Virtual Utility and the Core for Games with Incomplete Information
http://home.uchicago.edu/~rmyerson/research/vcore.pdf

1. Introduction
We consider here the question of how to extend our definition of the core to games with incomplete information.
Because of the importance of the core in economic theory and cooperative game theory, this question seems fundamental to information economics, and has regularly attracted effort from economic theorists for over 25 years.
See the recent survey by Forges, Minelli, and Vohra (J. Math Econ, 2002.)
In their framework, the concept that we are developing here may be characterized as an incentive-compatible interim fine core.
The hardest part of information economics is the analysis of informational incentive constraints that describe the difficulties that people have in trusting each other's statements about private information. So we insist here that all plans must satisfy incentive compatibility: no one can be made to share information against his interest.
Interim means that coalitions are assumed to consider the possibility of blocking when each players already knows his private information. The alternative of "exante" analysis assumes that players decide about joining alternative coalitions before learning their types, thus eliminating any problem of adverse selection in coalition formation ("if you want to join me in a coalition, then I don't want you").
Fine here means that players may share some information in the process of deciding whether to form a coalition.
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 and two-person games.
Transfering utility by sidepayments is a well-known simplifying assumption in cooperative game theory. Note that that planned sidepayments must satisfy incentive compatibility here.
Independent types simplifies notation, but makes incentive compatibility more difficult.
(We are not assuming independent private values.)
We assume that a risk-neutral mediator can offer monetary side-bets about other players' types such that the mediator's expected profit is nonnegative, which allow a kind of weak feasibility-in-expectation.
The mediator can also promise severance payments that would be paid to players even if they join a blocking coalition.
But these side-bets and severance payments are only allowed here if the mediator can be confident that the players would not cheat her by adverse selection (deviating to a blocking coalition in states where the mediator would have profited from the players).
With these assumptions, we show that existence of feasible mechanisms that can achieve core allocations, and we show that such mechanisms satisfy a coalitional-durability property (extending Holmstrom-Myerson 1983).

2. Formulation of the game.

Let N denote the set of players (nonempty, finite).

We let i denote a generic player in N, and let S denote a generic nonempty subset of N.

For each $S \subseteq N$, let C(S) denote the finite set of feasible joint actions for coalition S.

Let T_i denote the set of possible types of player i (nonempty, 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.

Notation: $t_S = (t_i)_{i \in S} \in T_S = \times_{i \in S} T_i$, $p(t_S) = \prod_{i \in S} p_i(t_i)$, for any $S \subseteq N$,

$t = t_N = (t_i)_{i \in N} \in T = T_N$, $p(t) = \prod_{i \in N} p_i(t_i)$,

$t_{-i} = (t_j)_{j \in N-i} \in T_{-i} = T_{N-i}$, $p(t_{-i}) = \prod_{j \in N-i} p_j(t_j)$, $t = (t_{-i}, t_i)$.

$\forall t \in T$, $\forall S \subseteq N$, $\forall i \in S$, $\forall c \in C(S)$, we let $u_i(c,t)$ denote the utility payoff to player i from joining coalition S which does c given that all players' types are as in t.

(<u>Orthogonal coalitions</u>: We assume that when members of S choose actions jointly feasible for them, the actions of others outside S do not matter to them, but types of others might.)

For any set X, $\Delta(X)$ denotes the set of probability distributions over X.

Let us say that $\theta = (\theta_{S,c})_{S \subseteq N, c \in C(S)}$ is a <u>balanced coalitional plan</u> iff

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

We assume that the game is <u>balanced</u>, in the sense that the grand coalition N can achieve anything that a balanced coalitional plan can achieve;

That is, for any balanced coalitional plan θ, N has a randomized strategy $\sigma \in \Delta(C(N))$ such that:

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

A collective choice <u>mechanism</u> is a pair of functions ($\mu: T \to \Delta(C(N))$, $x: T \to \mathbb{R}^N$).

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 transfer to player i when the players' types are t.

The expected utility of player i under mechanism (μ,x) is

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

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

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

A mechanism (μ,x) is <u>incentive compatible</u> iff $U_i(\mu,x|t_i) \geq \hat{U}_i(\mu,x,r_i|t_i)$, $\forall i \in N$, $\forall t_i \in T_i$, $\forall r_i \in T_i$.

The <u>mediator's expected payoff</u> from the plan (μ,x) is $-\sum_{t \in T} p(t) \sum_{i \in N} x_i(t)$.

A mechanism (μ,x) is <u>feasible</u> iff it is incentive compatible and yields a nonnegative expected payoff to the mediator; that is, $\sum_{t \in T} p(t) \sum_{i \in N} x_i(t) \leq 0$.

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 $(\bar{\mu},\bar{x})$ 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).

Let $\alpha_i(r_i|t_i)$ be the Lagrange multiplier for the constraint that type t_i should gain by reporting r_i.

Then the Lagrangean for this optimization problem can be written:

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

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

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

With this definition, the above Lagrangean can be rewritten

$L(\mu,x,\lambda,\alpha) = \sum_{t\in T} p(t) \sum_{c\in C(N)} \mu(c|t) \sum_{i\in N} v_i(c,t,\lambda,\alpha)$

$\qquad + \sum_{t\in T} p(t) \sum_{i\in N} x_i(t)((\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i)) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i))/p_i(t_i)$.

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:

Theorem 0. A feasible mechanism (μ,x) is incentive-efficient iff there exist vectors λ and α such that: $\lambda_i(t_i) \geq 0\ \forall i\in N, \forall t_i\in T_i$, $\alpha_i(r_i|t_i) \geq 0, \forall i\in N, \forall t_i\in T_i, \forall r_i\in T_i$,

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

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

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

Let us say that a type t_i of player i jeopardizes another type r_i of player i, in the incentive-efficient 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 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.

A Coasian believer in ex-post efficiency could "explain" inefficient signaling by the virtual-utility hypothesis: when incentive constraints bind, players act according to virtual utilities.

4. Blocking.

In the theory of the core, we think about an established plan that is able to inhibit players from joining any alternative blocking coalition.

With incomplete information, we should think about an established mediator implementing a mechanism that can inhibit players from deviating to cooperate with some 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. (Uninvited rejections do not count.)

But with incomplete information and interim coalition-formation, we must also consider the possibility that others 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 the probability of any particular coalition blocking and choosing some joint action can depend on their information but not on the information of others outside this coalition. So the blocking mediator can ask any random set S about their types and, based on their responses, either may invite all of S into a blocking coalition or may invite no blocking coalition.

To characterize a blocking plan, for any action c in C(S), we let $\nu(c|t_S)$ represent the probability that coalition S would be invited to block and do joint action c if they report types t_S.

The blocking mediator can also make monetary sidepayments with players who help block.

We 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 the type t_i to the blocking mediatior.

So we define a blocking plan to be any pair of vectors (ν,y) such that: $y_i(t_i) \in \mathbb{R}, \forall i\in N, \forall t_i\in T_i$,

$\nu(c|t_S) \geq 0, \forall c\in C(S), \forall t_S\in T_S$, and $\sum_{S\subseteq N} \sum_{c\in C(S)} \nu(c|t_S) \leq 1, \forall t\in T$.

[w.l.o.g. could require $\sum_{S\subseteq N} \sum_{t_S\in T_S}\sum_{c\in C(S)} \nu(c|t_S) \leq 1$] (For $t = (t_i)_{i\in N} \in T$, let $t_S = (t_i)_{i\in S}$.)

Let $\omega_i(t)$ denote the payoff allocation that player i would get from the established plan, but 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 tenable against ω iff

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

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

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

and y yields a positive expected profit to the blocking mediator: $-\sum_{i\in N} \sum_{t_i\in T_i} p_i(t_i)y_i(t_i) > 0$.

4.1 Example.
Player 1 is seller of a single good. 1's type $t_1 \in \{H,L\}$. $p_1(H) = p_1(L) = 0.5$.
Player 2 is the only potential buyer, and she has no private information.
If t_1=H then the good's value is \$5 to player 1, its value is \$6 to player 2.
If t_1=L then the good's value is \$1 to player 1, its value is \$2 to player 2.
In incentive-compatibility and participation constraints require $P(\text{trade}|t_1=H) < 1$.

Consider the following incentive compatible mechanism:
M1: If t_1=L then price=\$1.50, P(trade|L)=1; if t_1=H then priceIfTrade=\$5.50, P(trade|H)=1/9.

Notice $U_1(L) = \hat{U}_1(H|L)$, but $U_1(H) > \hat{U}_1(L|H)$ with this mechanism
With the given prices in each state, P(trade) cannot be higher.
This mechanism is incentive-efficient.
It seems fair, equalizes players' gains in each state: $\omega_1(H) = 1/18 = \omega_2(H)$, $\omega_1(L) = 0.50 = \omega_2(L)$.
But to support its incentive-efficiency, total virtual gains from trade must be \$0 when t_1=H,
because $0 < P(\text{trade}|t_1=H) < 1$.
Player 2's virtual utility is same as her actual utility, because she has only one type.
So when t_1=H, player 1's virtual valuation of the good must equal 2's valuation of \$6.
With $\alpha_1(L|H) = 0$ (because $U_1(H) > \hat{U}_1(L|H)$), this virtual valuation is achieved with
$\lambda_1(H) = 5/8$, $\lambda_1(L) = 3/8$, $\lambda_2 = 1$, $\alpha_1(H|L) = 1/8$, so $((5/8+0)\$5 - (1/8)\$1)/0.5 = \$6$.
So in terms of virtual utility, selling for \$5.50 seems unfair and unacceptable to 1 when t_1=H.
Under the virtual utility hypothesis, 1's virtual loss when t_1=H would block this mechanism!
(When t_1=L, 1's virtual valuation is $((3/8 + 1/8)\$1 - (0)\$5)/0.5 = \$1$, same as actual.)

Such virtual-utility calculations can be justified with random blocking mechanisms.
Here is a tenable blocking plan against this mechanism (yielding expected profit $(0.5/9)0.40 > 0$).
If player 1 reports t_1=H then:
with prob'y 8/9 {1} blocks alone and keeps good;
but with prob'y 1/9 {1,2} block together, good is traded, 2 pays \$5.90, 1 is paid \$5.50,
If player 1 reports t_1=L then return to fulfill the established mechanism M1.
This blocking plan always gives player 1 the same gains as the mechanism M1.
Player 2 does worse ex ante, but when her conditional expected gains are increased when is invited to join the blocking coalition:
2's conditional expected gain, given that 2 is invited and t_1=L, is 0.50, same as $\omega_2(L)$.
2's conditional expected gain, given that 2 is invited and t_1=H, is 0.10, more than $\omega_2(H)$=1/18!
Key is that blocking plan only invites player 2 when the random signaling cost is not applied.

Such a random blocking plan could be similarly constructed against any incentive-efficient mechanism where priceIfTrade is less than \$6 when t_1=H.
Notice the coincidence of two criteria: expected virtual gains for all types in the mechanism, and impossibility of random blocking against it. This coincidence is generalized in Theorem 1.

Here is a virtually-equitable incentive-efficient mechanism which cannot be blocked in this way:
M2: If t_1=L then price=\$1.50, P(trade|L)=1; if t_1=H then priceIfTrade=\$6.00, P(trade|H)=1/10.

5. Virtual utility and inhibitive allocations

We say ω is <u>inhibitive</u> iff there does not exist any blocking plan (ν,y) that is tenable against ω.

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

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

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

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

Thus 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: Consider the linear programming problem:

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

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

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

$\nu(c|t_S) \geq 0$, $\forall c\in C(S)$, $\forall t_S\in T_S$, 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. (The "...≤ 1" constraint 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 LP has optimal value 0 iff its dual also has optimal value 0. Let $\lambda_i(t_i)$ denote the dual variable for the first inequality, and let $\alpha_i(r_i|t_i)$ denote the dual variable for the second inequality. Then the dual problem can be written:

choose (λ,α) so as to maximize 0 subject to $\lambda_i(t_i) \geq 0$ and $\alpha_i(r_i|t_i) \geq 0$, $\forall i\in N$, $\forall t_i\in T_i$, $\forall r_i\in T_i$,

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

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

So ω is inhibitive iff these inequalities can be satisfied. QED

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

$$\sum_{i\in S} ((\lambda_i(t_i)+\sum_{r_i\in T_i} \alpha_i(r_i|t_i))\omega_i(t) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)\omega_i(t_{-i},r_i))/p_i(t_i) \geq \sum_{i\in S} v_i(c,t,\lambda,\alpha)$$
$$\forall t\in T, \forall S\subseteq N, \forall c\in C(S).$$

These fine-support conditions imply that the blocking mediator could not expect to make money if the action of any blocking coalition S could also depend on the information of others in N−S.

6. The Core

We say that ω is <u>achievable</u> by (μ,x) iff (μ,x) is an feasible mechanism and

$\omega_i(t) \leq x_i(t) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t), \forall i\in N, \forall t\in T.$

The difference $x_i(t) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t) - \omega_i(t)$ can be interpreted as the <u>severance payment</u> that player i would get from the established mediator (with reported types t) even if i joined a blocking coalition. These promised severance payments must be nonnegative.

So $\omega_i(t)$ is the residual stake in the established plan that i would lose if it is blocked in state t.

We say that ω is in the <u>core</u> iff ω is inhibitive and achievable by some feasible mechanism (μ,x).

<u>Theorem 2.</u> The core is nonempty.

Intuitive description of the proof:

Given any vectors (λ,α) (satisfying the hydraulic equations in Theorem 0), consider the problem of getting the highest expected net payment from players with an incentive-compatible mechanism that achieves some inhibitive allocation vector that is finely supported by (λ,α). This problem can be written as a linear programming problem:

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

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

$\sum_{t_{-i}\in T_{-i}} p(t_{-i})(x_i(t) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t))$

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

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

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

$\sum_{i\in S} ((\lambda_i(t_i)+\sum_{r_i\in T_i} \alpha_i(r_i|t_i))\omega_i(t_i) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)\omega_i(t_{-i},r_i))/p_i(t_i) \geq \sum_{i\in S} v_i(c,t,\lambda,\alpha)$

$\forall t\in T, \forall S\subseteq N, \forall c\in C(S).$

Notice that the pair (λ,α) are given parameters in the above problem.

We 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 show that the solution (μ,x,ω) at this (λ,α)-fixed point is an incentive-compatible mechanism achieving a core allocation.

Actually, in the proof, we decompose this problem, and we first consider a simpler optimization over (μ,x) where ω is taken as a given paramater.

6.1 [core existence proof]

Proof. Given any allocation vector ω, consider the linear programming problem:

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

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

(2) $\sum_{t_{-i}\in T_{-i}} p(t_{-i})(x_i(t) + \sum_{c\in C(N)} \mu(c|t)\, u_i(c,t))$

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

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

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

It is easy to see that this problem 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 minimum-payoff constraints (1) 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 the minimum-payoff constraints (1) 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, the dual of this linear programming problem also has an optimal solution for each ω.

To formulate this dual, we let the dual variables of (1) be $\eta_i(t)$, of (2) be $\beta_i(r_i|t_i)$, of (3) be $\gamma(t)$.

Then the dual problem is: choose vectors (η,β,γ) so as to

maximize $\sum_{t\in t} \sum_{i\in N} p(t_{-i})\, \omega_i(t)\, \eta_i(t) + \sum_{t\in T} p(t)\gamma(t)$ subject to

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

(6) $\sum_{i\in N} p(t_{-i})((\eta_i(t)+\sum_{r_i\in T_i} \beta_i(r_i|t_i))u_i(c,t) - \sum_{r_i\in T_i} \beta_i(t_i|r_i)u_i(c,(t_{-i},r_i))) + p(t)\gamma(t) \leq 0,$

$\forall t\in T,\ \forall c\in C(N),$

(7) $\eta_i(t) \geq 0$ and $\beta_i(r_i|t_i) \geq 0,\ \forall i\in N,\ \forall t\in T,\ \forall r_i\in T_i.$

(Constraints (5) are for the primal variables $x_i(t)$, and (6) are for the primal variables $\mu(c|t)$.)

Dividing both sides of (5) by $p(t_{-i})$, we get the equivalent constraints:

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

This implies that $\eta_i(t)$ is independent of t_{-i}, so there exists $h_i(t_i)$ such that $\eta_i(t) = h_i(t_i),\ \forall i,\ \forall t\in T.$

Notice that β actually determines the other two dual variables as follows:

$h_i(t_i) = p_i(t_i) + \sum_{r_i\in T_i} \beta_i(t_i|r_i) - \sum_{r_i\in T_i} \beta_i(r_i|t_i)$, and $\gamma(t) = -\max_{c\in C(N)} \sum_{i\in N} v_i(c,t,h,\beta).$

6.2 [core existence proof continued]

Substituting the vector h for η and using the definition of virtual utility v_i, the dual can be reformulated equivalently as: choose vectors (h,β,γ) so as to

maximize $\sum_{t\in t} \sum_{i\in N} p(t_{-i})\omega_i(t)h_i(t_i) + \sum_{t\in T} p(t)\gamma(t)$ subject to

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

(6') $\sum_{i\in N} v_i(c,t,h,\beta) + \gamma(t) \le 0, \forall t\in T,\ \forall c\in C(N),$

(7') $h_i(t_i) \ge 0$ 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 has an optimal solution for every allocation vector ω.

But ω only appears in the objective function. The dual constraints do not depend on ω.

The dual feasible set has a finite collection of extreme points. Let us number these extreme points from 1 to K and denote them as $\{(h^1,\beta^1,\gamma^1), ..., (h^K,\beta^K,\gamma^K)\}$.

Let Λ denote the convex hull of these K extreme points of the dual feasible set.

(Λ may be smaller than the whole feasible set, if it is unbounded.)

For any allocation ω, the smallest expected subsidy to implement ω is

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

Notice that, if $(\lambda,\alpha,\delta) \in \Lambda$ then (λ,α) satisfies the first two conditions required to support an allocations by in Theorem 1: $\lambda_i(t_i) \ge 0$ and $\alpha_i(r_i|t_i) \ge 0,\ \forall i\in N,\ \forall t_i\in T_i,\ \forall r_i\in T_i,$ and

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

Given any such (λ,α), the following LP problem finds the smallest expected subsidy that is required to achieve an inhibitive allocation ω that is finely supported by (λ,α):

choose (ω,z) to minimize z subject to

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

(9) $p(t) \sum_{i\in S} ((\lambda_i(t_i)+\sum_{r_i\in T_i} \alpha_i(r_i|t_i))\omega_i(t) - \sum_{r_i\in T_i} \alpha_i(t_i|r_i)\omega_i(t_{-i},r_i))/p_i(t_i) \ge p(t) \sum_{i\in S} v_i(c,t,\lambda,\alpha)$
$\forall t\in T, \forall S\subseteq N, \forall c\in C(S).$

To formulate the dual of this problem, let ρ^k denote the dual variables for (8), and let $\theta(S,c|t)$ denote the dual variables for (9). Then the dual is: choose (ρ,θ) so as to

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

(10) $\rho^1 + ... + \rho^K = 1,$

(11) $\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} ((\lambda_i(t_i) + \sum_{r_i\in T_i} \alpha_i(r_i|t_i))\theta(S,c|t) - \sum_{r_i\in T_i} \alpha_i(r_i|t_i)\theta(S,c|t_{-i},r_i)) - h_i(t_i) = 0,$
$\forall i\in N, \forall t\in T,$

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

Here (10) is the dual constraint for the primal variable z, and (11) is the dual constraint for the primal variable $\omega_i(t)$.

Given $(\lambda,\alpha,\delta)\in\Lambda$, let $F(\lambda,\alpha,\delta) \subseteq \Lambda$ be defined so that $(h,\beta,\gamma) \in F(\lambda,\alpha,\delta)$ iff there is an optimal solution (ρ,θ) to this dual problem such that $h = \sum_{k=1}^K \rho^k h^k,\ \beta = \sum_{k=1}^K \rho^k \beta^k,\ \gamma = \sum_{k=1}^K \rho^k \gamma^k.$

6.3 [core existence proof finished; severance payments and zero λ's]
By the Kakutani fixed-point theorem, the condition $(\lambda,\alpha,\gamma) \in F(\lambda,\alpha,\gamma)$ has a solution in Λ.
Let (ρ,θ) denote the dual solution that generaties this fixed point.
Using (5") and (11), we get, for every i in N and every t in T:

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

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

But these equations imply: $\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \theta(S,c,t) = 1, \ \forall i\in N, \forall t\in T.$
(If this failed for any t_{-i}, then we could sum these equations over all t_i such that $\sum_{S\supseteq\{i\}} \sum_{c\in C(S)} \theta(S,c,t) > 1$, cancel out the r_i terms that appear on both sides, and get a positive sum equalling a negative sum.)
So for each t, $\theta(\bullet|t)$ here satisfies the conditions for a balanced coalitional plan.
Thus, 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 the dual objective satisfies

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

But the optimal value of the dual is equal to the optimal value of the primal, which is the smallest expected subsidy that is required for an incentive-compatible mechanism that achieves some inhibitive alloction ω that is finely supported by (λ,α). Thus, this allocation ω is in the core, because it can be achieved by a feasible mechanism. QED

<u>Fact</u>. 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

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

So positive severance $\big(\sum_{c\in C(N)} \mu(c|t)\, u_i(c,t) + x_i(t) - \omega_i(t)\big)$ may be promised only when $\lambda_i(t_i)=0$.
<u>Proof of Fact</u>. The condition that (μ,x) achieves ω implies that every term in this sum is ≥ 0.
So it suffices to show that the sum is ≤ 0, as follows:

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

7. Coalitional durability

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 $D = \times_{i\in S} D_i$, such that each D_i is a nonempty finite set (representing the pure strategies for player i in this game), and the $z_i(d)$ numbers (representing the net sidepayment to i when players do d in the alternative game) satisfy $\sum_{i\in S} z_i(d) < 0$, $\forall d\in D$.

We are assuming that the alternative game always gives the blocking mediator some profit, and that the blocking mediator must announce the coalition that he is organizing, and cannot specify which equilibrium would be played in the alternative game.

We assume that the members of S vote independently about whether to form a blocking coalition and play Γ_S, which will happen iff they vote unanimously for it.

Then after such a vote, the players' beliefs about each other must be represented by some beliefs vector q in $\times_{i\in S} \Delta(T_i)$, where $q_i(t_i)$ denotes the probability of type t_i given that i 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: $q(t_{S-i}) = \prod_{j\in S-i} q_j(t_j)$, $\sigma(d|t_S) = \prod_{j\in S} \sigma_j(d_j|t_j)$,

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

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

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, 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.

We say that (μ,x) is coalitionally durable iff, for every $S\subseteq N$, and for alternative game Γ_S for the coalition S, there exists a beliefs vector q and a profile of strategies σ such that:

σ is an equilibrium of Γ_S with beliefs q,

$\forall i\in N$, $\forall t_i\in T_i$: if $U_i(\Gamma_S,\sigma|q,t_i) > U_i(\mu,x|q,t_i)$ 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 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 for whom all types would be willing to vote against this alternative.

Theorem 3. If a mechanism (μ,x) achieves an allocation in the core then (μ,x) is durable.

7.1 [proof of core implies durable]

Proof. Let ω be an allocation in the core that is achieved 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 (each knowing 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 (by trembling hand) vote for the alternative. 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

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

The equilibrium voting strategy must satisfy

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

$\tau_i(t_i) = 0$ 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.
We 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 $\psi_i = \prod_{j\in S-i} (\sum_{t_j\in T_j} p_j(t_j)\, \tau_j(t_j))$ denote the probability of all $S-i$ voting for the alternative.
If this limit of sequential equilibria does not satisfy the conditions for durability, it must be that, for every player i in S, there is at least one type t_i such that

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

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

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

$\nu(d|t_R) = 0$, 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, notice that, for any $i\in S$ and any $t_i\in T_i$,

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

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

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

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

The sequential equilibrium conditions on (τ,σ) imply that this expression is ≥ 0 and is

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

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

But tenability against ω contradicts the assumption that ω is inhibitive. So (μ,x) is durable. QED

8. Example 1. $N = \{1,2\}$, $T_1 = \{H,L\}$, $T_2=\{2\}$, $p(1H) = 1/2 = p(1L)$, so
$\lambda_2 = 1$, $\lambda_1(H) = 0.5+\alpha_1(H|L)-\alpha_1(L|H) \geq 0$, $\lambda_1(L) = 0.5+\alpha_1(L|H)-\alpha_1(H|L) \geq 0$.
1 is a seller with a single indivisible good. Value of 1's good to each depends on 1's type:

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	$1-8\alpha_1(L\|H)$

For such one-good, one-sided incomplete information games, the incentive-efficient set is characterized by in my paper in Roth, Game Theoretic Models of Bargaining, 1985.
Here this implies only that 1L must surely sell and the "1L not claim H" constraint must bind.
Let $\zeta = x_1(H) = E(\$ \text{ to } 1|H)$.
Let $\pi = P(\text{sell}|H) \in [0,1]$.
Then incentive efficiency implies $P(\text{sell}|L) = 1$, $E(\$ \text{ to } 1|L) = \zeta+1-\pi$.
All incentive-efficient mechanisms are verified (as in Thm 0) by
$\alpha_1(H|L) = 0.125$, $\alpha_1(L|H) = 0$, $\lambda_1(H) = 0.625$, $\lambda_1(L) = 0.375$, $\lambda_2 = 1$.
<u>Claim:</u> For this example, core solutions have no sidebets (so 2 pays what 1 is paid in each state) and no severance. (proof on next page).
So $\omega_1(H) =\zeta+5(1-\pi)$, $\omega_2(H) = 6\pi-\zeta$, $\omega_1(L) =\zeta+1-\pi$, $\omega_2(L) = 2-(\zeta+1-\pi)$.

With (λ,α) as above, 1's virtual value for his object is \$6 if H, \$1 if L, and so virtual core inequalities are: $\{1H\}$: $[0.625(\zeta+5-5\pi) - 0.125(\zeta+1-\pi)]/0.5 \geq 6$, so $\zeta - 6\pi + 6 \geq 6$,
$\{1H,2\}$: $[0.625(\zeta+5-5\pi) - 0.125(\zeta+1-\pi)]/0.5 + 6\pi - \zeta \geq 6$,
$\{1L\}$: $[(0.375+0.125)(\zeta+1-\pi)]/0.5 \geq 1$, so $\zeta+1-\pi \geq 1$,
$\{1L,2\}$: $[(0.375+0.125)(\zeta+1-\pi)]/0.5 + 2-(\zeta+1-\pi) \geq 2$,
$\{2\}$: $0.5(6\pi-\zeta) + 0.5(2-(\zeta+1-\pi)) \geq 0$, so $3.5\pi + 0.5 - \zeta \geq 0$.
(ζ,π) satisfies these conditions (and so is in core) iff $3.5\pi+0.5 \geq \zeta \geq 6\pi$ (implies $\pi \leq 0.2$).
We now show nothing else is in core:

Any incentive-efficient mechanism has $U_2 = 0.5(6\pi-\zeta) + 0.5(2-(\zeta+1-\pi)) = 3.5\pi + 0.5 - \zeta$.
So $3.5\pi+0.5 \geq \zeta$ is 2's participation constraint. If violated then $\{2\}$ would block alone.

The participation constraint for H (that $\{1\}$ should not block if H) is $\zeta \geq 5\pi$,
and this constraint with $3.5\pi+0.5 \geq \zeta$ implies that all core mechanisms must satisfy $1/3 \geq \pi$.

The core condition $\zeta \geq 6\pi$ says that 1H will not sell his good for less than his virtual value, which could be interpreted as a consequence of the virtual utility hypothesis.
With $1/3\geq\pi$, if the established mechanism had $6\pi > \zeta$ then a tenable blocking plan would be:
If H: with prob'y π, $\{1,2\}$ blocks, 2 gets good, 2 pays $\zeta+6(1-\pi)$, 1 paid ζ/π; and
with prob'y $1-\pi$, $\{1\}$ blocks, 1 keeps good , 1 is paid 0. If L: no block.
When 2 is invited, 2 expects $6-(\zeta+6(1-\pi)) = 6\pi-\zeta$, and blocking med'r gets $(6-\zeta/\pi)(1-\pi) > 0$.

8.1 Proof that no sidebets or severances can be paid in core mechanisms for Example 1.
Recall that we let $\zeta = x_1(H) = E(\$ \text{ to } 1|H)$, $\pi = P(\text{sell}|H) \in [0,1]$.
Then incentive efficiency implies $P(\text{sell}|L) = 1$, $E(\$ \text{ to } 1|L) = \zeta+1-\pi$.
Let $\varepsilon(t)\geq 0$ denote the severance payment offered to type t.
Let $\chi(t)$ denote the net tax 2 pays, above what 1 is paid, in the established plan if 1's type is t.
So in state t, 2 pays $\chi(t)$ plus whatever 1 is paid ($\chi(t)+\zeta$ if t=H, $\chi(t)+\zeta+1-\pi$ if t=L).
If $\chi(t) < 0$ then $-\chi(t)$ is a net subsidy from the mediator.
Step 1 In the core, we must have $\min\{\varepsilon(H),\varepsilon(L)\} = 0$ and $\chi(H)+\chi(L) = 0$.
If not, the blocking mediator could just imitate the established plan but 1 to pay an additional amount equal to $\min\{\varepsilon(H),\varepsilon(L)\}/2$.
Step 2 In core, if $\varepsilon(t) > 0$ then $\chi(t)+\varepsilon(t) \leq 0$.
Consider the blocking plan of imitating the established plan in state t but asking 1 to pay back an additional amount $\varepsilon(t)$, no blocking in the other state. This is tenable unless $\varepsilon(t) \leq -\chi(t)$.
Step 3 In core, we cannot have $\chi(L) > 0$.
Suppose $\chi(L) > 0$. By above steps, we must have $\varepsilon(L) = 0$, $\varepsilon(H) \leq \chi(H) = -\chi(L)$.
1H falsely claiming L in established plan would cost him $\zeta+5(1-\pi)-(\zeta+1-\pi) = 4(1-\pi)$.
Then the following blocking plan is tenable: the blocking mediator always has {1,2} form and imitates the established plan except, if 2's type is H then no blocking (return to established plan) occurs with probability $\min\{1,4(1-\pi)/\varepsilon(H)\}$. (This probability is 1 if $\varepsilon(H)=0$.)
This plan yields expected net tax $\min\{1,4(1-\pi)/\varepsilon(H)\}\chi(H)$ to the blocking mediator.
1H falsely claiming L would increase his expected severance by $\varepsilon(H)\min\{1,4(1-\pi)/\varepsilon(H)\}$ but would decrease his payment thereafter by $4(1-\pi)$ (which is not less than the severance gain).
Step 4 In core, we cannot have $\chi(H) > 0$.
Suppose $\chi(H) > 0$. By above steps, we must have $\varepsilon(L) = 0$, $\varepsilon(H) \leq \chi(H) = -\chi(L)$.
So the following blocking plan is tenable:
if H then {1,2} blocks, good is delivered to 2 with probability $\hat{\pi} = \max\{\pi-\chi(H)/4,0\}$,
2 pays $\zeta+\chi(H)-6(\pi-\hat{\pi})$, 1 is paid $\zeta-5(\pi-\hat{\pi})$;
if L then no block with prob'y $\max\{1,4\pi/\chi(H)\}$, else {1,2} blocks imitating established plan.
This plan gives the blocking mediator the expected net tax
$\chi(H) - \chi(H)(1-\min\{1,4\pi/\chi(H)\}) > 0$.
1L falsely claiming H would increase his expected severance pay by $\varepsilon(L)\min\{1,4\pi/\chi(H)\}$,
but it would decrease his expected payment after severance by more because
$(\zeta+1-\pi) - (\zeta-5(\pi-\hat{\pi}) + 1-\hat{\pi}) = 4(\pi-\hat{\pi}) = \max\{-\chi(H),0\} = \min\{\chi(H),4\pi\}$
$\geq \min\{\varepsilon(L), 4\pi\, \varepsilon(L)/\chi(H)\} = \varepsilon(L)\min\{1,4\pi/\chi(H)\}$.

9. Example 2. As in example 1, but now the value of 1's good to each player are as follows.

t_1	p	1's value of his good	2's value of 1's good	1's virtual value
H	0.5	3	7	$3+2\alpha_1(H\|L)$
L	0.5	2	0	$2-2\alpha_1(L\|H)$

For this example, my paper in Game Theoretic Models of Bargaining (1985) implies that incentive efficient mechanisms are those in which 1 always sells to 2.
So $P(\text{sell}|H) = P(\text{sell}|L) = 1$, and both 1's types must get the same expected payment.
Incentive-efficiency of all such mechanisms can be Thm0-verified by any (λ,α) such that
$\lambda_2 = 1$, $\lambda_1(H) = 1$, $0 = \lambda_2(L) = 0.5+\alpha_1(L|H)-\alpha_1(H|L)$, $3+2\alpha_1(H|L) \le 7$, $2-2\alpha_1(L|H) \le 0$.
That is, we let λ give zero weight to 1's bad type L, and we constrain α to make 1's virtual value for the good less than 2's value in each state, so that selling is always virtually efficient.
Let $\beta = \alpha_1(H|L) \le 2$. Then we must have $\alpha_1(L|H) = \beta-0.5 \ge 1$, and so $1.5 \le \beta \le 2$.
Let $\zeta = E(\$ \text{ paid to } 1)$ (same for both types in pooling mechanisms).
Let $\varepsilon(t) \ge 0$ denote the severance payment offered to type t of player 1.
Let $\chi(t)$ denote the net tax that 2 pays, above what 1 is paid, when 1's type is t.
So $\omega_1(H) = \zeta-\varepsilon(H)$, $\omega_2(H) = 7-(\zeta+\chi(H))$, $\omega_1(L) = \zeta-\varepsilon(H)$, $\omega_2(L) = -(\zeta+\chi(L))$.

In core, we must have $\min\{\varepsilon(H),\varepsilon(L)\} = 0$ and $\chi(H)+\chi(L) = 0$ (recall step 1 of 8.1).
With (λ,α) as above, let $\beta = \alpha_1(H|L) = 0.5+\alpha_1(L|H) \in [1.5,2]$. Suppose $\varepsilon(H) = 0$.
Then the virtual core inequalities are:
$\{1H\}$: $[(1+\beta-0.5)\zeta - \beta(\zeta-\varepsilon(L))]/0.5 \ge 3+2\beta$, so $\zeta+2\beta\varepsilon(L) \ge 3+2\beta$,
$\{1H,2\}$: $[(1+\beta-0.5)\zeta - \beta(\zeta-\varepsilon(L))]/0.5 + 7-(\zeta+\chi(H)) \ge 7$, so $2\beta\varepsilon(L)-\chi(H) \ge 0$
$\{1L\}$: $[\beta(\zeta-\varepsilon(L)) - (\beta-0.5)\zeta]/0.5 \ge 2-2(\beta-0.5)$, so $\zeta-2\beta\varepsilon(L) \ge 3-2\beta$
$\{1L,2\}$: $[\beta(\zeta-\varepsilon(L)) - (\beta-0.5)\zeta]/0.5 - (\zeta+\chi(L)) \ge 0$, so $\chi(H)-2\beta\varepsilon(L) \ge 0$,
$\{2\}$: $0.5(7-(\zeta+\chi(H))) - 0.5(\zeta+\chi(L)) \ge 0$, so $3.5 \ge \zeta$.

These conditions are satisfied and we have a core solution iff
$1.5 \le \beta \le 2$, $3 \le \zeta \le 3.5$, $\zeta-3 \ge 2\beta|1-\varepsilon(L)|$, $\chi(H) = 2\beta\varepsilon(L)$.
There are no solutions with $\chi(H)=0$ or $\varepsilon(L)=0$. In fact, all solutions have $\varepsilon(L) \ge 5/6$, $\chi(H) \ge 2.5$.
Solving for β in these conditions, we have found the core solutions satisfying:
$3 \le \zeta \le 3.5$, $\varepsilon(L)(\zeta-3) \ge \chi(H)|1-\varepsilon(L)|$, $4\varepsilon(L) \ge \chi(H) \ge 3\varepsilon(L)$, $\varepsilon(L) \ge 5/6$.
<u>Claim</u> There are no other core solutions for this game.

9.1 <u>Claim</u> In the core, $3 \le \zeta \le 3.5$, $\varepsilon(L)(\zeta-3) \ge \chi(H)|1-\varepsilon(L)|$, $4\varepsilon(L) \ge \chi(H) \ge 3\varepsilon(L) \ge 2.5$.
<u>Step 1</u> In the core, we must have $\min\{\varepsilon(H),\varepsilon(L)\} = 0$ and $\chi(H)+\chi(L) = 0$. [as in 8.1]
<u>Step 2</u> If $\varepsilon(t) > 0$ then $\chi(t)+\varepsilon(t) \le 0$. [as in 8.1]
<u>Step 3</u> $3 \le \zeta \le 3.5$.
If $\zeta > 3.5$ then {2} going alone in both states blocks. If $\zeta < 3$ then {1} alone if t=H blocks.
<u>Step 4</u> $\chi(H) \ge 0.5$, and so $\varepsilon(H) = 0$ and $\chi(L) = -\chi(H) \le -0.5$.
With $\zeta \le 3.5$, if $\chi(H) < 0.5$ then the following blocking plan is tenable:
If H then {1} blocks, 1 paid $\zeta-2$; if L then {1,2} blocks, 1 keeps, 2 pays $\zeta-\chi(H)$, 1 paid $\zeta-2$.
(Expected profit is $0.5(\zeta-\chi(H)) - (\zeta-2) = 0.5(4-\zeta-\chi(H)) \ge 0.5(4-3.5-\chi(H)) > 0$.)
<u>Step 5</u> $\varepsilon(L)(\zeta-3) \ge \chi(H)(1-\varepsilon(L))$. So $\varepsilon(L) \ge \chi(H)/(\chi(H)+\zeta-3) \ge 0.5/(0.5+3.5-3) = 0.5$.
From above, $\varepsilon(L)(\zeta-3) < \chi(H)(1-\varepsilon(L))$ would imply $\varepsilon(L) < 1$. But then following is tenable:
If L then no block. If H then with prob'y $\varepsilon(L)$, {1} blocks and 1 is paid $\zeta-3$;
and with prob'y $1-\varepsilon(L)$, {1,2} blocks, 2 gets good, 2 pays $\zeta+\chi(H)$, 1 is paid ζ.
The $\varepsilon(L)$ probability deters L from claiming H to get $\varepsilon(L)$ severance in this blocking plan.
<u>Step 6</u> $\varepsilon(L)(\zeta-3) \ge \chi(H)(\varepsilon(L)-1)$.
From above, $0 < \chi(H)(\varepsilon(L)-1) - \varepsilon(L)(\zeta-3)$ implies $\varepsilon(L) > 1$. Then following is tenable:
If L then {1} blocks, 1 is paid $\zeta-3$. If H: with prob'y $1/\varepsilon(L)$, there is no blocking; and with prob'y $1-1/\varepsilon(L)$, {1,2} blocks, 2 gets good, 2 pays $\zeta+\chi(H)$, 1 is paid ζ.
1L is willing to accept $\zeta-3$ to keep good worth 2 and have severance $\varepsilon(L)>1$, and lying would decrease 1L's probability of getting the $\varepsilon(L)$ severance payment by $1/\varepsilon(L)$.
Blocking mediator's expected profit is $0.5\,(1-1/\varepsilon(L))\chi(H)-0.5(\zeta-3) > 0$.
<u>Step 7</u> $4\varepsilon(L) \ge \chi(H)$.
Suppose $\chi(L) > 4\varepsilon(L)$. Then the following is a tenable blocking plan:
If H: with prob'y $\varepsilon(L)/(1+\varepsilon(L))$, {1,2} blocks, 1 keeps good, 2 pays $\zeta+\chi(H)-7$, 1 paid $\zeta-3$;
and with proby $1/(1+\varepsilon(L))$, {1,2} blocks, 2 gets good, 2 pays $\zeta+\chi(H)$, 1 is paid ζ.
If L: with prob'y $\varepsilon(L)/(1+\varepsilon(L))$, {1,2} blocks, 2 gets good, 2 pays $\zeta-\chi(H)$, 1 paid ζ;
with proby $1/(1+\varepsilon(L))$, there is no blocking (return to established plan).
Blocking mediator's expected profit is
$0.5\mathbf{(}\varepsilon(L)(\chi(H)-4) + \chi(H) - \varepsilon(L)\chi(H)\mathbf{)}/(1+\varepsilon(L)) = 0.5\mathbf{(}\chi(H)-4\varepsilon(L)\mathbf{)}/(1+\varepsilon(L)) > 0$.
The $\varepsilon(L)/(1+\varepsilon(L))$ probability of no sale when H is reported (which would hurt L by \$1) deters L from claiming H to get $1/(1+\varepsilon(L))$ increased probability of $\varepsilon(H)$ severance pay.
<u>Step 8</u> $\chi(H) \ge 3\varepsilon(L)$.
Suppose $\chi(H) < 3\varepsilon(L)$. Then the following is a tenable blocking plan:
If H: with prob'y $\min\{1,1/\varepsilon(L)\}$ there is no blocking; and
with prob'y $\max\{0,1-1/\varepsilon(L)\}$, {1,2} blocks, 2 gets good, 2 pays $\zeta+\chi(H)$, 1 paid ζ.
If L: with prob'y $Q=\min\{1,\varepsilon(L)\}$, {1,2} blocks, 1 keeps good, 2 pays $\zeta-\chi(H)$, 1 paid $\zeta-2-Q$;
and with prob'y $1-Q$, {1,2} blocks, 2 gets good, 2 pays $\zeta-\chi(H)$, 1 paid $\zeta-\varepsilon(L)$.
When $\varepsilon(L)\le 1$, blocking mediator's EProfit is $0.5(2\varepsilon(L)+\varepsilon(L)-\chi(H)) > 0$, and the randomization for L deters H from claiming L ($Q=\varepsilon(L)$ yields $Q(3-2-Q)+(1-Q)\varepsilon(L) = 0$).
When $\varepsilon(L)>1$, blocking medr's EProfit is $0.5(3-\chi(H)+(1-1/\varepsilon(L))\chi(H)) = 0.5(3-\chi(H)/\varepsilon(L)) > 0$, and the randomization for H deters L from claiming H (L gains $\varepsilon(L)-1$ either way).

9.2 More specific mechanisms for our example where severance pay is required.
Example 2:
Player 1 is seller of a single good. 1's type $t_1 \in \{H,L\}$. $p_1(H) = p_1(L) = 0.5$.
Player 2 is the only potential buyer, and she has no private information.
If t_1=H then the good's value is \$3 to player 1, its value is \$7 to player 2.
If t_1=L then the good's value is \$2 to player 1, its value is \$0 to player 2.

Consider the incentive-compatible mechanism, where
Both types sell for \$3.5 (no subsidies or taxes, no severance offer).
It is not in the core. The following is a tenable blocking plan against this mechanism:
If t_1=H then {1} blocks, keeps the good, is paid \$1.5 (by the blocking mediator);
If t_1=L then {1,2} block, 1 keeps the good, is paid \$1.5, but 2 pays \$3.5 to blocking mediator.

The following incentive compatible mechanism is in the core and has the smallest severance pay:
If t_1=H then 1 sells the good, is paid \$3.5, but 2 pays \$6;
If t_1=L then 1 sells the good, is paid \$3.5, but 2 pays \$1, and 1 is offered \$ε severance,
where $\varepsilon = \varepsilon(L) = \$5/6$.
If we had $\varepsilon < 5/6$, then the following blocking plan would be tenable against this mechanism:
If t_1=H then,
with probability ε, {1} blocks, 1 keeps the good but is paid \$0.5,
with probability 1−ε, {1,2} block, 1 sells the good, is paid \$3.5, but 2 pays \$6.
If t_1=L then do not block (return to fulfill the established mechanism).
(With $\varepsilon < 5/6$, the blocking mediator would expect $\varepsilon(-0.5)+(1-\varepsilon)(3.5) > 0$,
and 1's type L would expect $\varepsilon(-0.5)+(1-\varepsilon)(3.5-2) + \varepsilon = 1.5 \leq 1.5$.)

10. Example 3: N={1,2}, T_1={a,b}=T_2, each type has prob'y 1/2, independent of other's type. Each player has an endowment of one unit of his own good (1 has good 1, 2 has good 2).

t_1	t_2	p	1's value of good 1	2's value of good 1	1's value of good 2	2's value of good 2
a	a	0.25	7	6	6	7
a	b	0.25	3	6	1	7
b	a	0.25	7	1	6	3
b	b	0.25	2	1	1	2

The only gains from trade are from an a-type selling his good to a b-type, but the b is more eager than a to sell to b. Each thinks that the expected value of the other's good is 3.5. But each would only want to sell his own good for 3.5 if the other is b. Virtual values depend on α as below:

t_1	t_2	1's virtual value of good 1	2's virtual value of good 1	1's virtual value of good 2	2's virtual value of good 2
a	a	7	6	6	7
a	b	$3+2\alpha_1(a\|b)$	6	1	7
b	a	7	1	6	$3+2\alpha_2(a\|b)$
b	b	$2-2\alpha_1(b\|a)$	1	1	$2-2\alpha_2(b\|a)$

So incentive-efficient mechanisms can also have trade between b-types, with $\alpha_i(b|a) \geq 0.5$. Considering mechanisms that respect the symmetry between the two players, if there are no subsidies and no severance payments, then the only incentive-efficient mechanism has b's always buying the other's endowment for \$3.5, a's not buying.

t_1	t_2	1 consumes	\$ to 1	U_1	lie $\hat{U}_1$	2 consumes	\$ to 2	U_2	lie $\hat{U}_2$
a	a	(1,0)	0	7	9.5	(0,1)	0	7	9.5
a	b	(0,0)	3.5	3.5	1	(1,1)	−3.5	9.5	7
b	a	(1,1)	−3.5	9.5	7	(0,0)	3.5	3.5	1
b	b	(0,1)	0	1	3.5	(1,0)	0	1	3.5

But under this mechanism, when types are (b,b), they trade goods with net sidepayment \$0.
So a blocking mediator could make both better off with the tenable plan: "coalition {1,2} blocks with no trade if both are type-b (paying up to \$1 to blocking mediator), otherwise don't block."
A core allocation with the mechanism that achieves it is shown below. In this mechanism, b's have a \$2.50 sidebet that the other's type is b. Winnings at (b,b) tend to deter blocking there. Severances are needed at (b,b) so that blocking at (a,b) and (b,a) is not incentive compatible.
This core allocation is supported by $\lambda_i(a)=1$, $\lambda_i(b) = 0$, $\alpha_i(b|a) = 0.5$, $\alpha_i(a|b) = 1$, for each i.

t_1	t_2	1 consumes	\$ to 1	U_1	lie $\hat{U}_1$	ω_1	2 consumes	\$ to 2	U_2	lie $\hat{U}_2$	ω_2
a	a	(1,0)	0	7	7	7	(0,1)	0	7	7	7
a	b	(0,0)	3.5	3.5	3.5	3.5	(1,1)	−6	7	7	7
b	a	(1,1)	−6	7	7	7	(0,0)	3.5	3.5	3.5	3.5
b	b	(0,1)	2.5	3.5	3.5	2.25	(1,0)	2.5	3.5	3.5	2.25