Lecture 5: Expected Utility Theory

Instructor: Fei Tan

@econdojo @BusinessSchool101 Saint Louis University

Date: October 28, 2025

The Road Ahead

Why Numbers Matter in Game Theory

Challenge: Describing preferences with words is messy and hard to analyze.

Consider four possible outcomes:

- Outcome 3: "I love this"
- Outcome 1: "I like this"
- Outcome 4: "meh"
- Outcome 2: "This is worse than death"

Problems with descriptive language:

- Becomes an "absolute mess to memorize"
- Impossible to manage with many or infinite outcomes
- Cannot perform mathematical analysis

Solution: Use numerical utilities to represent preferences cleanly and compactly

From Preferences to Utilities

Utilities are a numerical system to represent rank-ordered preferences → Larger numbers = better outcomes, smaller numbers = worse outcomes

Preference Description	Assigned Utility
I love outcome 3	0
I like outcome 1	-1
Outcome 4 is meh	-8
Outcome 2 is worse than death	-10

Insight: Numbers instantly capture rank ordering: $0 > -1 > -8 > -10$

Preferences Come First

No one thinks in numbers. We don't care about utilities themselves, we care about what they represent

The one-way street: Preferences map to utilities, not the other way around

1. You have preferences over outcomes (this comes first)
2. We assign utility numbers to represent those preferences
3. Higher utility numbers don't cause preferences - preferences cause higher utility numbers

Example: In modeling international conflict, we first understand what outcomes leaders prefer, then construct utilities to represent those preferences

Axioms of Rational Choice

To predict behavior in strategic situations, we need rules governing preferences. Expected utility theory rests on **four key axioms**:

1. **Completeness**: For any two outcomes, you can state a preference
2. **Transitivity**: Preferences are logically consistent
3. **Independence**: Common components in lotteries don't affect choice
4. **Continuity**: No sudden jumps in preferences

These axioms form the logical bedrock of game theory by ensuring preferences can be represented with utility numbers that allow mathematical analysis.

Completeness Axiom

For any two outcomes X and Y, you must be able to state your preference

Three possibilities:

1. Prefer X to Y
2. Prefer Y to X
3. Be indifferent between X and Y

Example: $1 million, $0, and painful death. Typical preference ordering:

- $1 million > $0
- $0 > Painful death
- $1 million > Painful death

Key point: Completeness is about having **any** preference, not a "sensible" one

Transitivity Axiom

If X is preferred to Y and Y is preferred to Z, then X must be preferred to Z → Mathematical analogy: If A > B and B > C, then A > C

Example: Million dollars vs. $0 vs. dying

1. Prefer $1M to $0 (X > Y)
2. Prefer $0 to dying (Y > Z)
3. Must prefer $1M to dying (X > Z)

Transitivity works with indifference too:

- If indifferent between $1M and $0, and between $0 and dying
- Then must be indifferent between $1M and dying
- Transitivity: If A = B and B = C, then A = C

Key point: Transitivity eliminates illogical preference cycles

Rationality

Everyday rationality: Making sensible, logical choices

Game theory rationality: Having preferences that are complete and transitive

Everyday Rationality	Game Theory Rationality
Judges content of preferences	Only cares about structure
Subjective assessment of wisdom	Mechanical check of consistency
"Is this choice sensible?"	"Can you compare outcomes consistently?"

Insight: A preference for dying over $1 million can be **rational** (if complete and transitive) even if not **sensible**

Dealing with Uncertainty: Lotteries

Lottery: A probability distribution over outcomes. Why it matters:

- Mixed strategies create uncertainty for opponents
- Many real-world situations involve risk
- Need to compare certain outcomes with uncertain ones

Example:

- Option A: Get $0 for certain
- Option B: 50% chance of $1M, 50% chance of death

Key question: How do we compare these options rationally? → Expected utility theory provides the framework for consistent choice under uncertainty

Independence Axiom

When comparing lotteries, identical components should not affect your choice → If you prefer X to Y, then you should prefer:

- [X with probability p, Z with probability (1-p)] to
- [Y with probability p, Z with probability (1-p)]

Example:

- Lottery 1: 50% chance $1M, 50% chance death
- Lottery 2: 50% chance $0, 50% chance death

Analysis: The 50% chance of death is common to both → ignore it and focus on the difference: Do you prefer $1M or $0? If you prefer $1M, then choose Lottery 1

Key point: Independence enables consistent decision-making under uncertainty

Allais Paradox

Choice 1: A vs. B

- **A**: 11% chance of $1M, 89% chance of $0
- **B**: 10% chance of $5M, 90% chance of $0

Choice 2: C vs. D

- **C**: 100% chance of $1M
- **D**: 10% chance of $5M, 89% chance of $1M, 1% chance of $0

Paradox: Many people choose B and C, violating independence

- Both choices reduce to the same core decision
- Prefer certain $1M when certainty available (C over D)
- Prefer risky $5M gamble when risk unavoidable (B over A)

Psychological insight: "Certainty effect" - people overweight guaranteed outcomes

Continuity Axiom

For any three ranked outcomes (best, middle, worst), there exists a probability that makes you indifferent between the middle outcome for certain and a lottery on the best and worst

Example:

- Best: $1 million
- Middle: $0
- Worst: Painful death

What probability p makes you indifferent between:

- Getting $0 for certain
- p chance of $1M, (1-p) chance of death

Key point: As long as such a probability exists (even if p = 0.9999999), your preferences satisfy continuity → continuity rules out lexicographic preferences with infinite jumps

Utility Transformations

Utility numbers are representations, not absolute values. Consider **positive affine transformation**: $u' = au + b$ where $a > 0$

Example: Original Stag Hunt game

	Player 2: Stag	Player 2: Hare
Player 1: Stag	3, 3	1, 2
Player 1: Hare	2, 1	0, 0

Transformed game (multiply Player 1's payoffs by 2):

	Player 2: Stag	Player 2: Hare
Player 1: Stag	6, 3	2, 2
Player 1: Hare	4, 1	0, 0

Key point: Identical equilibria because preference ordering preserved

Rules for Valid Transformations

Three fundamental rules:

1. Use positive affine transformations only: $u' = au + b$ with $a > 0$
   - Never use $a \leq 0$ (reverses or eliminates preferences)
   - Avoid squaring, cubing, or other nonlinear transformations
2. Apply consistently within player: Same a and b for all payoffs of one player
3. Players can be transformed independently:
   - Player 1: $a = 2, b = 0$
   - Player 2: $a = 1, b = -1$
   - Or leave one player unchanged

Bottom line: Preserve preference ordering, maintain strategic equivalence

Pareto Efficiency

An outcome is Pareto efficient if there is no other outcome that makes at least one player better off without making any other player worse off

Example: Battle of the Sexes

	Player 2: Ballet	Player 2: Fight
Player 1: Ballet	1, 2	0, 0
Player 1: Fight	0, 0	100, 1

Player 1's payoffs ÷ 100:

	Player 2: Ballet	Player 2: Fight
Player 1: Ballet	0.01, 2	0, 0
Player 1: Fight	0, 0	1, 1

Key point: Sum of interpersonal utilities gives opposite conclusions, yet both are the same game!

Pareto Efficiency in Stag Hunt

	Player 2: Stag	Player 2: Hare
Player 1: Stag	3, 3	0, 1
Player 1: Hare	1, 0	1, 1

- **Pareto efficient**: (Stag, Stag) only
- **Pareto inefficient**: All other outcomes
- From (Hare, Hare), both players can improve by moving to (Stag, Stag)

Pareto Efficiency in Prisoner's Dilemma

	Player 2: Cooperate	Player 2: Defect
Player 1: Cooperate	3, 3	1, 4
Player 1: Defect	4, 1	2, 2

The unique equilibrium (Defect, Defect) is the only inefficient outcome!

Pareto efficient outcomes:

- (Cooperate, Cooperate) - both can improve from equilibrium
- (Cooperate, Defect) and (Defect, Cooperate) - can't help one without hurting the other

Key lesson: Equilibrium doesn't guarantee efficiency

Risk Preferences

Choose between:

- Lottery 1: 50% chance of $1,000,000, 50% chance of $0
- Lottery 2: Guaranteed payment of $X

Expected value of Lottery 1: (0.5 × $1,000,000) + (0.5 × $0) = $500,000

Question: For what value of $X are you indifferent? Your answer reveals your risk preference:

- **Risk-averse**: $X < $500,000 (prefer certainty)
- **Risk-neutral**: $X = $500,000
- **Risk-acceptant**: $X > $500,000 (prefer the gamble)

Utility Functions

Mathematical representation: $U(x) = x^a$

Risk-neutral ($a = 1$): $U(x) = x$

- Linear relationship between money and utility
- Each dollar provides same additional happiness

Risk-averse ($0 < a < 1$): $U(x) = x^{0.5}$ (square root)

- Diminishing marginal utility
- Each additional dollar provides less happiness

Risk-acceptant ($a > 1$): $U(x) = x^2$

- Increasing marginal utility
- Each additional dollar provides more happiness