



Information transmission in coalitional voting games

Roberto Serrano^{a,*}, Rajiv Vohra^b

^a*Department of Economics, Brown University, Providence, RI 02912, USA*

^b*Office of the Dean of the Faculty, Brown University, Providence, RI 02912, USA*

Received 12 April 2005; final version received 7 February 2006

Available online 30 March 2006

Abstract

A core allocation of a complete information economy can be characterized as one that would not be unanimously rejected in favor of another feasible alternative by any coalition. We use this test of coalitional voting in an incomplete information environment to formalize a notion of resilience. Since information transmission is implicit in the Bayesian equilibria of such voting games, this approach makes it possible to derive core concepts in which the transmission of information among members of a coalition is endogenous. Our results lend support to the credible core of Dutta and Vohra [Incomplete information, credibility and the core, *Math. Soc. Sci.* 50 (2005) 148–165] and the core proposed by Myerson [Virtual utility and the core for games with incomplete information, *Mimeo, University of Chicago*, 2005] as two that can be justified in terms of coalitional voting.

© 2006 Elsevier Inc. All rights reserved.

JEL classification: C71; C72; D51; D82

Keywords: Core; Incomplete information; Coalitional voting; Resilience; Mediation

1. Introduction

In the complete information setting, the core provides a natural way of formalizing coalitional stability. Simple as the core concept is, there is no unique, unambiguous way of extending it to an exchange economy with incomplete information at the interim stage; see Forges et al. [6] for a survey. This paper is concerned with the issue of making endogenous the amount of information that agents share in the process of cooperation. We shall formalize coalitional stability by means

* Corresponding author. Fax: +1 401 863 1970.

E-mail addresses: roberto_serrano@brown.edu (R. Serrano), rajiv_vohra@brown.edu (R. Vohra).

URLs: <http://www.econ.brown.edu/faculty/serrano> (R. Serrano), <http://www.econ.brown.edu/~rvohra> (R. Vohra).

of equilibria of voting games in which agents choose between the status-quo and another feasible alternative.

At the interim stage, there are two sets of assumptions to be made (amount of information sharing, and whether or not incentive constraints are imposed) before the standard definition of the core can be extended to the incomplete information model. For example, we may assume that there is no sharing of information; agents in a coalition can only rely on common knowledge events to construct potential objections. This approach leads to the notion of the coarse core, first formalized by Wilson [18] in a framework without incentive constraints, and subsequently extended by Vohra [16] to include them. The main idea underlying the coarse core, namely that a coalition's objection is focused on a common knowledge event, is motivated by standard issues of adverse selection; see the examples in Wilson [18] and Vohra [16].

Not surprisingly, the theory is quite different if one allows coalition members to share information. If information sharing is unrestricted among the members of a coalition, one arrives at the notion of the fine core, also proposed by Wilson [18]. Incentive constraints can be added to a fine objection by requiring that it be incentive compatible over the event that is relevant for the objection. While unrestricted sharing of information may seem arbitrary and, in many instances, unreasonable, there do exist cases in which some amount of information sharing seems natural.

Our aim is to make endogenous the amount of information that is shared among agents when a coalition forms in order to block a status-quo allocation.¹ While this concern is by no means new (see, for example, Dutta and Vohra [4]), our approach here is based on viewing the core in another, more positive (perhaps more primitive) way. We argue that non-cooperative equilibrium theory is ideally suited to deal with the question of how much private information agents transmit to each other. This is so even though we are interested in modeling cooperative behavior.² This leads to the idea of formalizing coalitional stability by making use of non-cooperative equilibrium behavior in a voting game.

Suppose a status-quo allocation is in the core, and a coalition compares it to some other feasible allocation. Clearly, if the two alternatives were to be voted upon by the agents in the coalition, we would not expect a unanimous acceptance of the alternative over the status-quo. Indeed, in the complete information setting this is a defining property of the core. We take this simple test of stability and apply it systematically to a model with incomplete information. The amount of information sharing as well as the incentive constraints will then emerge as equilibrium conditions of a voting game.

One might think that precise details of the voting game may turn out to be critical, perhaps leading to a plethora of different notions of core stability. Fortunately, we are able to show that this is not the case; the coalitional voting approach is quite sharp in its conclusions, and there are only several important aspects of the voting game that matter for core stability, leading to very few possible core concepts. Indeed, a version of the revelation principle is at work, as a Bayesian equilibrium of an arbitrary voting mechanism can be replaced by its outcome-equivalent truthful equilibrium in a direct voting game, where the message simply consists of a type report and a “yes–no” vote to the alternative. Thus, while information transmission is endogenous in the Bayesian

¹ Our approach is related to the notions of durability in Holmström and Myerson [7] and credibility in Dutta and Vohra [4], as we shall discuss below.

² While the ability to cooperate makes it reasonable to allow for unfettered, frictionless communication, it does not mean that agents will necessarily share their private information with others or believe others' claims that cannot be verified.

equilibrium actions of non-cooperative voting games, it is also robust to different specifications of their details.

If incentive constraints are not important (because types become verifiable) we show that the coalitional voting approach yields the fine core of Wilson [18] as the natural concept. In the general case where types are not verifiable, resilience to coalitional voting yields the credible core of Dutta and Vohra [4]. In this sense, the credible core can be seen as the appropriate generalization of the fine core when incentive constraints are relevant (as will be explained, the credible core is not simply the incentive compatible fine core described above).

An interesting modification of this simple voting game allows a mediator to construct a more sophisticated challenge to a status-quo by implementing the alternative as a function of the agents' reported types. Effectively, the only difference between this game and the simpler one corresponding to the credible core is that in this case a coalition can challenge a status-quo over an informational event that is not necessarily a product event. We call the corresponding set of resilient allocations the mediated core. The importance of the product/non-product structure of the event over which an objection takes place justifies the following analogy: a mediated objection is to a credible objection as a correlated equilibrium is to a Nash equilibrium.

Finally, we allow a mediator to randomize over the coalitions that it approaches for a vote. This leads to a form of resilience that yields the randomized mediated core, a notion that corresponds closely with a recent core concept suggested by Myerson [12]. Myerson's approach is based on virtual utility, and he is able to prove non-emptiness of his core concept in his class of games, provided that severance payments take place in an additional commodity, a numeraire, and feasibility is only required in expected terms. It follows from our results that his non-emptiness result also implies non-emptiness of the core concepts found here under his assumptions, since the blocking restrictions that define the mediated core and the credible core are less stringent than those in Myerson's. On the other hand, we also know that the non-emptiness question in general is difficult, given the results in Forges et al. [5] and Vohra [16], showing that the incentive compatible coarse core, a superset of all these other concepts, may be empty.

2. Preliminaries

The basic model of an exchange economy with asymmetric information can be formulated as follows. Let T_i denote the (finite) set of agent i 's types. The interpretation is that $t_i \in T_i$ denotes the *private information* possessed by agent i . With $N = \{1, \dots, n\}$ as the finite set of agents, let $T = \prod_{i \in N} T_i$ denote the set of all information states. We will use the notation t_{-i} to denote $(t_j)_{j \neq i}$. Similarly $T_{-i} = \prod_{j \neq i} T_j$, $T_S = \prod_{j \in S} T_j$ and $T_{-S} = \prod_{j \notin S} T_j$. We assume that agents have a common prior probability distribution q defined on T , and that no type is redundant, i.e., $q(t_i) > 0$ for all $t_i \in T_i$ for all i . At the interim stage, nature chooses $t \in T$, and each agent i knows her type, t_i . Hence, conditional probabilities will be important: for each $i \in N$ and $t_i \in T_i$, the conditional probability of $t_{-i} \in T_{-i}$, given t_i is denoted by $q(t_{-i}|t_i)$.

We assume that there are a finite number of commodities, and that the consumption set of agent i is $X_i \subseteq \mathbb{R}_+^l$. Agent i 's utility function in state t is denoted $u_i(\cdot, t) : X_i \times T \mapsto \mathbb{R}$. We shall assume that $u_i(x, t) \geq u_i(0, t)$ for all i , all $t \in T$ and all $x \in X_i$. The endowment of agent i of type t_i is $\omega_i \in X_i$ (assumed to be independent of the state—with this assumption, all private information concerns agents' preferences and beliefs).

We can now define an exchange economy as $\mathcal{E} = \langle (u_i, X_i, \omega_i, T_i)_{i \in N}, q \rangle$.

For coalition $S \subseteq N$, a feasible (state contingent) S -allocation, $x : T \mapsto \mathbb{R}^{ls}$ (where s denotes the cardinality of S), consists of a commodity bundle for each consumer in S in each state such that

$\sum_{i \in S} x_i(t) \leq \sum_{i \in S} \omega_i$ for all $t \in T$, and satisfying that $x(t_S, t'_{-S}) = x(t_S, t''_{-S})$ for all $t_S \in T_S$ and for all $t'_{-S}, t''_{-S} \in T_{-S}$. (The latter assumption is made to exclude basic externalities across coalitions, i.e., the set of feasible allocations to a coalition is independent of the information held by the complement). We will denote by \mathcal{A}_S the set of feasible state contingent allocations of S . Thus, we shall use \mathcal{A}_S to denote the set of feasible allocations in a given state: $\mathcal{A}_S = \{(x_i) \in \prod_{i \in S} X_i \mid \sum_{i \in S} x_i \leq \sum_{i \in S} \omega_i\}$. Similarly, state contingent N -allocations are simply referred to as allocations, and the set of state contingent allocations is denoted by \mathcal{A} .

Given $x \in \mathcal{A}_S$, the interim utility of agent $i \in S$ of type t_i is

$$U_i(x|t_i) = \sum_{t_{-i} \in T_{-i}} q(t_{-i}|t_i) u_i(x_i(t_{-i}, t_i), (t_{-i}, t_i)).$$

If agent i of type t_i pretends to be of type t'_i (while all other agents are truthful), she gets interim utility:

$$U_i(x, t'_i|t_i) = \sum_{t_{-i} \in T_{-i}} q(t_{-i}|t_i) u_i(x_i(t_{-i}, t'_i), (t_{-i}, t_i)).$$

An S -allocation $x \in \mathcal{A}_S$ is *incentive compatible* if for every $i \in S$ and for every $t_i, t'_i \in T_i$

$$U_i(x|t_i) \geq U_i(x, t'_i|t_i).$$

We shall denote the set of incentive compatible S -allocations by \mathcal{A}_S^* , and the set of incentive compatible allocations by \mathcal{A}^* .

For an event $E \subseteq T$ and $t_i \in T_i$, let

$$E_{-i}(t_i) = \{t_{-i} \in T_{-i} \mid (t_i, t_{-i}) \in E\}$$

and

$$E_i = \{t_i \in T_i \mid E_{-i}(t_i) \neq \emptyset\}.$$

Consider an allocation rule $x \in \mathcal{A}$, agent i of type t_i and an event E . Suppose $q(E_{-i}(t_i)) > 0$. Then the interim utility conditional on E can be expressed as

$$U_i(x|t_i, E) = \sum_{t_{-i} \in E_{-i}(t_i)} q(t_{-i}|t_i) u_i(x_i(t_{-i}, t_i), (t_{-i}, t_i)).$$

(Strictly speaking, the expression on the right-hand side should be divided by $q(E_{-i}(t_i)|t_i)$, but we will find it convenient not to do so.)

The corresponding interim utility (conditional on E) if type t_i pretends to be of type t'_i is

$$U_i(x, t'_i|t_i, E) = \sum_{t_{-i} \in E_{-i}(t_i)} q(t_{-i}|t_i) u_i(x_i(t_{-i}, t'_i), (t_{-i}, t_i)).$$

Given $E \subseteq T$, an S -allocation $x \in \mathcal{A}_S$ is *incentive compatible over E* if for every $i \in S$ and for every $t_i, t'_i \in E_i$

$$U_i(x|t_i, E) \geq U_i(x, t'_i|t_i, E).$$

The set of such allocations is denoted $\mathcal{A}_S^*(E)$.

In the next section, we will begin by considering events that refer to independent subsets of T_i , and are therefore product events of the form $E = \prod_i E_i$, where $E_i \subseteq T_i$ for all i . In later sections this restriction will be relaxed.

3. Coalitional voting

One way to assess the coalitional stability of a status-quo is to ask whether or not some coalition would vote unanimously in favor of another feasible “outcome.” This is a somewhat pedantic exercise in the complete information framework because there it is obvious that the core is precisely the set of allocations which no coalition would vote (unanimously) to give up in favor of some other feasible allocation. Moreover, this conclusion does not depend on precisely how the coalitional voting game is constructed. Our aim is to take this simple characterization of core stability in the complete information case and extend it to a model with incomplete information.³ Since several new issues arise in this transition, it is useful to begin by illustrating the coalitional voting approach in the simple setting where types are verifiable. The main focus of the paper, however, is the more general case in which types are not verifiable.

3.1. Verifiable types

We take verifiability of types to mean that eventually all private information becomes public, and contracts contingent on types can be enforced (by using prohibitive penalties if necessary). In this setting agents cannot misrepresent their types, and incentive constraints become unnecessary. Despite this simplicity, it is by no means obvious how the core should be defined in such a model. Indeed, Wilson [18] suggested two distinct notions of the core: the fine core (allowing arbitrary information sharing) and the coarse core (allowing agents to coordinate only on common knowledge events); see Forges et al. [6] for precise definitions based on the types formulation. Wilson’s notion of the core, and others that have since been suggested, all follow the traditional approach based on dominance and feasibility of a potential objection. The main conceptual issue in the construction of an objection concerns an exogenously specified information transmission within the coalition (see also Lee and Volij [8] and Volij [17]).⁴ The coalitional voting approach is different in that it seeks to justify the notion of an objection from more primitive principles: an equilibrium in a voting game is used as a means for defining what an “objection” is. As we shall see, in the verifiable types case, this approach leads to just one corresponding notion of the core—Wilson’s fine core. We see this as an illustration of the power of the coalitional voting approach in selecting an appropriate notion of core stability.

Suppose $x \in \mathcal{A}$ is the status-quo and coalition S can consider making use of a voting game to discard x in favor of $y \in \mathcal{A}_S$. More precisely, members of S vote to either “accept” or “reject” the new proposal y . It is important that y is not seen as a proposal by a particular agent, in order to avoid complications of signaling that such an act would imply.⁵ The coalition is simply asked to make a choice between x and y . In case of a unanimous “accept” vote, the outcome assigned will

³ The test being performed here is weaker than the implementation exercise. In the classical framework it is possible to construct a game form that implements the core of an economy; see, for example, Perry and Reny [13] and Serrano and Vohra [15]. The present paper may be seen as a preliminary step in developing game forms to implement the “core” in incomplete information economies.

⁴ An alternative route is taken by de Clippel [2], who proposes a competitive screening procedure with outside brokers that absorb bankruptcies in objections. This allows him to transform each agent’s decision problem of whether joining an objection into a one-person decision problem, independent of more complex considerations, such as the ones tackled here in terms of information transmission.

⁵ An extension to coalitional settings of the type of analysis in Myerson [10] or Maskin and Tirole [9] would be an important next step. In their approach, the informed principal proposes the agent a contract that, therefore, may signal some of his private information.

be $y(t)$ to coalition S , where t is the profile of actual types. Note that type announcements are not necessary because we are assuming that the types eventually become publicly known.⁶ In case of a “reject” vote from any member of S the mediator assigns $x(t)$ to the grand coalition. This defines a game $\Gamma_x^0(S, y)$, in which the only active players are all the players in S , who simultaneously must choose from $\{a, r\}$, (“accept” or “reject” the alternative y). Then, y is implemented if and only if it is unanimously accepted; otherwise, x is the outcome. A strategy for player i is a function $v_i : T_i \mapsto \{a, r\}$. Given a profile of strategies $(v_i(\cdot))$, the outcome $\phi(v)$ for members of S is defined as follows:

$$\phi_i(v(t)) = \begin{cases} y_i(t) & \text{if } v_i(t_i) = a \text{ for all } i \in S, \\ x_i(t) & \text{otherwise.} \end{cases}$$

A strategy profile \bar{v} of the voting game $\Gamma_x^0(S, y)$ is a *Bayesian Nash equilibrium* if, for all $i \in S$ and $t_i \in T_i$,

$$U_i(\phi_i(\bar{v}|t_i)) \geq U_i(\phi_i(\bar{v}_{-i}, v_i)|t_i) \quad \text{for all } v_i : T_i \mapsto \{a, r\}.$$

Given a Bayesian Nash equilibrium \bar{v} , define for each i the set of types who vote to accept the alternative, i.e.,

$$E_i(\bar{v}) = \{t_i \in T_i | v_i(t_i) = a\}.$$

For $i \in S$, let $E_{-i} = \prod_{j \in S, j \neq i} E_j \times T_{-S}$.

In a voting game $\Gamma_x^0(S, y)$, a Bayesian Nash equilibrium \bar{v} is said to be an *equilibrium rejection* of x if there is positive probability that in equilibrium all agents vote to accept the alternative mechanism, i.e., $q(\prod_{i \in S} E_i(\bar{v})) > 0$, and $\phi(\bar{v})$ is not interim equivalent to x for all $i \in S$ (in the sense that there exists $i \in S$ and $t_i \in T_i$ such that $U_i(\phi_i(\bar{v})|t_i) > U_i(x|t_i)$).

A status-quo x is *resilient to coalitional voting with verifiable types* if there does not exist a voting game $\Gamma_x^0(S, y)$ with an equilibrium rejection of x .

Proposition 1. *Given $x \in \mathcal{A}$, a coalition S and $y \in \mathcal{A}_S$, the following statements are equivalent:*

- (a) *There is an equilibrium rejection of x in the voting game $\Gamma_x^0(S, y)$.*
- (b) *There exists $y' \in \mathcal{A}_S$ and $E_i \subseteq T_i$ for all $i \in S$, where $q(E) > 0$ for the product event $E = \prod_{i \in S} E_i \times T_{-S}$, such that*

$$U_i(y'|t_i, E) \geq U_i(x|t_i, E) \quad \text{for all } i \in S \text{ and all } t_i \in E_i,$$

with strict inequality for some i and t_i .

Corollary 1. *x is resilient to coalitional voting with verifiable types if and only if it belongs to the fine core.*

The corollary follows from the fact that condition (b) is essentially the same as the definition of a fine objection; see, for example, Forges et al. [6]. The only difference is that an “objection” is usually defined with a strict inequality for all i . It should be clear that the characterization of

⁶ A model in which agents have to announce their types in order to implement the allocation, but types become known at a later stage (and prohibitive penalties are available), is operationally identical to this because it can be assumed that agents announce their types truthfully.

equilibria of voting games will be in the form of weak inequalities. For this reason, throughout this paper, when we refer to a certain core concept it should be understood to be the “strong” version of that core concept.⁷

It is also worth remarking that statement (a) in the proposition simply refers to an equilibrium of the voting game, and unlike statement (b) or the definition of a fine objection, it is not couched in terms of a particular event over which an “objection” takes place. This is a feature of our other equivalence results below as well.

Proof of Proposition 1. An equilibrium $\bar{\sigma}$ of the voting game $\Gamma_x^0(S, y)$ can be characterized equivalently in terms of y and x as follows. For all $i \in S$ and $t_i \in E_i = E_i(\bar{\sigma})$,

$$\begin{aligned} & \sum_{t_{-i} \in E_{-i}} q(t_{-i}|t_i)u_i(y, (t_i, t_{-i})) + \sum_{t_{-i} \notin E_{-i}} q(t_{-i}|t_i)u_i(x, (t_i, t_{-i})) \\ & \geq \sum_{t_{-i} \in T_{-i}} q(t_{-i}|t_i)u_i(x, (t_i, t_{-i})), \end{aligned}$$

and for all $i \in S, t_i \notin E_i$,

$$\begin{aligned} & \sum_{t_{-i} \in E_{-i}} q(t_{-i}|t_i)u_i(y, (t_i, t_{-i})) + \sum_{t_{-i} \notin E_{-i}} q(t_{-i}|t_i)u_i(x, (t_i, t_{-i})) \\ & \leq \sum_{t_{-i} \in T_{-i}} q(t_{-i}|t_i)u_i(x, (t_i, t_{-i})). \end{aligned}$$

These two conditions are equivalent to the following two conditions:

$$U_i(y|t_i, E) \geq U_i(x|t_i, E) \quad \text{for all } i \in S \text{ and } t_i \in E_i \tag{1}$$

and

$$U_i(y|t_i, E) \leq U_i(x|t_i, E) \quad \text{for all } i \in S \text{ and } t_i \notin E_i. \tag{2}$$

Clearly, this means that (a) implies (b).

To see that (b) implies (a) consider y' and E satisfying condition (b), and define

$$y(t) = \begin{cases} y'(t) & \text{if } t \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Since $U_i(y'|t_i, E) = U_i(y|t_i, E)$ for all $i \in S$ and $t_i \in E_i$, it follows that y satisfies both (1) and (2), and this completes the proof. \square

3.2. Non-verifiable types

When types are not verifiable, we need to be more careful about precisely what a status-quo means. Henceforth, we shall take a *status-quo* x to refer to an incentive compatible, state contingent allocation (thus $x \in \mathcal{A}^*$) with the interpretation that in every state t , the outcome is $x(t)$, unless there is an agreement to change it. This means, in particular, that if there is an attempt to change the status-quo but the attempt fails, the outcome in state t is $x(t)$, i.e., any *discussion* about a possible

⁷ Monotonicity of preferences will suffice to eliminate this difference as well, in the present result. However, when we turn to the case of non-verifiable types, this difference may be important.

change does not by itself allow any agent to strategically manipulate the status-quo x .⁸ Since x was assumed to be incentive compatible, this can be justified by a scenario in which agents have already sent their truthful reports to implement the status-quo via a direct mechanism. (In fact, this justification still holds even if an indirect mechanism is behind the status-quo, in which case the agents have sent a report about the action that each of them plans to take in that mechanism, and such actions are enforced.) The idea is then that these reports cannot be changed if the status-quo prevails. This assumption greatly simplifies the analysis, and will be used throughout the rest of this paper.

As before, coalitional voting may involve a comparison between a status-quo (a feasible allocation rule for the grand coalition) and an allocation rule that is feasible (both informationally and physically) for the given coalition. However, when types are not verifiable, an allocation rule is only the simplest example of what a coalition may consider as an alternative. More generally, we may think of a coalition constructing a communication mechanism of its own to map out its feasible set of alternatives. Indeed, this will capture the idea that a coalition is permitted as much latitude in constructing a competing proposal as is within the bounds of feasibility. A voting game can now be used to describe the situation in which members of a coalition vote to reject a status-quo in favor of a competing mechanism. The new mechanism is applied if all agents vote for it; otherwise, the status-quo remains as the outcome. We will say that a status-quo is resilient to coalitional voting if there does not exist a coalition and a Bayesian Nash equilibrium of a voting game in which an alternative mechanism is unanimously accepted in some positive probability state. Note that our approach yields a very strong notion of stability; a resilient allocation rule cannot be rejected in *any* Bayesian Nash equilibrium of *any* competing mechanism in any (positive probability) state. We now turn to formal definitions.

A mechanism for coalition S consists of message sets M_i for each $i \in S$ and an outcome function $g : M \mapsto \mathcal{A}_S$.

A status-quo $x \in \mathcal{A}^*$ and a mechanism for S , $((M_i)_{i \in S}, g)$, define a voting game in which each player $i \in S$ chooses an action in $\{a, r\} \times M_i$. The interpretation being that a refers to an “accept” vote in favor of the new mechanism while r refers to a rejection of the new proposal. A strategy for player i is therefore $\sigma_i : T_i \mapsto \{a, r\} \times M_i$. We will find it convenient to denote $\sigma_i(t_i) = (v_i(t_i), m_i(t_i))$, where the first element denotes the vote of agent i of type t_i . Given a strategy profile σ , let $\phi(\sigma)$ denote the corresponding outcome for coalition S . In state t , the commodity bundle assigned to each $i \in S$ is denoted $\phi_i(\sigma(t))$, and defined as follows⁹:

$$\phi_i(\sigma(t)) = \begin{cases} g_i(m(t)) & \text{for all } t \text{ such that } v_i(t_i) = a \text{ for all } i \in S, \\ x_i(t) & \text{otherwise.} \end{cases}$$

A coalitional voting game for coalition S corresponding to a status-quo x and a mechanism $((M_i), g)$ can now be denoted by $\Gamma_x(S, (M_i)_{i \in S}, g)$. This is the formulation first used by Holmström and Myerson [7], for the grand coalition, in defining their notion of durability. And as they pointed out, this way of formulating a voting game is more general than it may first seem. It includes any complex voting procedure for various alternatives, as long as each agent has the option, at the outset, to anonymously veto the new proposal and revert to the status-quo. The mechanism can then be interpreted as the normal form of the game following the accept/reject vote.

⁸ Cramton and Palfrey [1] consider the opposite case: if an alternative is accepted no manipulation is possible but a rejection yields information that can be used in manipulating the “status-quo.”

⁹ Note that $\phi(\sigma(t))$ does not necessarily belong to \mathcal{A}_S since $(x_i(t))_{i \in S}$ need not belong to \mathcal{A}_S .

A Bayesian Nash equilibrium of the voting game Γ_x is a strategy profile σ , for members of S such that, for all $i \in S$ and $t_i \in T_i$,

$$U_i(\phi_i(\sigma)|t_i) \geq U_i(\phi_i(\sigma_{-i}, \sigma'_i)|t_i) \quad \text{for all } \sigma'_i : T_i \mapsto \{a, r\} \times M_i.$$

In a direct mechanism $((T_i), y)$, a *truthful equilibrium* is an equilibrium where $m_i(t_i) = t_i$ for all $i \in S$ and all $t_i \in T_i$.

Given a Bayesian Nash equilibrium $\sigma = ((v_i(t_i), m_i(t_i)))_{i \in S}$, define for each $i \in S$ the set of types who vote a , i.e.,

$$E_i(\sigma) = \{t_i \in T_i | v_i(t_i) = a\}.$$

Let $E(\sigma) = \prod_{i \in S} E_i(\sigma) \times T_{-S}$ denote the states in which the alternative is adopted.

In a voting game $\Gamma_x(S, (M_i)_{i \in S}, g)$, a Bayesian Nash equilibrium σ is said to be an *equilibrium rejection* of x if there is positive probability that in equilibrium all agents vote to accept the alternative mechanism, i.e., $q(E(\sigma)) > 0$, and $\phi(\sigma)$ is not interim equivalent to x for all $i \in S$.

A status-quo x is *resilient to coalitional voting* if there does not exist a voting game $\Gamma_x(S, (M_i)_{i \in S}, g)$, with an equilibrium rejection of x .

Proposition 2. *Given $x \in \mathcal{A}^*$ and a coalition S , the following statements are equivalent:*

- (a) *There is an equilibrium rejection of x in voting game $\Gamma_x(S, (M_i)_{i \in S}, g)$.*
- (b) *There is a truthful equilibrium rejection of x in a (direct) voting game $\Gamma_x(S, (T_i)_{i \in S}, y)$.*
- (c) *There exists $y' \in \mathcal{A}_S$ and $E_i \subseteq T_i$ for all $i \in S$, where $q(E) > 0$ for the product event $E = \prod_{i \in S} E_i \times T_{-S}$, such that:*
 - (i) $U_i(y'|t_i, E) \geq U_i(x|t_i, E)$ for all $i \in S$ and all $t_i \in E_i$, with strict inequality for some i and t_i .
 - (ii) $U_i(y'|t_i, E) \leq U_i(x|t_i, E)$ for all $i \in S$ and all $t_i \notin E_i$.
 - (iii) $y' \in \mathcal{A}_S^*(E)$.

It is instructive to compare this proposition to the revelation principle. The equivalence between (a) and (b) follows from the usual argument, and allows us to restrict attention to direct mechanisms, without loss of generality. It shows that a wide variety of coalitional voting games are strategically equivalent to one in which agents vote and report their types. In terms of Bayesian Nash equilibria, it is unimportant whether the vote and type reports are simultaneous or sequential (votes followed by type reports or vice versa). In equilibrium, acceptance of an alternative mechanism is equivalent to truthful acceptance in a direct voting game.

The equivalence of these with (c) has a more novel interpretation: it can be viewed as a form of a revelation principle concerning mechanisms in the context of a status-quo. The existence of an equilibrium rejection (through *some* mechanism) of a status-quo is equivalent to the existence of a feasible allocation rule satisfying the three inequalities in condition (c): (i) there is a product event over which all members of the coalition gain, (ii) it is reasonable for them to believe this event since those types who do not belong to this event would not vote to accept the alternative mechanism, and (iii) the new proposed mechanism is incentive compatible over the event E .

Dutta and Vohra [4] define $x \in \mathcal{A}^*$ to be in the *credible core* if there does not exist a coalition S , $y' \in \mathcal{A}_S$ and an event $E = \prod_{i \in S} E_i \times T_{-S}$ (where $q(E) > 0$ and $E_i \subseteq T_i$ for all $i \in S$) such that

$$(i) \frac{U_i(y'|t_i, E)}{q(E_{-i}(t_i)|t_i)} > \frac{U_i(x_i|t_i, E)}{q(E_{-i}(t_i)|t_i)} \quad \text{for all } i \in S \text{ and all } t_i \in E_i.$$

- (ii) $\frac{U_i(y'|t_i, E)}{q(E_{-i}(t_i)|t_i)} \leq \frac{U_i(x_i|t_i, E)}{q(E_{-i}(t_i)|t_i)}$ for all $i \in S$ and $t_i \notin E_i$ such that $E_{-i}(t_i) \neq \emptyset$.
- (iii) $y' \in \mathcal{A}_S^*(E)$.

Note that if agent i is of type $t_i \notin E_i$ and $E_{-i}(t_i) = \emptyset$, then her vote has no bearing on the outcome. Thus condition (ii) above is the same as condition c(ii) of Proposition 2. This observation yields the following corollary.

Corollary 2. *x is resilient to coalitional voting if and only if it belongs to the credible core.*

Proof of Proposition 2. Suppose $\sigma = (v_i(t_i), m_i(t_i))$ is an equilibrium of $\Gamma_x(S, (M_i)_{i \in S}, g)$. Let $E_i = \{t_i \in T_i | v_i(t_i) = a\}$, and let $y(t) = g(m(t))$ for all t . The equilibrium interim utility of agent i of type t_i is

$$U_i(\sigma|t_i) = \begin{cases} U_i(y|t_i, E) + U_i(x|t_i, T \setminus E) & \text{if } t_i \in E_i, \\ U_i(x|t_i) & \text{otherwise.} \end{cases}$$

The fact that σ is an equilibrium means:

- (1) An agent $i \in S$ of type $t_i \in E_i$ cannot gain by rejecting the alternative,

$$U_i(y|t_i, E) \geq U_i(x|t_i, E) \quad \text{for all } i \in S \text{ and all } t_i \in E_i, \tag{1}$$

or by continuing to accept but changing the choice of $m_i(t_i)$ to $m_i(t'_i)$,

$$U_i(y|t_i, E) \geq U_i(y, t'_i|t_i, E) \quad \text{for all } i \in S, t_i \in E_i, t'_i \in T_i. \tag{2}$$

- (2) An agent i of type $t_i \notin E_i$ cannot gain by accepting the alternative, and choosing $m_i(t'_i)$,

$$U_i(x|t_i, E) \geq U_i(y, t'_i|t_i, E) \quad \text{for all } i \in S, t_i \notin E_i, t'_i \in T_i. \tag{3}$$

Since agent i of type $t_i \notin E_i$ cannot change the outcome by continuing to reject the alternative, this exhausts all possible unilateral deviations. Thus conditions (1)–(3) characterize an equilibrium of the voting mechanism. If the equilibrium involves a rejection of the status-quo, then in addition to these three conditions we must have $q(E) > 0$ and at least one inequality in (1) being strict. Clearly, then (a) implies (b) as well as (c). Of course, (b) implies (a), and so it remains only to show that (c) implies (b).

To see that (c) implies (b), consider y' and E satisfying condition (c). Let

$$y(t) = \begin{cases} y'(t) & \text{if } t \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Since $U_i(y'|t_i, E) = U_i(y|t_i, E)$ for all $i \in S$ and $t_i \in E_i$, it follows from (i) of condition (c) that y satisfies (1), with at least one strict inequality. From (ii) of condition (c) we know that y' (or y) satisfies (3) for all $t'_i \in E_i$. The fact that y also satisfies it for all $t'_i \notin E_i$ follows from construction. Similarly, (2) follows from (iii) of condition (c). \square

We proceed to give an example that illustrates the difference between the credible core and the incentive compatible fine core. It is a simple version of a “lemon” asymmetric information economy, and it is taken from Vohra [16].

Example 1. Let $N = \{1, 2\}$. Agent 1 is fully informed and has two possible types: $T_1 = \{t_H, t_L\}$. Agent 2 is uninformed about the true type of agent 1 and assigns equal probability to both. Let t_H

also denote the high state and t_L the low state. Agent 1 is the seller of an indivisible good to be traded for money, and agent 2 is the buyer. In state t_H , agents' valuations for the indivisible good are v_1 and v_2 , respectively, while they are 0 in state t_L . We shall assume that $v_2/2 < v_1 < v_2$. Under these assumptions, the credible core consists exclusively of no-trade contracts in state t_H and no transfers of money in state t_L . To see this, suppose there were trade in state t_H . Individual rationality for the seller implies that the money transfer should be at least v_1 . Using incentive compatibility, one can establish that the same transfer must happen in state t_L , which makes it impossible for the buyer to meet his individual rationality constraint. Note how any fine objection (constructed in state t_H) would not be credible because type t_L would like to join. However, as just observed, any credible core contract would not be in the incentive compatible fine core.

4. Mediated voting/blocking

The purpose of the next two sections is to connect our approach, in which information transmission occurs endogenously in equilibria of voting games, with a core concept recently proposed in Myerson [12]. Myerson derives his concept from very different considerations rooted in the idea of virtual utility. It is instructive to obtain Myerson's virtual utility core from our approach by adding two layers of generality to our voting games. First, in this section we continue to consider deterministic coalition formation—in each of the coalitional voting games considered, the players of a fixed coalition S are the only active players—but the type reports are used by the blocking mediator to construct objections over events that do not necessarily have a product structure. The next section will add to this the possibility of random coalitions.

There is one respect in which the voting mechanism of the previous section is *not* general enough. While agents are allowed a lot of flexibility in choosing a communication mechanism, the fact that the application of the alternative mechanism is based on independent accept/reject decisions may turn out to be an important restriction. One can argue that if a mediator can be used to translate messages into actions, presumably the mediator could also be delegated the accept/reject decision itself (perhaps as a function of additional inputs into the mechanism).

Recall that in the voting game $\Gamma_x(S, (M_i)_{i \in S}, g)$ of the previous section, the outcome function of the mechanism (M, g) maps from M to \mathcal{A}_S . However, an outcome of the game includes the possibility that the alternative is rejected, and the status-quo survives. A more general communication mechanism would allow the outcome function to map from M to $\mathcal{A}_S \cup \{x\}$.¹⁰ This is the interpretation of a mechanism we shall adopt in this section. As before, an agent votes for or against a new mechanism and also chooses a message to communicate in the mechanism. Given a status-quo x , coalition S and a mechanism (M, g) , where $g : \mapsto \mathcal{A}_S \cup \{x\}$, we now denote the coalitional voting game by $\Gamma'_x(S, (M_i)_{i \in S}, g)$.

Given a strategy profile σ , let $\phi(\sigma)$ denote the corresponding outcome for coalition S . Let $E(\sigma)$ denote the states in which the mechanism effectively adopts a new alternative, i.e.,

$$E(\sigma) = \{t \in T \mid \phi(\sigma(t)) \neq x(t)\}.$$

Note that $t \in E(\sigma)$ implies that $v_i(t_i) = a$ for all $i \in S$, but the converse may not hold, i.e., acceptance of an “alternative” by all agents in S does not necessarily mean that the mechanism accepts it. Indeed, there may be states in which, despite all types involved accept the alternative, the status-quo is implemented by the mechanism: it is as if over such states the given potential blocking plan is “turned off.”

¹⁰ There is some abuse of notation here because rejecting the new mechanism leads to $x(t)$ in state t .

Given an event E , let

$$E_i = \{t_i \in T_i | (t_i, t_{-i}) \in E \text{ for some } t_{-i} \in T_{-i}\}.$$

Note that now an event E need not be a product event.

In a voting game $\Gamma'_x(S, (M_i)_{i \in S}, g)$, a Bayesian Nash equilibrium σ is said to be an *equilibrium rejection* of x if $q(E(\sigma)) > 0$ and $\phi(\sigma)$ is not interim equivalent to x for all $i \in S$.

A status-quo x is *resilient to mediated coalitional voting* if there does not exist a voting game $\Gamma'_x(S, (M_i)_{i \in S}, g)$ with an equilibrium rejection of x .

Proposition 3. *Given $x \in \mathcal{A}^*$ and a coalition S , the following statements are equivalent:*

- (a) *There is an equilibrium rejection of x in a mediated voting game $\Gamma'_x(S, (M_i)_{i \in S}, g)$.*
- (b) *There is a truthful equilibrium rejection of x in a (direct) mediated voting game $\Gamma'_x(S, (T_i)_{i \in S}, y)$.*
- (c) *There exists $y' \in \mathcal{A}_S$ and $E \subseteq T$ (with $q(E) > 0$) such that:*
 - (i) $U_i(y'|t_i, E) \geq U_i(x|t_i, E)$ for all $i \in S$ and all $t_i \in E_i$, with strict inequality for some i and t_i .
 - (ii) $\sum_{t_{-i} \in E_{-i}(t_i)} q(t) [u_i(y_i(t), t) - u_i(x_i(t), t)] \geq \sum_{t_{-i} \in E_{-i}(t'_i)} q(t) [u_i(y_i(t_{-i}, t'_i), t) - u_i(x_i(t), t)]$ for all $i \in S, t_i, t'_i \in T_i$.

Again, we can use the equivalence between (a) and (b) to claim that multiple versions of the coalitional voting game are strategically equivalent. One version that we will find convenient in terms of fixing ideas is the following. Using the words of Myerson [12], suppose that the “blocking plan” constructed by the “blocking mediator” for coalition S involves departing from the status-quo only over an event $E_S \times T_{-S}$ for some $E_S \subseteq T_S$. Effectively, E_S is the set of type profiles for S over which the blocking mediator “turns on” the objection to abandon the status-quo, while elsewhere the objection is “turned off.” In the game, each agent $i \in S$ reports his type $t_i \in T_i$ and takes an action $m_i \in M_i$. Then, the mediator implements an outcome different from x if and only if the reported types belong to E_S . Of course, for us an E_S supporting an objection will be found endogenously, associated with an equilibrium of the voting game.

Proof of Proposition 3. Suppose $\sigma = (v_i(t_i), m_i(t_i))$ constitutes an equilibrium rejection of x in the voting game $\Gamma'_x(S, (M_i)_{i \in S}, g)$. Let $E(\sigma)$ be the positive probability event over which the outcome of σ is not x , and let $y(t) = \phi(m(t))$ for all t . Of course, $v_i(t_i) = \{a\}$ for all $t_i \in E_i(\sigma)$. Moreover, without loss of generality, one can modify, if necessary, the outcome function of the mechanism, so that following a change from σ to σ' consisting only in having all types vote a , the outcome is still different from x on $E(\sigma)$. It is easy to see that σ' is an equilibrium rejection of x in the modified game. It now follows that in the direct mediated mechanism $\Gamma'_x(S, (T_i)_{i \in S}, y)$, it is an equilibrium for all types of agents in S to accept, and report truthfully. To see this, note that voting to reject the alternative mechanism yields the status-quo—an option that was feasible in the original mechanism. A unilateral deception also yields an outcome that is feasible in the original mechanism. Clearly then, (a) and (b) are equivalent.

Suppose σ is a truthful equilibrium rejection of x in the game $\Gamma'_x(S, (T_i)_{i \in S}, y)$. This means that:

- No $i \in S$ of type t_i can gain by rejecting the alternative mechanism. For any $t_i \notin E_i(\sigma)$ the accept/reject decision is outcome equivalent. For $t_i \in E_i(\sigma)$, we must have $U_i(y|t_i, E) \geq$

$U_i(x|t_i, E)$. Thus,

$$U_i(y|t_i, E) \geq U_i(x|t_i, E) \quad \text{for all } i \in S \text{ and all } t_i \in E_i,$$

with strict inequality for some i of some type.

- No $i \in S$ of type t_i can gain by pretending to be of type t'_i (incentive compatibility). The interim utility from truthful reporting is

$$U_i(\sigma|t_i) = \sum_{t_{-i} \in E_{-i}(t_i)} q(t) u_i(y_i(t), t) + \sum_{t_{-i} \notin E_{-i}(t_i)} q(t) u_i(x_i(t), t).$$

If agent i of type t_i reports t'_i , the resulting interim utility is ¹¹

$$U_i(\sigma, t'_i|t_i) = \sum_{t_{-i} \in E_{-i}(t'_i)} q(t) u_i(y_i(t_{-i}, t'_i), t) + \sum_{t_{-i} \notin E_{-i}(t'_i)} q(t) u_i(x_i(t), t).$$

And we must have

$$U_i(\sigma|t_i) \geq U_i(\sigma, t'_i|t_i) \quad \text{for all } i \in S \text{ for all } t_i, t'_i \in T_i.$$

This general incentive compatibility constraint can also be expressed in a form that compares utilities only over the relevant subset of $E(\sigma)$. To do this, subtract $U_i(x|t_i)$ from the right-hand side of both $U_i(\sigma|t_i)$ and $U_i(\sigma, t'_i|t_i)$ to re-state the above inequality as

$$\begin{aligned} & \sum_{t_{-i} \in E_{-i}(t_i)} q(t) [u_i(y_i(t), t) - u_i(x_i(t), t)] \\ \geq & \sum_{t_{-i} \in E_{-i}(t'_i)} q(t) [u_i(y_i(t_{-i}, t'_i), t) - u_i(x_i(t), t)] \end{aligned} \tag{IC}$$

for all $i \in S, t_i, t'_i \in T_i$.

This proves that (b) implies (c). The converse follows easily from the same arguments as above by considering the direct mediated mechanism y' where

$$y'(t) = \begin{cases} y(t) & \text{if } t \in E, \\ x(t) & \text{otherwise.} \end{cases} \quad \square$$

To compare Proposition 3 with Proposition 2, consider $x \in \mathcal{A}^*$ and y, E satisfying condition (c) of Proposition 3. For $t'_i \notin E_i, E_{-i}(t'_i) = \emptyset$ (or more precisely, consists of types that have 0 probability given t_i). The RHS of (c.ii) is then 0, and the condition becomes the same as (c.i). Thus, we need only to consider (c.ii) for those cases in which $t'_i \in E_i$.

It may now be useful to write (c.ii) explicitly for the remaining two cases:

- (i) $t_i \notin E_i$ and $t'_i \in E_i$: In this case, the LHS is 0, and (c.ii) becomes the same as (c.ii) of Proposition 2; the self-selection condition used in defining a credible objection.
- (ii) $t_i \in E_i$ and $t'_i \in E_i$: In this case, if E has a product structure, $E_{-i}(t_i) = E_{-i}(t'_i)$ and it is easy to see that (c.ii) of Proposition 3 is the same as (c.iii) of Proposition 2.

Thus, we see that part (c) of Proposition 2 implies (c) of Proposition 3, and the converse holds if E is a product event. This leads us to define the *mediated core* as the set of all allocations in

¹¹ Note how the different type report affects the implementation of the blocking plan, but not that of the status-quo.

A^* for which there does not exist a coalition S , and event E and an allocation $y \in \mathcal{A}_S$ satisfying condition (c) of Proposition 3. Clearly then, the mediated core is a subset of the credible core. Recall the analogy from the introduction: a credible objection is to the independent choice of actions in a Nash equilibrium just like a mediated objection is to the action choice in a correlated equilibrium. The next example illustrates the difference between credible core and mediated core, and shows that the mediated core can be a strict subset of the credible core.

Example 2. Consider an exchange economy with three agents and three commodities. Agent 3 is uninformed and agents 1 and 2 have two types each: $T_1 = \{t_1, t'_1\}$, $T_2 = \{t_2, t'_2\}$, and each state is equally likely. Let $t = (t_1, t_2)$ and $t' = (t'_1, t'_2)$. The state independent endowments are

$$\omega_1 = (1, 0, 0), \quad \omega_2 = (0, 1, 0), \quad \omega_3 = (0, 0, 2).$$

The utility functions are

$$\begin{aligned} u_1((z_1, z_2, z_3), t) &= 0.9z_1 + z_2 + 1.05z_3, \\ u_1((z_1, z_2, z_3), t') &= 5z_3, \\ u_1((z_1, z_2, z_3), (t'_1, t_2)) &= u_1((z_1, z_2, z_3), (t_1, t'_2)) = 0.9z_1 + z_2, \\ u_2((z_1, z_2, z_3), t) &= z_1 + 0.9z_2 + 1.05z_3, \\ u_2((z_1, z_2, z_3), t') &= 5z_3, \\ u_2((z_1, z_2, z_3), (t'_1, t_2)) &= u_2((z_1, z_2, z_3), (t_1, t'_2)) = z_1 + 0.9z_2, \\ u_3((z_1, z_2, z_3), s) &= z_1 + z_2 \quad \text{for all } s. \end{aligned}$$

The status-quo allocation x , along with ex-post utilities, is the following:

	t_2	t'_2
t_1	$x_1 = (0, 0, 1); \quad u_1 = 1.05$ $x_2 = (0, 0, 1); \quad u_2 = 1.05$ $x_3 = (1, 1, 0); \quad u_3 = 2$	$x_1 = (0, 0.8, 0); \quad u_1 = 0.8$ $x_2 = (1, 0.1, 0); \quad u_2 = 1.09$ $x_3 = (0, 0.1, 2); \quad u_3 = 0.1$
t'_1	$x_1 = (0.1, 1, 0); \quad u_1 = 1.09$ $x_2 = (0.8, 0, 0); \quad u_2 = 0.8$ $x_3 = (0.1, 0, 2); \quad u_3 = 0.1$	$x_1 = (0, 0, 1); \quad u_1 = 5$ $x_2 = (0, 0, 1); \quad u_2 = 5$ $x_3 = (1, 1, 0); \quad u_3 = 2$

This allocation rule is incentive compatible, individually rational and interim efficient. Furthermore, coalition $\{1, 2\}$ does not have a coarse objection because type t'_1 and t'_2 cannot be better off without commodity z_3 . Also, it is easy to see that coalitions $\{1, 3\}$ or $\{2, 3\}$ do not have a coarse objection: there are not enough units of the first two goods to improve upon player 3's expected utility. This shows that x is in the incentive compatible coarse core.

Moreover, there is no credible objection to x either. To see this suppose there is a credible objection by agents 1 and 2 over some event E . If the event is t or t' they cannot do better. If the event is (t'_1, t_2) they can do better, but then type t_1 will want to lie; the preferences of agent 1 are the same in (t_1, t_2) and (t'_1, t_2) . The event E cannot be the first column either, because agent 1 will have to be given utility at least 1.09 in each state (to make sure that she is better-off in state (t'_1, t_2) , and to maintain incentive compatibility). But then t_2 would be worse off, getting no more than a utility of 0.9 in each state. For similar reasons, the argument holds for state (t_1, t'_2) , or for

the first row. This exhausts all possibilities for a credible objection. Thus x belongs to the credible core.

Now consider a mediated objection in which the mediator will use the mechanism δ described below for coalition $\{1, 2\}$ in states other than t' . (We know that there does not exist a credible objection over the event T , so it is not important what the allocation rule for the coalition specifies in state t' .) As we shall see, the event over which the objection is “turned on” is $E_{\{1,2\}} = T \setminus \{t'\}$, while in state t' the objection is “turned off” and the status-quo is to be implemented, despite the fact that types t'_1 and t'_2 would be voting to accept the plan δ in the corresponding equilibrium rejection of x . Here is the mediated objection:

	t_2	t'_2
t_1	$y_1 = (0, 1, 0); u_1 = 1$ $y_2 = (1, 0, 0); u_2 = 1$	$y_1 = (0, 0.895, 0); u_1 = 0.895$ $y_2 = (1, 0.105, 0); u_2 = 1.0945$
t'_1	$y_1 = (0.105, 1, 0); u_1 = 1.0945$ $y_2 = (0.895, 0, 0); u_2 = 0.895$	

One can show that there exists an equilibrium rejection of x in the voting game $\Gamma'_x(\{1, 2\}, (T_i)_{i \in \{1,2\}}, y)$ where the event E over which there is a departure from the status-quo is $T \setminus \{t'\}$. That is, we claim that σ is an equilibrium rejection of x , where each type of agents 1 and 2 accepts the alternative to the status-quo, reports his type truthfully, and $E(\sigma) = T \setminus \{t'\}$. We proceed to the analysis of such an equilibrium rejection of x .

Type t_1 prefers to accept the alternative, given that the two types of agent 2 are also accepting it. His expected utility from the blocking plan is higher than at the status-quo: he gets a utility equal to 1 in state (t_1, t_2) and equal to 0.895 in state (t_1, t'_2) (instead of status-quo utilities 1.05 and 0.8, respectively). In the words of Myerson [12], this type of agent 1 receives the call of the blocking mediator with probability 1 in this blocking plan. Thus, he maintains the same beliefs to sustain a higher expected utility in the blocking plan.

On the other hand, type t'_1 , who also wants to accept the alternative, compares the different bundles offered by the blocking plan and the status-quo only in state (t'_1, t_2) (because $t' \notin E(\sigma)$). In that state the blocking plan utility is 1.0945, higher than the status-quo utility of 1.09. The arguments for types t_2 and t'_2 are identical to those for t_1 and t'_1 , respectively.

It only remains to check that each type does not have an incentive to change his type report in the voting game. This amounts to showing incentive compatibility over the event $E(\sigma)$ of the blocking plan.

Let us begin with type t_1 . Again, by being truthful in equilibrium, his expected utility will be $(\frac{1}{2})(1 + 0.895)$. If he misreports his type by announcing t'_1 to the mediator, he will be invited to the blocking plan only with probability $\frac{1}{2}$, when the type of player 2 is t_2 , and he will then be offered to consume $(0.105, 1, 0)$, whereas with the rest of probability he will not be invited to the blocking plan and the status-quo will result. However, in this case, his false report will not interfere with the status-quo since, to implement it, the very first (truthful) reports are used. In conclusion, in this case he is allocated $(0, 0.8, 0)$. The corresponding expected utility is $(\frac{1}{2})(1.0945 + 0.8)$, smaller than what he gets by being truthful in equilibrium. In other words, type t_1 does not gain by lying, because the gain in the first column is 0.0945, but the loss in the second column is 0.095.

Finally, let us check the incentives of type t'_1 . If he reports truthfully, with probability $\frac{1}{2}$ he will be invited to the objection and be allocated $(0.105, 1, 0)$ [in state (t'_1, t_2)], while with probability

$\frac{1}{2}$ he will be allocated (0, 0, 1) [in state (t'_1, t'_2) from the status-quo, where he really values good z_3]. This is better than being allocated (0, 1, 0) and (0, 0.895, 0), respectively. Of course, the incentive compatibility arguments for types t_2 and t'_2 are identical to the corresponding types of player 1, and we omit them.

Deviations from σ consisting of an untruthful type report and rejection of the alternative have already been taken care of, since in this case the untruthful report is not taken into account, for the implementation of the status-quo. Hence, we have established that σ is an equilibrium rejection of x that sustains δ as a mediated objection over the event $E(\sigma) = T \setminus \{t'\}$.

Since the blocking inequalities we have written are all strict, we can conclude that x is not in the (weak) mediated core, and therefore not in the mediated core either.

5. Randomized mediation

In this section, we generalize the voting game of the previous section by considering a mechanism that involves an even more sophisticated role for the mediator. A mediator may construct a blocking plan in which the coalition that is asked to vote on a new alternative mechanism is chosen at random. This corresponds closely to the core concept in Myerson [12].¹²

Now, a proposal by a mediator, μ , consists of a probability distribution of feasible allocation mechanisms for various coalitions. In particular, $\mu(S, y^S, t)$, where $y^S \in \mathcal{A}_S$, denotes the probability with which coalition S is invited by the mediator to vote for y against the status-quo, when the (reported) state is t . Recall that for ease of comparison with the previous sections we will consider randomization only over coalitions, i.e., $\mu(S, y^S, t) > 0$ implies $\mu(S, z^S, t) = 0$ for all $z^S \neq y^S$. Thus, we associate with each coalition S one proposed allocation $y^S \in \mathcal{A}_S$. The mediator may choose with positive probability not to invite any coalition, thereby imposing the status-quo. Thus, for each $t \in T$, $0 \leq \mu(S, y^S, t) \leq 1$ for all coalitions S , and $\sum_S \mu(S, y^S, t) \leq 1$. This describes a “blocking plan” used in the definition and characterization of the *inner core*; see Myerson [11], Qin [14] and de Clippel and Minelli [3].

In the following paragraphs we describe a voting game, an extension of the game Γ'_x of the previous section. As before, at the outset, nature chooses state t with probability $q(t)$. Each player $i \in N$ of type t_i is endowed with his interim beliefs $q(t_{-i}|t_i)$. Also, an incentive compatible status-quo x is given, and truthful type reports (to be used in the implementation of x if the alternative is rejected) have already been sent prior to the voting game. Given Propositions 2 and 3, it should be clear that without loss of generality we may concentrate on direct voting games in which agents vote and report their types.

In each game of the previous sections, the set of active players was a fixed coalition S . Since now coalitions are chosen at random in each state, one has to specify when each player is “active” in the voting game, which will correspond to when he is invited by the blocking mediator. Suppose that coalitions are called according to the plan μ . Then, each player $i \in N$ has an action set $T_i \times \{a, r\}$ only if there exists $S \supseteq \{i\}$ such that $\mu(S, y^S, t) > 0$. Otherwise, agent i is not an active player in the voting game (his action set is empty). At the time an active agent i has to move, he knows only his type and the fact that there is a realization of the blocking plan μ in which he is invited. Since such realizations depend on the type reports, each player must form beliefs about how likely each of these realizations is: in equilibrium, he will expect that all types are being truthful, and hence he will use the true μ to perform all his expected utility calculations. Thus, all active agents,

¹² To maintain comparability with the previous sections we neglect the possibility that the alternative mechanism may be a random state contingent allocation. Myerson [12] allows for this form of randomization as well.

according to μ , vote whether to accept or reject μ and report their types to the mediator, who is forming coalitions according to the plan μ . If all voters accept, and given their type reports, say t , the mediator chooses coalition S with probability $\mu(S, y^S, t)$. If there is at least one rejection, the status-quo survives. Note that at the time of taking the vote each agent must, given the strategies of the others, use μ and Bayes' rule to update his beliefs at his information set. This allows him to figure out what is the probability that each S he is a member of is being invited, and therefore that he will receive each y^S .

Note that (1) if $\mu(S', y^{S'}, t) = 0$ for all coalitions $S' \neq S$, and (2) if $\mu(S, y^S, t) = 1$ for states $t \in E_S \times T_{-S}$, while $\mu(S, y^S, t) = 0$ for $t \notin E_S \times T_{-S}$, for some $E_S \subseteq T_S$, the voting game just described reduces to a game $\Gamma'_x(S, (T_i)_{i \in S}, y)$ of the previous section, in which the set of states $E_S \times T_{-S}$ is that event over which this non-random alternative is “turned on.”

Based on the interpretation of a blocking mediator given by this voting game with random coalitions, we are led to modifying the notion of mediated core of the previous section as follows.

An allocation $x \in \mathcal{A}^*$ is said to belong to the *randomized mediated core* if there does not exist a blocking plan μ such that

$$\sum_{t_{-i}} q(t) \sum_{S \supseteq \{i\}} \mu(S, y^S, t) [u_i(y_i^S(t), t) - u_i(x_i(t), t)] \geq 0 \quad \text{for all } i \in N \text{ and } t_i \in T_i$$

and

$$\begin{aligned} & \sum_{t_{-i}} q(t) \sum_{S \supseteq \{i\}} \mu(S, y^S, t) [u_i(y_i^S(t), t) - u_i(x_i(t), t)] \\ & \geq \sum_{t_{-i}} q(t) \sum_{S \supseteq \{i\}} \mu(S, y^S, t_{-i}, t'_i) [u_i(y_i^S(t_{-i}, t'_i), t) - u_i(x_i(t), t)] \\ & \text{for all } i \in N \text{ and } t_i, t'_i \in T_i. \end{aligned}$$

Suppose μ is a deterministic blocking plan that only uses coalition S and an S -allocation y over an event E . It follows that μ satisfies the above inequalities if and only if (S, y, E) satisfies condition (c) of Proposition 3. Equilibrium rejections of x given a deterministic blocking plan in the voting game of this section are precisely the equilibrium rejections of x in games Γ'_x defined in the previous section. Since in addition we can have equilibrium rejections associated with random blocking plans, we have:

Observation. The randomized mediated core is a subset of the mediated core, itself a subset of the credible core.

Therefore, as in the previous section, we can define resilience to randomized mediated voting by considering random alternative mechanisms μ and get an alternative characterization of the randomized mediated core. To avoid repetitions, we will omit the corresponding statements.

The concept of randomized mediated core of the present section is very similar to the core concept defined by Myerson [12]. The only differences are that: (1) he also allows for random allocation rules within each coalition, and (2) he assumes that transfers are possible and feasibility is weakened to require expected feasibility of the transferable commodity. Clearly, in terms of a definition of blocking, feature (1) may be significant, as the feasible set for coalitions is the set of lotteries over allocations, while feature (2) is a technical assumption on the environments that Myerson [12] uses to prove existence.

We close the section by going over another example, to illustrate the difference between the randomized mediated core and other cores previously mentioned.

Example 3. This example is an adaptation of Myerson's [12] Example 1 to our framework. The set of agents is $N = \{1, 2\}$, and agent 2 is uninformed. Let $T_1 = \{t_H, t_L\}$ be the set of types of agent 1. Both types are equally likely. Agent 1 is a seller of good z_1 , to be exchanged for money (good z_2). The state-independent endowment is

$$\omega_1 = (1, 0), \quad \omega_2 = (0, 10).$$

The utility functions are as follows:

$$u_1((z_1, z_2), t_H) = 5z_1 + z_2,$$

$$u_1((z_1, z_2), t_L) = z_1 + z_2,$$

$$u_2((z_1, z_2), t_H) = 6z_1 + z_2 - 10,$$

$$u_2((z_1, z_2), t_L) = 2z_1 + z_2 - 10.$$

Consider the following allocation x :

$$x_1(t_H) = (0.75, 1.3), \quad x_2(t_H) = (0.25, 8.7) \quad x_1(t_L) = (0, 2.05), \quad x_2(t_L) = (1, 7.95).$$

This allocation is interim incentive efficient and interim individually rational. Therefore, it is in the incentive compatible coarse core. Furthermore, because agent 1 is fully informed, it is easy to see that any allocation in the incentive compatible coarse core is also in the credible core. Since in this economy all events over which objections can be constructed have a product structure, there are no additional mediated objections. Hence, the mediated core also coincides with the credible core. Therefore, x is in the mediated core of the previous section.

Let us see now that x is not in the randomized mediated core. Consider the following blocking plan μ : $\mu(\{1, 2\}, y^{(1,2)}, t_H) = 0.25$, $\mu(\{1\}, y^{(1)}, t_H) = 0.75$, while $\mu(S, y^S, t_L) = 0$ for all S , where

$$y_1^{(1,2)}(t_H) = (0, 5.2), \quad y_2^{(1,2)}(t_H) = (1, 4.8) \quad y_1^{(1)}(t_H) = (1, 0).$$

In words, the blocking plan μ only concerns state t_H . In it, coalition $\{1, 2\}$ will be formed with probability 0.25 and exchange the entire unit of good z_1 for 5.2 units of money, while with probability 0.75 agent 1 will be instructed to keep his initial endowment.

In the corresponding voting game associated with μ , both agents 1 and 2 are "active" players. The following profile is a Bayesian Nash equilibrium: let each agent accept the blocking plan and report his type truthfully. Let us check this:

- Agent 1 of type t_L is being excluded from the blocking plan μ . Therefore, the expected utility that he receives from playing his equilibrium strategy is the status-quo utility 2.05. Changing his vote to a rejection would not alter this outcome. Finally, misreporting his type and pretending to be type t_H would "turn on" the blocking plan, but this would yield an expected utility $0.25 \cdot 5.2 + 0.75 \cdot 1 = 2.05$, so he is at a best response by telling the truth.
- Agent 1 of type t_H joins the blocking plan in this equilibrium, and his expected utility from doing so is $0.25 \cdot 5.2 + 0.75 \cdot 5 = 5.05$, which is also his status-quo expected utility. Thus, he would not benefit from rejecting the blocking plan. Misreporting his type would "turn off" the blocking plan, also resulting in the status-quo. Thus, he is also at a best response.

- As for agent 2, he updates his interim beliefs using μ and the equilibrium strategy of agent 1 conditioning on the fact that coalition $\{1, 2\}$ is being formed. His conditional expected utility from accepting the blocking plan is simply $6 - 5.2$. That is, he learns from having to vote that the state is t_H and that coalition $\{1, 2\}$ has been called to exchange one unit of the first good for 5.2 units of money; in no other realization of μ he is part of the blocking plan. Therefore, if he accepts the blocking plan when invited, he ends up with the bundle $(1, 4.8)$, which is better than what he would get from the status-quo by rejecting (the bundle $(0.25, 8.7)$).

Note finally how in this example the weak inequalities in the objection make a big difference. That is, while we have shown that x is not in the randomized mediated core, we claim that it is in its weak version defined by objections that use strict inequalities for every type. To see this, note that because of incentive compatibility, if one improves type t_H strictly, type t_L will also want to join. This means that he needs to be part of the blocking plan and also strictly improve, which renders improving the buyer impossible.

6. Relationship to durability

We close the paper by relating our approach to durability. In their analysis of interim efficiency, Holmström and Myerson [7] point out that agents may sometimes discard an interim (incentive) efficient allocation in favor of another allocation.¹³ Their notion of *durability* is meant to formalize the following:

‘The essential idea is that an incentive compatible decision rule δ should be considered *durable* iff the individuals in the economy would never unanimously approve a change from δ to any other decision rule.’ (Holmström and Myerson ([7, p. 1811])

Their actual definition of durability proceeds as follows. Given an allocation x and an alternative feasible allocation y , they consider a voting game $\Gamma_x(N, (T_i)_{i \in N}, y)$ as in our Section 3.2, in which type reports are made after it is determined whether or not y is unanimously accepted. An allocation x is said to *endure* y if there exists an equilibrium in which y is rejected in every state. Since there always exists a Bayesian equilibrium in which everyone rejects the alternative because they do not expect to be pivotal, it becomes necessary to refine the equilibrium notion. They do this by appealing to some of the conditions used in defining a trembling hand perfect equilibrium. An allocation $x \in \mathcal{A}^*$ is said to be *durable* if it endures every alternative mechanism in \mathcal{A} . Thus, a durable allocation has the property that for every alternative, there exists an equilibrium rejection of that alternative.

To compare our approach with durability, we apply the stability notion implicit in Γ_x of Section 3.2 to the grand coalition. An allocation $x \in \mathcal{A}^*$ is said to be *resilient to grand coalition voting*, or simply *resilient* if there is no equilibrium rejection of x in a voting game $\Gamma_x(N, (T_i)_{i \in N}, y)$.

Using the proof of Proposition 2 applied to $S = N$, it is easy to see that every resilient allocation is interim incentive efficient. In general, the other implication does not hold; to see this, note that one can find an event E , proper subset of T , over which the relevant types enjoy an improvement. It is true, though, that a uniformly incentive compatible interim incentive efficient allocation is resilient (see Dutta and Vohra [4]).

¹³ This phenomenon should be seen as an important consequence of incomplete information since it is, of course, impossible in the complete information setting.

The reader will notice that our notion of resilience seems to be more faithful to the idea described by the above Holmström–Myerson quote. A resilient allocation has the property that there is no equilibrium in which the alternative is ever unanimously accepted in place of the given status-quo. To meet the durability test it is enough that there exist a (trembling-hand perfect) equilibrium such that in every positive probability state the alternative is rejected. It does not rule out the possibility that there is also an equilibrium in which the alternative is accepted. Clearly, if there exists no equilibrium (perfect or not) in which the alternative is accepted with positive probability, then provided an equilibrium exists, it must involve a rejection of the alternative in each (positive probability) state. In this sense, a resilient allocation is durable. But the converse need not be true. This can be illustrated by the example in Section 9 of Holmström and Myerson [7] in which an inefficient allocation is durable because there is an equilibrium in which the status-quo endures the alternative. But there also exists, for an alternative that is an interim improvement, an equilibrium in which the alternative is unanimously accepted. Of course, this implies that the original allocation was not resilient. The basic idea in our approach, given our interest in a cooperative solution such as the core, is that a coalition may choose any feasible allocation as long as it can be supported by an equilibrium. In particular, a coalition should be able to resolve any coordination problem (potentially found in situations like those in Holmström and Myerson’s [7] example of Section 9). Compared to durability, this makes it easier to challenge a status-quo, and yields therefore a smaller set of stable allocations. Not surprisingly, this may lead to problems of non-existence. But in studying the core, unlike efficient allocations, this is a problem that we cannot avoid in general, anyway (see Vohra [16]).

Acknowledgments

We thank an Associate Editor and a referee for helpful comments. We acknowledge support from NSF Grant SES-0133113. Serrano also thanks Universidad Carlos III, CEMFI and the Institute for Advanced Study for their hospitality, and Fundación Banco Herrero, Universidad Carlos III and Deutsche Bank for research support.

References

- [1] P. Cramton, T. Palfrey, Ratifiable mechanisms: learning from disagreement, *Games Econ. Behav.* 10 (1995) 255–283.
- [2] G. de Clippel, The type-agent core of exchange economies with asymmetric information, *J. Econ. Theory*, in press
- [3] G. de Clippel, E. Minelli, Two remarks on the inner core, *Games Econ. Behav.* 50 (2005) 143–154.
- [4] B. Dutta, R. Vohra, Incomplete information, credibility and the core, *Math. Soc. Sci.* 50 (2005) 148–165.
- [5] F. Forges, J.-F. Mertens, R. Vohra, The ex ante incentive compatible core in the absence of wealth effects, *Econometrica* 70 (2002) 1865–1892.
- [6] F. Forges, E. Minelli, R. Vohra, Incentives and the core of an exchange economy: a survey, *J. Math. Econ.* 38 (2002) 1–41.
- [7] B. Holmström, R. Myerson, Efficient and durable decision rules with incomplete information, *Econometrica* 51 (1983) 1799–1819.
- [8] D. Lee, O. Volij, The core of economies with asymmetric information: an axiomatic approach, *J. Math. Econ.* 38 (2002) 43–63.
- [9] E. Maskin, J. Tirole, The principal-agent relationship with an informed principal. Part 2: common values, *Econometrica* 60 (1992) 1–42.
- [10] R. Myerson, Mechanism design by an informed principal, *Econometrica* 51 (1983) 1767–1798.
- [11] R. Myerson, *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, MA, 1991.
- [12] R. Myerson, Virtual utility and the core for games with incomplete information, Mimeo, University of Chicago, 2005.

- [13] M. Perry, P. Reny, A non-cooperative view of coalition formation and the core, *Econometrica* 62 (1994) 795–817.
- [14] C.-Z. Qin, The inner core and the strictly inhibitive set, *J. Econ. Theory* 59 (1993) 431–444.
- [15] R. Serrano, R. Vohra, Non-cooperative implementation of the core, *Soc. Choice Welfare* 14 (1997) 513–525.
- [16] R. Vohra, Incomplete information incentive, compatibility and the core, *J. Econ. Theory* 86 (1999) 123–147.
- [17] O. Volij, Communication, credible improvements and the core of an economy with asymmetric information, *Int. J. Game Theory* 29 (2000) 63–79.
- [18] R. Wilson, Information, efficiency and the core of an economy, *Econometrica* 46 (1978) 807–816.