# CS639: Algorithmic Game Theory & Learning

## Chapter 8: Cooperative Game Theory

Kirthevasan Kandasamy

UW-Madison

# Non-cooperative vs Cooperative Games

So far in class we have been studying **non-cooperative games**, where players choose actions (strategies) in an interaction with other players.

Broadly, non-cooperative games fall into one of two categories:

- *Normal form games:* A one-time interaction between players where players choose a single action (e.g. prisoner's dilemma).
- *Extensive form games:* A multi-round interaction where players repeatedly choose different actions. The set of available actions to each player could also change as the game proceeds (e.g. chess).

In **cooperative game theory**, instead of choosing actions, players make binding agreements with each other.

# Cooperative game theory

Overarching theme in cooperative game theory: when multiple agents can achieve more than the sum of their parts, how do you divide the value created?

In this chapter, we will study two models for cooperation.

- *Axiomatic bargaining:* Typically used when the collaboration breaks down if even one agent leaves.
- *Coalitional games (a.k.a cooperative games):* Typically used in when agents can form coalitions and continue collaboration even when an agent leaves.

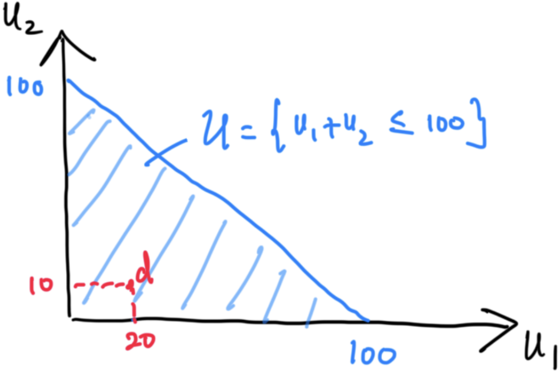
# Outline

1. Axiomatic bargaining
2. Coalitional Games
3. Case study: cooperation in collaborative learning

Slides are intended as teaching aids only and do not include all material discussed in class. Students are strongly encouraged to attend lectures and take their own notes.

# Ch 8.1: Axiomatic bargaining

**Motivating example.** An employer (Agent 1) can make \$20 on her own, and a worker (Agent 2) can make \$10 on her own. They can make \$100 if they work together. Clearly, it makes sense for them to collaborate. But how do they split the \$100?

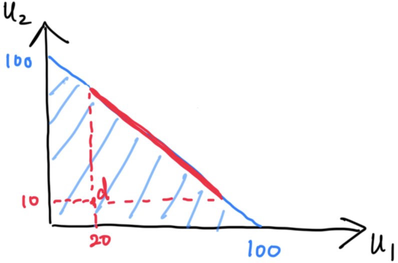


# Motivating example (cont'd)

**Discussion.** What properties would we like to have in a policy to allocate the $100?

*(i)* What if our policy allocates $u_1 < 20$ or $u_2 < 10$?

*(ii)* What if $u_1 = 45, u_2 = 45$?



*(iii)* Suppose $d_1 = 20$ and $d_2 = 20$? What should $u_1, u_2$ be?

*(iv)* What if agent 1 reports in cents, so that $d_1 = 2000$ and $d_2 = 10$?

# A model for bargaining

- There are $n$ agents.

- Let $\mathcal{U} \subseteq \mathbb{R}^n$ be the set of possible utilities for the agents via collaboration. This is called the *utility possibility set* (UPS).

- Let $d \in \mathcal{U}$ be the *disagreement point*, where the utilities default to when there is no collaboration.

- We will assume that there exists a point where everyone is strictly better off, *i.e.*, there exists $u \in \mathcal{U}$ such that $u > d$ (inequality holding pointwise). .

*E.g., In the previous example,* $\mathcal{U} = \{(u_1, u_2) : u_1 + u_2 \leq 100\}, \quad d = (20, 10).$

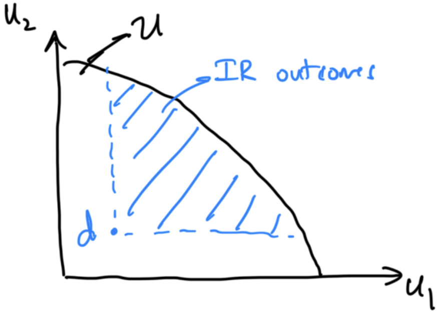
**Bargaining solution.** A *bargaining solution* is a rule $f$ that maps the utility possibility set $\mathcal{U}$ and a disagreement point $d \in \mathcal{U}$ to an agreement point $f(\mathcal{U}, d) \in \mathcal{U}$.

# A model for bargaining (cont'd)

**Individual rationality (IR).** Often, IR is a necessary requirement in bargaining, as, if any agent is worse off, she may simply not participate.

Formally, IR says that no agent should be worse off than at the disagreement point (initial state), *i.e.*,

$$f(\mathcal{U}, d) \geq d \qquad \text{(pointwise)}$$



# A model for bargaining (cont'd)

**Translation invariance (TI).** We say that a bargaining solution is TI if,

$$\text{for all } \beta \in \mathbb{R}^n, \text{ we have} \qquad f(\mathcal{U} + \beta, d + \beta) = f(\mathcal{U}, d) + \beta.$$

where, $\mathcal{U} + \beta = \{(u_1 + \beta_1, \ldots, u_n + \beta_n);\ u \in \mathcal{U}\}$.

Recall our motivating example to illustrate why TI is a desirable property.

**Example.** An employer and worker can respectively make \$20 and \$10 on their own. Together, they can make \$100. Hence, $\mathcal{U} = \{(u_1, u_2) : u_1 + u_2 \leq 100\}, \quad d = (20, 10)$. Say our bargaining solution chose $f(\mathcal{U}, d) = (60, 40)$.

Now, suppose that the employer decides to count some other \$100 income as part of this deal so that so $d' = (120, 10)$ and $\mathcal{U}' = \{(u_1, u_2); u_1 + u_2 \leq 200\}$. This should not change the outcome, *i.e.*, we should still choose $f(\mathcal{U}', d') = (160, 40)$.

# A model for bargaining (cont'd)

Recall the following definitions:

**Bargaining solution:** A rule $f$ that maps the utility possibility set $\mathcal{U}$ and a disagreement point $d \in \mathcal{U}$ to an agreement point $f(\mathcal{U}, d) \in \mathcal{U}$.

**IR:** $f(\mathcal{U}, d) \geq d$. **TI:** $f(\mathcal{U} + \beta, d + \beta) = f(\mathcal{U}, d) + \beta$.

We will only consider IR and TI bargaining solutions going forward.

Due to the TI property, we can assume, w.l.o.g, that $d = \mathbf{0}$, as we have

$$f(\mathcal{U}, d) = f(\mathcal{U}', \mathbf{0}) + d, \qquad \text{where, } \mathcal{U}' = \mathcal{U} - d$$

Then, IR becomes $f(\mathcal{U}', \mathbf{0}) \geq \mathbf{0}$, *i.e.*, the solution must lie in $\mathbb{R}_+^n$. Since we only care about points in $\mathcal{U}'$ that could be selected by an IR solution, we may restrict attention to $\mathcal{U}' \cap \mathbb{R}_+^n$.

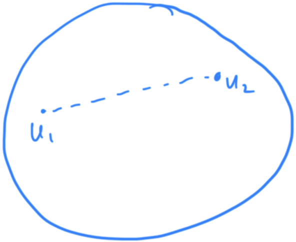
Hence, going forward, for simplicity, we will assume $\mathcal{U} \subset \mathbb{R}_+^n$ and simply write $f(\mathcal{U})$ instead of $f(\mathcal{U}, \mathbf{0})$.

# A model for bargaining (cont'd)

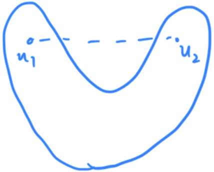
Finally, it is customary to assume that $\mathcal{U}$ is a closed and convex set.

- convex: for any $u_1, u_2 \in \mathcal{U}$, and $\lambda \in (0,1)$, we have $\lambda u_1 + (1-\lambda)u_2 \in \mathcal{U}$.

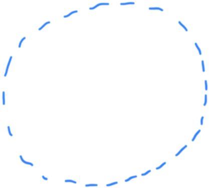
- closed: contains its boundary points (slightly more formally, every convergent sequence of points in the set converges to a point in the set).



closed and convex



not convex



not closed

# Designing bargaining solutions

A common (although not the only) approach to designing a bargaining solution is to define a welfare function $W : \mathcal{U} \to \mathbb{R}_+$ and choose

$$f(\mathcal{U}) = \operatorname*{argmax}_{u \in \mathcal{U}} W(u).$$

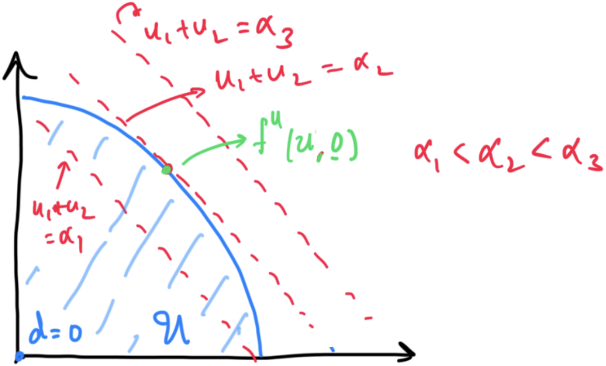
Some common welfare functions are the social (utilitarian) welfare $W_{\mathrm{soc}}$, egalitarian welfare $W_{\mathrm{egal}}$, and Nash welfare $W_{\mathrm{Nash}}$, defined below:

$$W_{\mathrm{soc}}(u) = \sum_{i \in [n]} u_i, \qquad W_{\mathrm{egal}}(u) = \min_{i \in [n]} u_i, \qquad W_{\mathrm{Nash}}(u) = \prod_{i \in [n]} u_i,$$

The resulting bargaining solutions are called utilitarian bargaining, egalitarian bargaining, and Nash bargaining.

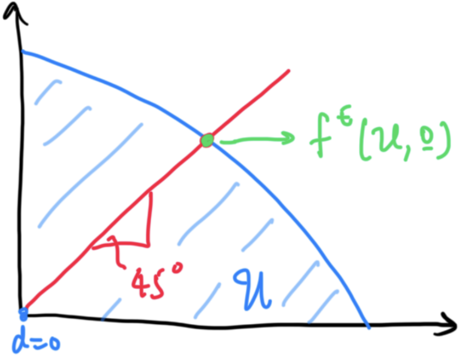
# Utilitarian bargaining

$$f^{\mathrm{U}}(\mathcal{U}) = \operatorname*{argmax}_{u \in \mathcal{U}} \sum_{i \in [n]} u_i .$$



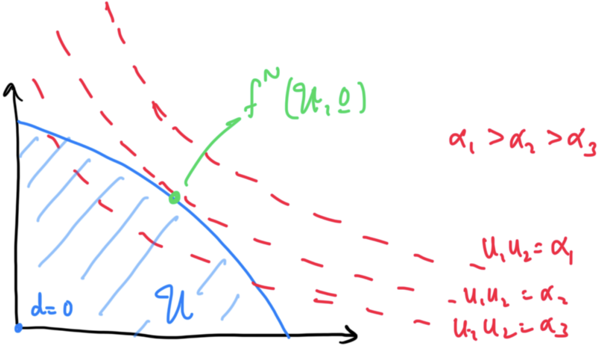
# Egalitarian bargaining

$$f^{\mathrm{E}}(\mathcal{U}) = \operatorname*{argmax}_{u \in \mathcal{U}} \min_{i \in [n]} u_i.$$



# Nash bargaining

$$f^{\mathrm{N}}(\mathcal{U}) = \operatorname*{argmax}_{u \in \mathcal{U}} \prod_{i \in [n]} u_i .$$



# Bargaining axioms

What is a reasonable bargaining solution? There is no "correct" answer to this, as it often depends on the application at hand.

The approach taken by John Nash was to formulate a set of four axioms (in addition to IR and TI) that a solution should satisfy.

**1. Pareto-efficiency.** A bargaining solution $f$ is PE if $f(\mathcal{U})$ is PE for all $\mathcal{U} \subset \mathbb{R}^n_+$. That is, for all $u \in \mathcal{U}$, if $\exists\, i \in [n]$ such that $u_i > f_i(\mathcal{U})$, then $\exists\, j \in [n]$ such that $u_j < f_j(\mathcal{U})$.

**2. Symmetry.** We say that a set $\mathcal{U}$ is symmetric if it is unaltered under permutation of the axes. That is, if $u = (u_1, \ldots, u_n) \in \mathcal{U}$, then $(u_{\sigma(1)}, \ldots, u_{\sigma(n)}) \in \mathcal{U}$ for any permutation $\sigma$ on $n$.

A bargaining solution $f$ is symmetric if, $f_i(\mathcal{U}) = f_j(\mathcal{U})\ \forall\, i, j$ when $\mathcal{U}$ is a symmetric set.

# Bargaining axioms (cont'd)

**3. Scale invariance.** For $\alpha \in \mathbb{R}^n_{++}$, let us define,

$$\phi_\alpha(u) = (\alpha_1 u_1, \ldots, \alpha_n u_n), \qquad \phi_\alpha(\mathcal{U}) = \{\phi_\alpha(u) : u \in \mathcal{U}\}.$$

Then $f$ is *scale invariant* if $\forall\, \alpha \in \mathbb{R}^n_{++}$, we have $f\big(\phi_\alpha(\mathcal{U})\big) = \phi_\alpha\big(f(\mathcal{U})\big)$.

*Example.* Consider our previous example where $d = (20, 10)$ and $\mathcal{U} = \{(u_1, u_2);\ u_1 + u_2 \leq 100\}$ (measured in dollars). Suppose we chose $f(\mathcal{U}, d) = (60, 40)$. Applying TI and IR, we have $f(\mathcal{U}') = (40, 30)$ where $\mathcal{U}' = \{(u_1, u_2); u_1 + u_2 \leq 70\}$.

Suppose the employer decides to count their income in cents, so that

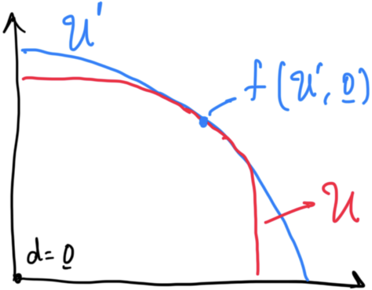
$$d = (2000, 10), \qquad \mathcal{U} = \{(u_1, u_2);\ u_1/100 + u_2 \leq 100\}$$
$$\mathcal{U}' = \{(u_1, u_2);\ u_1/100 + u_2 \leq 70\} \qquad \text{(after applying TI/IR).}$$

This should not change the outcome, *i.e.*, we should still choose $f(\mathcal{U}') = (4000, 30)$ or equivalently $f(\mathcal{U}, d) = (6000, 40)$.

# Bargaining axioms (cont'd)

**4. Independent of irrelevant alternatives .** Let $\mathcal{U} \subseteq \mathcal{U}'$. If $f(\mathcal{U}', d) \in \mathcal{U}$, then $f(\mathcal{U}, d) = f(\mathcal{U}', d)$.

Intuitively, removing an outcome $u\,(\neq f(\mathcal{U}, d))$ that you have not selected anyway from the utility possibility set (UPS) should not change your selection. This would indicate some inconsistency in the decision process.



# Nash bargaining

**Question:** Is there a bargaining solution which satisfies all axioms?
**Ans:** Yes! Nash bargaining, uniquely satisfies all four axioms.

$$f^{\mathrm{N}}(\mathcal{U}) = \operatorname*{argmax}_{u\in\mathcal{U}} \prod_{i\in[n]} u_i.$$

We will verify that Nash bargaining satisfies the four axioms. (We will not show uniqueness in class.) Let $\mathcal{U} \subset \mathbb{R}_+^n$ be given and let $u^\star = f^{\mathrm{N}}(\mathcal{U})$.

**Pareto-efficiency.** By way of contradiction, suppose $\exists\ u'$ such that $u' \geq u^\star$ (element wise) and $u'_i > u^\star_i$ for some $i \in [n]$. Then $\prod_{i=1}^n u'_i > \prod_{i=1}^n u^\star_i$ which contradicts the fact that $u^\star$ maximizes $W_{\mathrm{Nash}}$.

**Recall,** we say $f$ is PE if, for all $u \in \mathcal{U}$, if $\exists\ i \in [n]$ such that $u_i > f_i(\mathcal{U})$, then $\exists\ j \in [n]$ such that $u_j < f_j(\mathcal{U})$.

# Nash bargaining (cont'd)

**Recall:** A bargaining solution $f$ is symmetric if, $f_i(\mathcal{U}) = f_j(\mathcal{U})\ \forall\, i, j$ when $\mathcal{U}$ is a symmetric set. Here, a set $\mathcal{U}$ is symmetric if for any $u = (u_1, \ldots, u_n) \in \mathcal{U}$ and any permutation $\sigma$ on $n$, we have $(u_{\sigma(1)}, \ldots, u_{\sigma(n)}) \in \mathcal{U}$.

**Symmetry.** Let $\mathcal{U}$ be symmetric, and let $u^\star = (u_1^\star, \ldots, u_n^\star) = f^{\mathrm{N}}(\mathcal{U})$. Then, for any permutation $\sigma$, we have $u_\sigma^\star = (u_{\sigma(1)}^\star, \ldots, u_{\sigma(n)}^\star) \in \mathcal{U}$ and

$$W_{\mathrm{Nash}}(u^\star) = \prod_{i=1}^{n} u_i^\star = \prod_{i=1}^{n} u_{\sigma(i)}^\star = W_{\mathrm{Nash}}(u_\sigma^\star).$$

Therefore, $u^\star$ is also a maximizer of $W_{\mathrm{Nash}}$. However, $W_{\mathrm{Nash}}$ has a *unique* maximizer[1].

By uniqueness, $u^\star = u_\sigma^\star$ *for every* permutation $\sigma$, and hence

$$u_1^\star = u_2^\star = \cdots = u_n^\star.$$

---

[1] We will use this fact without a rigorous proof. It follows from the fact that $\log W_{\mathrm{Nash}}$ is a strictly concave function and $\mathcal{U}$ is a convex set.

# Nash bargaining (cont'd)

**Recall,** For $\alpha \in \mathbb{R}^n_{++}$, denote $\phi_\alpha(u) = (\alpha_1 u_1, \ldots, \alpha_n u_n)$, $\phi_\alpha(\mathcal{U}) = \{\phi_\alpha(u) : u \in \mathcal{U}\}$. Then $f$ is *scale invariant* if $\forall\, \alpha \in \mathbb{R}^n$, we have $f\big(\phi_\alpha(\mathcal{U})\big) = \phi_\alpha\big(f(\mathcal{U})\big)$.

**Scale invariance.** We begin with the following identity:

$$W_{\mathrm{Nash}}(\phi_\alpha(u)) = \prod_{i=1}^{n} (\alpha_i u_i) = \left(\prod_{i=1}^{n} \alpha_i\right) \cdot \left(\prod_{i=1}^{n} u_i\right) = \left(\prod_{i=1}^{n} \alpha_i\right) \cdot W_{\mathrm{Nash}}(u)$$

Hence,

$$f^{\mathrm{N}}(\mathcal{U}) = \operatorname*{argmax}_{u \in \mathcal{U}} W_{\mathrm{Nash}}(u) = \operatorname*{argmax}_{u \in \mathcal{U}} W_{\mathrm{Nash}}(\phi_\alpha(u)).$$

Therefore,

$$\begin{aligned} f^{\mathrm{N}}\big(\phi_\alpha(\mathcal{U})\big) &= \operatorname*{argmax}_{u' \in \phi_\alpha(\mathcal{U})} W_{\mathrm{Nash}}(u') = \operatorname*{argmax}_{\phi_\alpha(u) \in \phi_\alpha(\mathcal{U})} W_{\mathrm{Nash}}(\phi_\alpha(u)) \\ &= \phi_\alpha\left(\operatorname*{argmax}_{u \in \mathcal{U}} W_{\mathrm{Nash}}(\phi_\alpha(u))\right) = \phi_\alpha\left(f^{\mathrm{N}}(\mathcal{U})\right). \end{aligned}$$

# Nash bargaining (cont'd)

**Recall:** Let $\mathcal{U} \subseteq \mathcal{U}'$. $f$ is IIA if, whenever $f(\mathcal{U}', d) \in \mathcal{U}$, then $f(\mathcal{U}, d) = f(\mathcal{U}', d)$.

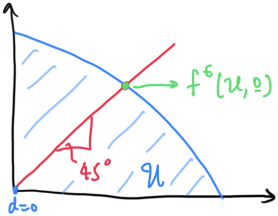
**Independent of irrelevant alternatives**[2]**.** If $u^\star \in \mathcal{U} \subseteq \mathcal{U}'$ maximizes $W_{\mathrm{Nash}}$ in $\mathcal{U}'$, it should also maximize $W_{\mathrm{Nash}}$ in $\mathcal{U}$.

Nash bargaining is not the only 'acceptable' solution. We may prefer other bargaining solutions depending on the application. We will (briefly) look at which of the other axioms, the utilitarian and egalitarian solutions satisfy.

---

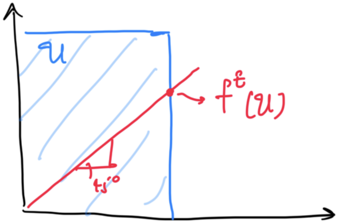
[2] Any bargaining solution where you maximize a welfare function and which has a unique maximizer satisfies IIA.

# Egalitarian bargaining



$$f^{\mathrm{E}}(\mathcal{U}) = \operatorname*{argmax}_{u \in \mathcal{U}} \min_{i \in [n]} u_i .$$

**Pareto-efficiency.** Yes, if $\mathcal{U}$ is a *strictly* convex set.

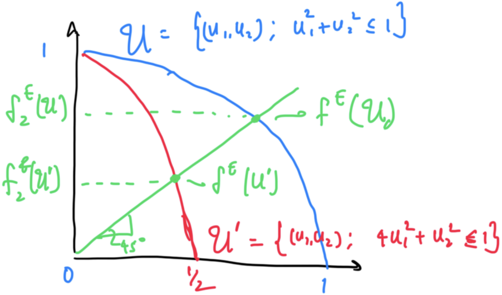


**Symmetry.** Yes.

# Egalitarian bargaining (cont'd)

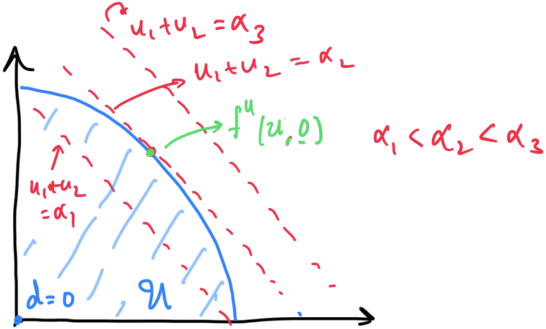
**Independent of irrelevant alternatives.** Yes, as we are maximizing a welfare function.

**Scale invariance.** No. To construct a counter example let $\mathcal{U} = \{(u_1, u_2);\ u_1^2 + u_2^2 \leq 1\}$ and consider the transformation $\phi(u) = \left(\frac{1}{2}u_1, u_2\right)$. From the figure below we see the transformation is not scale invariant since agent 2 (whose utility units have not changed) gets less utility.



# Utilitarian bargaining

$$f^{\mathrm{U}}(\mathcal{U}) = \operatorname*{argmax}_{u \in \mathcal{U}} \sum_{i \in [n]} u_i .$$



Utilitarian bargaining satisfies Pareto-efficiency, IIA, and symmetry. It does not satisfy scale invariance. *(try at home)*

# Ch 8.2: Coalitional games

**Example (Glove market, from Section 12.1 in KP).** There are 3 merchants (agents) who are selling gloves in a market. Agent 1 has a left glove, while Agents 2 and 3 have a right glove each. A buyer wishes to purchase a *pair* of gloves for \$1, but does not wish to engage with the agents individually.

*Question.* How should the merchants split the \$1?

# Coalitional games (cont'd)

**Coalitional game (a.k.a characteristic function game, cooperative game).** A *coalitional game* (with transferable utility) is defined by a set of $n$ agents, denoted $[n]$, and a *characteristic function* $v : 2^{[n]} \to \mathbb{R}_+$, where $v(S)$ denotes the value a *coalition* (subset) $S \subseteq [n]$ of players can achieve on their own.

Let $\mathcal{V}$ be the class of all characteristic functions satisfying *(i)* $v(\varnothing) = 0$, and *(ii)* $v(S) \leq v(S')$ whenever $S \subseteq S'$ (monotonicity, more agents can only add more value to a coalition).

**Definition (Solution/Allocation).** A *solution* (or *allocation*) $\psi$ to a coalitional game maps the characteristic function to a vector of payoffs for each player, *i.e.*,

$$\psi : \mathcal{V} \to \mathbb{R}_+^n, \qquad \text{where } \psi(v) = \{\psi_i(v)\}_{i \in [n]} \in \mathbb{R}_+^n$$

# Solution concepts in coalitional games

How does one choose a solution for a coalitional game?
Much like bargaining, there is no one "correct" answer. But there are many solution concepts that are commonly studied:

1. The core (will cover in class)
2. Shapley value (will cover in class)
3. Banzhaf Index (in HW5)
4. Kernel
5. Nucleolus
6. Bargaining set

# Ch 8.2.2: The core

The core captures a notion of stability in coalitional games. An allocation is in the core if no subset of players (coalition) can improve upon their utilities by leaving the group and acting on their own.

**Definition (Core).** Formally, an allocation $\psi(v)$ is in the *core* if it satisfies the following two properties:

*(i) Efficiency:* the players extract the maximum possible value, *i.e.*,

$$\sum_{i=1}^{n} \psi_i(v) = v([n]),$$

*(ii) Stability:* The total payoff assigned to a coalition $S$ is at least what it could guarantee itself. That is,

$$\forall\, S \subseteq [n], \quad \sum_{i \in S} \psi_i(v) \geq v(S).$$

# The Core (cont'd)

**Example 1 (Glove market).** Three agents are selling gloves in a market. Agent 1 has a left glove, while Agents 2 and 3 have a right glove each. A buyer wishes to purchase a *pair* of gloves for \$1.

Consider the allocation $\psi_1 = 0.6$, $\psi_2 = 0.2$, $\psi_3 = 0.2$. This is not in the core as agents 1 and 2 can split and divide agent 3's \$0.2 between them.

An allocation in the core must satisfy, (writing $\psi_i$ instead of $\psi_i(v)$ for brevity)

$$\begin{aligned}
&\psi_1 + \psi_2 + \psi_3 = 1 && \text{(efficiency)}\\
&\psi_1 + \psi_2 \geq 1, \qquad \psi_1 + \psi_3 \geq 1, \qquad \psi_2 + \psi_3 \geq 0 && \text{(stability)}\\
&\psi_1 \geq 0, \qquad \psi_2 \geq 0, \qquad \psi_3 \geq 0 && \text{(stability)}
\end{aligned}$$

The only solution in the core is

$$\psi_1 = 1, \quad \psi_2 = 0, \quad \psi_3 = 0.$$

# The Core (cont'd)

**Example 2 (Gold miners, Example 12.2.1 in KP).** There are $n$ miners. It takes 2 miners to mine a bar of gold. A single miner cannot mine any gold on her own. Hence,

$$v(S) = \left\lfloor \frac{|S|}{2} \right\rfloor .$$

However, if $n$ is odd, the core is empty. For instance, if $n = 3$, we require

$$\begin{aligned} &\psi_1 + \psi_2 + \psi_3 = 1 && \text{(efficiency)} \\ &\psi_1 + \psi_2 \geq 1, \quad \psi_1 + \psi_3 \geq 1, \quad \psi_2 + \psi_3 \geq 1, && \text{(stability)} \\ &\psi_1, \psi_2, \psi_3 \geq 0 && \text{(stability)} \end{aligned}$$

No solution satisfies this.

If $n$ is even, then $\psi(v) = (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$ is in the core. *(try at home)*

# The Core (cont'd)

**Example 3 (Employer vs employee).** An employer can make \$20 on her own, and a worker can make \$10 on her own. Together, they can make \$100. How do they split the \$100?

$$\psi_1 + \psi_2 = 100 \qquad \text{(efficiency)}$$

$$\psi_1 \geq 20, \quad \psi_2 \geq 10 \qquad \text{(stability)}$$

There are multiple solutions in the core.

**Drawbacks.** The core is a fairly strong requirement. But it also has some drawbacks:

1. The core may not always exist (*e.g.*, gold miners).
2. When it exists, it may not exist uniquely (*e.g.*, employer vs employee)
3. The allocations in the core may not seem "fair" (*e.g.*, glove market).

# Ch 8.2.2: Shapley value and Shapley's axioms

How else can we divide value in a coalitional game?
One approach, taken by Lloyd Shapley, is to define a set of reasonable axioms that a solution should satisfy. This gives rise to the Shapley value. We will first define the Shapley value. Later, we will study Shapley's axioms.

**Shapley value (motivation).** Consider the following intuitive idea:

- For an agent $i$, consider a subset $S \subseteq [n] \setminus \{i\}$ and see how much $i$ can contribute by joining: $v(S \cup \{i\}) - v(S)$.

- Do this for all possible subsets $S \subseteq [n]$ and then take the average to determine the average contribution.

For reasons that will become clear later, we will look at all possible orderings (permutations) of $[n]$, and see how much $i$ contributes when $i$ joins in that order.

# Shapley value

For a set of $n$ players $[n]$, let $\Pi_n$ denote all possible permutations of $[n]$. Note $|\Pi_n| = n!$. Let $S_\pi(i)$ denote the predecessors of $i$ in the permutation $\pi \in \Pi_n$.

The **Shapley value** $\psi_i(v)$ of player $i$ for characteristic function $v$ is given by:

$$\psi_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Big( v\big(S_\pi(i) \cup \{i\}\big) - v\big(S_\pi(i)\big) \Big).$$

Term inside the sum is the marginal increase in the coalition's value when $i$ joins.

---

[2] The Shapley value can also be defined for characteristic functions $v$ that do not satisfy monotonicity, and much of the theory we develop here can also be extended. In such cases, $\psi_i(v)$ can be negative, as the terms inside the summation can be negative.

# Shapley value examples

**Example (Employer vs employee).** An employer can make \$20 on her own, and a worker can make \$10. Together, they can make \$100. How do they split the \$100?

Characteristic function: $v(\{1\}) = 20,\ v(\{2\}) = 10,\ v(\{1,2\}) = 100.$

Permutations: $\Pi_2 = \{(1,2),(2,1)\}$.

Hence, the Shapley values are: (writing $\psi_i$ instead of $\psi_i(v)$ for brevity)

$$\psi_1 = \frac{1}{2!}\Big(20 + (100 - 10)\Big) = \frac{1}{2}(20 + 90) = 55,$$

$$\psi_2 = \frac{1}{2!}\Big((100 - 20) + 10\Big) = \frac{1}{2}(80 + 10) = 45.$$

Recall, Shapley value: $\psi_i(v) = \frac{1}{n!}\sum_{\pi\in\Pi_n}\Big(v\big(S_\pi(i)\cup\{i\}\big) - v\big(S_\pi(i)\big)\Big).$

# Shapley value examples (cont'd)

**Example (Glove market).** Three agents are selling gloves in a market. Agent 1 has a left glove, while Agents 2 and 3 have a right glove each. A buyer wishes to purchase a *pair* of gloves for \$1.

Characteristic function:
$$v(\{1\}) = v(\{2\}) = v(\{3\}) = v(\{2,3\}) = 0,$$
$$v(\{1,2\}) = v(\{1,3\}) = v(\{1,2,3\}) = 1,$$

Permutations: $\Pi_3 = \{(1,2,3),\ (1,3,2),\ (2,1,3),\ (2,3,1),\ (3,1,2),\ (3,2,1)\}$

Hence, the Shapley values are:

$$\psi_1 = \frac{1}{3!}\big(0+0+1+1+1+1\big) = \frac{4}{6},$$
$$\psi_2 = \frac{1}{3!}\big(1+0+0+0+0+0\big) = \frac{1}{6},$$
$$\psi_3 = \frac{1}{6}.$$

# Alternative expression for Shapley value

Note that the marginal contributions are the same for certain orderings. For instance in the glove market example, (2, 3, 1) and (3, 2, 1) are the same before 1 appears, and (1, 2, 3) and (1, 3, 2) are the same after 1 appears.

As $v(S_\pi(i) \cup \{i\}) - v(S_\pi(i))$ only depends on the agents in $S_\pi(i)$ and not on the ordering, we can write the Shapley value in the following alternative form:

$$\psi_i(v) = \sum_{S \subseteq [n] \setminus \{i\}} \frac{|S|!\,(n - |S| - 1)!}{n!} \Big( v(S \cup \{i\}) - v(S) \Big).$$

Here, $|S|!$ is the number of ways to order the agents before $i$, and $(n - |S| - 1)!$ is the number of ways to order the agents after $i$.

Note that the number of terms in the sum reduces from $n!$ to $2^n$.

Recall, Shapley value: $\psi_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Big( v\big(S_\pi(i) \cup \{i\}\big) - v\big(S_\pi(i)\big) \Big)$.

# Shapley value examples (cont'd) *(read at home)*

Let us compute the Shapley values for the glove market using the alternative formula.

**Example 1 (Glove market).** Three agents are selling gloves in a market. Agent 1 has a left glove, while Agents 2 and 3 have a right glove each. A buyer wishes to purchase a *pair* of gloves for \$1.

$$\begin{aligned}
\psi_1 &= \frac{1}{3!}\Big(0! \cdot 2! \cdot 0 + 1! \cdot 1! \cdot 1 + 1! \cdot 1! \cdot 1 + 2! \cdot 0! \cdot 1\Big) \\
&= \frac{1}{6}(0 + 1 + 1 + 2) = \frac{4}{6}. \\
\psi_2 &= \frac{1}{3!}\Big(0! \cdot 2! \cdot 0 + 1! \cdot 1! \cdot 1 + 1! \cdot 1! \cdot 0 + 2! \cdot 0! \cdot 0\Big) \\
&= \frac{1}{6}(0 + 1 + 0 + 0) = \frac{1}{6}. \\
\psi_3 &= \frac{1}{6}.
\end{aligned}$$

# Shapley's axioms

We can also motivate the Shapley value via an axiomatic approach. Consider the following 4 axioms, referred to as Shapley's axioms:

**1. Efficiency/feasibility.** The allocation should distribute exactly the total value created by the coalition, *i.e.*, $\sum_{i=1}^{n} \psi_i(v) = v([n])$.

**2. Dummy.** An agent who does not contribute to any coalition gets nothing. That is, if $v(S \cup \{i\}) = v(S)$ for all $S \subseteq [n] \setminus \{i\}$, then $\psi_i(v) = 0$.

**3. Additivity.** In two different games with characteristic functions $u$ and $v$, the sum of payoffs should be the same as in a combined game with characteristic function $u + v$. That is, for all $u, v \in \mathcal{V}$, we have $\psi(u + v) = \psi(u) + \psi(v)$.

**4. Symmetry.** If $i, j$ are 2 agents such that $v(S \cup \{i\}) = v(S \cup \{j\})$ for all $S \subseteq [n] \setminus \{i, j\}$, then $\psi_i(v) = \psi_j(v)$.

# Shapley's axioms (cont'd)

**Theorem.** The Shapley value uniquely satisfies Shapley's four axioms.

**Proof.** We will first show that the Shapley value satisfies the four axioms, and then show uniqueness. For brevity $\Delta_\pi^v(i) = v\big(S_\pi(i) \cup \{i\}\big) - v\big(S_\pi(i)\big)$. Therefore, we can write the Shapley value as

$$\psi_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Big( v\big(S_\pi(i) \cup \{i\}\big) - v\big(S_\pi(i)\big) \Big) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Delta_\pi^v(i).$$

**1. Efficiency.** We will write,

$$\sum_{i=1}^{n} \psi_i(v) = \sum_{i=1}^{n} \left( \frac{1}{n!} \sum_{\pi \in \Pi_n} \Delta_\pi^v(i) \right) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Bigg( \underbrace{\sum_{i=1}^{n} (v(S_\pi(i) \cup \{i\}) - v(S_\pi(i)))}_{=(\star)} \Bigg)$$

$$\overset{(a)}{=} \frac{1}{n!} \sum_{\pi \in \Pi_n} v([n]) = v([n])$$

Here, $(a)$ uses the observation that $(\star)$ is a telescoping sum.

# Shapley's axioms (cont'd)

**2. Dummy.** (We need to show, if $v(S \cup \{i\}) = v(S)$ for all $S \subseteq [n] \setminus \{i\}$, then $\psi_i(v) = 0$.)
If a player does not contribute to any group, $\Delta_\pi^v(i) = v(S_\pi(i) \cup \{i\}) - v(S_\pi(i)) = 0$ for all $\pi$. Therefore

$$\psi_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Delta_\pi^v(i) = 0.$$

**3. Additivity.** (We need to show for all $u, v \in \mathcal{V}$, we have $\psi(u + v) = \psi(u) + \psi(v)$.)
Let $v, u$ be two characteristic functions.

$$\begin{aligned}\Delta_\pi^{v+u}(i) &= (v+u)(S_\pi(i) \cup \{i\}) - (v+u)(S_\pi(i)) \\ &= v(S_\pi(i) \cup \{i\}) + u(S_\pi(i) \cup \{i\}) - \big(v(S_\pi(i)) + u(S_\pi(i))\big) \\ &= \Delta_\pi^v(i) + \Delta_\pi^u(i).\end{aligned}$$

By summing over all permutations $\pi$, we get: $\psi_i(u + v) = \psi_i(u) + \psi_i(v)$.

# Shapley's axioms (cont'd)

**4. Symmetry.** (We need to show that if $i, j \in [n]$ such that $v(S \cup \{i\}) = v(S \cup \{j\})$ for all $S \subseteq [n] \setminus \{i, j\}$, then $\psi_i(v) = \psi_j(v)$.)

Say $i, j$ are 2 agents such that $v(S \cup \{i\}) = v(S \cup \{j\})$ for all $S \subseteq [n] \setminus \{i, j\}$. Recall that in a permutation $\pi$, $S_\pi(i)$ refers to the predecessors of $i$.

For a permutation $\pi$, let $\pi^\star$ be the same as $\pi$ except with $i$ and $j$ switched. First, we will show $\Delta^v_\pi(i) = \Delta^v_{\pi^\star}(j)$ by studying two cases.

**Case 1:** Say $i$ precedes $j$ in $\pi$. Then $S_\pi(i) = S_{\pi^\star}(j)$ and this set contains neither $i$ nor $j$. Therefore, using symmetry of $v$ we have

$$v(S_\pi(i) \cup \{i\}) = v(S_\pi(i) \cup \{j\}) = v(S_{\pi^\star}(j) \cup \{j\}) \tag{$\star$}$$

Therefore,

$$\Delta^v_\pi(i) = v(S_\pi(i) \cup \{i\}) - v(S_\pi(i)) \underbrace{=}_{\text{by } (\star)} v(S_{\pi^\star}(j) \cup \{j\}) - v(S_{\pi^\star}(j)) = \Delta^v_{\pi^\star}(j).$$

# Shapley's axioms (cont'd)

**Case 2:** Say $j$ precedes $i$ in $\pi$. Let us denote $S = S_\pi(i) \setminus \{j\}$. Then,

$$\Delta_\pi^v(i) = v(S_\pi(i) \cup \{i\}) - v(S_\pi(i)) = v(S \cup \{i\} \cup \{j\}) - v(S \cup \{j\}) \qquad (\star\star)$$

In $\pi^\star$, $i$ precedes $j$, so $S_{\pi^\star}(j) = S \cup \{i\}$. Moreover, via symmetry we have $v(S \cup \{j\}) = v(S \cup \{i\}) = v(S_{\pi^\star}(j))$. Hence,

$$\Delta_\pi^v(i) \underbrace{=}_{\text{by } (\star\star)} v(S_{\pi^\star}(j) \cup \{j\}) - v(S_{\pi^\star}(j)) = \Delta_{\pi^\star}^v(j).$$

This concludes the proof of the claim $\Delta_\pi^v(i) = \Delta_{\pi^\star}^v(j)$. We now have:

$$\psi_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Delta_\pi^v(i) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Delta_{\pi^\star}^v(j) \overset{(a)}{=} \frac{1}{n!} \sum_{\pi^\star \in \Pi_n} \Delta_{\pi^\star}^v(j) = \psi_j(v).$$

Here, $(a)$ uses the fact that $\pi \mapsto \pi^\star$ is a bijection.

# Shapley's axioms: proof of uniqueness

It is sufficient to show that there is a unique $\psi$ which satisfies the four axioms. Since the Shapley value satisfies the four axioms, it has to be the one.

For $T \subseteq [n]$, let $\omega_T \in \mathcal{V}$ be the following characteristic function:

$$\omega_T(S) = \mathbb{1}(T \subseteq S) = \begin{cases} 1 & \text{if } S \text{ contains } T, \\ 0 & \text{otherwise.} \end{cases}$$

We will prove uniqueness via two claims.

**Claim 1.** Let $\psi$ be an allocation which satisfies the dummy, symmetry, and efficiency axioms. Then,

*(i)* $\psi_i(\omega_T) = \frac{1}{|T|}\mathbb{1}(i \in T) = \begin{cases} \frac{1}{|T|} & \text{if } i \in T, \\ 0 & \text{otherwise.} \end{cases}$

*(ii)* $\psi_i(c\,\omega_T) = c\,\psi_i(\omega_T)$ for all $c \geq 0$.

# Shapley's axioms: proof of uniqueness (cont'd)

**Claim 2.** For any $v \in \mathcal{V}$, there exists a set of unique coefficients $\{C_T\}_{T\subseteq[n],\, T\neq\varnothing}$ where $C_T \in \mathbb{R}$ such that, *(note $C_T \in \mathbb{R}$ and not $\mathbb{R}_+$)*

$$v(S) = \sum_{\substack{T\subseteq[n]\\ T\neq\varnothing}} C_T\, \omega_T(S) \qquad \forall\, S \subseteq [n].$$

We will first show that uniqueness follows from Claims 1 and 2.

Let $v \in \mathcal{V}$ and let $\{C_T\}_{T\subseteq[n],\, T\neq\varnothing}$ be the unique coefficients from Claim 2. Let $\psi$ be an allocation which satisfies Shapley's axioms. Let us define,

$$T^+ = \{T : C_T \geq 0\}, \qquad T^- = \{T : C_T < 0\},$$
$$v^+(S) = \sum_{T\in T^+} C_T\, \omega_T(S), \qquad v^-(S) = \sum_{T\in T^-} (-C_T)\, \omega_T(S).$$

Then $v^+, v^-$ are valid characteristic functions in $\mathcal{V}$ (*i.e.*, they are non-negative, satisfy $\tilde{v}(\varnothing) = 0$, and are monotone $\tilde{v}(A) \leq \tilde{v}(B)$ for $A \subset B$).

# Shapley's axioms: proof of uniqueness (cont'd)

Moreover, $v + v^- = v^+$. Hence, by the additivity axiom, $\psi_i(v) + \psi_i(v^-) = \psi_i(v^+)$, *i.e.*, $\psi_i(v) = \psi_i(v^+) - \psi_i(v^-)$.

Since $v^+$ is a sum of characteristic functions $C_T\,\omega_T$ with $C_T \geq 0$, the additivity axiom and Claim 1 gives

$$\psi_i(v^+) = \sum_{T \in T^+} \psi_i(C_T \omega_T) = \sum_{T \in T^+} C_T \cdot \frac{\mathbb{1}(i \in T)}{|T|}.$$

Similarly, $\psi_i(v^-) = \sum_{T \in T^-} (-C_T) \cdot \frac{\mathbb{1}(i \in T)}{|T|}$. Therefore,

$$\psi_i(v) = \psi_i(v^+) - \psi_i(v^-) = \sum_{\substack{T \subseteq [n] \\ T \neq \varnothing}} C_T \cdot \frac{\mathbb{1}(i \in T)}{|T|}.$$

By Claim 2, the $\{C_T\}$'s are unique. Hence, so is $\psi_i$. $\square$

# Shapley's axioms: proof of uniqueness (cont'd)

Recall, Claim 1: Let $\psi$ be an allocation which satisfies the dummy, symmetry, and efficiency axioms. Then, *(i)* $\psi_i(\omega_T) = \frac{1}{|T|}\mathbb{1}(i \in T)$, and *(ii)* $\psi_i(c\,\omega_T) = c\,\psi_i(\omega_T)$ for all $c \geq 0$.

**Proof of Claim 1.** First, the dummy axiom implies $\psi_i(\omega_T) = 0$ for any $i \notin T$ as $\omega_T(S \cup \{i\}) = \omega_T(S)$ for any $i \notin T$. By symmetry, $\psi_i(\omega_T) = \psi_j(\omega_T)$ for all $i, j \in T$. Finally, efficiency implies that,

$$\sum_{i=1}^{n} \psi_i(\omega_T) = \sum_{i \in T} \psi_i(\omega_T) = 1 \quad \Longrightarrow \quad \psi_i(\omega_T) = \frac{\mathbb{1}(i \in T)}{|T|} \quad \forall\, i \in [n].$$

This proves statement *(i)*. Statement *(ii)* follows similarly *(read at home)*.

The dummy axiom implies $\psi_i(c\omega_T) = 0$ for any $i \notin T$. By symmetry, $\psi_i(c\,\omega_T) = \psi_j(c\,\omega_T)$ for all $i, j \in T$. Therefore, by efficiency and statement *(i)*:

$$\sum_{i=1}^{n} \psi_i(c\,\omega_T) = \sum_{i \in T} \psi_i(c\,\omega_T) = c \quad \Longrightarrow \quad \psi_i(c\,\omega_T) = \frac{c \cdot \mathbb{1}(i \in T)}{|T|} = c\psi_i(\omega_T). \qquad \square$$

# Shapley's axioms: proof of uniqueness (cont'd)

Recall, Claim 2: For any $v \in \mathcal{V}$, there exists unique coefficients $\{C_T\}_{T\subseteq[n],\, T\neq\varnothing}$ such that $v(S) = \sum_{\substack{T\subseteq[n]\\ T\neq\varnothing}} C_T\,\omega_T(S)$ for all $S \subset [n]$.

**Proof of Claim 2.** Let $T_1, T_2, \ldots, T_{2^n-1}$ be all non-empty subsets of $[n]$, ordered in increasing cardinality, i.e. $|T_j| \leq |T_{j+1}|$ for all $j$.

Let $\mathbf{v} \in \mathbb{R}^{2^n-1}$ where $\mathbf{v}_i = v(T_i)$.
Let $W \in \mathbb{R}^{(2^n-1)\times(2^n-1)}$ where $W_{i,j} = \omega_{T_j}(T_i) = \mathbb{1}(T_j \subseteq T_i)$.

*An example: With $n = 2$, $W$ may look like this:*

$$\begin{array}{c|ccc} & \{1\} & \{2\} & \{1,2\} \\ \hline \{1\} & 1 & 0 & 0 \\ \{2\} & 0 & 1 & 0 \\ \{1,2\} & 1 & 1 & 1 \end{array}$$

*Observation.* $W$ is a lower triangular matrix, with 1's on the diagonal.

# Shapley's axioms: proof of uniqueness (cont'd)

*Proof of observation.* First note that $W_{i,i} = \omega_{T_i}(T_i) = \mathbb{1}(T_i \subseteq T_i) = 1$. For $i < j$, we have $W_{i,j} = \omega_{T_j}(T_i) = \mathbb{1}(T_j \subseteq T_i) = 0$, as $|T_j| \geq |T_i|$ and $T_j \neq T_i$. $\square$

Hence $W$ is invertible. (We will use this fact without proof: lower triangular matrices are invertible if its diagonal entries are nonzero.)

Let us define $C \triangleq W^{-1}\mathbf{v} \in \mathbb{R}^{2^n-1}$ (we will show that $C$ forms the coefficients set out in the claim). Then $\mathbf{v} = WC$. That is, for all $i \in \{1, \ldots, 2^n - 1\}$, we have

$$\mathbf{v}_i = \sum_{j=1}^{2^n-1} W_{i,j}\, C_j, \quad \Longrightarrow \quad v(T_i) = \sum_{j=1}^{2^n-1} \omega_{T_j}(T_i)\, C_j.$$

This is equivalent to $v(S) = \sum_{\substack{T \subseteq [n] \\ T \neq \varnothing}} C_T\, \omega_T(S)$ for all $S \subseteq [n]$.

Uniqueness follows from invertibility of $W$: precisely, there exists a unique $C \in \mathbb{R}^{2^n-1}$ such that $\mathbf{v} = WC$. $\square$

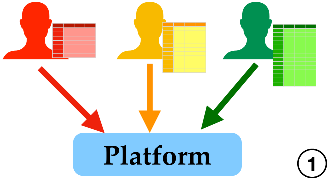
# Quiz: True or False

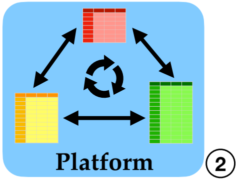
1. The Shapley value is always in the core.
2. The Shapley value is never in the core.
3. If we remove the efficiency axiom, the remaining three axioms (dummy, symmetry, additivity) still uniquely determine an allocation.
4. Recall the glove market example, where Agent 1 has a left glove, and Agents 2 and 3 have a right glove each. A buyer wishes to purchase a *pair* of gloves for \$1.

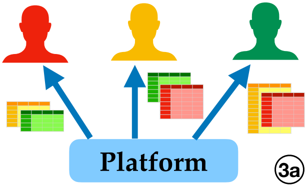
   Suppose now Agent 1 has two left gloves instead of one. Does Agent 1's Shapley value change? If so, does it increase or decrease?

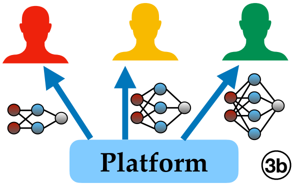
# Ch 8.3: Case study: Cooperative game theory in collaborative ML

**Platforms/consortia for collaborative ML.** Agents contribute data and receive others' data or models trained on others' data in return.
E.g. hospital consortia, federated learning



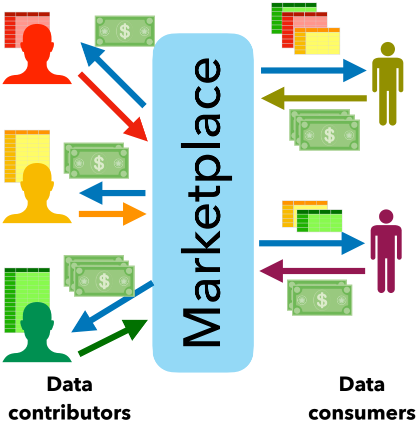






# Cooperative game theory in Collaborative ML (cont'd)

**Data marketplaces.** Sellers contribute data; buyers purchase data and pay the marketplace. The marketplace redistributes the revenue back to the sellers. *Examples:* AWS Data Exchange, Google Ads Data Hub.



# Cooperative game theory in Collaborative ML (cont'd)

In both settings, a central question is: *how do we fairly allocate the benefits (or costs) among agents?*

The literature models this using cooperative game theory in two different situations:

1. *Data has already been collected:* Each agent $i$ has already collected a dataset $D_i$. We wish to *value each agent's data* to determine how much reward (data, money, or model access) she should receive.

   $\longrightarrow$ **Coalitional games & Shapley value**.

2. *Data must be collected (at a cost):* Agent $i$ incurs a per-sample cost $c_i$ to collect data. We wish to *divide the data collection work* fairly, so that all agents benefit from collaboration.

   $\longrightarrow$ **Axiomatic bargaining**.

We will look at each scenario in turn.

# Ch 8.3.1: Valuing data via Shapley value

**Modeling collaborative ML as a coalitional game.** There are $n$ agents, each holding a dataset $D_i$. Let perf$(S)$ denote the performance (*e.g.*, test accuracy) of a model trained on the combined data $\bigcup_{i \in S} D_i$.

Define the characteristic function

$$v(S) = \mathrm{perf}(S), \qquad S \subseteq [n].$$

The **Shapley value** of agent $i$ is:

$$\psi_i(v) = \frac{1}{n!} \sum_{\pi \in \Pi_n} \Big( v\big(S_\pi(i) \cup \{i\}\big) - v\big(S_\pi(i)\big) \Big).$$

# Data Shapley: valuing individual data points

The formulation on the previous slide treats each *agent's dataset* as a player. One can also define a finer-grained game where each individual *data point* is a player. This is the *Data Shapley value* (Ghorbani & Zou, 2019).

**Setup.** Let $z_1, \ldots, z_N$ be the full pool of training points (across all agents), where $N = \sum_{i=1}^{n} |D_i|$. Define $v(S) = \text{perf}(\text{model trained on } \{z_j\}_{j \in S})$ for $S \subseteq [N]$. The Data Shapley value of data point $j$ is:

$$\begin{aligned}
\psi_j^{\mathrm{data}}(v) &= \frac{1}{N!} \sum_{\pi \in \Pi_N} \Big( v\big(S_\pi(j) \cup \{z_j\}\big) - v\big(S_\pi(j)\big) \Big) \\
&= \sum_{S \subseteq [N] \setminus \{j\}} \frac{|S|!\,(N - |S| - 1)!}{N!} \Big( v(S \cup \{z_j\}) - v(S) \Big).
\end{aligned}$$

**Agent-level vs. point-level Shapley values.** The Shapley value of agent $i$ (in the $n$-player game) is generally *not* equal to $\sum_{j \in D_i} \psi_j^{\mathrm{data}}(v)$ (the sum of Data Shapley values of her points in the $N$-player game).

# Data Shapley value vs (regular) Shapley value

Why study the Data Shapley value?

1. Often, another agent may only be interested in subset of an agent's data.
   *E.g. In a marketplace, a buyer may be willing to pay only for subset of a contributor's data that is relevant to their use case.*
2. The Data Shapley value can identify issues with individual corrupted examples, enabling fine-grained data cleaning.

**Quiz.** Suppose we set the allocation for an agent $i$ as,

$$\psi_i(v) = \sum_{j \in D_i} \psi_j^{\mathrm{data}}(v).$$

Which of the four axioms does this satisfy?

# Data Shapley: computation

Recall, data Shapley value:

$$\psi_j^{\mathrm{data}}(v) = \frac{1}{N!}\sum_{\pi\in\Pi_N}\Big(v\big(S_\pi(j)\cup\{z_j\}\big) - v\big(S_\pi(j)\big)\Big) = \sum_{S\subseteq[N]\setminus\{j\}}\frac{|S|!\,(N-|S|-1)!}{N!}\Big(v\big(S\cup\{z_j\}\big) - v(S)\Big).$$

**Computational challenge.** Exact computation requires evaluating $v(S)$ for all $2^N$ subsets, each requiring model retraining. This is infeasible for large datasets.

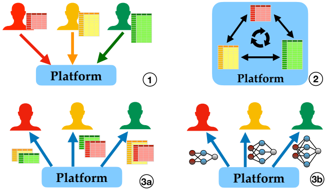
**Monte Carlo approximation.** Sample $T$ random permutations $\pi^{(1)},\dots,\pi^{(T)}$ and approximate:

$$\hat{\psi}_j^{\mathrm{data}}(v) = \frac{1}{T}\sum_{t=1}^{T}\Big(v\big(S_{\pi^{(t)}}(j)\cup\{z_j\}\big) - v\big(S_{\pi^{(t)}}(j)\big)\Big).$$

For each sampled permutation, we train models incrementally as data points are added one by one, requiring $N$ retrainings per permutation.

# Ch 8.3.2: Dividing data collection work via bargaining

Consider a *collaborative learning* setting where agents have *not yet* collected their data. Agent $i$ incurs a cost $c_i$ per data point collected. If they collaborate, all agents benefit from the pooled data. But how do you divide the data collection work?



**Set up.** Agent $i$'s model error when training her model with $n_i$ points is $\mathrm{err}_i(n_i)$. When working on her own, her *penalty* (negative utility) when she collects $n_i$ points is:

$$p_i(n_i) = \mathrm{err}_i(n_i) + c_i n_i$$

Typically, $\mathrm{err}_i$ decreases with $n_i$ while the cost $c_i n_i$ increases. An agent will choose $n_i$ to minimize her penalty. Let $p_i^{\mathrm{IND}} = \min_{n_i} p_i(n_i)$.

# Dividing data collection work via bargaining (cont'd)

**Data sharing.** Suppose there are $m$ agents. Let $n_i$ be the number of data points agent $i$ collects. Each agent's *penalty* (negative utility) is:

$$p_i(n_1, \ldots, n_m) = \mathrm{err}_i\big(\textstyle\sum_{j=1}^m n_j\big) + \ \ c_i\, n_i.$$

Clearly, collaboration can be helpful, as an agent has access to others' data. But, how do we choose $n_1, \ldots, n_m$ *fairly*?

**Minimizing cumulative penalty.** Suppose we decide to minimize the sum of penalties

$$\sum_{i=1}^m p_i(n_1, \ldots, n_m) = \sum_{i=1}^m \mathrm{err}_i\big(\textstyle\sum_{j=1}^m n_j\big) + \ \ \displaystyle\sum_{i=1}^m c_i\, n_i.$$

What is wrong with this approach?

# Modeling data collection as a bargaining problem

We can use axiomatic bargaining to divide the work more fairly. (Clinton et al, 2025)

**The bargaining problem.** Define the *utility* of agent $i$ as,

$$\widetilde{u}_i(n_1, \dots, n_m) \triangleq p_i^{\mathsf{IND}} - p_i(n_1, \dots, n_m).$$

*Individual rationality:* A collection scheme is individually rational (IR) if every agent is at least as well off as working alone: $p_i(n_1, \dots, n_m) \le p_i^{\mathsf{IND}} \ \forall\, i$, *i.e.*, $\widetilde{u}_i(n_1, \dots, n_m) \ge 0$ for all $i$.

*Utility possibility set and disagreement point:* Define the UPS as

$$\mathcal{U} = \big\{(u_1, \dots, u_m) \in \mathbb{R}^m : \text{there exists } (n_1, \dots, n_m) \text{ such that} \\ \forall\, i,\ u_i = \widetilde{u}_i(n_1, \dots, n_m) = p_i^{\mathsf{IND}} - p_i(n_1, \dots, n_m)\big\}.$$

The disagreement point is $d = \mathbf{0}$ (no collaboration means no gain over working individually).

# Modeling data collection as a bargaining problem (cont'd)

You can now maximize your favorite welfare function $W$ over $\mathcal{U}$ subject to the constraint $u \geq d = \mathbf{0}$. That is,

$$f(\mathcal{U}, \mathbf{0}) = \underset{u \in \mathcal{U};\; u \geq \mathbf{0}}{\operatorname{argmax}}\, W(u_1, \ldots, u_m) \qquad = (u_1^\star, \ldots, u_m^\star) \text{ (say)}$$

Once we do this, we can find the corresponding division of work $n_1^\star, \ldots, n_m^\star$ that gives rise to $(u_1^\star, \ldots, u_m^\star)$.