Royal Netherlands Academy of Arts and Sciences (KNAW)
Master Class
Amsterdam, February 8th, 2007

Epistemic Foundations of Game Theory

Lecture 1

Giacomo Bonanno

(http://www.econ.ucdavis.edu/faculty/bonanno/)
QUESTION:

What strategies can be chosen by rational players who know the structure of the game and the preferences of their opponents and who recognize each other’s rationality and knowledge?

Keywords: knowledge, rationality, recognition of each other’s knowledge and rationality
Modular approach

Module 1: representation of belief and knowledge of an individual (Hintikka, 1962; Kripke, 1963).

Module 2: extension to many individuals.
Common belief and common knowledge
(“recognition of each other’s belief / knowledge”)

Module 3: definition of rationality in games
(relationship between choice and beliefs)

QUESTION: what are the implications of rationality and common belief of rationality in games?
Module 1
representation of beliefs and knowledge of an individual

Finite set of states $\Omega$ and a binary relation $\mathcal{B}$ on $\Omega$.

$\alpha \mathcal{B} \beta$ means “at state $\alpha$ the individual considers state $\beta$ possible”

Notation: $\mathcal{B}(\omega) = \{ \omega' \in \Omega : \omega \mathcal{B} \omega' \}$ set of states considered possible at $\omega$

PROPERTIES $\forall \omega, \omega' \in \Omega,$

1. $\mathcal{B}(\omega) \neq \emptyset$ seriality
2. if $\omega' \in \mathcal{B}(\omega)$ then $\mathcal{B}(\omega') \subseteq \mathcal{B}(\omega)$ transitivity
3. if $\omega' \in \mathcal{B}(\omega)$ then $\mathcal{B}(\omega) \subseteq \mathcal{B}(\omega')$ euclideannes
Belief operator on events: \( B : 2^\Omega \rightarrow 2^\Omega \)

For \( E \subseteq \Omega, \; \omega \in BE \) if and only if \( B(\omega) \subseteq E \)

**EXAMPLE:**

\[
\begin{align*}
\neg p & \quad p \\
\alpha & \quad \beta \\
\neg p & \quad p \\
\gamma & \quad \delta
\end{align*}
\]

\( B(\alpha) = B(\beta) = \{ \alpha, \beta \} \)

\( B(\gamma) = B(\delta) = \{ \delta \} \)

Let \( E = \{ \beta, \delta \} \) : the event that represents the proposition \( p \)

Then \( BE = \{ \gamma, \delta \} \)
Properties of the belief operator: $\forall E \subseteq \Omega$

1. $BE \subseteq \neg B \neg E$ (consistency: follows from seriality of $\mathcal{B}$)
2. $BE \subseteq BBE$ (positive introspection: follows from transitivity of $\mathcal{B}$)
3. $\neg BE \subseteq B \neg BE$ (negative introspection: follows from euclideananness of $\mathcal{B}$)

*Mistaken beliefs are possible:* at $\gamma$ $p$ is false but the individual believes $p$

If $E = \{\beta, \delta\}$, then $\gamma \notin E$ but $\gamma \in BE = \{\gamma, \delta\}$
If - in addition to the previous properties - the "doxastic accessibility" relation $\mathcal{B}$ is **reflexive** ($\forall \omega \in \Omega, \; \omega \in \mathcal{B}(\omega)$) then it is an **equivalence relation** - giving rise to a **partition** of the set of states - and the associated belief operator satisfies the additional property that $\forall E \subseteq \Omega, \; BE \subseteq E$ (beliefs are correct). In this case we speak of **knowledge** and the associated operator is denoted by $K$ rather than $B$
Set of individuals $N$ and a binary relation $B_i$ for every $i \in N$

Let $E = \{\alpha, \beta, \gamma\}$: the event that represents the proposition $p$

Then $K_1E = \{\alpha, \beta, \gamma\}$, $K_2E = \{\alpha, \beta\}$

$K_1K_2E = \{\alpha\}$, $K_2K_1K_2E = \emptyset$
An event $E$ is *commonly believed* if (1) everybody believes it, (2) everybody believes that everybody believes it, (3) everybody believes that everybody believes that everybody believes it, etc.

Define the “everybody believes” operator $B^e$ as follows:

$$B^e E = B_1 E \cap B_2 E \cap \ldots \cap B_n E$$

The common belief operator $B_*$ is defined as follows:

$$B_* E = B^e E \cap B^e B^e E \cap B^e B^e B^e E \cap \ldots$$
Let $\mathcal{B}_*$ be the transitive closure of $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_n$. Thus $\omega' \in \mathcal{B}_*(\omega)$ if and only if there exists a sequence $\langle \omega_1, \ldots, \omega_m \rangle$ in $\Omega$ such that

1. $\omega_1 = \omega$
2. $\omega_n = \omega'$
3. for every $j = 1, \ldots, m$ there exists an individual $i \in N$ such that $\omega_{j+1} \in \mathcal{B}_i(\omega_j)$

$\mathcal{B}_1: \quad \begin{array}{cc}
\alpha & \beta \\
\bullet & \\
\gamma
\end{array} \quad \mathcal{B}_1(\alpha) = \mathcal{B}_1(\beta) = \{\alpha\}, \quad \mathcal{B}_1(\gamma) = \{\gamma\}$

$\mathcal{B}_2: \quad \begin{array}{ccc}
\alpha & \beta & \gamma \\
\bullet & \bullet & \\
\bullet
\end{array} \quad \mathcal{B}_2(\alpha) = \{\alpha\}, \quad \mathcal{B}_2(\beta) = \mathcal{B}_2(\gamma) = \{\beta, \gamma\}$

$\mathcal{B}_*: \quad \begin{array}{ccc}
\alpha & \beta & \gamma \\
\bullet & \bullet & \\
\bullet
\end{array} \quad \mathcal{B}_*(\alpha) = \{\alpha\}, \quad \mathcal{B}_*(\beta) = \mathcal{B}_*(\gamma) = \{\alpha, \beta, \gamma\}$
PROPOSITION. \( \omega \in B_\ast E \) if and only if \( B_\ast(\omega) \subseteq E \).

Let \( E = \{\beta, \gamma\} \) : the event that represents the proposition \( p \)

Then \( B_1E = \{\gamma\} \), \( B_2E = \{\beta, \gamma\} \), \( B_\ast E = \emptyset \)

In fact, while \( \gamma \in B_1B_2 E = \{\gamma\} \), \( \gamma \notin B_2B_1 E = \emptyset \)
Module 3
Models of games and Rationality

Definition. A finite strategic-form game with ordinal payoffs is a quintuple

\[ \langle N, \{ S_i \}_{i \in N}, O, \{ \succeq_i \}_{i \in N}, z \rangle \]

\( N = \{1, \ldots, n\} \) is a set of players

\( S_i \) is a finite set of strategies or choices of player \( i \in N \)

\( O \) is a set of outcomes

\( \succeq_i \) is player \( i \)'s ordering of \( O \) (\( o \succeq_i o' \) means that, for player \( i \),

outcome \( o \) is at least as good as outcome \( o' \))

\( z : S \rightarrow O \) (where \( S = S_1 \times \ldots \times S_n \)) associates an outcome with every

strategy profile \( s \in S \)
**Definition.** Given a strategic-form game with ordinal payoffs

\[ \langle N, \{S_i\}_{i \in N}, O, \{\succeq_i\}_{i \in N}, z \rangle \]

a reduced form of it is a triple

\[ \langle N, \{S_i\}_{i \in N}, \{u_i\}_{i \in N} \rangle \]

where \( u_i : S \to \mathbb{R} \) is such that \( u_i(s) \geq u_i(s') \) if and only if \( z(s) \succeq_i z(s') \)

player \( i \)'s utility function
**Definition.** An *epistemic model* of a strategic-form game is an interactive belief structure together with $n$ functions

$$\sigma_i : \Omega \rightarrow S_i \quad (i \in N)$$

**Interpretation:** $\sigma_i(\omega)$ is player $i$’s chosen strategy at state $\omega$

**Restriction:** if $\omega' \in B_i(\omega)$ then $\sigma_i(\omega') = \sigma_i(\omega)$

(no player has mistaken beliefs about her own strategy)
EXAMPLE

At every state each player knows his own strategy

1's strategy: A  C  C  D
2's strategy: f  f  g  g

At state β player 1 plays C (he knows this) not knowing whether player 2 is playing f or g and player 2 plays f (she knows this) not knowing whether player 1 is playing A or C
RATIONALITY

Non-probabilistic (no expected utility) and very weak notion of rationality

**Definition.** Player $i$ is *IRRATIONAL* at state $\omega$ if there is a strategy $s_i$ (of player $i$) which she believes to be better than $\sigma_i(\omega)$ (that is, if she believes that she can do better with another strategy)

Player $i$ is *RATIONAL* at state $\omega$ if and only if she is not irrational

---

Player 1 is rational at state $\beta$
Let $s_i$ and $t_i$ be two strategies of player $i$: $s_i, t_i \in S_i$

$s_i \succ_i t_i$ is interpreted as “strategy $s_i$ is better for player $i$ than strategy $t_i$”

$s_i \succ_i t_i$ is true at state $\omega$ if $u_i(s_i, \sigma_{-i}(\omega)) > u_i(t_i, \sigma_{-i}(\omega))$

that is, $s_i$ is better than $t_i$ against $\sigma_{-i}(\omega)$

---

**Table:**

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3,2</td>
<td>1,1</td>
<td>0,1</td>
</tr>
<tr>
<td>B</td>
<td>2,3</td>
<td>2,2</td>
<td>3,1</td>
</tr>
<tr>
<td>C</td>
<td>1,2</td>
<td>0,2</td>
<td>4,1</td>
</tr>
</tbody>
</table>

---

**Diagram:**

Player 1's strategy: A, C, C

Player 2's strategy: E, F, G

---

$A \succ_1 B$  $B \succ_1 A$  $C \succ_1 B$

$A \succ_1 C$  $B \succ_1 C$  $C \succ_1 A$

$B \succ_1 C$  $A \succ_1 C$  $B \succ_1 A$

$E \succ_2 F$  $F \succ_2 G$  $F \succ_2 G$  etc.
Let \( s_i \succ_i t_i = \{ \omega \in \Omega : u_i(s_i, \sigma_{-i}(\omega)) > u_i(t_i, \sigma_{-i}(\omega)) \} \) event that \( s_i \) is better than \( t_i \).

If \( s_i \in S_i \), let \( s_i = \{ \omega \in \Omega : \sigma_i(\omega) = s \} \) event that player \( i \) chooses \( s_i \).

Let \( R_i \) be the event representing the proposition “player \( i \) is rational”

\[
\|s_i\| \cap B_i \|t_i \succ_i s_i\| \subseteq \neg R_i
\]

\[
\neg R_i = \bigcup_{s_i \in S_i} \bigcup_{t_i \in S_i} (\|s_i\| \cap B_i \|t_i \succ_i s_i\|)
\]

\[
R = R_1 \cap \ldots \cap R_n \quad \text{all players are rational}
\]
At state $\alpha$ there is mutual knowledge of rationality but not common knowledge of rationality.
Let \( S_i = S_1 \times ... \times S_{i-1} \times S_{i+1} \times ... \times S_n \) set of strategy profiles of all players except \( i \).

**Definition.** Let \( s_i, t_i \in S_i \). We say that \( t_i \) is **strictly dominated** by \( s_i \) if \( u_i(t_i, s_{-i}) < u_i(s_i, s_{-i}) \) for all \( s_{-i} \in S_{-i} \).

**ITERATED DELETION OF STRICTLY DOMINATED STRATEGIES**

<table>
<thead>
<tr>
<th></th>
<th>e</th>
<th>f</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3, 2</td>
<td>3, 1</td>
<td>0, 1</td>
</tr>
<tr>
<td>B</td>
<td>2, 3</td>
<td>2, 2</td>
<td>3, 1</td>
</tr>
<tr>
<td>C</td>
<td>1, 2</td>
<td>1, 2</td>
<td>4, 1</td>
</tr>
<tr>
<td>D</td>
<td>0, 2</td>
<td>0, 3</td>
<td>1, 3</td>
</tr>
</tbody>
</table>

(by C)

<table>
<thead>
<tr>
<th></th>
<th>e</th>
<th>f</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3, 2</td>
<td>3, 1</td>
<td>0, 1</td>
</tr>
<tr>
<td>B</td>
<td>2, 3</td>
<td>2, 2</td>
<td>3, 1</td>
</tr>
<tr>
<td>C</td>
<td>1, 2</td>
<td>1, 2</td>
<td>4, 1</td>
</tr>
</tbody>
</table>

(by e)

<table>
<thead>
<tr>
<th></th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3, 2</td>
<td>3, 1</td>
</tr>
<tr>
<td>B</td>
<td>2, 3</td>
<td>2, 2</td>
</tr>
<tr>
<td>C</td>
<td>1, 2</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

(by B)
Let $G$ be a strategic-form game with ordinal payoffs and $G^\infty$ be the game obtained after applying the procedure of Iterated Deletion of Strictly Dominated Strategies.

Let $S^\infty$ denote the strategy profiles of game $G^\infty$.

Given a model of $G$, let $S^\infty$ denote the event $\{\omega \in \Omega : \sigma(\omega) \in S^\infty\}$
PROPOSITION 1. \( B^\ast R \subseteq S^\infty \)

If at a state it is commonly believed that all players are rational, then the strategy profile chosen at that state belongs to the game obtained after applying the iterated deletion of strictly dominated strategies.

At state \( \alpha \) there cannot be common knowledge of rationality since \( \sigma(\alpha) \neq (A, e) \)
Every normal operator $B$ satisfies the property that if $E \subseteq F$ then $BE \subseteq BF$.

$B_*$ is a normal operator. Thus from $B_*R \subseteq S^\infty$ it follows that $B_*B_*R \subseteq B_*S^\infty$.

By transitivity of $B_*$ we have that $B_*E \subseteq B_*B_*E$ for every event $E$.

Thus $B_*R \subseteq B_*B_*R$.

It follows that $B_*R \subseteq B_*S^\infty$.
REMARK. In general it is not true that $S^\infty \subseteq B^*_R$

\[ S^\infty = \{ \delta \} \]

\[ K^*_R = \emptyset \]

\[ R_1 = \{ \alpha, \delta \}, \quad R_2 = \{ \alpha, \beta, \gamma, \delta \} \]

\[ K_2 R_1 = \emptyset \]
**PROPOSITION 2.** Fix a strategic-form game with ordinal payoffs $G$ and let $s \in S^\infty$. Then there exists an epistemic model of $G$ and a state $\omega$ such that $\sigma(\omega) = s$ and $\omega \in B_\ast R$.

**EXAMPLE**

<table>
<thead>
<tr>
<th></th>
<th>Player 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>l</td>
<td>3, 2</td>
<td>2, 3</td>
</tr>
<tr>
<td>1</td>
<td>3, 3</td>
<td>4, 2</td>
</tr>
</tbody>
</table>

The game's matrix is shown above.

In this game every strategy profile survives iterative deletion.

In this model $R = B_\ast R = \Omega$ and every strategy profile occurs at some state.
REMARK. Given the above notion of rationality, there is no difference between common belief of rationality and common knowledge of rationality. The previous two propositions can be restated in terms of knowledge and common knowledge.

PROPOSITION 1'. $\mathcal{K} \ast \mathcal{R} \subseteq S^\infty$

PROPOSITION 2'. Fix a strategic-form game with ordinal payoffs $G$ and let $s \in S^\infty$. Then there exists an epistemic model of $G$ and a state $\omega$ such that $\sigma(\omega) = s$ and $\omega \in K \ast \mathcal{R}$. 

STRONGER NOTION OF RATIONALITY

Still non-probabilistic (no expected utility)

**Definition.** Player $i$ is **IRRATIONAL** at state $\omega$ if there is a strategy $s_i$ which she believes to be at least as good as $\sigma_i(\omega)$ and she considers it possible that $s_i$ is better than $\sigma_i(\omega)$

Player $i$ is **RATIONAL** at state $\omega$ if and only if she is not irrational

Player 1 is irrational at state $\beta$: $B$ is at least as good as $C$ at both $\beta$ and $\gamma$ and it is better than $C$ at $\gamma$

$$R_1 = \{\alpha\}, \quad R_2 = \emptyset$$
Player \( i \) is **IRRATIONAL** at state \( \omega \) if there is a strategy \( s_i \) which she believes to be at least as good as \( \sigma_i(\omega) \) and she considers it possible that \( s_i \) is better than \( \sigma_i(\omega) \)

\[
\left\| s_i \right\| \cap B_i \left\| t_i \geq i \right\| s_i \cap \neg B_i \neg \left\| t_i \succ i \right\| \subseteq \neg R_i
\]

\[
\neg R_i = \bigcup_{s_i \in S_i} \bigcup_{t_i \in S_i} \left( \left\| s_i \right\| \cap B_i \left\| t_i \geq i \right\| s_i \cap \neg B_i \neg \left\| t_i \succ i \right\| \right)
\]

\[
R = R_1 \cap \ldots \cap R_n \quad \text{all players are rational}
\]
Definition.

Given a game $G = \langle N, \{S_i\}_{i \in N}, O, \{\succeq\}_{i \in N}, z \rangle$, a subset of strategy profiles $X \subseteq S$ and a strategy profile $x \in X$, we say that $x$ is inferior relative to $X$ if there exist a player $i$ and a strategy $s_i \in S_i$ of player $i$ (thus $s_i$ need not belong to the projection of $X$ onto $S_i$) such that:

1. $z(s_i, x_{-i}) \succ_i z(x_i, x_{-i})$ and
2. for all $s_{-i} \in S_{-i}$, if $(x_i, s_{-i}) \in X$ then $z(s_i, s_{-i}) \succeq_i z(x_i, s_{-i})$.

Iterated Deletion of Inferior Profiles: for $m \in \mathbb{N}$ define $T^m \subseteq S$ recursively as follows: $T^0 = S$ and, for $m \geq 1$, $T^m = T^{m-1} \setminus I^{m-1}$, where $I^{m-1} \subseteq T^{m-1}$ is the set of strategy profiles that are inferior relative to $T^{m-1}$. Let $T^\infty = \bigcap_{m \in \mathbb{N}} T^m$. 
\( T^0 = S = \{(A,d),(A,e),(A,f),(B,d),(B,e),(B,f),(C,d),(C,e),(C,f)\}, I^0 = \{(B,e),(C,f)\} \) (the elimination of \((B,e)\) is done through player 2 and strategy \(f\), while the elimination of \((C,f)\) is done through player 1 and strategy \(B\));

\( T^1 = \{(A,d),(A,e),(A,f),(B,d),(B,f),(C,d),(C,e)\}, I^1 = \{(B,d),(B,f),(C,e)\} \) (the elimination of \((B,d)\) and \((B,f)\) is done through player 1 and strategy \(A\), while the elimination of \((C,e)\) is done through player 2 and strategy \(d\));

\( T^2 = \{(A,d),(A,e),(A,f),(C,d)\}, I^2 = \{(C,d)\} \) (the elimination of \((C,d)\) is done through player 1 and strategy \(A\));

\( T^3 = \{(A,d),(A,e),(A,f)\}, I^3 = \emptyset \); thus \( T^\infty = T^3 \).
PROPOSITION 3. $K \ast R \subseteq T^\infty$

If at a state it is commonly known that all players are rational, then the strategy profile chosen at that state belongs to the game obtained after applying the iterated deletion of Inferior strategy profiles.

PROPOSITION 4. Fix a strategic-form game with ordinal payoffs $G$ and let $s \in T^\infty$. Then there exists an epistemic model of $G$ and a state $\omega$ such that $\sigma(\omega) = s$ and $\omega \in K \ast R$. 
NOT TRUE if we replace common knowledge with common belief

\[
\begin{array}{c|cc}
\text{Player 1} & A & B \\
\hline
A & 1,1 & 1,0 \\
B & 1,1 & 0,1 \\
\end{array}
\]

\[
\begin{align*}
\mathcal{B}_1 : & \quad \alpha \rightarrow \beta \\
\mathcal{B}_2 : & \quad \alpha \quad \beta \\
\mathcal{B}_* : & \quad \alpha \rightarrow \beta \\
\sigma_1 : & \quad B \quad B \\
\sigma_2 : & \quad d \quad c \\
\end{align*}
\]

\[R_1 = \{\alpha, \beta\}, \quad R_2 = \{\alpha, \beta\}\]

There is common belief of rationality at every state and yet at state \(\alpha\) the strategy profile played is \((B,d)\) which is inferior

\[T^\infty = \{(A,c),(B,c)\}\]

\[S^\infty = \{(A,c),(A,d),(B,c),(B,d)\}\]
**PROBABILISTIC BELIEFS**

**Definition.** A *Bayesian frame* is an interactive belief frame together with a collection \( \{ p_{i,\omega} \}_{i \in N, \omega \in \Omega} \) of probability distributions on \( \Omega \) such that

1. if \( \omega' \in B_i(\omega) \) then \( p_{i,\omega'} = p_{i,\omega} \)
2. \( p_{i,\omega}(\omega') > 0 \) if and only if \( \omega' \in B_i(\omega) \)
   (the support of \( p_{i,\omega} \) coincides with \( B_i(\omega) \))

\[ B_1 : \begin{array}{c}
\bullet \quad 1/2 \quad 1/2 \bullet \\
\alpha \quad \beta \quad \gamma
\end{array} \]

\[ B_2 : \begin{array}{c}
\bullet \rightarrow \bullet \quad 1/3 \quad 2/3 \bullet
\end{array} \]
Definition. A strategic-form game \textit{with von Neumann-Morgenstern payoffs} is a quintuple

\[ \langle N, \{S_i\}_{i \in N}, O, \{U_i\}_{i \in N}, z \rangle \]

where

\[ N = \{1, \ldots, n\} \text{ is a set of players} \]

\[ S_i \text{ is the set of strategies of player } i \in N \]

\[ O \text{ is a set of outcomes} \]

\[ U_i : O \rightarrow \mathbb{R} \text{ is player } i \text{'s von Neumann-Morgenstern utility function} \]

\[ z : S \rightarrow O \text{ (where } S = S_1 \times \ldots \times S_n \text{) associates an outcome with every strategy profile } s \in S \]

Its reduced form is a triple \[ \langle N, \{S_i\}_{i \in N}, \{\pi_i\}_{i \in N} \rangle \text{ where } \pi_i(s) = U_i(z(s)). \]
An *epistemic model* of a strategic-form game is a Bayesian frame together with \( n \) functions

\[
\sigma_i : \Omega \rightarrow S_i \quad (i \in N)
\]

such that if \( \omega' \in B_i(\omega) \) then \( \sigma_i(\omega') = \sigma_i(\omega) \)

Stronger definition of Rationality than the previous ones

Player \( i \) is **RATIONAL** at state \( \alpha \) if her choice at \( \alpha \) maximizes her expected payoff, given her beliefs at \( \alpha \): for all \( t_i \in S_i \)

\[
\sum_{\omega \in B_i(\alpha)} \pi_i(\sigma_i(\alpha), \sigma_{-i}(\omega)) \ p_{i,\alpha}(\omega) \geq \sum_{\omega \in B_i(\alpha)} \pi_i(t_i, \sigma_{-i}(\omega)) \ p_{i,\alpha}(\omega)
\]
$R_1 = \{\delta, \varepsilon\}$

Player 1 is not rational at $\alpha$ because her expected payoff is $\frac{2}{3} \frac{1}{3} + \frac{1}{2} = \frac{4}{3}$

while if she had chosen strategy $A$ her payoff would have been $\frac{2}{3} \frac{1}{3} + \frac{1}{2} = 2$

On the other hand, Player 1 is rational at $\delta$ because her expected payoff is $\frac{1}{2} \frac{1}{2} + \frac{1}{2} 0 = \frac{3}{2}$

and if she had chosen strategy $B$ her payoff would have been $\frac{1}{2} \frac{1}{2} + \frac{1}{2} = \frac{3}{2}$

and if she had chosen strategy $C$ her payoff would have been $\frac{1}{2} \frac{1}{2} + \frac{1}{2} = \frac{3}{2}$
What are the implications of Common Belief of this stronger notion of rationality?

**Definition.** A mixed strategy of player $i$ is a probability distribution over $S_i$.

The set of mixed strategies of player $i$ is denoted by $\Delta(S_i)$.

Let $t_i \in S_i$ and $\nu_i \in \Delta(S_i)$. We say that $t_i$ is *strictly dominated* by $\nu_i$ if,

for every $s_{-i} \in S_{-i}$, $\pi_i(t_i, s_{-i}) < \sum_{s_i \in S_i} \nu_i(s_i) \pi_i(s_i, s_{-i})$

In this game strategy $B$ of player 1 is strictly dominated by the mixed strategy $\left( \begin{array}{cc} A & C \\ \frac{1}{6} & \frac{5}{6} \end{array} \right)$.
Iterative deletion of pure strategies that are strictly dominated by (possibly mixed) strategies

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>e</td>
</tr>
<tr>
<td>B</td>
<td>3,0</td>
</tr>
<tr>
<td>C</td>
<td>1,1</td>
</tr>
<tr>
<td>D</td>
<td>0,0</td>
</tr>
<tr>
<td></td>
<td>0,3</td>
</tr>
</tbody>
</table>

(a) The game $G$
B is strictly dominated by $(1/2 \text{ A}, 1/2 \text{ D})$

(b) The game $G^1$
Now f is strictly dominated by g

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>e</td>
</tr>
<tr>
<td>B</td>
<td>3,0</td>
</tr>
<tr>
<td>C</td>
<td>0,0</td>
</tr>
<tr>
<td>D</td>
<td>0,3</td>
</tr>
</tbody>
</table>

(c) The game $G^2$
Now C is strictly dominated by $(1/6 \text{ A}, 5/6 \text{ D})$

(d) The game $G^3 = G^\infty$
No strategy is strictly dominated
Let $G$ be a strategic-form game with von Neumann-Morgenstern payoffs and $G^\infty$ be the game obtained after applying the procedure of Iterated Deletion of Pure Strategies that are Strictly Dominated by Possibly Mixed Strategies.

Let $S^\infty_m$ denote the pure-strategy profiles of game $G^\infty$.

Given a model of $G$, let $S^\infty_m$ be the event $\{\omega \in \Omega : \sigma(\omega) \in S^\infty_m\}$

**PROPOSITION 5.** $B^*R \subseteq S^\infty_m$

**PROPOSITION 6.** Fix a strategic-form game with von Neumann-Morgenstern payoffs $G$ and let $s \in S^\infty_m$. Then there exists a Bayesian model of $G$ and a state $\omega$ such that $\sigma(\omega) = s$ and $\omega \in B_s R$. 

39
Given this stronger notion of rationality, there is a difference between common belief of rationality and common knowledge of rationality. The implications of common knowledge of rationality are stronger.

With knowledge, a player’s beliefs are always correct and are believed to be correct by every other player. Thus there is correctness and common belief of correctness of everybody’s beliefs.
**Definition.** Given a strategic-form game with von Neumann-Morgenstern payoffs $G$, a pure-strategy profile $x \in X \subseteq S$ is *inferior relative to* $X$ if there exists a player $i$ and a (possibly mixed) strategy $\nu_i$ of player $i$ (whose support can be any subset of $S_i$, not necessarily the projection of $X$ onto $S_i$) such that:

1. $\pi_i(x_i, x_{-i}) < \sum_{s_i \in S_i} \pi_i(s_i, x_{-i}) \nu_i(s_i)$ ($\nu_i$ yields a higher expected payoff than $x_i$ against $x_{-i}$)

2. for all $s_{-i} \in S_{-i}$ such that $(x_i, s_{-i}) \in X$, $\pi_i(x_i, s_{-i}) \leq \sum_{s_i \in S_i} \pi_i(s_i, s_{-i}) \nu_i(s_i)$

<table>
<thead>
<tr>
<th></th>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>Player A</td>
<td>2, 0</td>
<td>2, 2</td>
</tr>
<tr>
<td>1</td>
<td>2, 2</td>
<td>1, 2</td>
</tr>
<tr>
<td>C</td>
<td>2, 0</td>
<td>1, 0</td>
</tr>
</tbody>
</table>

Here $(C,F)$ is inferior relative to $S$ (for player 1, $B$ weakly dominates $C$ and is strictly better than $C$ against $F$) and $(A,D)$ is inferior relative to $S$ (for player 2, $E$ weakly dominates $D$ and is strictly better than $D$ against $A$)
(a) \( S_0^s = S, D_0^s = \{(A, D), (C, F)\} \)

(b) \( S_1^s = \{(A, E), (A, F), (B, D), (B, E), (B, F), (C, D), (C, E)\} \)
\( D_1^s = \{(C, E), (B, F)\} \)

(c) \( S_2^s = \{(A, E), (A, F), (B, D), (B, E), (C, D)\} \),
\( D_2^s = \{(B, E)\} \).

(d) \( S_3^s = S_\infty^s = \{(A, E), (A, F), (B, D), (C, D)\} \),
\( D_3^s = \emptyset \).
Let $G$ be a strategic-form game with von Neumann-Morgenstern payoffs and $G^\infty$ be the game obtained after applying the procedure of Iterated Deletion of Inferior Pure-Strategy Profiles.

Let $S^\infty_s$ denote the pure-strategy profiles of game $G^\infty$.

Given a model of $G$, let $S^\infty_s$ be the event $\{\omega \in \Omega : \sigma(\omega) \in S^\infty_s\}$

**PROPOSITION 7.** $K_\ast R \subseteq S^\infty_s$

**PROPOSITION 8.** Fix a strategic-form game with von Neumann-Morgenstern payoffs $G$ and let $s \in S^\infty_s$. Then there exists a Bayesian model of $G$ and a state $\omega$ such that $\sigma(\omega) = s$ and $\omega \in K_\ast R$. 
In this game $S^\infty = S^m = S$
while $S^s = \{(A, E), (A, F), (B, D), (C, D)\}$

Thus every strategy profile is compatible with *common belief* of rationality while only $(A, E), (A, F), (B, D)$ and $(C, D)$ are compatible with *common knowledge* of rationality
CREDITS

The link between the iterated deletion of strictly dominated strategies and the informal notion of common belief of rationality was first shown by Bernheim (1984) and Pearce (1984).

The first explicit epistemic characterization was provided by Tan and Werlang (1998) using a universal type space.

The state space formulation used in Propositions 5 and 6 is due to Stalnaker (1994), but it was implicit in Brandenburger and Dekel (1987).

Propositions 7 and 8 are due to Stalnaker (1994) (with a correction given in Bonanno and Nehring, 1996b).

To my knowledge, Propositions 1, 2, 3 and 4 have not been explicitly stated before.

References and further details can be found in


For a syntactic version of Propositions 1, 2, 3 and 4 see