EPICENTER Spring Course on Epistemic Game Theory
Advanced Topics III + IV: Psychological Games

Andrés Perea

Maastricht University

July 9, 2019

Psychological Games

- In "traditional" games, your utility only depends on your choice and your belief about the opponents' choices.
- In some situations, your utility may also depend on what you believe that the opponent believes that you do.
- For instance, surprise, meeting somebody's expectations, disappointment, guilt, and other emotions.
- Such situations can be modelled as psychological games.
- Psychological games were introduced by Geanakoplos, Pearce and Stacchetti (1989), and later extended to dynamic games by Battigalli and Dufwenberg (2009).

Example: Surprising Barbara

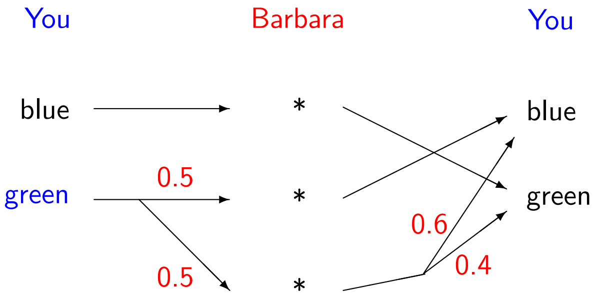
blue	green	red	yellow	no surprise
4	3	2	1	0

Story

- You are invited to a party together with Barbara, and you still have the same preferences over colors as before.
- But now your objective is to surprise Barbara by the color you wear.
- Let Prob(you not blue) be the expected probability you think that Barbara assigns to you not choosing blue.
- Your expected utility from choosing blue is

$$u_1(\text{blue}) = 4 \cdot \text{Prob}(\text{you not blue}).$$

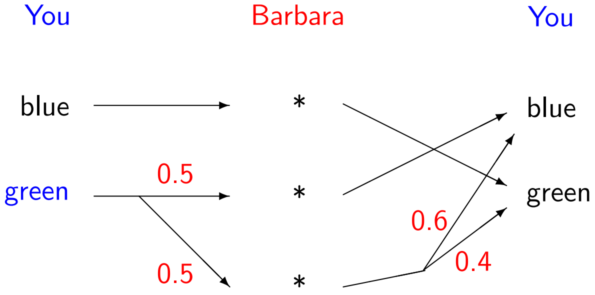
- Similarly for the other colors.



- Expected utility: $u_1(\text{blue}) = 4 \cdot \text{Prob}(\text{you not blue})$.
- Consider the belief hierarchy that starts at your choice green:

$$\text{Prob}(\text{you not blue}) = (0.5) \cdot 0 + (0.5) \cdot 0.4 = 0.2.$$

- Hence, $u_1(blue) = 4 \cdot (0.2) = 0.8$.



- Expected utility: $u_1(\text{green}) = 3 \cdot \text{Prob}(\text{you not green})$.
- Consider the belief hierarchy that starts at your choice green:

$$\text{Prob}(\text{you not green}) = (0.5) \cdot 1 + (0.5) \cdot 0.6 = 0.8.$$

- Hence, $u_1(green) = 3 \cdot (0.8) = 2.4$.
- Therefore, green is optimal for your belief hierarchy that starts at green.

blue	green	red	yellow	no surprise
4	3	2	1	0

- Expected utility: $u_1(\text{blue}) = 4 \cdot \text{Prob}(\text{you not blue})$.
- Your utility function can be represented by the following matrix:

	extreme second-order beliefs			
	blue	green	red	yellow
blue	0	4	4	4
green	3	0	3	3
red	2	2	0	2
yellow	1	1	1	0

- Your utility depends on your second-order belief.
- This cannot be modelled by traditional game theory.

	extreme second-order beliefs			
	blue	green	red	yellow
blue	0	4	4	4
green	3	0	3	3
red	2	2	0	2
yellow	1	1	1	0

- Matrix representation is possible because:
- your utility only depends on finitely many orders of belief: belief finite (Jagau and Perea (2017)),
- your utility only depends on the expected probability you think that Barbara assigns to each of your choices,
- your utility depends linearly on this expected probability.
- Last two conditions: expectation-based game (Jagau and Perea (2017)).

- In general, in a psychological game, your utility from making a choice may depend on your second-order belief, ...
- or your third-order belief, ...
- or even higher-order beliefs.
- That is, your utility may potentially depend on your full belief hierarchy.

Definition (Psychological game)

A psychological game specifies for every player i

a finite set of choices C_i,

a utility function u_i that assigns to every choice c_i and every belief hierarchy b_i some utility $u_i(c_i, b_i)$.

- Taken from Jagau and Perea (2017).
- Similar definitions can be found in Geanakoplos, Pearce and Stacchetti (1989) and Battigalli and Dufwenberg (2009).

Common Belief in Rationality

Definition (Psychological game)

A psychological game specifies for every player i

a finite set of choices C_i,

a utility function u_i that assigns to every choice c_i and every belief hierarchy b_i some utility $u_i(c_i, b_i)$.

- We have seen that belief hierarchies can be encoded by means of types in an epistemic model.
- Hence, we can write $u_i(c_i, t_i)$ instead of $u_i(c_i, b_i)$.
- We say that a choice c_i is optimal for a type t_i if

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

for all alternative choices $c_i' \in C_i$.

- Type t_i believes in the opponents' rationality if $b_i(t_i)$ only assigns positive probability to opponents' choice-type pairs (c_j, t_j) where c_j is optimal for t_j.

Definition (Common belief in rationality)

Type t_i expresses 1-fold belief in rationality if t_i believes in the opponents' rationality.

Type t_i expresses 2-fold belief in rationality if t_i only assigns positive probability to opponents' types that express 1-fold belief in rationality.

Type t_i expresses 3-fold belief in rationality if t_i only assigns positive probability to opponents' types that express 2-fold belief in rationality.

And so on.

Type t_i expresses common belief in rationality if t_i expresses k-fold belief in rationality for all k.

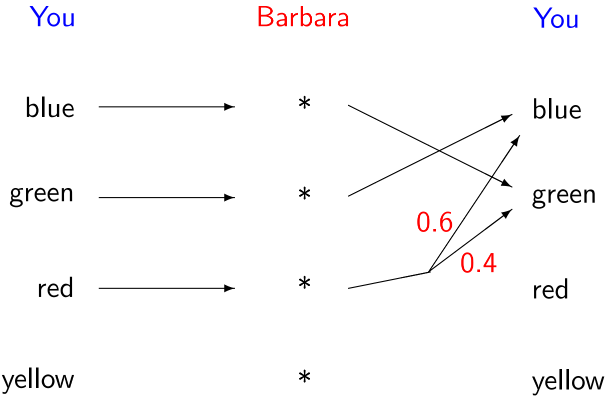
- Based on Jagau and Perea (2017). Similar definitions in Bjorndahl, Halpern and Pass (2017) and Battigalli and Dufwenberg (2009).

Example: Surprising Barbara

blue	green	red	yellow	no surprise
4	3	2	1	0

- Expected utility: $u_1(\text{blue}) = 4 \cdot \text{Prob}(\text{you not blue})$.
- Which colors can you rationally choose under common belief in rationality?
- Your choice yellow is not optimal for any belief hierarchy:
- If Prob(you blue) ≥ 0.5, then $u_1(\text{green}) \geq (0.5) \cdot 3 = 1.5 > 1$.
- If Prob(you blue) ≤ 0.5, then $u_1(\text{blue}) \geq (0.5) \cdot 4 = 2 > 1$.
- What about the other colors?

blue	green	red	yellow	no surprise
4	3	2	1	0



- All belief hierarchies for you express common belief in rationality.
- You can rationally choose blue, green and red under common belief in rationality.

Surprise Exam Paradox

- You announce to Barbara that you will give her a surprise exam on epistemic game theory some day next week.
- You would like to surprise Barbara with the day of the exam.
- Common sense reasoning tells you that this is not possible:
- If you have not given the exam by Thursday evening, Barbara knows that you must give it on Friday, and hence it cannot be a surprise.
- If you have not given the exam by Wednesday evening, Barbara knows that you must give it on Thursday, and hence it cannot be a surprise either.
- And so on.
- Hence, you could only give it on Monday.
- But Barbara expects this, and hence this cannot be a surprise either.
- However, you could then surprise Barbara by giving the exam on Wednesday.
- Surprise exam paradox.

- We model the surprise exam paradox as a psychological game.

- For simplicity, we assume that you can only put the exam on Monday or Tuesday.

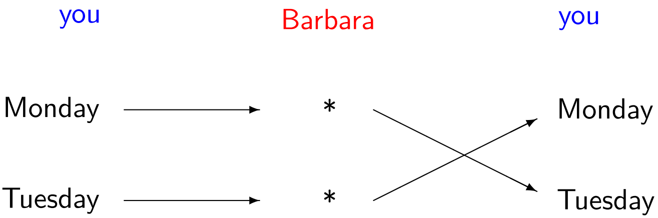
- Your utility function is given by

	extreme second-order beliefs Monday	extreme second-order beliefs Tuesday
Monday	0	1
Tuesday	0	0

- Is it possible for you to surprise Barbara under common belief in rationality?

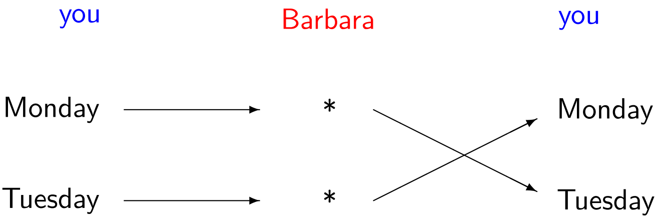
- Subsequent analysis based on Mourmans (2017) and Geanakoplos (1996).

	second-order beliefs	
	Monday	Tuesday
Monday	0	1
Tuesday	0	0



- Your belief hierarchy that starts at your choice Monday expresses common belief in rationality.
- Hence, you can believe to fully surprise Barbara under common belief in rationality.

	second-order beliefs	
	Monday	Tuesday
Monday	0	1
Tuesday	0	0



- Your belief hierarchy that starts at your choice Monday is not simple.
- A simple belief hierarchy that expresses common belief in rationality is called a psychological Nash equilibrium.
- Concept of psychological Nash equilibrium is due to Geanakoplos, Pearce and Stacchetti (1989).

	second-order beliefs	
	Monday	Tuesday
Monday	0	1
Tuesday	0	0

- We show: In a psychological Nash equilibrium, you believe you cannot surprise Barbara.
- Proof: Take a psychological Nash equilibrium generated by a probability distribution $\sigma_1 \in \Delta(\{\text{Mon, Tue}\})$.
- Suppose that $\sigma_1(\text{Tue}) > 0$.
- Then, you believe that Barbara believes that you believe that Barbara believes that you, with positive probability, will put the exam on Tuesday.
- Hence, you must believe that Barbara believes that you will put the exam on Monday.
- Therefore, $\sigma_1(\text{Tue}) = 0$, which is a contradiction.
- Conclusion: In a psychological Nash equilibrium, $\sigma_1(\text{Mon}) = 1$.

Common Belief in Rationality May be Impossible

- For standard games, where the utility only depends (linearly) on your first-order belief, and every player has finitely many choices, we have seen that common belief in rationality is always possible.
- Is this also true for psychological games?
- Not if the utility depends on all orders of belief.

Modified Surprise Exam Paradox

- Suppose you prefer to put the exam on Monday if ...
- you believe that Barbara assigns a positive probability to Tuesday, or
- you believe that Barbara believes that you believe that Barbara assigns a positive probability to Tuesday,
- and so on.
- However, if you believe that Barbara assigns probability 1 to Monday, and you believe that Barbara believes that you believe that Barbara assigns probability 1 to Monday, and so on, then you prefer to slightly surprise Barbara by putting the exam on Tuesday.
- Then, your utilities are given by

	belief hierarchies	
	b_1^{Mon}	other
Monday	0	1
Tuesday	1	0

where b_1^{Mon} is the simple belief hierarchy generated by $\sigma_1 = \text{Mon}$.

	belief hierarchies	
	b_1^{Mon}	other
Monday	0	1
Tuesday	1	0

- We show: There is no belief hierarchy b_1 that expresses common belief in rationality.
- Based on Jagau and Perea (2017). Similar example in Bjorndahl, Halpern and Pass (2017).
- Proof: Step 1: Show that belief hierarchy b_1^{Mon} does not express common belief in rationality.
- In b_1^{Mon}, you believe that Barbara believes that you choose Monday and that you have belief hierarchy b_1^{Mon}.
- But then, you believe that Barbara believes that you choose irrationally.

	belief hierarchies	
	b_1^{Mon}	other
Monday	0	1
Tuesday	1	0

- Step 2: Show that there is no belief hierarchy b_1 that expresses common belief in rationality.
- Suppose that b_1 expresses common belief in rationality.
- Then, you believe in b_1 that Barbara believes that you choose rationally while holding a belief hierarchy different from b_1^{Mon}.
- Hence, in b_1 you believe that Barbara believes that you choose Monday.

	belief hierarchies	
	b_1^{Mon}	other
Monday	0	1
Tuesday	1	0

- Suppose that b_1 expresses common belief in rationality.
- Then, you believe in b_1 that Barbara believes that you believe that Barbara believes that you choose rationally while holding a belief hierarchy different from b_1^{Mon}.
- Hence, you believe in b_1 that Barbara believes that you believe that Barbara believes that you choose Monday.
- And so on.
- Hence, $b_1 = b_1^{\text{Mon}}$. Contradiction.

When is Common Belief in Rationality Possible?

- We have seen a psychological game in which common belief in rationality is impossible.
- In that game, your utility depended on all levels of your belief hierarchy.
- What if the utility of all players only depends on finitely many belief levels?
- Belief finite games: See Jagau and Perea (2017).
- For such games, we can show that common belief in rationality is always possible.

Example: Surprising Barbara and Being Different

	blue	green	red	yellow	same color	no surprise
you	4	3	2	1	0	0
Barbara	2	1	4	3	0	0

Story

- You would like to surprise Barbara, but also to wear a different color than she does. Similarly for Barbara.
- Let Prob(Barbara not blue) be the probability you assign to Barbara not choosing blue.
- Let Prob(you not blue) be the expected probability you think that Barbara assigns to you not choosing blue.
- Your expected utility from choosing blue is

$$u_1(blue) = \underbrace{4 \cdot \text{Prob}(\text{Barbara not blue})}_{\text{being different / first-order belief}} + \underbrace{4 \cdot \text{Prob}(\text{you not blue})}_{\text{surprise / second-order belief}}.$$

	blue	green	red	yellow	same color	no suprise
you	4	3	2	1	0	0
Barbara	2	1	4	3	0	0

- Your expected utility from choosing blue is

$$u_1(blue) = \underbrace{4 \cdot \text{Prob}(\text{Barbara not blue})}_{\text{being different / first-order belief}} + \underbrace{4 \cdot \text{Prob}(\text{you not blue})}_{\text{surprise / second-order belief}}.$$

- Similarly for the other colors.

- Your utilities can be represented as follows:

first-order beliefs (being different)

You	b_2	g_2	r_2	y_2
b_1	0	4	4	4
g_1	3	0	3	3
r_1	2	2	0	2
y_1	1	1	1	0

being different

+

second-order beliefs

You	b_2	g_2	r_2	y_2
b_1	0	4	4	4
g_1	3	0	3	3
r_1	2	2	0	2
y_1	1	1	1	0

surprise

You	first-order beliefs b_2	g_2	r_2	y_2
b_1	0	4	4	4
g_1	3	0	3	3
r_1	2	2	0	2
y_1	1	1	1	0

being different

+

You	second-order beliefs b_1	g_1	r_1	y_1
b_1	0	4	4	4
g_1	3	0	3	3
r_1	2	2	0	2
y_1	1	1	1	0

surprise

$b_2 \; b_1 \; b_2 \; \ldots$

Barbara	first-order beliefs b_1	g_1	r_1	y_1
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

being different

+

Barbara	second-order beliefs b_2	g_2	r_2	y_2
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

surprise

g_1 b_2 b_1 b_2 ...

You	first-order beliefs b_2	g_2	r_2	y_2
b_1	0	4	4	4
g_1	3	0	3	3
r_1	2	2	0	2
y_1	1	1	1	0

being different

+

You	second-order beliefs b_1	g_1	r_1	y_1
b_1	0	4	4	4
g_1	3	0	3	3
r_1	2	2	0	2
y_1	1	1	1	0

surprise

r_2 g_1 b_2 b_1 b_2 ...

Barbara	first-order beliefs b_1	g_1	r_1	y_1
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

being different

+

Barbara	second-order beliefs b_2	g_2	r_2	y_2
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

surprise

$b_1 \; r_2 \; g_1 \; b_2 \; b_1 \; b_2 \; \ldots$

You	first-order beliefs b_2	g_2	r_2	y_2
b_1	0	4	4	4
g_1	3	0	3	3
r_1	2	2	0	2
y_1	1	1	1	0

being different

+

You	second-order beliefs b_1	g_1	r_1	y_1
b_1	0	4	4	4
g_1	3	0	3	3
r_1	2	2	0	2
y_1	1	1	1	0

surprise

$y_2 \;\; b_1 \;\; r_2 \;\; g_1 \;\; b_2 \;\; b_1 \;\; b_2 \;\; \ldots$

Barbara	first-order beliefs b_1	g_1	r_1	y_1
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

being different

+

Barbara	second-order beliefs b_2	g_2	r_2	y_2
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

surprise

$g_1\ y_2\ b_1\ r_2\ g_1\ b_2\ b_1\ b_2\ \ldots$

Barbara	first-order beliefs b_1	g_1	r_1	y_1
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

being different

+

Barbara	second-order beliefs b_2	g_2	r_2	y_2
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

surprise

$r_2 \ g_1 \ y_2 \ b_1 \ r_2 \ g_1 \ b_2 \ b_1 \ b_2 \ \ldots$

Barbara	first-order beliefs			
	b_1	g_1	r_1	y_1
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

being different

$+$

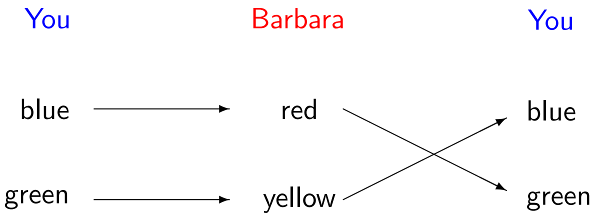
Barbara	second-order beliefs			
	b_2	g_2	r_2	y_2
b_2	0	2	2	2
g_2	1	0	1	1
r_2	4	4	0	4
y_2	3	3	3	0

surprise

$$\underbrace{r_2 \; g_1 \; y_2 \; b_1}_{\text{cycle}} \; r_2 \; g_1 \; b_2 \; b_1 \; b_2 \; \ldots$$

$\underbrace{r_2\ g_1\ y_2\ b_1}_{\text{cycle}}\ r_2\ g_1\ b_2\ b_1\ b_2\ \ldots$

- This cycle can be translated into a beliefs diagram where all belief hierarchies express common belief in rationality:



- This construction can be generalized to all psychological games where the players' utilities only depend on finitely many belief levels.

Theorem

Consider a psychological game with finitely many choices such that either

(a) all utility functions only depend on finitely many orders of belief, or

(b) all utility functions are continuous in the belief hierarchy.

Then, there is for every player i a belief hierarchy b_i that expresses common belief in rationality.

- Follows from Jagau and Perea (2017). They present a weaker condition that guarantees the possibility of common belief in rationality: preservation of rationality at infinity.
- In case (b), we can even find a psychological Nash equilibrium for every player. Follows from Geanakoplos, Pearce and Stacchetti (1989).

Theorem

Consider a psychological game with finitely many choices such that either

(a) all utility functions only depend on finitely many orders of belief, or

(b) all utility functions are continuous in the belief hierarchy.

Then, there is for every player i a belief hierarchy b_i that expresses common belief in rationality.

- Common belief in rationality may be impossible if (a) and (b) fail:

	belief hierarchies	
	b_1^{Mon}	other
Monday	0	1
Tuesday	1	0

Elimination of Choices

- For standard games, the choices that are possible under common belief in rationality can be characterized by the iterated elimination of strictly dominated choices.
- In every round, we proceed by elimination of choices only.
- We will see that elimination of choices only is no longer enough for psychological games.
- Hence, we must turn to procedures where we eliminate choices and beliefs in every round.

Example: Stay or Go?

Story

- Once again, you and Barbara are invited for a party tonight. You both must decide whether to go or not.
- Barbara only wants to join the party if you do. Otherwise, she prefers to stay at home.
- You only consider joining the party if you believe to surprise Barbara at the party by your presence.
- Possible utilities:

You	first-order beliefs: go	first-order beliefs: stay
go	1	0
stay	1	1

\+

You	second-order beliefs: go	second-order beliefs: stay
go	0	1
stay	1	1

Barbara	first-order beliefs: go	first-order beliefs: stay
go	1	0
stay	0	1

You	first-order beliefs go	stay
go	1	0
stay	1	1

+

You	second-order beliefs go	stay
go	0	1
stay	1	1

Barbara	first-order beliefs go	stay
go	1	0
stay	0	1

- For you, going to the party is only optimal if you believe that Barbara goes, and believe that Barbara believes that you stay at home.
- But then, you must necessarily believe that Barbara chooses irrationally.
- Therefore, going to the party cannot be optimal for you under common belief in rationality.
- However, every choice in this game is optimal for at least one belief hierarchy.
- Hence, elimination of choices alone is not enough to eliminate your choice go.

When is Elimination of Choices Enough?

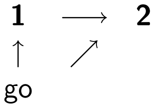
You	first-order beliefs go	first-order beliefs stay
go	1	0
stay	1	1

+

You	second-order beliefs go	second-order beliefs stay
go	0	1
stay	1	1

Barbara	first-order beliefs go	first-order beliefs stay
go	1	0
stay	0	1

- Problem: To make go optimal for you while believing in Barbara's rationality, we must impose conflicting restrictions on your second-order belief.



causality diagram

- In the causality diagram, the paths $(go, 2)$ and $(go, 1, 2)$ have an overlap at order 2.

Causality Diagrams

- The causality diagram for player 1 shows which orders of belief are directly, or indirectly, relevant for player 1 if he chooses rationally under common belief in rationality.
- Suppose, player 1 cares about 1st and 3rd order belief, whereas player 2 only cares about 2nd order belief.

$$\begin{array}{ccc} \mathbf{1} & \longrightarrow & \mathbf{3} \\ \uparrow & \nearrow & \\ c_1 & & \end{array}$$

causality diagram for player 1

overlap at order 3

- Suppose, player 1 only cares about 2nd order belief, but player 2 cares about 1st and 3rd order belief.

$$c_1 \longrightarrow \mathbf{2} \longrightarrow \mathbf{4} \longrightarrow \mathbf{6} \ \ldots$$ causality diagram for player 1

no overlapping paths

- From now on, we focus on psychological games with two players where

  (a) the utility of both players only depends on finitely many orders of belief,

  (b) the utility of both players only depends on a summary statistic of a belief hierarchy, called higher-order expectations (Jagau and Perea (2019)), and

  (c) the utility of both players depends linearly on the belief hierarchy.

- Such games are called two-player, belief-finite, expectation-based games. (Jagau and Perea (2017)).

- Look at all such games where player 1's utility depends on the orders of belief in N_1, and player 2's utility depends on the orders of belief in N_2.

- Look at all such games where player 1's utility depends on the orders of belief in N_1, and player 2's utility depends on the orders of belief in N_2.

- Mourmans (2019) has shown the following:

- Iterated elimination of strictly dominated choices characterizes, for all utility functions consistent with (N_1, N_2), the rational choices for player 1 under common belief in rationality, if and only if, the causality diagram for player 1 has no overlapping paths.

- The causality diagram for player 1 has no overlapping paths, if and only if,

  (a) both players only care about a single order of belief, or

  (b) player 1 only cares about a single, even order of belief, or

  (c) player 2 cares about a single even order of belief a, and player 1 cares about several odd orders of belief, such that for every two $b, c \in N_1$, the difference $c - b$ is not a multiple of a.

Theorem (When elimination of choices is enough)

Consider all two-player games above where player 1's utility depends on the orders of belief in N_1, and player 2's utility depends on the orders of belief in N_2.

Then, iterated elimination of strictly dominated choices characterizes, for all utility functions consistent with (N_1, N_2), the rational choices for player 1 under common belief in rationality, if and only if,

(a) both players only care about a single order of belief, or

(b) player 1 only cares about a single, even order of belief, or

(c) player 2 cares about a single even order of belief a, and player 1 cares about several odd orders of belief, such that for every two $b, c \in N_1$, the difference $c - b$ is not a multiple of a.

- Based on Mourmans (2019).

What if Elimination of Choices Does not Work?

- Suppose that player 1's utility depends on orders of belief in N_1, and player 2's utility depends on orders of belief in N_2.
- What if (a), (b) and (c) in the previous theorem are all violated?
- Then, iterated elimination of strictly dominated choices may not be enough to find the choices that player 1 can rationally make under common belief in rationality.
- In that case, we must turn to a procedure that eliminates combinations of choices and beliefs.

- Suppose the utilities of both players only depend on first- and second-order beliefs.
- Idea: Recursively eliminate pairs of choices and first-order beliefs.
- Formally, a second-order belief b_i^2 for player i is a probability distribution on j's choices and first-order beliefs.
- Hence, a second-order belief b_i^2 induces a first-order belief b_i^1.
- The utility for player i can be written as $u_i(c_i, b_i^2)$.

Algorithm (Iterated Elimination of Choices and First-Order Beliefs)

Suppose the utilities of both players only depend on the first-order and second-order beliefs.

Step 1. *For every player i, keep the choice-belief pairs in*

$$R_i^1 = \{(c_i, b_i^1) \mid c_i \text{ optimal for some } b_i^2 \text{ that induces } b_i^1\}.$$

Step 2. *For every player i, keep the choice-belief pairs in*

$$R_i^2 = \{(c_i, b_i^1) \mid c_i \text{ optimal for some } b_i^2 \in \Delta(R_j^1) \text{ that induces } b_i^1\}.$$

Step 3. *For every player i, keep the choice-belief pairs in*

$$R_i^3 = \{(c_i, b_i^1) \mid c_i \text{ optimal for some } b_i^2 \in \Delta(R_j^2) \text{ that induces } b_i^1\}.$$

And so on.

- Algorithm taken from Jagau and Perea (2017).

Theorem (Algorithm works)

Suppose the utilities of both players only depend on the first-order and second-order beliefs.

(1) The choice-belief pairs (c_i, b_i^1) that are possible for player i if he expresses up to k-fold belief in rationality are exactly the choice-belief pairs that survive $k+1$ rounds of iterated elimination of choices and first-order beliefs.

(2) The choice-belief pairs (c_i, b_i^1) that are possible for player i if he expresses common belief in rationality are exactly the choice-belief pairs that survive all rounds of iterated elimination of choices and first-order beliefs.

- Based on Jagau and Perea (2017).
- The algorithm is not guaranteed to terminate within finitely many rounds.

You	first-order beliefs go	stay
go	1	0
stay	1	1

+

You	second-order beliefs go	stay
go	0	1
stay	1	1

Barbara	first-order beliefs go	stay
go	1	0
stay	0	1

- **Step 1.** $R_1^1 = \{(\text{go}, \text{go})\} \cup \{(\text{stay}, b_1^1 | \ b_1^1 \text{ arbitrary}\}$.
- $R_2^1 = \{(\text{go}, b_2^1) \mid b_2^1(\text{go}) \geq \frac{1}{2}\} \cup \{(\text{stay}, b_2^1) \mid b_2^1(\text{stay}) \geq \frac{1}{2}\}$.
- **Step 2.** $R_1^2 = \{(\text{stay}, b_1^1 | \ b_1^1 \text{ arbitrary}\}$.
- $R_2^2 = \{(\text{go}, b_2^1) \mid b_2^1(\text{go}) \geq \frac{1}{2}\} \cup \{(\text{stay}, b_2^1) \mid b_2^1(\text{stay}) \geq \frac{1}{2}\}$.
- **Step 3.** $R_1^3 = \{(\text{stay}, b_1^1 | \ b_1^1 \text{ arbitrary}\}$.
- $R_2^3 = \{(\text{stay}, \text{stay})\}$.
- **Step 4.** $R_1^4 = \{(\text{stay}, \text{stay})\}$.
- $R_2^4 = \{(\text{stay}, \text{stay})\}$.

- Battigalli, P. and M. Dufwenberg (2009), Dynamic psychological games, *Journal of Economic Theory* **144,** 1–35.
- Bjorndahl, A., Halpern, J.Y. and R. Pass (2017), Language-based games, *Working paper.*
- Geanakoplos, J., Pearce, D. and E. Stacchetti (1989), Psychological games and sequential rationality, *Games and Economic Behavior* **1,** 60–79.
- Geanakoplos, J. (1996), The Hangman's Paradox and Newcomb's Paradox as psychological games, *Working paper.*

- Jagau, S. and A. Perea (2017), Common belief in rationality in psychological games, *EPICENTER Working Paper No. 10.*
- Jagau, S. and A. Perea (2019), Psychological expected utility, *In progress.*
- Mourmans, N. (2017), Reasoning about the surprise exam paradox: An application of psychological game theory, *EPICENTER Working Paper No. 12.*
- Mourmans, N. (2019), Reasoning in psychological games: When is iterated elimination of choices enough?, *EPICENTER Working Paper No. 20.*