Reasoning in Games

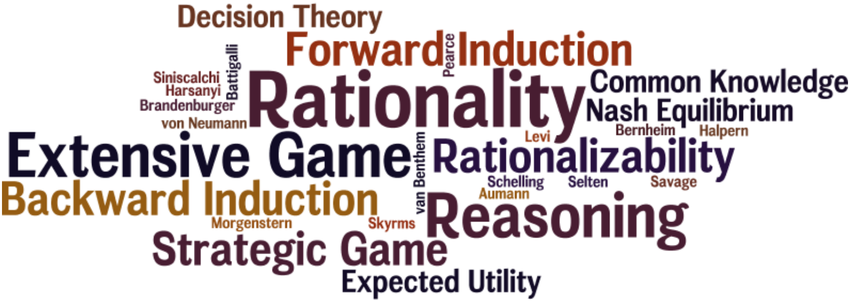
Lecture 1

Eric Pacuit

University of Maryland, College Park
pacuit.org
epacuit@umd.edu

August 10, 2015

Topics


Decision Theory
Forward Induction
Pearce
Battigalli
Siniscalchi
Harsanyi
Brandenburger
von Neumann
Rationality
Common Knowledge
Nash Equilibrium
Bernheim
Halpern
Levi
Extensive Game
van Benthem
Rationalizability
Backward Induction
Schelling
Selten
Savage
Aumann
Morgenstern
Skyrms
Reasoning
Strategic Game
Expected Utility

Information

Email:	epacuit@umd.edu
Website:	pacuit.org/esslli2015/reasgames
Reading:	plato.stanford.edu/entries/epistemic-game
Game Theory:	www.game-theory-class.org

Plan

- Day 1: From Decision Theory to Game Theory
- Day 2: Games and Game Models
- Day 3: Modeling Deliberation (in Games)
- Day 4: Backward and Forward Induction
- Day 5: Spill Over, Concluding Remarks (Language-Based Games/ Variable Frame Theory, Behavioral Game Theory, ...)

The Guessing Game

Guess a number between 1 & 100.
The closest to 2/3 of the average wins.

dev.pacuit.org/games/avg

The Guessing Game, again

Guess a number between 1 & 100.
The closest to 2/3 of the average wins.

The Guessing Game, again

Guess a number between 1 & 100.
The closest to 2/3 of the average wins.

dev.pacuit.org/games/avg

Traveler's Dilemma

1. You and your friend write down an integer between 2 and 100 (without discussing).

Traveler's Dilemma

1. You and your friend write down an integer between 2 and 100 (without discussing).
2. If both of you write down the same number, then both will receive that amount in Euros from the airline in compensation.

Traveler's Dilemma

1. You and your friend write down an integer between 2 and 100 (without discussing).
2. If both of you write down the same number, then both will receive that amount in Euros from the airline in compensation.
3. If the numbers are different, then the airline assumes that the smaller number is the actual price of the luggage.

Traveler's Dilemma

1. You and your friend write down an integer between 2 and 100 (without discussing).
2. If both of you write down the same number, then both will receive that amount in Euros from the airline in compensation.
3. If the numbers are different, then the airline assumes that the smaller number is the actual price of the luggage.
4. The person that wrote the smaller number will receive that amount plus 2 EUR (as a reward), and the person that wrote the larger number will receive the smaller number minus 2 EUR (as a punishment).

Traveler's Dilemma

1. You and your friend write down an integer between 2 and 100 (without discussing).
2. If both of you write down the same number, then both will receive that amount in Euros from the airline in compensation.
3. If the numbers are different, then the airline assumes that the smaller number is the actual price of the luggage.
4. The person that wrote the smaller number will receive that amount plus 2 EUR (as a reward), and the person that wrote the larger number will receive the smaller number minus 2 EUR (as a punishment).

Suppose that you are randomly paired with another person here at ESSLLI. What number would you write down?

dev.pacuit.org/games/td

Decision Theory

Decision Theory

Rational decision making is associated with both the capacity to order outcomes *and* to choose from the *top* of the order.

Are Walter's decisions *rational*?

Are Walter's decisions *rational*?

- What are his preferences?
- What does he believe?
- What type of choice problem is he facing?

Context of a decision

Context of a decision

Individual decision-making
(against nature)

- E.g., Gambling

Context of a decision

Individual decision-making (against nature)

- E.g., Gambling

Individual decision making in interaction

- E.g., Playing chess

Context of a decision

Individual decision-making (against nature)

- E.g., Gambling

Individual decision making in interaction

- E.g., Playing chess

Collective decision making

- E.g., Carrying a piano

Context of a decision

Individual decision-making (against nature)

- E.g., Gambling

Individual decision making in interaction

- E.g., Playing chess
- E.g., Voting in an election

Collective decision making

- E.g., Carrying a piano
- E.g., Voting in an election

Preference, Choice, and Utility

- Representing *preferences*: relations, preference axioms
- *Revealed* preference theory: WARP, Sen's α and β, Revelation Theorem
- *Utility*: Ordinal vs. cardinal utility, interval scale, ratio scale
- *Expected utility theory*: (probability), von Neumann-Morgenstern Theorem, Savage's Theorem, Aumann-Anscombe's Decision Theory, Allais paradox, Ellsberg paradox, ...
- Interpersonal comparison of utilities

Decision Problems

In many circumstances the decision maker doesn't get to choose outcomes directly, but rather chooses an instrument that affects what outcome actually occurs.

Decision Problems

In many circumstances the decision maker doesn't get to choose outcomes directly, but rather chooses an instrument that affects what outcome actually occurs.

Choice under

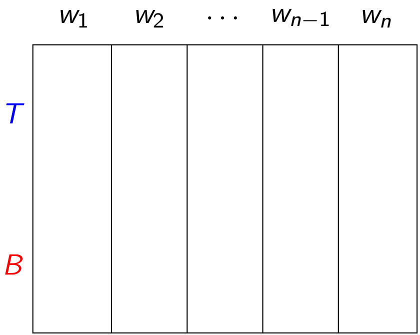
- *certainty*: highly confident about the relationship between actions and outcomes
- *risk*: clear sense of possibilities and their likelihoods
- *uncertainty*: the relationship between actions and outcomes is so imprecise that it is not possible to assign likelihoods

Decision Problems

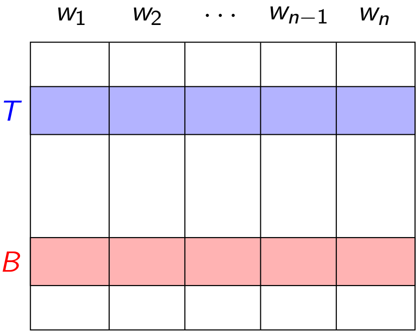
T

B

Decision Problems

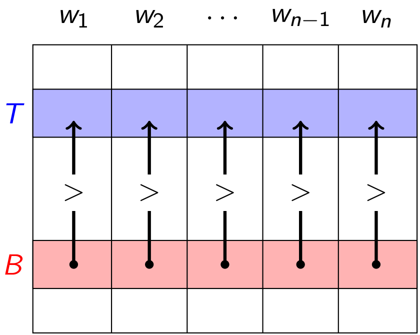


Decision Problems



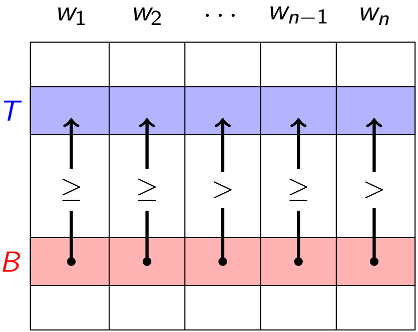
An **act** is a function $A : W \to O$

Strict Dominance



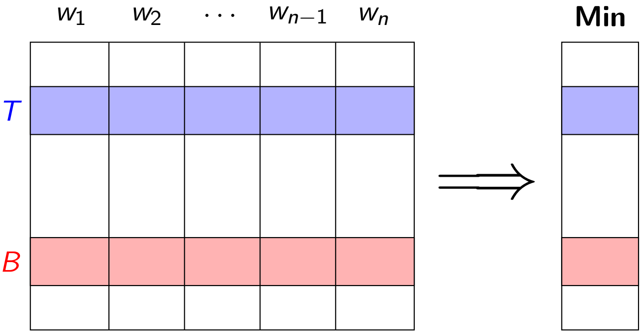
$$\forall\ w \in W,\ u(A(w)) > u(B(w))$$

Weak Dominance



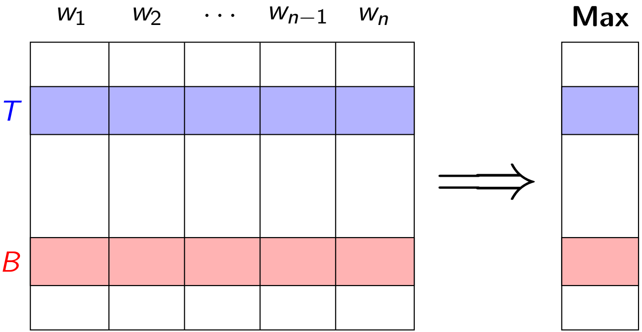
$$\forall\, w \in W,\ u(A(w)) \geq u(B(w)) \text{ and } \exists\, w \in W,\ u(A(w)) > u(B(w))$$

MaxMin (Security)



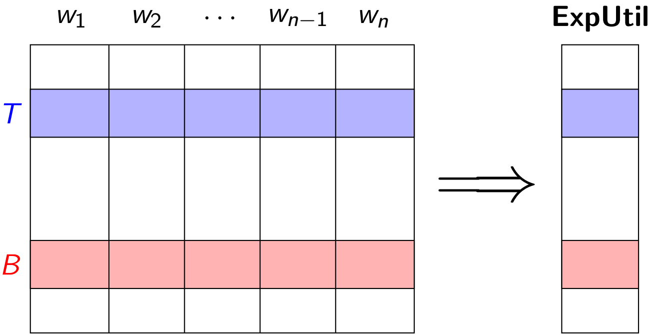
$$\min(\{A(w) \mid w \in W\})$$

MaxMax



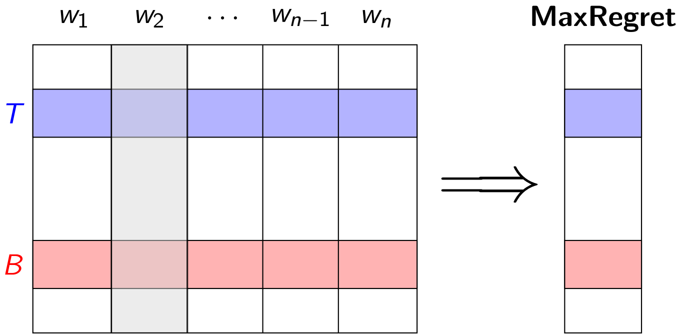
$$\max(\{A(w) \mid w \in W\})$$

Maximize (Subjective) Expected Utility



$$\sum_{w \in W} P_A(w) * u(A(w))$$

MinMax Regret



$$\max(\{u(A(w_i)) - \max(\{u(A_i(w_i)) \mid A_i \in \text{Act}\})$$

- Maximizing
- Utility/Preferences
- Probability

Maximizing

"The formulation of maximizing behavior in economics has often paralleled the modeling of maximization in physics and related disciplines.

Maximizing

"The formulation of maximizing behavior in economics has often paralleled the modeling of maximization in physics and related disciplines. But maximizing *behavior* differs from nonvolitional *maximization* because of the fundamental relevance of the choice act, which has to be placed in a central position in analyzing maximizing behavior.

Maximizing

"The formulation of maximizing behavior in economics has often paralleled the modeling of maximization in physics and related disciplines. But maximizing *behavior* differs from nonvolitional *maximization* because of the fundamental relevance of the choice act, which has to be placed in a central position in analyzing maximizing behavior. A person's preferences over *comprehensive* outcomes (including the choice process) have to be distinguished from the conditional preferences over *culmination* outcomes *given* the act of choice." (pg. 745)

A. Sen. *Maximization and the Act of Choice*. Econometrica, Vol. 65, No. 4, 1997, 745 - 779.

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it.

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it. You select a “less preferred” chair.

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it. You select a "less preferred" chair. Are you still a maximizer?

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it. You select a "less preferred" chair. Are you still a maximizer? Quite possibly you are, since your preference ranking for choice behavior may well be defined over "comprehensive outcomes", including choice processes (in particular, who does the choosing) as well as the outcomes at culmination (the distribution of chairs).

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it. You select a "less preferred" chair. Are you still a maximizer? Quite possibly you are, since your preference ranking for choice behavior may well be defined over "comprehensive outcomes", including choice processes (in particular, who does the choosing) as well as the outcomes at culmination (the distribution of chairs).

To take another example, you may prefer mangoes to apples, but refuse to pick the last mango from a fruit basket, and yet be very pleased if someone else were to "force" that last mango on you. " (Sen, pg. 747)

Preferences

Preferences

Preferring or choosing x is different than "liking" x or "having a taste for x": one can prefer x to y but *dislike* both options

Preferences

Preferring or choosing x is different than “liking” x or “having a taste for x”: one can prefer x to y but *dislike* both options

Preferences are always understood as comparative: “preference” is more like “bigger” than “big”

Preferences

Preferring or choosing x is different than "liking" x or "having a taste for x": one can prefer x to y but *dislike* both options

Preferences are always understood as comparative: "preference" is more like "bigger" than "big"

Revealed Preferences: Ann is said to have a preference for x over y iff Ann chooses x over y where choice is conceived of as overt behavior.

Preferences

Preferring or choosing x is different than "liking" x or "having a taste for x": one can prefer x to y but *dislike* both options

Preferences are always understood as comparative: "preference" is more like "bigger" than "big"

Revealed Preferences: Ann is said to have a preference for x over y iff Ann chooses x over y where choice is conceived of as overt behavior.

Deliberative Preferences: A person deliberates and (ideally) ranks all the possible "outcomes"

Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

1. $x \succeq y$ and $y \not\succeq x$: The agent *strictly prefers* x to y ($x \succ y$)

Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

1. $x \succeq y$ and $y \not\succeq x$: The agent *strictly prefers* x to y ($x \succ y$)
2. $y \succeq x$ and $x \not\succeq y$: The agent *strictly prefers* y to x ($y \succ x$)

Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

1. $x \succeq y$ and $y \not\succeq x$: The agent *strictly prefers* x to y ($x \succ y$)
2. $y \succeq x$ and $x \not\succeq y$: The agent *strictly prefers* y to x ($y \succ x$)
3. $x \succeq y$ and $y \succeq x$: The agent is *indifferent* between x and y ($x \approx y$)

Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

1. $x \succeq y$ and $y \not\succeq x$: The agent *strictly prefers* x to y ($x \succ y$)
2. $y \succeq x$ and $x \not\succeq y$: The agent *strictly prefers* y to x ($y \succ x$)
3. $x \succeq y$ and $y \succeq x$: The agent is *indifferent* between x and y ($x \approx y$)
4. $x \not\succeq y$ and $y \not\succeq x$: The agent *cannot compare* x and y ($x \perp y$)

Utility Functions

A **utility function** on a set X is a function $u : X \to \Re$

Utility Functions

A **utility function** on a set X is a function $u : X \to \Re$

A utility function $u : X \to \Re$ **represents** an ordering $\succeq$ on X provided for all $x, y \in X$, $x \succeq y$ iff $u(x) \geq u(y)$.

Ordinal Utility Theory: Axioms

Ordinal Utility Theory: Axioms

1. **Completeness**: The preference ordering is complete: the decision maker call always rank options (for any two options x and y, either the decision maker (1) strictly prefers x to y, (2) strictly prefers y to x or (3) is indifferent between x and y).

Ordinal Utility Theory: Axioms

1. **Completeness**: The preference ordering is complete: the decision maker call always rank options (for any two options x and y, either the decision maker (1) strictly prefers x to y, (2) strictly prefers y to x or (3) is indifferent between x and y).

2. **Transitivity**: Weak preference (and hence strict and indifference) is transitive

Ordinal Utility Theory

Theorem. Suppose that X is finite and $\succeq$ is a complete and transitive ordering over X. Then there is a utility function $u : X \to \mathfrak{R}$ that represents $\succeq$.

Cardinal Utility Theory

$x \succ y \succ z$ is represented by both $(3, 2, 1)$ and $(1000, 999, 1)$, so we cannot say y whether is "closer" to x than to z.

Cardinal Utility Theory

$x \succ y \succ z$ is represented by both $(3, 2, 1)$ and $(1000, 999, 1)$, so we cannot say y whether is "closer" to x than to z.

Key idea: Ordinal preferences over *lotteries* allows us to infer a cardinal scale (with some additional axioms).

John von Neumann and Oskar Morgenstern. *The Theory of Games and Economic Behavior*. Princeton University Press, 1944.

A Choice

R

B

W

S

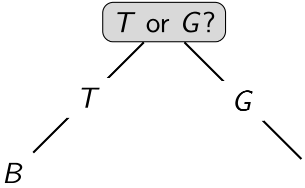
A Choice

R

B

W

S



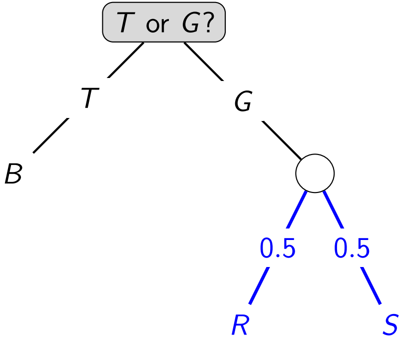
A Choice

R

B

W

S



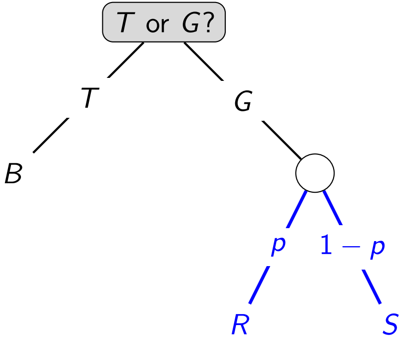
A Choice

R

B

W

S



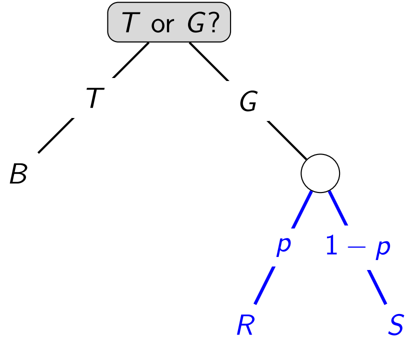
A Choice

R

B

W

S



$$[1 : B] \sim [p : R, 1-p : S]$$

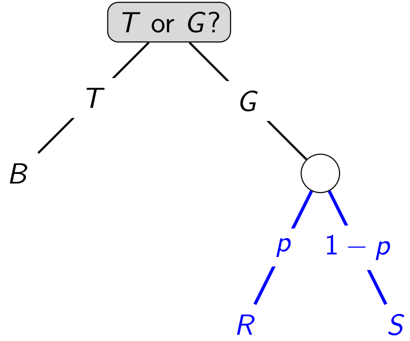
A Choice

R

B

W

S



$$1 * u(B) = p * u(R) + (1 - p) * u(S)$$

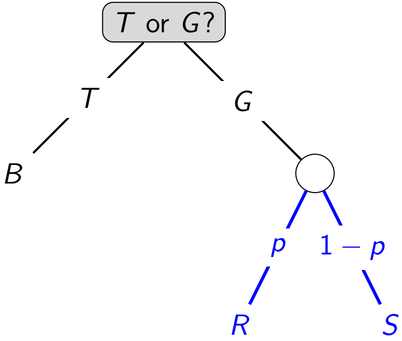
A Choice

R

B

W

S



$$u(B) = p * 1 + (1 - p) * 0 = p$$

Suppose that X is a set of outcomes.

A **(simple) lottery** over X is denoted $[p_1 : x_1, p_2 : x_2, \ldots, p_n : x_n]$ where for $i = 1, \ldots, n$, $x_i \in X$ and $p_i \in [0, 1]$, and $\sum_i p_i = 1$.

Let $\mathcal{L}$ be the set of (simple) lotteries over X. We identify elements $x \in X$ with the lottery $[1 : x]$.

Axioms

Preference $\succeq$ is reflexive, transitive and complete

Compound Lotteries The decision maker is indifferent between every compound lottery and the *corresponding* simple lottery.

Independence For all $L_1, L_2, L_3 \in \mathcal{L}$ and $a \in (0,1]$, $L_1 \succ L_2$
if, and only if,
$[L_1 : a, L_3 : (1-a)] \succ [L_2 : a, L_3 : (1-a)]$.

Continuity For all $L_1, L_2, L_3 \in \mathcal{L}$ and $a \in (0,1]$,
if $L_1 \succ L_2 \succ L_3$, then there exists $a, b \in (0,1)$ such that $[L_1 : a, L_3 : (1-a)] \succ L_2$ and $L_2 \succ [L_1 : b, L_3 : (1-b)]$.

$u : \mathcal{L} \to \Re$ is linear provided for all $L = [L_1 : p_1, \ldots, L_n : p_n] \in \mathcal{L}$,

$$u(L) = \sum_{i=1}^{n} p_i u(L_i)$$

von Neumann-Morgenstern Representation Theorem A binary relation $\succeq$ on $\mathcal{L}$ satisfies Preference, Compound Lotteries, Independence and Continuity iff $\succeq$ is representable by a linear utility function $u : \mathcal{L} \to \Re$.

Moreover, $u' : \mathcal{L} \to \Re$ represents $\succeq$ iff there exists real numbers $c > 0$ and d such that $u'(\cdot) = cu(\cdot) + d$. ("$u$ is unique up to linear transformations.")

Savage

Savage derives both a decision maker's utilities *and* probabilities from preferences over acts (a Savage act is a function from states to outcomes).

Difficulties

- Attitudes towards risk: the Allais Paradox
- Rabin's Theorem: the fact that people tend to avoid lotteries $[-\$100 : 0.5, \$110 : 0.5]$ is very hard to square with standard expected utility theory
- Ambiguity aversion: the Ellsberg Paradox
- Kahneman and Tversky: Framing, loss aversion, prospect theory
- Savage/Causal/Evidential Decision Theory: Newcomb's Paradox

Dominance Reasoning and Act-State Dependence

	w_1	w_2
A	1	3
B	2	4

Dominance Reasoning and Act-State Dependence

	w_1	w_2
A	1	3
B	2	4

Dominance Reasoning and Act-State Dependence

Dominance reasoning is appropriate only when probability of outcome is *independent of choice*.

(A nasty nephew wants inheritance from his rich Aunt. The nephew wants the inheritance, but other things being equal, does not want to apologize. Does dominance give the nephew a reason to not apologize? *Whether or not the nephew is cut from the will may depend on whether or not he apologizes*.)

Newcomb's Paradox

A very powerful being, who has been invariably accurate in his predictions about your behavior in the past, has already acted in the following way:

1. If he has predicted that you will open just box B, he has in addition put \$1,000,000 in box B
2. If he has predicted you will open both boxes, he has put nothing in box B.

What should you do?

R. Nozick. *Newcomb's Problem and Two Principles of Choice*. 1969.

Newcomb's Paradox

	B = 1M	B = 0
1 Box	1M	0
2 Boxes	1M + 1000	1000

Newcomb's Paradox

	B = 1M	B = 0
1 Box	1M	0
2 Boxes	1M + 1000	1000

	B = 1M	B = 0
1 Box	h	$1 - h$
2 Boxes	$1 - h$	h

Newcomb's Paradox

J. Collins. *Newcomb's Problem*. International Encyclopedia of Social and Behavorial Sciences, 1999.

Newcomb's Paradox

There is a conflict between maximizing your expected value (1-box choice) and dominance reasoning (2-box choice).

Newcomb's Paradox

There is a conflict between maximizing your expected value (1-box choice) and dominance reasoning (2-box choice).

What the Predictor did yesterday is *probabilistically dependent* on the choice today, but *causally independent* of today's choice.

$V(A) = \sum_w V(w) \cdot P_A(w)$
(the expected value of act A is a probability weighted average of the values of the ways w in which A might turn out to be true)

$V(A) = \sum_w V(w) \cdot P_A(w)$
(the expected value of act A is a probability weighted average of the values of the ways w in which A might turn out to be true)

Orthodox Bayesian Decision Theory: $P_A(w) := P(w \mid A)$ (Probability of w given A is chosen)

Causal Decision theory: $P_A(w) = P(A \,\Box\!\!\rightarrow w)$ (Probability of *if A were chosen then w would be true*)

Suppose 99% confidence in predictors reliability.

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

Suppose 99% confidence in predictors reliability.

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1)$$

Suppose 99% confidence in predictors reliability.

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1) =$
$1000000 \cdot 0.99 + 0 \cdot 0.01$

Suppose 99% confidence in predictors reliability.

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1) =$
$1000000 \cdot 0.99 + 0 \cdot 0.01 = 990,000$

Suppose 99% confidence in predictors reliability.

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1) = 1000000 \cdot 0.99 + 0 \cdot 0.01 = 990,000$$

$$V(B_2) = V(L)P(L \mid B_2) + V(K)P(K \mid B_2)$$

Suppose 99% confidence in predictors reliability.

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1) =$
$1000000 \cdot 0.99 + 0 \cdot 0.01 = 990,000$

$V(B_2) = V(L)P(L \mid B_2) + V(K)P(K \mid B_2) =$
$1001000 \cdot 0.01 + 1000 \cdot 0.99$

Suppose 99% confidence in predictors reliability.

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1) = 1000000 \cdot 0.99 + 0 \cdot 0.01 = 990,000$$

$$V(B_2) = V(L)P(L \mid B_2) + V(K)P(K \mid B_2) = 1001000 \cdot 0.01 + 1000 \cdot 0.99 = 11,000$$

Let μ be the assigned to the conditional $B_1 \Box\!\!\rightarrow M$ (and $B_2 \Box\!\!\rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

Let μ be the assigned to the conditional $B_1 \mathrel{\Box\!\!\rightarrow} M$ (and $B_2 \mathrel{\Box\!\!\rightarrow} L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \mathrel{\Box\!\!\rightarrow} M) + V(N)P(B_1 \mathrel{\Box\!\!\rightarrow} N)$$

Let μ be the assigned to the conditional $B_1 \mathrel{\Box\!\!\rightarrow} M$ (and $B_2 \mathrel{\Box\!\!\rightarrow} L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$V(B_1) = V(M)P(B_1 \mathrel{\Box\!\!\rightarrow} M) + V(N)P(B_1 \mathrel{\Box\!\!\rightarrow} N) =$
$1000000 \cdot \mu + 0 \cdot 1 - \mu$

Let μ be the assigned to the conditional $B_1 \Box\!\!\rightarrow M$ (and $B_2 \Box\!\!\rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \Box\!\!\rightarrow M) + V(N)P(B_1 \Box\!\!\rightarrow N) =$$
$$1000000 \cdot \mu + 0 \cdot 1 - \mu = 1000000\mu$$

Let μ be the assigned to the conditional $B_1 \Box\!\!\rightarrow M$ (and $B_2 \Box\!\!\rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \Box\!\!\rightarrow M) + V(N)P(B_1 \Box\!\!\rightarrow N) =$$
$$1000000 \cdot \mu + 0 \cdot 1 - \mu = 1000000\mu$$

$$V(B_2) = V(L)P(B_2 \Box\!\!\rightarrow L) + V(K)P(B_2 \Box\!\!\rightarrow K)$$

Let μ be the assigned to the conditional $B_1 \,\Box\!\!\rightarrow M$ (and $B_2 \,\Box\!\!\rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \,\Box\!\!\rightarrow M) + V(N)P(B_1 \,\Box\!\!\rightarrow N) =$$
$$1000000 \cdot \mu + 0 \cdot 1 - \mu = 1000000\mu$$

$$V(B_2) = V(L)P(B_2 \,\Box\!\!\rightarrow L) + V(K)P(B_2 \,\Box\!\!\rightarrow K) =$$
$$1001000 \cdot \mu + 1000 \cdot 1 - \mu$$

Let μ be the assigned to the conditional $B_1 \boxright M$ (and $B_2 \boxright L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1: one-box (open box B)
B_2: two-box choice (open both A and B)
N: receive nothing
K: receive \$1,000
M: receive \$1,000,000
L: receive \$1,001,000

$V(B_1) = V(M)P(B_1 \boxright M) + V(N)P(B_1 \boxright N) =$
$1000000 \cdot \mu + 0 \cdot 1 - \mu = 1000000\mu$

$V(B_2) = V(L)P(B_2 \boxright L) + V(K)P(B_2 \boxright K) =$
$1001000 \cdot \mu + 1000 \cdot 1 - \mu = 1000000\mu + 1000$

D. Lewis. *Prisoner's Dilemma Is a Newcomb Problem*. Philosophy and Public Affairs, 8, pgs. 235-240, 1979.

S. Brams. *Newcomb's Problem and Prisoners' Dilemma*. The Journal of Conflict Resolution, 19:4, pgs. 596 - 612, 1975.

S. Hurley. *Newcomb's Problem, Prisoner's Dilemma and Collective Action*. Synthese 86, pgs. 173 - 196, 1991.

Death in Damascus

A man in Damascus knows that he has an appointment with Death at midnight. He will escape Death if he manages at midnight not to be at the place of his appointment. He can be in either Damascus or Aleppo at midnight.

Death in Damascus

A man in Damascus knows that he has an appointment with Death at midnight. He will escape Death if he manages at midnight not to be at the place of his appointment. He can be in either Damascus or Aleppo at midnight. As the man knows, Death is a good predictor of his whereabouts. If he stays in Damascus, he thereby has evidence that Death will look for him in Damascus. However, if he goes to Aleppo he thereby has evidence that Death will look for him in Aleppo.

Death in Damascus

A man in Damascus knows that he has an appointment with Death at midnight. He will escape Death if he manages at midnight not to be at the place of his appointment. He can be in either Damascus or Aleppo at midnight. As the man knows, Death is a good predictor of his whereabouts. If he stays in Damascus, he thereby has evidence that Death will look for him in Damascus. However, if he goes to Aleppo he thereby has evidence that Death will look for him in Aleppo. Wherever he decides to be at midnight, he has evidence that he would be better off at the other place. No decision is stable.

A. Gibbard and W. Harper. *Counterfactuals and Two Kinds of Expected Utility*. In Ifs: Conditionals, Belief, Decision, Chance, and Time, pp. 153 - 190, 1978.

- The crucial distinction is between an act and a decision to perform the act.

- Before performing an act, an agent may assess the act in light of a decision to perform it. Information the decision carries may affect the act's expected utility and its ranking with respect to other acts.

- Decision makers should make self-ratifying, or ratifiable, decisions.

Two Forms of Ratificationism

- As an *elimination rule*: ratificationism requires you to reject all unratifiable acts, and to then choose among the ratifiable alternatives.
- As an *equilibrium rule*: ratificationism requires you to choose an act that is ratifiable relative to the beliefs and desires you will have when your deliberations cease ("reflective equilibrium").

Taking Stock

- Many choice rules: MEU, strict/weak dominance, maxmin, minmax regret
  - Which one is "best"?
  - What are the relationships between the different choice rules?

- Payoff is not the same as utility (von Neumann-Morgenstern utilities)

- Rational choice models should be applied with care (act-state dependence, deliberation, attitudes towards risk, attitudes toward ambiguity, ...

Material that we skipped during the lecture.

Allais Paradox

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$

Allais Paradox

	Options	Red (1)	White (89)	Blue (10)
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

Allais Paradox

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

Allais Paradox

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

$$A \succeq B \text{ iff } C \succeq B$$

Allais Paradox

We should not conclude either

Allais Paradox

We should not conclude either

(a) The axioms of cardinal utility fail to adequately capture our understanding of rational choice, or

Allais Paradox

We should not conclude either

(a) The axioms of cardinal utility fail to adequately capture our understanding of rational choice, or

(b) those who choose A in S_1 and D is L_2 are irrational.

Allais Paradox

We should not conclude either

(a) The axioms of cardinal utility fail to adequately capture our understanding of rational choice, or

(b) those who choose A in S_1 and D is L_2 are irrational.

Rather, people's utility functions (*their rankings over outcomes*) are often far more complicated than the monetary bets would indicate....

Ellsberg Paradox

	30	60	
Lotteries	Blue	Yellow	Green
L_1	$1M$	0	0
L_2	0	$1M$	0
L_3	$1M$	0	$1M$
L_4	0	$1M$	$1M$

Ellsberg Paradox

	30	60	
Lotteries	Blue	Yellow	Green
L_1	$1M$	0	0
L_2	0	$1M$	0
L_3	$1M$	0	$1M$
L_4	0	$1M$	$1M$

$$L_1 \succeq L_2 \quad \text{iff} \quad L_3 \succeq L_4$$