Chapter 15

Cooperative Game Theory

Cooperative (coalitional) theory abstracts from individual moves and asks two questions: *which coalitions form*, and *how are the gains divided*? Its solution concepts—the core and the Shapley value—are now everyday tools far beyond economics, from cost allocation to explainable machine learning.

15.1 Coalitional games

Definition 15.1: TU coalitional game

A *transferable-utility (TU) coalitional game* is a pair (N, v) where N is a finite player set and $v : 2^N \to \mathbb{R}$ with $v(\emptyset) = 0$ assigns each *coalition* $S \subseteq N$ its *worth*—the total payoff its members can guarantee by cooperating among themselves. The game is *superadditive* if $v(S \cup T) \geq v(S) + v(T)$ for disjoint S, T. An *allocation* $x \in \mathbb{R}^N$ is *efficient* if $\sum_i x_i = v(N)$.

Example 15.1: A gloves market

Players $L = \{1\}$ owns a left glove; $R = \{2, 3\}$ each own a right glove. A pair is worth \$1: $v(S) = \min(|S \cap L|, |S \cap R|)$. Which divisions of the \$1 generated by a pair are "stable"?

15.2 The core

Definition 15.2: Core

The *core* of (N, v) is the set of efficient allocations no coalition can improve upon:

$$C(v) \;=\; \Big\{ x \in \mathbb{R}^N : \textstyle\sum_{i \in N} x_i = v(N) \text{ and } \sum_{i \in S} x_i \geq v(S) \;\; \forall S \subseteq N \Big\}.$$

Example 15.2: Gloves, continued

In Example 15.1, $v(N) = 1$. Core constraints: $x_1 + x_2 \geq 1$, $x_1 + x_3 \geq 1$, $x_i \geq 0$, $\sum x_i = 1$. Adding the first two: $2x_1 + x_2 + x_3 \geq 2$, so $x_1 \geq 1$: the core is the *single point* $(1, 0, 0)$. The long side of the market gets nothing—competition between the two right-glove owners drives their "price" to zero. The core captures market discipline with brutal clarity.

Example 15.3: An empty core

Three players; any two-player coalition is worth 1; the grand coalition is worth 1; singletons worth 0. Core: $x_i + x_j \geq 1$ for all pairs sums to $2(x_1 + x_2 + x_3) \geq 3 > 2$—impossible. *No* division is stable: any agreement between two players can be undercut by the third. Empty cores model genuinely unstable situations (odd-player majority bargaining, some cost-sharing problems).

Theorem 15.1: Bondareva–Shapley (1963, 1967)

A TU game has a nonempty core if and only if it is *balanced*: for every collection of weights $\lambda_S \in [0,1]$ with $\sum_{S \ni i} \lambda_S = 1$ for each player i, we have $\sum_S \lambda_S v(S) \leq v(N)$. (Proof: LP duality applied to the core's linear program—the Minimax/duality toolkit of Chapter 12 again.)

15.3 The Shapley value

The core may be empty or huge; Shapley (1953) asked instead for a unique, fair one-point solution.

Definition 15.3: Shapley value

The *Shapley value* gives player i her average marginal contribution over all $|N|!$ orders of arrival:

$$\phi_i(v) \;=\; \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!}\,\big[v(S \cup \{i\}) - v(S)\big].$$

Theorem 15.2: Shapley's axiomatization (1953)

The Shapley value is the unique allocation rule satisfying: **efficiency** ($\sum_i \phi_i = v(N)$); **symmetry** (interchangeable players get the same); **null player** (a player adding nothing to any coalition gets 0); and **additivity** ($\phi(v + w) = \phi(v) + \phi(w)$).

Example 15.4: Airport cost game

Three airlines need runways of length small ($c = 1$), medium ($c = 2$), large ($c = 3$); one runway serving the largest need serves all. Fair division of the \$3 cost? Model the *cost* game and compute Shapley: the first kilometer is shared by all three ($\frac{1}{3}$ each), the second by the two larger airlines ($\frac{1}{2}$ each), the third only by the largest. Shapley cost shares: $(\frac{1}{3},\ \frac{1}{3} + \frac{1}{2},\ \frac{1}{3} + \frac{1}{2} + 1) = (0.33,\ 0.83,\ 1.83)$. This "divide each cost increment among its users" logic is used for actual airport landing fees, water infrastructure, and network cost sharing.

Example 15.5: Shapley values in machine learning (SHAP)

To explain a model's prediction, treat features as players and the prediction (relative to baseline) as $v(S)$ for each feature subset S; each feature's Shapley value is its fair credit for the prediction. The axioms translate to desirable explanation properties (local accuracy, missingness, consistency)—which is why SHAP became the standard attribution method.

Exact computation is exponential in the number of features; practical algorithms sample permutations—the same computational tension as everywhere in game theory (Chapter 17).

Remark: Core vs. Shapley; power indices

The Shapley value need not lie in the core (it does for *convex* games, where it is the core's center of gravity). For voting games ($v(S) \in \{0,1\}$: winning or losing coalitions), the Shapley value becomes the *Shapley–Shubik power index*; a close cousin is the Banzhaf index. Applied to the UN Security Council, the five permanent members hold over 98% of the power by Shapley–Shubik—formalizing the intuition that vetoes, not vote counts, are what matter.

15.4 The nucleolus and other solutions

The *nucleolus* (Schmeidler 1969) picks the allocation that lexicographically minimizes the largest coalition dissatisfaction $e(S,x) = v(S) - \sum_{i\in S} x_i$. It always exists, is unique, and lies in the core whenever the core is nonempty. Its pedigree is ancient: the nucleolus reproduces the disputed-garment divisions of the Babylonian Talmud (Aumann–Maschler 1985)—a 2000-year-old text anticipating a 1969 solution concept.

15.5 Matching markets (NTU cooperation)

Without transferable utility, stability remains the guiding idea. Recall deferred acceptance (Example 14.2); the coalitional view adds:

Theorem 15.3: Structure of stable matchings

In the marriage market, the set of stable matchings is a nonempty *lattice*: there is a man-optimal stable matching (simultaneously best for every man and worst for every woman among stable matchings) and dually a woman-optimal one. The core of the matching game equals the set of stable matchings.

Example 15.6: Kidney exchange

Patient–donor pairs that are immunologically incompatible can swap donors. With no money allowed (illegal almost everywhere), this is a pure matching-market design problem: top trading cycles and chain algorithms (Roth–Sönmez–Ünver) now run national kidney exchange programs, saving thousands of lives—cooperative game theory at its most concrete.

Exercises

Exercise 15.1: Core computation

$N = \{1,2,3\}$; $v(\{i\}) = 0$; $v(\{1,2\}) = 90$, $v(\{1,3\}) = 80$, $v(\{2,3\}) = 70$; $v(N) = 135$. Find the core (describe it as a polytope and give a concrete member), and compute the Shapley value. Is the Shapley value in the core?

Exercise 15.2: Production economy

A landowner (L) and w identical workers produce only together: $v(S) = 0$ if $L \notin S$, and $v(S) = g(k)$ if $L \in S$ with k workers, where g is increasing concave. Show core allocations pay each worker at most the marginal product $g(w) - g(w-1)$, and compute the Shapley value for $g(k) = \sqrt{k}$, $w = 3$.

Exercise 15.3: Shapley–Shubik index

A committee has voters with weights $(4, 3, 2, 1)$ and quota 6. List the pivotal positions and compute the Shapley–Shubik power index of each voter. Compare power with weight; comment.

Exercise 15.4: Instability in roommates

Show by a four-player example that the *roommate problem* (one-sided matching: everyone can be matched with anyone) can have an empty set of stable matchings, in contrast to the two-sided marriage market. Explain what structural feature of two-sided markets the example destroys.