Mini-course on Epistemic Game Theory
Lecture 1: Common Belief in Rationality

Andrés Perea
EpiCenter & Dept. of Quantitative Economics

EPICENTER
Research Center for Epistemic Game Theory

Maastricht University

Toulouse, June/July 2015

Introduction

- **Game theory** studies situations where you make a decision, but where the final outcome also depends on the choices of **others.**
- Before you make a choice, it is natural to **reason** about your opponents – about their **choices** but also about their **beliefs**.
- Oskar Morgenstern, in 1935, already stresses the importance of such reasoning for games.

- **Classical game theory** has focused mainly on the **choices** of the players.
- **Epistemic game theory** asks: Where do these choices come from?
- More precisely, it studies the **beliefs** that motivate these choices.
- Since the late 80's it has developed a broad spectrum of **epistemic concepts** for games.
- Some of these characterize **existing** concepts in classical game theory, others provide **new** ways of reasoning.

- **This course** studies some of these epistemic concepts.
- For every concept we present the **intuitive idea**, and show how it can be formalized as a collection of **restrictions on the players' beliefs**.
- For every concept we characterize the **choices** they induce.
- We also study **algorithms,** which can be used to compute these choices.

Outline

Part I: Static games

Lecture 1: Common belief in rationality

Lecture 2: Nash equilibrium

Part II: Dynamic games

Lecture 3: Backward induction reasoning

Lecture 4: Forward induction reasoning

The course is based on my **textbook** "*Epistemic Game Theory: Reasoning and Choice*".

Published by *Cambridge University Press* in July 2012.

Common belief in rationality: Idea

- In a game, you form a **belief** about the opponents' choices, and make a choice that is **optimal** for this belief.
- That is, you **choose rationally** given your belief.
- It seems reasonable to believe that your opponents will choose rationally as well, ...
- and that your opponents believe that the others will choose rationally as well, and so on.
- **Common belief in rationality.**

Example: Going to a party

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	2	1	4	3	0

Story

- This evening, you are going to a party together with your friend Barbara.
- You must both decide which color to wear: *blue, green, red* or *yellow.*
- You both dislike wearing the same color as the friend.

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	2	1	4	3	0

- Choosing *blue* is optimal if you believe that Barbara chooses *green.*
- Choosing *green* is optimal if you believe that Barbara chooses *blue.*
- Choosing *red* is optimal if you believe that, with probability 0.6, Barbara chooses *blue,* and that with probability 0.4 she chooses *green.*
- Choosing *yellow* is not optimal for you for any belief.
- So, *blue, green* and *red* are **rational** choices for you, *yellow* is **irrational** for you.

	blue	green	red	yellow	same color as friend
you	4	3	2	$\times$	0
Barbara	2	1	4	3	0

- If you believe that Barbara chooses rationally, and believe that Barbara believes that you choose rationally,

  then you believe that Barbara will **not** choose *blue* or *green*.

	blue	green	red	yellow	same color as friend
you	4	3	2	$\times$	0
Barbara	$\times$	$\times$	4	3	0

- But then, your unique **optimal** choice is *blue*.
- So, under **common belief in rationality,** you can only rationally wear *blue*.

New Scenario

- Barbara has same preferences over colors as you.
- Barbara **likes** to wear the same color as you, whereas you **hate** this.

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	4	3	2	1	5

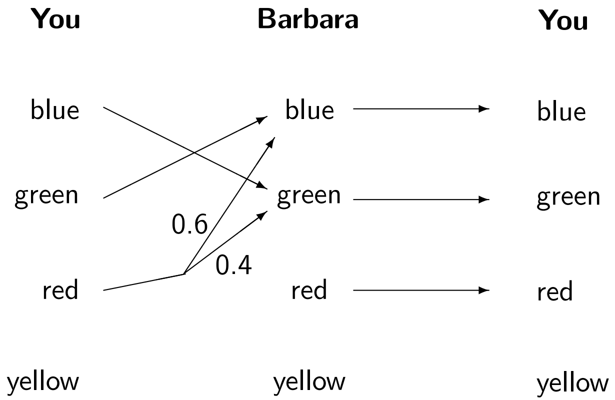
Which color(s) can you rationally choose under **common belief in rationality?**

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	4	3	2	1	5

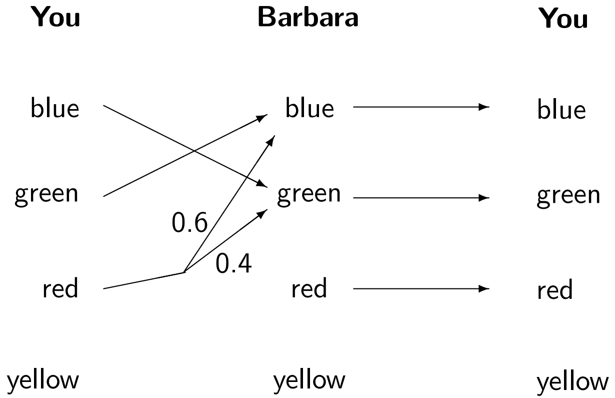
- If you choose rationally, you will not choose *yellow*.
- If you believe that Barbara chooses rationally, and believe that Barbara believes that you choose rationally, then you believe that Barbara will not choose *yellow* either.

	blue	green	red	yellow	same color as friend
you	4	3	2	$\times$	0
Barbara	4	3	2	$\times$	5

Beliefs diagram



	blue	green	red	yellow	same color as friend
you	4	3	2	$\times$	0
Barbara	4	3	2	$\times$	5



- The belief hierarchy that starts at your choice *blue* expresses **common belief in rationality.**
- Similarly, the belief hierarchies that start at your choices *green* and *red* also express **common belief in rationality.**
- So, you can rationally choose *blue, green* and *red* under **common belief in rationality.**

Epistemic model

In order to formally define **common belief in rationality,** we need to specify ...

- your belief about the opponents' choices,
- your belief about the opponents' beliefs about their opponents' choices,
- and so on, *ad infinitum.*

That is, we need to specify your complete **belief hierarchy.**

But how can we write down an **infinite** belief hierarchy?

In an infinite **belief hierarchy,** you hold a belief about ...

- the opponent's **choice,**
- the opponent's **first-order belief** about his opponents' choices,
- the opponent's **second-order belief** about his opponents' first-order beliefs,
- ...

That is, in an infinite **belief hierarchy,** you hold a belief about the opponent's **choice** and the opponent's infinite **belief hierarchy.**

Following Harsanyi (1967 / 1968), we call such a belief hierarchy a **type.**

Definition (Static game)

A **finite static game** Γ consists of

- a finite set of players $I = \{1, ..., n\}$,

- a finite set of choices C_i for every player, and

- a utility function $u_i : C_1 \times ... \times C_n \to \mathbb{R}$.

Definition (Epistemic model)

An **epistemic model** specifies for every player i a finite set T_i of possible **types**.

Moreover, for every type t_i it specifies a **probabilistic belief** $b_i(t_i)$ over the set $C_{-i} \times T_{-i}$ of opponents' choice-type combinations.

- **Implicit** epistemic model: For every type, we can **derive** the **complete belief hierarchy** induced by it.
- This is the model as used by Tan and Werlang (1988).
- Builds upon work by Harsanyi (1967 / 1968), Armbruster and Böge (1979), Böge and Eisele (1979), and Bernheim (1984).

Common Belief in Rationality: Definition

- **Remember:** A type t_i holds a belief $b_i(t_i)$ over the set $C_{-i} \times T_{-i}$ of opponents' choice-type combinations.

Definition (Belief in the opponents' rationality)

Type t_i **believes in the opponents' rationality** if his belief $b_i(t_i)$ only assigns positive probability to opponents' choice-type pairs (c_j, t_j), where choice c_j is optimal for type t_j.

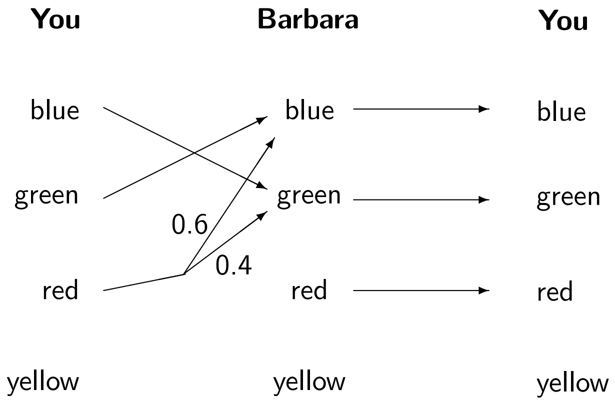
Definition (Common belief in rationality)

(Induction start) Type t_i expresses **1-fold** belief in rationality if t_i believes in the opponents' rationality.

(Inductive step) For every $k \geq 2$, type t_i expresses k-**fold** belief in rationality if t_i only assigns positive probability to opponents' types that express $(k-1)$-fold belief in rationality.

Type t_i expresses **common belief in rationality** if t_i expresses k-fold belief in rationality for all k.

- This definition is based on Tan and Werlang (1988) and Brandenburger and Dekel (1987).
- In terms of **choices** induced, it corresponds to the **pre-epistemic** concept of **rationalizability** (Bernheim (1984), Pearce (1984)).



- Can be transformed into **epistemic model** with types $t_1^{blue}, t_1^{green}, t_1^{red}$ and $t_2^{blue}, t_2^{green}, t_2^{red}$.
- Type t_1^{red} has belief

$$b_1(t_1^{red}) = (0.6) \cdot (blue, t_2^{blue}) + (0.4) \cdot (green, t_2^{green}).$$

	blue	green	red	yellow	same color as friend
you	4	3	2	1	0
Barbara	4	3	2	1	5

Beliefs for player 1	$b_1(t_1^{blue}) = (green,\ t_2^{green})$ $b_1(t_1^{green}) = (blue,\ t_2^{blue})$ $b_1(t_1^{red}) = (0.6)\cdot(blue,\ t_2^{blue}) + (0.4)\cdot(green,\ t_2^{green})$
Beliefs for player 2	$b_2(t_2^{blue}) = (blue,\ t_1^{blue})$ $b_2(t_2^{green}) = (green,\ t_1^{green})$ $b_2(t_2^{red}) = (red,\ t_1^{red})$

- Each of your types t_1^{blue}, t_1^{green} and t_1^{red} expresses **common belief in rationality.**
- So, you can rationally choose *blue, green* and *red* under **common belief in rationality**.

Algorithm

- Suppose we wish to find those choices you can rationally make under **common belief in rationality.**
- Is there an **algorithm** that helps us find these choices?

We start with a **basic question:** Which choices can be optimal for **some** belief about the opponents' choices?

Lemma (Pearce (1984))

A choice c_i is **optimal for some probabilistic belief** *about the opponents' choices, if and only if, c_i is* **not strictly dominated** *by any randomized choice.*

- Here, a **randomized choice** r_i for player i is a probability distribution on i's choices.

- Choice c_i is **strictly dominated** by the randomized choice r_i if

$$u_i(c_i, c_{-i}) < u_i(r_i, c_{-i})$$

for every opponents' choice-combination $c_{-i} \in C_{-i}$.

Step 1: 1-fold belief in rationality

- Which choices are rational for a type that expresses **1-fold** belief in rationality?
- Consider a type t_i that expresses 1-fold belief in rationality. Then, t_i only assigns positive probability to opponents' choice-type pairs (c_j, t_j) where c_j is optimal for t_j.
- We know from Pearce's Lemma that every such choice c_j is **not strictly dominated** within the game Γ.
- Let Γ^1 be the game obtained from Γ by **eliminating** all strictly dominated choices.
- So, t_i only assigns positive probability to choices in Γ^1.
- **Conclusion:** Every type t_i that expresses 1-fold belief in rationality, only assigns positive probability to choices in Γ^1.

Step 1: 1-fold belief in rationality

- **Conclusion:** Every type t_i that expresses 1-fold belief in rationality, only assigns positive probability to choices in Γ^1.
- Which choices can t_i rationally choose himself?
- By Pearce's Lemma, every choice c_i that is optimal for t_i must **not** be **strictly dominated** within Γ^1.
- Let Γ^2 be the game obtained from Γ^1 by eliminating all strictly dominated choices from Γ^1.
- So, every choice that is optimal for t_i must be in Γ^2.
- **Conclusion:** A type t_i that expresses 1-fold belief in rationality, can only rationally make choices from Γ^2.

Step 2: Up to 2-fold belief in rationality

- Which choices are rational for a type that expresses **up to 2-fold** belief in rationality?
- Consider a type t_i that expresses up to 2-fold belief in rationality. Then, t_i only assigns positive probability to opponents' choice-type pairs (c_j, t_j) where c_j is optimal for t_j, and t_j expresses 1-fold belief in rationality.
- By Step 1, every such choice c_j is in Γ^2.
- Hence, type t_i only assigns positive probability to opponents' choices in Γ^2.

Step 2: Up to 2-fold belief in rationality

- Type t_i only assigns positive probability to opponents' choices in Γ^2.
- Which choices can t_i rationally choose himself?
- By Pearce's Lemma, every choice c_i that is optimal for t_i must **not** be **strictly dominated** within Γ^2.
- Let Γ^3 be the game obtained from Γ^2 by eliminating all strictly dominated choices from Γ^2.
- So, every choice that is optimal for t_i must be in Γ^3.
- **Conclusion:** A type t_i that expresses up to 2-fold belief in rationality, can only rationally make choices from Γ^3.

Algorithm (Iterated elimination of strictly dominated choices)

Consider a finite static game Γ.

(Induction start) Let $\Gamma^0 := \Gamma$ *be the original game.*

(Inductive step) For every $k \geq 1$, *let* Γ^k *be the game which results if we eliminate from* Γ^{k-1} *all choices that are strictly dominated within* Γ^{k-1}.

- This algorithm terminates within finitely many steps. That is, there is some K with $\Gamma^{K+1} = \Gamma^K$.
- The choices in Γ^k are said to survive k-**fold elimination** of strictly dominated choices.
- The choices in Γ^K are said to survive **iterated elimination** of strictly dominated choices.

Theorem (Tan and Werlang (1988))

(1) For every $k \geq 1$, the choices that are optimal for a type that expresses **up to k-fold belief in rationality** *are exactly those choices that survive* **$(k+1)$-fold elimination** *of strictly dominated choices.*

(2) The choices that are optimal for a type that expresses **common belief in rationality** *are exactly those choices that survive* **iterated elimination** *of strictly dominated choices.*

Proof of part (2):

- We already know: If choice c_i is optimal for a type t_i that expresses **common belief in rationality,** then c_i must **survive** the algorithm.
- Still to show: If c_i **survives** the algorithm, then c_i is optimal for some type t_i that expresses **common belief in rationality.**

- Suppose that the algorithm terminates after K steps – that is, $\Gamma^{K+1} = \Gamma^K$. Let C_i^K be the set of surviving choices for player i.

- Then, every choice in C_i^K is not strictly dominated within Γ^K. Hence, by Pearce's Lemma, every choice c_i in C_i^K is optimal for some belief $b_i^{c_i} \in \Delta(C_{-i}^K)$.

- Define set of types $T_i = \{t_i^{c_i} : c_i \in C_i^K\}$ for every player i.

- Every type $t_i^{c_i}$ only deems possible opponents' choice-type pairs $(c_j, t_j^{c_j})$, with $c_j \in C_j^K$, and

$$b_i(t_i^{c_i})((c_j, t_j^{t_j})_{j \neq i}) := b_i^{c_i}((c_j)_{j \neq i}).$$

- Then, every type $t_i^{c_i}$ **believes in the opponents' rationality.**

- Hence, every type expresses **common belief in rationality.**

Example: The number machine

Story

- In a casino in Las Vegas you see a remarkable machine, that says "Guess two-thirds of the average and you will be rich!"
- After putting in 5 dollars, you must enter a number between 1 and 100.
- The closer your number is to **two-thirds of the average** of all the numbers previously entered, the higher your prize-money.
- What number should you choose?

- What numbers can you rationally choose under **common belief in rationality?** Use the **algorithm.**

- Clearly, any number above $(2/3) \cdot 100 \approx 67$ is strictly dominated. So, **eliminate** every number **above 67.**

- In the reduced game Γ^1, every number above $(2/3) \cdot 67 \approx 45$ is strictly dominated. So, **eliminate** every number **above 45.**

- In the reduced game Γ^2, every number above $(2/3) \cdot 45 = 30$ is strictly dominated. So, **eliminate** every number **above 30.**

- In the reduced game Γ^3, every number above $(2/3) \cdot 30 = 20$ is strictly dominated. So, **eliminate** every number **above 20.**

And so on.

- In this way, you will eliminate every number **except** for the lowest number 1.
- Hence, the **only** number you can rationally choose under **common belief in rationality** is **1!**
- But is this **realistic?**

- W. Armbruster and W. Böge, ‘Bayesian game theory ’ in O. Moeschlin and D. Pallaschke (eds.), *Game Theory and Related Topics* (North-Holland, Amsterdam, 1979)
- B.D. Bernheim, ‘Rationalizable strategic behavior’, *Econometrica*, 52 (1984), 1007–1028
- W. Böge and T.H. Eisele, ‘On solutions of bayesian games’, *International Journal of Game Theory*, 8 (1979), 193–215
- A. Brandenburger and E. Dekel, ‘Rationalizability and correlated equilibria’, *Econometrica*, 55 (1987), 1391–1402
- J.C. Harsanyi, ‘Games with incomplete information played by “bayesian” players, I–III’, *Management Science*, 14 (1967–1968), 159–182, 320–334, 486–502

- O. Morgenstern, ‘Vollkommene Voraussicht und wirtschaftliches Gleichgewicht’, *Zeitschrift für Nationalökonomie*, 6 (1935), 337–357. (Reprinted as ‘Perfect foresight and economic equilibrium’ in A. Schotter (ed.), *Selected Economic Writings of Oskar Morgenstern* (New York University Press, 1976), pp. 169–183)
- D. Pearce, ‘Rationalizable strategic behavior and the problem of perfection’, *Econometrica*, 52 (1984), 1029–1050
- T. Tan and S.R.C. Werlang, ‘The bayesian foundations of solution concepts of games’, *Journal of Economic Theory*, 45 (1988), 370–391