Game Theory

Phil Arena

Rationality

Introduction

- Four goals for this session

Introduction

- Four goals for this session
  1. Identify common critiques of "rational choice"
  2. Demonstrate that evidence of "irrationality" is overstated

Introduction

- Four goals for this session
  1. Identify common critiques of "rational choice"
  2. Demonstrate that evidence of "irrationality" is overstated
  3. Compare and contrast different views of human behavior

Introduction

- Four goals for this session
  1. Identify common critiques of "rational choice"
  2. Demonstrate that evidence of "irrationality" is overstated
  3. Compare and contrast different views of human behavior
  4. Discuss evidence from experiments

Critiques of "Rational Choice"

- "People are not rational"

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical
  - Can be restated as HPWC criterion

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical
  - Can be restated as HPWC criterion
  - A "rational" person would satisfy this criterion

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical
  - Can be restated as HPWC criterion
  - A "rational" person would satisfy this criterion
  - In this sense, and only this sense, we must assume rationality

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical
  - Can be restated as HPWC criterion
  - A "rational" person would satisfy this criterion
  - In this sense, and only this sense, we must assume rationality
  - Diversity of assumptions about what people want w/in "RC"

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical
  - Can be restated as HPWC criterion
  - A "rational" person would satisfy this criterion
  - In this sense, and only this sense, we must assume rationality
  - Diversity of assumptions about what people want w/in "RC"
- Minimalist interpretation non-falsifiable

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical
  - Can be restated as HPWC criterion
  - A "rational" person would satisfy this criterion
  - In this sense, and only this sense, we must assume rationality
  - Diversity of assumptions about what people want w/in "RC"
- Minimalist interpretation non-falsifiable
  - Allais Paradox purports to show evidence that it is false

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical
  - Can be restated as HPWC criterion
  - A "rational" person would satisfy this criterion
  - In this sense, and only this sense, we must assume rationality
  - Diversity of assumptions about what people want w/in "RC"
- Minimalist interpretation non-falsifiable
  - Allais Paradox purports to show evidence that it is false
  - Goal is not to test underlying assumptions, but hypotheses

Critiques of "Rational Choice"

- "People are not rational"
  - We require preferences to be complete, reflexive, and acyclical
  - Can be restated as HPWC criterion
  - A "rational" person would satisfy this criterion
  - In this sense, and only this sense, we must assume rationality
  - Diversity of assumptions about what people want w/in "RC"
- Minimalist interpretation non-falsifiable
  - Allais Paradox purports to show evidence that it is false
  - Goal is not to test underlying assumptions, but hypotheses
  - All research relies on non-falsifiable assumptions

Allais Paradox

	A	B
Experiment 1	pr($1 mil)= 1.00	pr($1 mil)= 0.89 pr($5 mil)= 0.10 pr($0)= 0.01
Experiment 2	pr($0)= 0.89 pr($1 mil)= 0.11	pr($0)= 0.90 pr($5 mil)= 0.10

Allais Paradox

	A	B
Experiment 1	pr($1 mil)= 1.00	pr($1 mil)= 0.89 pr($5 mil)= 0.10 pr($0)= 0.01
Experiment 2	pr($0)= 0.89 pr($1 mil)= 0.11	pr($0)= 0.90 pr($5 mil)= 0.10

- Majority of participants chose A in Experiment 1, B in 2

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B
  - A risk-averse actor strictly prefers A

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B
  - A risk-averse actor strictly prefers A
  - A risk-neutral actor values both equally

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B
  - A risk-averse actor strictly prefers A
  - A risk-neutral actor values both equally
- If one or more actions entails risk, we must calculate the expected utility of that action to identify optimal behavior

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
    - A always yields the same outcome, worth x
    - B yields zx with probability $\frac{1}{z}$, 0 otherwise
    - A risk-acceptant actor strictly prefers B
    - A risk-averse actor strictly prefers A
    - A risk-neutral actor values both equally

- If one or more actions entails risk, we must calculate the expected utility of that action to identify optimal behavior
    - Suppose there are N outcomes associated with action B

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B
  - A risk-averse actor strictly prefers A
  - A risk-neutral actor values both equally

- If one or more actions entails risk, we must calculate the expected utility of that action to identify optimal behavior
  - Suppose there are N outcomes associated with action B
  - Let p_k be the probability of outcome x_k for any $k \in \{1, 2 \ldots N\}$

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B
  - A risk-averse actor strictly prefers A
  - A risk-neutral actor values both equally

- If one or more actions entails risk, we must calculate the expected utility of that action to identify optimal behavior
  - Suppose there are N outcomes associated with action B
  - Let p_k be the probability of outcome x_k for any $k \in \{1, 2 \ldots N\}$
  - Let $u_i(x_k)$ be i's subjective utility for outcome x_k

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B
  - A risk-averse actor strictly prefers A
  - A risk-neutral actor values both equally

- If one or more actions entails risk, we must calculate the expected utility of that action to identify optimal behavior
  - Suppose there are N outcomes associated with action B
  - Let p_k be the probability of outcome x_k for any $k \in \{1, 2 \ldots N\}$
  - Let $u_i(x_k)$ be i's subjective utility for outcome x_k
  - Let τ_i be i's attitude towards risk

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B
  - A risk-averse actor strictly prefers A
  - A risk-neutral actor values both equally

- If one or more actions entails risk, we must calculate the expected utility of that action to identify optimal behavior
  - Suppose there are N outcomes associated with action B
  - Let p_k be the probability of outcome x_k for any $k \in \{1, 2 \ldots N\}$
  - Let $u_i(x_k)$ be i's subjective utility for outcome x_k
  - Let τ_i be i's attitude towards risk
  - i's expected utility of action B is denoted $E(u_i(B))$

Risk-Attitudes and Expected Utility

- Suppose an actor faces a choice between actions A and B
  - A always yields the same outcome, worth x
  - B yields zx with probability $\frac{1}{z}$, 0 otherwise
  - A risk-acceptant actor strictly prefers B
  - A risk-averse actor strictly prefers A
  - A risk-neutral actor values both equally
- If one or more actions entails risk, we must calculate the expected utility of that action to identify optimal behavior
  - Suppose there are N outcomes associated with action B
  - Let p_k be the probability of outcome x_k for any $k \in \{1, 2 \dots N\}$
  - Let $u_i(x_k)$ be i's subjective utility for outcome x_k
  - Let τ_i be i's attitude towards risk
  - i's expected utility of action B is denoted $E(u_i(B))$
  - $E(u_i(B)) = \sum_{k=1}^{N} p_k^{\tau} u_i(x_k)$

Numerical Examples

- Return to Allais Experiment 1

Numerical Examples

- Return to Allais Experiment 1
  - A always yields 1

Numerical Examples

- Return to Allais Experiment 1
  - A always yields 1
  - $E(u_i(B)) = (0.89)^{\tau}(1) + (0.10)^{\tau}(5) + (0.01)^{\tau}(0)$

Numerical Examples

- Return to Allais Experiment 1
  - A always yields 1
  - $E(u_i(B)) = (0.89)^\tau(1) + (0.10)^\tau(5) + (0.01)^\tau(0)$
  - Iff $\tau \geq \hat{\tau}_1$, i prefers A, where $\hat{\tau}_1 \approx 1.5$

Numerical Examples

- Return to Allais Experiment 1
  - A always yields 1
  - $E(u_i(B)) = (0.89)^\tau(1) + (0.10)^\tau(5) + (0.01)^\tau(0)$
  - Iff $\tau \geq \hat{\tau}_1$, i prefers A, where $\hat{\tau}_1 \approx 1.5$
- Return to Allais Experiment 2

Numerical Examples

- Return to Allais Experiment 1
  - A always yields 1
  - $E(u_i(B)) = (0.89)^\tau(1) + (0.10)^\tau(5) + (0.01)^\tau(0)$
  - Iff $\tau \geq \hat{\tau}_1$, i prefers A, where $\hat{\tau}_1 \approx 1.5$
- Return to Allais Experiment 2
  - $E(u_i(A)) = (0.89)^\tau(0) + (0.11)^\tau(1)$

Numerical Examples

- Return to Allais Experiment 1
  - A always yields 1
  - $E(u_i(B)) = (0.89)^\tau(1) + (0.10)^\tau(5) + (0.01)^\tau(0)$
  - Iff $\tau \geq \hat{\tau}_1$, i prefers A, where $\hat{\tau}_1 \approx 1.5$
- Return to Allais Experiment 2
  - $E(u_i(A)) = (0.89)^\tau(0) + (0.11)^\tau(1)$
  - $E(u_i(B)) = (0.90)^\tau(0) + (0.10)^\tau(5)$

Numerical Examples

- Return to Allais Experiment 1
  - A always yields 1
  - $E(u_i(B)) = (0.89)^\tau(1) + (0.10)^\tau(5) + (0.01)^\tau(0)$
  - Iff $\tau \geq \hat{\tau}_1$, i prefers A, where $\hat{\tau}_1 \approx 1.5$
- Return to Allais Experiment 2
  - $E(u_i(A)) = (0.89)^\tau(0) + (0.11)^\tau(1)$
  - $E(u_i(B)) = (0.90)^\tau(0) + (0.10)^\tau(5)$
  - Iff $\tau \geq \hat{\tau}_2$, i prefers A, where $\hat{\tau}_2 \approx 17$

What Paradox?

- If we take results at face value

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$
  - Still no evidence of cyclical preferences

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$
  - Still no evidence of cyclical preferences

- Should we take them at face value?

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$
  - Still no evidence of cyclical preferences
- Should we take them at face value?
  - Suppose S risk-neutral students participate in experiments

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$
  - Still no evidence of cyclical preferences
- Should we take them at face value?
  - Suppose S risk-neutral students participate in experiments
  - Each $s_i \in S$ receives $\Psi_i \geq 0$ for getting right answer

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$
  - Still no evidence of cyclical preferences

- Should we take them at face value?
  - Suppose S risk-neutral students participate in experiments
  - Each $s_i \in S$ receives $\Psi_i \geq 0$ for getting right answer
  - If s_i pays $c_{i,x} > 0$, learns $E(\$)$ for $x \in \{1B, 2A, 2B\}$

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$
  - Still no evidence of cyclical preferences
- Should we take them at face value?
  - Suppose S risk-neutral students participate in experiments
  - Each $s_i \in S$ receives $\Psi_i \geq 0$ for getting right answer
  - If s_i pays $c_{i,x} > 0$, learns $E(\$)$ for $x \in \{1B, 2A, 2B\}$
  - Otherwise, s_i must rely on some guess, denoted $\phi_{i,x}$

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$
  - Still no evidence of cyclical preferences
- Should we take them at face value?
  - Suppose S risk-neutral students participate in experiments
  - Each $s_i \in S$ receives $\Psi_i \geq 0$ for getting right answer
  - If s_i pays $c_{i,x} > 0$, learns $E(\$)$ for $x \in \{1B, 2A, 2B\}$
  - Otherwise, s_i must rely on some guess, denoted $\phi_{i,x}$
  - EQ for Ex1: iff $\Psi_i \geq c_{i,1B}$, calculate $E(\$)$ for B then select B; otherwise select A if $\phi_{i,1B}$ <\$1 mil and B if $\phi_{i,1B}$ >\$1 mil

What Paradox?

- If we take results at face value
  - Allais found that average $\tau \in (1.5, 17)$
  - Still no evidence of cyclical preferences

- Should we take them at face value?
  - Suppose S risk-neutral students participate in experiments
  - Each $s_i \in S$ receives $\Psi_i \geq 0$ for getting right answer
  - If s_i pays $c_{i,x} > 0$, learns $E(\$)$ for $x \in \{1B, 2A, 2B\}$
  - Otherwise, s_i must rely on some guess, denoted $\phi_{i,x}$
  - EQ for Ex1: iff $\Psi_i \geq c_{i,1B}$, calculate $E(\$)$ for B then select B; otherwise select A if $\phi_{i,1B} <$\$1 mil and B if $\phi_{i,1B} >$\$1 mil
  - EQ for Ex2: iff $\Psi_i \geq c_{i,2A} + c_{i,2B}$, calculate $E(\$)$ for both then select B; otherwise select A iff $\phi_{i,2A} \geq \phi_{i,2B}$

Alternative Views of Human Behavior

Stylized View	Greatest desire
Homo Economicus	Material wealth
Homo Biologicus	Healthy offspring
Homo Politicus	Control over policy
Homo Sociologicus	Respect/prestige
Homo Philosophicus	Virtue

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share
  - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1 - x)\pi$

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
    - Some pie of size π is to be divided b/w 1 & 2
    - 1 proposes a division, expressed as 1's share
    - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1 - x)\pi$
    - If 2 rejects, neither receives anything

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share
  - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1-x)\pi$
  - If 2 rejects, neither receives anything
  - If 1 & 2 both *H.E.*, 1 sets $x = 1$ & 2 accepts

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share
  - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1-x)\pi$
  - If 2 rejects, neither receives anything
  - If 1 & 2 both *H.E.*, 1 sets $x = 1$ & 2 accepts
- Deviations from pure *H.E.* abound

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share
  - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1-x)\pi$
  - If 2 rejects, neither receives anything
  - If 1 & 2 both *H.E.*, 1 sets $x = 1$ & 2 accepts
- Deviations from pure *H.E.* abound
  - x very rarely set to 1, but usually not 0.5 either

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share
  - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1-x)\pi$
  - If 2 rejects, neither receives anything
  - If 1 & 2 both *H.E.*, 1 sets $x = 1$ & 2 accepts
- Deviations from pure *H.E.* abound
  - x very rarely set to 1, but usually not 0.5 either
  - If choice completely anonymous, 40% set $x = 1$

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share
  - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1 - x)\pi$
  - If 2 rejects, neither receives anything
  - If 1 & 2 both *H.E.*, 1 sets $x = 1$ & 2 accepts
- Deviations from pure *H.E.* abound
  - x very rarely set to 1, but usually not 0.5 either
  - If choice completely anonymous, 40% set $x = 1$
  - If taking is an option in addition to giving, most take

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share
  - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1-x)\pi$
  - If 2 rejects, neither receives anything
  - If 1 & 2 both *H.E.*, 1 sets $x = 1$ & 2 accepts
- Deviations from pure *H.E.* abound
  - x very rarely set to 1, but usually not 0.5 either
  - If choice completely anonymous, 40% set $x = 1$
  - If taking is an option in addition to giving, most take
  - Behave more like *H.E.* when 2 is a computer

Evidence from Experiments

- Many have sought to disprove *H.E.* experimentally
  - Some pie of size π is to be divided b/w 1 & 2
  - 1 proposes a division, expressed as 1's share
  - If 2 accepts, 1 receives $x\pi$ and 2 receives $(1 - x)\pi$
  - If 2 rejects, neither receives anything
  - If 1 & 2 both *H.E.*, 1 sets $x = 1$ & 2 accepts
- Deviations from pure *H.E.* abound
  - x very rarely set to 1, but usually not 0.5 either
  - If choice completely anonymous, 40% set $x = 1$
  - If taking is an option in addition to giving, most take
  - Behave more like *H.E.* when 2 is a computer
  - Crossing the street to avoid the homeless person