EPICENTER Spring Course on Epistemic Game Theory
Advanced Topics II: Psychological Games

Andrés Perea

EPICENTER
Research Center for
Epistemic Game Theory

Maastricht University

June 2017

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, you want to surprise Barbara by the color you wear.
- Such situations can be modelled as psychological games.

Example: Surprising Barbara

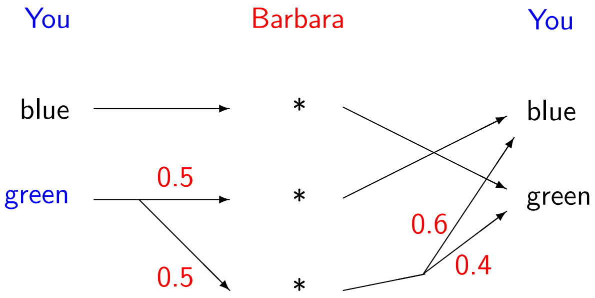
blue	green	red	yellow	no surprise
4	3	2	1	0

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

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 as follows:

	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.

- In general, in a psychological game, your expected 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 expected 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 $u_i(c_i, b_i)$.

- Definition based on 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.

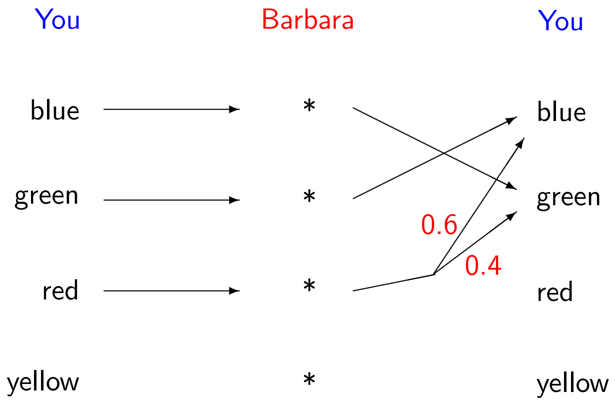
- Based on Jagau and Perea (2017). Similar definitions in Bjorndahl, Halpern and Pass (2017) and Battigalli and Dufwenberg (2009).

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 $\text{Prob}(\text{you blue}) \geq 0.5$, then $u_1(\text{green}) \geq (0.5) \cdot 3 = 1.5 > 1$.
- If $\text{Prob}(\text{you blue}) \leq 0.5$, 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.

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 above all 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

$$u_1(blue) = 3 \cdot (\underbrace{4 \cdot \text{Prob}(\text{Barbara not blue})}_{\text{being different / first-order belief}}) + 1 \cdot (\underbrace{4 \cdot \text{Prob}(\text{you not blue})}_{\text{surprise / second-order belief}}).$$

	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

$$u_1(blue) = 3 \cdot (\underbrace{4 \cdot \text{Prob}(\text{Barbara not blue})}_{\text{being different / first-order belief}}) + 1 \cdot (\underbrace{4 \cdot \text{Prob}(\text{you not blue})}_{\text{surprise / second-order belief}}).$$

- Similarly for the other colors.
- Your utilities can be represented as follows:

You	first-order beliefs: blue	green	red	yellow
blue	0	12	12	12
green	9	0	9	9
red	6	6	0	6
yellow	3	3	3	0

being different

+

You	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

surprise

	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

$$u_1(blue) = 3 \cdot (\underbrace{4 \cdot \text{Prob}(\text{Barbara not blue})}_{\text{being different / first-order belief}}) + 1 \cdot (\underbrace{4 \cdot \text{Prob}(\text{you not blue})}_{\text{surprise / second-order belief}}).$$

- Which colors can you rationally choose under common belief in rationality?
- Color yellow is not optimal for you for any belief hierarchy:
- If Prob(Barbara blue) ≥ 0.5, then $u_1(\text{green}) \geq 3 \cdot (0.5) \cdot 3 > 4$.
- If Prob(Barbara blue) ≤ 0.5, then $u_1(\text{blue}) \geq 3 \cdot (0.5) \cdot 4 > 4$.
- Similarly, green is not optimal for Barbara for any belief hierarchy.

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

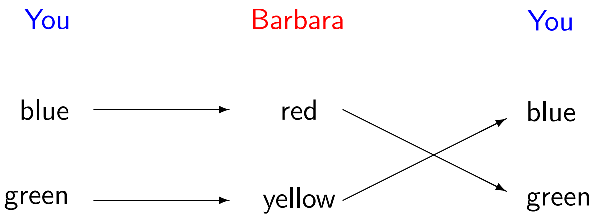
- Your expected utility from choosing blue is

$$u_1(blue) = 3 \cdot (\underbrace{4 \cdot \text{Prob}(\text{Barbara not blue})}_{\text{being different / first-order belief}}) + 1 \cdot (\underbrace{4 \cdot \text{Prob}(\text{you not blue})}_{\text{surprise / second-order belief}}).$$

- If you believe in Barbara's rationality, you believe that Barbara will not choose green.
- Then, $u_1(\text{green}) \geq 3 \cdot 3 = 9$, whereas $u_1(\text{red}) \leq 8$.
- Hence, choosing red can no longer be optimal for you.
- Similarly, choosing blue can no longer be optimal for Barbara if she believes in your rationality.

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

$$u_1(\textit{blue}) = 3 \cdot (\underbrace{4 \cdot \text{Prob}(\text{Barbara not blue})}_{\text{being different / first-order belief}}) + 1 \cdot (\underbrace{4 \cdot \text{Prob}(\text{you not blue})}_{\text{surprise / second-order belief}}).$$



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

Surprise Exam Paradox

Story

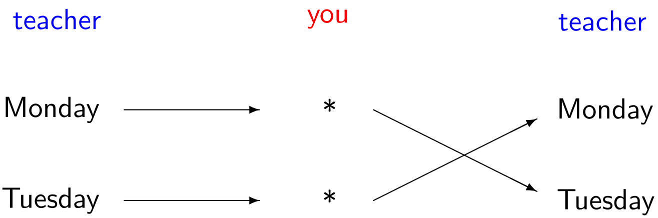
- It is Friday, and your biology teacher tells you (again) that he will give you a surprise exam some day next week.
- He would like to surprise you with the day of the exam.
- Common sense reasoning tells you that this is not possible:
- If he has not given the exam by Thursday evening, you know that he must give it on Friday, and hence it cannot be a surprise.
- If he has not given the exam by Wednesday evening, you know that he must give it on Thursday, and hence it cannot be a surprise either.
- And so on.
- Hence, the teacher could only give it on Monday.
- But you expect this, and hence this cannot be a surprise either.
- However, the teacher could then surprise you by giving the exam on Wednesday.
- Surprise exam paradox.

- We model the surprise exam paradox as a psychological game.
- For simplicity, we assume that the teacher can only put the exam on Monday or Tuesday.
- The teacher's utility function is given by

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

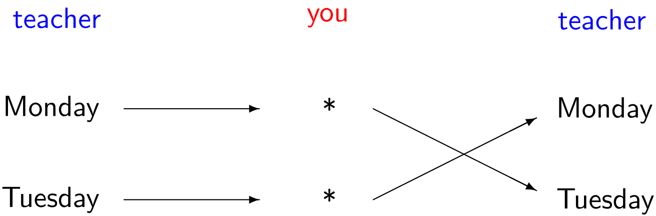
- Is it possible for the teacher to surprise you under common belief in rationality?
- Subsequent analysis based on Mourmans (2017) and Geanakoplos (1996).

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



- The teacher's belief hierarchy that starts at his choice Monday expresses common belief in rationality.
- Hence, the teacher can believe to surprise you under common belief in rationality.

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



- The teacher's belief hierarchy that starts at his 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, the teacher believes that he cannot surprise you.
- 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, the teacher believes that you believe that the teacher believes that you believe that the teacher, with positive probability, puts the exam on Tuesday.
- Hence, the teacher must believe that you believe that the teacher will put the exam on Monday.
- Therefore, $\sigma_1(\text{Tue}) = 0$, which is a contradiction.
- Conclusion: In a psychological Nash equilibrium, $\sigma_1(\text{Mon}) = 1$.

Is Common Belief in Rationality Always Possible?

- For standard games, where the utility only depends (linearly) on your first-order belief, and every player has finitely many choices, we have seen 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 that the teacher prefers to put the exam on Monday if ...
- he believes that you assign a positive probability to Tuesday, or
- he believes that you believe that he believes that you assign a positive probability to Tuesday,
- and so on.
- However, if he believes that you assign probability 1 to Monday, and he believes that you believe that he believes that you assign probability 1 to Monday, and so on, then he prefers to slightly surprise you by putting the exam on Tuesday.
- Then, the teacher's 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}, the teacher believes that you believe that the teacher chooses Monday and has belief hierarchy b_1^{Mon}.

- But then, the teacher believes that you believe that the teacher chooses 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, the teacher believes in b_1 that you believe that the teacher chooses rationally while holding a belief hierarchy different from b_1^{Mon}.
- Hence, in b_1 the teacher believes that you believe that the teacher chooses Monday.

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

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

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.*

In case (b), we can even find a psychological Nash equilibrium for every player.

- Part (a) due to Jagau and Perea (2017), part (b) due to Geanakoplos, Pearce and Stacchetti (1989).
- For case (a), there is a procedure to construct belief hierarchies that express common belief in rationality.

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.

In case (b), we can even find a psychological Nash equilibrium for every player.

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

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

Algorithm

- 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:

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

+

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

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

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

\+

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

	first-order beliefs	
Barbara	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.

Towards an Algorithm

- Suppose the utilities of all 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 the opponents' 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 all players only depend on the 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*

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

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

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

And so on.

- Algorithm taken from Jagau and Perea (2017).

Theorem (Jagau and Perea (2017))

Suppose the utilities of all players only depend on the 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.*

- We say that (c_i, b_i^1) is possible for player i if he expresses up to k-fold belief in rationality, if c_i is optimal for a belief hierarchy b_i that expresses up to k-fold belief in rationality, and that induces b_i^1.
- The algorithm is not guaranteed to terminate within finitely many rounds.

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

+

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

	first-order beliefs	
Barbara	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, *Working paper.*
- Mourmans, N. (2017), Master thesis on the Surprise Exam Paradox, Maastricht University.