Logistics

Quick note: My name is pronounced as “ling thô”

- **Timing**
  - Six weeks: last day of class is April 26.
  - Generals: May 5, four-hour long exam for microeconomic theory.
  - Review session: during reading period to be scheduled.

- **Generals preparation**
  - Start now: It is time to prioritize practicing for the generals until you feel comfortable
  - Old generals: [http://www.economics.harvard.edu/graduate/exams](http://www.economics.harvard.edu/graduate/exams)
  - Office hours: Jerry’s and mine
  - Questions: encouraged, and will be answered.

- **Sections**
  - Materials: For this part, our sections will be geared towards preparing for the generals. The lectures will go into more detail on proofs, which are important to understand, but in section we will focus on concepts and problem solving.
  - The first four section notes are re-written to clarify frameworks and concepts, together with examples. The fifth section on auctions will focus on problem solving.
  - Your feedback: Any comments that you’d like to share on this course are very valuable. Letting us know earlier will give us time to adjust to make the course better for you.

- **Problem sets**
  - Due: in class or mailbox by end of due date.
  - Grading: check plus/check/check minus. I will grade whatever you submit, but you don’t have to submit every question. You can get a check plus even if you submit only one problem. Problem sets are optional.
  - Grading: It’s really for feedback. If you spend time thinking about a problem and would like feedback, then you are encouraged to submit. You should not feel the need to copy answers from previous years’ solutions. Do as much or as little as you find helpful, not worrying about your grade on the problem set but rather focusing on how you can pass the generals. The purpose is for you to learn the material and understand it so that you can pass the generals.
  - Group work: encouraged, and please indicate who you work with.
1 What is Social Choice?

1.1 A definition

“Social choice theory, pioneered in its modern form by Arrow (1951), is concerned with the relation between individuals and the society. In particular, it deals with the aggregation of individual interests, or judgements, or well-beings, into some aggregate notion of social welfare, social judgement or social choice.” (Sen 2008, The New Palgrave Dictionary of Economics)

1.2 The basic components

How may we trade off a person well-being for that of another in the interests of society as a whole? In trying to make such trade-offs, does intensity of preference matter? If we think it does, other questions enter the picture. Can intensity of preference be known? Can people tell us how strongly they feel about different alternatives? Can different people’s intense desires be compared so that a balancing of gains and losses can be achieved?

Here, we start with a basic framework in which people consider binary comparisons with a complete and transitive preference relation.

- A finite set $X$ consisting of $n$ alternatives or choices: $X = \{1, 2, \ldots, n\}$
- A “society” $I$ consisting of $I$ individuals: $I = \{1, 2, \ldots, I\}$
- For each individual $i$, there is a weak preference relation $\succeq_i$ such that $a \succeq_i b$ if $a$ is weakly preferred to $b$, for any $a, b \in X$

We may be interested in the problem of eliciting these preferences. This will be discussed in mechanism design, a later part of the course. Here, we are only concerned with a problem of aggregation. We will start with the case in which we know (or utilize) less information about individual preferences (majority voting), to the case with full information (Arrow’s Theorem).

1.3 Types of solution concepts

- Point-valued solution concepts: a function that will give us a unique social choice given the available information on the underlying preferences that is optimal in some sense.
- Set-valued solution concepts: a procedure to select a set of possible choices that satisfy some desirable properties.
- Order-valued solution concepts: a social ranking of all choices.
2 Majority Tournament

2.1 Definition

Full information on the underlying preferences over all alternatives of the individuals in society is not available. However, between any two different alternatives, we do know if the majority of the population prefer one over the other. Implicitly, we are only considering populations with an odd number of individuals, so that there cannot be any tie between any two alternatives.

Definition. A tournament is a complete, asymmetric, and irreflexive binary relation on $X$. Denote the tournament relation as $T$:

- Complete: For any two alternatives $x$ and $y$ in $X$, either $xTy$ or $y Tx$.
- Asymmetric: If $xTy$ then not $yTx$.
- Irreflexive: For any $x \in X$, we cannot have $xTx$.

2.2 McGarvey’s Theorem: Any tournament is supported by some population

Theorem (McGarvey, 1953). Given an arbitrary complete, asymmetric, irreflexive binary relation $T$ on a set $X$ with $|X| = n$, there exists a society of $I = n(n-1)$ individuals with rational preferences over $X$ that induces $T$ as the tournament outcome.

Proof. McGarvey’s proof is constructive: for each of the $\frac{n(n-1)}{2}$ pairs $(x, y) \in T$, add two individuals to the society with preferences given by

$$\succeq_1: \quad x \succeq y \succeq 1 \succeq 2 \succeq \cdots \succeq n$$

$$\succeq_2: \quad n \succeq n-1 \succeq \cdots \succeq 1 \succeq x \succeq y.$$

Therefore, in the pairwise majority vote comparison, $x$ is preferred to $y$ by exactly $\frac{n(n-1)}{2} + 1$ individuals and $y$ is preferred to $x$ by exactly $\frac{n(n-1)}{2} - 1$ individuals, so that $xTy$, as desired. □

2.3 Condorcet Winner

Definition. Alternative $x \in X$ is a Condorcet winner if for all $y \in X$, we have $xTy$.

If a Condorcet winner exists then it is unique. We would like the Condorcet winner to be our solution if it exists (Condorcet consistency) but unfortunately one may not exist. In fact, it is very likely to be the case when there is a large number of alternatives, and the probability of not having a Condorcet winner is increasing in the number of voters (individuals) in the population.
2.4 The Top Cycle

The Top Cycle is the first solution concept of majority tournament you saw in class and the most lenient of all that we will study in the course.

**Definition.** For tournament $T$ on $X$,

$$TC(T) = \{ x \in X \mid \forall y \in X, \exists (x_1, \ldots, x_k) \text{ s.t. } x_1 = x, x_k = y, \text{ and } x \rightarrow Tx_{i+1} \}. $$

Heuristically the TC implies that any alternative that has indirect dominance over all elements is a permissible solution element.

**Theorem (Top Cycle Theorem).** For any $x \in TC(T)$ and $y \notin TC(T)$, $xTy$.

**Theorem (Top Cycle Equivalence).** Define $\mathcal{X}$ as the set of subsets of $X$ such that each set $Y \in \mathcal{X}$ satisfies the inside-beat-outside (IBO) property: $\forall x \in Y, \forall y \notin Y, xTy$.

The elements of $\mathcal{X}$ are nested sets and $TC(T)$ is the smallest non-empty set in $\mathcal{X}$.

**Theorem (Top Cycle Name).** The elements of the top cycle $TC(T)$ can be arranged in an order such that each element beats the next element, and the last element beats the first one, forming a cycle.

2.5 Example

**Problem (Spring 2007 — D1 — (a) and (b)).** Alice (A), Bob (B) and Charlie (C) are friends. They are such good friends that they are contemplating marriage. Specifically, Alice and Bob could get married ($x$), or Alice and Charlie could get married ($y$), or they could all remain single ($z$).

Alice prefers marrying Bob to marrying Charlie, and prefers marrying Charlie to remaining single. Bob prefers that everyone remain single (he wants to continue dating Alice), but he prefers marrying Alice himself to seeing Alice marry Charlie.

Charlie’s favorite alternative is to marry Alice himself. Because he values other’s happiness he would rather see Bob marry Alice, his second place outcome, than have everyone remain single.

a) Alice, Bob and Charlie decide to use majority rule to determine the outcome. Is there a Condorcet winner in the set $\{x, y, z\}$?

b) What is the top cycle of a majority relation? What is the top cycle in this example?
3 Properties of Solution Concepts

Given a solution concept, there are two categories of questions we might ask about its behavior. Normatively, we’re interested in solution concepts that reflect society preferences in “good” ways; in this regard, we have introduced a number of properties of arguable desirability. Positively, we’re interested in solution concepts that have strong predictive power for observed collective decision-making processes.

Normative properties can be partitioned into those that place restrictions on single tournaments and those that place cross-tournament restrictions. Below is a list of properties that we consider in this class.

This may appear similar to the equilibrium refinement approach in game theory that you have seen. However, keep in mind that by adding one property we may lose another, and it is not always straightforward how solution concepts are nested within each other. As solution concepts are constructed to satisfy certain criteria, we still need to check if they satisfy each individual property.

The properties below are stated for set-valued solution concepts, and you can think through how they may be adapted for order-valued solution concepts.

- Single-Tournament Properties
  - **Anonymity:** The solution concept is indifferent to the names of the individuals.
  - **Neutrality:** The solution concept is indifferent to the names of the alternatives.
  - **Condorcet Consistency:** If a Condorcet winner exists, it is uniquely selected by the solution concept.
  - **Pareto Efficiency:** The solution concept does not select any Pareto dominated alternative: if every individual prefers another alternative to the chosen one, then it is not selected by the solution concept.

  **Note:** For Pareto efficiency, we need to know if any possible underlying preference profile includes a unanimous ranking between two candidates. The proof that the Uncovered Set satisfies Pareto efficiency cleverly links unanimous vote results to majority vote results.

- Cross-Tournament Properties
  - **Monotonicity:** If alternative $x$ beats at least as many alternatives in pairwise majority voting at $T'$ as at $T$, and $x$ is selected by the solution concept at $T$, it is also selected by the solution concept at $T'$.

  **Note:** In the context when the full preference profile is known, monotonicity is stated as follows: A chosen alternative will still be chosen when it rises in individual preference rankings (while leaving everything else unchanged).
– *Independence of Losers*: The solution concept is indifferent to changes in the relationships among losers, when winners can only beat more losers and the relationships among winners are unchanged.

– *Composition Consistency*: If the solution concept selects the same set of winners for two different sets of individuals, it should select the same set when the two societies are pooled.

Note: This property is a more general property beyond majority tournaments. When restricting to majority tournaments, it is equivalent to independence of clones.

• Cross-Feasible-Set Properties

– *Independence of Clones*: The solution concept is indifferent to making “arbitrary” duplications of alternatives, or “merging” similar alternatives.

4 Majority Tournament Solutions

Anonymity, Neutrality, Condorcet Consistency are the “easier” properties that all the solution concepts below satisfy.

4.1 Set-valued solution concepts

• Top Cycle $TC(T)$: Definition above.

Properties: Monotonicity, Independence of Losers

• Uncovered Set $UC(T)$: $UC(T) = \{x \in X \mid \exists y \in X \text{ such that } yCx \text{ if } yTx \text{ and for all } z \in X, \text{ if } xTz \text{ then } yTz \}.$

Properties: Pareto Efficiency, Monotonicity, Independence of Clones

• Banks Set $B(T)$: $B(T)$ is the set of leading elements of maximal transitive chains in $T$ by set inclusion.

Alternately: $B(T)$ is the set of leading elements of maximal transitive chains in $T$ by pre-appending.

Properties: Pareto Efficiency, Monotonicity, Independence of Clones

• Bipartisan Set $BP(T)$: The set of $x \in X$ that are played with positive probability in the unique mixed equilibrium of the two-player zero-sum game $(X, X, g(T))$ where $g(T)$ is 1 if $xTy$, $-1$ if $yTx$ and 0 otherwise.

Properties: Pareto Efficiency, Monotonicity, Independence of Losers
- (Optional) Copeland’s Rule: The set of \( x \in X \) that wins against the most number of other alternatives.

Properties: Anonymity, Neutrality, Condorcet Consistency, Monotonicity

- Tournament Equilibrium Set \( TEQ(T) \): (Defined recursively) \( TEQ \) is the smallest subset of alternatives from which we can be sure that the contestation process will not escape.

Properties: Pareto Efficiency, Independence of Clones

**Note:** Computing the \( TEQ \) is NP-hard and (remains NP-hard even for 9 voters) due to its recursive nature and checking its axiomatic properties is very difficult. It has been checked computationally that the \( TEQ \) satisfies Monotonicity and Independence of Losers whenever the number of alternatives is fewer than 13.

A counterexample with 24 alternatives was found in 2013 with the help of a computer. In general, counterexamples such that the \( TEQ \) does not satisfy these other nice properties are extremely rare.

**Nested set-valued solution concepts:**

\[
TC(T) \supseteq UC(T) \supseteq B(T) \supseteq TEQ(T)
\]

\[
TC(T) \supseteq UC(T) \supseteq BP(T)
\]

### 4.2 Order-valued solution concepts

- Slater Order: The set of all orders that result from a minimal number of pairwise majority outcome reversals to obtain a transitive relation.

Properties: Pareto Efficiency

### 4.3 Example

**Problem** (Spring 2010 — D1). Consider the relation \( T \) depicted below, in which \( xTy \) if and only if \( x \) has a strict majority over \( y \) in a binary vote; \( xTy \) is denoted by a 1 in the \((x, y)\) cell of the table.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>—</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>—</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>—</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>—</td>
</tr>
</tbody>
</table>
a) What is the top cycle of $T$? Define it and compute it for this example.

b) Why might we be interested in the top cycle of the majority relation? (i.e. state any interesting theorem or property that applies to it)

c) What is the uncovered set of $T$? Define it and compute it for this example.

d) Why might we be interested in the uncovered set of the majority relation? (i.e. state any interesting theorem or property that applies to it)

e) (Extra) Show that the uncovered set does not satisfy Independence of Losers.

f) (Extra) What is (are) the Slater order(s)?

5 Tournament Solutions from Voting Equilibria

We are working with majority tournaments in which we know which candidate is the winner by a majority vote between any two candidates. A natural way to select a winning candidate is to get the candidates to be voted against a current winner in a sequential binary voting procedure. The order of the sequence may determine the winner, but a candidate who does not win given any sequence should not be part of a (set-valued) solution. This is the logic behind solution concepts from voting equilibria.

Formally, given an order of candidates, or an agenda $A = (x_1, x_2, \ldots, x_n)$ ($x_k$ is a member of $X$), the process is as follows:

1. $x_1^* = x_1$

2. For $2 \leq i \leq n$: $x_i^*$ is the winner between $x_{i-1}^*$ and $x_i$

5.1 Non-strategic voters: Forward proceed

In lecture 1, we already consider the case when voters are not strategic. They do not anticipate what will happen down the line, and therefore always truthfully pick their preferred candidate between the two candidates presented to them in each round, resulting in the majority tournament relation being the determination of the winner in each round.

Formally, the voting process is now decided as follows:

1. $x_1^* = x_1$

2. For $i = 2, 3, \ldots, n$:

$$x_i^* = \begin{cases} x_i, & \text{if } x_i T x_{i-1}^* \\ x_{i-1}^*, & \text{if } x_{i-1}^* T x_i \end{cases}$$
The final winner is $x_n$.

From lecture 1, we also know that any member of the Top Cycle can be elected in a non-strategic sequential voting procedure, and only the members of the Top Cycle can be elected in such a procedure. This is an alternative definition for the Top Cycle solution concept.

5.2 Strategic voters: Backward induction

Consider the case in which the current winner $x_{i-1}^*$ is pitted against $x_i$ with $x_{i-1}^*Tx_i$ but $x_{i-1}^*$ will ultimately lose to $x_j$ with $j > i$. If $x_iTx_k$ for all $k > i$, then the voters who prefer $x_i$ to $x_j$ can form a majority to strategically vote for $x_i$ against $x_{i-1}^*$, overturning the tournament result $x_{i-1}^*Tx_i$ and elect $x_i$ instead of $x_j$. Like with subgame perfect equilibrium in game theory, we only have to consider one-step deviations in finding the strategic voting equilibria.

This is the logic behind the theorem by Shepsle and Weingast (1984), and it suggests that, as with subgame perfect equilibrium, we can find the (unique) strategic voting equilibrium in a backward-induction process.

**Theorem** (Shepsle and Weingast (1984)). Given an agenda $A = (x_1, x_2, \ldots, x_n)$, the sophisticated voting outcome can be found as follows:

1. $x_n^* = x_n$
2. For $i = n - 1, n - 2, \ldots, 1$:

$$ x_i^* = \begin{cases} 
  x_i, & \text{if } x_iTx_j^*, \forall j > i \\
  x_{i+1}^*, & \text{otherwise} 
\end{cases} $$

The final winner is $x_1^*$.

The Banks theorem uses the same intuition to classify all possible winners in a strategic voting equilibrium by considering transitive chains, in which an element wins against all elements following it:

**Theorem.** Given tournament $T$ on alternatives $X$, the **Banks set** $B(T)$ of all possible sophisticated voting outcomes for some agenda $A$ of $X$ is equivalent to the set of the leading elements of the maximal transitive chains.

Furthermore, $B(T) \subseteq UC(T)$.

Banks defines maximal transitive chain in the set inclusion sense, but the intuition given here and in the proof in Lecture 2 suggest that we can consider maximal transitive chains in the pre-appending sense: We should not be able to add a leading element to the chain and still have a transitive chain. The two definitions are equivalent. You can use either one to help find the Banks set given a tournament relation.
5.3 Example

**Problem** (Adapted from Spring 2010 — D1). Consider the relation $T$ depicted below, in which $xTy$ if and only if $x$ has a strict majority over $y$ in a binary vote; $xTy$ is denoted by a 1 in the $(x, y)$ cell of the table.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

a) Who is the winner in a strategic voting equilibrium when the agenda is the sequence of candidates 54321, in that order.

b) What is the Banks set of $T$? Define it and compute it for this example.

c) Why might we be interested in the Banks set of the majority relation? (i.e. state any interesting theorem or property that applies to it)

d) Argue that candidate number 4 cannot be selected in a strong voting equilibrium without appealing to the Banks Theorem.
Note on Core and Stability in Social Choice

In the first part of Social Choice theory, we learn about solution concepts that use information from the majority tournament relation $T$. Some of these solution concepts are in turn built upon “stronger” relations, such as the covering relation ($C$, in the Uncovered Set) and contestation ($D$, in the TEQ). The idea behind these relations is to make it harder to overturn the status quo (the currently chosen alternative), making the chosen alternatives in some sense more stable.

The majority tournament relation by itself is quite weak: it is complete, and between every two alternatives, if we go by the majority relation then one can always overturn the other.

Given an overturning rule, an alternative is in the Core if it cannot be overturned by any other alternative. The overturning rule—or as called in the lecture, a voting game (given that rule)—is core-stable if for any profile of preferences of the individuals, there is always some alternative (perhaps more than one) that is in the core of that voting game.

If we use majority relation as the overturning rule, then the only time when there is an alternative that is in the Core is when there is an alternative that beats everyone else using the majority relation, i.e. the Condorcet winner. This does not happen very often. If on the other hand we use a very strong overturning rule, such as requiring that everyone has to prefer another alternative to overturn the existing one, then there are many alternatives in the Core given any preference profile. Note that if the rule requires more than majority to overturn something, then this new relation is not complete: between two alternatives, neither may overturn the other.

The idea for the Nakamura theorem is that, to achieve core-stability given some number of alternatives, you want a strong enough overturning rule so that there cannot be a situation when every alternative can be overturned. How strong the rule has to be depends on the number of alternatives. You want to avoid a situation with the overturning process going in a cycle around all the alternatives.

Any collection of winning coalitions (a winning coalition is a set of people who together can overturn an alternative if they have the right preferences) can have a non-empty intersection (i.e. at least one person being in all those coalitions). You are more likely to have an empty intersection with more winning coalitions. Let $v$ be the fewest number of winning coalitions needed to have an empty intersection. Note that $v$ is bigger with a stronger overturning rule.

Theorem (Nakamura theorem). A strong neutral voting game with $n$ alternatives is core-stable if and only if $v > n$.

The situation with the overturning process going in a cycle around all the alternatives is avoided because, if it happens, then the $n$ winning coalitions around the cycle have at least one shared member (since $n < v$), and this member will have a cyclical preference which is not rational, a contradiction. Recall that individuals are always assumed to be rational in all of our settings. It is
society’s job to aggregate these rational but conflicting preferences. This is not a full proof of the theorem, but a sketch of the main idea.

2 Social Choice with Vote Counts

In majority tournaments, you know that more voters prefer one candidate to another, but you do not know the vote counts or strength of majorities. In this section, we consider the case in which we do know the number of voters who prefer one candidate to another. A tournament table with binary information between any pair of candidates can be replaced with a vote count table in which each cell is the number of voters who prefers the row candidate to the column candidate. The total number of votes for each pair is the number of voters \( n \).

All of the solution concepts that we study in more depth below are order-valued solutions.

2.1 Properties of Solution Concepts: Order-valued

In section 1, we state the properties below for set-valued solution concepts. Here we adapt them for order-valued solution concepts.

- Single-Tournament Properties
  - Anonymity: The solution concept is indifferent to the names of the individuals.
  - Neutrality: The solution concept is indifferent to the names of the alternatives.
  - Condorcet Consistency: Whenever there is a transitive order of all alternatives under majority rule, the solution should be that transitive order.
  - Pareto Efficiency: If alternative \( x \) is preferred to alternative \( y \) by every individual, alternative \( x \) is preferred to alternative \( y \) in every order selected by the solution concept.

- Cross-Tournament Properties
  - Monotonicity: If alternative \( x \) beats at least as many alternatives in pairwise majority voting at \( T' \) as at \( T \), then \( x \) is preferred to at least as many alternatives at any ordering chosen by the solution concept at \( T' \) than at \( T \).
  - Composition Consistency: If the solution concept selects the same set of orders for two different sets of individuals, it should select the same set when the two societies are pooled.

- Cross-Feasible-Set Properties
  - Independence of Clones: The solution concept is indifferent to making “arbitrary” duplications of alternatives, or “merging” similar alternatives.
2.2 Solutions

Anonymity, Neutrality, Pareto Efficiency, Condorcet Consistency, and Monotonicity are the “easier” properties that all the solutions below satisfy.

- Kemeny Solution: The orders that result from the least number of switches in individual pairwise votes to obtain a transitive order in majority relation.
  
  Properties: Composition Consistency

- Method of Ranked Pairs: Rank all pairs of alternatives with a positive vote margin in decreasing order. Sequentially form a binary relation $R$ by adding in $xRy$ in decreasing order of their vote margin, as long as this does not create a cycle. If you encounter a pair that would create a cycle, simply skip this pair and continue down the list. The resulting ranking order is unique.

  Properties: Independence of Clones

- Schulze Method: For each pair of alternatives $x$ and $y$, a possible path of majority victories from $x$ to $y$ has its strength defined as the weakest victory along that path. Define the max-min strength from $x$ to $y$ as the maximal strength across all possible paths of majority victories from $x$ to $y$. Define a binary relation $B$ by $xBy$ if max-min strength from $x$ to $y$ is higher than max-min strength from $y$ to $x$. This binary relation will produce a unique transitive order, which is our solution.

  Properties: Independence of Clones

**Theorem** (Young-Levenglick). The Kemeny Solution is the only anonymous, neutral order-valued solution that is both composition consistent and Condorcet consistent.

2.3 Example

**Problem** (Spring 2013 — D1). This entire question concerns voting problems in which there are three alternatives $x, y, z$.

Suppose the binary votes between pairs of alternatives (the fraction of the population voting for the alternative in the row over the alternative in the column) are given by the following table:

<table>
<thead>
<tr>
<th></th>
<th>$x$</th>
<th>$y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>—</td>
<td>$\alpha$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>$y$</td>
<td>$1 - \alpha$</td>
<td>—</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>$z$</td>
<td>$1 - \beta$</td>
<td>$1 - \gamma$</td>
<td>—</td>
</tr>
</tbody>
</table>

Assume that $\alpha > \gamma > 1 - \beta > 1/2$. 

1 April 2016 Section 2 3 of 11
a) Prove that the Kemeny Method, the Method of Ranked Pairs, and the Schulze Method (also called the beat-path method) all lead to the same ranking of the alternatives.

b) Show that there are three Slater rankings, and that three of the six strict orders are not Slater rankings.

c) Give an example of a population of voters — a distribution of preferences over the six possible strict orders for three alternatives — that induces a binary vote table satisfying the assumption above and for which the ranking produced by Plurality with a Runoff differs from the common recommendation of the Kemeny, Ranked Pairs and Schulze Methods. (By the ranking produced by Plurality with a Runoff, I mean the order that puts the alternative with the fewest first-place votes last and ranks the remaining two alternatives first and second, respectively, in the same way as the majority vote between them.)

d) Is it necessarily the case that Plurality with a Runoff produces one of the three Slater rankings? Provide a proof or a counterexample.

3 Methods based on Full Rank Information

A common rule is Plurality Rule: Count the number of first-place votes (the number of people with an alternative on the top of their personal preference ranking) and the winner is the one with the most number of first-place votes.

There are several modified versions of Plurality Rule listed below. Each of them can be modified to produce either a point-valued solution or an order-valued solution by repeating the same procedure after getting a winner and repeat with the rest of the alternatives.

- Plurality with runoff: If no candidate has a majority of first-place votes (more than 50%), the top two are selected for a second round runoff where the winner is determined by direct competition: who has more support (coinciding with the $T$ relation).

- Alternative vote: If no candidate has a majority of first-place votes, drop the one with the lowest vote total and try a plurality vote again.

- Borda count: One round of voting with points assigned for each rank: $n, n-1, \ldots, 1$.

All of these methods do not satisfy Condorcet Consistency and Independence of Irrelevant Alternatives (defined below in Arrow’s theorem). Example 3.2 — Spring 2012 question D1 goes over these properties.
4 Methods based on Full Rank Information: Arrow’s Theorem

Balancing conflicting preferences becomes more difficult when we have more information on individuals’ preferences. In fact, when we take into account the full preference profiles of individuals, we have a negative result.

4.1 Arrow’s Impossibility Result

Recall the basic components in Section 1:

- A finite set $X$ consisting of $n$ alternatives or choices: $X = \{1, 2, \ldots, n\}$
- A “society” $I$ consisting of $I$ individuals: $I = \{1, 2, \ldots, I\}$
- For each individual $i$, there is a weak preference relation $\succeq_i$ such that $a \succeq_i b$ if $a$ is weakly preferred to $b$, for any $a, b \in X$. Denote a preference profile by $\succeq = (\succeq_1, \ldots, \succeq_I)$.

Denote $\mathcal{R}$ as the set of all weak orders on $X$ and $\mathcal{A} \subseteq \mathcal{R}^I$ the domain of preferences of all $I$ individuals. A Social Welfare Functional (SWF) is $F: \mathcal{A} \rightarrow \mathcal{R}$. Now the SWF takes into account all individual preferences $\succeq$, whereas in majority tournaments the solution only takes into account the binary relation $T$. SWF produces a unique order, not just a set of possible winners.

Definition (Arrow’s four axioms).

- **Unrestricted domain (UD)**: Any possible profile of weak order is allowed
  \[ \mathcal{A} = \mathcal{R}^I \]
- **Pareto (P)**: Respecting (strict) Pareto domination throughout the order
  For any $\{x, y\} \subseteq X$, $x \succ_i y$ for all $i$ implies $xF_P(\succeq)y$, where the subscript $P$ denotes a strict preference.
  **Note**: This is the same property as Pareto Efficiency above.
- **Independence of Irrelevant Alternatives (IIA)** (or pairwise independence): The SWF cannot reverse the ranking between a pair of alternatives unless some individual reverses her ranking between that particular pair of alternatives.
  For any $\{x, y\}$ and any two profiles $\succeq = (\succeq_1, \ldots, \succeq_I)$, $\succeq' = (\succeq'_1, \ldots, \succeq'_I)$ such that $x \succeq_i y \iff x \succeq'_i y$ for all $i$, then $xF_P(\succeq)y \iff xF_P(\succeq')y$.
  **Note**: IIA is a very strong condition. In particular, methods that satisfy Condorcet Consistency will not satisfy IIA (see proof below).
- **Non-dictatorship (ND)**: No individual decides the social order
  There exists no $h \in \{1, \ldots, I\}$ such that $x \succ_h y$ implies $xF_P(\succeq)y$
Any voting method that satisfies Condorcet Consistency does not satisfy IIA:

Why is this statement true? Construct a counterexample of 3 people and 3 alternatives as follows:

Person 1: $x > y > z$

Person 2: $y > z > x$

Person 3: $z > x > y$

Consider an arbitrary rule that satisfies Condorcet Consistency. There can only be 3 (symmetric) cases:

- If this rule puts $x$ on top: The social ordering has $x$ on top of $z$. Now move $y$ to the bottom of everyone’s preference ranking. People who preferred $x$ to $z$ still prefer $x$ to $z$, and people who preferred $z$ to $x$ still prefer $z$ to $x$. But now 2 people have preference ranking $z > x > y$ and 1 person has preference ranking $x > z > y$. Now $z$ is the Condorcet winner and $xyz$ is a transitive order. To satisfy Condorcet Consistency, $z$ should now be put before $x$ in the new social ordering, violating IIA.

- If this rule puts $y$ on top: The social ordering has $y$ on top of $x$. Now move $z$ to the bottom of everyone’s preference ranking. People who preferred $x$ to $y$ still prefer $x$ to $y$, and people who preferred $y$ to $x$ still prefer $y$ to $x$. But now 2 people have preference ranking $x > y > z$ and 1 person has preference ranking $y > x > z$. Now $x$ is the Condorcet winner and $xyz$ is a transitive order. To satisfy Condorcet Consistency, $x$ should now be put before $y$ in the new social ordering, violating IIA.

- If this rule puts $z$ on top: The social ordering has $z$ on top of $y$. Now move $x$ to the bottom of everyone’s preference ranking. People who preferred $z$ to $y$ still prefer $z$ to $y$, and people who preferred $y$ to $z$ still prefer $y$ to $z$. But now 2 people have preference ranking $y > z > x$ and 1 person has preference ranking $z > y > x$. Now $y$ is the Condorcet winner and $yzx$ is a transitive order. To satisfy Condorcet Consistency, $y$ should now be put before $z$ in the social ordering, violating IIA.

**Theorem** (Arrow’s Theorem). Under the conditions UD, P, IIA, and ND, there does not exist any $F: A \rightarrow R$.

Any SWF that satisfies three out of the four axioms must violate the fourth one. In particular, any SWF that satisfies UD, P, and IIA will be dictatorial. However, since P and ND are normatively very desirable, we might want to give up either of the other two axioms to avoid the impossibility result. One of the problem set questions will ask you to restrict $A$ to specific single-peaked preferences and deduce a P, IIA, and ND social welfare function. More importantly, Jerry very much likes to test your understanding of IIA and why/where it fails for specific SWFs.
4.2 Example

**Problem** (Spring 2012 — D1). This question is about voting rules that are “order-valued” — that is, for any profile of individuals’ preferences, they produce an ordering of all the alternatives. You can consider only strict orders, both in the domain and in the range, and disregard the possibility of ties in voting (due to the usual large-numbers considerations).

a) Give an example of a voting system that satisfies Arrow’s IIA (Pairwise Independence, the term used in MWG) but not Condorcet Consistency.

b) Give an example of a voting system that satisfies Condorcet Consistency but not Arrow’s IIA.

c) Give an example of a voting system that satisfies neither Condorcet Consistency nor Arrow’s IIA but is still reasonably well known and is in use. Why do you think this voting system is used? Does it have some other property that you think is valuable?

d) On Sunday France will hold an election for President — the second stage of a two-stage procedure known as Plurality with a Runoff. This system is described by a first stage in which everyone votes for their favorite candidate and a second stage in which the two highest vote getters from the first stage compete in a majority rule election. For the purposes of this question assume that everyone votes honestly in both stages. Plurality with a Runoff is designed to select a winner, not to produce an ordering. In this question we consider only the three-candidate case (whereas in France there were actually many candidates in the first stage). In the three-candidate case Plurality with a Runoff determines an ordering as follows: The candidate eliminated in the first stage is deemed to be last in the ordering and two surviving candidates are ordered in accordance with the outcome of the second-stage vote. Does this procedure satisfy Arrow’s IIA? How about Condorcet Consistency?

e) Assume that in France the original candidates are Hollande (left), Sarkozy (center) and many right-wing candidates (Le Pen, . . . ). (I know very little about French politics so this is purely hypothetical, for the purposes of this question.) Assume that the right-wing candidates are a set of clones, as defined in Lecture 3. What is the definition of “Independence of Clones” presented in that Lecture? Does the system of Plurality with a Runoff satisfy this definition?

5 Social Welfare Function with Cardinal Information

5.1 General Discussion

Given that Arrow’s Theorem only concerned ordinal preferences, perhaps we can circumvent the negative result by also accounting for cardinal information. More specifically, let $\mathcal{U}$ be the set of all possible utility functions on $X$. $\mathcal{R}$ is still the set of all preference relations on $X$. Redefine the
social welfare functional $F : \mathcal{U}^I \rightarrow \mathcal{R}$. We can then restate Arrow’s conditions, concerning utility rather than ordinal comparisons (see MWG Definitions 22.D.1–3) and get a positive result (see MWG Proposition 22.D.1). (The MWG definitions and proposition are not reproduced here since they are completely optional, the notation is heavy but the intuition is completely parallel to the existing situation when the social choice function maps individuals’ ordinal preference rankings instead of individuals’ utilities over alternatives to a social order. The difference is that, if there is no restriction on how $F$ should depend on transformations of $u$, then we do not have an impossibility result as in Arrow’s Theorem.)

However, for both normative and positive reasons, we may want to restrict the extent to which $F$ can depend on information contained in $u = (u_1, \ldots, u_I) \in \mathcal{U}^I$. In other words, certain changes in $u$ should leave $F(u)$ unchanged. We will express such changes by a set of person-specific transformations $g(u) = (g_1(u_1), \ldots, g_I(u_I))$ and require $\forall u \in \mathcal{U}^I, F(g \circ u) = F(u)$.

- All $g_i$’s are strictly increasing.
  That is, $g$ only preserves ordinal preferences and thus we are back at Arrow’s Theorem with preferences induced by intra-personal utility ranking.

- The set of positive affine transformations, i.e. $g_i(u_i) = \alpha_i u_i + \beta_i$, where $\alpha_i > 0$.
  That is, we allow $F$ to depend individuals’ risk attitudes but no inter-personal comparisons. Nonetheless, this risk-preference information does still not admit non-dictatorial SWF’s.

- $g(u) = (\phi(u_1), \ldots, \phi(u_I))$, where $\phi$ is strictly increasing.
  That is, we can rank order the welfare of people across outcomes, and within outcomes across the population. Indeed, this restriction on $g$ frees up $F$ sufficiently to weaken the dictatorial result of Arrow’s theorem; under this interpersonal comparability of utility any Paretian, pairwise-independent (or IIA) $F$ has to be a positional dictatorship.

**Definition** (Positional dictatorship). For some rank $k$, $F$ orders alternatives the same way as the person whose welfare $u(x) = u_i(x)$, where $u_i$’s utility is ranked $k$th among all the people for alternative $x$.

- $g(u) = (\phi(u^1), \ldots, \phi(u^I))$, where $\phi$ is a positive affine transformation.
  That is, we can rank order the differences in welfare between any two alternatives across people. This combination of interpersonal comparison and a comparison of differences across alternatives yields Utilitarianism (possibly with different weights for different people). Specifically, $F(u)$ orders $X$ the same way that $\sum_i (s_i \cdot u_i(x))$ does (see MWG Proposition 22.D.3).

---

1 Alternatively we might state this as $F$ being invariant on a set of Required Equivalence Classes.
2 $g_i(u_i)$ not only ranks $x \in X$ the same as $u^i$ but also ranks lotteries over $X$ the same as $u^i$
5.2 Relative Utilitarianism

Assume that for each individual there is a best and a worst element in $X$. Then we can “scale” everyone’s preferences so that $u_i(x_i^{\min}) = 0$ and $u_i(x_i^{\max}) = 1$. Relative utilitarianism is the social welfare function $F$ that ranks $X$ according to $\sum_i u_i(x)$.

Note: Jerry often assumes that individuals’ utility functions $u_i$ satisfy the Expected Utility Hypothesis. This is so that their scaled utility functions represent their risk-aversion and we can discuss lotteries over alternatives. It also allows us to define the utility possibility set and compare Relative Utilitarianism to the Nash Bargaining solution in lecture 7. See the example below and its continuation in problem set 4 for an illustration.

5.3 Example

Problem (Spring 2014 — D3 on Relative Utilitarianism). Two people have the opportunity to share one dollar. There are three non-stochastic sharing options that they can agree upon:

- $a$: person 1 gets the entire dollar
- $b$: person 2 gets the entire dollar
- $c$: person 1 gets 0.2 and person 2 gets 0.8

There are no other non-stochastic ways to share the dollar that are available to them and dividing the dollar in any proportions other than 100%, 20% or 0% for person 1 is not possible. However, the two people can make an agreement to randomize among the three options $a$, $b$, and $c$ using any probabilities.

If they do not make any agreement at all, both players get zero. They are each expected utility maximizers. Their Bernoulli utility functions are $u$ and $v$ respectively.

a) For this part of the problem, assume that $u$ and $v$ are both very slightly risk-averse, but almost risk neutral. This means that, for player 1, $c$ is as good as a randomization between $a$ and $b$ in which $a$ gets a probability weight of just slightly over .2 and, for player 2, $c$ is equivalent to a randomization between $a$ and $b$ in which $a$ gets a probability weight of just slightly under .2. What is the Relative Utilitarian solution to this bargaining problem. That is, what outcomes, non-stochastic or randomized, will these two people agree upon when they choose the Relative Utilitarian solution?

c) Now suppose that player 1 becomes significantly more risk averse while player 2’s utility remains only slightly risk averse as in part (a). For player 1, option $c$ is now indifferent to a mixture of $a$ and $b$ with probabilities 0.4 and 0.6 respectively. What is the Relative Utilitarian solution now?

h) Explain the behavior of the Relative Utilitarian solution in terms of the logical foundations of what it is trying to achieve.
6 Single-Peaked Preferences

Arrow’s impossibility result can be broken if we are willing to part with one of the four axioms. Single-peakedness restricts the domain of the preference profiles of individuals, violating Unrestricted Domain. We will define this restriction, and given such a restriction on the domain the majority tournament relation always produces a strict ordering with no cycling.

**Definition** (MWG 21.D.2. Linear order). A binary relation $\geq$ on the set of alternatives $X$ is a **linear order** on $X$ if it is reflexive (i.e., $x \geq x$ for every $x \in X$), transitive (i.e., $x \geq y$ and $y \geq z$ implies $x \geq z$), and complete (i.e., for any distinct $x, y \in X$, we have that either $x \geq y$ or $y \geq x$, but not both).

The simplest example of a linear order occurs when $X$ is a subset of the real line, $X \subset \mathbb{R}$, and $\geq$ is the natural “greater than or equal to” order of the real numbers. For example, if $X$ is a set of possible marginal tax rates and we are comparing tax rates in the natural way.

**Definition** (MWG 21.D.3. Single-peakedness). The rational preference relation $\leq$ is **single-peaked** with respect to the linear order $\geq$ on $X$ if there is an alternative $x \in X$ with the property that $\leq$ is increasing with respect to $\geq$ on the set $\{y \in X : x \geq y\}$ and decreasing with respect to $\geq$ on the set $\{y \in X : y \geq x\}$. That is, if $x \geq z > y$ or if $y > z \geq x$, then $z > y$.

In words: There is an alternative $x$ that represents a peak of satisfaction and, moreover, satisfaction increases as we approach this peak (so that, in particular, there cannot be any other peak of satisfaction).

**Definition** (MWG 21.D.4. Single-peakedness domain with respect to a linear order). Given a linear order $\geq$ on $X$, we denote by $\mathcal{R}_\geq \subset \mathcal{R}$ the collection of all rational preference relations that are single-peaked with respect to $\geq$.

For every individual $i \in \mathcal{I}$, we denote by $x_i \in X$ the maximal alternative for $\geq_I$ (we will say that $x_i$ is “$i$’s peak”).

**Definition** (MWG 21.D.5. Median agent). Agent $h \in \mathcal{I}$ is a **median agent for the profile** $(\geq_1, \ldots, \geq_I) \in \mathcal{R}_\geq^I$ if:

$$|\{i \in \mathcal{I} : x_i \geq x_h\}| \geq \frac{I}{2}$$

and

$$|\{i \in \mathcal{I} : x_h \geq x_i\}| \geq \frac{I}{2}$$

A median agent always exists.

Informally, the median voter is the one whose peak is the median of the set of all peaks.
Proposition (MWG 21.D.1. The median voter’s peak is a Condorcet winner). Suppose that $\succeq$ is a linear order on $X$ and consider a profile of preferences in $R_{\succeq}$. Let $h \in I$ be a median agent. Then $x_h \succeq y$ for every $y \in X$.

Proposition (MWG 21.D.2. Majority rule is transitive with single-peakedness). Suppose that $I$ is odd and $\succeq$ is a linear order on $X$. Then pairwise majority voting generates a well-defined social welfare function $F: P^I_{\succeq} \rightarrow R$. The resulting order is complete and transitive.

There will be no cycles with majority rule under single-peakedness, so we always get a well-defined social ordering without invoking any particular solution concept (solution concepts exist to make sure we always arrive at a social ordering or set of winners even when there are cycles).
1 Cooperative Games (Transferable Utility)

A cooperative game with transferable utility is given by an ordered pair \((N, v)\), where

- \(N = \{1, \ldots, n\}\) is the set of players
- \(v: 2^N \to \mathbb{R}\) assigns a value to each subset of \(N\), with \(v(\emptyset) = 0\) by convention
- Players have transferable utility (in the units of \(v(\cdot)\))

The solution to the cooperative game \((N, v)\) is given by an allocation function \(f(N, v) = x \in \mathbb{R}^n\), which is feasible if and only if \(\sum_{i \in N} x_i \leq v(N)\).

Most but not all games that we will consider fulfill the following property:

- **Super-Additivity**: For all \(S, T \subseteq N\) such that \(S \cap T = \emptyset\), \(v(S) + v(T) \leq v(S \cup T)\).

2 Properties of Solution Functions

We will be interested in solution functions \(f(N, v)\) that can be characterized by satisfying a number of different properties. The most important of these properties are stated below.

- **Dummy**: If a player does not add value to any coalition, the solution function should give him 0 value.
  If there exists a player \(i \in N\) such that \(v(S \cup \{i\}) = v(S)\) for all \(S \subseteq N\), then \(f_i(N, v) = 0\).

- **Symmetry**: If a pair of players \(i\) and \(j\) contribute equally to all coalitions, then the solution function should give them the same value.
  If there exist players \(i, j \in N\) such that \(v(S \cup \{i\}) = v(S \cup \{j\})\) for all \(S \subseteq N\) for which \(i \notin S\) and \(j \notin S\), then \(f_i(N, v) = f_j(N, v)\).

- **Efficiency**: No value left undistributed
  \[\sum_{i \in N} f_i(N, v) = v(N)\]

- **Additivity**: Playing two separate games should give the same total value for each player as playing the additive game.
  Given \((N, v_1)\) and \((N, v_2)\), \(f(N, v_1 + v_2) = f(N, v_1) + f(N, v_2)\).
3 Solutions for Cooperative Games

3.1 Shapley Value

Theorem (Shapley, 1953). There exists a unique solution $\phi(N, v) \rightarrow \mathbb{R}^n$ satisfying the dummy and symmetry properties, efficiency, and additivity, given by

$$\phi_i(N, v) = \frac{1}{n!} \sum_{\pi \in \Pi} [v(P(\pi, i) \cup \{i\}) - v(P(\pi, i))],$$

where $\Pi$ is the set of all permutations of enumerating $N$ and $P(\pi, i) = \{j \in N : \pi(j) < \pi(i)\}$ are predecessor sets.

The above is called the random order approach. Alternatively, instead of writing the formula as it depends on the list of marginal values of $i$ in different random orders, we can write it equivalently as a function of the marginal value of $i$ to different coalitions $S$. This is the combinatorial approach.

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

In words, the Shapley value of individual $i$ can be thought of as his expected marginal contribution to a randomly formed coalition, assuming all coalitions are equally likely to form in any order. Solving for the Shapley value therefore requires knowing when an individual makes a marginal contribution to a type of coalition and the associated probability of such a coalition forming.

3.2 Equal Allocation of Non-Separable Surplus

This solution is based on the idea of paying everyone their marginal product.

$$EANS_i = [v(N) - v(N \setminus \{i\})] + \frac{1}{n} \left[v(N) - \sum_j (v(N) - v(N \setminus \{j\}))\right]$$

$$= \frac{1}{n} \left[v(N) + \sum_j v(N \setminus \{j\})\right] - v(N \setminus \{i\})$$

Consider the first formulation before the simplification: Under EANS, individuals receive their marginal contribution to the worth of the entire population — the first bracketed term. The remaining surplus/deficit is split equally among all individuals — the second bracketed term. This is necessitated by efficiency. However, this second term necessarily violates the dummy property; dummy players share the unattributed surplus/cost despite having zero marginal contribution to any coalition.
A second thing to point out: the EANS rule has computational appeal; unlike the Shapley value it only considers the payoff values for \( N + 1 \) coalitions.

EANS is symmetric, efficient and additive. (It fails to satisfy dummy.)

### 3.3 The Pre-Nucleolus

**Definition** (Excess of coalition). For each coalition \( S \), define the **excess of coalition** \( S \) at \( x \) to be

\[
e(x, S) = v(S) - \sum_{i \in S} x_i
\]

The pre-nucleolus is the unique lexical minimizer of the excesses of any non-empty \( S \subseteq X \) over all feasible allocations. In particular, we first minimize the largest complaint (other than that for \( N \) which is zero at all feasible and efficient allocations). Then in the same spirit we continue the minimization process by asking for the allocation that, among those with the smallest maximal excess, exhibits the smallest next-largest excess until we converge to a unique solution.

*The pre-nucleolus satisfies efficiency, symmetry and dummy. It may or may not satisfy additivity. In the case of PS games, the pre-nucleolus coincides with the Shapley value and therefore will satisfy additivity.*

**Definition** (PS games). For every \( i \) and every \( S \) with \( i \notin S \), compute \( i \)'s marginal value to \( S \) and to \( N \setminus (S \cup \{ i \}) \):

\[
\begin{align*}
S : & \quad v(S \cup \{ i \}) - v(S) \\
N \setminus (S \cup \{ i \}) : & \quad v(N \setminus S) - v(N \setminus (S \cup \{ i \}))
\end{align*}
\]

A PS game is one where the sum of those two components is independent of \( S \).

The Shapley Value and the pre-nucleolus solutions coincide for this set of games.

If all the excesses can be made non-positive, then no coalition would have a valid complaint if this minimizer were implemented. This is the definition of the **core**.

**Definition** (The Core). The core of a transferable utility game is the set of utility vectors \( x = (x_1, \ldots, x_n) \) satisfying the linear inequalities \( \sum_{i \in S} x_i \geq v(S) \) for all \( S \subseteq N \) and \( \sum_{i \in N} x_i \leq v(N) \).

The core requires there to be no coalition that can deviate from the group and increase the total payoff of its members. Unlike in general equilibrium, not all cooperative games have a non-empty core. If the core exists, then the pre-nucleolus is in the core. If the core is empty when it is not possible to make all the excesses non-positive. **Thus, the non-emptiness of the core can be tested by calculating the pre-nucleolus.**
In general, for a large problem with many individuals, classifying the core fully is often not trivial. If you are required to show that the core is empty, think of what the core if exists should look like and/or what the pre-nucleolus should look like, and check if those allocations can be blocked (See Spring 2010 general question D2, part c for an example). If you are required to show that the core is non-empty, exhibiting one allocation in the core suffices. You can also invoke any result from lectures or problem sets. In particular, for any convex games, the core is non-empty and the Shapley value is in the core (See problem set 3, question 3).

**Definition** (Convex games). A cooperative game \((N,v)\) is convex if

\[
v(T) + v(S) \leq v(T \cup S) + v(T \cap S),\]

for any two coalitions \(T, S \subseteq N\).

**Note:** Convex games are a subclass of super-additive games. For convex games, \(S \cap T = \emptyset\) implies \(v(T) + v(S) \leq v(T \cup S)\).

### 4 Subgame Consistency

Given a game \((N,v)\) and a subset \(T \subseteq N\), if the subset of players \(N \setminus T\) “take their money and leave,” then we can define a subgame \((T,w)\), where \(w: 2^T \to \mathbb{R}\) depends on \(v(\cdot)\). We want the payoffs in this interaction to be the same as in the original game. Precisely how \(w(\cdot)\) depends on \(v(\cdot)\) requires an ethical stance on how cooperation between a coalition \(S \subseteq T\) and the players in \(N \setminus T\) should proceed, and should depend on the institutional structure being modeled.

- **Definition #1:** \(S\) must cooperate with all the members of \(N \setminus T\). \(S\) must pay for their participation as the solution \(f\) requires, as if the only players in the game were \(S \cup (N \setminus T)\). For all \(S \subseteq T\):

  \[
w(S) = v(S \cup (N \setminus T)) - \sum_{i \in N \setminus T} f_i(S \cup (N \setminus T), v)\]

- **Definition #2:** \(S\) must cooperate with all the members of \(N \setminus T\). \(S\) must pay for their participation as the solution \(f\) requires, as applied to the original game \((N,v)\). For all \(S \subseteq T\):

  \[
w(S) = v(S \cup (N \setminus T)) - \sum_{i \in N \setminus T} f_i(N,v),\]

- **Definition #3:** \(S\) can choose the best subset \(Q \subseteq N \setminus T\) with whom to cooperate. \(S\) must pay for their participation as the solution \(f\) requires, as applied to the original game \((N,v)\).
For all $S \subseteq T$:

$$w(S) = \max_{Q \subseteq N \setminus T} \left\{ v(S \cup Q) - \sum_{i \in Q} f_i(N, v) \right\},$$

**Definition.** A solution concept $f(\cdot)$ is **subgame consistent** if for every $T \subseteq N$ and every $i \in T$, $f_i(N, v) = f_i(T, w)$.

Each of the above definitions together with efficiency, symmetry, and dummy (but for definition 2) uniquely defines each of the three solution concepts as in the following table:

<table>
<thead>
<tr>
<th>Cooperation and compensation rules</th>
<th>$S$ can choose best $Q \subseteq N \setminus T$</th>
<th>$S$ must cooperate with all of $T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N \setminus T$ paid according to $f(N, v)$</td>
<td>Def. #3: Pre-nucleolus</td>
<td>Def. #2: EANS</td>
</tr>
<tr>
<td>$N \setminus T$ paid according to $f(S \cup (N \setminus T), v)$</td>
<td>No known solution</td>
<td>Def. #1: Shapley Value</td>
</tr>
</tbody>
</table>

## 5 Examples

**Problem** (Spring 2007 — D1 — (d) and (e)). Alice (A), Bob (B) and Charlie (C) are friends. They are such good friends that they are contemplating marriage. Specifically, Alice and Bob could get married ($x$), or Alice and Charlie could get married ($y$), or they could all remain single ($z$).

Alice, Bob and Charlie have utility functions that represent the intensity of the ordinal preferences described above. Moreover they can agree to make zero-sum monetary transfers among themselves and their preferences are quasi-linear in money. Their cardinal preferences for the alternatives in the absence of any monetary transfers are given in the following table:

<table>
<thead>
<tr>
<th></th>
<th>$x$</th>
<th>$y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

d) Now assume that these players decide to use Cooperative Game Theory to compute the outcome. The characteristic function is to be calculated as follows: For every coalition $S$, $v(S)$ is the largest value that $S$ can guarantee — that is it is the largest value that $S$ can achieve given that the members of $S$ can make a binding commitment to marry or not to marry, recognizing that the complementary coalition can take any action (whether or not it is rational for them), and $S$ must accept the lowest possible value that this action can induce. Thus, any coalition $S$ of which Alice is a member can agree to a marriage between Alice and any other member of the coalition, or it can agree that all members of the coalition remain
single. The coalition of Bob and Charlie can only achieve \( z \) since they do not have the right to assume that Alice will marry either of them (even though she will). Any coalition \( S \) consisting of a single player must assume that the complementary coalition will do whatever is worst for \( S \). What is the core of this cooperative game? If you were Bob, how would you argue against implementing one of the points in the core? Comment on the construction of the characteristic function as given above. What would you get if you used a different construction? (10)

e) Since things are not going well in this threesome’s marriage discussions, they decide to compute the Shapley Value of the game described above. At least that will be a single well-defined point. What is it? (8)

**Problem** (Spring 2010 — D2). There are two technologies for producing an output \( z \) that is used for consumption using inputs \((x, y)\). Technology A produces output according to \( z = 10x + 2y \). Technology B requires a fixed cost of **four million** units of \( y \) and has the output function \( z = 5x + 10(y - 3 \times 10^6) \). There is a population of **30 million** people, each of whom desires only the output \( z \). Half of them have a unit of \( x \) in their endowment. The other half have a unit of \( y \). Any group of people can choose between these these technologies, operating only one of them, so as to produce the largest amount of output they can from their collective endowment. The question is how should they divide the output that they produce.

They decide to use the Shapley value of the cooperative game where \( v(S) \) is the highest output that \( S \) can produce.

a) Explain the formula for the Shapley value. (6)

b) Using a “large numbers” approximation in this model with 30 million players, compute the payoffs in units of output for the owners of the two inputs. (15)

**Problem** (Spring 2012 — D2). A group of treasure hunters has found a map showing the location of buried treasure on a remote island. They believe that it is 100% accurate and the treasure, worth \( T \), will be recoverable by the first group to travel to the island. Unfortunately there are some pirates near this island. A group that is not sufficiently strong will be repelled by these pirates and will have to return home empty-handed. That outcome has the same value as never having set forth to find the treasure at all.

The set of treasure hunters is \( N \). The members of \( N \) denoted \( i = 1, \ldots, n \) are each characterized by two numbers a strength level \( y_i \) and a cost \( c_i \). Let \( S \subseteq N \) be a set of treasure hunters. The aggregate strength level needed to overcome the pirates is \( Y \). If \( \sum_{i \in S} y_i \geq Y \) the trip will succeed and the treasure will be recovered; if not the pirates will repel \( S \). If \( S \) makes a trip they come home with a net value \( T - \sum_{i \in S} c_i \) if \( \sum_{i \in S} y_i \geq Y \) and zero otherwise. Assume that the trip is worthwhile if undertaken by everyone, although that may not be the efficient way to do it: \( T - \sum_{i \in N} c_i > 0 \) and \( \sum_{i \in N} y_i \geq Y \).
The treasure hunters hold a meeting to plan the trip. They decide to use the Shapley value of the cooperative game that represents this situation to determine who should make the trip and how the net value should be divided among all of $N$. They can make any monetary transfers they desire.

a) Suppose that everyone is required or else the trip will not succeed. That is, $\sum_{i \in N} y_i \geq Y$ but $\sum_{i \in S} y_i < Y$ for all $S \subseteq N$. What is the Shapley value? How does it depend on $T, Y$ and the individuals’ parameters $(y_i, c_i)_{i \in N}$. (6)

For the remainder of this question, assume that the strengths $y_i$ are all the same, $y$. Also assume that there are only two values of $c_i$ in the population: $n_L$ people have $c_L$ and $n - n_L$ people have $c_H$, with $c_H > c_L$. Suppose that $k$ treasure hunters are required to overcome the pirates: $(k - 1)y < Y < ky$ and that the $c_L$ people alone cannot succeed: $n_L < k$.

b) Give a formula for the Shapley value in this problem. Which treasure hunters should actually go on a trip? Are the people not going on the trip receiving any payments from those who do go? Explain. (12)

c) Is the core of this game empty or non-empty? Explain in detail. (7)
# 1 Cooperative Games (Non-Transferable Utility)

Cooperative game theory studies situations where coalitions of individuals can make binding agreements and benefit from cooperation. In these games, we do not consider strategic interactions among individuals. A solution is an allocation of individual outcomes that satisfies desirable properties, usually including some notion of fairness.

In section 3, we study cooperative games with transferable utilities (TU games) when all outcomes are denominated in the same unit. Payoffs to each coalition can be represented by a single value and transfers can be made to divide the total payoff of the grand coalition among its members.

In cooperative games with non-transferable utilities (NTU games), especially when individuals are risk averse, group actions result in payoffs to individual group members. In this case, the set of feasible payoffs for a group of individuals can be any convex set instead of a set with a linear Pareto frontier (slope of negative one).

## 1.1 General Framework

A cooperative game with non-transferable utility is given by an ordered pair \((N, V)\), where

- \(N = \{1, \ldots, n\}\) is the set of players
- For any \(S \subseteq N\), the set of feasible payoff vectors for \(S\) is denoted by \(V(S) \subseteq \mathbb{R}^n\). An element \(x = (x_i)_{i \in S}\) of \(V(S)\) is interpreted as follows: there exists an outcome that is feasible for the coalition \(S\) such that the utility to player \(i\) is \(x_i\) (for each \(i\) in \(S\)). Thus \(V(S)\) is the set of utility combinations that are feasible for the coalition \(S\).

Compare this to a TU game \((N, v)\). The equivalent representation of \((N, v)\) is the NTU game \((N, V)\) such that \(V(S) = \{ x \in \mathbb{R}^n : \sum_{i \in S} x_i \leq v(S) \}\) for any \(S \subseteq N\). Because utilities are transferable in a TU game, the Pareto frontier is always a straight line (linear surface) at \(-45\) degree angle after normalizing each individual’s value to 0.

The solution to the cooperative game \((N, V)\) is given by an allocation function \(f(N, V) = x \in \mathbb{R}^n\), which is feasible if and only if \(x \in V(N)\). You can see the parallelism between NTU games and TU games.

## 1.2 Bargaining Framework

We are most often concerned with a specific type of NTU games: bargaining games. Here, the only surplus that can be generated is by the grand coalition \(N\): in other words, \(V(N) \neq \emptyset\), but \(V(S) = \emptyset\) for all \(S \subseteq N\). We are also concerned about a threat point (or disagreement point) of what each individual gets in the case the bargaining breaks down and there is no agreement.
• $N = \{1, \ldots, n\}$ is the set of players

• A set of feasible utilities $U \subseteq \mathbb{R}^n$ that is closed, convex, bounded above: utility possibility set by the grand coalition

• A vector of threat points, $u_0 = (u_{01}, \ldots, u_{0n}) \in U$

  The interpretation of $u_{0i}$ is that it forms a threat or a reservation utility that $i$ can achieve on his own. This forms a definite lower bound on $i$’s outcome.

• A bargaining solution $f(U, u_0) \in U$ is a rule that assigns a solution vector in the feasible set $U$ to every bargaining problem $(U, u_0)$

Most of the times, the bargaining problem is between two individuals: either they bargain and come to an agreement, or each gets their threat point utility. Applications include: divide-the-dollar game, or a bargain between an input supplier and a downstream firm when the input is firm-specific. In the latter case, the price is not determined by a competitive market but through bargaining. You have encountered this situation in Oliver’s part.

1.3 Properties of Solutions to Bargaining Games

• **Independence of utility origins:** The bargaining solution does not depend on absolute scales of utility.

  This property is always assumed. It allows us to normalize our problems to $u_0 = 0$. We can then denote $f(U, u_0)$ as $f(U)$.

• **Independence of utility units:** The bargaining solution does not involve interpersonal comparisons of utilities.

  For any $\beta = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$ with $\beta_i > 0$ for all $i$, we have: $f_i(U') = \beta_i f_i(U)$ for every $i$ whenever $U' = \{(\beta_1 u_1, \ldots, \beta_n u_n) : u \in U\}$.

  **Note:** These first two properties together are combined in Jerry’s definition of affine invariance.

• **Pareto:** The solution is on the Pareto frontier of the feasible set

  For every $U$, $f(U)$ is a (weak) Pareto optimum, that is, there is no $u \in U$ such that $u_i > f_i(U)$ for every $i$.

• **Symmetry:** If all agents are identical, then the gains from cooperation are split equally.

• **Individual rationality:** No one gets less than the threat point — $f(U) \geq 0$.

• **Independence of irrelevant alternatives:** If we remove only irrelevant alternatives from $U$, then the solution is unchanged

  Whenever $U' \subset U$ and $f(U) \in U'$, then $f(U') = f(U)$.
1.4 Nash Bargaining Solution

The Nash Bargaining Solution is defined as:

\[ f(U, u_0) = \arg \max_{u \in U} \prod_{i=1}^{I}(u_i - u_{0i}) \]

**Proposition** (MWG Proposition 22.E.1). *The Nash solution is the only bargaining solution that is independent of utility origins and units, Paretian, symmetric, and independent of irrelevant alternatives.*

Finding the Nash bargaining solution in the two-agent case after normalizing the threat point to 0 can be done graphically by considering the following cases:

- If the utility possibility set \( U \) is symmetric, then the solution is at the intersection between the 45-degree line and the Pareto frontier by Symmetry.
- If the Pareto frontier is linear, then the solution is the exact midpoint on the Pareto frontier: can be shown by Symmetry and Independence of Utility Units.
- With a generic Pareto frontier, consider all tangent lines to the Pareto frontier; if the midpoint of the first-quadrant segment of the tangent line is on the Pareto frontier, then that point is the Nash bargaining outcome: can be shown by IIA. Note that the “tangent” line at a kink point may not be unique unlike in the case with a smooth surface (see the Nash bargaining example in problem set 4).

There has been considerable effort in linking the cooperative and non-cooperative game approach. The “Nash program” is the research agenda for investigating on the one hand axiomatic bargaining solutions and on the other hand the equilibrium outcomes of strategic bargaining procedures. You may encounter non-cooperative bargaining solutions suggested by Rubinstein 1982 or Binmore, Rubinstein, and Wolinsky 1986.

1.5 The Shapley NTU Value

A solution to a generic NTU game (outside of the bargaining framework) can be given by the Shapley NTU value. This solution is not a point-valued solution and can result in a set of allocations.

For any \( \lambda \in \mathbb{R}_+^n \), define the \( \lambda \)-TU game that corresponds to our NTU game \((N, V)\) as follows: \( v_\lambda(S) = \max \{ \sum_{i \in S} \lambda_i x_i \mid (x_i)_{i \in S} \in V(S) \} \). Graphically, \( v_\lambda(S) \) can be represented by the tangent surface of \( V(S) \) with slope represented by \( \lambda \) in the \(|S|\)-dimensional space. If the Shapley value to this \( \lambda \)-TU game \( \phi(N, v_\lambda) \) is feasible in the original NTU problem, i.e. \( \phi(N, v_\lambda) \in V(N) \), then it
is an allocation that satisfies the Shapley NTU value solution. Note that it is not required that 
\( \{\phi_i(N, v_x)\}_{i \in S} \in V(S) \) for any \( S \subseteq N \).

In a two-people NTU game \((n = 2)\), the Shapley NTU value solution is unique and it coincides with 
the Nash bargaining solution. To see this, map the graphical representation of the corresponding 
\( \lambda \)-TU game to the graphical way to find the Nash bargaining solution in the previous subsection,
noting that the Shapley value also satisfies symmetry.

This solution satisfies affine invariance, Pareto, symmetry, individual rationality, and IIA as defined 
for the bargaining problem. Furthermore, because we consider all sub-coalitions in the general NTU 
framework, we can consider the dummy property as well.

One variant of the dummy property states that: If a dummy is defined as a player who does not 
gain any value himself when being added to any coalition that is already on their Pareto frontier, 
then he should get 0. This may not be desirable, because in an NTU game, it is possible that adding 
a player changes the utility possibility set in such a way that the Pareto frontier for the existing 
players expands even when the new player gets 0 himself (see example in Lecture 7, page 28). In 
that case, another variant of the dummy property does not consider such a player a dummy — he 
plays the role of a facilitator. We may not want to give such a player 0.

1.6 Example

Problem (Spring 2007 — D1 — (c)). Alice (A), Bob (B) and Charlie (C) are friends. They are 
such good friends that they are contemplating marriage. Specifically, Alice and Bob could get 
marrried \((x)\), or Alice and Charlie could get married \((y)\), or they could all remain single \((z)\).

Alice, Bob and Charlie have utility functions that represent the intensity of the ordinal preferences 
described above. Moreover they can agree to make zero-sum monetary transfers among themselves 
and their preferences are quasi-linear in money. Their cardinal preferences for the alternatives in 
the absence of any monetary transfers are given in the following table:

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

c) Alice, Bob and Charlie consider the intensities of their preference as described by these 
utilities. They realize that the efficient outcome is \(y\) — Alice should marry Charlie. Moreover, 
Bob objects to the outcome of part (a) because he says that everyone should enter marriage 
voluntarily. He proposes that they use the Nash Bargaining Solution to determine both the 
marriage outcome and a zero-sum monetary transfer among them, with the status quo point 
being outcome \(z\) (and zero monetary transfers). What would the resulting outcome be?
2 Claims Problems (a special TU game)

This is a specific type of problem, usually found in a legal context when the allocation problem involves claims that exceed the total value (estate) and bankruptcy occurs.

- **E** is the estate
- **N = {1, ..., n}** is the set of players
- Each player has a claim **c_i ∈ ℝ_+**
- Bankruptcy occurs, i.e. \( \sum_{i \in N} c_i \geq E \)

We want to allocate values to all the players assuming their claims are in good faith. The solution function is \( F(c, E) \in ℝ^n \) such that \( \sum_i F_i(c, E) = E \).

The desirable properties that we consider in this class are:

- **Claims boundedness**: No one gets more than their claim \( 0 \leq F_i(c, E) \leq c_i \)
- **Concede-and-divide** in any two-player problem: \( F_i(c, E) = \max \{ E - c_i, 0 \} + \frac{E - \max \{ E - c_1, 0 \} - \max \{ E - c_2, 0 \}}{2} \).
  This uniquely determines the solution in any two-player game.
- **Pairwise consistency**: At the solution of a problem with \( n > 2 \) players, every pair’s outcomes \( F_i \) and \( F_j \) should be the same as if they were a separate two-person problem with their original claims \( c_i \) and \( c_j \) and with an estate in that two-person problem being the amount they are actually sharing at the solution \( E_{\{i,j\}} = F_i + F_j \).

**Theorem** (Aumann and Maschler). *In any claims problem there is one unique allocation that satisfies Claims Boundedness, Concede-and-Divide, and Pairwise Consistency.*

Furthermore, it coincides with the pre-nucleolus of the TU game defined by \( N \) and \( v(S) = \min \left\{ E - \sum_{j \notin S} c_j, \sum_{j \in S} c_j \right\} \) for any \( S \subseteq N \).

The theorem describes how you will go from the description of a claims problem to a TU game such that the three desirable properties above are satisfied. The specified TU game is a “polite” one, in which any coalition \( S \) gives the outsiders all of their claims first and gets its own claim only if the total claim amount is less than the leftover.

In a problem when you have to find the pre-nucleolus (or Shapley value) with specific values, you should look for symmetry, dummy, and efficiency, which the pre-nucleolus solution satisfies. Often, if \( v(\{i\}) \neq 0 \) to some \( i \), it is harder to spot symmetry (think about a two-person game which is always symmetric if \( v(\{1\}) = v(\{2\}) \) but not otherwise). You may want to do a 0-normalization, reducing the marginal contribution of \( i \) to any coalition \( S \subseteq N \) by the same amount \( v(\{i\}) \), which will reduce the Shapley value and pre-nucleolus of \( i \) by \( v(\{i\}) \) and often an easier problem to solve.

The 0-normalization problem of \((N, v)\) is \((N, w)\) where \( w(S) = v(S) - \sum_{i \in S} v(\{i\}) \) for all \( S \subseteq N \). Note that \( f_i(N, v) = f_i(N, w) + v(\{i\}) \).
3 Mechanism Design — Setup

Game theory takes the rules of the game as given, and makes predictions about the behavior of strategic players. The theory of mechanism design is about designing the rules of the game to achieve outcomes that satisfy certain desirable properties. To achieve these outcomes, the designer often needs to know the individuals’ preferences.

In social choice theory, we deal with the problem of aggregating individual preferences, while mechanism design is about the problem of eliciting them given that individuals behave strategically.

3.1 Framework

- A set \( I = \{1, \ldots, I\} \) of players.
- A set of possible outcomes \( X \).
- Each player \( i \) has type \( \theta_i \in \Theta_i \) and we denote the vector of types \( \theta = (\theta_1, \ldots, \theta_I) \) and the type space \( \Theta = \Theta_1 \times \cdots \times \Theta_I \).
  - The distribution of types is assumed to be common knowledge. The random variable \( \tilde{\theta}_i \) will refer to player type \( i \)'s type when there is uncertainty about his type.
  - We will almost exclusively restrict attention to the case when the players’ types are drawn independently.
- A mechanism is given by \( \Gamma = (S_1, \ldots, S_I, g(\cdot)) \), where
  - \( S_i \) is player \( i \)'s action space, and \( s_i : \theta_i \rightarrow S_i \) is player \( i \)'s strategy.
  - \( g : S_1 \times \cdots \times S_I \rightarrow X \) is the outcome function.

Both \( S \) and \( g \) are designed. Different objectives of the designer can be specified but are not part of the general framework.

- Player \( i \) gains utility \( u_i(x, \theta) \) from outcome \( x \in X \), where utilities are generically allowed to depend on the types of other players.
  - We will often restrict attention to the private values case, in which \( u_i(x, \theta) = u_i(x, \theta_i) \) for all \( \theta \in \Theta \) and \( x \in X \), or utility of a person \( i \) only depends on his type \( \theta_i \) and not others’ types \( \theta_{-i} \).
  - Moreover, we will almost exclusively assume that individuals have utility functions that are **quasi-linear in monetary transfers**. Here, outcome \( x \) consists of the “real” aspect of the outcome \( k \) and money transfers to each agent \( (t_1, \ldots, t_I) \) \( (x = (k, t_1, \ldots, t_I) \in X) \). Utility function of agent \( i \) is given by
    \[
    u_i(x, \theta) = v_i(k, \theta) + t_i,
    \]
where \(v_i(k, \theta)\) is the value player \(i\) has for alternative \(k\) given type vector \(\theta\) and \(t_i\) is his monetary transfer. Sometimes we restrict \(\sum_{i=1}^{I} t_i = 0\): only a redistribution.

A **social choice function** is a mapping \(f: \Theta \to X\) that selects an outcome as a function of the players’ types: \(f(\theta) = g(s(\theta))\). This is the outcome function induced by the mechanism.

### 3.2 Timing and Participation Constraints

We need to make more precise assumptions about types and beliefs to turn the mechanism into a game. Once we do, we can apply what we know from game theory to make predictions.

- **Ex-ante:** The players know the structure of the game, but not their respective types.
- **Interim:** Each player knows his own type, but not the types of the other players.
- **Play:** Each player \(i\) chooses a strategy \(s_i\), and the mechanism takes the realized actions \(s(\theta)\) as input.
- **Ex-post:** All \((\theta_i, s_i(\theta_i))_{i \in I}\) and the outcome of the mechanism are common knowledge.

In some models of mechanism design, the set of participants is not fixed. People retain the right to quit, given some outside option. Usually the mechanism designer is constrained to make sure that no one exercises this right. Depending on the type of commitment the players can make, we will be interested in mechanisms that satisfy ex-ante, interim, and ex-post participation constraints in variable settings.

Note that all the definitions are stated in the private-value case. Suppose that player \(i\) has reservation utility \(\bar{u}_i\) and that \(s^*\) is an equilibrium strategy profile of the mechanism \(\Gamma = (S_1, \ldots, S_I, g(\cdot))\).

**Definition** (Ex-ante participation). The mechanism \(\Gamma\) satisfies the **ex-ante participation constraint** (or is ex-ante individually rational) under \(s^*\) if

\[
\mathbb{E}_{\tilde{\theta}}[u_i(g(s^*(\tilde{\theta}))), \tilde{\theta}_i)] \geq \bar{u}_i.
\]

Before knowing anyone’s type including their own, each player integrates out all the possible types according to some belief about the type distribution.

**Definition** (Interim participation). The mechanism \(\Gamma\) satisfies the **interim participation constraint** (or is interim individually rational) under \(s^*\) if

\[
\mathbb{E}_{\tilde{\theta}_{-i}}[u_i(g(s^*(\theta_i, \tilde{\theta}_{-i})), \theta_i)] \geq \bar{u}_i.
\]

The player knows their own type but takes an expectation over the other players’ types.
Definition (Ex-post participation). The mechanism $\Gamma$ satisfies the **ex-post participation constraint** (or is **ex-post individually rational**) under $s^*$ if

$$u_i(g(s^*(\theta)), \theta_i) \geq \bar{u}_i.$$ 

There is no more uncertainty in this case.

Note that participation constraints are easiest to achieve ex-ante and most difficult to achieve ex-post, since player $i$ may want to withdraw from the mechanism following unfavorable draws from $\Theta_i$, and he may be further inclined to withdraw following unfavorable realizations from $\Theta$.

### 3.3 Efficiency

Efficiency is a natural objective of the designer, and the definition of efficiency depends on timing. In this course, we often focus on the case of quasi-linear utilities with monetary transfers. In this environment, all players, conditional on participation, would prefer a mechanism that delivers an ex-post efficient outcome and redistributes the utility gains made from efficiency among the players. We therefore only state the condition for ex-post efficiency. The conditions for ex-ante efficiency and interim efficiency are defined analogously.

We again state the definition in the private-value case.

**Definition** (MWG 23.B.2). The social choice function $f: \Theta \rightarrow X$ is **ex-post efficient** (or **Pareto**) if for no type profile $\theta \in \Theta$ is there a feasible $x \in X$ such that $u_i(x, \theta_i) \geq u_i(f(\theta), \theta_i)$ for every $i \in I$ and $u_i(x, \theta_i) > u_i(f(\theta), \theta_i)$ for some $i \in I$.

### 3.4 Example: one-agent case

**Problem** (Spring 2007 — D2 — (a) to (c)). The federal government finally decides to do something about the health care situation in America. The population differs in their needs for health care according to a parameter $\theta \in \{\theta_L, \theta_H\}$ with $\theta_L < \theta_H$. Half of the population has each value of the parameter. Everyone’s utility depends on the health care allocated to them $q$ and the monetary transfer that they pay to the government $t$, according to the function

$$u(q, t, \theta) = \theta v(q) - t$$

Health care is a private good and thus the cost of providing a health care system which gives rise to the allocation $q(\theta_L)$, $q(\theta_H)$ in this population is $c(q(\theta_L) + q(\theta_H))$.

a) The current Republican administration wants to find a way to implement the efficient pattern $q(\theta)$. Moreover Republicans believe in individual responsibility and individual freedom to
choose: This means that (i) the payments $t(\theta)$ should be such that everyone pays for exactly what he or she uses, and (ii) everyone should choose their own $(q, t)$ from a menu with the same menu is offered to everyone. Is such an implementation possible? (5)

b) The Republican system is implemented and leads to great disparities in utility and in health care utilization. Thus when the Democrats get elected the first thing they do is to mandate an implementation of a pattern $q(\theta)$ that maximizes the average utility in the population. The Democrats also believe in individual freedom to choose. However they believe in fiscal responsibility (balanced budget) over all, but not necessarily that everyone should pay the exact cost of the health care that he or she selects. What allocation $(q(\theta_L), q(\theta_H))$ is implemented under the Democratic administration? (5)

c) Now the Socialists come to power. They are worried about the welfare of the worst-off members of society. Which type is worse off? Socialists also believe that the system must break even and give all people the same menu of choice. Their objective is the same as the Democrats, to maximize the welfare of the average person, but subject to the principle that the worst off people must get a certain minimum level of utility. The Socialists gradually raise the welfare of the worst off type of people by changing the menu of health care that is offered. Describe what happens to the optimal menu as this minimum level of utility is increased above what it was under the Democrats — starting at that level and raising it until it cannot be increased any more. (12)
Nash Bargaining Example

Linh T. Tô

Harvard University

14 April 2016
Setup

• Two agents dividing a dollar, but the only three possible deterministic outcomes (with any randomization among them allowed) are:
  ○ \(a\): Player 1 gets the dollar, player 2 does not
  ○ \(b\): Player 2 gets the dollar, player 1 does not
  ○ \(c\): Player 1 gets 0.2 dollar, player 2 gets 0.8 dollars

• Player 2 is always risk neutral. Player 1 can be risk averse.

• Utilities of the two players are normalized so that:
  ○ \(u_i(0) = 0\) with any \(i \in \{1, 2\}\)
  ○ \(u_i(1) = 1\) with any \(i \in \{1, 2\}\)

• Utilities satisfy the expected utility hypothesis.
Risk Neutral
Risk Neutral
Agent 1: $c \equiv \frac{1}{2}a + \frac{1}{2}b$
Agent 1: \[ c = \frac{1}{2}a + \frac{1}{2}b \]
Agent 1: \( c \equiv \frac{1}{2}a + \frac{1}{2}b \)
Agent 1: $c \equiv \frac{1}{2}a + \frac{1}{2}b$
Agent 1: Slightly risk-averse
Agent 1: Slightly risk-averse
Agent 1: Slightly risk-averse
Agent 1: Very risk-averse
Agent 1: Very risk-averse
Agent 1: Very risk-averse
MWG Proposition 23.D.2

If utility is of the form $u_i(x, \theta_i) = \theta_i w_i(x) + t_i(\theta)$, and the distribution of $\theta_i$ has the same support ranging from $\theta_{\min}$ to $\theta_{\max}$ for all $i$, then for an outcome $(k(\theta), t_1(\theta), \ldots, t_n(\theta))$ to be implementable in BNE (weaker than dominant strategies), theorem MWG 23.D.2 tells you that you need conditions (i) on $\bar{\omega}_i(\theta_i)$ (or equivalently on the allocation function $k(\theta)$), and (ii) on the interim utility $U_i(\theta_i)$ (or equivalently on the interim transfer $\bar{t}_i(\theta_i)$) to be satisfied.

Specifically:

(i) $\bar{\omega}_i(\theta_i) \equiv \mathbb{E}_{\theta_i \neq i}[w_i(k(\theta_i, \tilde{\theta}_{-i}))]$ is non-decreasing

(ii) Interim $U_i(\theta_i) \equiv \theta_i \bar{\omega}_i(\theta_i) + \bar{t}_i(\theta_i) = U_i(\theta_{\min}) + \int_{\theta_{\min}}^{\theta_i} \bar{\omega}_i(s)ds$, in which $\bar{t}_i(\theta_i) \equiv \mathbb{E}_{\theta_i \neq i}[t_i(\theta_i, \tilde{\theta}_{-i})].$

Denote the distribution of $\theta_i$ as $F_i$ (with domain ranging from $\theta_{\min}$ to $\theta_{\max}$ for all $i$). These distributions are common knowledge.

Mechanisms that implement the physical outcome $k(\theta)$ in BNE

If the physical outcome $k(\theta)$ induces $\bar{\omega}_i(\theta_i)$ which satisfies condition (i), then there are many ways we can set the transfers to satisfy condition (ii). In other words, knowing that condition (i) is satisfied is sufficient to say that there exists a mechanism which will implement $k(\theta)$ with some appropriate transfers. To show this, we just need to point out one way to choose the transfers.

Recall that one of the four equivalent ways to express condition (ii) is to have $\bar{t}_i(\theta_i) = C + \int_{\theta_{\min}}^{\theta_i} \bar{\omega}_i(s)ds - \theta_i \bar{\omega}_i(\theta_i)$. Thus, we can simply set the ex-post transfer function $t_i(\theta, \tilde{\theta}_{-i})$ to be $C + \int_{\theta_{\min}}^{\theta_i} \bar{\omega}_i(s)ds - \theta_i \bar{\omega}_i(\theta_i)$ (independent of $\theta_{-i}$) for some constant $C$. The lesson here is that knowing the interim transfer tells you that there are infinitely many ways to set the ex-post transfers such that they average out to be the same as the interim transfer — going forward we will only focus on the interim stage.

Note that here we assume that participation is enforced at all stages and we are not concerned about the budget. Note also that some ways to set the transfers are clearly more efficient than others. Given an appropriate $k(\theta)$, we can then explore the more interesting questions on interim participation, budget balance, and so on.

Bounds on total surplus (total ex-ante utility)

Assuming no subsidy from outside of the mechanism, sum of the total transfers ex-ante should be non-positive: $\sum_i \int_{\theta_{\min}}^{\theta_{\max}} \bar{t}_i(\theta_i)dF_i(\theta_i) \leq 0$
Given an allocation function \( k(\theta) \) and therefore \( \bar{w}(\theta) \), there is a bound on total ex-ante utility from the definition of interim utility (integrate out the interim utility and sum over \( i \)):

\[
\sum_i \int_{\theta_{\min}}^{\theta_{\max}} U_i(\theta_i) dF_i(\theta_i) = \sum_i \int_{\theta_{\min}}^{\theta_{\max}} [\theta_i \bar{w}_i(\theta_i) + \bar{l}_i(\theta_i)] dF_i(\theta_i)
\]

\[
= \sum_i \int_{\theta_{\min}}^{\theta_{\max}} \theta_i \bar{w}_i(\theta_i) dF_i(\theta_i) + \sum_i \int_{\theta_{\min}}^{\theta_{\max}} \bar{l}_i(\theta_i) dF_i(\theta_i)
\]

\[
\leq \sum_i \int_{\theta_{\min}}^{\theta_{\max}} \theta_i \bar{w}_i(\theta_i) d\theta_i
\]

Condition (ii) tells us that we can then bound \( \sum_i U_i(\theta_{\min}) \):

\[
\sum_i \int_{\theta_{\min}}^{\theta_{\max}} U_i(\theta_i) dF_i(\theta_i) = \sum_i \int_{\theta_{\min}}^{\theta_{\max}} \left[ U_i(\theta_{\min}) + \int_{\theta_{\min}}^{\theta_{\max}} \bar{w}_i(s) ds \right] dF_i(\theta_i)
\]

\[
= \sum_i U_i(\theta_{\min}) + \sum_i \int_{\theta_{\min}}^{\theta_{\max}} \int_{\theta_{\min}}^{\theta_i} \bar{w}_i(s) ds dF_i(\theta_i)
\]

Therefore, combining the two expressions (*) and (†) above, we have

\[
\sum_i U_i(\theta_{\min}) \leq \sum_i \int_{\theta_{\min}}^{\theta_{\max}} \theta_i \bar{w}_i(\theta_i) d\theta_i - \sum_i \int_{\theta_{\min}}^{\theta_{\max}} \int_{\theta_{\min}}^{\theta_i} \bar{w}_i(s) ds dF_i(\theta_i)
\]

The formula may look complicated written in its most general form, but everything on the right hand side is a fixed number — we have integrated out everything.

**The constant term in condition (ii)**

Condition (ii) tells us that given the same allocation function \( k(\theta) \), there is only one degree of freedom in the interim stage — the \( U_i(\theta_{\min}) \) or equivalently \( \bar{l}_i(\theta_{\min}) \). You need to know something else about the objective function to be able to say more about the constant.

To elaborate on what we have discussed so far throughout sections, some scenarios (may not be mutually exclusive) depending on the question:

1. Suppose you want ex-ante transfers to sum to 0. This is the best you can do to maximize ex-ante total utility. As we see with finding the bounds on total surplus above, this pins down \( \sum_i (U_i(\theta_{\min})) \), as the inequality in (*) and therefore in (**) both become an equality.

If we have symmetry, for example, when we try to emulate EEM and the players are symmetric, then we know what \( U_i(\theta_{\min}) \) is. Note that symmetry is not necessarily the case under the general scenario.
2. *Suppose you want interim-IR constraints to be satisfied*, meaning, each person should get at least as much as their outside option (at the interim stage) \( o_i(\theta_i) \) at any \( \theta_i \), i.e. \( U_i(\theta_i) \geq o(\theta_i) \) for any \( \theta_i \). The outside option may be getting 0, \( o_i(\theta_i) = 0 \), if you just leave the mechanism empty-handed, or the value of the good \( o_i(\theta_i) = \theta_i \) if you own it to start with, depending on the question.

Note that, it may not be possible to achieve interim-IR for all players, at all \( \theta_i \). You see in lecture 11, the case with a buyer and a seller, in which the seller has property rights, even when we constrain the total surplus to be the highest possible, we cannot achieve the interim-IR constraints for all players, at all types.

3. *Suppose you want to maximize the net transfers you receive* (e.g., you are a seller, rather than a social planner trying to make it 0). Then because you don’t know the types of the players when designing the mechanism, you want to give them the lowest possible utility such that no type will leave at the interim stage. If the outside option is 0, this means making \( U_i(\theta_i) \geq 0 \) for all \( \theta_i \), and since this is increasing, it’s enough to make \( U_i(\theta_{\text{min}}) = 0 \). This is another scenario in which you can pin down \( U_i(\theta_{\text{min}}) \) exactly.

This applies to the case with one agent (adverse selection) and a profit-maximizing seller — this is how we know \( W(0) \) (utility of the lowest type in equilibrium) in that case (which is equivalent to \( U_i(\theta_{\text{min}}) \) in this setup) to be 0.

Each problem you may encounter may be different — the details matter.

Examples from past general exams that belong to the categories above include:

- The setup in the Spring 2012 question on the Interim Payoff in an Auction is Scenario 1 — it’s stated that “this mechanism is symmetric across the bidders and returns all the revenue collected to the bidders, on average.” This is the same as to say that we want ex-ante transfers to sum to 0, and with symmetric transfer functions. Now the interim transfer for the lowest type is higher by exactly one third of the total surplus possible (given the same efficient allocation rule).

- The setup in the Spring 2006 question on Allocating an Innovation is Scenario 3 — after paying the fee ex-ante, the auctioneer wants to maximize the expected revenue in this auction, therefore he will drive down interim value of participating in this auction of the lowest type to 0, which is the same as giving the lowest type a transfer of 0 in the auction.