ECON813 Game Theory
Part A: Interactive Reasoning and Choice
Topic 2 Common Belief in Rationality

Christian W. Bach

University of Liverpool & EPICENTER

UNIVERSITY OF LIVERPOOL

EPICENTER
Research Center for
Epistemic Game Theory

Interactive Reasoning

- Since the outcome in a game for a player does not only depend on his own decision, but also on what his opponents are doing, it is crucial to model his belief about his opponents' choices.
- Due to this intuition the notion of conjecture was presented in T1.
- However, a full account of interactive thinking actually require mores:
  - *what a player thinks his opponents are conjecturing,*
  - *what he thinks his opponents are thinking their respective opponents are conjecturing,*
  - *etc.*
- Accordingly, **interactive reasoning** encompasses an (infinite) sequence of iterated beliefs.

Belief hierarchies

- More precisely, in **Epistemic Game Theory**, every player *i* is assumed to entertain a belief hierarchy:

  - *a belief of i about his opponents' choice-combinations,*

    (conjecture; also called first-order belief)

  - *a belief of i about his opponents' beliefs about their respective opponents', choice-combinations,*

    (second-order belief)

  - *a belief of i about his opponents' beliefs about their respective opponents' beliefs about their respective opponents' choice-combinations,*

    (third-order belief)

  - *etc.*

Thinking about Rationality Interactively

- A choice is **rational**, if it is optimal for *some* conjecture (cf. T1).
- The idea of **rationality** can be infused into interactive thinking.
- More formally speaking, belief in **rationality** can be iterated throughout the entire belief hierarchy of a player.
- Actually, the **epistemic condition** of **common belief in rationality** does exactly so:
  - *player i believes his opponents to choose **rationally**,*
  - *player i believes his respective opponents to believe their respective opponents to choose **rationally**,*
  - *player i believes his respective opponents to believe their respective opponents to believe their respective opponents to choose **rationally**,*
  - *etc.*

Example: Going to a Party

Story:

- *Alice* and *Bob* are going together to a party tonight.
- *Alice* asks herself what colour she should wear.
- *Alice* prefers *blue* to *green*, *green* to *red*, and *red* to *yellow*.
- However, *Alice* dislikes most to wear the same colour as *Bob*.
- Let the utilities be given as follows:
  - *blue*: *Alice*: 4 and *Bob*: 2
  - *green*: *Alice*: 3 and *Bob*: 1
  - *red*: *Alice*: 2 and *Bob*: 4
  - *yellow*: *Alice*: 1 and *Bob*: 3
  - same colour: *Alice*: 0 and *Bob*: 0
- Question: Which colours can *Alice* **rationally** choose for tonight's party under **common belief in rationality**?

Example: Going to a Party

- Rational choices for *Alice*: blue, green, and red.
- Rational choices for *Bob*: red, yellow, and blue.
  - Red is optimal for *Bob*, if he believes *Alice* to choose any other colour than red.
  - Yellow is optimal for *Bob*, if he believes *Alice* to choose red.
  - Blue is optimal for *Bob*, if he believes with probability 0.6 that *Alice* chooses red and with probability 0.4 that *Alice* chooses yellow.
  - Green is never optimal: red is better for all beliefs with probability of less than 0.5 for *Alice* choosing red and yellow is better for all beliefs with probability of at least 0.5 for *Alice* choosing red.
- If *Alice* believes in *Bob*'s rationality, then she assigns probability 0 to *Bob*'s choice green.
- Thus, restrict *Alice*'s belief about *Bob*'s choice to red, yellow, and blue.
  - blue is optimal, if *Alice* believes *Bob* to choose red.
  - green is optimal, if *Alice* believes *Bob* to choose blue.
  - green yields higher expected utility than red, if *Alice* believes *Bob* to choose from $\{red, yellow, blue\}$.
- Consequently, *Alice* can only rationally choose blue and green, if she believes in *Bob*'s rationality.

Example: Going to a Party

- Rational choices for *Alice* if she believes in *Bob*'s rationality: blue, and green.
- Rational choices for *Bob* if he believes in *Alice*'s rationality: red, and yellow.
  - red is optimal, if *Bob* believes *Alice* to choose blue.
  - yellow is optimal, if *Bob* believes *Alice* to choose red.
  - yellow yields higher expected utility than blue, if *Bob* believes *Alice* to choose from $\{blue, green, red\}$.
- Can *Alice* **rationally** choose blue and green under **common belief in rationality**?

Example: Going to a Party

- Note that blue is optimal for *Alice*, if she believes *Bob* to choose red, **and** that red is optimal for *Bob*, if he believes *Alice* to choose blue.
- Consider the following belief hierarchy h_{Alice} for *Alice*.
    - *Alice* believes *Bob* to choose red.
    - *Alice* believes *Bob* to believe her to choose blue.
    - *Alice* believes *Bob* to believe her to believe that he chooses red.
    - *Alice* believes *Bob* to believe her to believe him to believe that she chooses blue.
    - etc.
- Thus, *Alice* believes *Bob* to choose rationally, and believes *Bob* to believe her to choose rationally, etc.
- In other words, h_{Alice} does not contain any belief of any order in which the rationality neither of *Alice* nor of *Bob* is questioned.
- Consequently, h_{Alice} satisfies common belief in rationality, and blue is optimal for her given the first-order belief of h_{Alice}.

Example: Going to a Party

- What about *Alice*'s second most preferred colour green?

- If *Alice* believes in *Bob*'s rationality, and believes that he believes in her rationality, then she assigns probability 0 to *Bob*'s choices blue and green.

- However, blue then yields higher expected utility than green for *Alice*, if she believes *Bob* to choose from $\{red, yellow\}$.

- In particular, *Alice* can hence not **rationally** choose green – but only blue– under **common belief in rationality**.

- Analogously, it can be shown that *Bob* can only **rationally** choose red under **common belief in rationality**.

Outline

- Epistemic Model
- Common Belief in Rationality
- Iterated Strict Dominance
- Characterization

EPISTEMIC MODEL

Rewriting Belief Hierarchies

- A belief hierarchy involves infinitely many layers.
  - FIRST-ORDER BELIEF: *i's belief about his opponents' choices.*
  - SECOND-ORDER BELIEF: *i's belief about his opponents' beliefs about their resepctive opponents' choices.*
  - THIRD-ORDER BELIEF: *i's belief about his opponents' beliefs about their respective opponents' beliefs about their respective opponents' choices.*
  - FOURTH-ORDER BELIEF: *i's belief about his opponents' beliefs about their respective opponents' beliefs about their respective opponents' beliefs about their respective opponents' choices.*
  - *etc.*
- The above doxastic sequence can be rewritten as follows:
  - FIRST-ORDER BELIEF: *i's belief about his opponents' choices.*
  - SECOND-ORDER BELIEF: *i's belief about his opponents'* FIRST-ORDER BELIEFS.
  - THIRD-ORDER BELIEF: *i's belief about his opponents'* SECOND-ORDER BELIEFS.
  - FOURTH-ORDER BELIEF: *i's belief about his opponents'* THIRD-ORDER BELIEFS.
  - *etc.*
- In a way, a belief hierarchy thus consists of a **first-order belief** and a **belief** about the opponents´ **belief hierarchies**.

Finite Representation of Belief Hierarchies

- This is a crucial insight that actually enables a **compact representation** of belief hierarchies.
- The infinite doxastic sequences constituting a belief hierarchy is labelled by the abstract notion of **type**.
- A **type** induces a belief about the opponents' choice-type combinations.
- Every layer of the belief hierarchy that corresponds to the **type** can then be inferred.
- Types can thus be viewed as **implicit** belief hierarchies.

Epistemic Model

- Types and their beliefs are modelled in an additional mathematical structure called

  epistemic model

  that complements the game structure given by Γ.

Definition 1

Let Γ be a game. An epistemic model $\mathcal{M}^{\Gamma} = (T_i, b_i)_{i \in I}$ of Γ provides for every player $i \in I$,

- a finite set T_i of types,
- and for every type $t_i \in T_i$ a probability measure

$$b_i(t_i) \in \Delta\left((C_j \times T_j)_{j \in I \setminus \{i\}}\right)$$

on the opponents' choice-type combinations.

- Note that the probability measure b_i – the belief function of player i – provides for every type $t_i \in T_i$ a first-order belief as well as a belief about the opponents' types, i.e. belief hierarchies.

Illustration: An Epistemic Model for the Example

- Type Sets:

$$T_{Alice} = \{t^1_{Alice}, t^2_{Alice}, t^3_{Alice}\}$$
$$T_{Bob} = \{t^1_{Bob}, t^2_{Bob}, t^3_{Bob}, t^4_{Bob}\}$$

- Beliefs for *Alice*:

$$b_{Alice}(t^1_{Alice}) = (green, t^1_{Bob})$$
$$b_{Alice}(t^2_{Alice}) = (blue, t^2_{Bob})$$
$$b_{Alice}(t^3_{Alice}) = 0.6 \cdot (blue, t^3_{Bob}) + 0.4 \cdot (green, t^4_{Bob})$$

- Beliefs for *Bob*:

$$b_{Bob}(t^1_{Bob}) = (blue, t^1_{Alice})$$
$$b_{Bob}(t^2_{Bob}) = (green, t^2_{Alice})$$
$$b_{Bob}(t^3_{Bob}) = (red, t^3_{Alice})$$
$$b_{Bob}(t^4_{Bob}) = (yellow, t^1_{Alice})$$

Illustration: An Epistemic Model for the Example

- Type Sets:
  $T_{Alice} = \{t^1_{Alice}, t^2_{Alice}, t^3_{Alice}\}$
  $T_{Bob} = \{t^1_{Bob}, t^2_{Bob}, t^3_{Bob}, t^4_{Bob}\}$
- Beliefs for *Alice*:
  $b_{Alice}(t^1_{Alice}) = (green, t^1_{Bob})$
  $b_{Alice}(t^2_{Alice}) = (blue, t^2_{Bob})$
  $b_{Alice}(t^3_{Alice}) = 0.6 \cdot (blue, t^3_{Bob}) + 0.4 \cdot (green, t^4_{Bob})$
- Beliefs for *Bob*:
  $b_{Bob}(t^1_{Bob}) = (blue, t^1_{Alice})$
  $b_{Bob}(t^2_{Bob}) = (green, t^3_{Alice})$
  $b_{Bob}(t^3_{Bob}) = (red, t^2_{Alice})$
  $b_{Bob}(t^4_{Bob}) = (yellow, t^1_{Alice})$

Type t^3_{Alice} induces the following belief hierarchy:

- *Alice* believes with probability-0.6 *Bob* to wear blue and with probability-0.4 *Bob* to wear green. (**first-order belief**)
- *Alice* believes with probability-0.6 *Bob* to believe her to wear red and with probability-0.4 *Bob* to believe her to wear yellow. (**second-order belief**)
- *Alice* believes with probability-0.6 *Bob* to believe her to believe him to wear blue and with probability-0.4 *Bob* to believe her to believe him to wear green. (**third-order belief**)
- etc.

Optimality Defined for Types

Definition 2

Let Γ be a game, $\mathcal{M}^\Gamma$ an epistemic model of it, $i \in I$ some player, $c_i \in C_i$ some choice of player i, and $t_i \in T_i$ some type of player i. The choice c_i is optimal for t_i, if c_i is optimal given t_i's induced conjecture.

Note: to check whether some choice is optimal for a given type,

only the first-order belief

needs to be considered – not its higher-order beliefs.

Epistemic Models and Rationality

Definition 3

Let Γ be a game, $i \in I$ some player, and $c_i \in C_i$ some choice of player i. The choice c_i is rational, if there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ with a type $t_i \in T_i$ of player i such that c_i is optimal for t_i.

Common Belief in Rationality

Iterating Belief in Rationality

- Intuitively, a choice is rational, if it is optimal for some conjecture.
- A player can then be said to believe in rationality, if he only assigns positive probability to choices & conjectures of his opponents such that the choices are optimal for the conjectures.
- Correspondingly, a player believes his opponents to believe in rationality, if he only assigns positive probability to beliefs of his opponents that believe in rationality, etc.
- In this fashion, a restriction is imposed on every layer of a player´s belief hierarchy, and this gives rise to the **epistemic condition** of **common belief in rationality**.
- Intuititively, a player expressing **common belief in rationality** thus exhibits a state of mind, where
  - he believes in rationality,
  - he believes his opponents to believe in rationality,
  - he believes his opponents to believe that their respective opponents believe in rationality,
  - etc.
- These ideas are now formalized in **epistemic models**.

Belief in Rationality

Definition 4

Let Γ be a game, $\mathcal{M}^{\Gamma}$ an epistemic model of it, $i \in I$ some player, and $t_i \in T_i$ some type of player i. The type t_i believes in rationality, if t_i only assigns positive probability to choice-type combinations

$$\big((c_1, t_1), \ldots, (c_{i-1}, t_{i-1}), (c_{i+1}, t_{i+1}), \ldots (c_n, t_n)\big)$$

such that c_j is optimal for t_j for all $j \in I \setminus \{i\}$.

Higher-order Beliefs in Rationality

Definition 5

Let Γ be a game, $\mathcal{M}^\Gamma$ an epistemic model of it, $i \in I$ some player, and $t_i \in T_i$ some type of player i.

- The type t_i expresses *1-fold belief in rationality*, if t_i believes in rationality.
- Let $k > 1$. The 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.

Let $l \geq 1$. The type t_i expresses up to *l*-fold belief in rationality, if t_i expresses *k*-fold belief in rationality for all $k \leq l$.

Common Belief in Rationality

Definition 6

Let Γ be a game, $\mathcal{M}^{\Gamma}$ an epistemic model of it, $i \in I$ some player, and $t_i \in T_i$ some type of player i. The type t_i expresses common belief in rationality, if t_i expresses *k*-fold belief in rationality for all $k \geq 1$.

Rational Choice under Common Belief in Rationality

Definition 7

Let Γ be a game, $i \in I$ some player, and $c_i \in C_i$ some choice of player i. The choice c_i is rational under common belief in rationality, if there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ with some type $t_i \in T_i$ of player i such that

- t_i expresses common belief in rationality,
- c_i is optimal for t_i.

Illustration: An Epistemic Model for the Example

- Consider the following **epistemic model** of *Example I*.
  - Type Sets:

$$T_{Alice} = \{t_{Alice}\}$$
$$T_{Bob} = \{t_{Bob}\}$$

  - Beliefs for *Alice*:

$$b_{Alice}(t_{Alice}) = (red, t_{Bob})$$

  - Beliefs for *Bob*:

$$b_{Bob}(t_{Bob}) = (blue, t_{Alice})$$

- Observe that t_{Alice} expresses **common belief in rationality**.
  - *Alice* believes that *Bob* is of type t_{Bob} and chooses red, which is optimal for t_{Bob}.
    (*1*-fold belief in rationality)
  - *Alice* believes that *Bob* believes her to be of type t_{Alice} and to choose blue, which is optimal for t_{Alice}.
    (*2*-fold belief in rationality)
  - *Alice* believes that *Bob* believes her to believe him to be of type t_{Bob} and to choose red which is optimal for t_{Bob}.
    (*3*-fold belief in rationality)
  - etc.

Shortcut to Verifying Common Belief in Rationality

Theorem 8

Let Γ be a game and $\mathcal{M}^{\Gamma}$ an epistemic model of it. If all types express belief in rationality, then all types express common belief in rationality.

Proof:
(by INDUCTION on belief order k)

- INDUCTION BASE:
  - It directly holds that every type in $\mathcal{M}^{\Gamma}$ expresses 1-fold belief in rationality.
- INDUCTION STEP:
  - Suppose that every type expresses k^*-fold belief in rationality for some $k^* > 1$.
  - Every type thus only assigns positive probability to opponents' types that express k^*-fold belief in rationality, and consequently expresses $(k^* + 1)$-fold belief in rationality.
  - By induction, it then follows that every type expresses k-fold belief in rationality for all $k \in \mathbb{N}$.
- Therefore, every type expresses common belief in rationality.

ITERATED STRICT DOMINANCE

Iterating Strict Dominance Arguments

- Formally, a solution concept (*SC*) in **classical game theory** is a set of choice profiles, i.e. $SC \subseteq \times_{i \in I} C_i$.

- The solution concept of **iterated strict dominance** repeatedly applies strict dominance to the game:

  - **Step 1:** *within the original game, eliminate all choices that are strictly dominated.*

  - **Step 2:** *within the reduced game obtained after* **Step 1**, *eliminate all choices that are strictly dominated.*

  - **Step 3:** *within the reduced game obtained after* **Step 2**, *eliminate all choices that are strictly dominated.*

  - *etc.*

- The solution of the game then consists of all choice profiles that can be formed by the surviving choices.

Iterated Strict Dominance

Definition 9

Let Γ be a game.

- $SD^0 := \times_{i \in I} C_i$.
- $SD^{(n+1)} := \times_{i \in I} SD_i^{(n+1)}$, where

  $SD_i^{(n+1)} := SD_i^n \backslash$

  $\{c_i \in SD_i^n : \exists r_i \in \Delta(SD_i^n)$ s.t. $U_i(c_i, c_{-i}) < V_i(r_i, c_{-i}) \forall c_{-i} \in SD_{-i}^n\}$

  for all $i \in I$ and for all $n \geq 0$.

The set $SD^k = \times_{i \in I} SD_i^{(k+1)}$ is called *k-fold strict dominance* for all $k > 0$, and the set $ISD := \cap_{k \geq 0} SD^k$ is called *iterated strict dominance*.

Illustration: ISD in the Example

- **Step 1:** $C_{Alice} = \{blue, green, red, yellow\}$ and $C_{Bob} = \{red, yellow, blue, green\}$.
  - *Alice*: yellow is strictly dominated by 0.5blue $+ 0.5$green.
  - *Bob*: green is strictly dominated by 0.5red $+ 0.5$yellow.
- **Step 2:** $SD^1_{Alice} = \{blue, green, red\}$ and $SD^1_{Bob} = \{red, yellow, blue\}$.
  - *Alice*: red is strictly dominated by green.
  - *Bob*: blue is strictly dominated by yellow.
- **Step 3:** $SD^2_{Alice} = \{blue, green\}$ and $SD^2_{Bob} = \{red, yellow\}$.
  - *Alice*: green is strictly dominated by blue.
  - *Bob*: yellow is strictly dominated by red.
- **Iterated strict dominance** yields blue for *Alice* and red for *Bob*. (Formally, $ISD = \{(blue, red)\}$.)

Intelligibility

Theorem 10

Let Γ be a game.

$$ISD \neq \emptyset.$$

Proof:

- Towards a contradiction, suppose that $ISD = \emptyset$.
- Then, there exists a “final round” k^* such that $SD^{k^*} = \emptyset$, and thus $SD_i^{k^*} = \emptyset$ for some $i \in I$.
- Hence, every choice $c_i \in SD_i^{(k^*-1)}$ is strictly dominated by some mixed choice $r_i \in \Delta(SD_i^{(k^*-1)})$. i.e. $U_i(c_i, c_{-i}) < \sum_{c_i' \in \text{supp } (r_i)} r_i(c_i') \cdot U_i(c_i', c_{-i})$ for all $c_{-i} \in SD_{-i}^{(k^*-1)}$.
- Consider some $c_{-i} \in SD_{-i}^{(k^*-1)}$ and observe that, by Lemma 13 from T1, for every choice $c_i \in SD_i^{(k^*-1)}$ there exists some choice $c_i^* \in SD_i^{(k^*-1)}$ such that $U_i(c_i, c_{-i}) < U_i(c_i^*, c_{-i})$.
- Due to the finiteness of Γ there are only finitely many choices in $SD_i^{(k*-1)}$, which then implies that there must be some choice $\hat{c}_i \in SD_i^{(k^*-1)}$ such that $U_i(c_i, c_{-i}) \leq U_i(\hat{c}_i, c_{-i})$ holds for all $c_i \in SD_i^{(k^*-1)}$.
- However, $\hat{c}_i$ is then not strictly dominated in the reduced game $SD^{(k^*-1)}$, a contradiction.

$\lightning$

Effectiveness

Theorem 11

Let Γ be a game. There exists $k \in \mathbb{N}$ such that

$$SD^n = SD^k$$

for all $n > k$.

Proof:

- Towards a contradiction, suppose that $SD^n \neq SD^k$ for all $n > k$.
- Then, $SD^n \subsetneq SD^k$ for all $n > k$.
- However, since C_i is finite and at least one choice is deleted in every round, after maximally $(| C_i | -1)$ rounds no more strict dominance arguments can be formed.
- Therefore, $SD_i^n = SD_i^{(|C_i|-1)}$ for all $n > (| C_i | -1)$, a contradiction.

$$\lightning$$

Monotonicity

Theorem 12

Let Γ be a game, $i \in I$ some player, and $c_i \in C_i$ some choice of player i. If $c_i \in SD_i^k$ for some $k \geq 0$, then c_i is strictly dominated against $SD_{-i}^{k'}$ for all $k' > k$.

Proof:

- Suppose that $c_i \in SD_i^k$ for some $k \geq 0$
- Then, there exists some $r_i \in \Delta(SD_i^k)$ such that $U_i(c_i, c_{-i}) < V_i(r_i, c_{-i})$ for all $c_{-i} \in SD_{-i}^k$.
- Consider some $k' > k$.
- As $SD_{-i}^{k'} \subseteq SD_{-i}^k$, the inequality $U_i(c_i, c_{-i}) < V_i(r_i, c_{-i})$ also holds for all $c_{-i} \in SD_{-i}^{k'}$.
- Therefore, c_i is strictly dominated against $SD_{-i}^{k'}$.

Conceptual Upshots of the Three Properties

- INTELLIGIBILITY: **ISD** always returns a non-empty output and can thus be applied to any game.
- EFFECTIVENESS: **ISD** always stops after fintely many rounds and thus constitutes a finite procedure.
- MONOTONICTY: a choice identified by **ISD** as strictly dominated in some round remains strictly dominated in all succeeding rounds, and **ISD** can thus be viewed as order-independent.

CHARACTERIZATION

Motivation

- The epistemic and the classical perspectives are now related to each other.
- In the Example reasoning in line with common belief in rationality and the solution concept of ISD both lead to the same result.
- As it turns out this is not a coincidence, as common belief in rationality and ISD are equivalent.

Epistemic Characterization of Iterated Strict Dominance

Theorem 13

Let Γ be a game, $i \in I$ some player, and $c_i \in C_i$ some choice of player i. The choice c_i is rational under common belief in rationality, if and only if, $c_i \in ISD_i$.

- The epistemic characterization of ISD consists of two directions.
- **Epistemic Foundation:** *CBR* implies *ISD*.
- **Existence:** ISD can be supported by *CBR*.

Proof for: *Only If* Direction (Epistemic Foundation)

Lemma 14

Let Γ be a game, $i \in I$ some player, $c_i \in C_i$ some choice of player i, and $k \in \mathbb{N}$. If the choice c_i is rational under up to k-fold belief in rationality, then $c_i \in SD_i^{(k+1)}$.

- **Induction Base:**
  - Let $c_i \in C_i$ be a choice of some player $i \in I$ that is rational under up to *1*-fold belief in rationality.
  - Then, there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ with some type $t_i \in T_i$ of player i such that t_i believes in rationality and c_i is optimal for t_i.
  - Consequently, supp$(b_i(t_i))$ only contains choice type pairs (c_j, t_j) for every opponent $j \in I \setminus \{i\}$ such that c_j is optimal for t_j.
  - By PEARCE´S LEMMA it follows that supp$(b_i(t_i)) \subseteq SD^1_{-i}$.
  - As c_i is optimal for t_i, it cannot be – again via PEARCE´S LEMMA – strictly dominated against SD^1_{-i}.
  - Hence, $c_i \in SD_i^2$.

Proof for: *Only If* Direction (Epistemic Foundation)

- **Induction Step:**
  - Let $c_i \in C_i$ be a choice of some player $i \in I$ that is rational under up to k-fold belief in rationality.
  - Then, there exists an epistemic model $\mathcal{M}^{\Gamma}$ of Γ with some type $t_i \in T_i$ of player i such that t_i express up to k-fold belief in rationality and c_i is optimal for t_i.
  - Consequently, $\text{supp}(b_i(t_i))$ only contains choice type pairs (c_j, t_j) for every opponent $j \in I \setminus \{i\}$ such that t_j expresses up to *(k-1)*-fold belief in rationality and c_j is optimal for t_j.
  - Thus, for all $j \in I \setminus \{i\}$ the choice c_j is rational under up to *(k-1)*-fold belief in rationality, and the induction hypothesis then ensures that $c_{-j} \in SD^k_{-i}$ for all $c_{-j} \in \text{supp}(b_i(t_i))$.
  - Since c_i is optimal for t_i, it cannot be – by PEARCE´S LEMMA – strictly dominated against SD^k_{-i}.
  - Hence, $c_i \in SD_i^{k+1}$.
  - This establishes LEMMA 14.

Proof for: *Only If* Direction (Epistemic Foundation)

- Now suppose that c_i is rational under common belief in rationality.
- Then, there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ with some type $t_i \in T_i$ of player i such that t_i express common belief in rationality and c_i is optimal for t_i.
- Thus, t_i expresses up to *k*-fold belief in rationality for all $k \geq 1$.
- By Lemma 14, it follows that $c_i \in SD_i^{(k+1)}$ for all $k > 1$.
- Therefore, $c_i \in \cap_{k \geq 1} SD_i^k = ISD_i$, which establishes the *Only If* direction of Theorem 13.

Proof for: *If* Direction (Existence)

- By Theorem 12 there exists $k \in \mathbb{N}$ such that $SD^n = SD^k$ for all $n > k$, and thus $ISD = SD^k$.
- Consider the reduced game $\Gamma' = (I, (SD^k_j, U_j|_{SD^k})_{j \in I})$, where $U_j|_{SD^k}$ denotes the restriciton of U_j to SD^k for all $j \in I$.
- Since for all $j \in I$ every choice $c^k_j \in SD^k_j$ is not strictly dominated against SD^k_{-j}, it follows by PEARCE´S LEMMA applied to Γ' that every c^k_j is optimal for some conjecture $\beta^{c^k_j}_j \in \Delta(SD^k_{-j})$.
- Define an epistemic model $\mathcal{M}^{\Gamma} = (T_j, b_j)_{j \in I}$ of Γ, where

$$T_j := \{t^{c^k_j}_j : c^k_j \in SD^k_j\}$$

for all $j \in I$ as well as $b_j : T_j \to \Delta(C_{-j} \times T_{-j})$ such that

$$b_j(t^{c^k_j}_j)(c_{-j}, t_{-j}) := \begin{cases} \beta^{c^k_i}_i & \text{if } c_{-j} \in SD^k_{-j} \text{ and } t_l = t^{c_l}_l \text{ for all } l \in I \setminus \{j\}, \\ 0 & \text{otherwise} \end{cases}$$

for all $t^{c^k_j}_j \in T_j$ and for all $j \in I$.

Proof for: *If* Direction (Existence)

- By construction every type in $\mathcal{M}^\Gamma$ only deems possible opponent choice type pairs where the choice is optimal for the type.
- Thus, every type in $\mathcal{M}^\Gamma$ believes in rationality.
- By Theorem 8 it then follows that every type in $\mathcal{M}^\Gamma$ expresses common belief in rationality.
- Now consider player i and suppose that $c_i \in ISD_i$.
- Consequently there exists a type $t_i^{c_i}$ in $\mathcal{M}^\Gamma$ such that c_i is optimal for $t_i^{c_i}$ and $t_i^{c_i}$ expresses common belief in rationality.
- Therefore, c_i is rational under common belief in raitonality, which establishes the *If* direction of Theorem 13.

Intelligibility

Corollary 15

Let Γ be a game. There exists an epistemic model $\mathcal{M}^{\Gamma}$ of Γ in which all types express common belief in rationality.

- **Proof:** *Theorem 10 ensures that $ISD \neq \emptyset$ and the proof of the* If *direction of Theorem 13 then affirms there to be an epistemic model in which all types express common belief in rationality.*
- According to Corollary 15 it is always possible to reason in line with common belief in rationality in any game.
- The applicability of common belief in rationality does thus not depend on any particularities of the underlying game.
- **Intelligibility** thus takes shape classically (Theorem 10) as well as epistemically (Theorem 15).

Background Reading

- PEREA, A. (2012): *Epistemic Game Theory: Reasoning and Choice*. Cambridge University Press. **Chapter 3.**