An Epistemic Course in Game Theory

Andrés Perea

© 2007
1 Introduction

1.1 What is Game Theory about?

Game Theory is about situations in which you make a choice, but where the final outcome depends also on the choices of others.

Examples: Everyday life, Managerial decision making, Playing chess, etcetera.

Such situations are called games, and the persons involved are called players.
1.2 Epistemic models

Before you make a choice in a game, you must have a belief about what your opponents will do.

That is, you must reason about your opponents.

Epistemic models in game theory try to formalize possible ways in which players may reason about their opponents.
1.3 My approach in this course

I will investigate games from the **perspective of a single player**.

I will use the following general procedure:

1. Present an **intuitive way of reasoning** about my opponents.

2. Develop an **epistemic model** for this way of reasoning.

3. Which **choices** can I rationally make if I follow this way of reasoning?

4. Can these choices be computed by means of an **algorithm**?
1.4 Outline of this course

**Lecture 1:** Common belief in rationality

**Lecture 2:** Self-referential reasoning, cautious reasoning.

**Lecture 3:** Belief revision in dynamic games

**Lecture 4:** Belief in the opponents’ future rationality, strong belief in the opponents’ rationality, proper belief revision.
1.5 Surveys on Epistemic Game Theory

Adam Brandenburger (2007): “The power of paradox: Some recent results in interactive epistemology”


Andrés Perea: “Reasoning about your opponents: An epistemic course in game theory” (in progress)
2 Common Belief in Rationality

**Intuitive idea:**

When making a choice, **you believe that your opponents choose rationally.**

You believe that your opponents reason in this way, too.

You believe that your opponents believe that their opponents reason in this way, too.

Etcetera.
2.1 Examples

Where to locate my pub?

Which locations can you rationally choose?
Your choices

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>e</td>
<td>f</td>
</tr>
<tr>
<td>g</td>
<td></td>
</tr>
</tbody>
</table>

Opponent's choices

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td></td>
</tr>
</tbody>
</table>
Are all your rational choices $b, c, d, e$ and $f$ **reasonable**?
If you believe that the opponent chooses rationally.

You \begin{align*}
a & \rightarrow a \\
b & \rightarrow b \\
c & \rightarrow c \\
d & \rightarrow d \\
e & \rightarrow e \\
f & \rightarrow f \\
g & \rightarrow g
\end{align*}

Opponent

You

\begin{align*}
a & \rightarrow a \\
b & \rightarrow b \\
c & \rightarrow c \\
d & \rightarrow d \\
e & \rightarrow e \\
f & \rightarrow f \\
g & \rightarrow g
\end{align*}
If you believe that the opponent believes that you choose rationally.

You | Opponent | You
---|----------|-----
\(a\) | \(a\) | \(a\)
\(b\) | \(b\) | \(b\)
\(c\) | \(c\) | \(c\)
\(d\) | \(d\) | \(d\)
\(e\) | \(e\) | \(e\)
\(f\) | \(f\) | \(f\)
\(g\) | \(g\) | \(g\)
If you believe that the opponent chooses rationally, and believe that your opponent believes that you choose rationally.
So, **common belief in rationality** leads to a unique choice for you, namely \( d \).
Which color to wear?

<table>
<thead>
<tr>
<th></th>
<th>blue</th>
<th>green</th>
<th>red</th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

If your friend chooses the same color as you, your utility would be 0.

Which colors are **rational**?
**Yellow** is strictly dominated by the randomized choice in which you choose blue and green with probability 0.5.

<table>
<thead>
<tr>
<th></th>
<th>blue</th>
<th>green</th>
<th>red</th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>friend</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
If you believe that your friend chooses rationally.

<table>
<thead>
<tr>
<th></th>
<th>blue</th>
<th>green</th>
<th>red</th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>friend</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
If you **believe that your friend believes that you choose rationally.**

![Diagram]

<table>
<thead>
<tr>
<th></th>
<th>blue</th>
<th>green</th>
<th>red</th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>friend</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
If you believe that your friend chooses rationally, and believe that your friend believes that you choose rationally.

<table>
<thead>
<tr>
<th></th>
<th>blue</th>
<th>green</th>
<th>red</th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>friend</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
So, *common belief in rationality* leads to a unique choice, namely wearing blue (your most preferred color).

Things change if we change your friend’s preferences.
So, **common belief in rationality** allows for choosing blue, green and red.
2.2 Epistemic Model

Types: \( \{t^b_1, t^g_1, t^r_1\} \) for you, and \( \{t^r_2, t^b_2, t^g_2\} \) for friend.

Your type \( t^b_1 \) believes that friend chooses \( g \) and has type \( t^g_2 \).
Friend’s type $t^g_2$ believes that, with prob. 0.4, you choose $b$ and have type $t^b_1$, and with prob. 0.6 you choose $r$ and have type $t^r_1$. 
So, if you are of type $t_1^b$, then you believe that

- friend chooses $g$,

- friend believes that you choose $b$ with prob 0.4 and you choose $r$ with prob. 0.6.
In fact, your type $t_1^b$ induces a full hierarchy of beliefs.
In general:

Let $\Gamma = (C_i, u_i)_{i \in I}$ be a finite, static game where

- $I$ is the finite set of players,
- $C_i$ is the finite set of choices for player $i$, and
- $u_i$ is player $i$’s utility function.

$u_i$ assigns to every combination of choices $(c_j)_{j \in I}$ a utility $u_i((c_j)_{j \in I}) \in \mathbb{R}$.
A **finite epistemic model** for $\Gamma$ is a tuple $M = (T_i, b_i)_{i \in I}$ where

- $T_i$ is the finite set of **types** for player $i$, and
- $b_i$ is a function that assigns to every type $t_i \in T_i$ a **probabilistic belief** $b_i(t_i) \in \Delta(C_{-i} \times T_{-i})$.

Here, $C_{-i} := \times_{j \neq i} C_j$, and $T_{-i} := \times_{j \neq i} T_j$.

$\Delta(C_{-i} \times T_{-i})$ denotes the set of probability distributions on $C_{-i} \times T_{-i}$. 
$T_1 = \{ t_1^b, t_1^g, t_1^r \}$

<table>
<thead>
<tr>
<th>$b_1(t_1^b) = (g, t_2^g)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_1(t_1^g) = (b, t_2^b)$</td>
</tr>
<tr>
<td>$b_1(t_1^r) = 0.6(b, t_2^b) + 0.4(g, t_2^g)$</td>
</tr>
</tbody>
</table>

$T_2 = \{ t_2^b, t_2^g, t_2^r \}$

<table>
<thead>
<tr>
<th>$b_2(t_2^b) = (b, t_1^b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_2(t_2^g) = (r, t_1^r)$</td>
</tr>
<tr>
<td>$b_2(t_2^r) = 0.4(b, t_1^b) + 0.6(r, t_1^r)$</td>
</tr>
</tbody>
</table>
Consider a type $t_i \in T_i$ and a choice $c_i \in C_i$. Then,

$$u_i(c_i, t_i) := \sum_{c_{-i} \in C_{-i}} b_i(t_i)(c_{-i}) u_i(c_i, c_{-i})$$

is $t_i$’s expected utility from choosing $c_i$.

Choice $c_i$ is rational for type $t_i$ if

$$u_i(c_i, t_i) \geq u_i(c_i', t_i)$$

for all other choices $c_i'$. 
2.3 Common Belief in Rationality

Type $t_i$ believes in the opponents’ rationality if $b_i(t_i)$ only assigns positive probability to choice-type pairs $(c_j, t_j)$ where $c_j$ is rational for $t_j$. 
Types $t^b_1$, $t^g_1$ and $t^r_1$ believe in the opponent’s rationality.

Types $t^b_2$, $t^g_2$ and $t^r_2$ believe in the opponent’s rationality.
For every player $j$, let $E_j \subseteq T_j$ be a set of player $j$ types.

Type $t_i$ believes in $E_j$ if $b_i(t_i)$ only assigns positive probability to player $j$ types in $E$.

**For example:** Take $E_j :=$ set of player $j$ types that believe in the opponents' rationality.

If $t_i$ believes in $E_j$, then $t_i$ believes that player $j$ believes in his opponents' rationality.
Common belief in rationality (Tan and Werlang, 1988)

\[
T_i^1 : = \{t_i \in T_i \mid t_i \text{ believes in the opponents' rationality}\}
\]
\[
T_i^2 : = \{t_i \in T_i \mid t_i \text{ believes in } T_j^1 \text{ for all } j \neq i\}
\]
\[
\vdots
\]
\[
T_i^k : = \{t_i \in T_i \mid t_i \text{ believes in } T_j^{k-1} \text{ for all } j \neq i\}
\]
\[
\vdots
\]

Type \( t_i \) respects common belief in rationality if

\[
t_i \in T_i^k \text{ for every } k.
\]
Types $t^b_1, t^g_1$ and $t^r_1$ respect **common belief in rationality**.

So, you can rationally choose $b, g$ and $r$ under common belief in rationality.
In general:

Player $i$ can \textbf{rationally choose} $c_i \in C_i$ \textbf{under common belief in rationality} if:

there is an epistemic model $M = (T_j, b_j)_{j \in I}$, and

a type $t_i \in T_i$ such that

$t_i$ respects \textbf{common belief in rationality}, and

$c_i$ is \textbf{rational} for $t_i$. 
2.4 Existence

Does player $i$ always have a choice that can rationally be chosen under common belief in rationality?

If the game is finite, then the answer is “yes”.
You \hspace{1cm} \text{Opponent} \hspace{1cm} \text{You}

\begin{align*}
& a \quad d \quad a \\
& b \quad e \quad b \\
& c \quad f \quad c
\end{align*}
Types $t_1^b$ and $t_1^c$ respect common belief in rationality.

Choices $b$ and $c$ can rationally be chosen under common belief in rationality.
**Theorem 2.1:**

Let $\Gamma = \left( C_i, u_i \right)_{i \in I}$ be a finite, static game. Then, every player $i$ has at least one choice $c_i$ that can be chosen rationally under common belief in rationality.

Moreover, one can always construct a finite epistemic model $M = \left( T_i, b_i \right)_{i \in I}$ in which

- every type respects common belief in rationality, and
- every type holds a **point-belief** about the opponents’ choice-type pairs.
2.5 Algorithm

Question: Is there an algorithm that computes all choices for player $i$ that can rationally be chosen under common belief in rationality?
Yellow is strictly dominated by the randomized choice in which you chooses blue and green with probability 0.5.

<table>
<thead>
<tr>
<th></th>
<th>blue</th>
<th>green</th>
<th>red</th>
<th>yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>friend</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
A choice $c_i$ is **rational** if there is some probabilistic belief $b_i \in \Delta(C_{-i})$ such that

$$u_i(c_i, b_i) \geq u_i(c_i', b_i)$$

for all other choices $c_i'$.

A **randomized choice** is a probability distribution $\mu_i \in \Delta(C_i)$.

Choice $c_i$ is **strictly dominated** by the randomized choice $c_i$ if

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

for every $c_{-i} \in C_{-i}$. 

Lemma 2.2.

A choice $c_i$ is **rational** if and only if it is **not strictly dominated** by a randomized choice.

---

Proof:

(1) Suppose, $c_i$ is strictly dominated by randomized choice $\mu_i$. Then, for every belief $b_i \in \Delta(C_{-i})$

$$u_i(c_i, b_i) < u_i(\mu_i, b_i).$$

So, for every belief $b_i$ there is some $c'_i \in \text{supp}(\mu_i)$ with

$$u_i(c_i, b_i) < u_i(c'_i, b_i).$$

Hence, $c_i$ cannot be rational.
Lemma 2.2.

A choice $c_i$ is **rational** if and only if it is **not strictly dominated** by a randomized choice.

---

**Proof:**

(2) Suppose, $c_i$ is not rational. To show: $c_i$ is strictly dominated by some randomized choice.

For simplicity, assume there are 2 players, $C_1 = \{a, b, c\}$ and $C_2 = \{d, e\}$. Suppose that $c$ is not rational.
Suppose, you believe that player 2 chooses $d$ with prob. $p$ and $e$ with prob. $1 - p$.

Dotted line corresponds to randomized choice $\mu_1$ over $a$ and $b$.

So, $c$ is strictly dominated by $\mu_1$. 
For every player \( j \), let \( D_j \subseteq C_j \). Let \( D_{-i} := \times_{j \neq i} D_j \).

Choice \( c_i \) is **rational on** \( D_{-i} \) if there is some probabilistic belief \( b_i \in \Delta(D_{-i}) \) for which \( c_i \) is optimal.

Choice \( c_i \) is **strictly dominated on** \( D_{-i} \) if there is some randomized choice \( \mu_i \in \Delta(C_i) \) such that
\[
    u_i(c_i, c_{-i}) < u_i(\mu_i, c_{-i})
\]
for every \( c_{-i} \in D_{-i} \).
Lemma 2.3.

Let $D_j \subseteq C_j$, and $D_{-i} := \times_{j \neq i} D_j$.

A choice $c_i$ is **rational on** $D_{-i}$ if and only if it is **not strictly dominated on** $D_{-i}$.

\[ C^1_i := \{ c_i \in C_i \mid c_i \text{ not strictly dominated on } C_{-i} \} . \]

So, $i$ chooses **rationally**

if and only

if he chooses from $C^1_i$. 
Lemma 2.3.

Let \( D_j \subseteq C_j \), and \( D_{-i} := \times_{j \neq i} D_j \).

A choice \( c_i \) is rational on \( D_{-i} \) if and only if it is not strictly dominated on \( D_{-i} \).

\[
C_i^2 := \{ c_i \in C_i^1 \mid c_i \text{ not strictly dominated on } C_{-i}^1 \}.
\]

So, \( i \) chooses rationally and believes in the opponents’ rationality if and only if

\( i \) chooses from \( C_i^2 \).
Lemma 2.3.

Let $D_j \subseteq C_j$, and $D_{-i} := \times_{j \neq i} D_j$.

A choice $c_i$ is rational on $D_{-i}$ if and only if it is not strictly dominated on $D_{-i}$.

\[ C_i^3 := \{ c_i \in C_i^2 \mid c_i \text{ not strictly dominated on } C_{-i}^2 \}. \]

So, $i$ chooses rationally, believes in the opponents’ rationality, and believes that his opponents believe in the opponents’ rationality,

if and only if

$i$ chooses from $C_i^3$, and so on.
Algorithm: Iterated elimination of strictly dominated choices.

\[ C_1^i : = \{ c_i \in C_i \mid c_i \text{ not strictly dominated on } C_{-i} \} \]
\[ C_2^i : = \{ c_i \in C_1^i \mid c_i \text{ not strictly dominated on } C_{-i}^1 \} \]
\[ \vdots \]
\[ C_k^i : = \{ c_i \in C_{k-1}^i \mid c_i \text{ not strictly dominated on } C_{-i}^{k-1} \} \]
\[ \vdots \]

Choice \( c_i \) survives iterated elimination of strictly dominated choices if \( c_i \in C_k^i \) for every \( k \).
Theorem 2.4: (Tan and Werlang, 1988)

A choice $c_i$ can rationally be chosen under common belief in rationality if and only if

$c_i$ survives iterated elimination of strictly dominated choices.
**Example:** The traveler’s dilemma (Basu, 1994)

If you choose lower amount: average + 100.

If you choose higher amount: average - 225.

<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100,100</td>
<td>250,−75</td>
<td>300,−25</td>
<td>350,25</td>
</tr>
<tr>
<td>200</td>
<td>−75,250</td>
<td>200,200</td>
<td>350,25</td>
<td>400,75</td>
</tr>
<tr>
<td>300</td>
<td>−25,300</td>
<td>25,350</td>
<td>300,300</td>
<td>450,125</td>
</tr>
<tr>
<td>400</td>
<td>25,350</td>
<td>75,400</td>
<td>125,450</td>
<td>400,400</td>
</tr>
</tbody>
</table>
Choosing 400 is strictly dominated by randomized choice $0.45(100) + 0.55(300)$.
<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>200</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>You</td>
<td>100, 100</td>
<td>250, -75</td>
<td>300, -25</td>
</tr>
<tr>
<td>200</td>
<td>-75, 250</td>
<td>200, 200</td>
<td>350, 25</td>
</tr>
<tr>
<td>300</td>
<td>-25, 300</td>
<td>25, 350</td>
<td>300, 300</td>
</tr>
</tbody>
</table>

Choosing 300 is strictly dominated by randomized choice $0.5(100) + 0.5(200)$. 
Choosing 200 is strictly dominated by choosing 100.

Hence, **common belief in rationality** leads to a unique choice, namely 100.
2.6 Related Models

In common belief in rationality, player i’s belief about j’s choice may be dependent on his belief about k’s choice.

For instance,

$$b_1(t_1^b) = \frac{1}{2}((t_2^g, g), (t_3^g, g)) + \frac{1}{2}((t_2^r, r), (t_3^r, r)).$$
In the model of **rationalizability** (Bernheim (1984) and Pearce (1984)) it is assumed that player $i$’s belief about $j$’s choice should be **independent** from $i$’s belief about $k$’s choice.

A choice $c_i$ is **rationalizable** if it can rationally be chosen under

- **common belief in rationality**, and
- **common belief in event that players’ beliefs about opponents are independent**.

For two-player games, rationalizability is equivalent to common belief in rationality.
In the model by Brandenburger and Friedenberg (2006) it is assumed that player
i’s belief about j’s choice is **conditionally independent** from i’s belief about
k’s choice.

For instance, the belief

$$b_1(t_1^b) = \frac{1}{2} \left( (t_2^g, g), (t_3^g, g) \right) + \frac{1}{2} \left( (t_2^r, r), (t_3^r, r) \right)$$

is conditionally independent, but not independent.

Brandenburger and Friedenberg (2006) assume common belief in rationality and
common belief in event that players have conditionally independent beliefs.