The Center for Mathematical Studies in Economics and Management Science
Northwestern University, Evanston, Illinois 60208

Discussion Paper No. 907

# FICTITIOUS-TRANSFER SOLUTIONS IN COOPERATIVE GAME THEORY

by

by Roger B. Myerson[1]

September 1990

Abstract. The method of fictitious transfers (or $\lambda$-transfers) is used in cooperative game theory to extend solution concepts from the transferable utility case to the more general case of games without transferable utility. The importance of this idea is examined in the context of the history of its development and some recent results. A new interpretation of the inner core (or $\lambda$-transfer core) is discussed.

---

[1] J. L. Kellogg Graduate School of Management, Northwestern University, Evanston, Illinois 60208.

# FICTITIOUS-TRANSFER SOLUTIONS IN COOPERATIVE GAME THEORY

by Roger B. Myerson

It is easy to see that games with transferable utility (TU) really are just a special case of games with nontransferable utility (NTU), because transfer activities can be remodeled as strategic options in a game "without transferable utility." Thus, it is natural to ask why game theorists should have devoted substantial efforts to developing solution concepts for games with transferable utility. When I teach cooperative game theory now, more than 45 years after von Neumann and Morgenstern [1944], I try to motivate the old emphasis on coalitional games with transferable utility by two propositions. First, because coalitional interactions can be very complicated, we may initially want to simplify our analysis by assuming transferable utility, so that the set of feasible utility allocations for each coalition can be described by a single number. Second, by the method of fictitious transfers of weighted utility (or $\lambda$-transfers), we can easily generalize any solution concept for transferable-utility games to the case of games without transferable utility. Thus, from the perspective of 1990, the method of fictitious transfers appears to justify the original decision of von Neumann and Morgenstern [1944] to concentrate on games with transferable utility. However, the method of fictitious transfers was recognized only after the pivotal breakthrough of Harsanyi [1963]. My purpose in this paper is to reexamine the importance of this method, in the context of the history of its development and some recent results.

The assumption of transferable utility seems to have been adopted almost without discussion in the seminal work of von Neumann [1928] and von Neumann and Morgenstern [1944]. In the index of von Neumann and Morgenstern [1944], there are only two references to transferability of utility. Remarkably, in the book where it is first proven that cardinal utility is measurable from the risk preferences of a rational decision maker, the main justification of transferability appears to be to avoid measurement questions. Von Neumann and Morgenstern [1944, page 8] assert:

> We wish to concentrate on one problem -- which is not that of measurement of utilities and of preferences -- and we shall therefore attempt to simplify all other characteristics as far as reasonably possible. We shall therefore assume that the aim of all participants in the economic system, consumers as well as entrepreneurs, is money, or equivalently a single monetary commodity. This is supposed to be unrestictedly divisible and substitutable, freely transferable and identical, even in the quantitative sense, with whatever "satisfaction" or "utility" is desired by each participant.

But from von Neumann and Morgenstern's own theory of utility, we now understand that, with great generality, the options available to a coalition can be represented by a convex set of expected utility allocations. This result was used by Nash [1950]; but it was not until his second paper on bargaining (Nash [1953]) that Nash emphasized that, by studying games without transferable utility, he was dropping one of the restrictive assumptions of the previous work in game theory.

In the same volume of Econometrica, Shapley and Shubik [1953] published a one-page abstract that describes how von Neumann and Morgenstern's solution theory can be generalized to games without transferable utility, and they found that nontransferable utility did not seem to create major new conceptual problems. However, if Shapley and Shubik had also considered the problem of generalizing the Shapley [1953] value, no doubt they would have had more difficulties to report. By the late 1950s, published work in game theory reflected a much greater frustration with the prevailing assumption of transferable utility. Luce and Raiffa [1957, pp. 233-4] remark:

> The assumption ... that there exists a transferable utility in which sidepayments are effected is exceedingly restrictive -- for many purposes it renders n-person theory next to useless.

It is at this point that John Harsanyi began the search for a one point solution concept for n-person cooperative games without transferable utility. He had two partial solutions to build on: the Shapley value for n-person games with transferable utility and the Nash bargaining solution for two-person games without transferable utility. Each of these solution concepts had been compellingly derived as the unique solution satisfying a natural set of axioms, and each selects a unique payoff allocation for each game in its respective domain. In the fourth Princeton volume on game theory, Harsanyi [1959] published a general solution concept that determines a payoff allocation for any n-person cooperative game with nontransferable utility. He showed that this bargaining solution coincides with the Nash bargaining solution in the two-person case, and coincides with the Shapley value in the transferable utility case. Thus, for the first time, it was suggested that the Nash bargaining solution and the Shapley value might be understood as special cases of a more general solution concept.

There is a easy way to generalize the Nash bargaining solution to any number of players, by simply maximizing the product of all players' utility gains over the disagreement point. However, this simple n-person Nash bargaining solution is unacceptable because it neglects the power of subcoalitions, and thus it does not coincide with the Shapley value in the TU case. In particular, this simple multiplayer Nash bargaining solution would generally give positive payoff to a dummy player with no strategic options. In his first general solution concept, Harsanyi [1959] brought all coalitions into the analysis by assuming that every coalition (or nonempty set of players) would negotiate a vector of "dividends," such that the sum of all coalitions' dividend vectors would be a feasible allocation for the grand coalition of all players. Within each coalition S, Harsanyi [1959] tried to determine the dividend vector by applying the simple multiplayer Nash bargaining solution, where the disagreement point would be the sum of all dividend vectors for subsets of the coalition S. A difficulty with this approach arises because, for games with three or more players (such as the three-person majority game), the disagreement point for some coalitions may be outside of the feasible set, so that negative dividends may be required. Harsanyi's [1959] attempt to deal with this negative-dividend case seemed unsatisfactory, and Isbell [1960] showed that the solution can violate the basic individual-rationality constraints (that each player must get at least what he could earn on his own). Isbell [1960] suggested an alternative solution concept that always satisfies individual rationality and generalizes the Shapley value, but does not generalize the Nash bargaining solution. (From a modern perspective, Isbell's solution looks more like a generalization of the Kalai-Smorodinsky [1975] solution.)

In response to such objections, Harsanyi re-examined the Nash bargaining solution, to identify other ways of extending it to the multiplayer case. The key insight appears as Theorem 1 in Harsanyi [1963]. We may state it here as follows. In a two-person given game, let $V(\{1,2\})$ denote the set of feasible utility allocations for players 1 and 2 when they cooperate. Let $\Delta(C_i)$ denote the set of randomized strategies for player i, and let $u_i(\sigma_1,\sigma_2)$ denote the expected utility payoff for player i when players 1 and 2 use the randomized strategies $\sigma_1$ and $\sigma_2$ respectively. Then $(x_1,x_2)$ is the Nash bargaining solution and $(\tau_1,\tau_2)$ are the rational threats for the game with variable threats iff there exist nonnegative numbers $\lambda_1$ and $\lambda_2$, not both zero, such that

(1) $$(x_1,x_2) \in V(\{1,2\}),$$

(2) $$\lambda_1 x_1 + \lambda_2 x_2 = \max_{w\in V(\{1,2\})} (\lambda_1 w_1 + \lambda_2 w_2),$$

(3) $$\lambda_1(x_1 - u_1(\tau_1,\tau_2)) = \lambda_2(x_2 - u_2(\tau_1,\tau_2)),$$

(4) $$\tau_1 \in \operatorname{argmax}_{\sigma_1\in\Delta(C_1)} (\lambda_1 u_1(\sigma_1,\tau_2) - \lambda_2 u_2(\sigma_1,\tau_2)),$$

(5) $$\tau_2 \in \operatorname{argmax}_{\sigma_2\in\Delta(C_2)} (\lambda_2 u_2(\tau_1,\sigma_2) - \lambda_1 u_1(\tau_1,\sigma_2)).$$

These conditions have a natural interpretation. With nontransferable utility, we have no grounds for interpersonal comparison of utility, so we may feel free to rescale either player's utility separately by a positive scaling factor or utility weight $\lambda_i$. Now, in a rescaled version of the game, pretend that the weighted-utility payoffs are transferable. Suppose that each player chooses a threat that maximizes the difference between his own and the other player's weighted-utility payoff (4,5), and they then divide among themselves the maximal transferable weighted-utility worth that they can earn together (2), in such a way that each player gets the same weighted-utility gain over the threat point (3). If the resulting utility allocation would actually be feasible even without these fictitious transfers (1), then it is a Nash bargaining solution.

Given a game in strategic (or normal) form, Von Neumann and Morgenstern [1944] defined the worth of a coalition S to be the maximum total payoff that the members of S can guarantee themselves against the most offensive threat from the complementary coalition N\S (where N denotes the set of all players). Conditions (4) and (5), derived from Nash's [1953] rational threats criterion, pose a challenge to von Neumann and Morgenstern's definition of the characteristic function. For a game with transferable utility, condition (2) just implies that $\lambda_1 = \lambda_2$, and then (4) and (5) assert that the worth of each one-person coalition should be defined in terms of a threat game where each player seeks to maximize the difference between the worth of his (one-person) coalition and the complementary coalition. Harsanyi [1963] generalized this criterion to create a new way to derive coalitional-form games from games in strategic form, for both the TU and NTU cases. Around the same time, Aumann and Peleg [1960] and Jentzsch [1964] developed other ways of deriving NTU coalitional-form games from strategic-form games, but Aumann and Peleg and Jentzsch's derivations coincide with von Neumann and Morgenstern's minimax definition in the TU case.

For our present purposes, however, the most important consequence of these conditions (1)-(5) is that they revealed new ways to extend the Nash bargaining solution to games with more than two players. We may present these ideas most simply by assuming that we are given a game in NTU coalitional form. When N denotes the finite set of players, an <u>NTU game in coalitional form</u> is a function V such that, for each coalition S that is a nonempty subset of N, V(S) is a convex subset of $\mathbb{R}^S$ (the set of all vectors of real numbers indexed on the members of S). We interpret V(S) as the set of expected utility allocations that the members of S could cooperatively guarantee themselves.

We assume throughout this paper that V has the following properties, for all coalitions S and T:

(6) $V(S)$ is a nonempty closed convex subset of $\mathbb{R}^S$;

(7) if $x \in \mathbb{R}^S$, $y \in V(S)$, and $x_i \le y_i$, $\forall i \in S$, then $x \in V(S)$;

(8) if $S \cap T = \emptyset$ then $V(S \cup T) \supseteq \{x \in \mathbb{R}^{S \cup T} \mid (x_i)_{i\in S} \in V(S),\ (x_j)_{j\in T} \in V(T)\}$.

(Here $x_i$ and $y_i$ respectively denote the i-components of the vectors x and y.) Condition (7) asserts that each set V(S) is <u>comprehensive</u> and (8) asserts that V is <u>superadditive</u>. As an additional technical condition, we may say that V is <u>finitely generated</u> or <u>polyhedral</u> iff, for every coalition S, there exist a finite set of points W(S) such that V(S) is the smallest convex comprehensive set containing W(S); that is,

(9) $x \in V(S)$ iff there exists a function $q:W(S) \to \mathbb{R}$ such that $q(y) \ge 0$ for every y in W(S), $\Sigma_{z\in W(S)}\, q(z) = 1$, and $x_i \le \Sigma_{z\in W(S)}\, q(z)z_i$, $\forall i \in S$.

Let $\partial V(S)$ denote the boundary of V(S), so that $x \in \partial V(S)$ iff $x \in V(S)$ and there does not exist any y in V(S) such that $y_i > x_i$ for all i in S.

A feasible payoff allocation $\bar{x}$ is a <u>Harsanyi NTU value</u>, in the sense of Harsanyi's [1963] simplified bargaining solution, if there exists a nonnegative vector $\lambda = (\lambda_i)_{i\in N}$ and, for each coalition S, there exists a coalitional allocation $x^S = (x_i^S)_{i\in S}$, such that:

(10) $\bar{x} = x^N \in \operatorname{argmax}_{w\in V(N)} \Sigma_{i\in N}\, \lambda_i w_i$;

(11) $x^S \in \partial V(S)$, $\forall S \subseteq N$;

(12) $\lambda_i(x_i^S - x_i^{S-j}) = \lambda_j(x_j^S - x_j^{S-i})$, $\forall S \subseteq N$, $\forall i \in S$, $\forall j \in S-i$.

Condition (10) generalizes condition (2) of the Nash bargaining solution.

Condition (12), called balanced contributions by Myerson [1980], asserts that each pair of players make equal weighted-utility contributions to each other in each coalition, and generalizes the $\lambda$-equity condition (3) of the Nash bargaining solution. In the domain of two-person games, the Harsanyi NTU value coincides with the Nash bargaining solution. In the domain of coalitional games with transferable utility (that is, where there exists some TU coalitional game v such that

$$V(S) = \{x \in \mathbb{R}^S \mid \Sigma_{i \in S} x_i \leq v(S)\}$$

for each coalition S), the Harsanyi NTU value coincides with the Shapley value.

Shapley [1964] (see also Aumann [1967] and Shapley [1969]) found Harsanyi's NTU value to be hard to work with in large games, and so he proposed another NTU value that is also naturally derived from Harsanyi's conditions (1)-(3) for the Nash bargaining solution. Let $\mathbb{R}^N_{++}$ denote the set of all strictly positive vectors indexed on N, so

$$\mathbb{R}^N_{++} = \{\lambda = (\lambda_i)_{i \in N} \mid \lambda_i > 0, \ \forall i \in N\}.$$

For any positive vector $\lambda$ in $\mathbb{R}^N_{++}$, let $F(V,\lambda)$ be a set of TU coalitional games such that $v \in F(V,\lambda)$ iff

(13) $$v(N) \geq \max_{y \in V(N)} \Sigma_{i \in N} \lambda_i y_i, \quad \text{and} \quad v(S) = \max_{y \in V(S)} \Sigma_{i \in S} \lambda_i y_i, \ \forall S \subset N.$$

That is, $F(V,\lambda)$ is the set of TU coalitional game that we would get if we enlarged the feasible set for the grand coalition N, rescaled each player i's utility by the weighting factor $\lambda_i$, and then (fictitiously) allowed every coalition to make interpersonal transfers of these weighted-utility payoffs. Now let

$$\phi(v) = (\phi_i(v))_{i \in N}$$

denote the Shapley value of any game v in $F(V,\lambda)$. If we believe in the Shapley value as a predictive solution concept for TU games, then we should predict the weighted-utility allocation $\phi(v)$ in the fictitious game with transferable weighted utility. To convert each player i's payoff $\phi_i(v)$ back to the original utility scale, we must divide by weight $\lambda_i$. So let

(14) $$\Phi(V,\lambda) = \{(\phi_i(v)/\lambda_i)_{i\in N} \mid v \in F(V,\lambda)\}.$$

It can be shown that, for any such positive vector $\lambda$, $\bar{x} \in \Phi(V,\lambda)$ if and only if there exist coalitional allocations $x^S$, for each coalition S, such that

(15) $$\bar{x} = x^N, \quad \forall i \in N;$$

(16) $$\Sigma_{i\in N} \lambda_i x_i^N \geq \max_{y\in V(N)} \Sigma_{i\in N} \lambda_i y_i;$$

(17) $$\Sigma_{i\in S} \lambda_i x_i^S = \max_{y\in V(S)} \Sigma_{i\in S} \lambda_i y_i, \quad \forall S \subset N;$$

(18) $$\lambda_i(x_i^S - x_i^{S-j}) = \lambda_j(x_j^S - x_j^{S-i}), \quad \forall S \subseteq N, \forall i \in S, \forall j \in S-i.$$

Thus, the allocations in $\Phi(V,\lambda)$ satisfy an analogue of the conditions (10)-(12) for the TU games that is derived from V by expanding the grand coalition's feasible set and pretending that $\lambda$-weighted utility is transferable for all coalitions. An allocation $\bar{x}$ in $\Phi(V,\lambda)$ is a Shapley NTU value of V if it is actually feasible without any such fictitious transfers, that is, if $\bar{x} \in \Phi(V,\lambda) \cap V(N)$.

To be able to prove a general existence theorem, we must define Shapley NTU values more broadly, in a way that allows some $\lambda_i$ components to go to zero. However, condition (14) is undefined when $\lambda_i$ is zero, so some limit-condition is needed. There are several ways of constructing such a limit condition; here we consider a technical formulation slightly different from that of Shapley [1969]. Let us say that $\bar{x}$ is a <u>Shapley NTU value</u> of V iff $\bar{x} \in V(N)$ and, for every positive number $\varepsilon$, there exists vector $\lambda$ in $\mathbb{R}^N_{++}$ such that and a TU coalitional game v in $F(V,\lambda)$ such that

(19) $$\bar{x}_i + \varepsilon \geq \phi_i(v)/\lambda_i, \quad \forall i \in N.$$

That is, a Shapley NTU value is a feasible allocation $\bar{x}$ such that, by slightly expanding the grand coalition's feasible set and making weighted utility transferable, we could generate a game for which the Shapley value is not substantially better than $\bar{x}$ (in the original utility scales) for any player.

Owen [1972] proposed a third NTU value that also extends by the Nash bargaining solution and the TU Shapley value, but Owen's NTU value requires the solution of a complicated set of differential equations and has not attracted much further interest. Shapley's and Harsanyi's NTU values have been subsequently derived axiomatically by Aumann [1985] and Hart [1985] for games where the boundary of V(N) is differentiable. On the other hand, Roth [1980] has given an example of a game in which the boundary of V(N) is not differentiable and the Shapley and Harsanyi NTU values seem quite unreasonable.

Shapley [1969] remarked that the method of fictitious transfers could be used to extend any TU solution concept to the more general NTU case. To see how, suppose that we are given a solution concept that specifies a set of solutions G(v) for each TU coalitional game v. Then to extend this solution concept G to NTU games by the <u>method of fictitious transfers</u>, we may say that an allocation $\bar{x}$ is a solution for V iff $\bar{x} \in V(N)$ and, for every (small) positive number $\varepsilon$, there exists a positive vector $\lambda$ in $\mathbb{R}^N_{++}$, a TU coalitional game v in $F(V,\lambda)$, and a vector z in G(v) such that

$$(20) \qquad \bar{x}_i + \varepsilon \geq z_i/\lambda_i, \quad \forall i \in N.$$

That is, a solution of V is a feasible allocation $\bar{x}$ such that, by (slightly) increasing the feasible set for the grand coalition and making weighted utility transferable for all coalitions, we could generate a game for which there exist solutions that are not substantially better than $\bar{x}$ for any player.

When we use the core as our TU solution concept, the NTU solution concept generated by the method of fictitious transfers is called the inner core (see Shubik [1982], page 155). As the name suggests, the inner core is a subset of the core as originally defined for NTU games. The core of V has been defined (see Aumann and Peleg [1960]) to be the set of all allocations x such that $x \in V(N)$ and there does not exist any nonempty coalition S and any allocation y in V(S) such that $y_i > x_i$ for every player i in S. That is, x is in the core iff it is feasible for the grand coalition N, and there does not exist any coalition that could guarantee a higher payoff to all its members.

The comparison between the core and the inner core can provide us with a new test of power of the method of fictitious transfers. As we have seen, fictitious transfers were first used by Shapley in the early 1960s, to define a tractable generalization of the Shapley value for NTU games. At that time, there was not a controversy about the definition of the core for NTU coalitional games (although there was controversy about how to construct a game in NTU coalitional form to represent any given strategic-form game). So now, unasked-for, the method of fictitious transfers gives us an alternative way to define cores for NTU games. If the method of fictitious transfers is really an appropriate way to generate NTU solution concepts, then there should be some fundamental rationale for considering the inner core to be the appropriate extension of the TU core to NTU games.

To pose this question more concretely, let us consider the Banker game analyzed by Owen [1972]. In this three-person game, player 1 is the worker, player 2 is the helper, and player 3 is the banker. By himself, each player can get zero payoff. The pairs {1,3} and {2,3} cannot do any better, because the worker and helper are both needed to create anything of positive value. Players 1 and 2 together can generate a payoff of 100 for player 1, and player 1 can try to transfer any fraction of this payoff to player 2 but, without the banker, three-quarters of the transfered amount will be lost. With the banker, the players can transfer utility among themselves in any way. So we may let

$$V(\{i\}) = \{x_i \mid x_i \le 0\}, \quad \forall i \in N = \{1,2,3\},$$

$$V(\{1,2\}) = \{(x_1,x_2) \mid x_1 \le 100,\ x_2 \le (100 - x_1)/4,\ x_2 \le 25\},$$

$$V(\{1,3\}) = \{(x_1,x_3) \mid x_1 + x_3 \le 0\},$$

$$V(\{2,3\}) = \{(x_2,x_3) \mid x_2 + x_3 \le 0\}, \quad \text{and}$$

$$V(\{1,2,3\}) = \{(x_1,x_2,x_3) \mid x_1 + x_2 + x_3 \le 100\}.$$

The core for this game is

$$\{(x_1,x_2,x_3) \mid x_1 + x_2 + x_3 = 100,\ x_1 + 4x_2 \ge 100,\ x_i \ge 0,\ \forall i\},$$

which is a triangle with corners (100,0,0), (0,100,0), and (0,25,75). To compute the inner core, let $v^\lambda$ denote the minimal TU game in $F(V,\lambda)$, so

$$v^\lambda(S) = \max_{y \in V(S)} \Sigma_{i \in S}\ \lambda_i y_i.$$

Notice that $v^\lambda(\{1,2,3\})$ would be $+\infty$ unless $\lambda_1 = \lambda_2 = \lambda_3$. When all $\lambda_i$ are equal and positive,

$$\begin{aligned} v^\lambda(S) &= 100\lambda_1 \quad \text{if} \quad S \supseteq \{1,2\}, \\ &= 0 \qquad\quad \text{otherwise}, \end{aligned}$$

and so the core of the TU game $v^\lambda$ is

$$\{(w_1,w_2,w_3) \mid w_1 + w_2 = 100\lambda_1,\ w_1 \ge 0,\ w_2 \ge 0,\ w_3 = 0\}.$$

Dividing payoffs by $\lambda_1$ to translate them back into the original utility scales, the core of $v^\lambda$ then maps to the set

$$\{(x_1,x_2,x_3) \mid x_1 + x_2 = 100,\ x_1 \ge 0,\ x_2 \ge 0,\ x_3 = 0\},$$

which is in V(N) and therefore is the inner core. Increasing the worth of the grand coalition would not generate any other core allocations that are feasible in the original game. Thus, the inner core is the portion of the core where player 3 (the banker) gets zero.

The unique Shapley NTU value for this example is (50,50,0), which is the Shapley value of the TU game $v^{\lambda}$ when $\lambda_1 = \lambda_2 = \lambda_3 = 1$. The Harsanyi NTU value is (40,40,20), which satisfies conditions (10)-(12) when

$$\lambda_1 = \lambda_2 = \lambda_3 = 1,$$

$$x_1^{\{1,2\}} = 20, \quad x_2^{\{1,2\}} = 20$$

$$x_1^{\{1,2,3\}} = 40, \quad x_2^{\{1,2,3\}} = 40, \quad x_3^{\{1,2,3\}} = 20,$$

and $x_i^S = 0$ for each i in any other coalitions S.

The Owen NTU value for this game is (51.88, 47.57, 0.58). Owen [1972] suggests intuitively that a reasonable solution should give a positive payoff to player 3, but the inner core and the Shapley NTU value contradict this intuition. To support the inner core, we now construct an intuitive justification for giving zero to the banker.

To be specific, suppose that the players expect to get the Harsanyi NTU allocation (40,40,20) unless some blocking coalition forms. The allocation (40,40,20) is in the core, as originally defined, so it might seem that no such blocking coalition could be organized. But suppose that a blocking agent (an outside mediator who is not a player) can try to organize a blocking coalition according to a pre-announced randomized plan. That is, suppose that an agent will approach some set of players and ask them to sign over power of attorney to him for all bargaining purposes. If any player who is approached by the agent refuses to give him power of attorney, then the agent will not block the (40,40,20) allocation; but if he gets power of attorney from all the players that he approaches, then the agent will block (40,40,20) and implement instead some allocation that is feasible for the coalition that he controls. Suppose that, when the agent approaches a player, the agent does not have to reveal the identity of the other players whom he has decided to approach, but he does have to describe the randomized plan by which he has made this decision and the expected utility allocation that he would implement for each such possible coalition.

Such a randomized blocking plan could be as follows: with probability .41, the agent will ask players 1 and 2 to join the blocking coalition and implement the allocation $(x_1,x_2) = (100,0)$; and with probability .59, the agent will ask all players to join the blocking coalition and implement the allocation $(x_1,x_2,x_3) = (0,79,21)$. Under this rule, players 1 and 2 are always approached by the blocking agent and get expected payoffs of $.41 \times 100 = 41$ and $.59 \times 79 = 46.61$ respectively. Furthermore, conditionally on being approached by the blocking agent, the expected payoff to player 3 is 21, which is strictly greater than 20. So each player should be willing to accept the agent's invitation to block (40,40,20). Thus, the core allocation of (40,40,20) can be blocked by a randomized blocking rule. Of course, this randomized blocking plan leaves player 3 worse off in expected value. Ex ante, player 3 expects to get $.59 \times 21 + .49 \times 0 = 12.39$ if the agent always succeeds. However, player 3 cannot prevent the blocking plan unless the agent approaches player 3, in which case accepting the plan and giving the agent power of attorney can only help player 3. Thus, unless player 3 can commit himself to reject the blocking coalition before he is invited to join it, the unique equilibrium in undominated strategies is for every player to join the agent's blocking coalition when invited to do so.

Similar tricks can be used to randomly block any other core allocation that gives player 3 a positive payoff. The idea is to randomize between two alternatives: either do not use player 3 and let player 1 keep everything; or use player 3 and transfer everything from player 1 while giving player 3 a slightly higher banker's commission than he would get in the allocation being blocked.

This random-blocking property actually characterizes the inner core quite generally. Let V be any NTU coalitional game on the finite set of players N. Define a randomized blocking plan to be any pair $(\eta, Y)$ such that $\eta$ is a probability distribution over the set of all possible coalitions, and Y is a function such that

$$Y(S) \in V(S)$$

for every coalition S. Here $\eta(S)$ represents the probability that the blocking agent will set out to form the coalition S, and Y(S) denotes the expected payoff allocation that he would implement if he succeeded in forming the coalition S.

Let x denote an allocation in $\mathbb{R}^N$ that is interpreted as the status quo, which is to be implemented unless the blocking agent is successful in forming his coalition. Then a randomized blocking plan $(\eta, Y)$ is viable against x iff

$$\Sigma_{S \supseteq \{i\}} \eta(S)(Y_i(S) - x_i) \geq 0, \quad \forall i \in N.$$

That is, $(\eta, Y)$ is viable against x iff for each player i, conditionally on player i's being invited to join the blocking coalition, his expected payoff in the blocking coalition would be at least $x_i$. We may say that x is strongly inhibitive iff there do not exist any viable randomized blocking plans against $x_i$. The following result follows from the work of Myerson [1988, 1991] and by Qin [1990].

Theorem 1. Let V be a finitely generated NTU coalitional game and let $\bar{x}$ be a feasible allocation in V(N). Then $\bar{x}$ is in the inner core of V if and only if there exists a sequence of allocations $(x^k)_{k=1}^{\infty}$ such that $\lim_{k\to\infty} x^k = \bar{x}$, and, for every k, $x^k$ is strongly inhibitive for V.

Proof. Let $\Omega = \{(S,y) \mid S \text{ is a coalition and } y \in W(S)\}$, where W(S) is the finite set of which V(S) is the comprehensive convex hull. A vector z is strongly inhibitive iff there is no probability distribution $\mu$ such that

$$\mu(S,y) \geq 0, \quad \forall(S,y) \in \Omega,$$

$$\Sigma_{S\supseteq\{i\}} \Sigma_{y\in W(S)} \mu(S,y)(z_i - y_i) \leq 0, \quad \forall i \in N,$$

$$\Sigma_{(S,y)\in\Omega} \mu(S,y) = 1.$$

By theorems of the alternative for linear systems, or the duality theorem for linear programming, the nonexistence of such a distribution $\mu$ holds iff there exists some vector $\lambda$ in $\mathbb{R}^N$ such that

$$\lambda_i \geq 0, \quad \forall i \in N,$$

$$\Sigma_{i\in S} \lambda_i(z_i - y_i) > 0, \quad \forall(S,y) \in \Omega.$$

The last of these inequalities when $S = \{i\}$ implies that each $\lambda_i$ must be strictly positive. Thus, z is strongly inhibitive iff there exists some positive vector $\lambda$ in $\mathbb{R}^N_{++}$ such that

$$\text{(21)} \qquad \Sigma_{i\in S} \lambda_i z_i > v^{\lambda}(S), \quad \forall S \subseteq N$$

where

$$\text{(22)} \qquad v^{\lambda}(S) = \max_{y\in V(S)} \Sigma_{i\in S} \lambda_i y_i, \quad \forall S \subseteq N.$$

Now suppose that $\bar{x}$ is the limit of strongly inhibitive points $x^k$. For each k, let $\lambda^k$ satisfy condition (21) for $z = x^k$, let $w_i^k = \lambda_i^k x_i^k$ for each player i, and let $u^k$ be the TU coalitional game such that

$$u^k(N) = \Sigma_{i\in N} \lambda_i^k x_i^k, \quad \text{and} \quad u^k(S) = v^{\lambda^k}(S), \quad \forall S \subset N.$$

Then $u^k \in F(V,\lambda^k)$, $w^k \in \text{Core}(u^k)$, and $\lim_{k\to\infty} w_i^k/\lambda_i^k = \bar{x}_i$ for each i; and so $\bar{x}$ is in the inner core.

Conversely, suppose that $\bar{x}$ is in the inner core. Then there exists a sequence $(\lambda^k,v^k,w^k)_{k=1}^{\infty}$ such that $\limsup_{k\to\infty} w_i^k/\lambda_i^k \le \bar{x}_i$ for every player i, and, for each k, $\lambda^k \in \mathbb{R}^N_{++}$, $v^k \in F(V,\lambda^k)$, and $w^k \in \text{Core}(v^k)$. By increasing the components of $w^k$ and increasing the $v^k(N)$ numbers where necessary, we can assume without loss of generality that the $(w_i^k/\lambda_i^k)_{k=1}^{\infty}$ sequence is decreasing in k and converges to $\bar{x}_i$, for each i. Furthermore, by increasing the components of each $w^k$ allocation slightly, we can assume that the core inequalities are all strictly satisfied for each $w^k$. Let $x_i^k = w_i^k/\lambda_i^k$ for each i and k. Then

$$\Sigma_{i\in S} \lambda_i^k x_i^k > v^{\lambda^k}(S), \quad \forall S \subseteq N.$$

So $x^k$ is strongly inhibitive (by condition (21)). Thus $\bar{x}$ is a limit of strongly inhibitive allocations if it is in the inner core. Q.E.D.

Harsanyi [1967-8] laid the foundations for the analysis of Bayesian games with incomplete information. Since his paper, perhaps because of rising interest in noncooperative game theory, the extension of cooperative game theory to games with incomplete information has been relatively neglected. (See Harsanyi and Selten [1972], Myerson [1984a, 1984b].)

For any game theorist who wants to develop solution concepts for cooperative games with incomplete information, a natural methodological question is how to extend the method of fictitious transfers to this case. Our technical result about the inner core for complete information games can be generalized and used to give an answer to this question.

Let us consider a simple model in which N is the finite set of players; $T_i$ is the finite set of possible types of player i; $C_S$ denotes the finite set of actions that are jointly feasible for each coalition $S \subseteq N$; $u_i(c_S, t_N)$ denotes the expected utility payoff that player i would get if he joined coalition S, which jointly implemented action $c_S$, and the profile of all players' types was $t_N = (t_i)_{i \in N}$; and $p_i(t_i)$ denotes positive probability that player i's true type is $t_i$. We assume here that players' types are independent random variables, but each player learns his own type before the play of the game begins. That is, a player's type is defined to be the state of his private information at the beginning of the game. We also assume that types are unverifiable, in the sense that there is nothing to prevent a player from lying about his type if he is given an incentive to do so.

For any possible coalition S, let $T_S = \times_{i \in S} T_i$ denote the set of all possible combinations of players' types in coalition S, and let

$$p_S(t_S) = \Pi_{j \in S}\, p_j(t_j), \quad \forall t_S \in T_S.$$

(Here $t_S = (t_j)_{j \in S}$.)

The status quo that an agent might try to block can be represented by an allocation rule $x: T_N \to \mathbb{R}^N$ such that, for every possible types profile $t_N$ and every player i, $x_i(t_N)$ is the expected payoff to player i in the status quo if $t_N$ is the profile of all players' types.

Now suppose that an outside agent or mediator attempts to block such a status quo x by forming a blocking coalition according to some random rule. For notational simplicity, we consider here only a very simple (but quite versatile) class of randomized blocking rules. Let us assume that the agent will randomly designate a coalition $S \subseteq N$, a profile of types $t_S$ for this coalition, and a feasible joint action $c_S$ for this coalition. Let $\mu(S, c_S, t_S)$ denote the probability of designating this coalition-types-action triple, under the random blocking plan $\mu$. Then the agent will separately and confidentially approach the members of his designated coalition S and will request that they give him power of attorney in all subsequent bargaining. If any player in the designated coalition rejects the agent's request, then the agent will not block the status quo. Each player who gives the agent power of attorney will then be asked to confidentially tell his type to the agent. If the profile of reported types matches the designated profile $t_S$, then the agent will implement the designated joint action $c_S$ for the coalition; if the profile of reported types does not match the designated types profile, then the agent will not block the status quo. We assume here that the status quo will be implemented if the agent does not block it. (When x results from the equilibrium of some game, there is a problem about whether the fact that a blocking coalition has not formed might convey information to the players that would change their behavior and thus change the status quo payoffs. One way to get around such considerations is to think of the $\mu(S,c_S,t_S)$ probabilities as infinitesimally small, so that very little information is conveyed by the nonformation of a blocking coalition.)

The agent has two problems in the design of his random blocking rule. He needs that the players in the blocking coalition should expect to do better by joining it than under the status quo, and he needs to give the players in the blocking coalition an incentive to share their type information honestly. So a random blocking plan $\mu$ is <u>viable</u> against x iff

(23) $\mu(S,c_S,t_S) \geq 0, \quad \forall S \subseteq N, \quad \forall c_S \in C_S, \quad \forall t_S \in T_S;$

(24) $\Sigma_{S\subseteq N} \Sigma_{c_S\in C_S} \Sigma_{t_S\in T_S} \mu(S,c_S,t_S) > 0;$

(25) $\Sigma_{S\supseteq\{i\}} \Sigma_{t_{N-i}\in T_{N-i}} \Sigma_{c_S\in C_S} \mu(S,c_S,t_S)\, p_{N-i}(t_{N-i})(u_i(c_S,t_N) - x_i(t_N)) \geq 0,$

$\forall i \in N, \quad \forall t_i \in T_i;$

$$(26) \quad \Sigma_{S\supseteq\{i\}} \Sigma_{t_{N-i}\in T_{N-i}} \Sigma_{c_S\in C_S} \mu(S,c_S,t_S)\, p_{N-i}(t_{N-i})(u_i(c_S,t_N) - x_i(t_N))$$

$$\geq \Sigma_{S\supseteq\{i\}} \Sigma_{t_{N-i}\in T_{N-i}} \Sigma_{c_S\in C_S} \mu(S,c_S,(t_{S-i},r_i))\, p_{N-i}(t_{N-i})(u_i(c_S,t_N) - x_i(t_N)),$$

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

We say that the status quo x is <u>strongly inhibitive</u> iff there do not exist any randomized blocking plans that are viable against $\mu$. Extending the preceding result, we may naturally define the inner core to be the set of status-quo allocation rules that can be implemented by incentive-compatible mechanisms and are equal or Pareto-superior to limits of strongly inhibitive allocation rules. (Of course, such a core may be empty for many examples. However, Myerson [1988] showed a class of replicated games in which a related core concept is always nonempty.) Now we may state the generalization of Theorem 1 which shows us how to characterize such strongly inhibitive allocations in terms of fictitious transfers of <u>virtual utility</u>, as defined in Myerson [1984a, 1984b].

<u>Theorem 2.</u> An allocation rule $x:T_N \to \mathbb{R}^N$ is strongly inhibitive iff there exists nonnegative numbers $\lambda_i(t_i)$ and $\alpha_i(r_i|t_i)$, for every player i in N and all possible types $t_i$ and $r_i$ in $T_i$, such that

$$\Sigma_{i\in S} \Sigma_{t_{N\backslash S}\in T_{N\backslash S}} p_{N\backslash S}(t_{N\backslash S})\, \hat{x}_i(t_N,\lambda,\alpha)$$

$$> \Sigma_{i\in S} \Sigma_{t_{N\backslash S}\in T_{N\backslash S}} p_{N\backslash S}(t_{N\backslash S})\, \hat{u}_i(c_S,t_N,\lambda,\alpha), \quad \forall S \subseteq N,\ \forall c_S \in C_S,\ \forall t_S \in T_S;$$

where $\hat{u}_i(c_S,t_N,\lambda,\alpha) =$

$$[(\lambda_i(t_i) + \Sigma_{r_i\in T_i} \alpha_i(r_i|t_i))u_i(c_S,t_N) - \Sigma_{r_i\in T_i} \alpha_i(t_i|r_i)u_i(c_S,t_{N-i},r_i)]/p_i(t_i)$$

and $\hat{x}_i(t_N,\lambda,\alpha) =$

$$[(\lambda_i(t_i) + \Sigma_{r_i\in T_i} \alpha_i(r_i|t_i))x_i(t_N) - \Sigma_{r_i\in T_i} \alpha_i(t_i|r_i)x_i(t_{N-i},r_i)]/p_i(t_i).$$

(Notice that, if player i has only one possible type $t_i$ then $\hat{x}_i(t_N,\lambda,\alpha) = \lambda_i(t_i)x_i(t_N)$ and $\hat{u}_i(c_S,t_N,\lambda,\alpha) = \lambda_i(t_i)u_i(c_S,t_N)$, and so these virtual-utility formulas are a generalization of $\lambda$-weighted utility for games with incomplete information.)

<u>Proof</u>. Consider the linear programming problem of maximizing the sum in line (24) over all nonnegative vectors $\mu$, subject to the constraints (25) and (26). The optimal value of this homogeneous linear program is either 0 or $+\infty$. The allocation rule x is strongly inhibitive iff the optimal value is zero. So, by the duality theory of linear programming, x is strongly inhibitive iff the dual problem has a feasible solution. For each i, $t_i$, and $r_i$, let $\lambda_i(t_i)$ and $\alpha_i(r_i|t_i)$ denote the dual variables for constraints (25) and (26) respectively. Then the constraints of the dual problem can be written

$$\Sigma_{i\in S}\ \Sigma_{t_{N\backslash S}\in T_{N\backslash S}}\ p_{N-i}(t_{N-i})[(u_i(c_S,t_N) - x_i(t_N))(\lambda_i(t_i) + \Sigma_{r_i\in T_i}\ \alpha_i(r_i|t_i))$$
$$- \Sigma_{r_i\in T_i}\ (u_i(c_S,(t_{N-i},r_i)) - x_i(t_{N-i},r_i))\alpha_i(t_i|r_i)] \geq 1,$$
$$\forall S \subseteq N,\quad \forall t_S \in T_S,\quad \forall c_S \in C_S.$$

It is straightforward to show that these dual constraints can be satisfied iff the constraints in the theorem can be satisfied by some nonnegative vectors $\lambda$ and $\alpha$. Q.E.D.

## REFERENCES


Aumann, R. J. [1967]. "A Survey of Cooperative Games Without Side Payments." In M. Shubik, ed., Essays in Mathematical Economics, pp. 3-27. Princeton: Princeton University Press.

Aumann, R. J. [1985]. "An Axiomatization of the Non-Transferable Utility Value." Econometrica 53:599-612.

Aumann, R. J., and B. Peleg [1960]. "Von Neumann-Morgenstern Solutions to Cooperative Games Without Sidepayments." Bulletin of the American Mathematical Society 66:173-179.

Harsanyi, J. C. [1959]. "A Bargaining Model for the Cooperative n-Person Game." In R. D. Luce and A. W. Tucker, eds., Contributions to the Theory of Games IV, Princeton: Princeton University Press, pages 325-355.

Harsanyi, J. C. [1963]. "A Simplified Bargaining Model for the n-Person Cooperative Game." International Economic Review 4:194-220.

Harsanyi, J. C. [1967-8]. "Games with Incomplete Information Played by 'Bayesian' Players." Management Science 14:159-182, 320-334, 486-502.

Harsanyi, J. C., and R. Selten. [1972]. "A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information." Management Science 18:80-106.

Hart, S. [1985]. "An Axiomatization of Harsanyi's Nontransferable Utility Solution." Econometrica 53:1295-1313.

Isbell, J. R. [1960]. "A Modification of Harsanyi's Bargaining Model." Bulletin of the American Mathematical Society 66:70-73.

Jentzsch, G. [1959]. "Some Thoughts on the Theory of Cooperative Games." In R. D. Luce and A. W. Tucker, eds., Contributions to the Theory of Games IV, Princeton: Princeton University Press, pages 407-442.

Kalai, E., and M. Smorodinsky [1975]. "Other Solutions to Nash's Bargaining Problem." Econometrica 45:513-518.

Luce, R. D., and H. Raiffa [1957]. Games and Decisions. New York: Wiley.

Myerson, R. B. [1980]. "Conference Structures and Fair Allocation Rules." International Journal of Game Theory 9:169-182.

Myerson, R. B. [1984a]. "Two-Person Bargaining Problems with Incomplete Information." Econometrica 52:461-487.

Myerson, R. B. [1984b]. "Cooperative Games with Incomplete Information." International Journal of Game Theory 13:69-96.

Myerson, R. B. [1988]. "Sustainable Matching Plans with Adverse Selection." Northwestern University discussion paper. To appear in Games and Economic Behavior.

Nash, J. F. [1950]. "The Bargaining Problem." Econometrica 18:155-162.

Nash, J. F. [1953]. "Two-Person Cooperative Games." Econometrica 21:128-140.

Neumann, J. von. 1928. "Zur Theorie der Gesellschaftsspiele." Mathematische Annalen 100:295-320. English translation in R. D. Luce and A. W. Tucker, eds., Contributions to the Theory of Games IV, Princeton: Princeton University Press, pages 13-42.

Neumann, J. von, and O. Morgenstern [1944]. Theory of Games and Economic Behavior. Princeton: Princeton University Press. 2nd ed., 1947.

Owen, G. [1972]. "A Value for Non-Transferable Utility Games." International Journal of Game Theory 1:95-109.

Qin, C.-Z. [1990]. "The Inner Core and the Strongly Inhibitive Set." CORE discussion paper, Louvain-la-Neuve, Belgium.

Roth, A. E. [1980]. "Values for Games Without Side-Payments: Some Difficulties with Current Concepts." Econometrica 48:457-465.

Shapley, L. S. [1953]. "A Value for n-Person Games." In H. Kuhn and A. W. Tucker, eds., Contributions to the Theory of Games 2, pp. 307-317. Princeton: Princeton University Press.

Shapley, L. S. [1964]. ""Values of Large Market Games: the Status of the Problem," RAND Corporation paper RM-3957PR.

Shapley, L. S. [1969]. "Utility Comparison and the Theory of Games." In La Decision, Editions du CNRS, Paris, pp 251-263. Reprinted in A. E. Roth, ed. The Shapley Value, pp. 307-319, Cambridge: Cambridge University Press.

Shapley, L. S., and M. Shubik [1953]. "Solutions of n-Person Games with Ordinal Utilities" (abstract). Econometrica 21:348.

Shubik, M. [1982]. Game Theory in the Social Sciences: Concepts and Solutions. Cambridge, MA: M.I.T. Press.