Common Belief in Rationality in Psychological Games

Stephan Jagau

Nottingham University Business School
UK | CHINA | MALAYSIA

EPICENTER
Research Center for Epistemic Game Theory

July 8, 2024

Introduction

- So far preferences over choices only depended on first-order beliefs wrt opponent behavior.
- This lecture: What if players care about opponent behavior and beliefs?
- Two examples with second-order beliefs:
  - If aiming to **meet opponent's expectations** (aka guilt aversion) you prefer a choice to the extent that you believe the opponent expects you to make that choice.
  - If aiming to **surprise opponent** you prefer a choice to the extent that you believe the opponent expects you to *not* make that choice.
- **Notes:**
  - Here, guilt/surprise emerge as reflections wrt (not) matching expectations. Such insights make psychological game useful.
  - No new tools needed here. Instead, different notion of optimal choice leads to more complex setting.

Introductory Example

Surprising Barbara, *baseline decision problem*

- *You* and *Barbara* are invited to a party. Each of you simultaneously choose from dress colors *blue, green, red*.
- Personally, you prefer *blue* to *green* to *red*. In addition, you seek to wear *different* color than Barbara.
- Same for Barbara with color preference *red* to *blue* to *green*.

You	*blue*	*green*	*red*
blue	0	3	3
green	2	0	2
red	1	1	0

Barbara	*blue*	*green*	*red*
blue	0	2	2
green	1	0	1
red	3	3	0

Introductory Example

Surprising Barbara, *surprise utilities*

- Additionally, you seek to **surprise** Barbara, deriving additional utility for surprising choices proportional to your color preference. Same is true for Barbara.

	Barbara expects		
You	*blue*	*green*	*red*
blue	0	3	3
green	2	0	2
red	1	1	0

	You expect		
Barbara	*blue*	*green*	*red*
blue	0	2	2
green	1	0	1
red	3	3	0

Introductory Example

Surprising Barbara, *full decision problem*

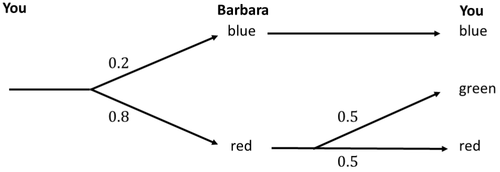
- Finally, suppose your overall utility is the sum of your baseline and surprise utilities.
- This yields decision problem with **choice-belief combinations** replacing choices for opponent.

You	(b,b)	(b,g)	(b,r)	(g,b)	(g,g)	(g,r)	(r,b)	(r,g)	(r,r)
blue	0	3	3	3	6	6	3	6	6
green	4	2	4	2	0	2	4	2	4
red	2	2	1	2	2	1	1	1	0

Barbara	(b,b)	(b,g)	(b,r)	(g,b)	(g,g)	(g,r)	(r,b)	(r,g)	(r,r)
blue	0	2	2	2	4	4	2	4	4
green	2	1	2	1	0	1	2	1	2
red	6	6	3	6	6	3	3	3	0

Introductory Example: Expected Utility

How to calculate **utility at a second-order belief**? Take following example:



- You believe w. 0.2: Barbara chooses $blue$ and believes you choose $blue$. $\Rightarrow$ State (b, b) in decision problem.
- Similarly, you assign $0.8 \cdot 0.5 = 0.4$ each to states (r, g) and (r, r).
- Then, for example, choosing $blue$ yields expected utility $0.2 \cdot 0 + 0.4 \cdot 6 + 0.4 \cdot 6 = 4.8$.

Introductory Example: Rationality

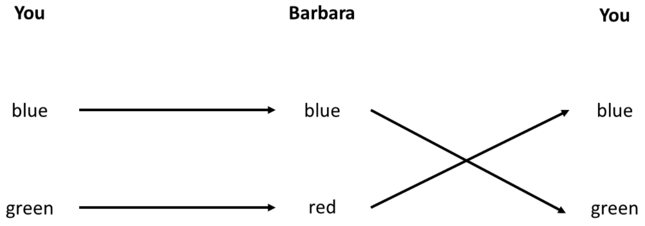
You	(b,b)	(b,g)	(b,r)	(g,b)	(g,g)	(g,r)	(r,b)	(r,g)	(r,r)
blue	0	3	3	3	6	6	3	6	6
green	4	2	4	2	0	2	4	2	4
red	2	2	1	2	2	1	1	1	0

Barbara	(b,b)	(b,g)	(b,r)	(g,b)	(g,g)	(g,r)	(r,b)	(r,g)	(r,r)
blue	0	2	2	2	4	4	2	4	4
green	2	1	2	1	0	1	2	1	2
red	6	6	3	6	6	3	3	3	0

- Your choice red is strictly dominated by (e.g.) $0.4 \cdot blue + 0.6 \cdot green$.Similarly, $green$ strictly dominated for Barbara by (e.g.) $0.4 \cdot red + 0.6 \cdot blue$.
- Hence, no *second-order belief* makes these choices optimal for you and Barbara. $\Rightarrow$ irrational

Introductory Example: Rationality

Remaining choices *blue* and *green* rational for you:



- *blue* strictly optimal if you believe Barbara chooses *blue* and believes you choose *green* (state (b, g)). Similar for *green* at (r, b).
- Also, *blue* is optimal for Barbara at (g, r) and *red* is optimal for her at (b, b).

$\Rightarrow$ Common belief in rationality.

- **Note:** Both can choose at least 2 colors, so surprise possible at CBR.

Agenda

- Psychological Games and Common Belief in Rationality
- Procedural Characterization
- Possibility
- Variants of the Procedure

Second-Order Expectations

Definition

A ***second-order expectation*** for player i is a probability distribution $e_i \in \Delta(C_i \times C_j)$.

- Second-order expectations concern events of form "player j chooses c_j and believes player i chooses c_i" ($\hat{=}\ e_i(c_j, c_i)$).
- Formally, every second-order belief $b_i^2 \in \Delta(C_j \times \Delta(C_i))$ induces a second-order expectation e_i via

$$e_i(c_j, c_i) = b_i^1(c_j) \int_{\Delta(C_i)} b_j^1(c_i') \, \mathrm{d}b_i^2(|c_j),$$

where $b_i^2(E|c_j) = b_i^2(\{c_j\} \times E)/b_i^2(\{c_j\} \times \Delta(C_i))$ for every $E \subseteq \Delta(C_i)$.

(Linear) Psychological Games (of Order 2)

Definition

A ***psychological game*** with two players specifies

a) finite set of choices C_i for both players i,

b) utility function $u_i : C_i \times \Delta(C_j \times C_i) \to \mathbb{R}$ for both players i,

where
$$u_i(c_i, e_i) = \sum_{(c_j, c_i') \in C_j \times C_i} e_i(c_j, c_i) u_i(c_i, (c_j, c_i')).$$

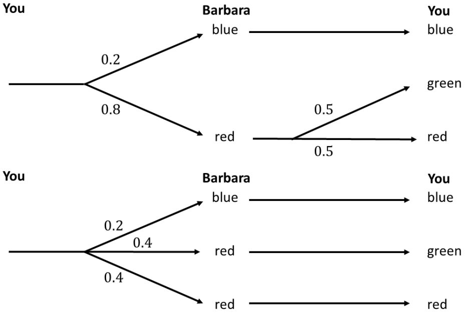
Notes:

- u_i generalizes standard expected utility using expectations.
- Assumptions: (i) u_i depends on **second-order beliefs** only,
  (ii) u_i is **linear in up to level-2 uncertainty**.

  $\Rightarrow$ Decision problems with set of states $C_j \times C_i$ iso C_j.
- General psychological games: u_i *non-linear* in *full belief hierarchy*.

Linearity in up to Second-Order Uncertainty

Reconsider introductory example:



- Both second-order beliefs above induce the same expectation $e_i = 0.2 \cdot (b, b) + 0.4 \cdot (r, g) + 0.4 \cdot (r, r)$.
- Intuitively, it does not matter whether uncertainty emanates at level 1 (other's behavior) or level 2 (other's beliefs about behavior).

Epistemic Model for Introductory Example

- **Types:** $T_1 = \{t_1^{blue}, t_1^{green}\}$, $T_2 = \{t_2^{blue}, t_2^{red}\}$
- **Beliefs for *You*:** $b_1(t_1^{blue}) = 0.8 \cdot (blue, t_2^{blue}) + 0.2 \cdot (red, t_2^{red})$,
  $b_1(t_1^{green}) = (red, t_2^{red})$.
- **Beliefs for *Barbara*:** $b_2(t_2^{blue}) = (green, t_1^{green})$,
  $b_2(t_2^{red}) = 0.9 \cdot (blue, t_1^{blue}) + 0.1 \cdot (green, t_1^{green})$.

Types, Optimal and Rational Choices

- Consider epistemic models like in Chapter 3, but now possibly with **infinitely** many types.
- Main change in psychological games: optimality is wrt exectations.

Definition

Take type t_i with expectation e_i. Choice $c_i \in C_i$ is *optimal* for t_i if

$$u_i(c_i, t_i) = u_i(c_i, e_i) = \sum_{(c_j, c_i') \in C_j \times C_i} e_i(c_j, c_i') u_i(c_i, (c_j, c_i')) \geq u_i(c_i'', e_i)$$

for all $c_i'' \in C_i$.

(Common) Belief in Rationality

Up to k-fold/common belief in rationality now defined like in standard game:

Definition

Type t_i,

- *believes in the opponents' rationality* if $b_i(t_i)$ only deems possible (c_j, t_j) where c_j is optimal for t_j,
- *expresses up to k-fold belief in rationality* for $k \geq 1$ if $b_i(t_i)$ only deems possible (c_j, t_j) where c_j is optimal for t_j expressing up to $(k-1)$-fold belief in rationality,
- *expresses common belief in rationality* if $b_i(t_i)$ expresses up to k-fold belief in rationality for all $k \geq 1$.

Agenda

- Psychological Games and Common Belief in Rationality
- Procedural Characterization
- Possibility
- Variants of the Procedure

Towards an Iterative Procedure

- To find all choices consistent with common belief in rationality, we generalize iterated strict dominance.
- As seen in following example, eliminating strictly dominated choices and corresponding (standard) states in decision problems is not enough.
- More surprisingly, also eliminating choices and full states (deterministic second-order expectations) is not enough.

Example: “Black and White Dinner with a Twist”

- *You* and *Barbara* go to a dinner an simultaneously choose from dress colors $black$ and $white$.
- Personally, you prefer $white$ to $black$. However, to the degree that you believe Barbara wears $white$ and expects you to wear $white$, you slightly prefer $black$.
- Barbara’s preferences are the same with $black$ and $white$ reversed.

You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	(w_2, w_1)
black	0	0	0	3
white	2	2	2	2

Barbara	(b_1, b_2)	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	3	0	0	0

- Note that no choice is strictly dominated for you or Barbara!

"Black and White Dinner with a Twist": Rationality

You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	(w_2, w_1)
black	0	0	0	3
white	2	2	2	2

Barbara	(b_1, b_2)	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	3	0	0	0

- Even though no strategy is dominated, we are not done yet.
- **Why?**
  - Utilities depend on **second-order expectations**.
  - Hence, need to track **choices and first-order beliefs**.
- *black* rational for you iff $e_1(w_2, w_1) \geq 2/3$.
- Similarly, *white* rational for Barbara iff $e_2(b_1, b_2) \geq 2/3$.

"Blk and Wt Dinner w Twist": Belief in Rationality

You	(b_2,b_1)	(b_2,w_1)	(w_2,b_1)	(w_2,w_1)
black	0	0	0	3
white	2	2	2	2

Barbara	(b_1,b_2)	(b_1,w_2)	(w_1,b_2)	(w_1,w_2)
black	2	2	2	2
white	3	0	0	0

- How does belief in rationality affect states you deem possible?
  - For Barbara to rationally play *white*, need $b_2^1(b_1) \geq \frac{2}{3}$.
    (If not, could never have $e_2(b_1,b_2) \geq \frac{2}{3}$.)
  - But then, using Bayes' rule, belief in Barbara's rationality implies $e_1(w_1|w_2) = \frac{e_1(w_2,w_1)}{e_1(w_2,b_1)+e_1(w_2,w_1)} \leq \frac{1/3}{2/3+1/3} = 1/3.$

  ⇒ Conditional on Barbara rationally choosing w_2, you must believe Barbara assigns at most $1/3$ to your choice w_1.
- Similarly, belief in rationality implies $e_2(b_1|b_2) \leq 1/3$ for Barbara.

"Blk and Wt Dinner w Twist": Belief in Rationality

You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	(w_2, w_1)
black	0	0	0	3
white	2	2	2	2

Barbara	(b_1, b_2)	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	3	0	0	0

- But then, *black* is not rational for you under belief in rationality! **Why?**
  - Rationality of *black* for you requires $e_1(w_2, w_1) \geq 2/3$
  - Belief in Barbara's rationality requires $e_1(w_1|w_2) \leq 1/3$.
  - The latter implies $e_1(w_2, w_1) = b_1^1(w_2)\,[e_1(w_1|w_2)] \leq 1/3$.

  $\Rightarrow \perp$.
- Similarly, *white* is not rational for Barbara under belief in rationality.

"Blk and Wt Dinner w Twist": Belief in Rationality

- Clearly, cannot capture reasoning using strict dominance and elimination of standard states.
- However, also no **full** state among (b_2, b_1), (b_2, w_1), (w_2, w_1), (w_1, w_2) can be eliminated here (and similarly for Barbara).
- **Why?** Barbara's rational choice $white$ puts **probabilistic** upper bound $1/3$ on her belief in w_1 (and analogously for you).
- Hence, correct decision problems for belief in rationality:

You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	$(w_2, 2/3 \cdot b_1 + 1/3 \cdot w_1)$
black	0	0	0	1
white	2	2	2	2

Barbara	$(b_1, 2/3 \cdot w_2 + 1/3 \cdot b_2)$	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	1	0	0	0

“Blk and Wt Dinner w Twist”: CBR

- Eliminating *black* for you and *white* for Barbara (and one more round of eliminating states) yields:

You	(b_2, w_1)
white	2

Barbara	(w_1, b_2)
black	2

$\Rightarrow$ *white* for you and *black* for Barbara uniquely rational under CBR.

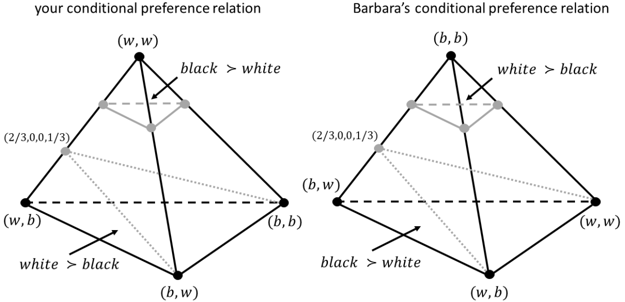
Elimination of Second-Order Expectations

- Crucial step in example: Eliminate e_i inconsistent w. j's rationality.
- More generally, following recipe:
  1) For every undominated c_j, find expectations $E_j(c_j)$ making c_j optimal.
  2) Let $B_j(c_j) = \{b_j \in \Delta(C_j) | b_j = \text{marg}_{C_j} e_j$ for some $e_j \in E_j(c_j)\}$ be corresponding first-order beliefs.
  3) Then, conditional on c_j, i must believe j's first-order belief is in $B_j(c_j)$. Formally, $e_i(|c_j) \in B_j(c_j)$, where

$$e_i(c_i|c_j) = \frac{e_i(c_j, c_i)}{\sum_{c_i' \in C_i} e_i(c_j, c_i')} \text{ for all } c_i \in C_i.$$

- **Notes:**
  - Let E_i be i's expectations satisfying (3). E_i is convex combination of finitely many extreme $e_i \in \Delta(C_j \times C_i)$.
  - Repeat steps above for e_i (in)consistent w. up to k-fold belief in rationality, $k > 1$.

"Blk and Wt Dinner w Twist": Eliminating Second-Order Expectations



- Tetrahedron: $\Delta(C_j \times C_i)$-probability simplex.
- Solid triangle: Indifference hyperplane for choices *black* and *white*.
- Dotted triangle and below: Expectations consistent with belief in rationality.

It. Elim. of Choices and Second-Order Expectations

Definition

Round 1. For both players i, eliminate all strictly dominated choices. For all other c_i, let $E_i^1(c_i)$ be supporting expectations.

Round $k \geq 1$. For each player i and opp. choice c_j, let $B_j^{k-1}(c_j)$ be first-order beliefs induced by $E_j^{k-1}(c_j)$, and let E_i^k be i's expectations s.th. $e_i(|c_j) \in B_j^{k-1}(c_j)$ f. all c_j deemed possible by e_i. Eliminate all choices c_i that are not optimal for any $e_i \in E_i^k$. For all other c_i, let $E_i^k(c_i)$ be supporting expectations.

Proceed until no more choices/expectations can be eliminated.

Theorem

For any $k \geq 1$, choice c_i is rational for player i under up to k-fold (common) belief in rationality iff c_i survives $(k+1)$-fold (iterated) elimination of choices and expectations.

stephan.jagau@nottingham.ac.uk

Example: "Dinner w Strong Preference f Surprise"

- *You* and *Barbara* go to a dinner an simultaneously choose from dress colors $black$ and $white$.
- Your preferences are the same as before, except each of you more strongly prefers your less liked choice if you mismatch with your opponent and surprise them as well.

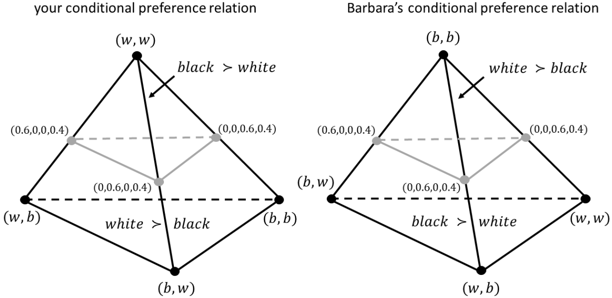
You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	(w_2, w_1)
$black$	0	0	0	5
$white$	2	2	2	2

Barbara	(b_1, b_2)	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
$black$	2	2	2	2
$white$	5	0	0	0

- We use iterated elimination of choices and expectations to find choices consistent with common belief in rationality.

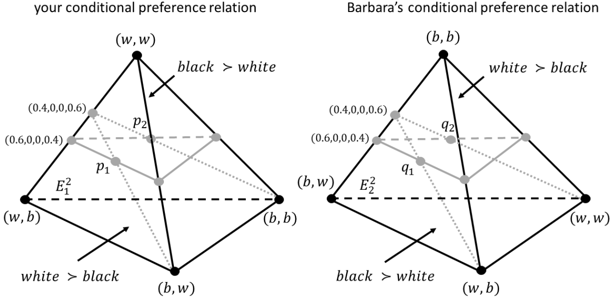
"Dinner w Str Pref f Surprise": Rationality

- As before, no choices strictly dominated.
- *black* rational for you iff $e_1(w_2, w_1) \geq 2/5$ and *white* rational for Barbara iff $e_1(b_2, b_1) \geq 2/5$.



"Dinner w Str Pref f Surprise": Belief in Rationality

- With belief in rationality, must have $e_1(w_1|w_2) \leq 3/5$. Hence, state (w_2, w_1) in your decision problem replaced by $2/5 \cdot (w_2, b_1) + 3/5 \cdot (w_2, w_1)$.
- Similarly, state (b_1, b_2) in Barbara's decision problem replaced by $2/5 \cdot (b_1, w_2) + 3/5 \cdot (b_1, b_2)$.



- As seen in the figure, no choices are eliminated at belief in rationality.

"Dinner w Str Pref f Surprise": Belief in Rationality

- Decision problems after 2-fold elimination of choices and expectations:

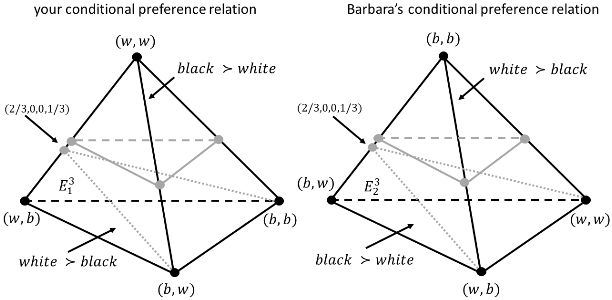
You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	$(w_2, 2/5 \cdot b_1 + 3/5 \cdot w_1)$
black	0	0	0	3
white	2	2	2	2

Barbara	$(b_1, 2/5 \cdot w_2 + 3/5 \cdot b_2)$	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	3	0	0	0

- As follows from the table, *black* rational for you under belief in rationality iff $e_1(w_2, 2/5 \cdot b_1 + 3/5 \cdot w_1) \geq 2/3$.
- Analogously, *white* rational for Barbara under belief in rationality iff $e_2(b_1, 2/5 \cdot w_2 + 3/5 \cdot b_2) \geq 2/3$.

"Dinner w Str Pref f Surp": Up to 2-Fold Bel in Rat

- With up to 2-fold belief in rationality (given new extreme state), must now have $e_1(w_1|w_2) \leq 1/3$. Hence, state $2/5 \cdot (w_2, b_1) + 3/5 \cdot (w_2, w_1)$ in your decision problem replaced by $2/3 \cdot (w_2, b_1) + 1/3 \cdot (w_2, w_1)$.
- Similarly, state $2/5 \cdot (b_1, w_2) + 3/5 \cdot (b_1, b_2)$ in Barbara's decision problem replaced by $2/3 \cdot (b_1, w_2) + 1/3 \cdot (b_1, b_2)$.



- As seen in figure, *black* eliminated for you and *white* for Barbara.

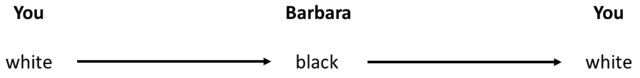
"Dinner w Str Pref f Surp": Common Belief in Rat

- Decision problems after 3-fold elimination of choices and expectations:

You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	$(w_2, 2/3 \cdot b_1 + 1/3 \cdot w_1)$
black	0	0	0	1
white	2	2	2	2

Barbara	$(b_1, 2/3 \cdot w_2 + 1/3 \cdot b_2)$	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	1	0	0	0

- With 4-fold elimination of choices and expectations, states involving w_2 are eliminated for you and states involving b_1 are eliminated for Barbara.
- Then, with 5-fold elimination of choices and expectations, state (b_2, b_1) is eliminated for you and state (w_1, w_2) is eliminated for Barbara.
- Beliefs diagram for CBR:



Example: “Dinner w Huge Preference f Surprise”

- Different from previous procedures, elimination of choices and expectations is **not** finite, even with finitely many choices for both players.
- This is seen in following variation of previous examples:

You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	(w_2, w_1)
black	0	0	0	8
white	2	2	2	2

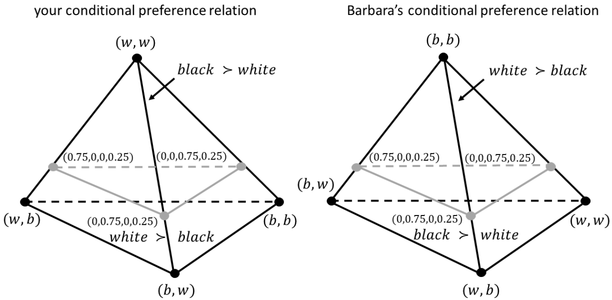
Barbara	(b_1, b_2)	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	8	0	0	0

- We use iterated elimination of choices and expectations to find choices consistent with common belief in rationality.

stephan.jagau@nottingham.ac.uk

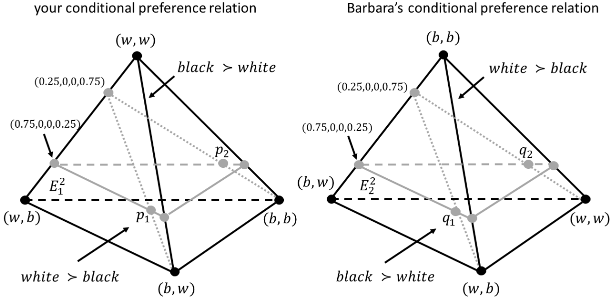
“Dinner w Huge Pref f Surprise”: Rationality

- Again, no choices strictly dominated.
- *black* rational for you iff $e_1(w_2, w_1) \geq 1/4$ and *white* rational for Barbara iff $e_1(b_2, b_1) \geq 1/4$.



"Dinner w Huge Pref f Surp": Belief in Rationality

- With belief in rationality, must have $e_1(w_1|w_2) \leq 3/4$. Hence, state (w_2, w_1) in your decision problem replaced by $1/4 \cdot (w_2, b_1) + 3/4 \cdot (w_2, w_1)$.
- Similarly, $1/4 \cdot (b_1, w_2) + 3/4 \cdot (b_1, b_2)$ replaces (b_1, b_2) for Barbara.



- As seen in figure, more expectations supporting *black* for you and *white* for Barbara survive initial restrictions.

"Dinner w Huge Pref f Surp": Common Bel in Rat

- It turns out that some beliefs supporting *black* for you and *white* for Barbara are **never** eliminated.
- To see this write $(1-e^{k-1})$ for maximum weight on $(w_2, w_1)/(b_1, b_2)$ after round $k-1$ and consider reduced decision problems at round k:

You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	$(w_2, (1-e^{k-1})\cdot w_1 + e^{k-1}\cdot b_1)$
black	0	0	0	$(1-e^{k-1})8$
white	2	2	2	2

Barbara	$(b_1, (1-e^{k-1})\cdot b_2 + e^{k-1}\cdot w_2)$	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	$(1-e^{k-1})8$	0	0	0

- New minimum weight e^k on $(w_2, w_1)/(b_1, b_2)$ solves $e^k \geq \frac{2}{8(1-e^{k-1})}$.
- $e^k \neq e^{k-1}$ for any finite k.
- Furthermore, at common belief in rationality/iterated elimination of choices and expectations, one has $e^k = e^{k-1} = 1/2$.

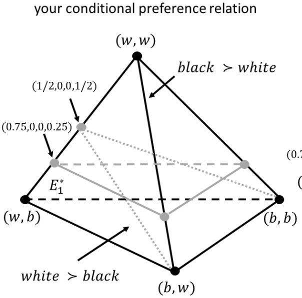
"Dinner w Huge Pref f Surp": Common Bel in Rat

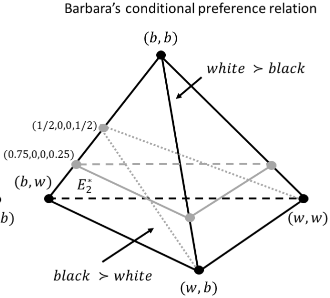
- Reduced decision problems after countably many rounds:

You	(b_2, b_1)	(b_2, w_1)	(w_2, b_1)	$(w_2, 1/2 \cdot w_1 + 1/2 \cdot b_1)$
black	0	0	0	4
white	2	2	2	2

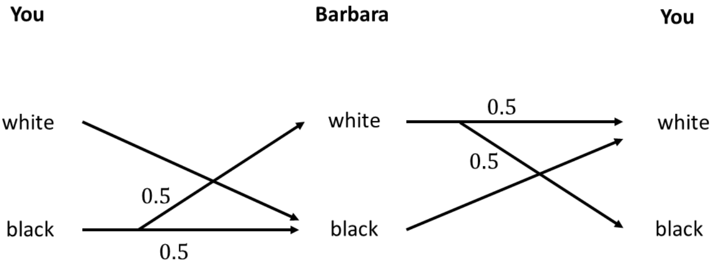
Barbara	$(b_1, 1/2 \cdot b_2 + 1/2 \cdot w_2)$	(b_1, w_2)	(w_1, b_2)	(w_1, w_2)
black	2	2	2	2
white	4	0	0	0

- Expectations consistent with CBR:





“Dinner w Huge Pref f Surp”: Beliefs Diagram



Agenda

- Psychological Games and Common Belief in Rationality
- Procedural Characterization
- Possibility
- Variants of the Procedure

Possibility of Common Belief in Rationality

- An important question is whether psychological games as defined here are always consistent with common belief in rationality.
- In other words, for any such game Γ, can we find a model M^{Γ} such that some type t_i for every i expresses common belief in rationality?
- The answer is non-obvious in view of the procedure's countable length (see previous example).

Possibility of Common Belief in Rationality

- Using that E_i^k is a convex polytope for both players i and any k, standard techniques (Cantor's intersection theorem) imply that $\bigcap_{k\geq 1} E^k$ is non-empty for both players.
- For similar reasons, any choice elimination must occur within finite steps.
- However, between two consecutive choice eliminations, the procedure may take any finite number of steps.
- **Note:**
  - General psychological games can feature both **non-existence** and eliminations after **countable** steps.
  - Linearity ensures all choice eliminations are after finite steps. Dependence of u_i on finite orders of beliefs ensures existence. Both conditions can be weakened.

Agenda

- Psychological Games and Common Belief in Rationality
- Procedural Characterization
- Possibility
- Variants of the Procedure

stephan.jagau@nottingham.ac.uk

Order Independence

- Similar to standard iterated strict dominance, iterated elimination of choices and expectations is *order-independent*.

- Intuitively, this is true for two reasons:

  1) If a choice is strictly dominated in a decision problem, it is also strictly dominated in any reduced version of that problem.
  2) If an expectation is not eliminated a some step, it can still be eliminated at a later step.

- As a consequence, we can start off eliminating strictly dominated choices and probability-one second-order expectations and then apply the full procedure to the simplified problem.

- **Caution:** Correct **intermediate** outputs (k-fold elim of chs and exps, $k \geq 1$) only found when eliminating **full-speed** in the **original order**.

States-First Procedure

The following procedure is output-equivalent to the original one:

Definition

Round 1. For both players i, eliminate all strictly dominated choices.

Round $k \geq 1$. For each player i's decision problem, eliminate all states (c_j, c_i) such that either choice has been eliminated for the respective player at the previous round. In the reduced problem, eliminate all strictly dominated choices.

Proceed until no more choices/states can be eliminated.
Subsequently *perform elimination of choices and expectations.*

Theorem

The states-first procedure always yields the same final output as iterated elimination of choices and expectations.

Example: "Exceeding Barbara's Expectations"

- *You* and *Barbara* record a song together, each practicing 1, 3, 5, or 7 weeks.
- Investing w_i weeks costs w_i^2 for both players i.
- Direct benefits of practice are given by $w_i \cdot w_j$ with own investment w_i and opponent investment w_j.
  Additionally, each of you wants to **exceed other's expectations** w_i', giving you added benefit of $(w_i - w_i')$ for $w_i > w_i'$.
- Utility functions: $u_i(w_i, (w_j, w_i')) = \begin{cases} w_i \cdot w_j - w_i^2 + (w_i - w_i'), & \text{if } w_i > w_i', \\ w_i \cdot w_j - w_i^2, & \text{otherwise.} \end{cases}$

You/Barbara	(1,1)	(1,3)	(1,5)	(1,7)	(3,1)	(3,3)	(3,5)	(3,7)	(5,1)	(5,3)	(5,5)	(5,7)	(7,1)	(7,3)	(7,5)	(7,7)
1	0	0	0	0	2	2	2	2	4	4	4	4	6	6	6	6
3	-4	-6	-6	-6	2	0	0	0	8	6	6	6	14	12	12	12
5	-16	-18	-20	-20	-6	-8	-10	-10	4	2	0	0	14	12	10	10
7	-36	-38	-40	-42	-22	-24	-26	-28	-8	-10	-12	-14	6	4	2	0

- We use states-first procedure to find choices consistent with common belief in rationality.

 stephan.jagau@nottingham.ac.uk

"Exceeding Barbara's Expectations": Rationality

You/Barbara	(1,1)	(1,3)	(1,5)	(1,7)	(3,1)	(3,3)	(3,5)	(3,7)	(5,1)	(5,3)	(5,5)	(5,7)	(7,1)	(7,3)	(7,5)	(7,7)
1	0	0	0	0	2	2	2	2	4	4	4	4	6	6	6	6
3	-4	-6	-6	-6	2	0	0	0	8	6	6	6	14	12	12	12
5	-16	-18	-20	-20	-6	-8	-10	-10	4	2	0	0	14	12	10	10
7	-36	-38	-40	-42	-22	-24	-26	-28	-8	-10	-12	-14	6	4	2	0

- 7 strictly dominated by 5 for you and Barbara.

“Exceeding Barbara’s Exp”: States-First Proc Rd 2

You/Barbara	(1,1)	(1,3)	(1,5)	(3,1)	(3,3)	(3,5)	(5,1)	(5,3)	(5,5)
1	0	0	0	2	2	2	4	4	4
3	-4	-6	-6	2	0	0	8	6	6
5	-16	-18	-20	-6	-8	-10	4	2	0

- All states of form $(7, \cdot)$ and $(\cdot, 7)$ eliminated.
- Then, 3 strictly dominates 5.

"Exceeding Barbara's Exp": States-First Proc Rd 3

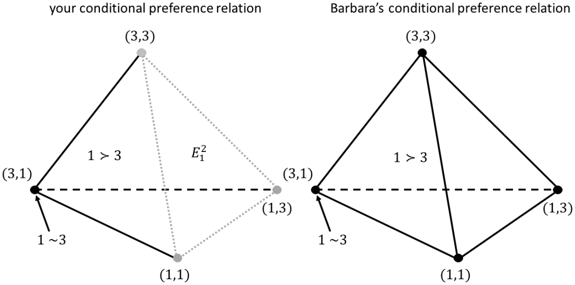
You/Barbara	(1,1)	(1,3)	(3,1)	(3,3)
1	0	0	2	2
3	-4	-6	2	0

- All states of form $(5,\cdot)$ and $(\cdot,5)$ eliminated.
- No more choices strictly dominated.

  $\Rightarrow$ Switch to elimination of choices and expectations.
- 1 *weakly* dominates 3.
- Hence, 3 is optimal iff $e_i(3,1)=1$ and 1 is optimal for any expectation.

"Exceeding Barb's Exp": States-First Proc Rd 4 ff

- Given 3-fold reduced decision problem, belief in Barbara's rationality requires that $e_1(3|3) = 1$.
- Hence, surviving states at rd 4 in $\mathrm{Conv}\{(1,1),(1,3),(3,3)\}$.



- Since state $(3,1)$ is eliminated, choice 3 is also eliminated.

  $\Rightarrow$ 1 uniquely rational under CBR for both players.

Interacting Belief Restrictions & Strict Dominance

- In “Black and White Dinner with a Twist” and other examples, standard iterated strict dominance is **insufficient** for CBR.
- This is due to interacting belief restrictions.
- E.g., in “Dinner w twist” your choosing *black* requires sufficiently high expectation of (w_2, w_1).
- But any such expectation for you goes beyond Barbara’s maximum belief in w_1 while rationally choosing w_2.
- Hence, belief in Barbara’s rationality eliminates these expectations and your choice *black*.

Interacting Belief Restrictions & Strict Dominance

- Interacting belief restrictions are the reason why iterated strict dominance does not work in psychological games.
- Conversely, special psychological games may exclude such interactions, allowing us to use strict dominance.
- In psychological games as studied here, this will be true for player i if:
  - i cares only about j's behavior and j only cares about i's first-order beliefs.
  - i cares only about j's first-order beliefs.
- In particular, iterated strict dominance works for **both** players if one player only cares about behavior and the other only cares about first-order beliefs.

Example: "Barbara's Birthday"

- *You* choose to buy a *necklace*, *ring*, or *bracelet* as a gift for *Barbara*.
- You personally prefer *necklace* over *ring* over *bracelet*. In addition, you seek to surprise Barbara with your gift. Meanwhile, Barbara seeks to guess which gift you bought her.

You	$(\cdot, n)$	$(\cdot, r)$	$(\cdot, b)$
necklace	0	3	3
ring	2	0	2
bracelet	1	1	0

Barbara	$(n, \cdot)$	$(r, \cdot)$	$(b, \cdot)$
necklace	1	0	0
ring	0	1	0
bracelet	0	0	1

- Your behavior matters for Barbara but not vice versa. Similarly, you care what Barbara expect you to do but not vice versa.
- Hence, no belief restrictions for you and Barbara interact in this game.

  $\Rightarrow$ Iterated strict dominance finds choices consistent with CBR.

"Barbara's Birthday": Rationality

You	$(\cdot, n)$	$(\cdot, r)$	$(\cdot, b)$
necklace	0	3	3
ring	2	0	2
bracelet	1	1	0

Barbara	$(n, \cdot)$	$(r, \cdot)$	$(b, \cdot)$
necklace	1	0	0
ring	0	1	0
bracelet	0	0	1

- *bracelet* strictly dominated for you by (e.g.) $0.4 \cdot \mathit{necklace} + 0.6 \cdot \mathit{ring}$.
- No choice dominated for Barbara.

"Barbara's Birthday": Belief in Rationality

You	$(\cdot, n)$	$(\cdot, r)$	$(\cdot, b)$
necklace	0	3	3
ring	2	0	2

Barbara	$(n, \cdot)$	$(r, \cdot)$
necklace	1	0
ring	0	1
bracelet	0	0

- Under belief in rationality, Barbara discards all states of form $(b, \cdot)$.
- Then, *bracelet* strictly dominated by (e.g.) $0.5 \cdot necklace + 0.5 \cdot ring$.
- No choice or state eliminated for you.
  Caution: $(\cdot, b)$ eliminated for you at up to 2-fold belief in rationality!

stephan.jagau@nottingham.ac.uk

"Barbara's Bday": Common Belief in Rationality

You	$(\cdot, n)$	$(\cdot, r)$
necklace	0	3
ring	2	0

Barbara	$(n, \cdot)$	$(r, \cdot)$
necklace	1	0
ring	0	1

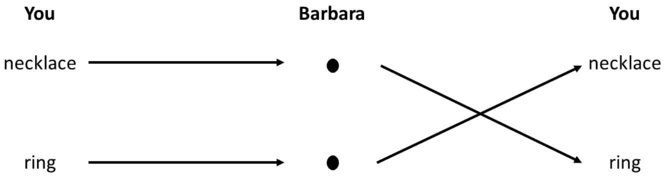
- Under up to 2-fold belief in rationality, you discard $(\cdot, b)$ as well as (b, n) and (b, r).
- Finally, under up to 3-fold belief in rationality, Barbara discards (n, b) and (r, b).
- No further choices are eliminated, so the procedure stops.
- Reduced decision problems:

You	(n, n)	(n, r)	(r, n)	(r, r)
necklace	0	3	0	3
ring	2	0	2	0

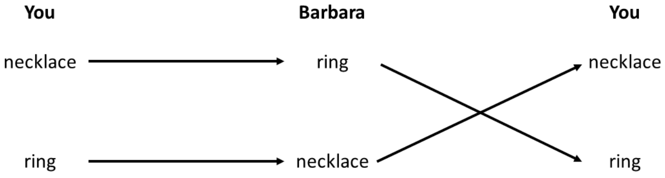
Barbara	(n, n)	(n, r)	(r, n)	(r, r)
necklace	1	1	0	0
ring	0	0	1	1

"Barbara's Birthday": Beliefs Diagram

- To support your choices, only need **partial** beliefs diagram, omitting beliefs about Barbara's behavior:



- Now complete diagram to also support Barbara's choices:



stephan.jagau@nottingham.ac.uk