Virtual Utility and the Core for Games with Incomplete Information

by Roger B. Myerson

February 2005, revised August 2006

Abstract. The core is extended to games with incomplete information. The feasible set is characterized by incentive-compatible mechanisms. Blocking is organized at the interim stage by an incentive-compatible mediation plan. Membership of the blocking coalition itself may be determined randomly by the blocking mediator. Nonemptiness of an interim fine core is proven for games with a balanced structure, independent types, and sidepayments. An offer of severance payments may be needed to inhibit blocking. Core allocations are characterized in terms of virtual-utility scales that generalize the weighted-utility scales of the inner core. Mechanisms that achieve core allocations are coalitionally durable.

JEL classification numbers C71, D82

(Running title: Virtual-Utility Core)

Author's address:

Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637.

Email: myerson@uchicago.edu

Phone: 1-773-834-9071. Fax: 1-773-702-8490

1. Introduction

We consider here the question of how to extend our definition of the core to games with incomplete information. The core is intuitively defined as the set of payoff allocations such that no coalition could make all its members better off. As such, the core is the most natural formulation of coalitional rationality in cooperative game theory, and it has a broad equivalence with Walrasian equilibria in large market games. Thus, the question of how to extend the core to games where players have different information seems fundamental to information economics, and it has regularly attracted effort from economic theorists for decades. Forges, Minelli, and Vohra [3] offer an excellent survey of the many papers which have been written on this question. For other recent contributions in this area, see Forges, Mertens, and Vohra [4], Lee and Volij [7], Serrano and Vohra [19], and de Clippel [1].

In defining a core for games with incomplete information, there are three basic distinctions that must be made about how the players' information may be used. First, in evaluating whether a coalition can make all its members better off, we must ask whether players' welfare should be evaluated at the ex ante stage, before any player learns his private information about his type, at the interim stage, when each player has learned his own type but not the types of others, or at the ex post stage, after all players' types have become common knowledge (see Holmström and Myerson, 1983). Second, in defining what a coalition can do, we might assume that a coalition's blocking plan can use all the type-information of all its members, leading to a fine definition of the core, in the terminology of Wilson [23]; or we might assume that a coalition can only use the information that is common knowledge among all its members, leading to what Wilson called a coarse core. Third, when we assume that members of a coalition can share information, we must specify whether their communication is verifiable, or subject to incentive constraints that no player should have any incentive to lie about his type.

In this conceptual framework, the core concept that is developed in this paper may be characterized as an incentive-compatible interim fine core. We concentrate on the interim stage, because, in Harsanyi's [5] general formulation of Bayesian games with incomplete information, each player's type is defined to be the random variable that summarizes all of the player's private information at the start of the game. We require incentive compatibility, because the hardest part of information economics is the analysis of the difficulties that people have in trusting each other's statements about their private information.

The core that we develop here can also be distinguished from other definitions in that it is a generalized version of the inner core that is defined by Shapley's [21] method of fictitious transfers, a method which has been used quite generally to extend cooperative solution concepts from games with transferable utility (TU) to games with nontransferable-utility (NTU). In the method of fictitious transfers, we consider rescaled versions of the game, in which each player's utility function is multiplied by some positive weight. For each rescaled version, we identify the allocations which would satisfy some given TU solution concept when the weighted utilities are assumed to be transferable among the players. Then we say that such allocations are solutions for the original NTU game when they would actually be feasible without such fictitious transfers.

Although this method of fictitious transfers was first applied to define the NTU Shapley value, Qin [18] and Myerson [14; 16, Section 9.8] have shown that the inner core which this method generates can be naturally interpreted in terms of randomized coalition formation. That is, we consider the possibility that a blocking coalition might be organized by a mediator who does not tell the players which other players are being invited into the coalition. The players are assumed to know the probability distribution that the blocking mediator is using to choose his random coalition, and so each invited player can update his beliefs about who else has been invited by Bayes rule, given his own inclusion in the coalition. Could such randomized blocking eliminate from the core some allocations that would not be blocked without such randomization? Qin and Myerson showed that, although the answer to this question is No when utility is transferable, it becomes Yes when utility is not transferable. Not only does randomized blocking yield a smaller core for NTU games, but it actually simplifies our analysis by linearizing the inequalities that characterize coalitional objections against an allocation. So a straightforward duality argument shows that these randomized blocking conditions are equivalent to the inner core that is defined by the method of fictitious transfers. This paper extends this result to the case of incomplete information.

With incomplete information, the role of weighted utility is taken by virtual utility, which is defined by a formula that takes informational incentive constraints into account. This author has explored the use of virtual utility as a mathematical tool for understanding cooperation under uncertainty in several earlier papers. Myerson [9, 13]considered negotiations that are completely controlled by one player who has private information, [10] formulated a generalized Nash bargaining solution for two-person bargaining problems, [11] formulated a generalized Shapley NTU-value, and [15] formulated a core-like equilibrium concept for dynamic matching problems. These past formulations are reviewed in Myerson [16, Chapter 10]. Virtual utility arises in these solutions in a subtle and complex way that may become clearer here in the case of the core, because the rationale for the core is simpler and more intuitive than the Nash bargaining solution or Shapley value.

The great drawback of the core is that it may be empty. To get existence, we consider balanced games with sidepayments and independent types. Balanced games are the games where cores are nonempty in the complete-information case, and they include exchange economies with linear utility as well as all two-person games. Transferring utility by sidepayments is a well-known simplifying assumption in cooperative game theory. Any planned sidepayments here will be required to satisfy incentive compatibility, and thus our model may be described as a game with sidepayments, but not with fully transferable utility (when the latter is taken to mean that any type-contingent plan of sidepayments would be feasible). Assuming independent types simplifies our notation, but it actually makes incentive compatibility more difficult to satisfy (because a player's type claim cannot be tested against correlated information of other players). We are not assuming independent private values here: we allow that a player's utility payoff can depend on other players' types.

Our existence theorem also relies on two other assumptions that enlarge the feasible set: side-bets and severance payments. We assume that a risk-neutral mediator may offer the players side-bets about each others' types, as long as the mediator's expected profit is nonnegative. This possibility allows a kind of weak feasibility-in-expectation. But such side-bets will be constrained by a requirement that they must never create adverse-selection problems for the mediator (in the sense that the players could incentive-compatibly cheat the mediator, by deviating to a blocking coalition in type-states where the mediator would have gained from the side-bets, so that the mediator is left with an expected loss when the players do not deviate). We also assume that the mediator can also promise nonnegative severance payments that would be paid to a player who joins a blocking coalition. Such severance payments could never expand the core of a game with complete information. But with incomplete information, a promise of severance payment to a "bad" type who joins a blocking coalition could effectively undermine the blocking coalition, by exacerbating its adverse-selection problems.

Such severance payments would require a kind of centralization in the mediator's plan, so that a player who claims the severance payment can then be excluded from rejoining the cooperative coalition (see Myerson, 1995). So core allocations that rely on such promises of severance payments may not correspond to decentralizable economic equilibria in large market games. In this sense, the equivalence of core and competitive equilibrium may not extend to the case of incomplete information (see also Serrano, Vohra, and Volij, 2001, and Forges, Heifetz, and Minelli, 2001).

2. Formulation of the game

Let N denote the set of players. We let i denote a generic player in N, and we let S denote a generic nonempty subset of N. For each $S \subseteq N$, let C(S) denote the set of feasible joint actions for coalition S. Let T_i denote the set of possible types of player i. We assume that these sets N, C(S), and T_i are all nonempty and finite.

For each possible type t_i in T_i, let $p_i(t_i)$ denote the probability that player i is type t_i. As the players' types are assumed to be independent random variables, we may write

$$p(t_S) = \prod_{i \in S} p_i(t_i), \ \ \forall S \subseteq N, \ \ \forall t_S \in T_S;$$

$$p(t_{-i}) = \prod_{j \in N-i} p_j(t_j), \ \ \forall i \in N, \ \ \forall t \in T;$$

$$p(t) = \prod_{j \in N} p_j(t_j), \ \ \forall t \in T.$$

Here we use the notation

$$t_S = (t_i)_{i \in S} \in T_S = \times_{i \in S} T_i;$$

$$t_{-i} = (t_j)_{j \in N-i} \in T_{-i} = T_{N-i};$$

$$t = t_N = (t_i)_{i \in N} = (t_{-i}, t_i) \in T = T_N.$$

For any t in T, any $S \subseteq N$, any i in S, and c in C(S), let $u_i(c,t)$ denote the utility payoff to player i from joining coalition S when S implements its feasible actions c given that all players' types are as in t. We are assuming here that, when members of a coalition S choose actions jointly feasible for them, the actions of others outside S do not matter to them, but types of others might matter to them. For any finite set X, we let $\Delta(X)$ denote the set of probability distributions over X.

A collective-choice mechanism is a pair of functions $(\mu:T\rightarrow\Delta(C(N)),\ x:T\rightarrow\mathbb{R}^N)$. Here $\mu(c|t)$ represents the probability of choosing joint action c when the players' types are t, and $x_i(t)$ denotes the expected net monetary sidepayment to player i when the players' types are t. The expected utility of player i under mechanism (μ,x) is

$$U_i(\mu,x|t_i) = \sum_{t_{-i}\in T_{-i}} p(t_{-i})\,[x_i(t) + \sum_{c\in C(N)} \mu(c|t)\,u_i(c,t)].$$

If player i has type t_i but pretends to have type r_i in mechanism (μ,x), then his expected utility is

$$\hat{U}_i(\mu,x,r_i|t_i) = \sum_{t_{-i}\in T_{-i}} p(t_{-i})\,[x_i(t_{-i},r_i) + \sum_{c\in C(N)} \mu(c|t_{-i},r_i)\,u_i(c,t)].$$

A mechanism (μ,x) is incentive compatible iff

$$U_i(\mu,x|t_i) \geq \hat{U}_i(\mu,x,r_i|t_i),\ \ \forall i\in N,\ \forall t_i\in T_i,\ \forall r_i\in T_i.$$

That is, if a trustworthy mediator implemented the mechanism (μ,x) by asking each player to confidentially report his type, then honest reporting by all players would be an equilibrium. By the revelation principle, there is no loss of generality in restricting our attention to such incentive-compatible mechanisms.

The mediator's expected payoff from the mechanism (μ,x) is $-\sum_{t\in T} p(t) \sum_{i\in N} x_i(t)$. Let us say that a mechanism (μ,x) is feasible iff it is incentive compatible and yields a nonnegative expected payoff to the mediator. That is, a feasible mechanism must also satisfy the constraint

$$\sum_{t\in T} p(t) \sum_{i\in N} x_i(t) \leq 0.$$

3. Incentive efficiency and virtual utility

A mechanism is (weakly) incentive-efficient iff it is feasible and no other feasible mechanism yields higher expected utilities for all types of all players. By convexity of all constraints, a feasible mechanism $(\bar{\mu},\bar{x})$ is incentive-efficient iff there exists some vector

$$\lambda = (\lambda_i(t_i))_{i\in N, t_i\in T_i}$$

such that $\lambda_i(t_i) \geq 0\ \ \forall i\in N,\ \forall t_i\in T_i$, with at least one strict inequality, and

(1) $(\bar{\mu},\bar{x})$ maximizes $\sum_{i\in N} \sum_{t_i\in T_i} \lambda_i(t_i)\, U_i(\mu,x|t_i)$ over all feasible mechanisms (μ,x).

The optimization problem is a linear-programming problem in (μ,x). Let $\alpha_i(r_i|t_i)$ be the Lagrange multiplier (or dual variable) for the constraint that type t_i should not gain by reporting r_i. Then the Lagrangean for this optimization problem can be written

(2) $$L(\mu,x,\lambda,\alpha) = \sum_{i\in N} \sum_{t_i\in T_i} \Big(\lambda_i(t_i)\, U_i(\mu,x|t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i)(U_i(\mu,x|t_i) - \hat{U}_i(\mu,x,r_i|t_i))\Big) .$$

For any joint action c, and types-profile t, and any vectors λ and α as above, the <u>virtual utility</u> for player i from action c with types t and parameters λ and α is defined to be

(3) $$v_i(c,t,\lambda,\alpha) = [(\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i))\, u_i(c,t) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)\, u_i(c,(t_{-i},r_i))]/p_i(t_i).$$

With this definition, the above Lagrangean can be rewritten

(4) $$L(\mu,x,\lambda,\alpha) = \sum_{t\in T} p(t) \sum_{c\in C(N)} \mu(c|t) \sum_{i\in N} v_i(c,t,\lambda,\alpha)$$
$$+ \sum_{t\in T} p(t) \sum_{i\in N} x_i(t)[(\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i)) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)]/p_i(t_i).$$

A maximization of this Lagrangean is possible only if the coefficients of $x_i(t)$ here are constant over all i and all t, and this constant can be set equal to 1 without loss of generality. Thus, standard Lagrangean analysis yields the following fact.

<u>Theorem 0.</u> A feasible mechanism (μ,x) is incentive-efficient iff there exist vectors λ and α such that:

$\lambda_i(t_i) \geq 0$ and $\alpha_i(r_i|t_i) \geq 0$, $\forall i\in N$, $\forall t_i\in T_i$, $\forall r_i\in T_i$,

(5) $\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i) = p_i(t_i)$, $\forall i\in N$, $\forall t_i\in T_i$,

(6) $\alpha_i(r_i|t_i)[U_i(\mu,x|t_i) - \hat{U}_i(\mu,x,r_i|t_i)] = 0$, $\forall i\in N$, $\forall t_i\in T_i$, $\forall r_i\in T_i$ (complementary slackness),

(7) $\mu(c|t) > 0$ implies $c \in \text{argmax}_{d\in C(N)} \sum_{i\in N} v_i(d,t,\lambda,\alpha)$, $\forall t\in T$, $\forall c\in C(N)$.

Let us say that a type t_i of player i <u>jeopardizes</u> another type r_i of player i, in the incentive-efficient mechanism μ, iff the constraint that t_i should not want to imitate r_i ($U_i(\mu,x|t_i) \geq \hat{U}_i(\mu,x,r_i|t_i)$) is binding and has a positive Lagrange multiplier in the optimality conditions for μ. Multiplying a type's utility function by a positive constant is decision-theoretically inessential. So the essential difference between virtual utility $v_i(c,t,\lambda,\alpha)$ and actual utility $u_i(c,t)$ is that virtual utility of type t_i exaggerates the difference from the utilities of i's other types that jeopardize t_i. Thus, we have the following general proposition to help us to qualitatively understand the ex-post inefficiencies (signaling costs) that may be incurred in an incentive-efficient mechanism:

The incentive-efficient mechanism will be ex-post efficient in terms of the players' virtual utilities, where the virtual utility of any type t_i differs from the actual utility by exaggerating the difference from the other possible types that jeopardize t_i.

In this sense, a Coasian believer in ex-post efficiency could "explain" inefficient signaling by the virtual-utility hypothesis that, when incentive constraints bind, players act according to their virtual utilities.

Equation (5) can be interpreted as a system of hydraulic equations, for flow in a network that contains a node for each type t_i. If we interpret $p_i(t_i)$ as flow into the network at t_i, $\lambda_i(t_i)$ as flow out of the network at t_i, and $\alpha_i(r_i|t_i)$ as flow from t_i to r_i, then equation (5) says that these flows balance at each node. (See also Malakhov and Vohra, 2004.) Equation (5) implies

$$\sum_{t_i \in T_i} \lambda_i(t_i) = \sum_{t_i \in T_i} p_i(t_i) = 1,$$

and so some $\lambda_i(t_i)$ must be strictly positive for each player i.

4. Blocking

In the theory of the core, we think about an established plan that can inhibit players from joining any alternative blocking coalition. With incomplete information, we should think about an established mediator who implements a mechanism that can inhibit players from deviating to cooperate with some other blocking mediator.

In the theory of the core, players compare a blocking plan with the established plan under an assumption that any player who rejects an invitation to block gets his established-plan allocation, even if others accept the blocking coalition (established payoffs are guaranteed, no bank runs). To justify this assumption, we must assume that, if any one player rejects an invitation to block, then all players must stay in the established plan. That is, there is no blocking without unanimity among all invited blockers. (Of course rejections from players who have not been invited do not count.)

But with incomplete information and interim coalition formation, we must also consider the possibility that other players might accept the blocking coalition only for certain types, and so a player who offered to block and was then returned by someone else's refusal would learn new information that might enable him to find profitable opportunities to lie in the established plan.

So to justify the core assumption that established payoffs are guaranteed, we should think about the blocking question being raised after the players have sent in their reports to the established plan, but before they are committed to implement the established plan.

We allow that the blocking mediator may invite different coalitions according to some known randomized plan, and so the probability of any particular coalition blocking and choosing some joint action can depend on their information. But it may seem unreasonable to allow the blocking coalition to depend on the information of other players outside this coalition. So we assume that the blocking mediator can ask any random set S about their types and, based on their responses, must either invite all of S into a blocking coalition or invite no blocking coalition.

To characterize such a blocking plan, for any action c in C(S) and any types t_S in T_S, we may let $\nu(S,c|t_S)$ represent the probability that coalition S would be invited to block and do their jointly-feasible action c if these players report the types t_S. In our description of the monetary sidepayments that a blocking mediator could offer, we only need to specify the expected sidepayment for each possible type of each player (because the additive separability of utility for money means that players do not care how their sidepayments are correlated with the random coalition). So for each type t_i of each player i, we may let $y_i(t_i)$ denote the blocking mediator's expected net sidepayment to player i if i would be willing to help block and report type t_i to the blocking mediator. Thus, we define a <u>blocking plan</u> to be any pair of vectors (ν,y) such that:

$$y_i(t_i) \in \mathbb{R},\ \ \forall i\in N,\ \forall t_i\in T_i,$$

$$\nu_S(c|t_S) \geq 0,\ \ \forall c\in C(S),\ \forall t_S\in T_S,\ \ \text{and}$$

(8) $$\textstyle\sum_{S\subseteq N} \sum_{c\in C(S)} \nu_S(c|t_S) \leq 1,\ \ \forall t\in T.$$

Constraint (8) says that, given any profile of types t, the probability of forming a blocking coalition can never be greater than one. This condition actually plays almost no role in our analysis. In fact, we would have no loss of generality if we tightened this constraint and considered only blocking plans that satisfy the constraint

(9) $$\textstyle\sum_{S\subseteq N} \sum_{t_S\in T_S}\sum_{c\in C(S)} \nu_S(c|t_S) \leq 1.$$

With constraint (9), the blocking plan (ν,y) could be implemented as follows. First, according to the probability distribution ν, the mediator chooses a random coalition $S \subseteq N$, a random profile of types $t_S \in T_S$, and a random action c in C(S). Next the mediator asks each of the players in S whether he is willing to join the blocking coalition and, if so, what his type is. If the players in S all express willingness to join the blocking coalition and their reported types match the profile t_S then the blocking mediator forms the coalition S and implements their jointly feasible action c. But if anyone in S refuses to block or has a type different from his type in t_S then the blocking mediator tells them all to remain in the established plan. When a blocking coalition forms, the planned monetary sidepayment to a player i in the blocking coalition could be depend on the blocking coalitions S, their types t_S, and their action c, according to any function $\hat{y}_i(c,t_S)$ such that

$$\sum_{S\supseteq\{i\}} \sum_{t_{S-i}\in T_{S-i}} \sum_{c\in C(S)} \nu_S(c|t_S)\hat{y}_i(c,t_S) = y_i(t_i).$$

(Here we use the notation tS = (t_{S-i},t_i).)

Let $\omega_i(t)$ denote the payoff allocation from the established plan that player i would lose if he joined a blocking coalition when the types are t. Let $\omega = (\omega_i(t))_{i\in N,t\in T}$. A tenable blocking plan must give players an incentive to accept an invitation to block, and then to report honestly to the blocking mediator. So the blocking plan (ν,y) is <u>tenable</u> against ω iff it satisfies

(10) $y_i(t_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \nu_S(c|t_S)(u_i(c,t)-\omega_i(t)) \geq 0, \ \forall i\in N, \forall t_i\in T_i,$

(11) $y_i(t_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \nu_S(c|t_S)(u_i(c,t)-\omega_i(t))$

$\geq y_i(r_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \nu_S(c|t_{S-i},r_i)(u_i(c,t)-\omega_i(t)), \ \forall i\in N, \forall t_i\in T_i, \forall r_i\in T_i,$

(12) $-\sum_{i\in N} \sum_{t_i\in T_i} p_i(t_i)\, y_i(t_i) > 0.$

Conditions (10) says that each type of each player should get a nonnegative expected gain from accepting an invitation into the blocking coalition, and condition (11) say that no type t_i of any player i should be able to get a higher expected gain from reporting some false type r_i, when all other invited players are expected to accept the blocking mediator and report their types honestly. That is, conditions (10)-(11) make it an equilibrium the players to all accept any invitation from the blocking mediator and then report honestly to the blocking mediator, when their losses from leaving the established plan are as specified by ω. Condition (12) says that the blocking mediator should get a positive expected profit from the blocking plan (ν,y). (The addition of constraint (9) above would be without loss of generality in the sense that anything that is blocked by a tenable blocking plan that satisfies constraint (8) can also be blocked by some tenable blocking plan that also satisfies condition (9).)

5. Example 1

To illustrate such random blocking with incomplete information, it may be helpful now to consider a simple example.

In Example 1, there are two players, with N = {1,2}. Player 1 is seller of a single good, and player 1's may be H or L, that is $T_1 = \{H,L\}$. Player 2 is the only potential buyer, and has no private information. We assume that player 1's types are equally likely, so p(H) = p(L) = 1/2. If 1's type is L then the value of the good is $1 to player 1 and is $2 to player 2. If 1's type is H then the value of the good is $5 to player 1 and is $6 to player 2. Each player alone has can only choose a "no trade" action (keeping his initial endowment), but together the players can also choose to trade the good. We may normalize payoffs from the no-trade option to be 0, so

$$u_i(\text{no-trade}, t) = 0, \forall i \; \forall t,$$

$$u_1(\text{trade},L) = -1, \;\; u_2(\text{trade},L) = 2, \;\; u_1(\text{trade},H) = -5, \;\; u_2(\text{trade},H) = 6.$$

If player 1 always expected to sell the good to player 2 (who always values it more), then incentive compatibility would require that the price he gets must be independent of his reported type. But player 2 would not pay more than the ex-ante expected value of the good to player 2, 0.5(2+6) = $4, which is less than the value of the good to 1's type H. So incentive compatibility and participation constraints in this game imply that the probability of trade must be strictly less than 1 when player 1's type is H. Some positive probability of H not trading is necessary as a random signaling cost.

Let us consider a specific mechanism (μ,x) that might seem reasonable:

$$\mu(\text{trade}|L) = 1, \;\; x_1(L) = 1.5 = -x_2(L), \;\; \mu(\text{trade}|H) = 1/9, \;\; x_1(H) = 5.5/9 = -x_2(H).$$

That is, if 1's type is L then with probability 1 they trade at price $1.50, but if 1's type is H then with probability 1/9 they trade at the price $5.50. This mechanism has the intuitively appealing property that, whenever trade occurs, the price gives each player an equal gain from trade. With these type-dependent prices, the probability of trade cannot be higher than 1/9 when 1's type is H, because $(5.5-1)(1/9) = (1.5-1)(1)$ so that any higher probability would tempt 1's type L to report H. This mechanism is incentive-efficient, and yields the expected payoff allocation:

$$\omega_1(H) = 1/18 = \omega_2(H), \;\; \omega_1(L) = 0.50 = \omega_2(L).$$

It is easy to see what virtual utility would support the incentive-efficiency of this mechanism. Virtual utility for player 2 is the same as actual utility, because player 2 has only one type. Furthermore, 1's type L is not jeopardized by his type H, because $U_1(\mu,x|H) > \hat{U}_1(\mu,x,L|H)$, and so $\alpha_1(L|H) = 0$. So 1's type L also must have virtual utility that is the same as actual utility. The only difference between virtual utility and actual utility is for 1's type H, which is jeopardized by L. But notice that trade and no-trade each have a positive probability in this mechanism when 1's type is H, and so the ex-post virtual efficiency condition (7) requires that the players' total virtual gains from trade must be 0 when $t_1 = H$. That is, when t_1=H, the virtual value of the good to player 1 must equal the \$6 value of the good to player 2:

$$[\lambda_1(H)(5)-\alpha_1(H|L)(1)]/0.5 = 6.$$

The hydraulic equations (5) give us

$$\lambda_1(H) - \alpha_1(H|L) = 0.5,\ \ \lambda_1(L) + \alpha_1(H|L) = 0.5,\ \ \lambda_2 = 1.$$

So the (λ,α) that support this mechanism as incentive-efficient in Theorem 0 must be

$$\alpha_1(L|H) = 0,\ \ \alpha_1(H|L) = 1/8,\ \ \lambda_1(H) = 5/8,\ \ \lambda_1(L) = 3/8,\ \ \lambda_2 = 1.$$

In fact, it can be shown that these same (λ,α) parameters also satisfy Theorem 0 for all incentive-efficient mechanisms in this example. (See Myerson, 1985.)

Now consider the implications of the virtual-utility hypothesis, that players who are bargaining subject to incentive constraints will act as if they were maximizing their virtual utilities. If type H of player 1 acts according to his virtual value of the good, which is \$6, then he would never agree to sell the good for \$5.50. So the virtual-utility hypothesis suggests that this mechanism should not be in the core, even though it is incentive efficient. Of course, the reader who feels no intellectual commitment to the virtual-utility hypothesis might be unimpressed by this argument.

But indeed there are random blocking mechanisms that are tenable against the allocation ω that is generated by this mechanism. For example, consider the following blocking plan: The blocking mediator always asks player 1 his type. If player 1 reports H then, with probability 8/9, {1} blocks alone and keeps the good, but with probability 1/9, {1,2} block together, the good is traded, player 2 pays \$5.90, player 1 is paid \$5.50, and the blocking mediator keeps \$0.40 profit. If player 1 reports $t_1 = L$, then the blocking mediator sends them back to the established plan. This blocking plan can be represented by (ν,y) such that

$\nu_{\{1\}}(\text{no-trade}|H) = 8/9$, $\nu_{\{1,2\}}(\text{trade}|H) = 1/9$, $y_1(H) = 5.5/9$, $y_2 = 5.9/18$,

$\nu_S(c|L) = 0$ $\forall S, \forall c$.

Player 1 is willing to participate honestly in the blocking plan, whenever invited, because it treats him the same as the given mechanism (μ,x). Conditional on being invited to join the blocking coalition, player 2 knows that 1's type is H and that the blocking plan would give player 2 the profit $(6-5.90)$ = \$0.10, which is strictly better than the expected profit $\omega_2(H) = (6-5.50)(1/9)$ in the established plan. So player 2, when invited, will also be willing to participate in this blocking plan. Ex ante, the blocking plan makes 2 worse off, but that calculation is irrelevant when player 2 is given the option to join the blocking coalition. The key is that the blocking plan only invites player 2 when the random signaling cost is not applied, so that player 2 is willing to accept a higher price (closer to \$6), thus allowing a positive profit for the blocking mediator.

Tenable blocking plans like this could be similarly constructed against any incentive-efficient mechanism in which the price when player 2 buys the good from 1's type H would be strictly less than \$6. That is, any mechanism where type H of player 1 is not paid his virtual value for selling the good can be blocked by a randomized blocking plan. For an example of a virtually-equitable incentive-efficient mechanism that cannot be blocked in this way, consider the following: if 1's type is L then with probability 1 they trade at price \$1.50, but if 1's type is H then with probability 0.1 they trade at the price \$6, so that no-trade has probability 0.9 when $t_1 = H$, but player 1 never sells at a price below his virtual value of the good.

This relationship between virtual utility and the existence of tenable blocking plans is quite general, as we show in Theorem 1.

6. Virtual utility and inhibitive allocations

We say a utility-allocation vector ω is inhibitive iff there does not exist any blocking plan (ν,y) that is tenable against ω. Let $V_i(\omega,t,\lambda,\alpha)$ be the transformation of player i's payoffs in ω into virtual utility in state t, according to the formula (3) with parameters λ and α. That is, let

$$V_i(\omega,t,\lambda,\alpha) = \Big[(\lambda_i(t_i)+\textstyle\sum_{r_i\in T_i} \alpha_i(r_i|t_i))\omega_i(t) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)\omega_i(t_{-i},r_i)\Big]/p_i(t_i)\ . \tag{13}$$

Theorem 1 An allocation vector ω is inhibitive iff there exist vectors λ and α such that

(5) $$\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i) = p_i(t_i), \ \forall i\in N, \forall t_i\in T_i,$$

(14) $$\sum_{t_{N-S}\in T_{N-S}} p(t_{N-S}) \sum_{i\in S} V_i(\omega,t,\lambda,\alpha) \geq \sum_{t_{N-S}\in T_{N-S}} p(t_{N-S}) \sum_{i\in S} v_i(c,t,\lambda,\alpha),$$

$$\forall S\subseteq N, \forall c\in C(S), \forall t_S\in T_S,$$

$$\lambda_i(t_i) \geq 0 \ \text{ and } \ \alpha_i(r_i|t_i) \geq 0, \ \forall i\in N, \ \forall t_i\in T_i, \ \forall r_i\in T_i,$$

That is, ω is inhibitive iff there exist parameters λ and α such that, for any coalition S, the sum of virtual utilities that the members of S can expect with an action that is feasible for them, given all their information, is not more than the virtual-utility transformation of what they expect from ω.

Proof of Theorem 1. Consider the following linear-programming problem:

(15) choose (ν,y) to minimize $\sum_{i\in N} \sum_{t_i\in T_i} p_i(t_i)\, y_i(t_i)$ subject to the constraints

$$y_i(t_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \nu_S(c|t_S)(u_i(c,t)-\omega_i(t)) \geq 0, \ \forall i\in N, \forall t_i\in T_i,$$

$$y_i(t_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \nu_S(c|t_S)(u_i(c,t)-\omega_i(t))$$

$$\geq y_i(r_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \nu_S(c|t_{S-i},r_i)\,(u_i(c,t)-\omega_i(t)),$$

$$\forall i\in N, \forall t_i\in T_i, \forall r_i\in T_i,$$

$$\nu_S(c|t_S) \geq 0, \ \forall c\in C(S), \ \forall t_S\in T_S, \ \text{ and } \ y_i(t_i) \in \mathbb{R}, \ \forall i\in N, \ \forall t_i\in T_i.$$

The allocation ω is inhibitive iff this linear-programming problem has an optimal value 0. (Constraint (8) in the definition of a tenable blocking plan can be omitted because, if we find any feasible solution of the above problem that yields a negative value of the objective, then some small positive multiple of the this feasible solution would be a tenable blocking plan.)

By the duality theorem of linear programming, this linear-programming problem has optimal value 0 iff its dual also has optimal value 0. Let $\lambda_i(t_i)$ denote the dual variable for the first constraint line in (15), and let $\alpha_i(r_i|t_i)$ denote the dual variable for second constraint line in (15). Then the dual problem can be written:

(16) choose (λ,α) so as to maximize 0 subject to

$$\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i) = p_i(t_i),\ \ \forall i\in N,\ \forall t_i\in T_i,$$

$$\sum_{t_{N-S}\in T_{N-S}} p(t) \sum_{i\in S} [(\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i))\omega_i(t_i) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)\omega_i(t_{-i},r_i)]/p_i(t_i)$$
$$\geq \sum_{t_{N-S}\in T_{N-S}} p(t) \sum_{i\in S} [(\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i))u_i(c,t) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)u_i(c,(t_{-i},r_i))]/p_i(t_i)$$
$$\forall S\subseteq N,\ \forall c\in C(S),\ \forall t_S\in T_S,$$

$$\lambda_i(t_i) \geq 0 \ \text{ and } \ \alpha_i(r_i|t_i) \geq 0,\ \forall i\in N,\ \forall t_i\in T_i,\ \forall r_i\in T_i,$$

(In this dual problem, the first constraint line corresponds to the primal variables $y_i(t_i)$, and the second constraint line corresponds to the primal variables $\nu_S(c|t_S)$ in (15).) So ω is inhibitive iff these inequalities can be satisfied. But the virtual-utility formulas in equations (3) and (13) reduce the second constraint line in (16) to condition (14) in the theorem QED

We may say that an inhibitive allocation ω is finely supported by (λ,α) iff they satisfy

(17) $$\sum_{i\in S} V_i(\omega,t,\lambda,\alpha) \geq \sum_{i\in S} v_i(c,t,\lambda,\alpha),\ \ \forall t\in T,\ \forall S\subseteq N,\ \forall c\in C(S).$$

By a similar duality argument, these fine-support conditions imply that the blocking mediator could not expect to make money even if the action of any blocking coalition S could also depend on the information of other players in $N-S$. Allowing the blocking mediator to talk to all players would mean replacing the variable $\nu_S(c|t_S)$ in (15) by $\nu_S(c|t)$ (that is, $\nu_S(c|t_N)$). Then the second constraint line in the dual (16) would correspondingly become condition (17) (quantifying over all t, instead of all t_S, and dropping the summation over t_{N-S}). We may say that ω is finely inhibitive when this condition (17) holds.

7. The core

As mentioned in the introduction, we will allow the established mediator to promise a nonnegative severance payment to any player who departs to join a blocking mediator. Such severance payments could never expand the core with complete information, but they will be essential for our existence proof in the case of incomplete information. So let $\varepsilon_i(t)$ denote the severance payment that player i would get from the established mediator if player i joined a blocking coalition after types t had been reported to the established mediator.

We say that a payoff allocation ω is achievable by (μ,x) iff (μ,x) is a feasible mechanism, and there exists a vector of promised severance payments $\varepsilon = (\varepsilon_i(t))_{i\in N,t\in T}$ such that

(18) $\varepsilon_i(t) \geq 0$ and $\omega_i(t) = x_i(t) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t) - \varepsilon_i(t),\ \forall i\in N,\ \forall t\in T.$

Thus, $\omega_i(t)$ is the residual stake in the established plan that player i would lose by joining a blocking coalition in state t. We say an allocation ω is in the core iff ω is inhibitive and achievable by some feasible mechanism (μ,x).

Even with complete information, the core is empty for many games. Standard proofs of nonempty cores generally rely on some kind of balancedness assumption about the game. So let us say that $\theta = (\theta_{S,c})_{S\subseteq N,c\in C(S)}$ is a balanced coalitional plan iff

$$\theta_{S,c} \geq 0,\ \forall S\subseteq N,\ \forall c\in C(S),$$

$$\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \theta_{S,c} = 1,\ \forall i\in N.$$

(Such balanced coalitional plans could be naturally interpreted in a process where the game is repeated over time, one copy of each player i arrives at each period, and players depart when they join some coalition that does some jointly feasible action. If we interpret $\theta_{S,c}$ as the rate at which S-coalitions are formed to do action c, then a balanced coalitional plan is one that is sustainable over time, because arrival and departure rates are equal for each player.) We say that the game is balanced iff the grand coalition N can achieve anything that a balanced coalitional plan can achieve, independently of types. That is, balancedness means that, for any balanced coalitional plan θ, there is some randomized strategy $\sigma\in\Delta(C(N))$ such that

$$\sum_{d\in C(N)} \sigma(d)\, u_i(d,t) = \sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \theta_{S,c}\, u_i(c,t),\ \forall i\in N, \forall t\in T.$$

This balancedness condition is always satisfied in games with two players (#N = 2), and it is satisfied in linear exchange games, where the actions of each coalition are exchanges of assets among its members and each player has linearly additive utility for these assets.

Theorem 2. If the game is balanced then the core is nonempty.

Proof of Theorem 2. Given any vectors (λ,α) that satisfy the hydraulic equations (5) in Theorem 0, consider the problem of getting the highest expected net payment for the established mediator with an incentive-compatible mechanism that achieves some inhibitive allocation vector that is finely supported by (λ,α). This problem can be formulated as follows:

(19) choose (μ,x,ω) to minimize $\sum_{t\in T} p(t) \sum_{i\in N} x_i(t)$ subject to

$$p(t_{-i})[x_i(t) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t) - \omega_i(t)] \geq 0,\ \forall i\in N,\ \forall t\in T,$$

$$U_i(\mu,x|t_i) - \hat{U}_i(\mu,x,r_i|t_i) \geq 0,\ \forall i\in N,\ \forall t_i\in T_i,\ \forall r_i\in T_i,$$

$$\sum_{c\in C(N)} \mu(c|t) = 1,\ \forall t\in T,$$

$$\mu(c|t) \geq 0,\ \forall c\in C(N),\ \forall t\in T,$$

$$\sum_{i\in S} V_i(\omega,t,\lambda,\alpha) \geq \sum_{i\in S} v_i(c,t,\lambda,\alpha),\ \forall t\in T,\ \forall S\subseteq N,\ \forall c\in C(S).$$

Notice that the pair (λ,α) are given parameters in the above problem. We will use the Kakutani fixed-point theorem to guarantee the existence of some (λ,α) pair such that the components of these λ and α vectors are equal to the dual Lagrange multipliers of the first and second constraint lines in this linear-programming problem. Then we will show that the solution (μ,x,ω) at this (λ,α)-fixed point is an incentive-compatible mechanism achieving a core allocation.

To begin, let us consider a reduced part of this problem (19) when the vector ω is taken as a given parameter. Then the ω-reduced problem becomes

(20) choose (μ,x) to minimize $\sum_{t\in T} p(t) \sum_{i\in N} x_i(t)$ subject to

$$p(t_{-i})[x_i(t) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t)] \geq p(t_{-i})\, \omega_i(t),\ \forall i\in N,\ \forall t\in T,$$

$$\sum_{t_{-i}\in T_{-i}} p(t_{-i})[x_i(t) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t)]$$
$$- \sum_{t_{-i}\in T_{-i}} p(t_{-i})[x_i(t_{-i},r_i) + \sum_{c\in C(N)} \mu(c|t_{-i},r_i)\, u_i(c,t)] \geq 0,\ \forall i\in N,\ \forall t_i\in T_i,\ \forall r_i\in T_i,$$

$$\sum_{c\in C(N)} \mu(c|t) = 1,\ \forall t\in T,$$

$$\mu(c|t) \geq 0,\ \forall c\in C(N),\ \forall t\in T.$$

It is easy to see that this problem (20) always has an optimal solution. The constraints always have a feasible solution, because we can satisfy them by taking any incentive-compatible (μ,x), and then increasing the $x_i(t)$ payments to each player by a constant amount, independent of his type, until the ω-payoff requirements in the first constraint line are all satisfied. On the other hand, the expected payoff to the mediator $-\sum_{t\in T} p(t) \sum_{i\in N} x_i(t)$ cannot be increased infinitely, because these ω-payoff requirements put a lower bound on each player's expected payoff after sidepayments, and the mediator's expected payoff is the difference between the sum of the players' expected payoffs before monetary sidepayments (which is bounded above from the finiteness of C(N)) and the sum of their expected payoffs after sidepayments.

Thus, for any ω, the dual of this ω-reduced linear-programming problem (20) also has an optimal solution. To formulate this dual, let the dual variables be $\zeta_i(t)$ for the first constraint line in (20), $\beta_i(r_i|t_i)$ for the second constraint line, and $\gamma(t)$ for the third constraint line. Then the dual problem is:

(21) choose (ζ,β,γ) to maximize $\sum_{t\in T}\sum_{i\in N} p(t_{-i})\,\omega_i(t)\,\zeta_i(t) + \sum_{t\in T} p(t)\gamma(t)$ subject to

$$p(t_{-i})[\zeta_i(t) + \textstyle\sum_{r_i\in T_i}\beta_i(r_i|t_i) - \sum_{r_i\in T_i}\beta_i(t_i|r_i)] = p(t),\ \ \forall i\in N,\ \forall t\in T,$$

$$\textstyle\sum_{i\in N} p(t_{-i})[(\zeta_i(t)+\sum_{r_i\in T_i}\beta_i(r_i|t_i))u_i(c,t) - \sum_{r_i\in T_i}\beta_i(t_i|r_i)u_i(c,(t_{-i},r_i))] + p(t)\gamma(t) \le 0,\ \ \forall t\in T,\ \forall c\in C(N),$$

$$\zeta_i(t) \ge 0 \text{ and } \beta_i(r_i|t_i) \ge 0,\ \ \forall i\in N,\ \forall t\in T,\ \forall r_i\in T_i.$$

Here the first constraint line is for the primal variables $x_i(t)$, and the second constraint line is for the primal variables $\mu(c|t)$ in (20). Dividing the first constraint line by $p(t_{-i})$, we get the equivalent constraints:

$$\zeta_i(t) + \textstyle\sum_{r_i\in T_i}\beta_i(r_i|t_i) - \sum_{r_i\in T_i}\beta_i(t_i|r_i) = p_i(t_i),\ \ \forall i\in N,\ \forall t\in T.$$

This implies that $\zeta_i(t)$ is independent of t_{-i}, so there exists $\eta_i(t_i)$ such that

$$\zeta_i(t) = \eta_i(t_i),\ \forall i,\ \forall t\in T.$$

Substituting the vector η for ζ and using the definition of virtual utility v_i, the dual problem (21) can be reformulated equivalently as:

(22) choose (η,β,γ) to maximize $\sum_{t\in T}\sum_{i\in N} p(t_{-i})\omega_i(t)\eta_i(t_i) + \sum_{t\in T} p(t)\gamma(t)$ subject to

$$\eta_i(t_i) + \textstyle\sum_{r_i\in T_i}\beta_i(r_i|t_i) - \sum_{r_i\in T_i}\beta_i(t_i|r_i) = p_i(t_i),\ \ \forall i\in N,\ \forall t_i\in T_i,$$

$$\textstyle\sum_{i\in N} v_i(c,t,\eta,\beta) + \gamma(t) \le 0,\ \forall t\in T,\ \ \forall c\in C(N),$$

$$\eta_i(t_i) \ge 0 \ \text{ and }\ \beta_i(r_i|t_i) \ge 0,\ \ \forall i\in N,\ \ \forall t_i\in T,\ \ \forall r_i\in T_i.$$

This dual problem (22) has an optimal solution for every allocation vector ω. But ω only appears in its objective function, and its constraints do not depend on ω. So the feasible set of (22) has a finite collection of extreme points, independently of ω. Let us number these extreme points from 1 to K and denote them as $\{(\eta^1,\beta^1,\gamma^1), ..., (\eta^K,\beta^K,\gamma^K)\}$. For any ω, the dual problem (22) must have an optimal solution among these extreme points. So for any allocation ω, the optimal value of the ω-reduced problem (20) (the smallest expected subsidy to implement ω) is

$$\text{maximum }\{\textstyle\sum_{t\in T}\sum_{i\in N} p(t_{-i})\omega_i(t)\eta_i^k(t_i) + \sum_{t\in T} p(t)\gamma^k(t)|\ k\in\{1,...,K\}\}.$$

Thus, for any given (λ,α) vectors, our original linear-programming problem (19) has the same optimal value as

(23) choose (ω,z) to minimize z subject to

$$z - \sum_{t\in T} \sum_{i\in N} p(t_{-i})\omega_i(t)\eta_i^k(t_i) \geq \sum_{t\in T} p(t)\gamma^k(t), \ \ \forall k\in\{1,...,K\},$$

$$p(t) \sum_{i\in S} [(\lambda_i(t_i)+\sum_{r_i\in T_i} \alpha_i(r_i|t_i))\omega_i(t) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)\omega_i(t_{-i},r_i)]/p_i(t_i)$$

$$\geq p(t) \sum_{i\in S} v_i(c,t,\lambda,\alpha), \ \ \forall t\in T, \forall S\subseteq N, \forall c\in C(S).$$

To formulate the dual of this problem (23), let ρ^k denote the dual variables for the first constraint line, and let $\theta(S,c|t)$ denote the dual variables for the second constraint line. Then the dual of (23) is:

(24) choose (ρ,θ) to

$$\text{maximize} \sum_{t\in T} p(t) [\sum_{k=1}^K \rho^k \gamma^k(t) + \sum_{S\subseteq N} \sum_{c\in C(S)} \theta(S,c|t) \sum_{i\in S} v_i(c,t,\lambda,\alpha)]$$

subject to $\rho^1 + ... + \rho^K = 1$,

$$\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} [(\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i))\theta(S,c|t) - \sum_{r_i\in T_i}\alpha_i(r_i|t_i)\theta(S,c|t_{-i},r_i)]$$

$$- \sum_{k=1}^K \rho^k \eta_i^k(t_i) = 0, \ \ \forall i\in N, \forall t\in T,$$

$\rho^k \geq 0$, $\forall k\in\{1,...,K\}$, and $\theta(S,c|t) \geq 0$, $\forall S\subseteq N, \forall c\in C(S), \forall t\in T$.

Here the first constraint line corresponds to the primal variable z in (23), and the second constraint line corresponds to the primal variables $\omega_i(t)$ in (23).

Let Λ denote the convex hull of our dual extreme points $\{(\eta^1,\beta^1,\gamma^1), ..., (\eta^K,\beta^K,\gamma^K)\}$. Let us define a point-to-set correspondence $F:\Lambda\rightarrow\rightarrow\Lambda$ so that, for any (λ,α,δ) in Λ, (η,β,γ) is in the set $F(\lambda,\alpha,\delta)$ iff there exists an optimal solution (ρ,θ) of problem (24) with (λ,α) such that

$$\sum_{k=1}^K \rho^k \eta^k = \eta, \ \ \sum_{k=1}^K \rho^k \beta^k = \beta, \ \ \sum_{k=1}^K \rho^k \gamma^k = \gamma.$$

By the Kakutani fixed-point theorem, there exists some (λ,α,γ) in Λ such that

$$(\lambda,\alpha,\gamma) \in F(\lambda,\alpha,\gamma).$$

So let (λ,α,γ) be such a fixed point, and take ω and θ from the optimal solutions of (23) and (24) with this fixed point. Because (λ,α,γ) is feasible for the dual problem (22), the second constraint line in (22) gives us

$$\gamma(t) \leq -\max_{c\in C(N)} \sum_{i\in N} v_i(c,t,\lambda,\alpha), \ \ \forall t\in T.$$

The first constraint line in (22) tells us that (λ,α) satisfy the hydraulic equations (5).

The second constraint line in (24) implies that θ satisfies

$$\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} [(p_i(t_i) + \sum_{r_i\in T_i} \alpha_i(t_i|r_i))\theta(S,c|t) - \sum_{r_i\in T_i} \alpha_i(r_i|t_i)\theta(S,c|t_{-i},r_i)]$$

$$= \lambda_i(t_i) = p_i(t_i) + \sum_{r_i\in T_i} \alpha_i(t_i|r_i)) - \sum_{r_i\in T_i} \alpha_i(r_i|t_i), \ \forall i\in N, \forall t\in T.$$

Thus,

(25) $$(p_i(t_i) + \sum_{r_i\in T_i} \alpha_i(t_i|r_i))[\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \theta(S,c|t) - 1]$$

$$= \sum_{r_i\in T_i} \alpha_i(r_i|t_i)[\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \theta(S,c|t_{-i},r_i) - 1], \ \forall i\in N, \forall t\in T.$$

But these equations imply:

(26) $$\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \theta(S,c|t) = 1, \ \forall i\in N, \forall t\in T.$$

(If this equation (26) ever failed for some i with some t_{-i}, then we could sum equations (25) over all t_i such that $\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \theta(S,c,t) > 1$ [or <1], cancel out the r_i terms that appear on both sides, and get a positive sum equaling negative sum.) So for each t, $\theta(\bullet|t)$ here satisfies the conditions for a balanced coalitional plan. By the assumption that the game is balanced, there exists some $\sigma(\bullet|t)$ in $\Delta(C(N))$ that yields the same utility and virtual utility for every type.

Thus, the optimal value of (24) satisfies

$$\sum_{t\in T} p(t) [\sum_{k=1}^{K} \rho^k \gamma^k(t) + \sum_{S\subseteq N} \sum_{c\in C(S)} \theta(S,c|t) \sum_{i\in S} v_i(c,t,\lambda,\alpha)]$$

$$= \sum_{t\in T} p(t) [\gamma(t,\alpha) + \sum_{i\in N} \sum_{d\in C(N)} \sigma(d|t) v_i(d,t,\lambda,\alpha)]$$

$$\leq \sum_{t\in T} p(t) [-\max_{c\in C(N)} \sum_{i\in N} v_i(c,t,\lambda,\alpha) + \sum_{i\in N} \sum_{d\in C(N)} \sigma(d|t) v_i(d,t,\lambda,\alpha)] \leq 0.$$

But the optimal value of (19) is the same as the optimal value of (24). So we have found a finely inhibitive allocation ω that is achievable by some feasible mechanism (μ,x), which is incentive compatible and yields nonnegative expected profit for the established mediator. Q.E.D.

We now show that the severance payments for core allocations can be positive only for types t_i that have weight $\lambda_i(t_i) = 0$ in the (λ,α) parameters that support the incentive-efficient mechanism. The fact that the mechanism maximizes an interim social welfare function that gives these types zero weight means that, in the mechanism, they are in some sense "getting more than they deserve." That is, we may intuitively expect positive severance offers only to bad types that are getting more than they deserve because of their ability to imitate good types.

Theorem 3. Suppose ω is an inhibitive allocation satisfying the conditions of Theorem 1 for (λ,α), and suppose that (μ,x) is a feasible mechanism that achieves ω. Then

$$\sum_{i\in N} \sum_{t\in T} p(t_{-i})\, \lambda_i(t_i)[\sum_{c\in C(N)} \mu(c|t)\, u_i(c,t) + x_i(t) - \omega_i(t)] = 0.$$

So the severance payment $\varepsilon_i(t) = [\sum_{c\in C(N)} \mu(c|t)\, u_i(c,t) + x_i(t) - \omega_i(t)]$ can be strictly positive only if $\lambda_i(t_i) = 0$.

Proof of Theorem 3. The condition that (μ,x) achieves ω implies that every term in this sum is nonnegative. So it suffices to show that the sum is nonpositive, as follows:

$$\begin{aligned}
&\sum_{i\in N} \sum_{t\in T} p(t_{-i})\, \lambda_i(t_i)\, \omega_i(t) = \\
&\quad = \sum_{i\in N} \sum_{t\in T} p(t)\, [(\lambda_i(t_i) + \textstyle\sum_{r_i\in T_i} \alpha_i(r_i|t_i))\, \omega_i(t) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)\, \omega_i(t_{-i},r_i)]/p_i(t_i) \\
&\quad \geq \sum_{t\in T} p(t) \max_{c\in C(N)} \sum_{i\in N} v_i(c,t,\lambda,\alpha) \\
&\quad \geq \sum_{t\in T} p(t) \sum_{c\in C(N)} \mu(c|t) \sum_{i\in N} (v_i(c,t,\lambda,\alpha) + x_i(t)) \\
&\quad = \sum_{t\in T} p(t_{-i}) \sum_{c\in C(N)} \mu(c|t) \sum_{i\in N} [(\lambda_i(t_i) + \textstyle\sum_{r_i\in T_i} \alpha_i(r_i|t_i))(u_i(c,t)+x_i(t)) \\
&\qquad\qquad - \textstyle\sum_{r_i\in T_i} \alpha_i(t_i|r_i)(u_i(c,(t_{-i},r_i)) + x_i(t_{-i},r_i))] \\
&\quad = \sum_{i\in N} \sum_{t\in T} p(t_{-i})\, \lambda_i(t_i)[\sum_{c\in C(N)} \mu(c|t) u_i(c_N,t) + x_i(t)] + \\
&\qquad\qquad + \textstyle\sum_{i\in N} \sum_{t_i\in T_i} \sum_{r_i\in T_i} \alpha_i(r_i|t_i)[U_i(\mu,x|t_i) - \hat{U}_i(\mu,x,r_i|t_i)] \\
&\quad \geq \sum_{i\in N} \sum_{t\in T} p(t_{-i})\, \lambda_i(t_i)[\sum_{c\in C(N)} \mu(c|t)\, u_i(c_N,t) + x_i(t)].
\end{aligned}$$

Q.E.D.

8. Example 1, continued

This example was introduced in Section 5. The players are $N = \{1,2\}$, player 1's possible types are $T_1 = \{H,L\}$, but player 2 has only one possible type $T_2=\{2\}$. Player 1's types are equally likely, $p_1(H) = 1/2 = p_1(L)$. So the hydraulic equations (5) here become

$$\lambda_1(H) = 0.5+\alpha_1(H|L)-\alpha_1(L|H) \geq 0,\ \ \lambda_1(L) = 0.5+\alpha_1(L|H)-\alpha_1(H|L) \geq 0,\ \ \lambda_2 = 1.$$

Player 1 here is a seller with a single indivisible good. The value of 1's good to each player depends on 1's type as in the following table.

t_1	p	1's value of his good	2's value of 1's good	1's virtual value
H	0.5	5	6	$5+8\alpha_1(H\|L)$
L	0.5	1	2	$1-8\alpha_1(L\|H)$

As shown in the last column of the table, the virtual value of the good to type H of player 1 is

$$[(\lambda_1(H)+\alpha_1(L|H))5-\alpha_1(H|L)1]/p_1(H) =$$

$$= [(0.5+\alpha_1(H|L))5 - \alpha_1(H|L)1]/0.5 = 5 + 8\alpha_1(H|L).$$

Similarly, the virtual value of the good to type L of player 1 is

$$[(\lambda_1(L)+\alpha_1(H|L))1-\alpha_1(L|H)5]/p_1(L) =$$

$$= [(0.5+\alpha_1(L|H))1 - \alpha_1(L|H)5]/0.5 = 1 - 8\alpha_1(L|H).$$

For such one-good, one-sided incomplete information games, the incentive-efficient set can be shown to include all incentive-compatible mechanisms that satisfy two properties: the conditional probability of trading the good must be 1 when player 1's type is L, and the incentive constraint that 1's type L would not do better by reporting H must be binding. (See Myerson, 1985.) In fact, all incentive-efficient mechanisms can be supported in Theorem 0 by the same (λ,α) vectors that we saw in Section 5

$$\alpha_1(L|H) = 0,\ \ \alpha_1(H|L) = 1/8,\ \ \lambda_1(H) = 5/8,\ \ \lambda_1(L) = 3/8,\ \ \lambda_2 = 1.$$

With this (λ,α), the virtual value of the good to 1's type H is $5 + 8\alpha_1(H|L) = 6$, which is equal to the value of the good to player 2 when t_1=H, so that a randomization between trading and not-trading is consistent with the virtual-utility hypothesis. On the other hand, the virtual value of the good to 1's type L is 1, so the players can get virtual gains from trading equal to $2-1$ when t_1 = L. All incentive-efficient mechanisms then achieve the same maximized weight-sum of expected utilities subject to incentive constraints,

$$(5/8)U_1(\mu,x|H) + (3/8)U_1(\mu,x|L) + U_2(\mu,x) = 0.5$$

which is equal to the expected virtual gains from trade, $0.5(2-1)+0.5(6-6) = 0.5$.

With this (λ,α), the virtual constraints for an inhibitive allocation ω are

$$[(5/8 + 0)\omega_1(H)-(1/8)\omega_1(L)]/0.5 \geq 0,\ \text{ for } S=\{1\} \text{ with } t_1 = H,$$

$$[(3/8 + 1/8)\omega_1(L)-(0)\omega_1(H)]/0.5 \geq 0,\ \text{ for } S=\{1\} \text{ with } t_1 = L,$$

$$0.5\omega_2(H) + 0.5\omega_2(L) \geq 0,\ \text{ for } S = \{2\},$$

$$[(5/8 + 0)\omega_1(H)-(1/8)\omega_1(L)]/0.5 + \omega_2(H) \geq 0,\ \text{ for } S=\{1,2\} \text{ with } t_1 = H,$$

$$[(3/8 + 1/8)\omega_1(L)-(0)\omega_1(H)]/0.5 + \omega_2(L) \geq 1,\ \text{ for } S=\{1,2\} \text{ with } t_1 = L.$$

We are normalizing payoffs here so that no-trade gives payoff 0 to each player, and the coalition

{1,2} can get no virtual gains from trading when 1's type is H. The right-hand side of the last constraint is the virtual gain (2−1) that {1,2} can achieve from trading when 1's type is L. To be achievable, ω must satisfy

$$(5/8)\ \omega_1(H) + (3/8)\omega_1(L) + 0.5(\omega_2(H)+\omega_2(L)) \le 0.5.$$

This achievability constraint can only be satisfied when it and the last two inhibitive constraints are binding. Incentive compatibility implies that 1's expected profit from trade cannot be higher for type H than for type L. Thus, we get the system of inequalities

$$\omega_1(L) \ge \omega_1(H) \ge (1/5)\omega_1(L),\ \ (5/8)\ \omega_1(H) + (3/8)\omega_1(L) \le 0.5$$

$$\omega_2(L) = 1-\omega_1(L),\ \ \omega_2(H) = (1/4)\omega_1(L) - (5/4)\omega_1(H).$$

The extreme points $(\omega_1(L),\omega_1(H),\omega_2(L),\omega_2(H))$ of the allocations that satisfy these inequalities are

$$\{(0, 0, 1, 0), (0.5,0.5, 0.5, -0.5), (1, 0.2, 0, 0)\}.$$

Each of these three allocation vectors can be achieved by an incentive-efficient mechanism without sidepayments or severance pay. The first allocation is achieved by the mechanism:

if t_1 = L then player 2 buys the good for \$1, if t_1 = H then no trade.

The second allocation is achieved by the mechanism:

if t_1 = L then player 2 pays \$1.5 to buy the good,

if t_1 = H then player 2 pays \$0.5 and gets nothing.

The third allocation is achieved by the mechanism:

if t_1 = L then player 2 buys the good for \$2,

if t_1 = H then with probability 1/5 player 2 buys the good for \$6, else no trade.

So these three mechanisms, and all convex combinations of them, yield allocations that are in the core for this game. It can be shown (see notes available from the author) that these are the only core allocations for this game.

9. Side-bets and severance pay in Example 2

Vohra [22] and Forges, Mertens, and Vohra [4] have found that, when side-bets and severance offers are not allowed, concepts of incentive-compatible core that are broad enough to include ours as a subset can be empty. So our second example is a game where side-bets and severance offers are necessary to achieve allocations in the core.

The basic structure of Example 2 is the same as Example 1: there are two players, player 1's type is equally likely to be H or L, player 2 has only one possible type, and player 1 has a single indivisible good that he can sell to player 2. But in Example 2, the value of the good depends on 1's type as in the following table.

t_1	p	Value of good to 1	Value of good to 2
H	0.5	3	7
L	0.5	2	0

For this example, it can be shown (Myerson, 1985) that all incentive-efficient mechanisms are pooling: player 1 always sells the good and is paid a price χ that is between \$3 and \$3.50, and all incentive-efficient mechanisms are supported by (λ,α) with $\lambda_1(L) = 0$. Although player 2 cannot profit directly from trading with 1's type L, player 2 finds it unproductive to try to separate 1's types because their values of the good are so close. In this context, 1's type L is a "bad" type that can get profits just because of its similarity to 1's other type. (The problem of pooling with the bad type here is similar to the insurance examples where Rothschild and Stiglitz, 1976, found nonexistence of competitive equilibria.)

Letting $\beta = \alpha_1(L|H)$, the hydraulic equations (5) with $\lambda_1(L) = 0$ yield

$$\lambda_1(L) = 0,\ \ \lambda_1(H) = 1,\ \ \alpha_1(H|L)-0.5 = \alpha_1(L|H) = \beta \geq 0,\ \ \ \lambda_2 = 1.$$

Then the virtual value of the good to 1's type H is

$$[(1 + \beta - 0.5)3 - (\beta)2]/0.5 = 3+2\beta,$$

and the virtual value of the good to 1's type L is

$$[(0 + \beta)2 - (\beta-0.5)3]/0.5 = 3-2\beta.$$

To make trade virtually ex-post efficient for both types, we need $3+2\beta \leq 7$ and $3-2\beta \leq 0$, and so

$$1.5 \leq \beta \leq 2.$$

With $\lambda_1(L) = 0$, Theorem 3 allows that severance offers might be made to 1's type L, because its λ-weight is 0. An offer of ε in severance pay to type L would reduce the stakes $\omega_1(L)$ by ε but would leave $\omega_1(H)$ unaffected, and the result would be to decrease $V_1(\omega,L,\lambda,\alpha)$ by $\varepsilon[\lambda_1(L)+\alpha_1(H|L)]/0.5 = 2\beta\varepsilon$ and to increase $V_1(\omega,H,\lambda,\alpha)$ the same amount $\varepsilon\alpha_1(H|L)/0.5 = 2\beta\varepsilon$.

In fact, we can show that no mechanism without severance pay can achieve a core allocation for this example. Consider any mechanism where player 1 always sells the good for the price $\chi = x_1(H) = x_1(L)$ without side-bets or severance pay. The absence of side-bets means that player 2 pays exactly what 1 receives ($x_2 = -x_1$). We must have $3 \leq \chi \leq 3.5$, because 1's type H would refuse to sell for less than 3, and player 2 would refuse to pay more than 3.5. Then the following is a tenable blocking plan against this mechanism:

If t_1=H then {1} blocks, player 1 keeps the good and is paid $\chi-2$ by the blocking mediator;

but if t_1=L then {1,2} block, player 1 keeps the good and is paid $\chi-2$, and player 2 pays χ to blocking mediator.

The blocking mediator's expected profit is $0.5\chi - (\chi-2) = 2-0.5\chi > 0$ because $\chi \leq 3.5$. Player 1's type L gets the same profit $\chi-2$ as in the established plan, and player 1's type H does strictly better than in the established plan, but neither type could do better by lying. When player 2 is invited, 1's type is L, in which case paying χ for nothing is not worse for player 2 than the established plan (where 2 will be asked to pay χ for a good that is worth 0).

To find the core, let us consider more general pooling mechanisms where 1 is always paid $\chi = x_1(H) = x_1(L)$ to sell his good, but 2 pays either $\chi+\tau$ if t_1=H or $\chi-\tau$ if t_1=L, and player 1's type L is offered $\varepsilon = \varepsilon_1(L)$ severance pay. Here τ is the amount of a side-bet between player 2 and the established mediator, which player 2 pays if 1's type is H. If there were no severance offer (ε=0) then such a mechanism could be easily blocked, because the blocking mediator could plan to imitate the established plan when 1's type is H but to send the players back to the established plan when 1's type is L. In fact, we can show that a core allocation cannot be achieved with any severance ε less than 5/6.

A mechanism in this pooling family yields the following stake allocation ω:

$$\omega_1(L) = \chi-2-\varepsilon,\ \ \omega_1(H) = \chi-3,\ \ \omega_2(L) = \tau-\chi,\ \ \omega_2(H) = 7-\chi-\tau.$$

With (λ,α) as above, the virtual transforms of player 1's payoffs are

$$V_1(\omega,H,\lambda,\alpha) = [(1 + \beta - 0.5)(\chi-3) - (\beta)(\chi-2-\varepsilon)]/0.5 = \chi-2\beta(1-\varepsilon)-3$$

$$V_1(\omega,L,\lambda,\alpha) = [(0 + \beta)(\chi-2-\varepsilon) - (\beta-0.5)(\chi-3)]/0.5 = \chi+2\beta(1-\varepsilon)-3$$

So the virtual constraints for an inhibitive allocation ω are

$\chi - 2\beta(1-\varepsilon) - 3 \geq 0$, for S={1} with t_1 = H,

$\chi + 2\beta(1-\varepsilon) - 3 \geq 0$, for S={1} with t_1 = L,

$0.5(7-\chi-\tau) + 0.5(\tau-\chi) \geq 0$, for S = {2},

$\chi - 2\beta(1-\varepsilon) - 3 + (7-\chi-\tau) \geq 7 - (3+2\beta)$, for S={1,2} with t_1 = H,

$\chi + 2\beta(1-\varepsilon) - 3 + (\tau-\chi) \geq 0 - (3-2\beta)$, for S={1,2} with t_1 = L.

As noted above, we require here that $\beta = \alpha_1(L|H)$ satisfies $1.5 \leq \beta \leq 2$, so that virtual gains from trade on the right-hand sides of the last two constraints are both nonnegative. Then our conditions for a core allocation can be simplified to

(27) $3 + 2\beta|1-\varepsilon| \leq \chi \leq 3.5$, $\tau = 2\beta\varepsilon$, where $1.5 \leq \beta \leq 2$.

There are many such core solutions, but the one with the smallest severance offer ε is

$\beta = 1.5$, $\chi = 3.5$, $\varepsilon = 5/6$, $\tau = 2.5$.

10. Coalitional durability

In this final section, we consider an alternative concept of stability against coalitional renegotiation, based on the concept of durability proposed by Holmström and Myerson [6].

Let (μ, x) be any incentive-compatible mechanism, and let $S \subseteq N$ be any coalition. An alternative game for coalition S is any

$$\Gamma_S = ((D_i)_{i\in S}, f{:}D \to \Delta(C(S)), z{:}D \to \mathbb{R}^S),$$

where each D_i is a nonempty finite set, $D = \times_{i\in S} D_i$, and the $z_i(d)$ numbers satisfy

(28) $\sum_{i\in S} z_i(d) < 0$, $\forall d \in D$.

Here D_i represents the set of the pure strategies for player i in this alternative game. When players choices in this game are $d = (d_j)_{j\in N}$, f(d) is the feasible action that coalition S would implement, and $z_i(d)$ is the net sidepayment that player i would get.

Now suppose that such an alternative game can be proposed to the members of some coalition S by a blocking mediator. So we are now assuming that the blocking mediator's invited coalition is publicly known, and the blocking mediator cannot specify which equilibrium would be played in the alternative game. So condition (28) requires that the blocking mediator is guaranteed a strictly positive profit when the alternative game is played, no matter what strategies the players in S might choose in the alternative game.

Suppose that the players in S will vote independently about whether to form S as a blocking coalition and then play the alternative game Γ_S. We assume that they will play the alternative game Γ_S if and only if they vote unanimously for it, otherwise they will remain with the mechanism (μ,x). But the players' beliefs about each other might be affected by the news that everyone was unanimous for Γ_S. That is, there must be some beliefs vector q in $\times_{i\in S}\,\Delta(T_i)$ that represents what the players would believe about each others' types after a unanimous vote for Γ_S. Here $q_i(t_i)$ denotes the probability of player i being type t_i given that he voted for Γ_S.

An equilibrium of alternative Γ_S with beliefs q is a profile of strategies $\sigma = (\sigma_i{:}T_i\to\Delta(D_i))_{i\in S}$ such that

$$U_i(\Gamma_S,\sigma|q,t_i) \geq \hat{U}_i(\Gamma,\sigma,e_i|q,t_i),\ \ \forall i\in S,\ \forall t_i\in T_i,\ \forall e_i\in D_i,$$

where we use the notation

$$U_i(\Gamma_S,\sigma|q,t_i) = \sum_{t_{-i}\in T_{-i}} \sum_{d\in D} p(t_{N-S})\, q(t_{S-i})\, \sigma(d|t_S)\, [z_i(d) + \sum_{c\in C(S)} f(c|d)\, u_i(c,t)],$$

$$\hat{U}_i(\Gamma_S,\sigma,e_i|q,t_i) =$$

$$= \sum_{t_{-i}\in T_{-i}} \sum_{d\in D} p(t_{N-S})\, q(t_{S-i})\, \sigma(d|t_S)\, [z_i(d_{-i},e_i) + \sum_{c\in C(S)} f(c|d_{-i},e_i)\, u_i(c,t)],$$

$$q(t_{S-i}) = \prod_{j\in S-i} q_j(t_j),\ \text{ and }\ \sigma(d|t_S) = \prod_{j\in S} \sigma_j(d_j|t_j).$$

The expected payoffs under mechanism (μ,x) with the same beliefs about S would be

$$U_i(\mu,x|q,t_i) = \sum_{t_{-i}\in T_{-i}} p(t_{N-S})\, q(t_{S-i})\, [x_i(d) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t)].$$

The definition of sequential equilibrium would allow (with #S>1) that each player votes against Γ_S because he thinks that all others will vote against it, so that his vote is meaningless. To avoid this trivial result, we require also that any type of any player who would expect to do better under the alternative, given unanimity of the others for it, must be expected to vote for the alternative.

So we say that (μ,x) is coalitionally durable iff, for any alternative game Γ_S for any coalition $S\subseteq N$, there always exists a beliefs vector q and a profile of strategies σ such that σ is an equilibrium of Γ_S with beliefs q, and for each type t_i of each player i we have

$$\text{if }\ U_i(\Gamma_S,\sigma|q,t_i) > U_i(\mu,x|q,t_i)\ \text{ then }\ q_i(t_i)/p_i(t_i) = \max\{q_i(r_i)/p_i(r_i)|\ r_i\in T_i\},$$

and there exists some player j in S such that

$$U_j(\mu,x|q,t_j) \geq U_j(\Gamma_S,\sigma|q,t_j),\ \forall t_j\in T_j.$$

That is, we can find beliefs and an equilibrium for any alternative game such that types who would strictly gain by playing this alternative are considered maximally likely to vote for it, but there is at least one player in S who is always willing to vote against this alternative.

Theorem 4. If an allocation in the core is achievable by a feasible mechanism (μ,x) then (μ,x) is coalitionally durable.

Proof of Theorem 4. Let ω be an allocation in the core that is achievable by (μ,x). Given any Γ_S, consider the extensive-form game where the members of S first vote independently for or against the alternative Γ_S, when each player knows only his own type. If anyone votes against Γ_S, then everybody gets their allocation in ω, but otherwise the players in S play Γ_S. Perturb this game by adding a small positive probability that each player might vote for the alternative (by a trembling hand). This finite game must have a sequential equilibrium.

Now take the limit of these sequential equilibria as the trembling probabilities go to zero. In such a limit of sequential equilibria, each type t_i of each player i in S has some probability $\tau_i(t_i)$ of voting for the alternative, and has some strategy $\sigma_i(\bullet|t_i)$ for playing the alternative if it is unanimously voted. Also everyone would have some consistent beliefs $q_i \in \Delta(T_i)$ about player i's type given that he voted for the alternative, which must satisfy

$$q_i(t_i) = p_i(t_i)\tau_i(t_i) / \sum_{r_i \in T_i} p_i(r_i)\tau_i(r_i), \text{ if any } \tau_i(r_i) > 0.$$

The equilibrium voting strategy must satisfy

$$\tau_i(t_i) = 1 \text{ if } U_i(\Gamma_S,\sigma|q,t_i) > \sum_{t_{-i} \in T_{-i}} p(t_{N-S})\, q(t_{S-i})\, \omega_i(t),$$

$$\tau_i(t_i) = 0 \text{ if } U_i(\Gamma_S,\sigma|q,t_i) < \sum_{t_{-i} \in T_{-i}} p(t_{N-S})\, q(t_{S-i})\, \omega_i(t).$$

Also, σ must satisfy the conditions for an equilibrium of the alternative game Γ_S with beliefs q. Let $\tau(t_S) = \prod_{j \in S} \tau_j(t_j)$ denote the probability of the alternative when types are t_S, and let

$$\psi_i = \prod_{j \in S-i} [\sum_{t_j \in T_j} p_j(t_j)\, \tau_j(t_j)] \;.$$

So ψ_i denotes the probability of all $S-i$ voting for the alternative. We now show that this limit of sequential equilibria must satisfy the conditions for coalitional durability.

If the limit of sequential equilibria does not satisfy the conditions for coalitional durability then, for every player i in S, there must be at least one type t_i such that

$$U_i(\Gamma_S,\sigma|q,t_i) > U_i(\mu,x|q,t_i) \geq \sum_{t_{-i} \in T_{-i}} p(t_{N-S})\, q(t_{S-i})\, \omega_i(t), \text{ and so } \tau_i(t_i) > 0.$$

Then we can define the blocking plan (ν,y) by

$$\nu_S(c|t_S) = \tau(t_S) \sum_{d\in D} \sigma(d|t_S)\, f(c|d), \ \ \forall c\in C(S), \ \ \forall t_S\in T_S,$$

$$\nu_R(d|t_R) = 0, \text{ for any coalition } R\neq S,$$

$$y_i(t_i) = \sum_{t_{-i}\in T_{-i}} p(t_{-i})\, \tau(t_S) \sum_{d\in D} \sigma(d|t_S)\, z_i(d).$$

But the sequential equilibrium conditions for (q,τ,σ) in Γ_S then imply that this blocking plan (ν,y) is tenable against ω. To verify tenability, we apply the sequential-equilibrium conditions on (τ,σ) to get, for any $i\in S$ and any $t_i\in T_i$,

$$y_i(t_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{c\in C(S)} \nu(c|t_S)(u_i(c,t)-\omega_i(t))$$

$$= \sum_{t_{-i}\in T_{-i}} p(t_{-i})\, \tau(t_S) \sum_{d\in D} \sigma(d|t_S)\, [z_i(d) + \sum_{c\in C(S)} f(c|d)u_i(c,t) - \omega_i(t)]$$

$$= \psi_i\, \tau_i(t_i) \sum_{t_{-i}\in T_{-i}} p_{N-S}(t_{N-S}\, q_{S-i}(t_{S-i}) \sum_{d\in D} \sigma(d|t_S)[z_i(d) + \sum_{c\in C(S)} f(c|d)u_i(c,t) - \omega_i(t)]$$

$$= \psi_i\, \tau_i(t_i)[U_i(\Gamma_S,\sigma|q,t_i) - \sum_{t_{-i}\in T_{-i}} p(t_{N-S})\, q(t_{S-i})\, \omega_i(t)] \geq 0$$

and similarly

$$y_i(t_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{c\in C(S)} \nu(c|t_S)(u_i(c,t)-\omega_i(t))$$

$$= \sum_{t_{-i}\in T_{-i}} p(t_{-i})\, \tau(t_S) \sum_{d\in D} \sigma(d|t_S)\, [z_i(d) + \sum_{c\in C(S)} f(c|d)u_i(c,t) - \omega_i(t)]$$

$$\geq \sum_{t_{-i}\in T_{-i}} p(t_{-i})\, \tau(t_{S-i},r_i) \sum_{d\in D} \sigma(d|t_{S-i},r_i)\, [z_i(d) + \sum_{c\in C(S)} f(c|d)u_i(c,t) - \omega_i(t)]$$

$$= y_i(r_i) + \sum_{t_{-i}\in T_{-i}} p(t_{-i}) \sum_{c\in C(S)} \nu(c|t_{S-i},r_i)\, (u_i(c,t)-\omega_i(t)).$$

But tenability of (ν,y) against ω contradicts the assumption that ω is inhibitive. So the mechanism (μ,x) must be coalitionally durable. QED

REFERENCES

[1] G. de Clippel, "The type-agent core of exchange economies with asymmetric information," working paper, Department of Economics, Brown University (2004).

[2] F. Forges, A. Heifetz, and E. Minelli, "Incentive-compatible core and competitive equilibria in differential information economies," Economic Theory 18 (2001), 349-365.

[3] F.Forges, E. Minelli, and Rajiv. Vohra, "Incentives and the core of an exchange economy: a survey," Journal of Mathematical Economics 38 (2002), 1-41.

[4] F. Forges, J.-F. Mertens, and Rajiv Vohra, "The ex-ante incentive-compatible core in the

absence of wealth effects," Econometrica 70 (2002), 1865-1892.

[5] J.C. Harsanyi, "Games with incomplete information played by 'Bayesian' players," Management Science 14 (1967-8), pp 159-182, 320-334, 486-502.

[6] B. Holmström and R.B. Myerson, "Efficient and Durable Decision Rules with Incomplete Information," Econometrica 51 (1983), 1799-1819.

[7] D. Lee and O. Volij, "The core of economies with asymmetric information: an axiomatic approach," Journal of Mathematical Economics 38 (2002), 43-63.

[8] A. Malakhov and Rakesh Vohra, "Single and multi-dimensional optimal auctions, a network approach," working paper, MEDS Department, Northwestern University (2004).

[9] R.B. Myerson, "Mechanism design by an informed principal," Econometrica 51 (1983), 1767-1797.

[10] R.B. Myerson, "Two-person bargaining problems with incomplete information," Econometrica 52 (1984), 461-487.

[11] R.B. Myerson, "Cooperative games with incomplete information," International Journal of Game Theory 13 (1984), 69-86

[12] R.B. Myerson, "Analysis of two bargaining problems with incomplete information," in Game-Theoretic Models of Bargaining, edited by A.E. Roth, Cambridge U. Press, 1985, pp. 115-147.

[13] R.B. Myerson, "Credible negotiation statements and coherent plans," Journal of Economic Theory 48 (1989), 264-303.

[14] R.B. Myerson, "Fictitious-transfer solutions in cooperative game theory," in Rational Interaction, edited by Reinhard Selten, Springer-Verlag, 1992, pp. 13-33.

[15] R.B. Myerson, "Sustainable matching plans with adverse selection," Games and Economic Behavior 9 (1995), 35-65.

[16] R.B. Myerson, Game Theory: Analysis of Conflict, Harvard U. Press, 1991.

[17] M. Rothschild and J. Stiglitz, "Equilibrium in competitive insurance markets," Quarterly Journal of Economics 90 (1976), 629-650.

[18] C.-Z. Qin, "The inner core and the strictly inhibitive set," Journal of Economic Theory 59 (1993), 96-106.

[19] R. Serrano and Rajiv Vohra, "Information transmission in coalitional voting," working paper, Department of Economics, Brown University (2004).

[20] R. Serrano, Rajiv Vohra, and O. Volij, "On the failure of core convergence in economies with asymmetric information," Econometrica 69 (2001), 1685-1696.

[21] L.S. Shapley, "Utility comparison and the theory of games," in La Decision, Editions du CNRS, Paris, 1969, pp. 251-263. Reprinted in A.E. Roth, editor, The Shapley Value, Cambridge U. Press, 1988, pp. 307-319.

[22] Rajiv Vohra, "Incomplete information, incentive compatibility and the core," Journal of Economic Theory 86 (1999), 123-147.

[23] R. Wilson, "Information, efficiency and the core of an economy," Econometrica 46 (1978), 807-816