# Epistemic Game Theory

## Lecture 1

Eric Pacuit

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

January 26, 2014

# Game Theory

“We wish to find the mathematically complete principles which define ‘rational behavior’ for the participants.” (pg. 31)

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

# Game Theory

"We wish to find the mathematically complete principles which define 'rational behavior' for the participants." (pg. 31)

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

"Game theory is a bag of analytical tools designed to help us understand the phenomena that we observe when decision-makers interact." (pg. 1)

M. Osborne and A. Rubinstein. *Introduction to Game Theory*. MIT Press, 2004.

# The Guessing Game

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

# The Guessing Game

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

What number should you guess?

# The Guessing Game

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

What number should you guess? 100

# The Guessing Game

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

What number should you guess? ~~100~~, 99

# The Guessing Game

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

What number should you guess? ~~100~~, ~~99~~, . . . , 67

## The Guessing Game

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

What number should you guess? $\cancel{100}$, $\cancel{99}$, . . . , $\cancel{67}$, . . . , 2, 1

# The Guessing Game

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

What number should you guess? $\cancel{100}$, $\cancel{99}$, . . . , $\cancel{67}$, . . . , $\cancel{2}$, 1

# Plan for Today

- Just enough decision theory
- Just enough game theory
- Setting the stage: Epistemic game theory

# Just enough decision theory

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

# Just enough decision theory

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

"The only normative content implied by the use of an expected utility model is that preferences should be coherent." (pg. 11, Skyrms)

# Just enough decision theory

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

"The only normative content implied by the use of an expected utility model is that preferences should be coherent." (pg. 11, Skyrms)

# Maximizing

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

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

# Maximizing

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

"The formulation of maximizing behavior in economics has often paralleled the modeling of maximization in physics an 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

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

"The formulation of maximizing behavior in economics has often paralleled the modeling of maximization in physics an 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)

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)

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

# 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"

# Preliminaries: Orderings

An ordering is a *relation* $R$ on a set $X$: a subset of the set of pairs of elements from $X$: $R \subseteq X \times X$

Write $aRb$ iff $(a, b) \in R$

# Preliminaries: Orderings

An ordering is a *relation* $R$ on a set $X$: a subset of the set of pairs of elements from $X$: $R \subseteq X \times X$

Write $aRb$ iff $(a, b) \in R$

Properties of orderings:

- Reflexivity: for all $a \in X$, $aRa$
- Transitivity: for all $a, b, c \in X$, $aRb$ and $bRc$ then $aRc$
- Symmetry: for all $a, b \in X$, $aRb$ implies $bRa$
- Asymmtery: for all $a, b \in X$, $aRb$ implies not-$bRa$
- Completeness: for all $a, b \in X$, $aRb$ or $bRa$ (or $a = b$)

# Preliminaries: Orderings

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

# Preliminaries: 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:

## Preliminaries: 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 $y$ ($x \succ y$)

# Preliminaries: 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 $x$ ($y \succ x$)

# Preliminaries: 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 $y$ ($x \approx y$)

# Preliminaries: 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 $y$ ($x \perp y$)

## Preliminaries: Utility Function

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

# Preliminaries: Utility Function

A **utility function** on a set $X$ is a function $u : X \to \mathfrak{R}$

A utility function $u : X \to \mathfrak{R}$ **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. **Reflexivity**: Weak preference is reflexive: the agent always thinks $x$ is at least as good as $x$.

## 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. **Reflexivity**: Weak preference is reflexive: the agent always thinks $x$ is at least as good as $x$.

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

# Why should we accept these axioms?

"Rather than trying to provide instrumental or pragmatic justifications for the axioms of ordinal utility, it is better...to see them as constitutive of our conception of a fully rational agent....those disposed to blatantly ignore transitivity are unintelligible to us: we can't understand their pattern of actions as sensible"

(Gaus, *On Philosophy, Politics and Economics*, pg. 39)

# Ordinal Utility Theory

**Fact**. Suppose that $X$ is finite and $\succeq$ is a complete and transitive ordering over $X$, then there is a utility function $u : X \rightarrow \Re$ that represents $\succeq$.

# Ordinal Utility Theory

**Fact**. 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 $\succeq$.

Utility is *defined* in terms of preference (so it is an error to say that the agent prefers $x$ to $y$ *because* she assigns a higher utility to $x$ than to $y$).

# Ordinal Utility Theory

**Fact**. Suppose that $X$ is finite and $\succeq$ is a complete and transitive ordering over $X$, then there is a utility function $u : X \rightarrow \mathfrak{R}$ that represents $\succeq$.

Utility is *defined* in terms of preference (so it is an error to say that the agent prefers $x$ to $y$ *because* she assigns a higher utility to $x$ than to $y$).

**Important point**: consider $x \succ y \succ z$

# Ordinal Utility Theory

**Fact**. 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 $\succeq$.

Utility is *defined* in terms of preference (so it is an error to say that the agent prefers $x$ to $y$ *because* she assigns a higher utility to $x$ than to $y$).

**Important point**: consider $x \succ y \succ z$, all three utility functions represent this ordering:

| Preference | $u_1$ | $u_2$ | $u_3$ |
|---|---|---|---|
| $x$ | 3 | 10 | 1000 |
| $y$ | 2 | 5 | 99 |
| $z$ | 1 | 0 | 1 |

# Cardinal Utility Theory

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

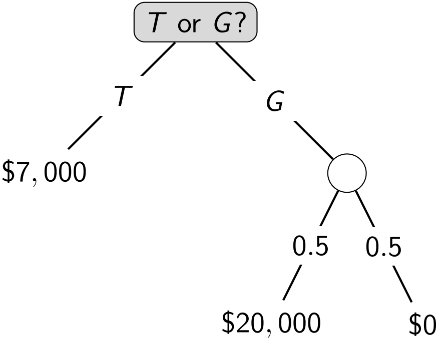
# Cardinal Utility Theory

$x \succ y \succ z$ is represented by both $(3, 2, 1)$ and $(1000, 999, 1)$, so cannot say $y$ 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



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 $[1 : x]$.

# Axiom 1

$\succeq$ is a preference ordering over $\mathcal{L}$: i.e., $\succeq$ is a reflexive, transitive and complete relation on $\mathcal{L}$.

# Axiom 2

**Continuity**: For every triple $x, y, z \in X$, if $x \succeq y \succeq z$, then there exists a $p \in [0, 1]$ such that

$$y \approx [p : x, (1 - p) : z]$$

# Axiom 3

**Monotonicity**: Suppose that $p, q \in [0, 1]$ and suppose that $x \succ y$. Then,

$$[p : x, (1 - p) : y] \succeq [q : x, (1 - q) : y]$$

if, and only if, $p \geq q$.

# Compound Lotteries

Suppose that $L_1, \ldots, L_n$ are lotteries. A compound lottery is the lottery $[p_1 : L_1, \ldots, p_n : L_n]$ where $p_i \in [0,1]$ and $\sum_i p_i = 1$.

Let $\hat{\mathcal{L}}$ be the set of compound lotteries.

# Axiom 4

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

*This eliminates utility from the thrill of gambling and so the only ultimate concern is the prizes.*

# Axiom 5

**Independence**: Suppose that $\hat{L} = [p_1 : L_1, \ldots, L_i, \ldots, p_n : L_n]$ is a compound lottery and $M$ is a simple lottery. If $L_i \approx M$, then

$$\hat{L} \approx [p_1 : L_1, \ldots, p_{i-1} : L_{i-1}, M, p_{i+1} : L_{i+1}, \ldots, p_n : L_n]$$

**Fact**. For $x, y \in X$, if $x \succeq y$, then $x \succeq [0.5 : y, 0.5 : x]$

# Are the Axioms Reasonable?

Suppose you have a kitten, which you plan to give away to either Ann or Bob. Ann and Bob both want the kitten very much. Both are deserving, and both would care for the kitten. You are sure that giving the kitten to Ann ($x$) is at least as good as giving the kitten to Bob ($y$) (so $x \succeq y$). But you think that would be unfair to Bob. You decide to flip a fair coin: if the coin lands heads, you will give the kitten to Bob, and if it lands tails, you will give the kitten to Ann. (J. Drier, "Morality and Decision Theory" in *Handbook of Rationality*)

# Are the Axioms Reasonable?

Does this contradict the von Neumann-Morgenstern Axioms?

# Are the Axioms Reasonable?

Does this contradict the von Neumann-Morgenstern Axioms? Consider the lottery which is $x$ for sure ($L_1$) and the lottery which is 0.5 for $y$ and 0.5 for $x$ ($L_2$).

# Are the Axioms Reasonable?

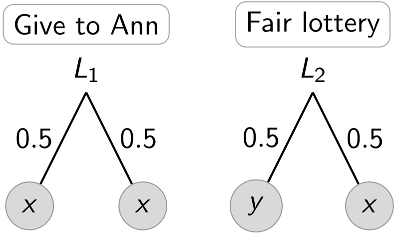
Does this contradict the von Neumann-Morgenstern Axioms? Consider the lottery which is $x$ for sure ($L_1$) and the lottery which is 0.5 for $y$ and 0.5 for $x$ ($L_2$). The previous fact implies that $L_1 \succeq L_2$

## Are the Axioms Reasonable?

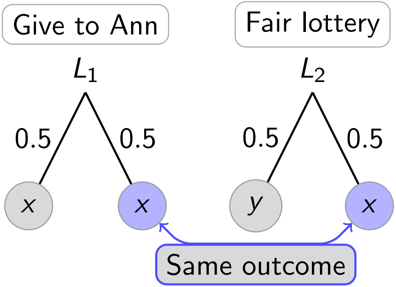
Does this contradict the von Neumann-Morgenstern Axioms? Consider the lottery which is $x$ for sure ($L_1$) and the lottery which is 0.5 for $y$ and 0.5 for $x$ ($L_2$). The previous fact implies that $L_1 \succeq L_2$ but a person concerned with fairness may have $L_2 \succeq L_1$.

# Are the Axioms Reasonable?

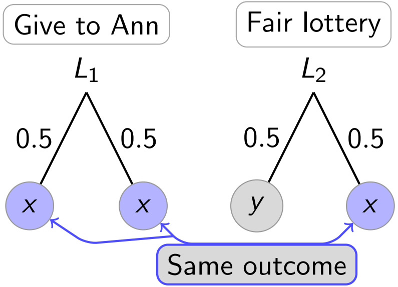
Does this contradict the von Neumann-Morgenstern Axioms? Consider the lottery which is $x$ for sure ($L_1$) and the lottery which is 0.5 for $y$ and 0.5 for $x$ ($L_2$). The previous fact implies that $L_1 \succeq L_2$ but a person concerned with fairness may have $L_2 \succeq L_1$. *But if fairness is important then that should be part of the description of the outcome!*



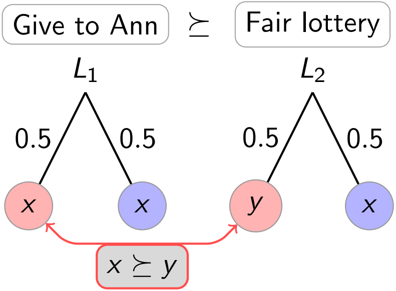

- $x$ is the outcome "Ann gets the kitten"
- $y$ is the outcome "Bob gets the kitten"



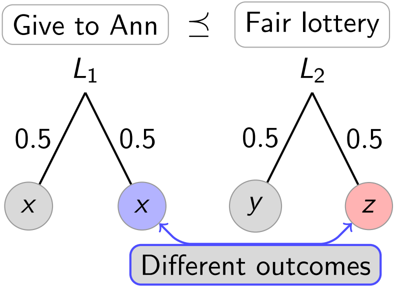

- $x$ is the outcome "Ann gets the kitten"
- $y$ is the outcome "Bob gets the kitten"



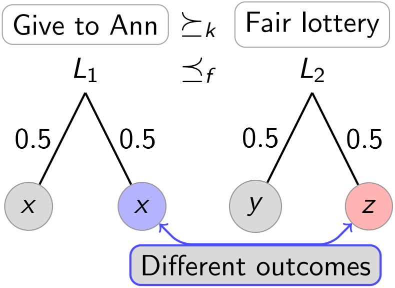

- $x$ is the outcome "Ann gets the kitten"
- $y$ is the outcome "Bob gets the kitten"




- $x$ is the outcome "Ann gets the kitten"
- $y$ is the outcome "Bob gets the kitten"




- $x$ is the outcome "Ann gets the kitten"
- $z$ is the outcome "Ann gets the outcome, *fairly*
- $y$ is the outcome "Bob gets the kitten, *fairly*"




If all the agent cares about is who gets the kitten, then $L_1 \succeq L_2$

If all the agent cares about is being fair, then $L_1 \preceq L_2$

A utility function $u : \hat{\mathcal{L}} \to \mathfrak{R}$ is **linear** provided

$$u([p_1 : L_1, \ldots, p_n : L_n]) = \sum_i p_i u(L_i)$$

**Theorem** (von Neumann & Morgenstern) If a relation $\succeq$ on $\hat{\mathcal{L}}$ satisfies axioms 1-5, then there exists a linear utility function $u : \hat{\mathcal{L}} \to \mathfrak{R}$ that represents $\succeq$.

# Subjective Expected Utility

Given an agent's beliefs (probabilities) and desires (utilities), the **expected utility** of an **action** leading to a set of outcomes $X$ is:

$$\sum_{x \in X} [\text{the probability that the act will lead to } x] \times [\text{the utility of } x]$$

# 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
- Causal vs. Evidential Decision Theory: Newcomb's Paradox

Just enough game theory

# Game Situations

1. a group of *self-interested* agents (players) involved in some interdependent decision problem

# Game Situations

| | Bob | |
|---|---|---|
| | $L$ | $R$ |
| | 1 | 0 |

1. a group of *self-interested* agents (players) involved in some interdependent decision problem

# Game Situations

| | Bob | |
|---|---|---|
| | $L$ | $R$ |
| | 1 | 0 |
| | 0 | 1 |

1. a group of *self-interested* agents (players) involved in some interdependent decision problem

# Game Situations

| | | Bob | |
|---|---|---|---|
| | | $L$ | $R$ |
| Ann | $U$ | 1 1 | 0 0 |
| | $D$ | 0 0 | 1 1 |

1. a group of *self-interested* agents (players) involved in some interdependent decision problem

# Game Situations

| | | Bob $L$ | Bob $R$ |
|---|---|---|---|
| Ann | $U$ | 1,1 | 0,0 |
| Ann | $D$ | 0,0 | 1,1 |

1. a group of *self-interested* agents (players) involved in some interdependent decision problem

# Game Situations

| | | Bob $L$ | $R$ |
|---|---|---|---|
| Ann | $U$ | 1,1 | 0,0 |
| | $D$ | 0,0 | 1,1 |

1. a group of *self-interested* agents (players) involved in some interdependent decision problem

# Just Enough Game Theory

A **game** is a mathematical model of a strategic interaction that includes

- the actions the players *can* take
- the players' interests (i.e., preferences),
- the "structure" of the decision problem

# Just Enough Game Theory

A **game** is a mathematical model of a strategic interaction that includes

- the actions the players *can* take
- the players' interests (i.e., preferences),
- the "structure" of the decision problem

*It does not specify the actions that the players do take.*

# Solution Concepts

A **solution concept** is a systematic description of the outcomes that may emerge in a family of games.

This is the starting point for most of game theory and includes many variants: Nash equilibrium, backwards induction, or iterated dominance of various kinds.

These are usually thought of as the embodiment of “rational behavior” in some way and used to analyze game situations.

# Nash Equilibrium

A strategy profile $\sigma$ is a **(pure strategy) Nash equilibrium** provided no player has an incentive to deviate from his/her choice: for all $i$ and all $s_i \neq \sigma_i$:

$$u_i(\sigma) \geq u_i(s_i, \sigma_{-i})$$

# Pure Coordination Game

| | | Bob | |
|---|---|---|---|
| | | $L$ | $R$ |
| Ann | $U$ | 1,1 | 0,0 |
| | $D$ | 0,0 | 1,1 |

The profiles **(U, L) and (D, R)** are Nash equilibria.

# Zero-Sum Games

| | | Bob $L$ | Bob $R$ |
|---|---|---|---|
| Ann | $U$ | 1,4 | 4,1 |
| Ann | $D$ | 2,3 | 3,2 |

What should Ann *do*?

# Zero-Sum Games

| | | Bob | |
|---|---|---|---|
| | | $L$ | $R$ |
| Ann | $U$ | 1,4 | 4,1 |
| | $D$ | 2,3 | 3,2 |

What should Ann *do*? *Bob best choice in Ann's worst choice*

## Zero-Sum Games

| Ann \ Bob | $L$ | $R$ | |
|---|---|---|---|
| $U$ | 1,4 | 4,1 | 1 |
| $D$ | 2,3 | 3,2 | 2 |

What should Ann *do*? *Security strategy: minimize over each row and choose the maximum value*

# Zero-Sum Games

| | | Bob | |
|---|---|---|---|
| | | $L$ | $R$ |
| Ann | $U$ | 1,4 | 4,1 |
| | $D$ | 2,3 | 3,2 |
| | | 3 | 1 |

What should Bob *do*? *Security strategy: minimize over each column and choose the maximum value*

# Zero-Sum Games

| | | | Bob | |
|---|---|---|---|---|
| | | | $L$ | $R$ |
| Ann | $U$ | | 1,4 | 4,1 |
| | $D$ | | 2,3 | 3,2 |

The profile of security strategies $(D, L)$ is a Nash equilbirium

# Matching Pennies

| | | Bob $H$ | Bob $T$ |
|---|---|---|---|
| Ann | $H$ | 1,-1 | -1, 1 |
| Ann | $T$ | -1,1 | 1,-1 |

There are no pure strategy Nash equilibria.

# Mixed Strategies

Bob

| Ann | $H$ | $T$ |
|---|---|---|
| $H$ | 1,-1 | -1, 1 |
| $T$ | -1,1 | 1,-1 |

A **mixed strategy** is a probability distribution over the set of pure strategies. For instance:

- $(1/2H, 1/2T)$
- $(1/3H, 2/3T)$
- ...

# Matching Pennies

| | | Bob $H$ | Bob $T$ |
|---|---|---|---|
| Ann | $H$ | 1,-1 | -1, 1 |
| Ann | $T$ | -1,1 | 1,-1 |

The mixed strategy $([1/2 : H, 1/2 : T], [1/2 : H, 1/2 : T])$ is the only Nash equilibrium.

**Theorem** (Von Neumann). For every two-player zero- sum game with finite strategy sets $S_1$ and $S_2$, there is a number $v$, called the **value** of the game such that:

1. $v = \max_{p\in\Delta(S_1)} \min_{q\in\Delta(S_2)} U_1(p, q) = \min_{q\in\Delta(S_2)} \max_{p\in\Delta(S_1)} U_1(p, q)$
2. The set of mixed Nash equilibria is nonempty. A mixed strategy profile $(p, q)$ is a Nash equilibrium if and only if

$$p \in \mathrm{argmax}_{p\in\Delta(S_1)} \min_{q\in\Delta(S_2)} U_1(p, q)$$

$$q \in \mathrm{argmax}_{q\in\Delta(S_2)} \min_{p\in\Delta(S_1)} U_1(p, q)$$

3. For all mixed Nash equilibria $(p, q)$, $U_1(p, q) = v$

# Why play such an equilibrium?

"Let us now imagine that there exists a complete theory of the zero-sum two-person game which tells a player what to do, and which is absolutely convincing. If the players knew such a theory then each player would have to assume that his strategy has been "found out" by his opponent. The opponent knows the theory, and he knows that the player would be unwise not to follow it... a satisfactory theory can exist only if we are able to harmonize the two extremes...strategies of player 1 'found out' or of player 2 'found out.' " (pg. 148)

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

"Von Neumann and Morgenstern are assuming that the *payoff matrix* is common knowledge to the players, but presumably the players' subjective probabilities might be private. Then each player might quite reasonably act to maximize subjective expected utility, believing that he will *not* be found out, with the result *not* being a Nash equilibrium."

(Skyrms, pg. 14)

| | | Bob | |
|---|---|---|---|
| | | $L$ | $R$ |
| Ann | $U$ | 1,4 | 4,1 |
| | $D$ | 2,3 | 3,2 |

| Ann \ Bob | $L$ | $R$ |
|---|---|---|
| $U$ | 1,4 | 4,1 |
| $D$ | 2,3 | 3,2 |

- Suppose that Ann believes Bob will play $L$ with probability $1/4$, *for whatever reason*. Then,

$$1 \times 0.25 + 4 \times 0.75 = 3.25 \geq 2 \times 0.25 + 3 \times 0.75 = 2.75$$

| | | Bob | |
|---|---|---|---|
| | | $L$ | $R$ |
| Ann | $U$ | 1,4 | 4,1 |
| | $D$ | 2,3 | 3,2 |

- Suppose that Ann believes Bob will play $L$ with probability $1/4$, *for whatever reason*. Then,

$$1 \times 0.25 + 4 \times 0.75 = 3.25 \geq 2 \times 0.25 + 3 \times 0.75 = 2.75$$

- But, $L$ is maximizes expected utility no matter what belief Bob may have:

$$p + 3 = 4 \times p + 3 \times (1 - p) \geq 1 \times p + 2 \times (1 - p) = 2 - p$$

In zero-sum games

- There exists a mixed strategy Nash equilibrium
- There may be more than one Nash equilibria
- Security strategies are always a Nash equilibrium
- Components of Nash equilibria are interchangeable: If $\sigma$ and $\sigma'$ are Nash equilibria in a 2-player game, then $(\sigma_1, \sigma_2')$ is also a Nash equilbiria.

| | | Bob | | |
|---|---|---|---|---|
| | | $L$ | $C$ | $R$ |
| Ann | $T$ | 1, -1 | 0,0 | -1, 1 |
| | $M$ | 0,0 | 0,0 | 0,0 |
| | $B$ | -1,1 | 0,0 | 1,-1 |

$(M, C)$ is the unique Nash equilibria. Suppose that both player's subjective probabilities are $(1/3, 1/3, 1/3)$, and this is *common knowledge*. Then, any choice maximizes the players' expected utility.

Suppose that $G = (S_1, \ldots, S_n, u_1, \ldots, u_n)$ is a strategic game.

A strategy $s_i \in S_i$ is a **best response** to a joint probability $m_{-i} \in \Pi_{j \neq i} \Delta(S_j)$ iff $U_i(s_i, m_{-i}) \geq U_i(s_i', m_{-i})$ for all $s_i' \in S_i$ (here $U_i(\cdot, m_{-i})$ is the expected utility with respect to the joint probability $m_{-i}$).

# Prisoner's Dilemma

Two people commit a crime.

# Prisoner's Dilemma

Two people commit a crime. The are arrested by the police, who are quite sure they are guilty but cannot prove it without at least one of them confessing.

# Prisoner's Dilemma

Two people commit a crime. The are arrested by the police, who are quite sure they are guilty but cannot prove it without at least one of them confessing. The police offer the following deal. Each one of them can confess and get credit for it.

# Prisoner's Dilemma

Two people commit a crime. The are arrested by the police, who are quite sure they are guilty but cannot prove it without at least one of them confessing. The police offer the following deal. Each one of them can confess and get credit for it. If only one confesses, he becomes a state witness and not only is he not punished, he gets a reward.

# Prisoner's Dilemma

Two people commit a crime. The are arrested by the police, who are quite sure they are guilty but cannot prove it without at least one of them confessing. The police offer the following deal. Each one of them can confess and get credit for it. If only one confesses, he becomes a state witness and not only is he not punished, he gets a reward. If both confess, they will be punished but will get reduced sentences for helping the police.

# Prisoner's Dilemma

Two people commit a crime. The are arrested by the police, who are quite sure they are guilty but cannot prove it without at least one of them confessing. The police offer the following deal. Each one of them can confess and get credit for it. If only one confesses, he becomes a state witness and not only is he not punished, he gets a reward. If both confess, they will be punished but will get reduced sentences for helping the police. If neither confesses, the police honestly admit that there is no way to convict them, and they are set free.

# Prisoner's Dilemma

Two options: Confess ($C$), Don't Confess ($D$)

# Prisoner's Dilemma

Two options: Confess ($C$), Don't Confess ($D$)

Possible outcomes:

# Prisoner's Dilemma

Two options: Confess ($C$), Don't Confess ($D$)

Possible outcomes: We both confess ($C, C$),

# Prisoner's Dilemma

Two options: Confess ($C$), Don't Confess ($D$)

Possible outcomes: We both confess ($C, C$), I confess but my partner doesn't ($C, D$),

# Prisoner's Dilemma

Two options: Confess ($C$), Don't Confess ($D$)

Possible outcomes: We both confess ($C, C$), I confess but my partner doesn't ($C, D$), My partner confesses but I don't ($D, C$),

# Prisoner's Dilemma

Two options: Confess ($C$), Don't Confess ($D$)

Possible outcomes: We both confess ($C, C$), I confess but my partner doesn't ($C, D$), My partner confesses but I don't ($D, C$), neither of us confess ($D, D$).

# Prisoner's Dilemma

| | | Bob | |
|---|---|---|---|
| | | $D$ | $C$ |
| Ann | $D$ | | |
| | $C$ | | |

# Prisoner's Dilemma

| | | Bob | |
|---|---|---|---|
| | | $D$ | $C$ |
| Ann | $D$ | 3 | 1 |
| | $C$ | 4 | 2 |

Ann's preferences

# Prisoner's Dilemma

| | | Bob | |
|---|---|---|---|
| | | $D$ | $C$ |
| Ann | $D$ | 3 | 4 |
| | $C$ | 1 | 2 |

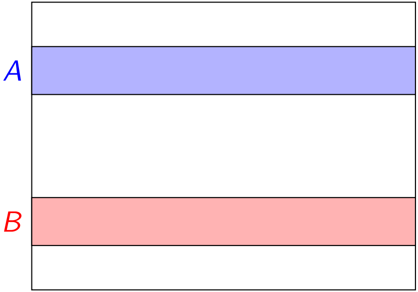
Bob's preferences

# Prisoner's Dilemma

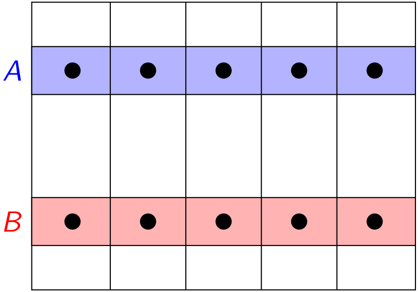
| | | Bob | |
|---|---|---|---|
| | | $D$ | $C$ |
| Ann | $D$ | 3,3 | 1,4 |
| | $C$ | 4,1 | 2,2 |

What should Ann (Bob) do?

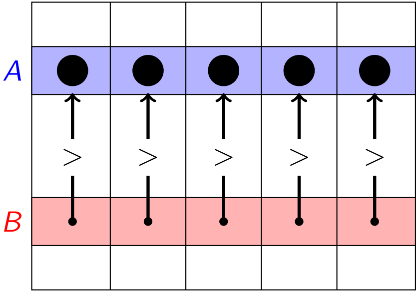
# Dominance Reasoning



# Dominance Reasoning



# Dominance Reasoning



# Prisoner's Dilemma

| | | Bob | |
|---|---|---|---|
| | | $D$ | $C$ |
| Ann | $D$ | 3,3 | 1,4 |
| | $C$ | 4,1 | 2,2 |

What should Ann (Bob) do?

# Prisoner's Dilemma

| | | Bob | |
|---|---|---|---|
| | | $D$ | $C$ |
| Ann | $D$ | 3,3 | 1,4 |
| | $C$ | 4,1 | 2,2 |

What should Ann (Bob) do? *Dominance reasoning*

# Prisoner's Dilemma

| | | Bob | |
|---|---|---|---|
| | | $D$ | $C$ |
| Ann | $D$ | 3,3 | 1,4 |
| | $C$ | 4,1 | 2,2 |

What should Ann (Bob) do? *Dominance reasoning*

# Prisoner's Dilemma

| | | Bob | |
|---|---|---|---|
| | | $D$ | $C$ |
| Ann | $D$ | 3,3 | 1,4 |
| | $C$ | 4,1 | 2,2 |

What should Ann (Bob) do? *Dominance reasoning* is not **Pareto**!

In an arbitrary (finite) games (that are not zero-sum)

- There exists a mixed strategy Nash equilibrium
- Security strategies are not necessarily a Nash equilibrium
- There may be more than on Nash equilibrium
- Components of Nash equilibrium are not interchangeable.

# Chicken

| | | Bob $D$ | $S$ |
|---|---|---|---|
| Ann | $D$ | -4,-4 | 1,-1 |
| | $S$ | -1,1 | 0,0 |

# Chicken

Bob / Ann

| | | $D$ | $S$ |
|---|---|---|---|
| Ann | $D$ | -4,-4 | 1,-1 |
| | $S$ | -1,1 | 0,0 |

$(D, S)$ and $(S, D)$ are Nash equilibria. If both choose their components of these equilibria, we may end up at $(D, D)$.

# Chicken

| | | Bob $D$ | Bob $S$ |
|---|---|---|---|
| Ann | $D$ | -4,-4 | 1,-1 |
| Ann | $S$ | -1,1 | 0,0 |

$(D, S)$ and $(S, D)$ are Nash equilibria. Their security strategies are $(S, S)$.

# Battle of the Sexes

| | | Bob $B$ | Bob $M$ |
|---|---|---|---|
| Ann | $B$ | 2, 1 | 0, 0 |
| Ann | $M$ | 0, 0 | 1, 2 |

# Battle of the Sexes

| Ann \ Bob | $B$ | $M$ |
| --- | --- | --- |
| $B$ | 2, 1 | 0, 0 |
| $M$ | 0, 0 | 1, 2 |

$(B, B)$ $(M, M)$, and $([2/3 : B, 1/3 : M], [1/3 : B, 2/3 : M])$ are Nash equilibria.

# Perfect equilibrium

| | | Bob | |
|---|---|---|---|
| | | $L$ | $R$ |
| Ann | $U$ | 1,1 | 0,0 |
| | $D$ | 0,0 | 0,0 |

# Perfect equilibrium

Bob

| Ann | | $L$ | $R$ |
|---|---|---|---|
| | $U$ | 1,1 | 0,0 |
| | $D$ | 0,0 | 0,0 |

# Perfect equilibrium

| | | Bob $L$ | Bob $R$ |
|---|---|---|---|
| Ann | $U$ | 1,1 | 0,0 |
| Ann | $D$ | 0,0 | 0,0 |

Isn't $(U, L)$ more "reasonable" than $(D, R)$?

# Perfect equilibrium

Bob

| Ann | | $L$ | $R$ |
|---|---|---|---|
| | $U$ | 1,1 | 0,0 |
| | $D$ | 0,0 | 0,0 |

**Completely mixed strategy**: a mixed strategy in which every strategy gets some positive probability

# Perfect equilibrium

| Ann \ Bob | $L$ | $R$ |
|---|---|---|
| $U$ | 1,1 | 0,0 |
| $D$ | 0,0 | 0,0 |

**Completely mixed strategy**: a mixed strategy in which every strategy gets some positive probability

$\epsilon$-**perfect equilibrium**: a completely mixed strategy profile in which any pure strategy that is not a best reply receives probability less than $\epsilon$

**Prefect equilibrium**: the mixed strategy profile that is the limit as $\epsilon$ goes to 0 of $\epsilon$-prefect equilibria.

# Proper equilibrium

Bob

| Ann | | $L$ | $C$ | $R$ |
|---|---|---|---|---|
| | $U$ | -9,-9 | -7,-7 | -7,-7 |
| | $M$ | 0,0 | 0,0 | -7,-7 |
| | $D$ | 1,1 | 0,0 | -9,-9 |

# Proper equilibrium

| | | Bob | | |
|---|---|---|---|---|
| | | $L$ | $C$ | $R$ |
| Ann | $U$ | -9,-9 | -7,-7 | -7,-7 |
| | $M$ | 0,0 | 0,0 | -7,-7 |
| | $D$ | 1,1 | 0,0 | -9,-9 |

## Proper equilibrium

| | | Bob | | |
|---|---|---|---|---|
| | | $L$ | $C$ | $R$ |
| Ann | $U$ | -9,-9 | -7,-7 | -7,-7 |
| | $M$ | 0,0 | 0,0 | -7,-7 |
| | $D$ | 1,1 | 0,0 | -9,-9 |

$\epsilon$-**proper equilibrium**: a completely mixed strategy profile such that if strategy $s$ is a better response than $s'$, then $\frac{p(s)}{p(s')} < \epsilon$

**Proper equilibrium**: the mixed strategy profile that is the limit as $\epsilon$ goes to 0 of $\epsilon$-proper equilibria.

## Proper equilibrium

| Ann \ Bob | $L$ | $C$ | $R$ |
|---|---|---|---|
| $U$ | -9,-9 | -7,-7 | -7,-7 |
| $M$ | 0,0 | 0,0 | -7,-7 |
| $D$ | 1,1 | 0,0 | -9,-9 |

$\epsilon$-**proper equilibrium**: a completely mixed strategy profile such that if strategy $s$ is a better response than $s'$, then $\frac{p(s)}{p(s')} < \epsilon$

**Proper equilibrium**: the mixed strategy profile that is the limit as $\epsilon$ goes to 0 of $\epsilon$-proper equilibria.

# Trembling Hands

"There cannot be any mistakes if the players are absolutely rational. Nevertheless, a satisfactory interpretation of equilibrium points in extensive games seems to require that the possibility of mistakes is not completely excluded. This can be achieved by a point of view which looks at complete rationality as the limiting case of incomplete rationality." (pg. 35)

R. Selten. *Reexamination of the Perfectness Concept of Equilibrium in Extensive Games*. International Journal of Game Theory, 4, pgs. 25 - 55, 1975.

Setting the stage: Epistemic game theory

# Knowledge and beliefs in game situations


J. Harsanyi. *Games with incomplete information played by "Bayesian" players I-III*. *Management Science Theory* **14***: 159-182, 1967-68.*

Robert Aumann. *Agreeing to Disagree*. Annals of Statistics **4** (1976).

R. Aumann. *Interactive Epistemology I & II*. International Journal of Game Theory (1999).

P. Battigalli and G. Bonanno. *Recent results on belief, knowledge and the epistemic foundations of game theory*. Research in Economics (1999).

R. Myerson. *Harsanyi's Games with Incomplete Information*. Special 50th anniversary issue of *Management Science*, 2004.

John C. Harsanyi, nobel prize winner in economics, developed a theory of games with **incomplete information**.

J. Harsanyi. *Games with incomplete information played by "Bayesian" players I-III. Management Science Theory* **14**: *159-182, 1967-68.*

John C. Harsanyi, nobel prize winner in economics, developed a theory of games with **incomplete information**.

1. incomplete information: uncertainty about the *structure* of the game (outcomes, payoffs, strategy space)
2. imperfect information: uncertainty *within the game* about the previous moves of the players

J. Harsanyi. *Games with incomplete information played by "Bayesian" players I-III. Management Science Theory* **14**: *159-182, 1967-68.*

# Harsanyi's Problem

A natural question following any game-theoretic analysis is

# Harsanyi's Problem

A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*?

## Harsanyi's Problem

A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

# Harsanyi's Problem

A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

1. Suppose there is a parameter that some player $i$ does not know

# Harsanyi's Problem

A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

1. Suppose there is a parameter that some player $i$ does not know
2. $i$'s uncertainty about the parameter must be included in the model (first-order beliefs)

# Harsanyi's Problem

A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

1. Suppose there is a parameter that some player $i$ does not know
2. $i$'s uncertainty about the parameter must be included in the model (first-order beliefs)
3. this is a new parameter that the other players may not know, so we must specify the players beliefs about this parameter (second-order beliefs)

# Harsanyi's Problem

A natural question following any game-theoretic analysis is *how would the players react if some parameters of the model are not known to the players*? How do we completely specify such a model?

1. Suppose there is a parameter that some player $i$ does not know
2. $i$'s uncertainty about the parameter must be included in the model (first-order beliefs)
3. this is a new parameter that the other players may not know, so we must specify the players beliefs about this parameter (second-order beliefs)
4. but this is a new parameter, and so on....

# Harsanyi's Problem

A (game-theoretic) **type** of a player summarizes everything the player knows privately at the beginning of the game which could affect his beliefs about payoffs in the game and about all other players' types.

(Harsanyi argued that all uncertainty in a game can be equivalently modeled as uncertainty about payoff functions.)

# Information in games situations

- imperfect information about the play of the game
- incomplete information about the structure of the game

# Information in games situations

- imperfect information about the play of the game
- incomplete information about the structure of the game
- strategic information (what will the other players do?)
- higher-order information (what are the other players thinking?)

# Epistemic Game Theory

*Formally, a game is described by its strategy sets and payoff functions.*

# Epistemic Game Theory

*Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game.*

# Epistemic Game Theory

*Formally, a game is described by its strategy sets and payoff functions. But in real life, may other parameters are relevant; there is a lot more going on. Situations that substantively are vastly different may nevertheless correspond to precisely the same strategic game....*
*The difference lies in the attitudes of the players, in their expectations about each other, in custom, and in history, though the rules of the game do not distinguish between the two situations.* *(pg. 72)*

R. Aumann and J. H. Dreze. *Rational Expectations in Games*. American Economic Review 98 (2008), pp. 72-86.

# The Epistemic Program in Game Theory

"...the analysis constitutes a fleshing-out of the textbook interpretation of equilibrium as 'rationality plus correct beliefs.'...this suggests that equilibrium behavior cannot arise out of strategic reasoning alone. "

E. Dekel and M. Siniscalchi. *Epistemic Game Theory*. manuscript, 2013.

A. Brandenburger. *The Power of Paradox*. International Journal of Game Theory, 35, pgs. 465 - 492, 2007.

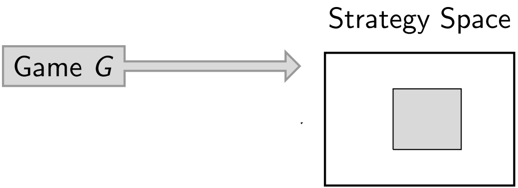
EP and O. Roy. *Epistemic Game Theory*. Stanford Encyclopedia of Philosophy, forthcoming, 2013.

# The Epistemic Program in Game Theory

$\boxed{\text{Game } G}$

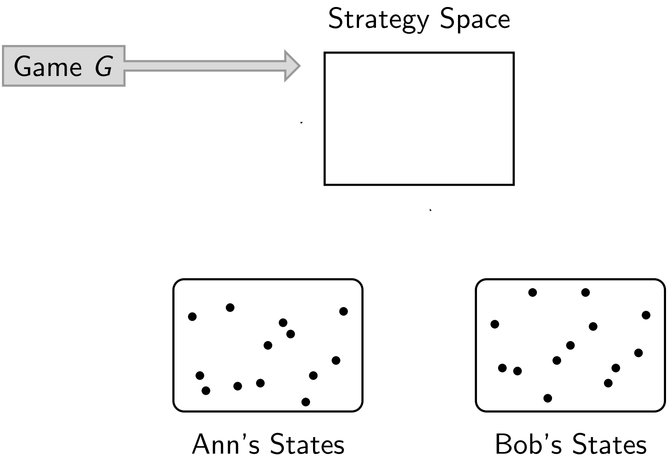
$G$: available actions, payoffs, structure of the decision problem

# The Epistemic Program in Game Theory



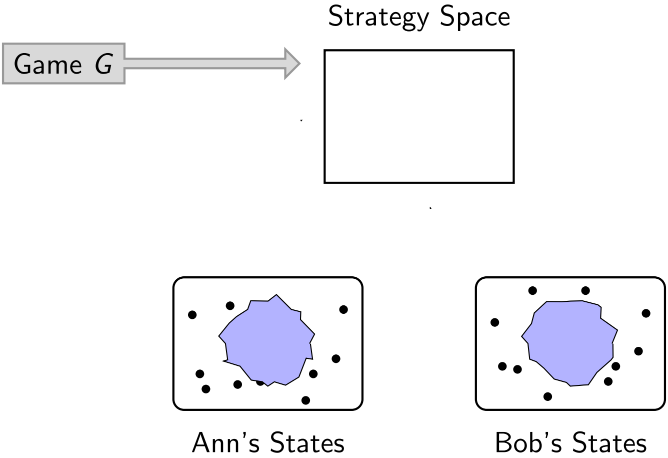

**solution concepts** are systematic descriptions of what players *do*

# The Epistemic Program in Game Theory



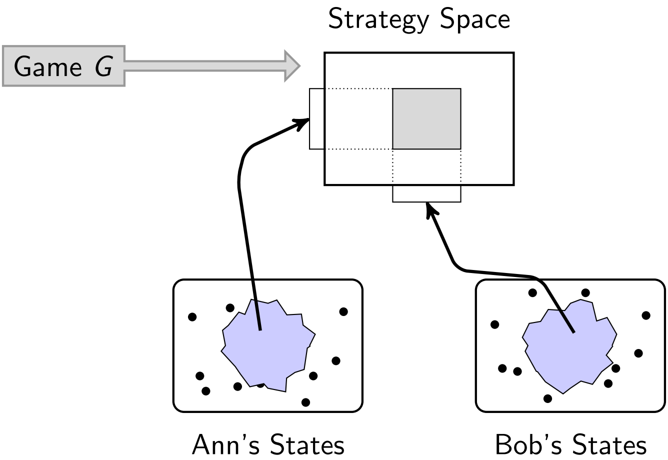

The game model includes *information states* of the players

# The Epistemic Program in Game Theory




Restrict to information states satisfying some rationality condition

# The Epistemic Program in Game Theory




Project onto the strategy space

# Information in games situations

- Various states of information disclosure.

# Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*

# Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*

- Various "types" of information:

# Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*

- Various "types" of information:
  - imperfect information about the play of the game
  - incomplete information about the structure of the game
  - strategic information (what will the other players do?)
  - higher-order information (what are the other players thinking?)

# Information in games situations

- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*

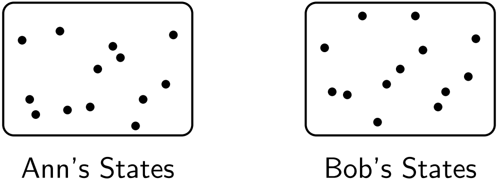
- Various "types" of information:
  - imperfect information about the play of the game
  - incomplete information about the structure of the game
  - strategic information (what will the other players do?)
  - higher-order information (what are the other players thinking?)

- Varieties of informational attitudes

# Information in games situations

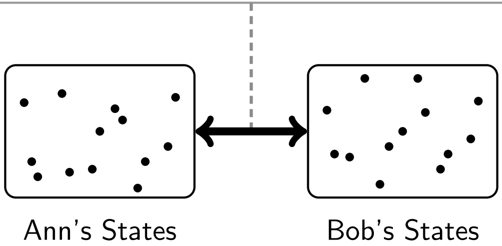
- Various states of information disclosure.
  - *ex ante*, *ex interim*, *ex post*
- Various "types" of information:
  - imperfect information about the play of the game
  - incomplete information about the structure of the game
  - strategic information (what will the other players do?)
  - higher-order information (what are the other players thinking?)
- Varieties of informational attitudes
  - hard ("knowledge")
  - soft ("beliefs")

# Two key assumptions



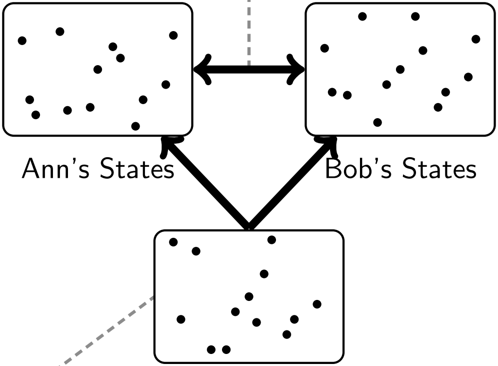
# Two key assumptions

1. The players recognize that they are in a game situation

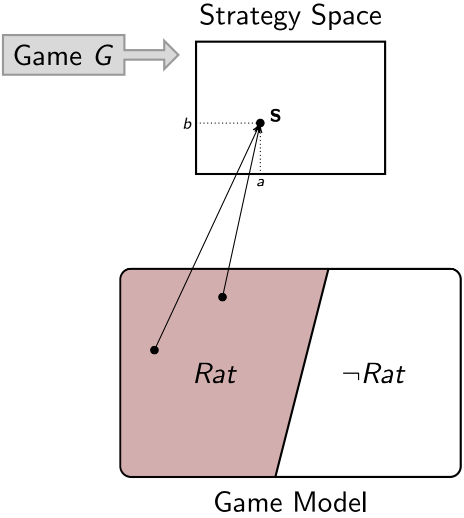


# Two key assumptions

1. The players recognize that they are in a game situation




2. The players *agree* on a common initial model




- Each state in a game model is associated with a strategy profile *and* a description of the players beliefs.
- *Rat* is event that the players optimize (and there is common belief that they optimize)
- "The viewpoint is *descriptive*. Not 'why,' not 'should,' just *what*. Not that $i$ does $a$ *because* he believes $E$; simply that he does $a$ and believes $E$."

# Harsanyi Type Space

Based on the work of John Harsanyi on games with *incomplete information*, game theorists have developed an elegant formalism that makes precise talk about beliefs, knowledge and rationality:

## Harsanyi Type Space

Based on the work of John Harsanyi on games with *incomplete information*, game theorists have developed an elegant formalism that makes precise talk about beliefs, knowledge and rationality:

- A **type** is everything a player knows privately at the beginning of the game which could affect his beliefs about payoffs and about all other players' possible types.

# Harsanyi Type Space

Based on the work of John Harsanyi on games with *incomplete information*, game theorists have developed an elegant formalism that makes precise talk about beliefs, knowledge and rationality:

- A **type** is everything a player knows privately at the beginning of the game which could affect his beliefs about payoffs and about all other players' possible types.
- Each **type** is assigned a joint probability over the space of types and actions

$$\lambda_i : T_i \to \Delta(T_{-i} \times S_{-i})$$

# Harsanyi Type Space

Based on the work of John Harsanyi on games with *incomplete information*, game theorists have developed an elegant formalism that makes precise talk about beliefs, knowledge and rationality:

- A **type** is everything a player knows privately at the beginning of the game which could affect his beliefs about payoffs and about all other players' possible types.

- Each **type** is assigned a joint probability over the space of types and actions

$$\lambda_i : T_i \rightarrow \Delta(T_{-i} \times S_{-i})$$

Player $i$'s types

# Harsanyi Type Space

Based on the work of John Harsanyi on games with *incomplete information*, game theorists have developed an elegant formalism that makes precise talk about beliefs, knowledge and rationality:

- A **type** is everything a player knows privately at the beginning of the game which could affect his beliefs about payoffs and about all other players' possible types.

- Each **type** is assigned a joint probability over the space of types and actions

$$\lambda_i : T_i \rightarrow \Delta(T_{-i} \times S_{-i})$$

The set of all probability distributions

## Harsanyi Type Space

Based on the work of John Harsanyi on games with *incomplete information*, game theorists have developed an elegant formalism that makes precise talk about beliefs, knowledge and rationality:

- A **type** is everything a player knows privately at the beginning of the game which could affect his beliefs about payoffs and about all other players' possible types.

- Each **type** is assigned a joint probability over the space of types and actions

$$\lambda_i : T_i \to \Delta(T_{-i} \times S_{-i})$$

The other players' types

# Harsanyi Type Space

Based on the work of John Harsanyi on games with *incomplete information*, game theorists have developed an elegant formalism that makes precise talk about beliefs, knowledge and rationality:

- A **type** is everything a player knows privately at the beginning of the game which could affect his beliefs about payoffs and about all other players' possible types.

- Each **type** is assigned a joint probability over the space of types and actions

$$\lambda_i : T_i \to \Delta(T_{-i} \times S_{-i})$$

The other players' choices

# Returning to the Example: A Game Model

| | | Bob $H$ | Bob $M$ |
|---|---|---|---|
| Ann | $H$ | 3,3 | 0,0 |
| Ann | $M$ | 0,0 | 1,1 |

$t_A$

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$u_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

# Returning to the Example: A Game Model

**Bob**

| Ann | $H$ | $M$ |
|---|---|---|
| $H$ | 3,3 | 0,0 |
| $M$ | 0,0 | 1,1 |

- One type for Ann ($t_A$) and two types for Bob ($t_B, u_B$)

$t_A$

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$u_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

# Returning to the Example: A Game Model

| | | Bob $H$ | Bob $M$ |
|---|---|---|---|
| Ann | $H$ | 3,3 | 0,0 |
| Ann | $M$ | 0,0 | 1,1 |

- One type for Ann ($t_A$) and two types for Bob ($t_B, u_B$)
- A **state** is a tuple of choices and types: $(M, M, t_A, t_B)$

$t_A$

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$u_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

# Returning to the Example: A Game Model

Bob

| Ann | $H$ | $M$ |
|---|---|---|
| $H$ | 3,3 | 0,0 |
| $M$ | 0,0 | 1,1 |

- One type for Ann ($t_A$) and two types for Bob ($t_B, u_B$)
- A **state** is a tuple of choices and types: $(M, t_A, M, u_B)$
- Calculate **expected utility** in the usual way...

$t_A$

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$u_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

# Returning to the Example: A Game Model

| | | Bob $H$ | Bob $M$ |
|---|---|---|---|
| Ann | $H$ | 3,3 | 0,0 |
| Ann | $M$ | 0,0 | 1,1 |

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_A$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$t_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

$u_B$

# Returning to the Example: A Game Model

Bob

| | | $H$ | $M$ |
|---|---|---|---|
| Ann | $H$ | 3,3 | 0,0 |
| | $M$ | 0,0 | 1,1 |

- $M$ is **rational** for Ann ($t_A$)
  $0 \cdot 0.2 + 1 \cdot 0.8 \geq 3 \cdot 0.2 + 0 \cdot 0.8$

$t_A$

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$u_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

# Returning to the Example: A Game Model

| Ann \ Bob | $H$ | $M$ |
|---|---|---|
| $H$ | 3,3 | 0,0 |
| $M$ | 0,0 | 1,1 |

- $M$ is **rational** for Ann ($t_A$)
  $0 \cdot 0.2 + 1 \cdot 0.8 \geq 3 \cdot 0.2 + 0 \cdot 0.8$
- $M$ is **rational** for Bob ($t_B$)
  $0 \cdot 0 + 1 \cdot 1 \geq 3 \cdot 0 + 0 \cdot 1$

$t_A$

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$u_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

# Returning to the Example: A Game Model

| | | Bob $H$ | Bob $M$ |
|---|---|---|---|
| Ann | $H$ | 3,3 | 0,0 |
| Ann | $M$ | 0,0 | 1,1 |

- $M$ is **rational** for Ann ($t_A$)
  $0 \cdot 0.2 + 1 \cdot 0.8 \geq 3 \cdot 0.2 + 0 \cdot 0.8$
- $M$ is **rational** for Bob ($t_B$)
  $0 \cdot 0 + 1 \cdot 1 \geq 3 \cdot 0 + 0 \cdot 1$
- Ann thinks Bob may be irrational

$t_A$:

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_B$:

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$u_B$:

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

# Returning to the Example: A Game Model

| Ann \ Bob | $H$ | $M$ |
|---|---|---|
| $H$ | 3,3 | 0,0 |
| $M$ | 0,0 | 1,1 |

- $M$ is **rational** for Ann ($t_A$)
  $0 \cdot 0.2 + 1 \cdot 0.8 \geq 3 \cdot 0.2 + 0 \cdot 0.8$
- $M$ is **rational** for Bob ($t_B$)
  $0 \cdot 0 + 1 \cdot 1 \geq 3 \cdot 0 + 0 \cdot 1$
- Ann thinks Bob may be irrational
  $P_A(Irrat[B]) = 0.3$, $P_A(Rat[B]) = 0.7$

$t_A$

| | $H$ | $M$ |
|---|---|---|
| $t_B$ | 0 | 0.5 |
| $u_B$ | 0.2 | 0.3 |

$t_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0 | 1 |

$u_B$

| | $H$ | $M$ |
|---|---|---|
| $t_A$ | 0.4 | 0.6 |

"*Common Knowledge*" is informally described as what any fool would know, given a certain situation: It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record.

"*Common Knowledge*" is informally described as what any fool would know, given a certain situation: It encompasses what is relevant, agreed upon, established by precedent, assumed, being attended to, salient, or in the conversational record.

*It is not Common Knowledge who "defined" Common Knowledge!*

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

**Fixed-point definition**: $\gamma :=$ *i* and *j* know that ($\varphi$ and $\gamma$)

G. Harman. *Review of* Linguistic Behavior. Language (1977).

J. Barwise. *Three views of Common Knowledge*. TARK (1987).

The first formal definition of common knowledge?

M. Friedell. *On the Structure of Shared Awareness*. Behavioral Science (1969).

R. Aumann. *Agreeing to Disagree*. Annals of Statistics (1976).

The first rigorous analysis of common knowledge

D. Lewis. *Convention, A Philosophical Study*. 1969.

**Fixed-point definition**: $\gamma := \ i$ and $j$ know that ($\varphi$ and $\gamma$)

G. Harman. *Review of* Linguistic Behavior. Language (1977).

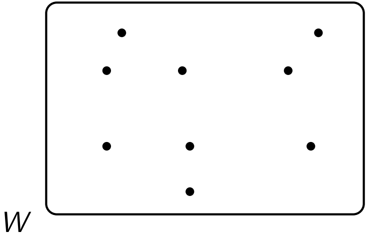
J. Barwise. *Three views of Common Knowledge*. TARK (1987).

**Shared situation**: There is a *shared situation s* such that (1) *s* entails $\varphi$, (2) *s* entails everyone knows $\varphi$, plus other conditions

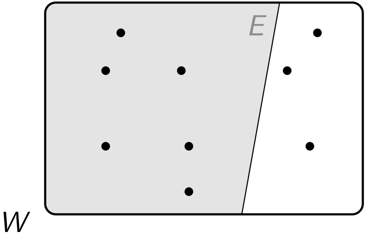
H. Clark and C. Marshall. *Definite Reference and Mutual Knowledge*. 1981.

M. Gilbert. *On Social Facts*. Princeton University Press (1989).

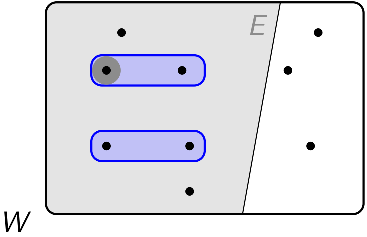
P. Vanderschraaf and G. Sillari. *"Common Knowledge", The Stanford Encyclopedia of Philosophy (2009)*.
`http://plato.stanford.edu/entries/common-knowledge/`.



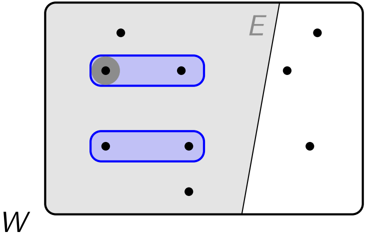

$W$ is a set of **states** or **worlds**.



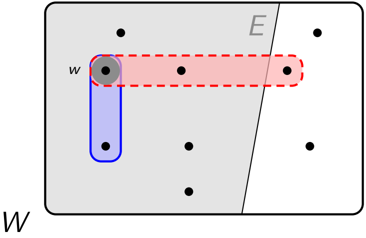

An **event**/**proposition** is any (definable) subset $E \subseteq W$



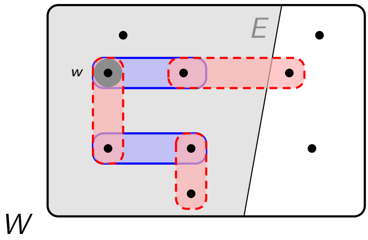

The agents receive signals in each state. States are considered equivalent for the agent if they receive the same signal in both states.



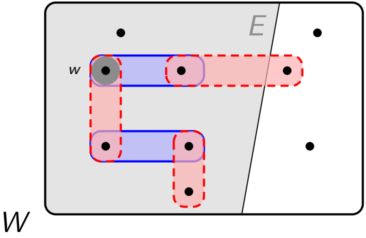

**Knowledge Function**: $K_i : \wp(W) \rightarrow \wp(W)$ where $K_i(E) = \{w \mid R_i(w) \subseteq E\}$



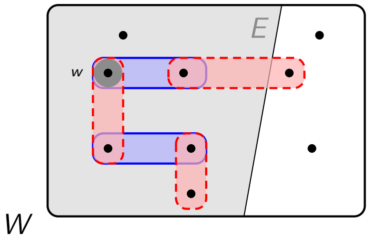

$w \in K_A(E)$ and $w \notin K_B(E)$




The model also describes the agents' **higher-order knowledge/beliefs**

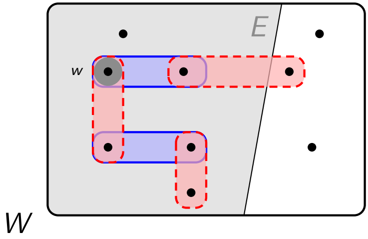



**Everyone Knows**: $K(E) = \bigcap_{i\in\mathcal{A}} K_i(E)$, $K^0(E) = E$, $K^m(E) = K(K^{m-1}(E))$

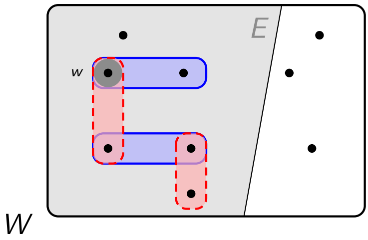



**Common Knowledge**: $C : \wp(W) \rightarrow \wp(W)$ with

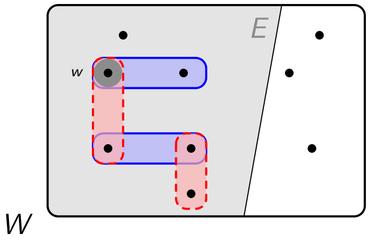
$$C(E) = \bigcap_{m \geq 0} K^m(E)$$




$w \in K(E)$ $\qquad$ $w \notin C(E)$




$w \in C(E)$




**Fact.** $w \in C(E)$ if every finite path starting at $w$ ends in a state in $E$

# An Example

Two players Ann and Bob are told that the following will happen. Some positive integer $n$ will be chosen and *one* of $n$, $n + 1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

# An Example

Two players Ann and Bob are told that the following will happen. Some positive integer $n$ will be chosen and *one* of $n$, $n + 1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

# An Example

Two players Ann and Bob are told that the following will happen. Some positive integer $n$ will be chosen and *one* of $n$, $n + 1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?

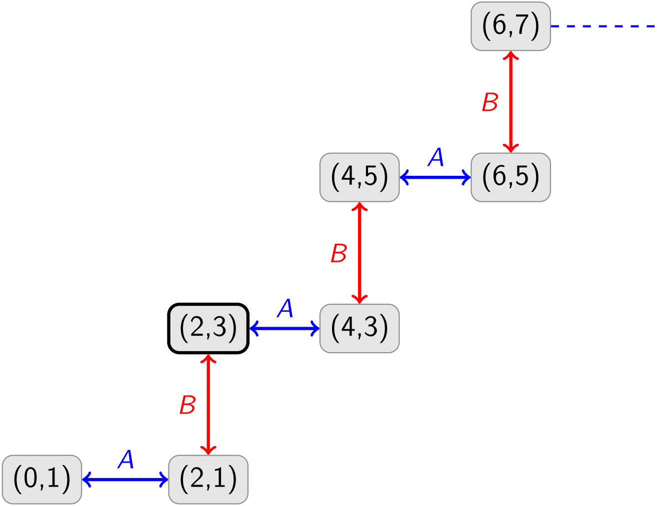
# An Example

Two players Ann and Bob are told that the following will happen. Some positive integer $n$ will be chosen and *one* of $n$, $n+1$ will be written on Ann's forehead, the other on Bob's. Each will be able to see the other's forehead, but not his/her own.

Suppose the number are (2,3).

Do the agents know there numbers are less than 1000?

Is it common knowledge that their numbers are less than 1000?


(6,7)
B
(4,5)
A
(6,5)
B
(2,3)
A
(4,3)
B
(0,1)
A
(2,1)

**Fact.** For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

**Fact.** For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

> Suppose you are told "Ann and Bob are going together,"' and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. ...the event "Ann and Bob are going together" — call it $E$ — is common knowledge if and only if some event — call it $F$ — happened that entails $E$ and also entails all players' knowing $F$ (like all players met Ann and Bob at an intimate party). *(Aumann, pg. 271, footnote 8)*

**Fact.** For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event $F$ is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

**Fact.** An event $E$ is commonly known iff some self-evident event that entails $E$ obtains.

**Fact.** For all $i \in \mathcal{A}$ and $E \subseteq W$, $K_i C(E) = C(E)$.

An event $F$ is **self-evident** if $K_i(F) = F$ for all $i \in \mathcal{A}$.

**Fact.** An event $E$ is commonly known iff some self-evident event that entails $E$ obtains.

**Fact.** $w \in C(E)$ if every finite path starting at $w$ ends in a state in $E$

The following axiomatize common knowledge:

- $C(\varphi \to \psi) \to (C\varphi \to C\psi)$
- $C\varphi \to (\varphi \wedge EC\varphi)$ (Fixed-Point)
- $C(\varphi \to E\varphi) \to (\varphi \to C\varphi)$ (Induction)

# The Fixed-Point Definition

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E))$

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E))$

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E))$

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$

## The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of $f_E$: $f_E(\bot) = \bot$

## The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i\in\mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i\in\mathcal{A}} K_i(C(E)) = \bigcap_{i\in\mathcal{A}} C(E) = C(E)$
- The are other fixed points of $f_E$: $f_E(\perp) = \perp$
- $f_E$ is monotonic:

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of $f_E$: $f_E(\bot) = \bot$
- $f_E$ is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$.

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of $f_E$: $f_E(\bot) = \bot$
- $f_E$ is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$. Then $f_E(E \cap A) = K(E \cap A) \subseteq K(E \cap B) = f_E(E \cap B)$

## The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of $f_E$: $f_E(\bot) = \bot$
- $f_E$ is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$. Then $f_E(E \cap A) = K(E \cap A) \subseteq K(E \cap B) = f_E(E \cap B)$
- (Tarski) Every monotone operator has a greatest (and least) fixed point

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of $f_E$: $f_E(\bot) = \bot$
- $f_E$ is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$. Then $f_E(E \cap A) = K(E \cap A) \subseteq K(E \cap B) = f_E(E \cap B)$
- (Tarski) Every monotone operator has a greatest (and least) fixed point
- Let $K^*(E)$ be the greatest fixed point of $f_E$.

# The Fixed-Point Definition

$f_E(X) = K(E \cap X) = \bigcap_{i \in \mathcal{A}} K_i(E \cap X)$

- $C(E)$ is a fixed point of $f_E$: $f_E(C(E)) = K(E \cap C(E)) = K(C(E)) = \bigcap_{i \in \mathcal{A}} K_i(C(E)) = \bigcap_{i \in \mathcal{A}} C(E) = C(E)$
- The are other fixed points of $f_E$: $f_E(\perp) = \perp$
- $f_E$ is monotonic: $A \subseteq B$ implies $E \cap A \subseteq E \cap B$. Then $f_E(E \cap A) = K(E \cap A) \subseteq K(E \cap B) = f_E(E \cap B)$
- (Tarski) Every monotone operator has a greatest (and least) fixed point
- Let $K^*(E)$ be the greatest fixed point of $f_E$.
- **Fact**. $K^*(E) = C(E)$.

# The Fixed-Point Definition

Separating the fixed-point/iteration definition of common knowledge/belief:


J. Barwise. *Three views of Common Knowledge*. TARK (1987).

J. van Benthem and D. Saraenac. *The Geometry of Knowledge*. Aspects of Universal Logic (2004).

A. Heifetz. *Iterative and Fixed Point Common Belief*. Journal of Philosophical Logic (1999).