one individual is able to affect the aggregate outcome, i.e., the level of public-good provision. We obtain the following main results:

- First, in a large economy, individual incentive compatibility poses no problems for efficient public-good provision. The Samuelson rule for first-best provision is easily implemented. For example, for a public good that comes as a single indivisible unit, it suffices to ask people what the public good is worth to them and to have the public good provided if and only if the aggregate per-capita valuation exceeds the per-capita cost. If the public good is provided, the participants can share the costs equally, so there is no problem with budget balance.[4] Thus, in a large economy, most of the concerns of the standard textbook treatment of public-good provision are moot.

- Second, there may be substantial reasons for individuals to form coalitions with a view to coordinating messages so as to manipulate overall outcomes. When coalition incentive compatibility is an issue, public-good provision according to the Samuelson rule is typically impossible. For *monotonic* social-choice functions, we show that, to achieve coalition-proofness as well as individual incentive compatibility, it is necessary and sufficient to use a voting mechanism, i.e., a mechanism that conditions on numbers of votes for or against different alternatives, rather than on expressions of

[4]Participation may not be voluntary however. 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.

how much people value a public good.[5] Mechanisms that condition on preference intensities can be manipulated by coalitions.

Economists have traditionally been critical of voting, which does not take account of preference intensities.[6] If a public good is provided because many people vote in favor and few people against, the outcome can still be inefficient if the proponents' benefits are relatively small and they only vote in favor because they do not have do bear the full cost. However, our analysis shows that this criticism may be irrelevant if public-good provision mechanisms must be robustly incentive-compatible and coalition-proof as well as anonymous.[7]

We use a robust Bayesian approach to implementation. At the level of mechanism design, we allow neither the public-good provision rule, nor the participants' payments to depend on the specification of beliefs that people have about each other. At the level of coalition formation, we do not allow collective deviations to depend on the specification of individual beliefs.

In the following, Section 2 gives an overview of related work. Section 3 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 4 presents our formal model and shows that a

[5]Monotonicity requires that the public-good provision level must not move down if the distribution of public-good valuations moves up in the sense of first-order stochastic dominance. With finitely many individuals, monotonicity is implied by individual incentive compatibility. In a large economy, it is implied by the notion that the continuum of agents is merely an idealization of a large, but finite set of agents.

[6]Thus, 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.

[7]When 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.

first-best public-good provision rule with equal cost sharing is robustly incentive compatible. The requirement of coalition-proofness is formally introduced in Section 4. Sections 5 and 6 introduce the notion of immunity to robust blocking and show that an anonymous, monotonic, robustly incentive compatible social choice function has this property if and only if it can be implemented by a voting mechanism. Further discussions of modelling choices and extensions are given in the Supplementary Material.

Robust implementability and immunity to robust blocking imply that an individual's payments are independent of the individual's valuation for the public good and depend only on the level at which th epublic good is provided. In this paper, we assume that the function relating provision levels to payments is linear, an assumption that follows from equal cost sharing if the marginal cost of public-good provision is constant. In a sequel to this paper, Bierbrauer and Hellwig (2016b), we allow for nonlinear payment and cost functions, studying in particular the case of increasing marginal costs when payment functions provide for equal cost sharing. The case of constant marginal costs is special in that, with equal cost sharing, only three coalitions need to be considered, the grand coalition of all participants, the coalition of agents who like and the coalition of agents who dislike increases in the level of public-good provision. With nonlinear payment functions, the set of relevant coalitions is much larger. We nevertheless confirm the principle that, if an anonymous and monotonic social choice function is to achieve immunity against robust blocking as well as robust incentive compatibility, it must condition on numbers of votes rather than preference intensities.

2 Related work - the concept of coalition-proofness

The concept of a public good, which many people can enjoy without rivalry, was introduced by Lindahl (1919) and Samuelson (1954), who gave conditions for efficient provision. Whereas Lindahl introduced the concept to provide a contractarian foundation for government activities, Samuelson emphasized the incentive problems associated with public-good provision and doubted that efficient outcomes would be implementable by contracting.

Since the 1970s, these incentive problems have been the subject of a large literature on mechanism design for public-good provision. For models with finitely many participants, Clarke (1971), Groves (1973), and Green and Laffont (1977) characterized direct mechanisms that would implement the Samuelson rule for first-best provision in dominant strategies. These mechanisms can be designed so that the mechanism never runs a deficit but they cannot be designed to achieve budget balance, see Clarke (1971), Green and Laffont (1979b). First-best implementation with budget balance can be achieved in a Bayesian framework with independent private values, see d'Aspremont and Gérard-Varet (1979).[8] However, the mechanisms used in Bayesian implementation are highly dependent on the specification of the participants' beliefs and are thus incompatible with the postulate of Ledyard (1979) or Bergemann and Morris (2005) that mechanisms should be robust, i.e. independent of the specification of beliefs, which the mechanism designer cannot be presumed to know.

Bennett and Conn (1977), Green and Laffont (1979 a), and Crémer (1996) showed that Groves mechanisms are vulnerable to certain deviations by coalitions of agents acting together and are therefore not coalition-proof.[9] The failures of coalition-proofness in these papers are closely linked to situations where some agents are pivotal and incentive payments add up to more than the cost of public-good provision, so that the mechanism generates a budget surplus. As we explain in the next section, this type of failure of coalition-proofness becomes less likely when there are many agents and plays no role at all in a large economy where nobody is ever pivotal.

However, we also identify a second type of failure of coalition-proofness that occurs even when no agent is pivotal and a Groves mechanism requires equal payments from all agents. For suppose that there are people whose benefits from the public good, while positive, are smaller than the payments they have to make. Also suppose that, when taken as a whole, this set of people is large enough for their reports to affect the overall outcome. Then, coalition-proofness fails because these individuals can benefit from jointly understating their valuations of the public good. Unlike the failures of coalition-proofness

[8]However, it is not possible to achieve both budget balance and interim individual rationality; see Güth and Hellwig (1986), and Mailath and Postlewaite (1990).

[9]There also is an extensive literature in social choice theory on the implications of having coalitions coordinate their members' behaviors in order to manipulate overall outcomes. For examples see Barberà (1979), Section 4.5 in Dasgupta et al. (1979) and Moulin (1980).

discussed in the literature, this type of failure is present even in a large economy.

The formalization of coalition-proofness in Bennett and Conn (1977), Green and Laffont (1979 a), and Crémer (1996) requires that truthtelling be a dominant strategy for coalitions as well as individuals, with a Pareto criterion for assessing the advantages of deviations for coalition members. The dominant-strategy approach pays no attention to the information and incentive problems that might prevent a deviating coalition from being effective. Some concerns of this type have been considered in subsequent work. Thus, for normal-form games of complete information, Bernheim et al. (1986) introduce a concept of coalition-proof Nash equilibrium, in which coalitions are required to be themselves immune to deviations by sub-coalitions (which in turn must be immune...). In a framework involving Bayes-Nash implementation, Laffont and Martimort (1997, 2000) and Che and Kim (2006) study "collusion-proofness" with particular attention to the incentives that coalition members have to divulge their private information to "the coalition" and to do what "the coalition" asks them to do.

We follow the early literature in imposing coalition-proofness axiomatically, without considering coalition formation and communication between coalition members as part of a strategic game. Unlike the early literature, however, we use a Bayesian, rather than a dominant-strategy, approach to implementation. We do so because the Bayesian approach is explicit about the role of beliefs and information in decision making, at the level of coalition formation as well as individual behavior.

However, we share the view, expressed by Ledyard (1978) and Bergemann and Morris (2005), that mechanism design should not condition on individual beliefs. This is why we impose the Bergemann and Morris (2005) condition of robustness.[10]

We impose this condition on potential collective deviations as well as the overall mechanism. Collective deviation must be designed without any knowledge about the actual distribution of the participants' valuations and without any knowledge about the participants' beliefs. We require that, regardless of their beliefs, (almost) all prospective

[10]Börgers and Smith (2014) suggest that this condition might be too strong. Allowing for indirect mechanisms, they suggest that the arguments of Ledyard (1978) and Bergemann and Morris (2005) support a requirement of independence of the incentive mechanism from the specification of the belief system, but not a requirement of independence of equilibrium outcomes from the specification of the belief system.

members of a coalition expect to benefit from the collective deviation. A collective deviation that satisfies this condition is said to block the social choice function robustly. If no such deviation exists, the social choice function is said to immune against robust blocking.

Immunity against robust blocking is a weak form of coalition-proofness. This requirement neglects all collective deviations by coalitions under which coalition members are made better off in some states of the economy and worse off in others, and people join the coalition because they assign more weight to the states in which the collective deviation makes them better off. In contrast, the requirement of robust coalition-proofness that we used in Bierbrauer and Hellwig (2016a) requires that, for any belief system, a social function must be immune against blocking by any coalition with a collective deviation from which all coalition members expect to benefit given the beliefs that they have. Under this requirement, collective deviations can be conditioned on the belief system.

The analysis in Bierbrauer and Hellwig (2016a) makes much use of "complete information belief systems", i.e., belief systems that are based on degenerate common priors. With such a belief system, every participant may be taken to know the public-good valuations of the other participants so a coalition considering a collective deviation knows what distribution of reported valuations will be sent to the overall mechanism, and there is no uncertainty about the outcome the deviation generates.

By contrast, if the collective deviation cannot be conditioned on the belief system, coalition design is hampered by the lack of information about the people who do not join the coalition. This lack of information leaves coalition members with some uncertainty about the effects of the collective deviation, which depend not only on the manipulated reports of the coalition members but also on the reports of others that the coalition members do not know. Because of this uncertainty, the arguments used in Bierbrauer and Hellwig (2016a) are not available here.

The characterization of social choice functions that are immune to robust blocking that we obtain in this paper is similar to the characterization of robustly coalition-proof social choice functions in Bierbrauer and Hellwig (2016a), namely in both settings, coalition-proofness is obtained if and only if the social choice function under consideration can be implemented by voting mechanisms.[11] Because the coalition-proofness condition in this paper is weaker, the "only if" part of this characterization is more demanding and requires a new argument.[12] For this result, in fact, we need an additional restriction that the social choice function be monotonic in the sense that a first-order increase in the distribution of valuations does not make the public-good provision level go down. This monotonicty condition serves to mitigate the effects of a coalition's being uncertain about the implications of a collective deviation. Thus, with monotonicity, a coalition of people who dislike public-good provision know that they cannot hurt themselves if they report valuations of zero even though their true valuations might actually be positive.

Another difference between this paper and Bierbrauer and Hellwig (2016a) concerns the number of participants. Whereas Bierbrauer and Hellwig (2016a) considers economies with finitely many participants, our analysis in this paper studies economies with a continuum of agents, in which no one individual is ever pivotal. In such economies, robust incentive compatibility implies that agents' payments must not depend on what they report about their public-good valuations. With anonymity and immunity to robust blocking, these payments must be the same for all agents and can only depend on the level of public-good provision. Together with monotonicity, this simplification, which does not available in finite economies, greatly facilitates our analysis here.

The large-economy approach to mechanism design has been pioneered by Hammond (1979, 1987), Mas-Colell and Vives (1993) and Guesnerie (1995), who pointed out the simplifications that are obtained, relative to the case of a large, but finite economy. 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 and therefore about the desired level of public-good provision.

Previous papers by Bierbrauer (2009b, 2014) and Bierbrauer and Sahm (2010) explored the interaction of optimal taxation and public-goods provision in large economies in which

[11]For economies with finitely many agents having single-peaked preferences over a linearly ordered set of alternatives, already Moulin (1980) shows that the median-voter mechanism, i.e. majority voting is "group strategy-proof" as well as individually strategy-proof, i.e., dominant-strategy implementable.

[12]The "if" part is of course less demanding. In Bierbrauer and Hellwig (2016a), this part of the characterization result actually presumes that the requirement of coalition-proofness is limited to coalitions that are themselves sub-coalition-proof in the sense of Bernheim et al. (1986).

individuals differ not only in their preferences for public goods, but also in their productive abilities, as in the theory of optimal income taxation. Bierbrauer (2009a) notes that, in a large economy, the requirement of robust incentive compatibility has no bite because nobody is ever pivotal. In contrast, Bierbrauer (2009b, 2014) finds that a requirement of robust coalition-proofness has bite even in a large economy and considers some of its implications for public-good provision and taxation. Bierbrauer and Sahm (2010) study voting on public-good provision. None of these papers considers the relation between coalition-proofness requirements and voting. Building on the results of this paper, in future work we plan to study this relation in an integrated model of public-good provision and income taxation.

3 Why coalition-proofness? Why Robustness?

We use an example to explain why public-goods analysis should not only involve a requirement of individual incentive compatibility but also a requirement of coalition-proofness. In the example, the public good comes as a single indivisible unit, which can be provided or not. The per-capita cost of providing it 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 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.[13] The public

[13]For a formal treatment of the law of large numbers in a model with a continuum of agents, see Sun

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.

But why should people with a valuation of 3 report this valuation honestly? Reporting a valuation of 3 contributes to making provision of the public good more likely, if only infinitesimally. If the public good is provided, these people enjoy a benefit of 3 and have to pay 4 for a net payoff equal to -1. Each one of them would be better off if the public good was not provided. Moreover, the public good would indeed not be provided if each one of these people reported a valuation of 0. Why, then, should they report honestly, rather than claiming that the public good is worth nothing to them?

If individual incentive compatibility is the only requirement for the public-good provision mechanism, the answer to this question is that nobody minds reporting his or her valuation honestly because nobody feels that his or her report will make a difference to anything anyway. We 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. In political economy, such consid-

(2006) and Qiao et al. (2014).

erations are routinely used to justify the assumption of *sincere voting*, i.e. people voting to promote the alternative they prefer even though, as individuals, they do not expect their votes to have an effect on aggregate outcomes.[14] In our example, such behavior would lead people with public-good valuation 3 to report the valuation 0, thereby preventing an implementation of the efficient provision rule with equal cost sharing.

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.

Which of these objections 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.

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 consider a version of our example with a finite number n of agents. As before, 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 $3S_3 + 10S_{10} = 4n$.

[14]For instance, this notion underlies the Downsian analysis of political competition where person i is assumed to vote for politician 1 or party 1 if and only if i prefers the policy proposed by politician 1 or party 1 over the policy proposed by politician 2 or party 2.

Dominant-Strategies and Coalition-Proofness in Finite Economies. 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.[15] Table 1 describes the special case of a Clarke-Groves mechanism for our example with an arbitrary but finite number n of individuals. The table shows how the payment p_i of agent i depends on the agent's reported valuation $\hat{v}_i$ and the reported aggregate per-capita valuation $\hat{\bar{v}}$. If $\hat{S}_3$ people report the valuation 3 and $\hat{S}_{10}$ people the valuation 10, $\hat{\bar{v}} = (3\hat{S}_3 + 10\hat{S}_{10})/n$. The mechanism also stipulates that the public good is provided if $\hat{\bar{v}} \geq 4$ and is not provided if $\hat{\bar{v}} < 4$.

Table 1.

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

The table exhibits four cases, no payments and no provision if $\hat{\bar{v}}$ is very low; payments by agents with valuation zero and no provision if $\hat{\bar{v}}$ is close to but less than 4; provision with extra payments by agents with valuation 10 if $\hat{\bar{v}}$ is close to but not less than 4; finally, provision with equal cost sharing if $\hat{\bar{v}}$ is very high. If everyone tells the truth, the reported aggregate per-capita valuation $\hat{\bar{v}}$ coincides with the actual aggregate per-capita valuation $\bar{v} = (3S_3 + 10S_{10})/n$ and the provision decision is efficient.

Truth-telling is a dominant strategy, and moreover, whenever an individual is pivotal for public-goods provision, lying to change the aggregate outcome is strictly inferior to truthtelling. For example, if an agent with the true valuation $v_i = 3$ is honest when the reported average valuation $\hat{\bar{v}}$ is $4+\frac{1}{n}$, the public good is provided, the agent pays 4 and has a net payoff equal to -1. If the agent lies and reports the valuation $\hat{v}_i = 0$, the reported average valuation is changed to $\hat{\bar{v}} = 4 - \frac{2}{n}$, the public good is not provided, and the agent pays $n(\hat{\bar{v}} - 4) + 4 = 2$, for a net payoff equal to -2. In this instance, the agent is pivotal and, in contrast to the large-economy model, has a strict preference for truthtelling. In

[15] See Clarke (1971) and Groves (1973).

other instances, where he is not pivotal, the agent is simply indifferent whether he reports $\hat{v}_i = v_i = 3$ or not. Hence, "sincere voting", in the sense of communicating a valuation of 0 whenever the true valuation is 3, is not incentive compatible.

However, there is a problem with coalition-proofness. Coalition-proofness can actually fail in two ways. First, if $4 - \frac{4}{n} < \bar{v} < 4$ 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 $4 - \frac{4}{n} < \bar{v} < 4$) 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 (1979b). The second type of failure does not seem to have been noted before and is an exact analogue of the failure of coalition-proofness in the continuum model.

Bayes-Nash Implementation, Coalition-Proofness, and Robustness. The large-economy approach is based on the notion that individual agents are too insignificant to affect aggregate outcomes. In finite economies, however, there always are states in which some individual is pivotal. To relate the large-economy and finite-economy models, the set of such states must be considered to be in some sense unimportant when there are many participants. The dominant-strategy approach has no room for such an assessment. The Bayesian approach does.

To see this, recall our 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 $.7 - \alpha, \alpha,$ 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 $4 - \frac{4}{n}$ with a probability close to one. In both cases, if $\alpha > \frac{1}{3}$ and if $\alpha < \frac{1}{3}$, the events $4 - \frac{4}{n} < \bar{v} < 4$ 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 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 $4 - \frac{4}{n} < \bar{v} < 4$ 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_is from their means are positive or negative.

Most papers on public-good provision in Bayesian models with finitely many agents assume independent private values. In our example, this corresponds to the case where α is fixed and commonly known. In a large economy, this case is not so interesting because, if α is known, it is also clear whether the public good should be provided or not, i.e., efficient public-good provision does not require any information from participants.

Efficient public-good provision does require information from participants if α is the realization of a nondegenerate random variable $\tilde{\alpha}$. In this case we have correlated values as each agent's valuation is correlated with $\tilde{\alpha}$, and therefore the different agents' valuations are correlated with each other.

If agents' valuations are correlated with $\tilde{\alpha}$ their beliefs about $\tilde{\alpha}$ will vary with their valuations. A high value of $\tilde{\alpha}$ makes a valuation of 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 $\tilde{\alpha}$.

As was observed by Crémer and McLean (1988), such differences in beliefs induce differences in preferences over lotteries whose outcomes depend on the value of $\tilde{\alpha}$. These differences in preferences over lotteries can be exploited to provide for a coalition-proof, Bayesian-incentive-compatible implementation of first-best provision rules.[16] However, theses schemes are very sensitive to the precise specification of how the agents' beliefs depend on their valuations or, in our example, the specification of the distribution of $\tilde{\alpha}$.

Following the arguments of Ledyard (1978) and Bergemann and Morris (2005), we consider it unreasonable to suppose that a mechanism designer has the requisite information about the participants' beliefs that is needed to implement the appropriate Crémer-McLean payment scheme. With a robustness requirement, i.e., a requirement that the public-good provision and payment rules must not be conditioned on the specification of belief systems, the possibility of using Crémer-McLean payment schemes to achieve

[16]For an example, see the Supplementary Material for this paper.

coalition-proofness is eliminated.

Because robust Bayes-Nash and dominant-strategy implementability requirements impose the same restrictions on the design of incentive mechanisms,[17] our findings about the incompatibility of coalition-proofness with first-best implementation in dominant strategies apply to robust Bayes-Nash implementation as well. However, the Bayes-Nash framework gives rise to the additional observation that failures of coalition-proofness related to individuals' being pivotal, such as those identified by Bennett and Conn (1977) and Green and Laffont (1979), are unlikely to matter when there are many agents. By contrast, failures of coalition-proofness related to large interest groups' being able to manipulate provision levels are relevant in large economies as well as small.

In the following, we study the implications of coalition-proofness for robust Bayes-Nash implementation. Relative to the dominant-strategy approach, the Bayes-Nash approach has three advantages. First, 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. Second, as mentioned above, it gives meaning to the assertion that, if the number of participants is large, the set of states in which one agent is pivotal is relatively unimportant. Third, it allows for a richer discussion of coalition-proofness, in which the information available to a coalition is restricted. As far as we know, the notion of robustly blocking coalitions that we propose has no counterpart in a dominant-strategy framework.[18]

[17]For environments with private values, Corollary 1 in Bergemann and Morris (2005) shows that robust Bayes-Nash imlementability and dominant strategy implementability of a social choice function, as opposed to a social choice correspondence, are equivalent.

[18]As developed by Bennett and Conn (1977), Green and Laffont (1979b), and Moulin (1999), in a dominant strategy setting, coalition-proofness 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 equivalent to 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.

4 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 installation of Q units of the public good requires aggregate resources (per-capita) equal to kQ units of the private good. The provision level Q is bounded above by some number $\bar{Q}$. Given a public-good provision level Q, 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 lies in the interior of V.

A social choice function determines how the provision level of the public good and the payments of the different individuals depend on the state of the economy. We identify the state of the economy with the cross-section distribution of public-good valuations. Following Guesnerie (1995), we impose an anonymity requirement by which the level of public-good provision and 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 makes the public-good provision level and the payment rule depend on the cross-section distribution of valuations.

Formally, the state of the economy is an element s of the set $\mathcal{M}(V)$ of 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, $Q_F(s) \in [0, \bar{Q}]$ 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.[19]

[19]Anonymity is a substantive constraint. Using the idea of sampling, that has been developed by Green and Laffont (1979a), Bierbrauer and Sahm (2010) show that first-best outcomes can be implemented by a

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

$$\int_V P_F(s,v)ds(v) \geq kQ_F(s) \tag{1}$$

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 nondecrasing 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. A typical element of $\mathcal{M}(T)$ will be denoted by δ. For any event $X \subset \mathcal{M}(T)$, $\beta(X \mid t_i)$ is the probability that type t_i of agent i assigns to the event that the cross-section distribution of types δ belongs to the set X. We refer to the map $\beta : T \to \mathcal{M}(\mathcal{M}(T))$ as the *belief system* of the economy.[20]

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.

[20]We do not assume that the belief system is compatible with a common prior. By the arguments in Bierbrauer and Hellwig (2010), our analysis would be unchanged if we restricted ourselves to belief

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 $s(V' \mid \delta)$ for the mass of individuals that the distribution $s(\delta)$ assigns to payoff types in V'.

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 type distribution δ^* is *admissible* if and only if the associated payoff type distribution $s(\delta^*)$ belongs to the set $\mathcal{M}^{na}(V)$ of measures on V that do not have atoms. We also say that a belief system $\beta : T \to \mathcal{M}(\mathcal{M}(T))$ is *admissible* if and only if for all $t \in T$, the induced belief function satisfies $\beta(\mathcal{M}^*(T))| \; t) = 1$, where $\mathcal{M}^*(T) := \{\delta^* | \; s^{-1}(\delta^*) \in \mathcal{M}^{na}(V)\}$ is the set of admissible measures on T. 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. In this special case, we obviously have $\mathcal{M}^*(V) = \mathcal{M}^{na}(V)$, i.e., a cross-section distribution of payoff types is admissible if and only if it is atomless.

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 $t, t' \in T$

$$U(t \mid t) \geq U(t \mid t') \; , \tag{2}$$

where

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

systems that are compatible with common priors. Hellwig (2011) gives conditions for the existence and uniqueness of common priors in our setup.

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 $\tau : T \to V$, the inequalities in (2) hold for every admissible belief system β. Adapting an argument of Bergemann and Morris (2005), the following proposition provides a characterization of robustly implementable social choice functions in the present setup.[21]

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 and v' in V and all $s \in \mathcal{M}(V)$,*

$$vQ_F(s) - P_F(v, s) \geq vQ_F(s) - P_F(v', s) \ . \tag{3}$$

Robust incentive compatibility is thus equivalent to *ex post incentive compatibility*: once s has become known, no individual regrets having revealed his type to the mechanism. By inspection of (3), in our setting, *ex post* incentive compatibility in turn 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. Thus we obtain:

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 $\bar{P}_F : \mathcal{M}(V) \to \mathbb{R}$ such that P_F takes the form $P_F(v, s) = \bar{P}_F(s)$ for all $v \in V$ and all $s \in \mathcal{M}(V)$.*

In view of 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 Implementability 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,

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

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 (1). By standard arguments, this requires that the public good should be provided at the maximal level, i.e., that $Q_F(s)$ should be set equal to $\bar{Q}$, if the aggregate valuation $\bar{\mathbf{v}}(s) := \int_V v \; ds(v)$ exceeds the cost k and should not be provided at all, i.e., that $Q_F(s)$ should be set equal to 0, if $\bar{\mathbf{v}}(s)$ is less than k. There also should be no slack in the feasibility constraint, i.e., aggregate payments should exactly cover the cost of public-good provision. This requirement is satisfied by setting $P_F(v, s) = k \; Q_F(s)$ for all s, which provides for equal cost sharing. Since this payment rule satisfies the condition of Corollary 1, we obtain:

Theorem 1 *Any first-best social choice function with equal cost sharing is robustly implementable.*

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 parallels a result of Bierbrauer (2009a) but 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 violations of budget balance arise from the need to provide proper incentives to individuals when they are pivotal. In a large economy, this concern is moot because no agent is ever pivotal.

5 Coalition-proofness

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 Section 3, we consider robust incentive compatibility to be too weak a requirement to do full justice to the information and incentive problems of public-good provision in such an economy. In the following, we therefore impose an additional requirement of coalition-proofness.

Following Bennett and Conn (1977), Green and Laffont (1979 a), and Crémer (1996), we abstract from all the details of how a coalition might be formed and simply impose the constraint that there must not exist a group of agents who would all benefit from a false communication of preferences. However, whereas the dominant-strategy approach followed by these authors involves a concept of coalition-proofness ex post, we apply the concept ex interim, with agents evaluating coalition outcomes in terms of their type-dependent beliefs. Moreover, we impose a robustness requirement on the desirability of collective deviations.

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, most of these concerns are moot. For a detailed discussion, we refer to the Supplementary Material for this paper.

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.

Given an abstract space $(T, \mathcal{T})$, a payoff type function $\tau : T \to V$ and an anonymous, robustly implementable social choice function $F = (Q_F, \bar{P}_F)$, we specify a collective deviation as a pair $(T', \ell_{T'})$ such that T' is a subset of T and $\ell_{T'}$ is a reporting strategy (a "lying" strategy) for people with types in T'.[22] Formally, $\ell_{T'}$ is a function from the set

[22]We neglect the possibility of side payments between coalition members. In the Supplementary Ma-

$\hat{\mathcal{M}}(T')$ of measures on T' that have total measure less than or equal to one to the set $\mathcal{M}(T)$ of probability measures on T. We think of $\ell_{T'}(\delta_{T'})$ as a lottery that determines the report or "lie" that any one individual with a type in T' sends to the overall mechanism when the distribution of types in T' is $\delta_{T'}$. Thus, for any measurable set $\tilde{T} \subset T$, $\ell_{T'}(\tilde{T} \mid \delta_{T'})$ is the probability that the agent's report belongs to $\tilde{T}$. [23] We assume that, by a law of large numbers, $\ell_{T'}(\tilde{T} \mid \delta_{T'})$ is also the share of reports from coalition members that lie in $\tilde{T}$.[24] The lies are a function of $\delta_{T'}$, rather than $\delta = (\delta_{T'}, \delta_{T \setminus T'}) \in \hat{\mathcal{M}}(T') \times \hat{\mathcal{M}}(T \setminus T')$, 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

$$\hat{\delta}(\ell_{T'}, \delta) = \delta(T') \cdot \ell_{T'}(\delta_{T'}) + \delta_{T \setminus T'}, \tag{4}$$

where $\delta_{T \setminus T'} \in \hat{\mathcal{M}}(T \setminus T')$ is the (non-normalized) restriction of the true type distribution δ to the non-deviating 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 $(V, \mathcal{V})$, 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. We use the shorthand

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

for the utility realized by an individual with payoff type $\tau(t)$ under manipulation $\ell_{T'}$ if the true distribution of types equals δ. Analogously, we denote by $u(\tau(t), s(\delta))$ the utility that is realized if everybody behaves truthfully.

Given a type space $[(T, \mathcal{T}), \tau, \beta]$, we say that a collective deviation $(T', \ell_{T'})$ *blocks* the social choice function F on $[(T, \mathcal{T}), \tau, \beta]$ if it satisfies the following two conditions:

terial, we show that, in our large-economy model with robustness at the level of coalition design as well as mechanism design, side payments that satisfy the relevant conditions for incentive compatibility, voluntary participation in coalitions, and feasibility must actually be zero.

[23]The 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.

[24]For the measure theoretic foundations of this assumption, see Qiao et al. (2014), Sun (2006).

Regardless of what the actual distribution of types in the economy is, with beliefs given by the belief system β, (almost) no person with a type in T' expects to lose from the collective deviation; formally, for all $\delta^* \in \mathcal{M}^*(T)$ and δ^*-almost all $t' \in T'$,

$$\int_{\mathcal{M}(T)} u(\tau(t'), s(\hat{\delta}(\ell_{T'}, \delta)))\, d\beta(\delta \mid t') \geq \int_{\mathcal{M}(T)} u(\tau(t''), s(\delta))\, d\beta(\delta \mid t') \,. \tag{5}$$

Second, for some actual distribution of types in the economy, there is a non-negligible set of people with types in T' who expect to gain from the collective deviation; formally, there exists some $\delta^* \in \mathcal{M}^*(T)$ and some set $T'' \subset T'$ such that $\delta^*(T'') > 0$ and, for all $t'' \in T''$,

$$\int_{\mathcal{M}(T)} u(\tau(t''), s(\hat{\delta}(\ell_{T'}, \delta)))\, d\beta(\delta \mid t'') > \int_{\mathcal{M}(T)} u(\tau(t''), s(\delta))\, d\beta(\delta \mid t'') \,. \tag{6}$$

Given $(T, \mathcal{T})$ and a payoff type function $\tau : T \to V$, we say that a collective deviation $(T', \ell_{T'})$ blocks the 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 all $t' \in T'$, the belief $\beta(t')$ assigns positive probability to the event that the deviation $(T', \ell_{T'})$ departs from truthtelling. The social choice function F is *immune against robust blocking*, if there is no $[(T, \mathcal{T}), \tau]$ and no collective deviation $(T', \ell_{T'})$ that blocks F robustly.

For naive type spaces, conditions (5) and (6) are much simpler. Given a naive type space $[(V, \mathcal{V}), \beta_s]$, a collective deviation $(V', \ell_{V'})$ *blocks* the social choice function F if, first,

$$\int_{\mathcal{M}(V)} u(v', \hat{s}(\ell_{V'}, s)) d\beta_s(s \mid v') \geq \int_{\mathcal{M}(V)} u(v', s) d\beta_s(s \mid v') \,, \tag{7}$$

for almost all v'', and, second, for some nonnegligible set $V'' \subset V'$, the inequality in (7) is strict.[25]

As mentioned above, immunity against robust blocking is much weaker property than the requirement of robust coalition-proofness that we used in Bierbrauer and Hellwig

[25]Using the fact that V is a subset of the set of real numbers, the terms "almost" and "non-negligible" here are defined in terms of Lebesgue measure. If the inequality (7) holds for Lebesgue almost all $v \in V'$, then it also holds for s^*-almost all $v \in V'$, where s^* is any atomless measure on V. Moreover, if the inequality (7) is strict for a Lebesgue-non-negligible subset of V', then trivially there exists a measure $s^* \in \mathcal{M}^*(V)$ (namely uniform distribution) such that the inequality (7) is strict on an s^*-non-neglligible subset of V'.

(2016a). 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 $[(T,\mathcal{T}),\tau,\beta]$. In this condition, the collective deviation $(T',\ell_{T'})$ is adapted to the particular type space on which it is used. For a belief system β that has all beliefs concentrated on a single cross-section distribution δ, allowing a coalition to condition its deviation on β is equivalent to presuming that the coalition knows the distribution of types among non-members. Immunity against robust blocking involves no such presumption.

6 Coalition-proofness and Voting

In the following, we characterize anonymous and robustly implementable social choice functions that are immune against robust blocking. We begin with the observation that, with anonymity, robust implementability and immunity to robust blocking, payments can only depend on the public-goods provision level, i.e., for any two states s and s', $\bar{P}_F(s)=\bar{P}_F(s')$ if $Q_F(s)=Q_F(s')$.

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

$$P_F(v,s)=P_F(v,s') \quad \textit{if} \quad Q_F(s)=Q_F(s'). \tag{8}$$

Proof. By Corollary 1, the payment rule P_F takes the form $P_F(v,s)=\bar{P}_F(s)$ for all v and s. Suppose that the proposition is false. Then there exist s and s' in $\mathcal{M}^{na}(V)$ such that $Q_F(s)=Q_F(s')$ and $\bar{P}_F(s)<\bar{P}_F(s')$. Consider the naive type space $(V,\mathcal{V})$ and a collective deviation (V,ℓ_V) by the grand coalition of all agents such that $\ell_V(s'')=s$ for $s''\in\{\hat{s}\in\mathcal{M}^{na}(V)\mid Q_F(\hat{s})=Q_F(s)\quad\text{and}\quad\bar{P}_F(\hat{s})>\bar{P}_F(s)\ \}$ and $\ell_V(s'')=s''$, otherwise. Whenever this coalition deviates from truth-telling, it lowers everybody's payment. Thus, this collective deviation blocks F robustly, so F is not immune to robust blocking. ∎

6.1 Immunity to Robust Blocking Requires Voting

Proposition 2 and Corollary 1 imply that the payment rule can be written in the form

$$P_F(v,s)=\Pi_F(Q_F(s)). \tag{9}$$

As we continue to discuss the implications of robust incentive compatibility and immunity to robust blocking, we impose the additional assumption that P_F takes the affine form

$$P_F(v, s) = \pi_F^0 + \pi_F^1 \cdot Q_F(s). \tag{10}$$

This assumption will be relaxed in Bierbrauer and Hellwig (2016b), where we allow for nonlinear payment functions.

Imposing the affine form (10) actually involves no loss of generality if the public good comes as a single indivisible unit so that the choice is only whether to provide it or not. In this case, there are just two payment levels, P_F^0 and P_F^1, and we can write

$$P_F(v, s) = \begin{cases} P_F^0, & \text{if} \quad Q_F(s) = 0 \;, \\ P_F^1, & \text{if} \quad Q_F(s) = 1 \;, \end{cases} \tag{11}$$

which reduces to (10) with $\pi_F^0 = P_F^0$ and $\pi_F^1 = P_F^1 - P_F^0$.

With linear provision costs, the affine form (10) is also implied by equal cost sharing. In this case, $\pi_F^0 = 0$ and $\pi_F^1 = k$.

With a payment rule satisfying (10), an agent with payoff type v achieves the payoff

$$(v - \pi_F^1) \cdot Q_F(s) - \pi_F^0$$

when the state of the economy is s. There are thus two natural interest groups: Agents with valuations in the set $V_0(\pi_F^1) := [v_{\min}, \pi_F^1)$ dislike any increase in Q, agents with valuations in the set $V_1(\pi_F^1) := (\pi_F^1, v_{\max}]$ dislike any decrease in Q. The remaining set of individuals who are indifferent can be neglected. For $s \in \mathcal{M}^{na}(V)$, this set has measure $s(\{\pi_F^1\})$, which is zero if the distribution s is admissible. For any $s \in \mathcal{M}^{na}(V)$, we have

$$s(V_1(\pi_F^1)) + s(V_0(\pi_F^1)) = 1. \tag{12}$$

regardless of the value of π_F^1. To simplify the notation, we will drop the argument π_F^1 in referring to $V_1(\pi_F^1)$ and $V_0(\pi_F^1)$ whenever we can do so without creating confusion.

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

$$s(V_1) \geq s'(V_1) \quad \textit{implies} \quad Q_F(s) \geq Q_F(s'). \tag{13}$$

Corollary 2 *If an anonymous and robustly implementable monotonic social choice function F is immune to robust blocking, then it takes the form of a mechanism that asks people to vote for or against public-good provision and that makes the level of public-good provision an increasing function of the share of "YES"-votes.*

Proof of Theorem 2. Consider any monotonic, anonymous and robustly implementable social choice function F with an affine payment rule that is immune to robust blocking. Consider the collective deviations (V_1, ℓ_{V_1}) and (V_0, ℓ_{V_0}) such that, for all $s \in \mathcal{M}(V)$,

$$\ell_{V_1}(s_{V_1}) = s_{v_{\max}} \quad \text{and} \quad \ell_{V_0}(s_{V_0}) = s_{v_{\min}} \ , \tag{14}$$

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 $v_{\min}$. Under the collective deviation (V_1, ℓ_{V_1}), all coalition members always say that their public-good valuation is $v_{\max}$, and under the collective deviation (V_0, ℓ_{V_0}), all coalition members always say that their public-good valuation is $v_{\min}$. By the monotonicity of Q_F, we have

$$Q_F(\hat{s}(\ell_{V_0}, s)) \leq Q_F(s) \leq Q_F(\hat{s}(\ell_{V_1}, s)) \ , \tag{15}$$

for all s, so the collective deviation (V_1, ℓ_{V_1}) is never harmful for agents with valuations in V_1 and is beneficial for them at any state s at which the right-hand inequality in (15) is strict; similarly, the collective deviation (V_0, ℓ_{V_0}) is never harmful for agents with valuations in V_0 and is beneficial for them at any state s at which the left-hand inequality in (15) is strict. Immunity to robust blocking therefore implies that both inequalities in (15) must hold as equations, i.e., we must have

$$Q_F(\hat{s}(\ell_{V_0}, s)) = Q_F(s) = Q_F(\hat{s}(\ell_{V_n}, s)) \ , \tag{16}$$

for all s. Upon replacing s by $\hat{s}(\ell_{V_0}, s)$ in the right-hand equation in (16), we also obtain

$$Q_F(\hat{s}(\ell_{V_0}, s)) = Q_F(\hat{s}(\ell_{V_n}, \hat{s}(\ell_{V_0}, s))) \ ,$$

and hence

$$Q_F(s) = Q_F(\hat{s}(\ell_{V_n}, \hat{s}(\ell_{V_0}, s))) \ ,$$

for all s. One easily verifies that, for any s,

$$\hat{s}(\ell_{V_n}, \hat{s}(\ell_{V_0}, s)) = s(V_n) \ s_{v_{\max}} + s(V_0) \ s_{v_{\min}}$$

or, for $s \in \mathcal{M}^{na}(V)$,

$$\hat{s}(\ell_{V_n}, \hat{s}(\ell_{V_0}, s)) = s(V_n)\; s_{v_{\max}} + (1 - s(V_n))\; s_{v_{\min}}.$$

For any s, we thus have

$$Q_F(s) = Q_F(s(V_n)\; s_{v_{\max}} + (1 - s(V_n))\; s_{v_{\min}}), \tag{17}$$

By the monotonicity of Q_F, the theorem follows immediately. ■

The 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 preference intensities 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 (17).

Technically, the monotonicity condition ensures that, for any split $(\sigma, 1-\sigma)$ of the population between the sets V_1 and V_0, the pair $(s_{v_{\min}}, s_{v_{\max}})$ is a saddle-point of the function $(s_0, s_1) \to Q_F(\sigma s_1 + (1-\sigma)s_0)$ i.e. that

$$Q_F(\sigma s_{v_{\max}} + (1-\sigma)s_0) \geq Q_F(\sigma s_{v_{\max}} + (1-\sigma)s_{v_{\min}}) \geq Q_F(\sigma s_1 + (1-\sigma)s_{v_{\min}})$$

$(s_0, s_1) \in \mathcal{M}(V)^2$. This property underlies the proof of Theorem 2.[26]

6.2 Voting Mechanisms are Immune to Robust Blocking

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

[26]The 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. In this equilibrium, the organizer of V_1 claims that everybody who belongs to V_1 has the maximal valuation v_{max} and the organizer of V_0 claims that everybody who belongs to V_0 has the valuation v_{min}. A truthful communication of preferences is a Nash equilibrium only if it yields the same outcome as these strategies, i.e., only if (17) holds.

Theorem 3 *If an anonymous, robustly implementable social choice function F with an affine payment rule satisfies condition (13) for all s and s' in $\mathcal{M}^{na}(V)$, then F is immune to robust blocking.*

Proof. Let F be a robustly incentive compatible anonymous social choice function with an affine payment rule. If F satisfies condition (13) there exists a function Q_F^* on $[0,1]$ such that for all s, $Q_F(s) = Q_F^*(s(V_1))$. We will show that F is also immune to robust blocking. For suppose that this claim is false. Then there exist $(T, \mathcal{T})$ and a payoff type function τ such that F is robustly blocked on $[(T, \mathcal{T}), \tau]$. Thus, there exists a collective deviation $(T', \ell_{T'})$ such that, first, for all admissible belief systems β, all admissible type distributions $\delta^* \in \mathcal{M}^*(T)$ and δ^*-almost all $t' \in T'$,

$$\int_{\mathcal{M}(T)} u(\tau(t'), s(\hat{\delta}(\ell_{T'}, \delta)))\, d\beta(\delta \mid t') \geq \int_{\mathcal{M}(T)} u(\tau(t'), s(\delta))\, d\beta(\delta \mid t')\ , \tag{18}$$

and, second, for some admissible belief system β^{**} and some admissible type distribution $\delta^{**} \in \mathcal{M}^*(T)$, we have

$$\int_{\mathcal{M}(T)} u(\tau(t''), s(\hat{\delta}(\ell_{T'}, \delta)))\, d\beta(\delta \mid t'') > \int_{\mathcal{M}(T)} u(\tau(t''), s(\delta))\, d\beta(\delta \mid t'') \tag{19}$$

for all t'' in a set $T'' \subset T'$ that satisfies $\delta^{**}(T'') > 0$.

If we set $\delta^* = \delta^{**}$ in the first condition, we find that (18) holds for all belief systems β and δ^{**}-almost all $t' \in T'$. This is only possible if the inequalities

$$u(\tau(t'), s(\hat{\delta}(\ell_{T'}, \delta))) \geq u(\tau(t'), s(\delta)),$$

and

$$(\tau(t') - \pi_F^1)(Q_F(s(\hat{\delta}(\ell_{T'}, \delta))) - Q_F(s(\delta))) \geq 0 \tag{20}$$

hold for all $\delta \in \mathcal{M}^*(T)$ and δ^{**}-almost all $t' \in T'$. From (19), we can also infer that we must have

$$(\tau(t'') - \pi_F^1)(Q_F(s(\hat{\delta}(\ell_{T'}, \delta))) - Q_F(s(\delta))) > 0\ ,$$

for all $\delta \in \mathcal{M}^*(T)$ and all $t'' \in T''$, where T'' is the specified subset of T'.

Let $V' := \tau(T')$ be the set of the coalition members' payoff types. From (20), we infer that

$$(v' - \pi_F^1)(Q_F(s(\hat{\delta}(\ell_{T'}, \delta))) - Q_F(s(\delta))) \geq 0\ , \tag{21}$$

for all $\delta \in \mathcal{M}^*(T)$ and almost all $v' \in V'$. Moreover, for all $\delta \in \mathcal{M}^*(T)$, the inequality in (21) is strict for all v' in a non-negligible set $V'' \subset V'$.

The strictness of the inequality (21) for $v'' \in V''$ implies that we cannot have $Q_F(s(\hat{\delta}(\ell_{T'}, \delta))) - Q_F(s(\delta)) = 0$ for all $\delta \in \mathcal{M}^*(T)$, i.e., for some $\bar{\delta} \in \mathcal{M}^*(T)$, we must have either $Q_F(s(\hat{\delta}(\ell_{T'}, \bar{\delta}))) - Q_F(s(\bar{\delta})) > 0$ or $Q_F(s(\hat{\delta}(\ell_{T'}, \bar{\delta}))) - Q_F(s(\bar{\delta})) < 0$.

Suppose that the first alternative is true, and let $Q_F(s(\hat{\delta}(\ell_{T'}, \bar{\delta}))) - Q_F(s(\bar{\delta})) > 0$. Then (13) implies that

$$s(V_1|\hat{\delta}(\ell_{T'}, \bar{\delta})) > s(V_1|\bar{\delta}), \tag{22}$$

where again $V_1 := (\pi_F^1, v_{\max}]$ is the set of valuations at which an agent is always in favour of increasing the public-good provision level.

By (21), $Q_F(s(\hat{\delta}(\ell_{T'}, \bar{\delta}))) - Q_F(s(\bar{\delta})) > 0$ also implies $v' \geq \pi_F^1$ for all $v' \in V'$ and hence $V' \subset V_1 \cup \{\pi_F^1\}$ or

$$T' \cap \tau^{-1}(V_1 \cup \{\pi_F^1\}) = T'. \tag{23}$$

Because a coalition cannot make itself larger than it is, we also have

$$s(V_1|\hat{\delta}(\ell_{T'}, \bar{\delta})) \leq \bar{\delta}((T \backslash T') \cap \tau^{-1}(V_1)) + \bar{\delta}(T'). \tag{24}$$

By standard set theory, $(T \backslash T') \cap \tau^{-1}(V_1) = \tau^{-1}(V_1) \backslash T'$, so (23) and (24) yield

$$s(V_1|\hat{\delta}(\ell_{T'}, \bar{\delta})) \leq \bar{\delta}(\tau^{-1}(V_1 \cup \{\pi_F^1\})) = s(V_1 \cup \{\pi_F^1\}|\bar{\delta}). \tag{25}$$

Because $\bar{\delta}$ is admissible, we also have $s(\{\pi_F^1\}|\bar{\delta}) = 0$, so (25) implies

$$s(V_1|\hat{\delta}(\ell_{T'}, \bar{\delta})) \leq s(V_1|\bar{\delta}). \tag{26}$$

Since (26) is obviously incompatible with (22), the assumption that $Q_F(s(\hat{\delta}(\ell_{T'}, \bar{\delta}))) - Q_F(s(\bar{\delta})) > 0$ must be false.

The second alternative, $Q_F(s(\hat{\delta}(\ell_{T'}, \bar{\delta}))) - Q_F(s(\bar{\delta})) < 0$, also leads to a contradiction. The argument is the same but now involves the set $V_0 := [v_{\min}, \pi_F^1)$ of payoff types who are in favor of decreased public-good provision.

In either case, if $Q_F(s(\hat{\delta}(\ell_{T'}, \bar{\delta}))) - Q_F(s(\bar{\delta})) > 0$ and if $Q_F(s(\hat{\delta}(\ell_{T'}, \bar{\delta}))) - Q_F(s(\bar{\delta})) < 0$, the assumption that the collective deviation $(T', \ell_{T'})$ blocks F robustly on $[(T, \mathcal{T}), \tau]$ leads to a contradiction and must be false. Hence F is immune to robust blocking. ■

In the proof of Theorem 3, in moving from (25) to (26), we have used the fact that, if $\bar{\delta}$ is an admissible cross-section distribution of types, then $s(\bar{\delta})$ has no atoms, and therefore $s(\{\pi_F^1\}|\bar{\delta}) = 0$. If $s(\{\pi_F^1\}|\delta)$ were strictly positive, this argument would not go through, and indeed the theorem would not be true.

For distributions of public-good valuations that satisfy $s(\{\pi_F^1\}) > 0$, there are actually two distinct versions of the argument given in the proof of Theorem 2 and two distinct necessary conditions for immunity to robust blocking. In one version, the set V_1 of valuations at which an agent prefers an increase in public-good provision is replaced by the set $\hat{V}_1 := V_1 \cup \{\pi_F^1\}$ valuations at which an agent prefers an increase in public-good provision or is indifferent about it, while the set of opponents is the same, agents with valuations in V_0. With this split of the population, the argument in the proof of Theorem 2 yields

$$s(\hat{V}_1) \geq s'(\hat{V}_1) \quad \text{implies} \quad Q_F(s) \geq Q_F(s'). \tag{27}$$

or, equivalentlyy,

$$s(V_0) \leq s'(V_0) \quad \text{implies} \quad Q_F(s) \geq Q_F(s'). \tag{28}$$

If, instead, we apply the argument with the sets V_1 and $\hat{V}_0 := V_0 \cup \{\pi_F^1\}$, we obtain (13), as before.

For type distributions satisfying $s(\{\pi_F^1\}) = 0$, all these monotonicity properties are equivalent. For type distributions satisfying $s(\{\pi_F^1) > 0$, however, conditions (13) and (27) or, equivalently, (28), may not be compatible. In fact, unless $Q_F(s)$ is constant on the set of distributions satisfying $s(V_1) > 0$, one can always find distributions s_1 and s_2 for which either (13) or (27) is violated.[27] In the Supplementary Material, we provide a more detailed discussion of the issue.

In the Supplementary Material, we also discuss the applicability of our analysis in models with finite type sets in which payoff type distributions necessarily have atoms. In such models, the condition $s(\{\pi_F^1\}|\delta) = 0$ is necessarily satisfied, if the π_F^1, i.e., the slope of the payment function is not itself an element of the set of payoff types. If this is the case, our results hold for models with finite type sets as well as the continuum model

[27]Let s^1, s^2 be such that $Q_F(s_1) > Q_F(s_2)$ and suppose that (27) holds. Then $s_1(\hat{V}_1) > s_2(\hat{V}_1) > 0$. However, if $s_1(\hat{V}_1) = s_1(\{\pi_F^1\})$ and $s_2(\hat{V}_1) = s_2(V_1)$, we also have $s_1(V_1) = 0 < s_2(V_1)$, which contradicts (13).

studied here. In contrast, the conclusion of Theorem 3 is not true if π_F^1 is an element of the set of payoff types.

To conclude this discussion, we note that Theorem 3 would not be true if we replaced immunity to robustly blocking coalitions by robust coalition-proofness. 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 of robust blocking eliminates this possibility.[28]

6.3 Implications and Extensions

Welfare Implications. An anonymous, robustly implementable social choice function that implements a first-best public-good provision rule is monotonic and involves equal cost sharing, i.e. an affine payment rule. From Theorem 2, we therefore obtain:

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

This corollary is related to the well known observation that the number of votes in favor of an alternative may be a poor indicator of the social surplus from that alternative. Deploring distortions from the use of voting mechanisms is not very useful however, if coalition-proofness concerns must be taken seriously. Monotonic social choice functions that condition on measures of social surplus, rather than numbers of votes are not immune to robust blocking.

The literature on public-good provision in small economies without participation constraints has mainly focused on the impossibility of robustly implementing first-best public-good provision with budget balance. In large economies, there is no problem with budget balance, but first-best implementation is vulnerable to manipulations by groups with common interests.

[28]This is why the analogue of Theorem 3 in Bierbrauer and Hellwig (2016a) restricts coalition formation by the Bernheim et al. (1986) requirement that a deviating coalition must itself be subcoalition-proof.

Given the impossibility of implementing a first-best provision rule, second-best considerations involve tradeoffs between overprovision of the public good in some states, underprovision of the public good in some states, and waste from payments exceeding the costs of public-good provision. Thus, in the example of Section 3, one might choose not to provide the public good under any circumstances, to provide the public good under all circumstances, or to provide the public good if the population share of people with valuation 3 exceeds a specified threshold, using the payment function $P_F(s) = 1.1 + 4Q_F(s)$, with excessive payments in non-provision states in order to change incentives for collective deviations. The Supplementary Material contains a more detailed discussion of these tradeoffs and shows that sometimes deviations from equal cost sharing with wasteful payments by participants may be preferable to the alternatives.

Nonlinear Payment and Provision Cost Functions. Theorems 2 and 3 show that an anonymous, monotonic robustly implementable social choice function with an affine payment function is immune to robust blocking if and only of it can be implemented by a voting mechanism. The proof of Theorem 2 makes essential use of the restriction to affine payment functions. As we observed above, this condition is automatically satisfied if the public good comes as a single indivisible unit. If the public good can be provided at multiple levels, the condition that payment functions be affine is still natural if the provision cost function is linear and the costs are shared equally among the participants.

Even so, it is important to consider what becomes of our analysis if payment functions and provision cost functions are nonlinear. We do this in a sequel to this paper, Bierbrauer and Hellwig (2016b). There we show that, in models with finitely many provision levels $Q = 1, 2, ..., n$, with increasing marginal provision costs, an anonymous and monotonic social choice function that is immune to robust blocking as well as robustly implementable must again be implementable by a voting mechanism. For an important class of social choice functions, these mechanisms rely on sequences of binary votes, where at each step, the question is whether to raise the provision level by another unit and the thresholds for doing so become ever larger as the level of public-good provision is increased. In this case, voting involves more than a single binary vote, but the basic principle remains valid that robust implementability and immunity to robust blocking preclude social choice functions from conditioning on preference intensities.

Integrating the Analysis of Public-Good Provision and Taxation. As mentioned in the introduction, one reason for developing a model of public-good provision in large economies is to provide a framework for integrating the analysis of public-good provision and taxation. In the present paper, taxation plays only a rudimentary role. The payment function P_F can be understood as a scheme for levying lump sum taxes that are the same for all and that serve to finance the public good. Given the requirements of anonymity and robust implementability, of course the given model does not admit any other form of taxation.

For a more interesting analysis of public-good provision and taxation, one needs a richer model. For example, along the lines of Bierbrauer (2009a, 2014), one might introduce the utility specification

$$U_i = v_i Q + c_i - \psi(y_i, \omega_i) \tag{29}$$

where, as before, Q is the public-good provision level, $v_i \in V$ is the agent's public-good valuation; c_i is the agent's consumption of the private good, and y_i the agent's output. The term $\psi(y_i, \omega_i)$ in (29) provides a measure of the agent's effort cost of producing y_i. As in Mirrlees (1971), this cost depends on a productivity parameter ω_i, which is privately known by the agent, like the public-good valuation v_i.

In this setting, a state of the economy is given by a cross-section distribution s of public goods valuations and productivity parameters. An anonymous social choice function is given by a triple $F = (Q_F, c_F, y_F)$ of functions such that $Q_F(s)$ is the level of public-good provision and $c_F(v, \omega, s), y_F(v, \omega, s)$ are the private-good consumption and production levels of an agent with hidden characteristics v, ω when the state of the economy is s. The difference $y_F(v, \omega, s) - c_F(v, \omega, s)$ corresponds to the payment $P_F(v, s)$ in the preceding analysis, but now this agent's payment may depend on ω as well as v and s. The functions c_F and y_F will be chosen so that, on aggregate, the net payments suffice to cover the cost of public-good provision and moreover provide whatever redistribution is considered desirable.

In this setting, as shown in Bierbrauer (2014), robust incentive compatibility amounts to the condition that payments must be independent of public-good valuations; payments may depend on individual productivity parameters (as well as the state of the economy) but must satisfy a Mirrlees (1971) incentive compatibility condition. This finding extends

Corollary 1. If coalition-proofness is also imposed, then, under certain conditions, public-good provision and redistribution can still be separated, and our finding that coalition-proof public-good provision requires voting remains valid. The conditions for separability of the two problems are restrictive and preclude, in particular, the possibility that changes in redistributive taxation might be used as a tool to improve the performance of voting mechanisms by changing the sets of opponents and proponents of public-good provision. The implications of this interference of public-good provision and income taxation must be a subject for further research.

7 Concluding Remarks

Our paper has several 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. To be sure, in a large finite economy, the view that individuals are not pivotal is strictly speaking not correct, but with a million people, the probability of being pivotal is on the order of 10^{-4}, an order of magnitude where it is unlikely to make much of a difference to people's behaviors.[29]

Second, 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. This mechanism, which satisfies budget balance, implements first-best outcomes robustly when there are no participation constraints.

Third, 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. While the literature has considered failures of coalition-proofness in connection with fail-

[29] As 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.

ures of budget balance, which are moot in the large economy, we identify a second type of failure of coalition-proofness that matters in large economies as well as small. This failure is due to the fact that first-best implementation may require information from people who are hurt by the use that is made of it. Providing this information is incentive-compatible because, as individuals, people feel that their reports do not make a difference to the outcome. However, collectively, they can make a difference. A requirement of coalition-proofness takes account of this collective interest (without violating individual incentive compatibility). In a large economy, such concerns about collective misrepresentations of preferences may preclude first-best implementation even if the coalitions themselves must satisfy a robustness condition, which prevents them from conditioning on people's beliefs about each other.

We consider it desirable to impose a robustness requirement on blocking coalitions as well as the social choice functions. Coalition formation involves a mechanism design problem of its own, and it seems natural to assume that a coalition organizer faces the same informational limitations as the overall mechanism designer. If we require the overall mechanism to be robustly incentive-compatible because the overall mechanism designer does not know the participants' beliefs, it would be somewhat incongruous to presume that collective deviations can condition on 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. The requirement of immunity to robust blocking in this paper takes account of the limitations in the information available for collective deviations.

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. Monotonicity is not a very restrictive condition. In models with finitely many participants, it is actually implied by incentive compatibility. Monotonicity is therefore a natural condition to impose if we think of the continuum model as an idealization of models with finitely many participants.

The focus on numbers of votes for the different alternatives entails a loss of information and a loss of efficiency. Because of these efficiency losses, economists tend to criticize the use of voting mechanisms. Our analysis suggests that these losses may be an unavoidable consequence of the fact that first-best implementation is vulnerable to collective misreporting. 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 beneftis 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 provision of the public good. The challenge now is to obtain a better understanding of the tradeoffs involved in designing second-best mechanisms.

References

Barberà, S. (1979). A Note on Group Strategy-Proof Decision Schemes. *Econometrica*, 47:637–640.

Barro, R. J. (1990). Government Spending in a Simple Model of Endogeneous Growth. *Journal of Political Economy*, 98(5):103–125.

Battaglini, M. and Coate, S. (2008). A Dynamic Theory of Public Spending, Taxation, and Debt. *American Economic Review*, 98(1):201–226.

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. (2016). Robustly Coalition-Proof Incentive Mechanisms for public-good provision are Voting Mechanisms and Vice Versa. *The Review of Economic Studies,* forthcoming.

Bierbrauer, F. and Hellwig, M. (in preparation). Public-Good Provision in Large Economies 2: The case of multiple provision levels. 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. (2014). Optimal Tax and Expenditure Policy with Aggregate Uncertainty. *American Economic Journal: Microeconomics*, 6:205–257.

Boadway, R. and Keen, M. (1993). Public Goods, Self-Selection and Optimal Income Taxation. *International Economic Review*, 34(3): 463–478.

Buchanan, J. and Tullock, G. (1962). The Calculus of Consent. Univerysity of Michigan Press, Ann Arbor.

Börgers, 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. (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.

Dasgupta, P., Hammond, P., and Maskin, E. (1979). The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility. *The Review of Economic Studies*, 46:185–216.

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. (1977). Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods. *Econometrica*, 45: 472-487.

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

Green, J. and Laffont, J.-J. (1979b). 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.

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

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

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

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. (1980). On Strategy-Proofness and Single Peakedness. *Public Choice*, 35:437–455.

Qiao, L., Sun, Y., and Zhang, Z. (2014). Conditional exact law of large numbers and asymmetric information economies with aggregate uncertainty. *Economic Theory*, DOI 10.1007/s00199-014-0855-6.

Samuelson, P. (1954). The Pure Theory of Public Expenditure. *Review of Economics and Statistics*, 36:387–389.

Sun, Y. (2006). The Exact Law of Large Numbers via Fubini extension and Characterization of Insurable Risks. *Journal of Economic Theory*, 126:31–69.

A Supplementary Material

A.1 A Coalition-Proof Crémer-McLean Mechanism for the Example in Section 2

We show how a Crémer-McLean machanism can be used to provide for coalition-proof Bayes-Nash implementation of a first-best public-good provision rule in the example of Section 2 with correlated values. 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 the random variablke $\tilde{\alpha}$. 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

$$\beta_L(0) = \frac{5}{6},\ \beta_H(0) = \frac{1}{6}, \beta_L(3) = \frac{1}{4},\ \beta_H(3) = \frac{3}{4}, \beta_L(10) = \frac{1}{2},\ \beta_H(10) = \frac{1}{2}. \tag{30}$$

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: 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.[30]

[30] The 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 $v_i = 10$.

The given incentive mechanism is also coalition-proof. Under the payment scheme in Table 2, 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 (30) 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 (30). For the posterior probabilities given by (30), 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.

A.2 Issues in Modeling Coalition-Proofness

As mentioned in the text, there are many ways to model coalition-proofness. We briefly discuss some of the major modelling issues.

- Axiomatic versus strategic approaches: As mentioned in the text, the first generation of papers on coalition-proofness in social choice and mechanism design proceeded axiomatically, by imposing an additional restriction without considering the strategic aspects of coalition formation itself.[31] By contrast, a more recent literature studies coalition formation itself as part of the overall strategic interaction that is being analysed. Thus Laffont and Martimort (1997, 2000) introduce communication between a coalition organizer and participants into the extensive form game that determines the constraints faced by the overall mechanism designer.[32]

- Normal versus extensive-form approaches: 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.

- To what extent are coalitions constrained by subcoalition formation? Bernheim et al. (1986) impose the condition that any deviating coalition must itself be robust

[31] See, in particular, Bennett and Conn (1977), Barberà (1979), Dasgupta et al. (1979), Green and Laffont (1979b), Moulin (1980).

[32] In Bierbrauer and Hellwig (2010), we also followed this approach. In the game considered in that paper, 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.

to the formation of subcoalitions, which in turn must be robust to the formation of subsubcoalitions, and so on. This condition plays an important role in Peleg and Sudhölter (1999) and in Bierbrauer and Hellwig (2016a).[33] Because of its recursive nature, it is not clear how this condition should be applied in a large economy. In a model with finitely many agents, one can proceed inductively, beginning with one-person coalitions, and then considering coalitions with k participants to be admissible if they are not in turn blocked by sub-coalitions involving j participants where $j < k$. In a model with a continuum of participants, this procedure does not work because the "smallest" sets of participants are null sets, which are negligible.

- What is the appropriate solution concept? As mentioned in the text, the first generation of papers relied on dominant-strategy implementation and modelled coalition-proofness through a requirement of coalitionally dominant-strategy proofness. By contrast, Bernheim et al. (1986) work with a concept of Nash equilibrium, Laffont and Martimort (1997, 2000) with a refinement of sequential equilibrium. In Bierbrauer and Hellwig (2016a), as in the present paper, we work with a normal-form Bayes-Nash approach but require that individual incentive compatibility holds robustly, i.e. for every possible specification of agents' beliefs.

- What information is available to deviating coalitions? Coalitionally dominant-strategy proofness assumes that in designing a collective deviations the participants fully know the state of the economy. The same assumption underelies the requirement of ex post coalition-proofness, which is why coalitionally dominant-strategy proofness and ex post coalition-proofness are equivalent. However, ex post coalition-proofness is weaker than robust coalition-proofness. Depending on how beliefs are specified, interim uncertainty may provide room for coalitions to exploit belief-based

[33]This restriction on coalition formation plays an important role in Peleg and Sudhölter's (1999) result on the coalition-proofness of generalized median voter rules for models with single-peaked preferences over multidimensional sets of alternatives. Whereas Moulin (1980) found that, in models with single-peaked preferences overe unidimensional choice sets, simple median-voter rules are "group strategy-proof", the Peleg-Sudhölter generalization of this result requires the Bernheim et al. (1986) restriction to coalitions that are themselves immune to blocking by subcoalitions that are... In Bierbrauer and Hellwig (2016a), we use the Bernheim et al. (1986) restriction to eliminate coalitions that involve both, adherents and opponents of public-good provission.

gains from "trade" between groups of agents with conflicting interests. In public-good provision, for example, a coalition of adherents and opponents might jointly manipulate overall outcomes so that the level of public-good provision is lowered in states that the opponents consider to be relatively likely and is raised in states that the adherents consider to be relatively likely. As discussed in Bierbrauer and Hellwig (2016a), such coalitions do not however meet the Bernheim et al. (1986) condition of subcoalition-proofness. They also do not satisfy the criterion, introduced in this paper, that the advantageousness of joining a coalition should be "robust", i.e. independent of how the participants' beliefs are specified.

The requirement of robust coalition-proofness fudges 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 $[(T,T),\tau,\beta_\delta]$ 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 $[(T,T),\tau,\beta_\delta]$. Consequently, 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 has become known among the participants. In Bierbrauer and Hellwig (2016a), we use these constraints to show that every robust and coalition-proof social choice function is a voting mechanism. In dealing with robustly blocking coalitions, this proof strategy is not available because collective deviations must be attractive to the participants for every admissible belief systems.

- What mechanisms do coalitions use to ensure the incentive compatibility of intra-coalition communication? In particular, is there a role for side-payments in facilitating coalition formation? With finitely many participants, not allowing for side-payments typically involves a loss of generality. It restricts what coalitions can do and therefore tends to enlarge the set of "coalition-proof" social choice functions. With a continuum of participants, individual incentive compatibility is equivalent to the requirement that payments inside the coalition should be independent of messages; in the next section, we show that, with a robustness requirement on the coalition design itself, such side payments must actually be zero.

- What overall mechanisms are used to implement the social choice function? With a requirement of coalition-proofness, the validity of the revelation principle cannot generally be taken for granted.[34] However, one can show that, with a requirement of robust implementation in a large economy, indirect mechanisms do not improve the scope for overall mechanism design.[35]

A.3 Irrelevance of Side Payments in Robust Coalition Design

We argue that, in our large-economy model with robust incentive, participation, and feasibility constraints, there is no role for side payments in robust coalition design. 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 $\delta_{T'}$ of types in T' but not on the belief system β or on the distribution of types in $T \backslash T'$. Given a belief system β, an individual of type $t' \in T'$ who participates in this coalition receives the expected side payment $\int_{\mathcal{M}(T')} z_{T'}(t', \delta_{T'})\, d\, \beta(\delta_{T'} \mid t')$ and achieves the expected utility

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

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,

$$\int_{\mathcal{M}(T')} z_{T'}(t', \delta_{T'})\, d\, \beta(\delta_{T'} \mid t') \geq \int_{\mathcal{M}(T')} z_{T'}(\hat{t}', \delta_{T'})\, d\, \beta(\delta_{T'} \mid t') \,, \tag{31}$$

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

[34] See Boylan (1998) or Bierbrauer (2014).

[35] See the analysis in Bierbrauer and Hellwig (2010).

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

$$\int_{\mathcal{M}(T')} z_{T'}(t', \delta_{T'})\ d\ \beta(\delta_{T'} \mid t') \geq 0\ , \tag{32}$$

for all $t' \in T'$.

Feasibility Constraint. The coalition organizers must not expect to lose money.[36]

We require incentive and participation constrains to hold robustly, i.e. for every admissible belief system.[37] 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 $\bar{z}_{T'} : \mathcal{M}(T') \to \mathbb{R}$ so that, for all $t' \in T'$, $z_{T'}(t', \delta_{T'}) = \bar{z}_{T'}(\delta_{T'})$, and the participation constraint implies that

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

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.

■

[36]Our 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.

[37]The arguments of Crémer and McLean (1988), McAfee and Reny (1992) and Gizatulina and Hellwig (2015) imply that, for generic belief functions β, one can find 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.

A.4 Allowing for Mass Points in Payoff Type Distributions

Throughout the paper, we assumed that cross-section distributions of public-good valuations do not have mass points. As we discussed in the text, this assumption is needed in the proof of Theorem 3. Without it, Theorem 3 would not be true; in fact, only social choice functions with trivial provision rules for the public good would be immune to robust blocking.

As we discussed in the text, a straightforward modification of the arguments in the proof of Theorem 2 can be used to show that, if an anonymous, monotonic and robustly incentive compatible social choice function with an affine payment rule is immune to robust blocking, then the implications

$$s(V_1) \geq s'(V_1) \quad \text{implies} \quad Q_F(s) \geq Q_F(s'). \tag{33}$$

and

$$s(V_0) \leq s'(V_0) \quad \text{implies} \quad Q_F(s) \geq Q_F(s'). \tag{34}$$

must both be true. These implications are equivalent if the distribution s assigns zero mass to the singleton $\{\pi_F^1\}$ so that $s(V_1) + s(V_0) = 1 - s(\{\pi_F^1\}) = 1$. This equivalence breaks down if $s(\{\pi_F^1\}) > 0$.

If (33) and (34) must hold simultaneously for all s, with $s(\{\pi_F^1\}) > 0$ as well as $s(\{\pi_F^1\}) = 0$, the public-good provision rule must actually be constant on the set of distributions satisfying $0 < s(V_1) < 1$. To see why, consider the case where the public good comes as a single indivisible unit. In this case, (34) implies that, for some $\bar{s}_0$ and any cross-section payoff type distribution s,

$$Q_F(s) = 1 \quad \text{implies} \quad s(V_0) < \bar{s}_0, \tag{35}$$

and

$$Q_F(s) = 0 \quad \text{implies} \quad s(V_0) > \bar{s}_0. \tag{36}$$

Similarly, (33) implies that there exists $\bar{s}_1$ so that

$$Q_F(s) = 1 \quad \text{implies} \quad s(V_1) > \bar{s}_1, \tag{37}$$

and

$$Q_F(s) = 0 \quad \text{implies} \quad s(V_1) < \bar{s}_1. \tag{38}$$

However, if s is not restricted to be atomless, 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 satisfying $0 < s(V_1) < 1$. Thus, if we allow for payoff type distributions with mass points at $v = \pi_f^1$, 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.

This destructive conclusion is avoided if, somewhat arbitrarily, we assume that people with payoff type $v = \pi_F^1$ 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.

In Theorems 2 and 3, 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 conclusions of Theorems 2 and 3 hold for any payoff type space V and any social choice function F with a payment function taking the form (10) with slope parameter $\pi_F^1 \notin V$.*

This extension of Theorems 2 and 3 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 conclusions of Theorems 2 and 3 hold anyway unless the slope parameter π_F^1 is an element of V. In the example of Section 3, with $V = \{0, 3, 10\}$, this requirement is satisfied if the social choice function stipulates equal cost sharing, i.e., $\pi_F^0 = 0$, and $\pi_F^1 = k = 4$.

A.5 Second-Best Welfare Considerations

In the following, we use a further example to discuss second-best considerations. In this example, $\bar{Q} = 1$, so, as in example of Section 3, there is a simple binary choice between provision and non-provision of the public good. We set $V = \{0, 5, 10\}$, so there are three possible payoff types. The per-capita cost of public-good provision is $k = 4.5$. There are two possible cross-section distributions s^j, $j = 0, 1$ of payoff types. The population shares s_v^j of the different payoff types under these two cross-section distributions are given in Table 3.[38]

Table 3.

j	s_0^j	s_5^j	s_{10}^j	$\bar{\mathbf{v}}(s^j)$
0	0.3	0.7	0	3.5
1	0.4	0.1	0.5	5.5

(39)

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

In this example, first-best implementation requires that the public good should not be provided in state 0 and that the public good should be provided in state 1. With equal cost sharing, the associated payments would be $P_F(s^0) = 0$ and $P_F(s^1) = 4.5$. Given these payments, the set of opponents of public-good provision consists of all types with valuations 0 and the set of net beneficiaries 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 0 and of 0.6 in state 1. Because the population share of the set of net beneficiaries is larger in state 0 than in state 1, 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

[38] The 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.

the population share of the cooalition 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.

Because 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 funds than he needs, e.g., by asking for a payment $P_F(s^1) = 5.1$ if the public good is provided, rather than $P_F(s^1) = k = 4.5$, 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 π_F^0, π_F^1, and $Q_F : \mathcal{M}(V) \to \{0, 1\}$ so as to maximize the expected aggregate surplus

$$E^M[(\bar{\mathbf{v}}(s) - \pi_F^1)Q_F(s) - \pi_F^0(1 - Q_F(s))] \tag{40}$$

subject to the feasibility constraints that $\pi_F^0 \geq 0$, $\pi_F^1 \geq k$, 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 (40) indicates that expectations over s are taken with respect to the mechanism designer's subjective beliefs.

The solution to this second-best problem depends on the probabilities ρ_0^M and ρ_0^M that the mechanism designer assigns to the different states. In the example of Table 3, 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 and no payments in state 1 and provision with a payment $\pi_F^1 = 5.1 > k$ 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 $\pi_F^0 = 0$ and $\pi_F^1 = k = 4.5$ 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.

References

Boylan, R. (1998). Coalition-Proof Implementation. *Journal of Economic Theory*, 82:132–143.

Gizatulina, A. and Hellwig, M. (2015). The Genericity of the McAfee-Reny Condition for Full Surplus Extraction in Models with a Continuum of Types. Mimeo, Max Planck Institute for Research on Collective Goods.

McAfee, P. and Reny, P. (1992). Correlated Information and Mechanism Design. *Econometrica*, 60: 395–421.

Peleg, B. and Sudhölter, P. (1999). Single-peakedness and coalition-proofness. *Review of Economic Design*, 4: 381–387.