# COOPERATIVE GAMES

MIHAI MANEA

Department of Economics, MIT

## 1. Definitions

A *coalitional* (or *cooperative*) *game* is a model of interacting decision-makers that focuses on the behavior of groups of players.

$N$ denotes the set of *players*. A *coalition* is a group of players $S \subset N$. We refer to $N$ as the *grand coalition*. Every coalition $S$ has a set of available *actions* $A_S$.

An *outcome* consists of a partition of $N$ (coalition structure) and one action associated with each coalition in the partition,

$$(S_k, a_k)_{k=1,\ldots,\bar{k}} \text{ with } S_j \cap S_k = \emptyset, \forall j \neq k, \cup_k S_k = N, a_k \in A_{S_k}.$$

Each agent $i$ has *preferences* over the set of all outcomes represented by a utility function $u_i$. We restrict attention to settings *without externalities*. Agent $i$ cares only about the action of the coalition he belongs to, i.e., $\exists U_i : \cup_{S \ni i} A_S \to \mathbb{R}$ s.t.

$$u_i((S_k, a_k)_{k=1,\ldots,\bar{k}}) = U_i(a_j) \text{ if } i \in S_j.$$

**Example 1** (Three-player majority game)**.** Three agents have access to a unit of output. Any majority—coalition of two or three agents—may control the allocation of the output. The output may be shared among the members of the winning coalition any way they wish. No agent can produce any output by himself. Each agent only cares about the amount of output he receives (prefers more to less). Actions for each coalition? Feasible outcomes?

**Example 2** (Firm and workers)**.** Consider a firm and $n$ potential workers. The firm generates profit $f(k)$ from hiring $k$ workers, where $f$ is some exogenously given function. Any coalition consisting only of workers produces profit 0. Each agent only cares about his own share of the profit. Actions for each coalition? Feasible outcomes?

---

*Date*: January 19, 2017.

I thank Gabriel Carroll for proofreading.

**Example 3** (Marriage market)**.** A group of men and a group of women can be matched in pairs. A matching of a coalition of men and women is a partition of the coalition into man-woman pairs and single individuals. Each person cares only about her partner (and how she/he compares to being single). Actions for each coalition? Feasible outcomes?

**Definition 1.** A game is *cohesive* if for every outcome $(S_k, a_k)_{k=1,\ldots,\bar{k}}$ there exists an outcome generated by the grand coalition which is at least as desirable as $(S_k, a_k)_{k=1,\ldots,\bar{k}}$ for every player.

The solution concepts we consider assume that the grand coalition forms (rather than outcomes involving a non-trivial coalition structure) and have attractive interpretations only for cohesive games.

**Definition 2.** A game with *transferable payoffs* associates to any coalition $S$ a real number $v(S)$, which is interpreted as the *worth* of the coalition $S$. We assume $v(\emptyset) = 0$. The set of actions available to coalition $S$ consists of all possible divisions $(x_i)_{i\in S}$ of $v(S)$ among the members of $S$, $\sum_{i\in S} x_i = v(S)$.

When is a game with transferable payoffs cohesive? Which of the games above have transferable payoffs? What are the corresponding worth functions?

## 2. The Core

Which action may we expect the grand coalition to choose? We seek actions that withstand the pressures imposed by the opportunities of each coalition. We define an action of the grand coalition to be "stable" if no coalition can break away and choose an action that all its members prefer. Formally, a coalition $S$ *blocks* an action $a_N$ of the grad coalition if there is an action $a_S \in A_S$ that all members of $S$ strictly prefer to $a_N$. The set of all actions that cannot be blocked forms the core.

**Definition 3.** The *core* of a coalitional game is the set of actions $a_N$ of the grand coalition $N$ that are not blocked by any coalition.

If a coalition $S$ has an action that all its members prefer to an action $a_N$ of the grand coalition, we say that $S$ *blocks* $a_N$.

For a game with transferable payoffs with payoff function $v$, a coalition $S$ can block the allocation $(x_i)_{i\in N}$ iff $x_S < v(S)$, where $x_S = \sum_{i\in S} x_i$. Hence the allocation $x$ is in the core of the game iff $x_S \geq v(S), \forall S \subset N$.

**Example 4** (Two-player split the dollar with outside options)**.** $v(\{1\}) = p, v(\{2\}) = q, v(\{1,2\}) = 1$. The core is given by the set of allocations

$$\{(x_1, x_2)|x_1 + x_2 = 1, x_1 \geq p, x_2 \geq q\}.$$

What happens for $p = q = 0$? What if $p + q > 1$?

Argue that the core of the three-player majority game is empty.

When is the core non-empty? A vector $(\lambda_S)_{S\in 2^N}$ of non-negative numbers is a *balanced collection* of weights if

$$\sum_{\{S\subset N|i\in S\}} \lambda_S = 1, \forall i \in N.$$

A payoff function $v$ is *balanced* if

$$\sum_{S\subset N} \lambda_S v(S) \leq v(N) \text{ for every balanced collection of weights } \lambda.$$

Interpretation: each player has a unit of time, which he needs to distribute among all his coalitions. For coalition $S$ to be active for $\lambda_S$ time and generate payoff $\lambda_S v(S)$, all its members need to be active in $S$ for this fraction of time $\lambda_S$. A game is *balanced* if there is no allocation of time across coalitions that yields a total value greater than that of the grand coalition.

**Theorem 1** (Bondareva 1963; Shapley 1967)**.** *A coalitional game with transferable payoffs has a non-empty core iff it is balanced.*

*Proof.* Consider the linear program

$$\min \sum_{i\in N} x_i \text{ s.t. } \sum_{i\in S} x_i \geq v(S), \forall S \subset N.$$

The core is non-empty iff the minimized sum is not greater than $v(N)$.[1] The dual of the latter linear program is given by

$$\max \sum_{S\in 2^N} \lambda_S v(S) \text{ s.t. } \lambda_S \geq 0, \forall S \subset N \;\&\; \sum_{S\ni i} \lambda_S \leq 1, \forall i \in N.$$

[1]Actually, the optimal value needs to be exactly $v(N)$ in this case.

By definition, the game is balanced iff the optimal value for the dual does not exceed $v(N)$.[2] Note that the primal linear program has an optimal solution. Then the conclusion follows from the duality theorem of linear programming, which implies that both problems have solutions and the optimal values of the two objective functions are identical.

[Easy proof of "only if" part. Let $x$ be a core allocation and suppose that $(\lambda_S)_{S\in 2^N}$ is a balanced collection of weights. Then

$$\sum_{\{S|S\subset N\}} \lambda_S v(S) \leq \sum_{\{S|S\subset N\}} \lambda_S x_S = \sum_{i\in N} x_i \sum_{S\ni i} \lambda_S = \sum_{i\in N} x_i = v(N).$$

Hence $v$ is balanced.] □

We next introduce a simpler condition, convexity, which guarantees a non-empty core.

**Definition 4.** A game with transferable payoffs $v$ is *convex*[3] if for any two coalitions $S$ and $T$,

$$v(S\cup T) + v(S\cap T) \geq v(S) + v(T).$$

The game $v$ is *superadditive* if for any disjoint coalitions $S$ and $T$,

$$v(S\cup T) \geq v(S) + v(T).$$

The game $v$ is *monotone* if for any coalitions $S \subset T$,

$$v(S) \leq v(T).$$

Note that convexity implies superadditivity. If $v$ is nonnegative, then superadditivity implies monotonicity. Also, you can check that superadditivity implies cohesiveness. Convexity owes its name to the following implication

$$S \subset T \,\&\, i \notin T \implies v(T\cup\{i\}) - v(T) \geq v(S\cup\{i\}) - v(S), \tag{2.1}$$

which says that the marginal contribution of an individual $i$ to a coalition is (weakly) increasing as the coalition gets larger. To see this, consider the coalitions $S\cup\{i\}$ and $T$. Indeed, by convexity,

$$v((S\cup\{i\})\cup T) + v((S\cup\{i\})\cap T) \geq v(S\cup\{i\}) + v(T).$$

[2] See footnote 1.
[3] Convexity is sometimes referred to as as super-modularity.

which can be rewritten as

$$v(T \cup \{i\}) - v(T) \geq v(S \cup \{i\}) - v(S).$$

**Proposition 1.** *Any convex game with transferable payoffs has a non-empty core.*

*Proof.* We show that the allocation $x$, with $x_i = v(\{1, \ldots, i\}) - v(\{1, \ldots, i-1\})$, belongs to the core. For all $i_1 < i_2 < \cdots < i_k$,

$$\begin{aligned}
\sum_{j=1}^{k} x_{i_j} &= \sum_{j=1}^{k} v(\{1, \ldots, i_j - 1, i_j\}) - v(\{1, \ldots, i_j - 1\}) \\
&\geq \sum_{j=1}^{k} v(\{i_1, \ldots, i_{j-1}, i_j\}) - v(\{i_1, \ldots, i_{j-1}\}) \\
&= v(\{i_1, i_2, \ldots, i_k\}),
\end{aligned}$$

where the inequality follows from $\{i_1, \ldots, i_{j-1}\} \subset \{1, \ldots, i_j - 1\}$ and 2.1. □

It is easy to construct examples which illustrate that convexity is not a necessary condition for the non-emptiness of the core (balancedness is).

## 3. Core Implementation (Perry and Reny 1994)

We restrict attention to games with transferable payoffs. An allocation which is not in the core is regarded as unstable. For every such allocation, some players can form a coalition and obtain an allocation that each strictly prefers. Almost sounds like a non-cooperative game! But what's the dynamics/timing of forming and breaking coalitions? What sort of agreements are feasible/binding and what offers and counteroffers are allowed following a history of outstanding proposals and binding agreements? Perry and Reny (1994) provide a dynamic game in continuous time that implements (all and only) core allocations in equilibrium. The game is designed to match the ideas behind the definition of the core.

The game starts at $t = 0$, at which time one player can choose to make a proposal or be quiet. At any $t > 0$, a player can choose to make a proposal, accept the currently active proposal, be quiet, or leave. A proposal, $(x, S)$, suggests a coalition $S$ and an allocation of $v(S)$ or less among the players in $S$ (such proposals can be made by all players, including ones outside $S$). A proposal $(x, S)$ remains active until either it is accepted by all players in $S$, in which case it becomes binding, or another proposal is made, in which case the new proposal becomes active and the old one disappears.

At any time there can thus be only one active proposal, but there can also be a number of binding proposals that have been previously accepted. Whenever a binding proposal $(x, S)$ exists, any new proposal's coalition has to be either disjoint from $S$ or else fully include it. If the latter happens, then if the new proposal is accepted, the original proposal $(x, S)$ disappears. Thus, at all times, the coalitions associated with currently binding proposals are mutually disjoint.

Once a player accepts a proposal, he must remain quiet until it is binding or a new proposal displaces it. After a proposal becomes binding, each player involved can leave and consume or wait for better proposals. If one involved player decides to leave, all other players in the coalition have to leave too.

Each player's payoff is the amount allocated to him by the proposal that applies to him when he leaves. There is *no discounting.*

A technical assumption on strategies is necessary. For every $t$ and every history up to $t$, players are not allowed to choose to make a "non-quiet" decision either "just before" or "right after" $t$. Technically, for any history up to time $t$, there is an $\varepsilon > 0$ such that each player is quiet in the intervals $(t - \varepsilon, t)$ and $(t, t + \varepsilon)$. First, this ensures that given any history up to time $t$, the strategies induce a unique continuation path (and well-defined payoffs). Second, if we want to obtain only core outcomes in equilibrium, it is important that no player want to take a "non-quiet" action as near as possible, yet before or after, any time $t$. If a player had that option then there would not be any way for another player to convince him to stick around with an appropriately timed blocking proposal.

The equilibrium concept is stationary subgame perfect equilibrium (SSPE), that is, an SPE in stationary strategies. A strategy is stationary if the player's action depends only on the set of players who have not yet left, the currently active proposal with the set of players who accepted it, and the currently binding proposals.

**Theorem 2.** *Every SSPE outcome belongs to the core.*

*Proof.* Suppose that $x$ is an SSPE outcome that does not belong to the core. There exists a proposal $(y, S)$ such that $y_i > x_i, \forall i \in S$. WLOG, assume $S = \{1, \ldots, k\}$.

Consider a time $t$ history where nothing has happened except that $(y, S)$ was proposed and $1, 2, \ldots, k-1$ accepted. After this history, player $k$ accepts before anything else happens. Suppose not, let $(z, T)$ be the next proposal. It must be that in equilibrium $k$ gets at least $y_k > x_k$ in the continuation game. By stationarity, it must be that in any subgame when only $(z, T)$ is on the table (and there has been no acceptance), $k$ obtains more than $x_k$. Then $k$ can deviate from the putative equilibrium by proposing $(z, T)$ close to time 0, thereby obtaining more than $x_k$, a contradiction with the equilibrium requirements.

We prove by induction on $l$ that after any history where nothing happened except that $1, 2, \ldots, k-l$ accepted $(y, S)$, player $k-l+1$ accepts. We showed that the base case is true. If $k-l+1$ accepts, then by the induction hypothesis everyone accepts and $k-l+1$ obtains $y_{k-l+1}$. If he does not accept, it must be that he can do even better and can obtain strictly more than $y_{k-l+1}$ making a proposal close to time 0.

The induction hypothesis for $l = k-1$ implies that if 1 offers $(y, S)$ close to time 0 and accepts it before any other action is taken, then he obtains a payoff $y_1 > x_1$. This contradicts $x$ being an SSPE outcome. $\square$

The result above also holds for subgames in which a subset of the players left and consumed. An outcome for the remaining players must be in the core of the cooperative game they induce. One consequence of the result is that a necessary condition for an SSPE to *exist* is that the game be totally balanced, i.e., its restriction to any coalition is balanced (or any subgame has a non-empty core).

**Theorem 3.** *If the game is totally balanced, then any element of its core can be supported as an SSPE outcome.*

The construction of an SSPE implementing any core point is involved, particularly off the equilibrium path.

## 4. The Core and Competitive Equilibria

Consider an *exchange economy* with a set of consumers $N$ and a set of goods $G$. A *consumption bundle* is an element $x \in \mathbb{R}_+^G$. Consumer $i$ enjoys utility $u_i(x_i)$ from a bundle $x_i$. Each consumer starts with an *endowment* $\omega_i \in \mathbb{R}_+^G$.

An *allocation* $\{x_i\}_{i\in N}$, with $x_i \in \mathbb{R}^G_+$, is *feasible* if $\sum_{i\in N} x_i = \sum_{i\in N} \omega_i$, where $\omega_i$ is consumer $i$'s initial endowment of goods.

Suppose that consumers interact in a market. A price is determined for every good and the consumers take these prices as given. The price-taking assumption is reasonable for markets with many participants and a high degree of competition among them.

**Definition 5** (Competitive Equilibrium)**.** A *competitive equilibrium* is a price vector $(p^*_g)_{g\in G}$ and a feasible allocation $x^* = (x^*_i)_{i\in N}$ such that

$$p^* \cdot x_i \leq p^* \cdot \omega_i \implies u_i(x^*_i) \geq u_i(x_i),$$

i.e. $x^*_i$ maximizes consumer $i$'s utility among all the consumption bundles that he can afford given the prices and his initial endowment.

4.1. **Aside on Existence.** Does a competitive equilibrium exist? Yes it does, if one imposes some regularity conditions on the utility functions. The standard proof involves Kakutani's fixed point theorem, which you are already familiar with. However, equilibrium existence is not our focus here.

4.2. **Competitive Equilibria and Cooperative Games.** We can view the market as a cooperative game. The actions available for any coalition of players $S \subseteq N$ is the set of all distributions of their total endowment $\sum_{i\in S} \omega_i$ among themselves,

$$A_S = \left\{ (x_i)_{i\in S} \mid \sum_{i\in S} x_i = \sum_{i\in S} \omega_i \right\}.$$

Of course, the payoff of consumer $j \in S$ corresponding for the action $(x_i)_{i\in S}$ is given by $u_j(x_j)$. Note that for most utility functions the resulting game does not have transferable utility.

**Example 5.** Let $G = \{g\}$, $N = \{1, 2\}$, $\omega_1 = \omega_2 = 1$, and $u_i(x_i) = \sqrt{x_i}$. Then the set of feasible payoffs (utilities) for the coalition $\{1, 2\}$ is given by

$$\{(u_1, u_2) \mid u_1^2 + u_2^2 = 2; u_1, u_2 \geq 0\}.$$

**Example 6.** Consider a 2-consumer, 2-good economy. We can use the Edgeworth box to illustrate allocations, competitive equilibria, and the core. See pp. 515-525 in Mas-Colell, Whinston, and Green (posted).

We know how to define the core for any cooperative game, including the one derived here. When can a coalition $S$ block a feasible allocation $x^*$? It must be that the members of $S$ can redistribute their total endowment $\sum_{i \in S} \omega_i$ amongst themselves so as to make each of them better off. Formally, there exists an allocation $(x_i)_{i \in S}$ that is feasible for $S$ ($\sum_{i \in S} x_i = \sum_{i \in S} \omega_i$), which every agent prefers to $x^*$ ($u_i(x_i) > u_i(x_i^*), \forall i \in S$).

**Theorem 4.** *Any competitive equilibrium is in the core.*

*Proof.* Let $x^*$ be a competitive equilibrium allocation corresponding to a price vector $p$. Suppose that a coalition $S$ can block $x^*$. Then there exists $(x_i)_{i \in S}$ such that $\sum_{i \in S} x_i = \sum_{i \in S} \omega_i$ and

$$u_i(x_i) > u_i(x_i^*) \ \forall \ i \in S$$

By definition of equilibrium the latter inequality implies that no $i \in S$ can afford $x_i$, so

$$p \cdot x_i > p \cdot \omega_i \ \forall \ i \in S$$

Adding up these conditions we obtain

$$p \cdot \sum_{i \in S} x_i > p \cdot \sum_{i \in S} \omega_i$$

which contradicts $\sum_{i \in S} x_i = \sum_{i \in S} \omega_i$. □

The converse of this theorem is not true. That is, not every point in the core is a competitive equilibrium allocation. For instance, one can see an example in the Edgeworth box in which the competitive equilibrium is unique but the core has a continuum of elements.

However, it is true that as we increase the market by replicating every consumer a large number of times the non-equilibrium allocations gradually drop from the core, until in the limit only equilibria survive. We will not prove this result.

4.3. **Aside on Equilibrium Tatonnement.** Even in environments where an equilibrium exists, the market may start at a price level which is not an equilibrium. What sort of price dynamics should we expect then?

There are some intuitive dynamic principles—if the demand for a good is larger than the supply, then one may expect the price of that good to increase.

There are, however, many possible disequilibium dynamics that one can consider, and it is hard to argue that one is better than all others.

I will provide an example just to illustrate the ideas. Let $z(p)$ be the excess demand function at $p$. That is, we assume that the utility function $u_i$ is such that for any non-zero price vector $p$ and initial endowment $\omega_i$, player $i$ has a unique optimal demand $x_i(p)$, where

$$x_i(p) = \arg \max_{x_i \cdot p \leq \omega_i \cdot p} u_i(x_i)$$

The excess demand function for good $g$ is then given by

$$z_g(p) = \sum_{i \in N} (x_{ig}(p) - \omega_{ig})$$

We say that good $g$ is in excess demand (supply) if $z_g(p) > 0$ ($< 0$).

Note that $p^*$ is an equilibrium price vector if and only if $z(p^*) = 0$. Suppose we start from a disequilibrium price vector $p$. One price dynamic that adjusts prices upwards for goods in excess demand and downwards for goods in excess supply is given by the following process

$$\frac{dp_g}{dt} = c_g z_g(p)$$

where $\frac{dp_g}{dt}$ is the rate of change of the price of good $g$, and $c_g > 0$ is a constant that determines the speed of convergence. This process is guaranteed to converge to an equilibrium if several restrictions are imposed on the utility functions. The paper we discuss next applies the idea of tatonnement in the context of the core.

## 5. Core Tatonnement

See slides.

## 6. The Nash Bargaining Solution

We revisit the bargaining problem.[4] The *non-cooperative approach* involves explicitly modeling the bargaining process as an extensive form game. The Rubinstein (1982) alternating offer bargaining model discussed earlier constitutes a prominent example. We next adopt the *axiomatic approach*, which abstracts away from the details of the bargaining process. The latter approach attempts to determine directly what "reasonable" or "natural" properties the outcomes should satisfy. Nash (1950) provided a seminal contribution in this direction.

The immediate question is: What are "reasonable" axioms? Consider a situation where two players must split \$1. If no agreement is reached, then the players receive nothing. If the preferences over monetary prizes are identical, then we might expect that each player obtains 50 cents. This example reflects two desirable properties of an allocation: efficiency and symmetry of the outcome for identical preferences.

A *bargaining problem* is a pair $(U, d)$ where $U \subset \mathbb{R}^2$ and $d \in U$. We assume that U is convex and compact and that there exists some $u \in U$ such that $u > d$. We denote the set of all possible bargaining problems by $\mathcal{B}$. A *bargaining solution* is a function $f : \mathcal{B} \to \mathbb{R}^2$ with $f(U, d) \in U$.

**Definition 6.** The *Nash (1950) bargaining solution* $f^N$ is defined by

$$\{f^N(U,d)\} = \arg\max_{u \in U, u \geq d} (u_1 - d_1)(u_2 - d_2). \tag{6.1}$$

Given the assumptions on $(U, d)$, the solution to the optimization problem above *exists* and is *unique*.

We will show that the Nash bargaining solution is the unique solution that satisfies the following axioms.

**Axiom 1** (Pareto Efficiency)**.** A bargaining solution $f$ is *Pareto efficient* if for any bargaining problem $(U, d)$, there does not exist $(u_1, u_2) \in U$ such that $u_1 \geq f_1(U, d)$ and $u_2 \geq f_2(U, d)$, with at least one strict inequality.

The motivation for this axiom is straightforward—an inefficient outcome is unlikely because it leaves space for renegotiation.

[4]This section builds on lecture notes by Asu Ozdaglar.

**Axiom 2** (Symmetry)**.** A bargaining solution $f$ is *symmetric* if for any symmetric bargaining problem $(U, d)$ ($(u_1, u_2) \in U$ if and only if $(u_2, u_1) \in U$ and $d_1 = d_2$), we have $f_1(U, d) = f_2(U, d)$.

The intuition for this axioms is that if players are indistinguishable, the agreement should not discriminate between them.

**Axiom 3** (Invariance to Linear Transformations)**.** A bargaining solution $f$ is *invariant* if for any bargaining problem $(U, d)$ and all $\alpha_i \in (0, \infty), \beta_i \in \mathbb{R}$ $(i = 1, 2)$, if we consider the bargaining problem $(U', d')$ with

$$\begin{aligned} U' &= \{(\alpha_1 u_1 + \beta_1, \alpha_2 u_2 + \beta_2) \mid (u_1, u_2) \in U\} \\ d' &= (\alpha_1 d_1 + \beta_1, \alpha_2 d_2 + \beta_2) \end{aligned}$$

then $f_i(U', d') = \alpha_i f_i(U, d) + \beta_i$ for $i = 1, 2$.

The motivation for this axiom is that the feasible payoffs are derived from an underlying outcome space where agents have expected utility over lotteries. Linear transformations of the utility functions lead to equivalent preferences over lotteries, and the bargaining outcome should not be sensitive to the preference representation. The axiom effectively amounts to a normalization of the bargaining problem.

**Axiom 4** (Independence of Irrelevant Alternatives)**.** A bargaining solution $f$ is *independent* if for any two bargaining problems $(U, d)$ and $(U', d)$ with $U' \subseteq U$ and $f(U, d) \in U'$, we have $f(U', d) = f(U, d)$.

**Theorem 5.** *$f^N$ is the unique bargaining solution that satisfies the four axioms.*

*Proof.* We first check that the Nash bargaining solution satisfies the axioms. We then show that if a bargaining solution satisfies the axioms, then it must be identical to $f^N$.

(1) Pareto efficiency: This follows immediately from the fact that the objective function in 6.1 is increasing in $u_1$ and $u_2$.
(2) Symmetry: Assume that $(U, d)$ is a symmetric bargaining problem. Then $(f_2^N(U, d), f_1^N(U, d)) \in U$ also solves the optimization problem 6.1. By the uniqueness of the optimal solution, we must have $f_1^N(U, d) = f_2^N(U, d)$.

(3) Independence of irrelevant alternatives: Suppose that $f^N(U,d) \in U' \subseteq U$. The value of the objective function in 6.1 for $(U',d)$ cannot exceed that for $(U,d)$. Since $f^N(U,d) \in U'$, the two values must be equal, and by the uniqueness of the optimal solution for 6.1, we have $f^N(U,d) = f^N(U',d)$.

(4) Invariance to linear transformations: Suppose that $(U,d)$ and $(U',d')$ are related as in the statement of the axiom. By definition, $f^N(U',d')$ is an optimal solution of the problem

$$\max_{\{(u_1',u_2')|u_1'=\alpha_1 u_1+\beta_1, u_2'=\alpha_2 u_2+\beta_2, (u_1,u_2)\in U\}} (u_1' - \alpha_1 d_1 - \beta_1)(u_2' - \alpha_2 d_2 - \beta_2)$$

It follows immediately that $f_i^N(U',d') = \alpha_i f_i^N(U,d) + \beta_i$ for $i = 1, 2$.

We next show that $f^N$ is the only bargaining solution that satisfies the axioms. We show that for any $f$ with this property, $f(U,d) = f^N(U,d)$ for all $(U,d)$.

Fix a bargaining problem $(U,d)$ and let $z = f^N(U,d)$. There exists $\alpha_i > 0, \beta_i$ such that the transformation $u_i \to \alpha_i u_i + \beta_i$ takes $d_i$ to 0 and $z_i$ to 1/2. Define

$$U' = \{(\alpha_1 u_1 + \beta_1, \alpha_2 u_2 + \beta_2)|(u_1,u_2) \in U\}.$$

Since both $f$ and $f^N$ satisfy the invariance to linear transformations axiom, we have $f(U,d) = f^N(U,d)$ if and only if $f(U',0) = f^N(U',0) = (1/2,1/2)$. Hence, to establish the desired claim, it is sufficient to prove that $f(U',0) = (1/2,1/2)$.

Note that the line $\{(u_1,u_2)|u_1+u_2 = 1\}$ is tangent to the hyperbola $\{(u_1,u_2)|u_1u_2 = 1/4\}$ at the point $(1/2,1/2)$. Given that $f^N(U',0) = (1/2,1/2)$, we can argue that $u_1+u_2 \le 1$ for all $u \in U'$. Assume there is a $u \in U'$ with $u_1 + u_2 > 1$. Let $t = (1-\lambda)(1/2,1/2) + \lambda(u_1,u_2)$ for some $\lambda \in (0,1)$. Since $U'$ is convex, we have $t \in U'$. We can choose $\lambda$ sufficiently small so that $t_1t_2 > 1/4$, a contradiction with the optimality of $f^N(U',0) = (1/2,1/2)$ in 6.1 for the bargaining problem $(U',0)$.

Since $U'$ is bounded, we can find a rectangle $U''$ with one side along the line $u_1 + u_2 = 1$, symmetric with respect to the line $u_1 = u_2$, such that $U' \subseteq U''$ and $(1/2,1/2)$ is on the boundary of $U''$.

Since $f$ is efficient and symmetric, it must be that $f(U'',0) = (1/2,1/2)$. We assumed that $f$ also satisfies the independence of irrelevant alternatives axiom. Then $f(U'',0) = (1/2,1/2) \in U' \subseteq U''$ leads to $f(U',0) = (1/2,1/2)$, which completes the proof. $\square$

## 7. Literature Discussion

- Nash (1953) demand game; Abreu and Pearce's (2015) endogenous threats in a dynamic setting
- Nash bargaining solution used as reduced form for equilibrium analysis in the macro and search literature (Diamond-Mortensen-Pissarides, Shimer-Smith)
- the Kalai-Smorodinsky (1975) solution
- axiomatic approach in matching theory

## 8. The Shapley Value

The core of a coalitional game may be empty or quite large, which compromises its role as a predictive theory in certain situations. Ideally, we would like to develop a theory selecting a unique outcome for every cooperative game. A *value* for cooperative games is a function from the space of games to outcomes. Note that we allow the player set to vary and we implicitly expect some "consistency" in the outcomes of "related" games. Here we restrict attention to games with transferable utility and assume that the set of feasible outcomes for a game $(N, v)$ consists of all divisions of $v(N)$ among the players in $N$.

Shapley (1953) proposed a solution that has many properties that are economically desirable and mathematically elegant.

**Definition 7** (Shapley value)**.** The Shapley value of a game with worth function $v$ is given by

$$\varphi_i(v) = \sum_{S \subset N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}(v(S \cup \{i\}) - v(S)).$$

For an interpretation, suppose that all players are randomly ordered in a line, all orders being equally likely. Then $\varphi_i(v)$ represents the expected value of player $i$'s contribution to the coalition formed by the players preceding him in line. The values across players sum to $v(N)$ because they do for every realization of the ordering.

Note that, by the proof of Proposition 1, for *convex* games, the Shapley value is a convex combination of core allocations. Since the core is a convex set, the Shapley value of a convex game belongs to its core.

What is special about the Shapley value? The following axioms describe some simple properties one might want a value to have, and it will turn out that they completely characterize the Shapley value. Before stating the axioms, we need to introduce some new definitions. Player $i$ is a *dummy* in $v$ if $v(S \cup \{i\}) = v(S)$ for all $S$. Players $i$ and $j$ are *interchangeable* in $v$ if $v(S \cup \{i\}) = v(S \cup \{j\})$ for all $S$ disjoint from $\{i, j\}$.

**Axiom 5** (Symmetry)**.** If $i$ and $j$ are interchangeable in $v$ then $\varphi_i(v) = \varphi_j(v)$.

**Axiom 6** (Dummy Player)**.** If $i$ is a dummy in $v$ then $\varphi_i(v) = 0$.

**Axiom 7** (Additivity)**.** For any two games $v$ and $w$, we have $\varphi(v + w) = \varphi(v) + \varphi(w)$.

The first two axioms are quite straightforward. The additivity axiom is mathematically convenient, but difficult to motivate. The structure of $v + w$ may induce behavior that does not arise when $v$ or $w$ are considered separately.[5] If we rescale additivity to require that $\varphi(pv + (1-p)w) = p\varphi(v) + (1-p)\varphi(w)$, we can obtain an interpretation in terms of bargaining over uncertain outcomes and independence of the bargaining process from the timing of resolution of uncertainty.

**Theorem 6.** *A value satisfies the three axioms above iff it is the Shapley value.*

*Proof.* The "if" part is easy to check. The only step that is not immediate is showing that $\varphi$ satisfies the symmetry axiom. For that, suppose that $i$ and $j$ are interchangeable. Then

$$\begin{aligned}
\varphi_i(v) &= \sum_{S \subset N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}(v(S \cup \{i\}) - v(S)) \\
&= \sum_{S \subset N \setminus \{i,j\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}(v(S \cup \{i\}) - v(S)) \\
&+ \sum_{S \subset N \setminus \{i,j\}} \frac{(|S|+1)!(|N|-(|S|+1)-1)!}{|N|!}(v(S \cup \{i,j\}) - v(S \cup \{j\})) \\
&= \sum_{S \subset N \setminus \{i,j\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}(v(S \cup \{j\}) - v(S)) \\
&+ \sum_{S \subset N \setminus \{i,j\}} \frac{(|S|+1)!(|N|-(|S|+1)-1)!}{|N|!}(v(S \cup \{i,j\}) - v(S \cup \{i\})) \\
&= \varphi_j(v).
\end{aligned}$$

[5]This criticism is similar to that of the independence of irrelevant alternatives axiom for the Nash bargaining solution.

We next prove the "only if" part. Suppose that $\psi$ is a value satisfying the three axioms. We need to argue that $\psi = \varphi$.

For any non-empty coalition $T$, define the payoff function $v^T$ with $v^T(S) = 1(0)$ if $S \supset T$ ($S \not\supset T$). Such games are sometimes referred to as *carrier* games. Fix $a \in \mathbb{R}$. Note that by the *symmetry* axiom, $\psi_i(av^T) = \psi_j(av^T)$ for all $i, j \in T$. By the *dummy player* axiom, $\psi_i(av^T) = 0$ for all $i \notin T$. Hence $\psi_i(av^T) = a/|T|(0)$ for $i \in T$ ($i \notin T$), so $\psi(av^T) = \varphi(av^T)$.

We show that the $(2^{|N|}-1)$ payoff functions $v^T$ span the linear space of all payoff functions. If we view payoff functions as $(2^{|N|}-1)$-dimensional vectors, it is sufficient to show that the vectors corresponding to the $(2^{|N|}-1)$ functions are linearly independent. For a contradiction, suppose that $\sum_{T \subset N} \alpha^T v^T = 0$ with not all $\alpha$'s equal to zero. Let $S$ be a set (one of the sets) with minimal cardinality satisfying $\alpha^S \neq 0$. Then $\sum_{T \subset N} \alpha^T v^T(S) = \alpha^S \neq 0$, a contradiction.

Thus for any $v$ there exist $\alpha$'s s.t. $v = \sum_{T \subset N} \alpha^T v^T$. The additivity of $\psi$ and $\varphi$ immediately imply that

$$\psi(v) = \psi\left(\sum_{T \subset N} \alpha^T v^T\right) = \sum_{T \subset N} \psi(\alpha^T v^T) = \sum_{T \subset N} \varphi(\alpha^T v^T) = \varphi\left(\sum_{T \subset N} \alpha^T v^T\right) = \varphi(v).$$

□

An alternative characterization of the Shapley value can be obtained in terms of the following *equity* requirement—for any pair of players, the amounts that each player gains or loses from the other's withdrawal from the game are equal. The new characterization relates the outcomes of games with different sets of players. For a coaltional game with transferable payoffs $(N, v)$, we denote by $v|M$ its restriction to the players in $M$.

**Definition 8** (Balanced contributions)**.** A value $\psi$ has balanced contributions if for every coaltional game with transferable payoffs $(N, v)$ we have

$$\psi_i(v|N) - \psi_i(v|N \setminus \{j\}) = \psi_j(v|N) - \psi_j(v|N \setminus \{i\}), \forall i, j \in N.$$

Note that for $N = \{i, j\}$ the condition above becomes

$$\psi_i(v) - v(\{i\}) = \psi_j(v) - v(\{j\}),$$

which along with the constraint $\psi_i(v) + \psi_j(v) = v(\{i, j\})$ leads to the *Nash bargaining solution* for the game with bargaining set $U = \{(x_i, x_j) | x_i + x_j = v(\{i, j\})\}$ and disagreement payoffs given by $d = (v(\{i\}), v(\{j\}))$. Hence, for 2-player games, the Shapley value coincides with the Nash bargaining solution of the underlying bargaining problem. The next result shows that we can view the Shapley value as an extension of the Nash bargaining solution to multi-player games.

**Theorem 7.** *The unique value that has balanced contributions is the Shapley value.*

*Proof.* First, one can easily show that at most one value has balanced contributions. For a contradiction, let $\varphi'$ and $\varphi''$ be two different such values. Let $(N, v)$ be a game with minimal $|N|$ for which the two values yield different outcomes. Then for all $i, j \in N$, $\varphi'_i(v|N \setminus \{j\}) = \varphi''_i(v|N \setminus \{j\})$ and $\varphi'_j(v|N \setminus \{i\}) = \varphi''_j(v|N \setminus \{i\})$, along with the balancedness of $\varphi'$ and $\varphi''$, imply $\varphi'_i(v|N) - \varphi''_i(v|N) = \varphi'_j(v|N) - \varphi''_j(v|N)$. Since $\sum_{i \in N}(\varphi'_i(v|N) - \varphi''_i(v|N)) = 0$, we immediately obtain $\varphi'_i(v|N) - \varphi''_i(v|N) = 0, \forall i \in N$, or $\varphi'(v|N) = \varphi''(v|N)$, a contradiction.

We next argue that the Shapley value has balanced contributions. The Shapley value $\varphi$ is a linear function of the game, so the set of games for which $\varphi$ satisfies balanced contributions is closed under taking linear combinations. Since any game can be written as a linear combination of "carrier" games (from the proof of the previous result), it is sufficient to show that carrier games satisfy balanced contributions. The latter assertion is checked without difficulty. □

## 9. Values for Network Cooperation Structures

Under the Shapley value, the payoff of every player depends on the values of all coalitions. There are some situations where such symmetric treatment of coalitions may be unrealistic. There may be some exogenous factors—e.g., location, social relationships, trade agreements—that make some coalitions intrinsically more important/potent/feasible than others. In such situations only a subset of coalitions can actually form and influence the outcome.

Myerson (1977) looks at cooperation structures defined by networks. Fix $N$. A *network* $G$ (with vertex set $N$) consists of a set of unordered pairs of players, which we call *links*; we use the notation $ij \in G$ to represent the fact that $i$ and $j$ are linked in $G$. It is assumed that only coalitions that are internally connected can negotiate effectively. A coalition is (internally) *connected* if any two players in $S$ are connected by a path of links in $G$ among players in $S$.

For each coalition $S$, let $S|G$ denote the partition of $S$ into sets of players that are connected by $G$ within $S$,

$$S|G = \{\{i|i \text{ and } j \text{ are connected by } G \text{ within } S\}|j \in S\}.$$

Hence a coalition $S$ is internally connected if $S|G = \{S\}$.

Consider a worth function $v$. How does the outcome of the game generated by $v$ depend on the network $G$? An *allocation rule* $\psi$ specifies an allocation $\psi_i(G)$ for each $i \in N$ for all networks $G$, with the property that

$$\sum_{i \in S} \psi_i(G) = v(S), \forall S \in N|G, \forall G, \tag{9.1}$$

This condition asserts that if $S$ is a connected component of $G$ then the members of $S$ ought to share the total wealth $v(S)$ available to them.

We say that an allocation rule $\psi$ is *fair* if

$$\psi_i(G) - \psi_i(G \ominus ij) = \psi_j(G) - \psi_j(G \ominus ij), \forall ij \in G, \forall G. \tag{9.2}$$

(Here $G \ominus ij$ is the network remaining when $ij$ is removed from $G$.) Fairness requires that any two linked players benefit equally from their bilateral relationship.

We need one further definition to state the main result. What is the effective worth of a coalition $S$ that is internally disconnected in $G$? We denote by $v|G$ the worth function given by

$$(v|G)(S) = \sum_{T \in S|G} v(T), \forall S \subset N.$$

This game can be interpreted as the result of altering the situation described by $v$ to take into account the communication constraints entailed by $G$.

**Theorem 8** (Myerson 1977)**.** *For any worth function $v$, there is a unique fair allocation rule. The unique fair allocation rule is defined by*

$$\psi(G) = \varphi(v|G), \forall G,$$

*where $\varphi$ is the Shapley value.*

*Proof.* We can establish that there is at most one fair allocation rule by an argument similar to that of the previous proof. The minimal counterexample involves a network with the smallest number of links rather than vertices.

We are only left to show that the Myerson value is a fair allocation rule. First, we need to prove it is an allocation rule, that is, 9.1 holds. Note that

$$T|G = \cup_{S \in N|G} (T \cap S)|G,$$

hence

$$v|G = \sum_{S \in N|G} u^S,$$

where $u^S$ is defined by

$$u^S(T) = \sum_{R \in (T \cap S)|G} v(R), \forall T \subset N.$$

All players outside $S$ are dummies for $u^S$, so

$$\begin{aligned} \sum_{i \in S} \varphi_i(u^S) &= u^S(N) = v(S) \\ \sum_{i \in T} \varphi_i(u^S) &= 0, \forall T \in N|G, T \neq S \end{aligned}$$

because $\varphi$ satisfies the dummy player axiom. By the additivity of $\varphi$, we have

$$\varphi(v|G) = \sum_{S \in N|G} \varphi(u^S).$$

Therefore, for any $T \in N|G$,

$$\sum_{i \in T} \varphi_i(v|G) = \sum_{S \in N|G} \sum_{i \in T} \varphi_i(u^S) = u^T(N) = v(T).$$

Second, we need to prove that the Myerson value is fair, that is, it satisfies 9.2. Define the game $w = v|G - v|(G \ominus ij)$. Note that $i$ and $j$ are interchangeable in $w$. Indeed, $w(S \cup \{i\}) = w(S \cup \{j\}) = 0$ for all $S$ that do not contain $i$ and $j$. Since $\varphi$ satisfies the symmetry axiom, it must be that $\varphi_i(w) = \varphi_j(w)$. By the linearity of $\varphi$, we obtain $\varphi_i(v|G) - \varphi_i(v|(G \ominus ij)) = \varphi_j(v|G) - \varphi_j(v|(G \ominus ij))$. □

As homework, you will be asked to show stability (no player has incentives to drop links) for superadditive games and will look at an example.

## 10. Non-Cooperative Implementation of the Shapley Value

Gul (1989) introduces a dynamic model of bargaining that, under a certain set of assumptions, implements the Shapley value (in the limit, as players become patient or offers can be made frequently). Consider a game with transferable utility $(N, v)$. Assume that $v$ is *strictly superadditive*,

$$L \cap M = \emptyset \Rightarrow v(L) + v(M) < v(L \cup M) \text{ for all } L, M \subset N.$$

We interpret $v$ as follows. Each player owns a resource, and various combinations of these resources produce payoffs according to the function $v$. When pooled together, the resources of a coalition $M$ produce a payoff flow with discounted *present value* of $v(M)$ (corresponding to a constant utility flow of $(1-\delta)v(M)$ per period). At any point in time an player controls a bundle of resources he bought off other players (who sold everything and left the market), which he may choose to sell to another player who is still active in the market.

In every period $t = 0, 1, \ldots$ two active players are randomly (with equal probability) matched. One of the two matched players is randomly (with equal probability) chosen to make an offer for the *entire* bundle of the other player. If the offer is accepted, the responder leaves the market (becomes inactive) and the proposer inherits his bundle. The game proceeds without change to the next period in case of rejection. The game is over when only one active player is left. Players have a common discount factor $\delta$. The utility of player $i$ is given by

$$U^i = \sum_{t=0}^{\infty} [(1-\delta)v(M_t^i) - r_t^i]\delta^t,$$

where $M_t^i$ denotes the coalition of players whose resources $i$ holds at time $t$ and $r_t^i$ is the payment $i$ makes at $t$. The equilibrium notion is *stationary subgame perfect equilibrium (SSPE)*. At any point in time each player's behavior depends only on the distribution of resources in the market and the current offer.

A state $q = (M_1, M_2, \ldots, M_k)$ refers to the situation in which, after various rounds of trading, there are $k$ players left in the game and each player (who is still in the game) $i$ owns resources $M_{k(i)} \subset N$. $Q$ is the set of all possible states (partitions of $N$). $\bar{N} \in Q$ denotes the finest partition of $N$, that is, $\bar{N} = (\{1\}, \ldots, \{n\})$. For all $L, M \in q$, let $R(q, L, M) \in Q$

denote the partition obtained from $q$ by replacing the elements $L$ and $M$ by $L \cup M$. That is, $R(q, L, M) = (q \backslash \{L, M\}) \cup \{L \cup M\}$.

Fix $q = (M_1, M_2, \ldots, M_k)$. Define the *generalized Shapley* value for the partition $q$ as follows

$$S(M_m, q) = \sum_{\{i_1, i_2, \ldots, i_s\} \subseteq \{1,2,\ldots,k\} \backslash \{m\}} \frac{s!(k-s-1)!}{k!} (v(M_{i_1} \cup M_{i_2} \cup \ldots M_{i_s} \cup M_m) - v(M_{i_1} \cup M_{i_2} \cup \ldots \cup M_{i_s})).$$

Observe that $S(i, \bar{N})$ is the Shapley value of $i \in N$ in the game $(N, v)$. In general $S(M, q)$ is the Shapley value of the player who owns the resource bundle $M$ in a game in which initial endowments are distributed according to $q$.

**Theorem 9.** *The payoffs of any family of equilibria in which every match leads to trade converges to the Shapley value as $\delta \to 1$.*

The proof of the theorem follows from the following two lemmas.

**Lemma 1.** *Assume that, after various rounds of trade, the economy has reached a situation in which only two players remain. The continuation game is a two-player bargaining game with randomly selected proposer. If at this stage $q = (M, N \backslash M)$, then the expected continuation payoffs of the two remaining players are*

$$U(M, q) = \frac{v(N) - v(N \backslash M) + v(M)}{2} = S(M, q)$$
$$U(N \backslash M, q) = \frac{v(N) - v(M) + v(N \backslash M)}{2} = S(N \backslash M, q)$$

*Proof.* Refer to the "representative" of $M$ as player 1 and to that of $N \backslash M$ as player 2. By stationarity, player 1 (2) always makes the same offer $a$ ($b$) to 2 (1). Under the assumption that the equilibrium leads to immediate agreement, we obtain the following equations for equilibrium payoffs

$$u_1 = \frac{1}{2}(v(N) - a) + \frac{b}{2}.$$

$$u_2 = \frac{a}{2} + \frac{1}{2}(v(N) - b).$$

If player 2 refuses the offer, then he enjoys the payoff flow $(1-\delta)v(N\backslash M)$ and continuation payoff $\delta u_2$. In equilibrium, player 1 makes the offer that makes 2 exactly indifferent between accepting and rejecting,

$$a = (1-\delta)v(N\backslash M) + \delta u_2.$$

Similarly,

$$b = (1-\delta)v(M) + \delta u_1.$$

Substituting the values of $a$ and $b$ in the formulae for $u_1$ and $u_2$, we find

$$\begin{aligned} u_1 &= \frac{v(N) - v(N\backslash M) + v(M)}{2} \\ u_2 &= \frac{v(N) - v(M) + v(N\backslash M)}{2}. \end{aligned}$$

□

**Lemma 2.** *Suppose that for any $i, j \in N$, equilibrium payoffs in a game in which players $i$ and $j$ traded satisfy $\lim_{\delta\to 1} U(M, R(\bar{N}, i, j)) = S(M, R(\bar{N}, i, j))$ for all $M \in R(\bar{N}, i, j)$. Then $\lim_{\delta\to 1} U(i, \bar{N}) = S(i, \bar{N})$.*

*Proof.* Taking into account all possible matches in the first period of the game, $U(i, \bar{N})$ is given by the following expectation

$$\begin{aligned} U(i,\bar{N}) = \frac{2}{n(n-1)} \Bigg[ \sum_{j\neq i} \frac{1}{2} \Big[ &(1-\delta)v(\{i\}) + \delta U(i,\bar{N}) + \\ &(1-\delta)v(\{i,j\}) + \delta U(ij, R(\bar{N}, i, j)) - (1-\delta)v(j) - \delta U(j, \bar{N}) \Big] + \\ &\sum_{\{j,k\}\subset N\backslash\{i\}} \Big((1-\delta)v(\{i\}) + \delta U(i, R(\bar{N}, j, k))\Big) \Bigg] \end{aligned}$$

The first term captures offers accepted from other players $j$, the second the gains from trade when $i$ is the proposer, and the third term continuation payoffs when other pairs trade.

Then $(U(i, \bar{N}))_{i\in N}$ is the solution of a linear system of equations with coefficients that are linear functions of $\delta$. Using Cramer's rule, each payoff can be expressed as a fraction of polynomials in $\delta$. One can easily show that if $|N| \geq 3$ then the system in non-singular, and hence the solution converges as $\delta \to 1$ to the solution of the limit system,

$$U(i,\bar{N}) = \frac{1}{n(n-1)}\left(\sum_{j\neq i}[U(i,\bar{N}) + U(ij, R(\bar{N},i,j)) - U(j,\bar{N})] + 2\sum_{\{j,k\}\subset N\setminus\{i\}} U(i, R(\bar{N},j,k))\right).$$

We have to show that $U(ij, R(\bar{N},i,j)) = S(ij, R(\bar{N},i,j))$ for all $i \neq j \in N$, then $U(i,\bar{N}) = S(i,\bar{N})$, which is equivalent to showing that the Shapley value satisfies

$$S(i,\bar{N}) = \frac{1}{n(n-1)}\left(\sum_{j\neq i}[S(i,\bar{N}) + S(ij, R(\bar{N},i,j)) - S(j,\bar{N})] + 2\sum_{\{j,k\}\subset N\setminus\{i\}} S(i, R(\bar{N},j,k))\right).$$

In order to prove this last equation, we can again use carrier games and the linearity of the Shapley value. A clean argument, showing that the desired formula leads to an alternative characterization of the Shapley value, can be found in Haviv (1995). The Shapley value satisfies a consistency property with respect to amalgamations of pairs of players. □

Lemma 2 states that if after one transaction occurs, the equilibrium is such that for $\delta$ close to 1 it yields an expected payoff equal to his Shapley value for every remaining player (relative to a new distribution of resources), then for $\delta$ close enough 1 the equilibrium will yield each player expected payoff equal to his Shapley value in the original game before any transaction takes place. But this is equivalent to saying that, if the equilibrium yields expected payoffs according to the Shapley values in all $n-1$ player games, then the equilibrium will yield payoffs according to the Shapley value in all $n$ player games.