Lexicographic Beliefs

Part I: Primary Belief in Rationality

Christian W. Bach

EPICENTER & University of Liverpool

Introduction

- Thus far, a player's belief about his opponents' choices has been modelled by a probability distribution.
- Ways of reasoning have been described in which some choices are completely discarded by receiving probability 0.
- Now, cautious reasoning is considered: some choices can be deemed much more likely than others, while at the same time no choice is completely discarded.
- Tool: lexicographic beliefs

Agenda

- Lexicographic Beliefs
- Lexicographic Epistemic Models
- Common Full Belief in (Caution & Primary Belief in Rationality)
- Existence
- Algorithm

Agenda

- **Lexicographic Beliefs**
- Lexicographic Epistemic Models
- Common Full Belief in (Caution & Primary Belief in Rationality)
- Existence
- Algorithm

Example: Should I call or not?

Story

- Tonight *Barbara* will go to the cinema.
- *You* can join if you wish, but *Barbara* decides on the movie.
- There is the choice between *The Godfather* and *Casablanca.*
- *You* prefer *The Godfather* (utility 1) to *Casablanca* (utility 0).
- *Barbara*'s movie preferences are inverse to yours.
- Staying at home yields *you* utility 0.
- *Barbara* goes to the cinema in any case.
- **Question:** Should *you* call *Barbara* or not?

Example: Should I call or not?

		Barbara: *Godfather*	Barbara: *Casablanca*
You	*call*	1, 0	0, 1
You	*not call*	0, 0	0, 1

Example: Should I call or not?

			Barbara	
			Godfather	*Casablanca*
You		*call*	1, 0	0, 1
		not call	0, 0	0, 1

- Intuitively, the **unique best choice** for *you* is *call*!
  - **standard beliefs**
    - However, if *you* believe in *Barbara*'s rationality with **standard beliefs**, then *you* must assign **probability 0** to her choice *Godfather*.
    - Consequently, both of your choices would be optimal for *you*.
  - **lexicographic beliefs**
    - A state of mind can be modelled in which *you* deem *Barbara* choosing *Casablanca* **infinitely more likely** than her picking *Godfather*.
    - Yet, the possibility of *Barbara* choosing *Godfather* is not completely discarded.

Example: Should I call or not?

		Barbara	
		Godfather	*Casablanca*
You	*call*	1, 0	0, 1
	not call	0, 0	0, 1

- Suppose *you* hold the following **lexicographic belief** on *Barbara's* choice:
  - **primary belief**: *you* believe *Barbara* to choose *Casablanca*.
  - **secondary belief**: *you* believe *Barbara* to choose *Godfather*.
- *You* then deem the event that *Barbara* chooses *Casablanca* **infinitely more likely** than the event that she picks *Godfather*.
  - Yet, given this lexicographic belief, the **unique optimal choice** for *you* is then *call*!

Lexicographic Beliefs

Definition

A ***lexicographic belief*** on some set *S* is a finite sequence

$$b^{lex} = (b^1, b^2, \ldots, b^k)$$

of distinct probability measures on S, where

- b^1 is called *level-1 belief*,
- b^2 is called *level-2 belief*,
- $\ldots$
- b^k is called *level-k belief*.

Remark.
Some authors require the probability measures in b^{lex} to have disjoint supports.

Intuition

- An event can be deemed infinitely more likely than another event, without completely discarding the latter!
- **Example:** lexicographic beliefs about the solar system
  - primary belief: the earth rotates around the sun
  - secondary belief: the sun rotates around the earth
  - tertiary belief: the sun and the earth both rotate around a hidden star
- A player i is said to deem an opponent j's choice c_j infinitely more likely than some choice c_j' for j, if c_j receives positive probability at an earlier lexicographic level than c_j' under his lexicographic belief b_i^{lex}.

Example: Where to read my book?

Story

- *You* would like to go to a pub to read your book.
- *Barbara* is going to a pub as well, but *you* forgot to ask her to which one.
- Your only objective is to avoid *Barbara*, since *you* would like to read your book in silence.
- *Barbara* prefers *Pub A* to *Pub B*, and *Pub B* to *Pub C*.
- **Question:** Which pub should *you* go to?

Example: Where to read my book?

		Barbara		
		Pub A	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

Example: Where to read my book?

		Barbara		
		Pub A	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- Intuitively, the **unique best choice** for *you* is *Pub C*, since it is the least preferred pub for *Barbara*!

- However, if *you* believe in *Barbara's* rationality with **standard beliefs**, then *you* must assign **probability 0** to her choosing *Pub B* and *Pub C*.
- Consequently, both *Pub B* and *Pub C* are optimal for *you*.
- Indeed, with **standard beliefs** *you* cannot believe in *Barbara's* rationality, while at the same time deeming her choice *Pub C* less likely than *Pub B*.

Example: Where to read my book?

		Barbara		
		Pub A	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- **Scenario 1:** Consider the **lexicographic belief** $(Pub\ A; Pub\ B; Pub\ C)$ for *you* about *Barbara's* choice
  - **primary belief**: *you* believe *Barbara* to choose *Pub A*.
  - **secondary belief**: *you* believe *Barbara* to choose *Pub B*.
  - **tertiary belief**: *you* believe *Barbara* to choose *Pub C*.
  - **Interpretation:** *you* deem *Barbara's* choice *Pub A* infinitely more likely than *Pub B* and *Pub B* infinitely more likely than *Pub C*, yet *you* consider all her choices possible.
- Given this lexicographic belief, the **unique optimal choice** for *you* is *Pub C!*

Example: Where to read my book?

		Barbara		
		Pub A	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- **Scenario 2:** Consider the **lexicographic belief** $(Pub\ A; Pub\ C; Pub\ B)$ for *you* about *Barbara's* choice
  - **primary belief**: *you* believe *Barbara* to choose *Pub A*.
  - **secondary belief**: *you* believe *Barbara* to choose *Pub C*.
  - **tertiary belief**: *you* believe *Barbara* to choose *Pub B*.
- Given this lexicographic belief, the **unique optimal choice** for *you* is *Pub B*!

Example: Where to read my book?

		Barbara		
		Pub A	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- **Scenario 3:** Consider the **lexicographic belief** $(Pub\ A; \frac{1}{3}Pub\ B + \frac{2}{3}Pub\ C)$ for for *you* about *Barbara's* choice
  - **primary belief**: *you* believe *Barbara* to choose *Pub A*.
  - **secondary belief**: *you* believe with probability $\frac{1}{3}$ *Barbara* to choose *Pub B* and with probability $\frac{2}{3}$ her to choose *Pub C*.
- Given this lexicographic belief, the **unique optimal choice** for *you* is *Pub B*!

Expected Utility under Lexicographic Beliefs

- Let $\Gamma = \big(I, (C_i)_{i \in I}, (U_i)_{i \in I}\big)$ be a game with two players.

- Suppose that player i entertains a lexicographic belief $b_i^{lex} = (b_i^1, b_i^2, \ldots, b_i^K)$ about j's choice.

- For every level $k \in \{1, 2, \ldots, K\}$ and for every choice $c_i \in C_i$ the k-level expected utility for player i of picking c_i is given by

$$u_i^k(c_i, b_i^{lex}) = \sum_{c_j \in C_j} \big(b_i^k(c_j) \cdot U_i(c_i, c_j)\big)$$

- Hence, every choice $c_i \in C_i$ for player i induces a sequence of expected utilities: lexicographic expected utility

$$u_i^{lex}(c_i, b_i^{lex}) = \big(u_i^1(c_i, b_i^{lex}), u_i^2(c_i, b_i^{lex}), \ldots, u_i^K(c_i, b_i^{lex})\big)$$

Preferences Induced by Lexicographic Beliefs

Definition

A player i with lexicographic belief b_i^{lex} ***prefers*** some choice c_i to c_i', if there exists some lexicographic level k such that

1. $u_i^k(c_i, b_i^{lex}) > u_i^k(c_i', b_i^{lex})$ and
2. $u_i^l(c_i, b_i^{lex}) = u_i^l(c_i', b_i^{lex})$ for all lexicographic levels $l < k$.

Useful Fact: Note that the binary relation ***prefer*** is transitive on the respective agent's choice set!

Definition

Given a lexicographic belief b_i^{lex} a choice c_i is called ***optimal***, if there exists no choice $c_i^* \in C_i$ such that i prefers c_i^* to c_i.

Rationality under Lexicographic Beliefs

Definition

A choice c_i is called ***rational***, if there exists some lexicographic belief b_i^{lex} such that c_i is optimal.

Example: Where to read my book?

		Barbara		
		Pub A	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- Consider **lexicographic belief** $b_{you}^{lex} = (Pub\ A; Pub\ B; Pub\ C)$
  - under the **primary belief**:
    $u_{you}^{1}(Pub\ A, b_{you}^{lex}) = 0,\ u_{you}^{1}(Pub\ B, b_{you}^{lex}) = 1,\ u_{you}^{1}(Pub\ C, b_{you}^{lex}) = 1$
  - under the **secondary belief**:
    $u_{you}^{2}(Pub\ B, b_{you}^{lex}) = 0,\ u_{you}^{2}(Pub\ C, b_{you}^{lex}) = 1$
- Hence, *you* prefer *Pub C* to *Pub B*, and *Pub B* to *Pub A*.
- Given b_{you}^{lex} the **unique optimal choice** is *Pub C* for *you*!

Example: Where to read my book?

		Barbara		
		Pub A	*Pub B*	*Pub C*
	Pub A	0, 3	1, 2	1, 1
You	*Pub B*	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- Consider **lexicographic belief** $b_{you}^{lex\,\prime} = (Pub\ A; \frac{1}{3} Pub\ B + \frac{2}{3} Pub\ C)$
  - under the **primary belief**: $u_{you}^1(Pub\ A, b_{you}^{lex\,\prime}) = 0,\ u_{you}^1(Pub\ B, b_{you}^{lex\,\prime}) = 1,\ u_{you}^1(Pub\ C, b_{you}^{lex\,\prime}) = 1$
  - under the **secondary belief**: $u_{you}^2(Pub\ B, b_{you}^{lex\,\prime}) = \frac{2}{3},\ u_{you}^2(Pub\ C, b_{you}^{lex\,\prime}) = \frac{1}{3}$
- Hence, *you* prefer *Pub B* to *Pub C*, and *Pub C* to *Pub A*.
- Given $b_{you}^{lex\,\prime}$, the **unique optimal choice** is *Pub B* for *you*!

Example: Where to read my book?

		Barbara: *Pub A*	Barbara: *Pub B*	Barbara: *Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
You	*Pub B*	1, 3	0, 2	1, 1
You	*Pub C*	1, 3	1, 2	0, 1

- Consider **lexicographic belief** $b_{you}^{lex\,''} = (\frac{1}{2} Pub\ A + \frac{1}{2} Pub\ B; \frac{1}{3} Pub\ B + \frac{2}{3} Pub\ C)$
  - under the **primary belief**: $u_{you}^1(Pub\ A, b_{you}^{lex\,''}) = \frac{1}{2},\ u_{you}^1(Pub\ B, b_{you}^{lex\,''}) = \frac{1}{2},\ u_{you}^1(Pub\ C, b_{you}^{lex\,''}) = 1$
  - under the **secondary belief**: $u_{you}^2(Pub\ A, b_{you}^{lex\,''}) = 1,\ u_{you}^2(Pub\ B, b_{you}^{lex\,''}) = \frac{2}{3}$
- Hence, *you* prefer *Pub C* to *Pub A*, and *Pub A* to *Pub B*.
- Given $b_{you}^{lex\,''}$, the **unique optimal choice** is *Pub C* for *you*!

Agenda

- Lexicographic Beliefs
- **Lexicographic Epistemic Model**
- Common Full Belief in (Caution & Primary Belief in Rationality)
- Algorithm
- Algorithm

Reasoning with Lexicographic Beliefs

- When reasoning about his opponents a player does not only entertain a belief about his opponents' choices but also about their beliefs, their beliefs about their opponents' beliefs, etc., i.e. a full belief hierarchy.

- A full belief hierarchy with standard beliefs is modelled by types in an epistemic model: a type induces a standard belief about his opponents' choice-type combinations.

- Analogously, a full belief hierarchy with lexicographic beliefs is now modelled by types in a lexicographic epistemic model: a type induces a lexicographic belief about his opponents' choice-type combinations.

Epistemic Model with Lexicographic Beliefs

Definition

A ***lexicographic epistemic model*** is a tuple $\mathcal{M}_l = \langle (T_i)_{i\in I}, (b_i^{lex})_{i\in I} \rangle$ such that

- T_i is a set of types for player i,
- every type $t_i \in T_i$ induces a lexicographic belief $b_i^{lex}(t_i)$ on the opponents' choice-type combinations $\times_{j\in I\setminus\{i\}} \left(C_j \times T_j\right)$.

Formalizing Caution

- **Intuition:** No opponent's choice is excluded from consideration, yet some opponent's choice can be deemed infinitely more likely than some other choice of his.

- A type t_i is said to deem possible an opponent's type t_j, whenever there exists some lexicographic level k such that t_j receives positive probability under b_i^k.

Definition

A type t_i is ***cautious***, whenever, if t_i deems possible some opponent's type t_j, then t_i also deems possible the choice-type pair (c_j, t_j) for all $c_j \in C_j$.

Interpretation

- Agent i is cautious, if for every mental set-up (“type”) that i deems possible for j to entertain, i does not exclude any feasible act.

Example: Where to read my book?

		Barbara: *Pub A*	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- Consider the following **lexicographic epistemic model**:
  - **Type Spaces:**
    $T_{you} = \{t_y, t'_y\}$
    $T_{Barbara} = \{t_B, t'_B\}$
  - **Beliefs for *You*:**
    $b^{lex}_{you}(t_y) = ((Pub\ A, t_B);\ \frac{1}{3}(Pub\ B, t'_B) + \frac{2}{3}(Pub\ C, t'_B))$
    $b^{lex}_{you}(t'_y) = (\frac{1}{2}(Pub\ A, t_B) + \frac{1}{2}(Pub\ B, t'_B); (Pub\ C, t'_B))$
  - **Beliefs for *Barbara*:**
    $b^{lex}_{Barbara}(t_B) = ((Pub\ A, t_y);\ \frac{3}{4}(Pub\ A, t'_y) + \frac{1}{4}(Pub\ C, t_y))$
    $b^{lex}_{Barbara}(t'_B) = ((Pub\ A, t'_y); (Pub\ B, t_y); (Pub\ C, t'_y))$
- No type in this lexicographic epistemic model is cautious!

Example: Where to read my book?

- A **lexicographic epistemic model** with a cautious type for *you*:
  - **Type Spaces:**

    $T_{you} = \{t_y, t'_y, t''_y\}$

    $T_{Barbara} = \{t_B, t'_B\}$

  - **Beliefs for *You*:**

    $b^{lex}_{you}(t_y) = ((Pub\ A, t_B);\ \frac{1}{3}(Pub\ B, t'_B) + \frac{2}{3}(Pub\ C, t'_B))$

    $b^{lex}_{you}(t'_y) = (\frac{1}{2}(Pub\ A, t_B) + \frac{1}{2}(Pub\ B, t'_B);\ (Pub\ C, t'_B))$

    $b^{lex}_{you}(t''_y) = ((Pub\ A, t_B);\ (Pub\ A, t'_B);\ \frac{1}{3}(Pub\ B, t_B) + \frac{2}{3}(Pub\ C, t'_B);\ \frac{1}{3}(Pub\ B, t'_B) + \frac{2}{3}(Pub\ C, t_B))$

  - **Beliefs for *Barbara*:**

    $b^{lex}_{Barbara}(t_B) = ((Pub\ A, t_y);\ \frac{3}{4}(Pub\ A, t'_y) + \frac{1}{4}(Pub\ C, t_y))$

    $b^{lex}_{Barbara}(t'_B) = ((Pub\ A, t'_y);\ (Pub\ B, t_y);\ (Pub\ C, t'_y))$

- Your type t''_y is cautious!

Agenda

- Lexicographic Beliefs
- Lexicographic Epistemic Models
- **Common Full Belief in (Caution & Primary Belief in Rationality)**
- Existence
- Algorithm

Being Cautious and Believing in Rationality

- Caution and belief in the opponents' rationality at all lexicographic levels is generally impossible!

- Indeed, caution requires every choice – including non-rational ones (i.e. choices that are not optimal for any belief) – to receive positive probability at some lexicographic level.

Primary Belief in Rationality

- A type t_i is said to primarily believe in some property, if t_i's primary belief only assigns positive probability to j's choice-type pairs that satisfy this property.

Definition

A type t_i ***primarily believes in rationality***, whenever t_i's level-1 belief only assigns positive probability to opponent choice-type pairs (c_j, t_j) such that c_j is optimal for t_j.

- **Remark.**
  Note that no conditions are put on any lexicographic level deeper than the primary one!

Example: Where to read my book?

		Barbara: *Pub A*	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- **Type Spaces:**

  $T_{you} = \{t_y, t'_y\}$

  $T_{Barbara} = \{t_B, t'_B\}$

- **Beliefs for *You*:**

  $b_{you}(t_y) = ((Pub\ A, t_B); \frac{1}{3}(Pub\ B, t'_B) + \frac{2}{3}(Pub\ C, t'_B))$

  $b_{you}(t'_y) = (\frac{1}{2}(Pub\ A, t_B) + \frac{1}{2}(Pub\ B, t'_B); (Pub\ C, t'_B))$

- **Beliefs for *Barbara*:**

  $b_{Barbara}(t_B) = ((Pub\ B, t_y); \frac{3}{4}(Pub\ A, t'_y) + \frac{1}{4}(Pub\ C, t_y))$

  $b_{Barbara}(t'_B) = ((Pub\ A, t'_y); (Pub\ B, t_y); (Pub\ C, t'_y))$

- If *you* primarily believe in *Barbara*'s rationality, then your primary belief must only assign positive probability to *Barbara's* choice *Pub A*.
- Type t_y primarily believes in *Barbara's* rationality and t'_y does not.
- Type t_B primarily believes in *your* rationality and t'_B does not.

Common Full Belief in (Caution & Primary Belief in Rationality)

Definition

A type t_i expresses ***common full belief in (caution & primary belief in rationality)***, whenever

- t_i expresses 1-fold full belief in caution and primary belief in rationality, i.e. t_i primarily believes in j's rationality and only deems possible types t_j that are cautious,
- t_i expresses 2-fold full belief in caution and primary belief in rationality, i.e. t_i only deems possible types t_j that express 1-fold belief in caution and primary believe in rationality,
- t_i expresses 3-fold full belief in caution and primary belief in rationality, i.e. t_i only deems possible types t_j that express 2-fold belief in caution and primary believe in rationality,
- etc.

Example: Should I call or not?

		Barbara	
		Godfather	*Casablanca*
You	*call*	1, 0	0, 1
	not call	0, 0	0, 1

- **Type Spaces:**

  $T_{you} = \{t_y\}$
  $T_{Barbara} = \{t_B\}$

- **Beliefs for** ***You*****:**

  $b_{you}(t_y) = ((Casablanca, t_B); (Godfather, t_B))$

- **Beliefs for** ***Barbara*****:**

  $b_{Barbara}(t_B) = ((call, t_y); (not\ call, t_y))$

- If you are cautious then your only optimal choice is *call*.
- Your type t_y is cautious – thus *call* is optimal for him – and expresses common full belief in caution and primary belief in rationality.
- Hence, you can rationally and cautiously choose *call* under common full belief in caution and primary belief in rationality.

Example: Where to read my book?

		Barbara *Pub A*	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- If you primarily believe in Barbara's rationality, then your primary belief must assign probability 1 to Barbara's choice *Pub A*.
- Hence, *Pub A* cannot be optimal for you.
- Which of your remaining choices – *Pub B* and *Pub C* – can you rationally choose under caution and common full belief in caution and primary belief in rationality?

Example: Where to read my book?

		Barbara: *Pub A*	Barbara: *Pub B*	Barbara: *Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- **Type Spaces:**

  $T_{you} = \{t_y\}$
  $T_{Barbara} = \{t_B\}$

- **Beliefs for *You*:**

  $b_{you}(t_y) = ((Pub\ A, t_B); \frac{1}{3}(Pub\ B, t_B) + \frac{2}{3}(Pub\ C, t_B))$

- **Beliefs for *Barbara*:**

  $b_{Barbara}(t_B) = ((Pub\ B, t_y); \frac{1}{2}(Pub\ A, t_y) + \frac{1}{2}(Pub\ C, t_Y))$

- Your type t_y is cautious and expresses common full belief in caution and primary belief in rationality.
- Your choice *Pub B* is optimal for type t_y.
- Hence, you can rationally and cautiously choose *Pub B* under common full belief in caution and primary belief in rationality.

Example: Where to read my book?

		Barbara: *Pub A*	*Pub B*	*Pub C*
You	*Pub A*	0, 3	1, 2	1, 1
	Pub B	1, 3	0, 2	1, 1
	Pub C	1, 3	1, 2	0, 1

- **Type Spaces:**

  $T_{you} = \{t_y\}$
  $T_{Barbara} = \{t_B\}$
- **Beliefs for *You*:**

  $b_{you}(t_y) = ((Pub\ A, t_B); \frac{2}{3}(Pub\ B, t_B) + \frac{1}{3}(Pub\ C, t_B))$
- **Beliefs for *Barbara*:**

  $b_{Barbara}(t_B) = ((Pub\ C, t_y); \frac{1}{2}(Pub\ A, t_y) + \frac{1}{2}(Pub\ C, t_Y))$

- Your type t_y is cautious and expresses common full belief in caution and primary belief in rationality.
- Your choice *Pub C* is optimal for type t_y.
- Hence, you can rationally and cautiously choose *Pub C* under common full belief in caution and primary belief in rationality.

Agenda

- Lexicographic Beliefs
- Lexicographic Epistemic Models
- Common Full Belief in (Caution & Primary Belief in Rationality)
- **Existence**
- Algorithm

A Way of Cautious Reasoning

- A lexicographic cautious way of reasoning – Common Full Belief (in Caution & Primary Belief in Rationality) – has been introduced.

- Accordingly, a type

  - only deems possible cautious opponent types and primarily believes in his opponents' rationality,
    [= 1-fold full belief in (caution & primary belief in rationality]
  - only deems possible opponent types that only deem possible cautious opponent types and primarily believe in their opponents' rationality,
    [= 2-fold full belief in (caution & primary belief in rationality)]
  - only deems possible opponent types that only deem possible opponent types that only deem possible cautious opponent types and primarily believe in their opponents' rationality,
    [= 3-fold full belief in (caution & primary belief in rationality)]
  - etc.

- **Two remaining key questions:**
  existence and algorithmic characterization

Example: Hide and Seek

Story

- *You* would like to go to a pub to read your book.
- *Barbara* is going to a pub as well, but *you* forgot to ask her to which one.
- *You* would like to avoid *Barbara*, in order to enjoy reading your book in silence.
- *Barbara* prefers *Pub A* to *Pub B*, and *Pub B* to *Pub C*, and would also like to talk to *you*.
- Question: Which pub should *you* go to?

Example: Hide and Seek

		Barbara		
		Pub A	*Pub B*	*Pub C*
You	*Pub A*	0, 5	1, 2	1, 1
	Pub B	1, 3	0, 4	1, 1
	Pub C	1, 3	1, 2	0, 3

Example: Hide and Seek

		Barbara A_B	B_B	C_B
You	A_y	0, 5	1, 2	1, 1
	B_y	1, 3	0, 4	1, 1
	C_y	1, 3	1, 2	0, 3

Is common full belief in (caution & primary belief in rationality) possible in this game?

- Consider some arbitrary cautious lexicographic belief for *you* about Barbara's choice, e.g. $(A_B; B_B; C_B)$.
- Given this belief, the choice C_y is optimal for *you*.
- Consider the belief $(C_y; A_y; B_y)$ for *Barbara* about your choice.
- Given this belief, the choice A_B is optimal for *Barbara*.
- Consider the belief $(A_B; B_B; C_B)$ for *you* about Barbara's choice.
- A **chain of lexicographic beliefs** has thus been formed which has entered in a cylce:
  $(A_B; B_B; C_B) \rightarrow (C_y; A_y; B_y) \rightarrow (A_B; B_B; C_B)$

Example: Hide and Seek

		Barbara		
		A_B	B_B	C_B
You	A_y	0, 5	1, 2	1, 1
	B_y	1, 3	0, 4	1, 1
	C_y	1, 3	1, 2	0, 3

- The cycle $(A_B; B_B; C_B) \rightarrow (C_y; A_y; B_y) \rightarrow (A_B; B_B; C_B)$ is now transformed into a **lexicographic epistemic model**.
- **Type Spaces:** $T_{you} = \{t_y\}$ and $T_{Barbara} = \{t_B\}$
- **Beliefs for** *You*: $b^{lex}_{you}(t_y) = ((A_B, t_B); (B_B, t_B); (C_B, t_B))$
- **Beliefs for** *Barbara*: $b^{lex}_{Barbara}(t_B) = ((C_y, t_y); (A_y, t_y); (B_y, t_y))$
- Both types in the epistemic model t_y and t_B are cautious and primarily believe in rationality.
- Hence, both types t_y and t_B express common full belief in caution and primary belief in rationality.
- Concluding, common full belief in (caution & primary belief in rationality) is indeed possible in the *Hide and Seek* game.

Generalizing the Construction for Existence

- Fix some finite game and consider an arbitrary cautious lexicographic belief b_i^{lex1} for player i about j's choice.
- Let c_i^1 be optimal given this belief.
- Consider some cautious lexicographic belief b_j^{lex2} for player j about i's choice such that the primary belief assigns probability 1 to c_i^1 and also probability 1 to some choice at all deeper levels.
- Let c_j^2 be optimal given this belief.
- Consider some cautious lexicographic belief b_i^{lex3} for player i about j's choice such that the primary belief assigns probability 1 to c_j^2 and also probability 1 to some choice at all deeper levels..
- Let c_i^3 be optimal given this belief.
- etc.
- The sequence of lexicographic beliefs thus constructed bears the following property:
  The unique choice in the support of the primary belief of any element of the sequence is optimal given the immediate predecessor lexicographic belief in the sequence.
- Since there are only finitely many choices and the same choices can always be specified for the support of all belief levels beyond level 1, respectively, the sequence of lexicographic beliefs must eventually enter into a cycle of lexicographic beliefs.

From Lexicographic Beliefs to Types

- Suppose some cycle of lexicographic beliefs:
  $b_i^{lex\,1} \rightarrow b_j^{lex\,2} \rightarrow b_i^{lex\,3} \rightarrow \ldots \rightarrow b_j^{lex\,K} \rightarrow b_i^{lex\,1}$

- This cycle can be transformed into an lexicographic epistemic model:
  - $b_i(t_i^1) = (b_i^{lex\,1}, t_j^K)$, where $b_i^{lex\,1} = (c_j^K; \ldots)$
  - $b_j(t_j^2) = (b_j^{lex\,2}, t_i^1)$, where $b_j^{lex\,2} = (c_i^1; \ldots)$
  - $b_i(t_i^3) = (b_i^{lex\,3}, t_j^2)$, where $b_i^{lex\,3} = (c_j^2; \ldots)$
  - $b_j(t_j^4) = (b_j^{lex\,4}, t_i^3)$, where $b_j^{lex\,4} = (c_i^3; \ldots)$
  - etc.

- In such an epistemic model, every type is cautious and primarily believes in rationality.

- Hence, all types express common full belief in (caution & primary belief in rationality)!

Existence

Theorem

Let Γ be some finite two player game. Then, there exists a lexicographic epistemic model such that

- *every type in the model is cautious and expresses common full belief in (caution & primary belief in rationality),*
- *every type in the model deems possible only one opponent's type, and assigns at each lexicographic level probability 1 to one of the opponent's choices.*

Agenda

- Lexicographic Beliefs
- Lexicographic Epistemic Models
- Common Full Belief in (Caution & Primary Belief in Rationality)
- Existence
- **Algorithm**

Towards Characterizing Cautious Reasoning

Definition

A choice c_i of player i is ***weakly dominated*** by some randomized choice $r_i \in \Delta(C_i)$, whenever

- $U_i(c_i, c_j) \leq V_i(r_i, c_j)$ for all $c_j \in C_j$,
- there exists $c_j^* \in C_j$ such that $U_i(c_i, c_j^*) < V_i(r_i, c_j^*)$.

Characterizing Cautious Reasoning

An analogy to **Pearce's Lemma** for lexicographic beliefs:

Theorem

A choice c_i of player i can **optimally** *be chosen under a* **cautious lexicographic belief** *if and only if c_i is* **not weakly dominated** *by some randomized choice r_i.*

Randomized Choices and Lexicographic Expected Utility

The k-level expected utility $v_i^k(r_i, b_i^{lex})$ of a randomized choice $r_i \in \Delta(C_i)$ is defined as

$$v_i^k(r_i, b_i^{lex}) := \sum_{c_j \in C_j} b_i^k(c_j)\Big(\sum_{c_i \in C_i} \big(r_i(c_i) \cdot U_i(c_i, c_j)\big)\Big)$$

A basic lemma

Basic-Lemma I

Let I be some index set, $0 \leq \alpha_i \leq 1$ for all $i \in I$ such that $\sum_{i \in I} \alpha_i = 1$, $x \in \mathbb{R}$, and $y_i \in \mathbb{R}$ for all $i \in I$. If $x < \sum_{i \in I} \alpha_i y_i$, then there exists $i^* \in I$ such that $x < y_{i^*}$.

Proof:

- Towards a contradiction suppose that $x \geq y_i$ for all $i \in I$.
- Then, $\alpha_i x \geq \alpha_i y_i$ holds for all $i \in I$.
- It directly follows that $1 \cdot x = \sum_{i \in I} \alpha_i x \geq \sum_{i \in I} \alpha_i y_i$, a contradiction.

A second basic lemma

Basic-Lemma II

Let I be some index set, $0 < \alpha_i < 1$ for all $i \in I$ such that $\sum_{i \in I} \alpha_i = 1$, $x \in \mathbb{R}$, and $y_i \in \mathbb{R}$ for all $i \in I$. If $x \leq \sum_{i \in I} \alpha_i y_i$, then (there exists $i^* \in I$ such that $x < y_{i^*}$) or ($x = y_i$ for all $i \in I$).

Proof:

- By contraposition, suppose that $x \geq y_i$ for all $i \in I$ and that there exists $i' \in I$ such that $x \neq y_{i'}$.
- Then, $x > y_{i'}$.
- As $0 < \alpha_i < 1$ holds for all $i \in I$, it is the case that $\alpha_{i'} x > \alpha_{i'} y_{i'}$ and $\alpha_i x \geq \alpha_i y_i$ for all $i \in I \setminus \{i'\}$.
- It follows that $x = \sum_{i \in I} \alpha_i x > \sum_{i \in I} \alpha_i y_i$.

Proof of the *only if* ($\Rightarrow$) Direction of the Theorem

- The proof proceeds by contraposition.
- Let $c_i \in C_i$ be weakly dominated by some randomized choice $r_i \in \Delta(C_i)$.
- Thus, $U_i(c_i, c_j) \leq \sum_{c_i \in C_i} \left(r_i(c_i) \cdot U_i(c_i, c_j)\right)$ for all $c_j \in C_j$ and there exists some choice $c_j^* \in C_j$ such that $U_i(c_i, c_j^*) < \sum_{c_i \in C_i} \left(r_i(c_i) \cdot U_i(c_i, c_j^*)\right)$.
- Suppose that player i holds some cautious lexicographic belief $b_i^{lex} = (b_i^1, b_i^2, \ldots, b_i^K)$.
- Then, for all levels k

$$\sum_{c_j \in C_j} \left(b_i^k(c_j) \cdot U_i(c_i, c_j)\right) \leq \sum_{c_j \in C_j} \left(b_i^k(c_j) \sum_{c_i \in C_i} \left(r_i(c_i) \cdot U_i(c_i, c_j)\right)\right)$$

i.e.

$$u_i^k(c_i, b_i^{lex}) \leq \sum_{c_i' \in C_i} r_i(c_i') u_i^k(c_i', b_i^{lex}) = v_i^k(r_i, b_i^{lex}),$$

and, by caution there exists a level k^* such that $c_j^* \in \text{supp}(b_i^{k^*})$ and thus

$$\sum_{c_j \in C_j} \left(b_i^{k^*}(c_j) \cdot U_i(c_i, c_j)\right) < \sum_{c_j \in C_j} \left(b_i^{k^*}(c_j) \sum_{c_i \in C_i} \left(r_i(c_i) \cdot U_i(c_i, c_j)\right)\right)$$

i.e.

$$u_i^{k^*}(c_i, b_i^{lex}) < \sum_{c_i' \imath C_i} r_i(c_i') u_i^{k^*}(c_i', b_i^{lex}) = v_i^{k^*}(r_i, b_i^{lex}).$$

Proof of the *only if* (⇒) Direction of the Theorem (continued)

- Consider the set $\text{supp}(r_i) \subseteq C_i$ of i's choices to which r_i assigns positive probability and level-1 belief b_i^1.
- Then, by Basic-Lemma II, either **(a)** there exists some $c_i' \in \text{supp}(r_i)$ such that $u_i^1(c_i, b_i^{lex}) < u_i^1(c_i', b_i^{lex})$, or **(b)** $u_i^1(c_i, b_i^{lex}) = u_i^1(c_i', b_i^{lex})$ for all $c_i' \in \text{supp}(r_i)$.
- If case **(a)** holds, then player i prefers c_i' to c_i, and c_i is thus not optimal.
- If case **(b)** holds, i.e., $u_i^1(c_i, b_i^{lex}) = u_i^1(c_i', b_i^{lex})$ for all $c_i' \in \text{supp}(r_i)$, then consider b_i^2.
- Then, again by Basic-Lemma II, either **(a)** there exists some $c_i' \in \text{supp}(r_i)$ such that $u_i^2(c_i, b_i^{lex}) < u_i^2(c_i', b_i^{lex})$, or **(b)** $u_i^2(c_i, b_i^{lex}) = u_i^2(c_i', b_i^{lex})$ for all $c_i' \in \text{supp}(r_i)$.
- If case **(a)** holds, then $u_i^1(c_i, b_i^{lex}) = u_i^1(c_i', b_i^{lex})$ and $u_i^2(c_i, b_i^{lex}) < u_i^2(c_i', b_i^{lex})$, and consequently player i prefers c_i' to c_i, implying that c_i is not optimal.
- If case **(b)** holds, i.e., $u_i^1(c_i, b_i^{lex}) = u_i^1(c_i', b_i^{lex})$ and $u_i^2(c_i, b_i^{lex}) = u_i^2(c_i', b_i^{lex})$ for all $c_i' \in \text{supp}(r_i)$, then consider b_i^3.
- etc.
- As $u_i^{k^*}(c_i, b_i^{lex}) < v_i^{k^*}(r_i, b_i^{lex})$ there must eventually be some level l' such that – by Basic-Lemma I – it is the case that $u_i^l(c_i, b_i^{lex}) < u_i^{l'}(c_i', b_i^{lex})$ for some $c_i' \in \text{supp}(r_i)$.
- Hence, there exists some choice $c_i' \in \text{supp}(r_i)$ that player i prefers to c_i, and therefore c_i is not optimal.

Towards an Algorithm

It is desirable to algorithmically characterize the choices under

- rationality (=optimality given the agent's beliefs),
- caution,
- common full belief in (caution & primary belief in rationality).

Lexicographic Optimality and Standard Optimality

Lemma

If a choice c_i is lexicographically-optimal given a lexicographic belief b_i^{lex}, then c_i is standard-optimal given b_i^1.

Proof:

- Towards a contradiction suppose that c_i is lexicographically-optimal given b_i^{lex}, but not standard-optimal given b_i^1.
- Then, there exists a choice $c_i^* \in C_i$ such that $u_i^1(c_i, b_i^{lex}) = u_i(c_i, b_i^1) < u_i(c_i^*, b_i^1) = u_i^1(c_i^*, b_i^{lex})$.
- However, this contradicts lexicographic optimality of c_i according to which there exists no choice $c_i' \in C_i$ such that $u_i^k(c_i, b_i^{lex}) < u_i^k(c_i', b_i^{lex})$ for some level k and $u_i^k(c_i, b_i^{lex}) = u_i^k(c_i', b_i^{lex})$ for all levels $l < k$.

Step 1

1-fold full belief in caution and primary belief in rationality

- Which choices can optimally and cautiously be made under 1-fold full belief in caution and primary belief in rationality?
- Suppose that type t_i is cautious and expresses 1-fold full belief in caution and primary belief in rationality.
- Then, by the Theorem, t_i's primary belief assigns probability 0 to all weakly dominated choices for j.
- Note that due to t_i being cautious, t_i cannot optimally choose any weakly dominated choice himself.
- Let Γ^1 be the reduced game that remains after eliminating all weakly dominated choices from the game: t_i's primary belief is concentrated on Γ^1.
- Hence, every optimal choice for t_i must be optimal for some lexicographic belief with primary belief restricted to Γ^1, i.e. standard-optimal given the primary belief.
- Thus, by Pearce's Lemma applied to Γ^1, every optimal choice for t_i must not be strictly dominated on Γ^1.
- Let Γ^2 be the reduced game that remains after eliminating all strictly dominated choices from Γ^1.
- Then, every optimal choice for t_i must be in Γ^2.
- **Conclusion:** If type t_i is cautious and expresses 1-fold full belief in caution and primary belief in rationality, then every optimal choice for t_i must be in Γ^2.
- Note that Γ^2 is obtained by first eliminating all weakly dominated choices, and then eliminating all strictly dominated choices.

Step 2

Up to 2-fold full belief in caution and primary belief in rationality

- Which choices can optimally and cautiously be made under up to 2-fold full belief in caution and primary belief in rationality?
- Suppose that type t_i is cautious and expresses up to 2-fold full belief in caution and primary belief in rationality.
- Then, t_i's primary belief only assigns positive probability to choice-type pairs (c_j, t_j) such that c_j is optimal for t_j, and t_j expresses 1-fold full belief in caution and primary belief in rationality.
- From Step 1 it follows that all such choices c_j receiving positive probability by t_i's primary belief are in Γ^2.
- As t_i satisfies 1-fold full belief in caution and primary belief in rationality, every optimal choice for t_i is in Γ^2.
- Hence, every optimal choice for t_i must be optimal for some lexicographic belief with primary belief restricted to Γ^2, i.e. standard-optimal given the primary belief.
- Thus, by Perace's Lemma applied to Γ^2, every optimal choice for t_i must not be strictly dominated in Γ^2.
- Let Γ^3 be the reduced game that remains after eliminating all strictly dominated choices from Γ^2.
- Then, every optimal choice for t_i must be in Γ^3.
- **Conclusion:** If type t_i is cautious and expresses up to 2-fold full belief in caution and primary belief in rationality, then every optimal choice for t_i must be in Γ^3.
- Note that Γ^3 is obtained by first eliminating all weakly dominated choices, and then applying two-fold strict dominance.

Algorithm

Definition (Dekel-Fudenberg-Procedure)

Step 1. Eliminate all choices that are weakly dominated in the game.

Step 2. Within the reduced game after Step 1, apply iterated strict dominance.

- The algorithm stops after finitely many steps.
- The algorithm returns a non-empty set.
- The order and speed in which choices are eliminated after Step 1 is not relevant for the set it returns.

Algorithmic Characterization

Theorem

For all $k \geq 1$, the choices that can rationally be made by a cautious type that expresses up to k-fold full belief in caution and primary belief in rationality are exactly those choices that survive the first $k+1$ steps of the Dekel-Fudenberg-Procedure.

Corollary

The choices that can rationally be made by a cautious type that expresses common full belief in caution and primary belief in rationality are exactly those choices that survive the Dekel-Fudenberg-Procedure.

Remark

- In fact, the epistemic concept can be weakened and *still* be characterized by the **Dekel-Fudenberg-Procedure**.
- The weaker concept of common primary belief in (caution & rationality *only* puts conditions on the *first lexicographic level*.

Common Primary Belief in (Caution & Rationality)

Definition

A type t_i expresses ***common primary belief in (caution & rationality)***, whenever

- t_i expresses 1-fold primary belief in caution and rationality, i.e. t_i primarily believes in j's caution and rationality,
- t_i expresses 2-fold primary belief in caution and rationality, i.e. t_i primarily believes that j expresses 1-fold belief in caution and rationality,
- t_i expresses 3-fold primary belief in caution and rationality, i.e. t_i primarily believes that j that express 2-fold belief in caution and rationality,
- etc.

Example: Teaching a Lesson

Story

- It is Friday and your teacher announces a surprise exam for next week.
- You must decide on what day you will start preparing for the exam.
- In order to pass the exam you must study for at least two days.
- For a perfect exam and a subsequent compliment by your father you need to study for at least six days.
- Passing the exam increases your utility by 5.
- Failing the exam increases the teacher's utility by 5.
- Every day you study decreases your utility by 1, but increases the teacher's utility by 1.
- A compliment by your father increases your utility by 4.

Example: Teaching a Lesson

Teacher

		Mon	*Tue*	*Wed*	*Thu*	*Fri*
	Sat	3, 2	2, 3	1, 4	0, 5	3, 6
	Sun	−1, 6	3, 2	2, 3	1, 4	0, 5
You	*Mon*	0, 5	−1, 6	3, 2	2, 3	1, 4
	Tue	0, 5	0, 5	−1, 6	3, 2	2, 3
	Wed	0, 5	0, 5	0, 5	−1, 6	3, 2

Example: Teaching a Lesson

		Teacher				
		Mon	*Tue*	*Wed*	*Thu*	*Fri*
	Sat	3, 2	2, 3	1, 4	0, 5	3, 6
	Sun	−1, 6	3, 2	2, 3	1, 4	0, 5
You	*Mon*	0, 5	−1, 6	3, 2	2, 3	1, 4
	Tue	0, 5	0, 5	−1, 6	3, 2	2, 3
	Wed	0, 5	0, 5	0, 5	−1, 6	3, 2

- With standard beliefs under common belief in rationality you can rationally choose any day.
- With standard beliefs under common belief in rationality and a simple belief hierarchy you can only rationally pick *Saturday* or *Wednesday*.
- What days can you rationally and cautiously choose under common full belief in caution and primary belief in rationality?

Example: Teaching a Lesson

You \ Teacher	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	−1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	−1, 6	3, 2	2, 3	1, 4
Tue	0, 5	0, 5	−1, 6	3, 2	2, 3
Wed	0, 5	0, 5	0, 5	−1, 6	3, 2

Step 1.

- *Your* choice *Wednesday* is weakly dominated by *your* choice *Saturday*.
- Eliminate *your* choice *Wednesday* from the original game.

Example: Teaching a Lesson

		Teacher				
		Mon	*Tue*	*Wed*	*Thu*	*Fri*
You	*Sat*	3, 2	2, 3	1, 4	0, 5	3, 6
	Sun	−1, 6	3, 2	2, 3	1, 4	0, 5
	Mon	0, 5	−1, 6	3, 2	2, 3	1, 4
	Tue	0, 5	0, 5	−1, 6	3, 2	2, 3

Step 2.

- The *teacher*'s choice *Thursday* is strictly dominated by *Friday*.
- Eliminate the *teacher*'s choice *Friday* from the reduced game after Step 1.

Example: Teaching a Lesson

		Teacher			
		Mon	*Tue*	*Wed*	*Fri*
You	*Sat*	3, 2	2, 3	1, 4	3, 6
	Sun	−1, 6	3, 2	2, 3	0, 5
	Mon	0, 5	−1, 6	3, 2	1, 4
	Tue	0, 5	0, 5	−1, 6	2, 3

Step 3.

- *Your* choice *Tuesday* is strictly dominated by *Saturday*.
- Eliminate the *your* choice *Tuesday* from the reduced game after Step 2.

Example: Teaching a Lesson

		Teacher			
		Mon	*Tue*	*Wed*	*Fri*
You	*Sat*	3, 2	2, 3	1, 4	3, 6
	Sun	−1, 6	3, 2	2, 3	0, 5
	Mon	0, 5	−1, 6	3, 2	1, 4

Step 4.

- The *teacher*'s choice *Wednesday* is strictly dominated by *Friday*.
- Eliminate the *teacher*'s choice *Wednesday* from the reduced game after Step 3.

Example: Teaching a Lesson

		Teacher		
		Mon	*Tue*	*Fri*
You	*Sat*	3, 2	2, 3	3, 6
	Sun	−1, 6	3, 2	0, 5
	Mon	0, 5	−1, 6	1, 4

Step 5.

- *Your* choice *Monday* is strictly dominated by *Saturday*.
- Eliminate *your* choice *Monday* from the reduced game after Step 4.

Example: Teaching a Lesson

		Teacher		
		Mon	*Tue*	*Fri*
You	*Sat*	3, 2	2, 3	3, 6
	Sun	−1, 6	3, 2	0, 5

Step 6.

- The *teacher*'s choice *Tuesday* is strictly dominated by *Friday*.
- Eliminate the *teacher*'s choice *Tuesday* from the reduced game after Step 5.

Example: Teaching a Lesson

		Teacher	
		Mon	*Fri*
You	*Sat*	$3, 2$	$3, 6$
	Sun	$-1, 6$	$0, 5$

Step 7.

- *Your* choice *Sunday* is strictly dominated by *Saturday*.
- Eliminate *your* choice *Sunday* from the reduced game after Step 6.

Example: Teaching a Lesson

		Teacher	
		Mon	*Fri*
You	*Sat*	3, 2	3, 6

Step 8.

- The *teacher*'s choice *Monday* is strictly dominated by *Friday*.
- Eliminate the *teacher*'s choice *Monday* from the reduced game after Step 7.

The algorithm stops.

		Teacher
		Fri
You	*Sat*	3, 6

Example: Teaching a Lesson

You \ Teacher	Mon	Tue	Wed	Thu	Fri
Sat	3, 2	2, 3	1, 4	0, 5	3, 6
Sun	−1, 6	3, 2	2, 3	1, 4	0, 5
Mon	0, 5	−1, 6	3, 2	2, 3	1, 4
Tue	0, 5	−1, 6	3, 2	2, 3	1, 4
Wed	0, 5	0, 5	0, 5	−1, 6	3, 2

- **Type Spaces:**

  $T_{you} = \{t_y\}$
  $T_{Barbara} = \{t_B\}$

- **Beliefs for You:**

  $b_{you}(t_y) = ((Fri, t_B); \frac{1}{4}(Mon, t_B) + \frac{1}{4}(Tue, t_B) + \frac{1}{4}(Wed, t_B) + \frac{1}{4}(Thu, t_B))$

- **Beliefs for Teacher:**

  $b_{Barbara}(t_B) = ((Sat, t_y); \frac{1}{4}(Sun, t_y) + \frac{1}{4}(Mon, t_y) + \frac{1}{4}(Tue, t_y) + \frac{1}{4}(Wed, t_y))$

- Your type t_y is cautious and expresses common full belief in caution and primary belief in rationality.
- Your choice *Saturday* is optimal for type t_y.
- Hence, you can indeed cautiously and rationally choose *Saturday* under common full belief in caution and primary belief in rationality.

Thank you!