# *EPICENTER* Spring Course on Epistemic Game Theory
# Chapters 2 and 3: Common Belief in Rationality

Andrés Perea

EPICENTER
Research Center for Epistemic Game Theory

Maastricht University

July 1, 2019

# What is game theory about?

- In game theory, we study situations where you must make a choice, but where the final outcome also depends on the choices of others.
- Examples are everywhere:
- Negotiating about the price of a car,
- choosing a marketing strategy for your firm,
- bidding in an auction,
- discussing with your partner about what TV program to watch this evening.
- Key question: What choice would you make, and why?
- This depends crucially on how you reason about the opponent!

# Example: Going to a party

| blue | green | red | yellow | same color as Barbara |
|---|---|---|---|---|
| 4 | 3 | 2 | 1 | 0 |

Story

- This evening, you are going to a party together with your friend Barbara.
- You must both decide which color to wear: blue, green, red or yellow.
- Your preferences for wearing these colors are as in the table. These numbers are called utilities.
- You dislike wearing the same color as Barbara: If you both would wear the same color, your utility would be 0.
- What color should you choose, and why?

| blue | green | red | yellow | same color as Barbara |
|---|---|---|---|---|
| 4 | 3 | 2 | 1 | 0 |

- What color is optimal for you depends on your belief about Barbara's choice:

- If you believe that Barbara wears blue, then green is optimal for you.
- If you believe that Barbara wears green, then blue is optimal for you.
- If you believe that Barbara wears red, then blue is optimal for you.
- If you believe that Barbara wears yellow, then blue is optimal for you.

- We call blue and green rational choices for you, because they are optimal for some belief about Barbara's choice.

- Does this mean that red and yellow are irrational for you?

| blue | green | red | yellow | same color as Barbara |
|---|---|---|---|---|
| 4 | 3 | 2 | 1 | 0 |

- Suppose you believe that, with probability 0.6, Barbara chooses blue, and that, with probability 0.4, she chooses green.

- If you would choose blue, your expected utility would be $(0.6) \cdot 0 + (0.4) \cdot 4 = 1.6.$
- If you would choose green, your expected utility would be $(0.6) \cdot 3 + (0.4) \cdot 0 = 1.8.$
- If you would choose red, your utility would be 2.
- If you would choose yellow, your utility would be 1.

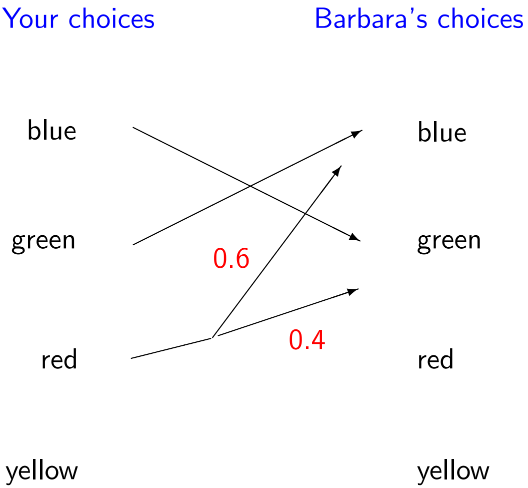
- So, choosing red is optimal for you if you hold this probabilistic belief about Barbara's choice. In particular, red is a rational choice for you.

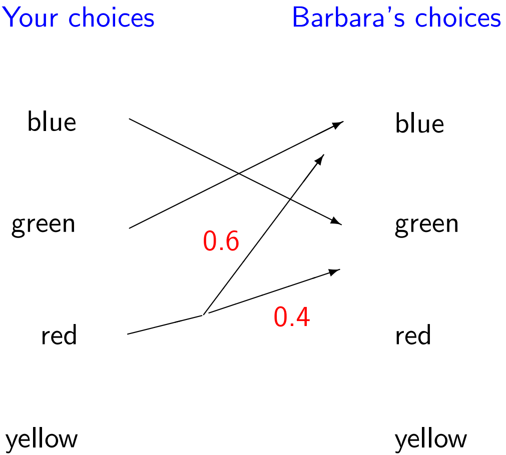
| blue | green | red | yellow | same color as Barbara |
|---|---|---|---|---|
| 4 | 3 | 2 | 1 | 0 |

- Choosing yellow can never be optimal for you, even if you hold a probabilistic belief about Barbara's choice.
- If you assign probability less than 0.5 to Barbara's choice blue, then by choosing blue yourself, your expected utility will be at least $(0.5) \cdot 4 = 2$.
- If you assign probability at least 0.5 to Barbara's choice blue, then by choosing green yourself your expected utility will be at least $(0.5) \cdot 3 = 1.5$.
- Hence, whatever your belief about Barbara, you can always guarantee an expected utility of at least 1.5.
- So, yellow can never be optimal for you, and is therefore an irrational choice for you.

# Beliefs diagram

| blue | green | red | yellow | same color as Barbara |
|---|---|---|---|---|
| 4 | 3 | 2 | 1 | 0 |






- The choices blue, green and red are rational for you.
- But are all of these choices also reasonable? This depends on Barbara's preferences!

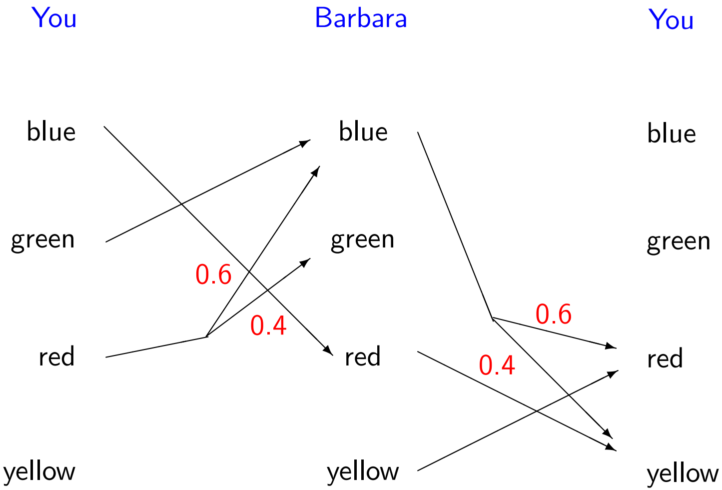
| | blue | green | red | yellow | same color as friend |
|---|---|---|---|---|---|
| you | 4 | 3 | 2 | 1 | 0 |
| Barbara | 2 | 1 | 4 | 3 | 0 |

- For Barbara, the choices red, yellow and blue are rational, whereas green is irrational.
- Choosing red is optimal for her if she believes that you choose yellow.
- Choosing yellow is optimal for her if she believes that you choose red.
- Choosing blue is optimal for her is she believes that, with probability 0.6, you choose red, and with probability 0.4 you choose yellow.

| | blue | green | red | yellow | same color as friend |
|---|---|---|---|---|---|
| you | 4 | 3 | 2 | 1 | 0 |
| Barbara | 2 | × | 4 | 3 | 0 |

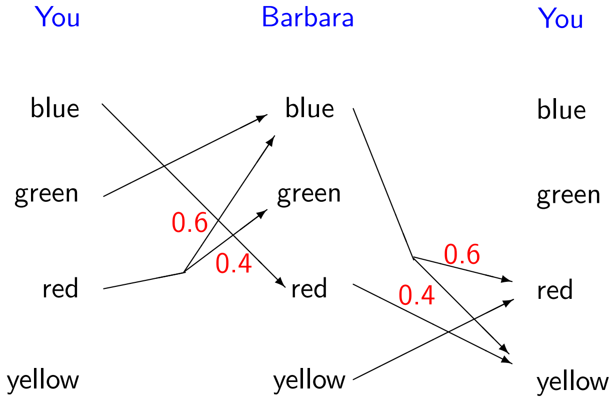
- If you believe that Barbara chooses rationally, you believe that Barbara will choose red, yellow or blue.
- But then, choosing red will no longer be optimal for you, as choosing green will always be better in this case.
- Choosing blue is optimal for you if you believe that Barbara rationally chooses red.
- Choosing green is optimal for you if you believe that Barbara rationally chooses blue.

# Beliefs diagram

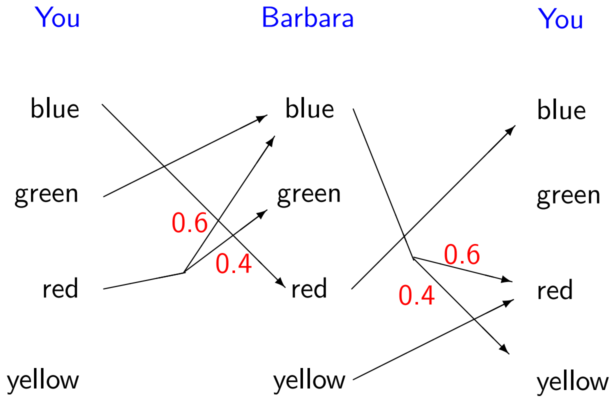


| | blue | green | red | yellow | same color as friend |
|---|---|---|---|---|---|
| you | 4 | 3 | 2 | 1 | 0 |
| Barbara | 2 | 1 | 4 | 3 | 0 |

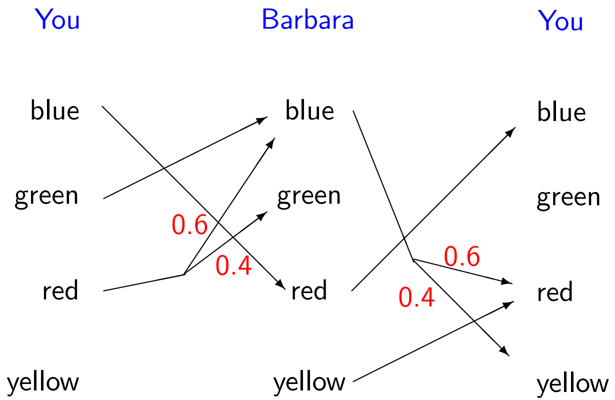
- The color yellow is irrational for you.
- The color red is rational for you, but you can no longer rationally choose it if you believe that Barbara chooses rationally.
- If you believe that Barbara chooses rationally, you can still rationally choose the colors blue and green.
- But are both blue and green reasonable choices for you?



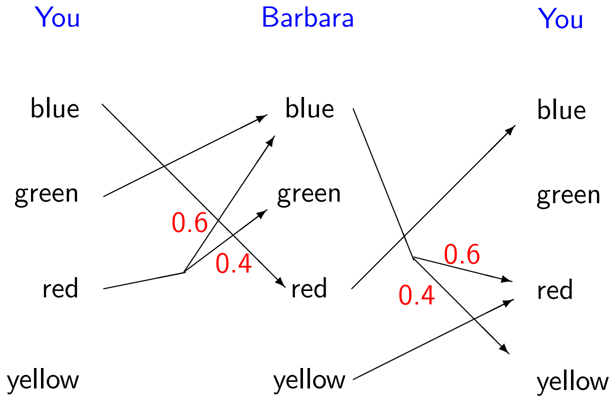

- Consider the belief hierarchy that starts at your choice blue:
- You believe that Barbara chooses red.
- You believe that Barbara believes that you choose yellow.
- You believe that Barbara believes that you choose irrationally (yellow), so this belief hierarchy is not reasonable.




- In this alternative beliefs diagram, consider the belief hierarchy that starts at your choice blue.
- You believe that Barbara rationally chooses red.
- You believe that Barbara believes that you rationally choose blue.
- You believe that Barbara believes that you believe that Barbara rationally chooses red. And so on.




- The belief hierarchy that supports your choice blue expresses common belief in rationality.
- So, you can rationally choose blue under common belief in rationality!




- What about your choice green? Consider the belief hierarchy that starts at your choice green.
- You believe that Barbara chooses blue.
- You believe that Barbara believes that, with probability 0.6, you choose red, and with probability 0.4 you irrationally choose yellow.
- It does not express common belief in rationality.

| | blue | green | red | yellow | same color as friend |
|---|---|---|---|---|---|
| you | 4 | 3 | 2 | × | 0 |
| Barbara | 2 | 1 | 4 | 3 | 0 |

- In fact, you cannot rationally choose green under common belief in rationality:

- If Barbara believes that you choose rationally, then she believes that you will not choose yellow.
- But then, she cannot rationally choose blue, as yellow would always be better for her.
- So, if you believe that Barbara chooses rationally, and that Barbara believes that you choose rationally, you must believe that she will only choose red or yellow.
- But then, you should choose blue, and not green.

| | blue | green | red | yellow | same color as friend |
|---|---|---|---|---|---|
| you | 4 | 3 | 2 | 1 | 0 |
| Barbara | 2 | 1 | 4 | 3 | 0 |

## Summarizing

- Your choice yellow is irrational.
- Your choice red is rational, but can no longer be optimal if you believe that Barbara chooses rationally.
- You can rationally choose green if you believe that Barbara chooses rationally, but not if you believe, in addition, that Barbara believes that you choose rationally.
- You can rationally choose blue under common belief in rationality. In fact, blue is the only color you can rationally choose under common belief in rationality.

- The idea of common belief in rationality first appears in Friedell (1969).
- Later, Armbruster and Böge (1979) and Böge and Eisele (1979) implicitly used the idea of common belief in rationality.
- Spohn (1982) explicitly discusses the idea of common belief in rationality.
- Bernheim (1984) and Pearce (1984) implicitly incorporate the idea of common belief in rationality in their concept of rationalizability.
- Aumann (1987) uses the idea of common belief in rationality as a foundation for correlated equilibrium.
- Brandenburger and Dekel (1987) use the idea of common belief in rationality as a foundation for correlated rationalizability.
- Tan and Werlang (1988) formally define common belief in rationality.
- A historical overview can be found in Perea (2014).

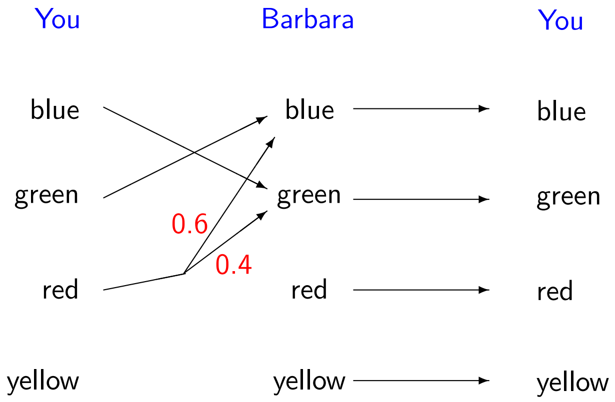
# New Scenario

- Barbara has same preferences over colors as you.
- Barbara likes to wear the same color as you, whereas you dislike this.

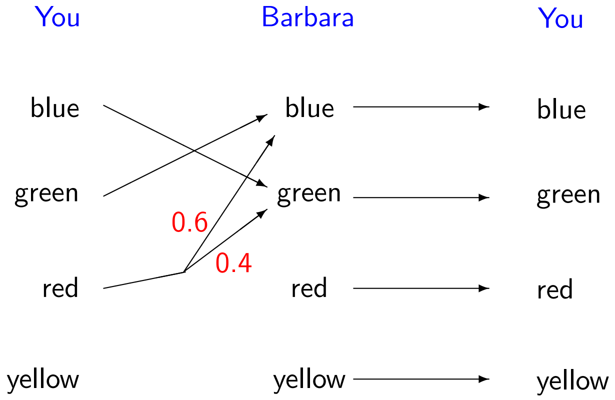
| | blue | green | red | yellow | same color as friend |
|---|---|---|---|---|---|
| you | 4 | 3 | 2 | 1 | 0 |
| Barbara | 4 | 3 | 2 | 1 | 5 |

- Which color(s) can you rationally choose under common belief in rationality?

# Beliefs diagram




| | blue | green | red | yellow | same color as friend |
|---|---|---|---|---|---|
| you | 4 | 3 | 2 | 1 | 0 |
| Barbara | 4 | 3 | 2 | 1 | 5 |




- The belief hierarchy that starts at your choice blue expresses common belief in rationality.
- Similarly, the belief hierarchies that start at your choices green and red also express common belief in rationality.
- So, you can rationally choose blue, green and red under common belief in rationality.

# Choosing rationally

We will now define formally what we mean by a rational choice.

- $I = \{1, 2, ..., n\}$ : set of players.
- $C_i$: set of choices for player $i$.
- A choice-combination for $i$'s opponents is a combination $(c_1, ..., c_{i-1}, c_{i+1}, ..., c_n)$.
- By $C_{-i}$ we denote the set of all choice-combinations for $i$'s opponents.
- A belief for player $i$ about his opponents' choices is a probability distribution $b_i$ over the set $C_{-i}$ of opponents' choice-combinations.
- For every choice-combination $c_{-i} \in C_{-i}$, the number $b_i(c_{-i})$ specifies the probability that player $i$ assigns to the event that his opponents make precisely this combination of choices.

- A utility function for player $i$ is a function $u_i$ that assigns to every combination of choices $(c_1, ..., c_n)$ some number $u_i(c_1, ..., c_n)$.
- The number $u_i(c_1, ..., c_n)$ indicates how desirable player $i$ finds the outcome induced by $(c_1, ..., c_n)$.
- In the example "Going to a party":
- $u_1(green, \ red) = 3$,
- $u_1(green, \ blue) = 3$,
- $u_1(green, \ green) = 0$,
- $u_1(blue, \ red) = 4$.

- Suppose that player $i$ holds a belief $b_i$ about the opponents' choices.
- The expected utility of making choice $c_i$, while having the belief $b_i$, is

$$u_i(c_i, b_i) = \sum_{c_{-i} \in C_{-i}} b_i(c_{-i}) \cdot u_i(c_i, c_{-i}).$$

- The choice $c_i$ is optimal for player $i$ given his belief $b_i$, if

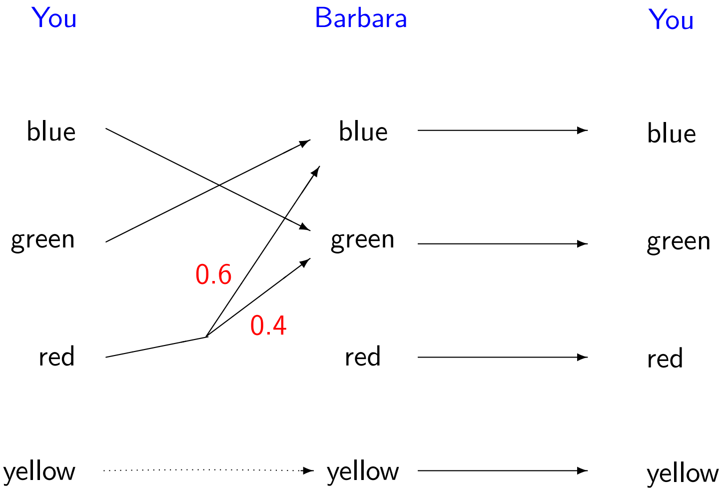
$$u_i(c_i, b_i) \geq u_i(c_i', b_i)$$

  for all other choices $c_i' \in C_i$.
- The choice $c_i$ is rational for player $i$ if it is optimal for some belief $b_i$ about the opponents' choices.

# Belief hierarchies

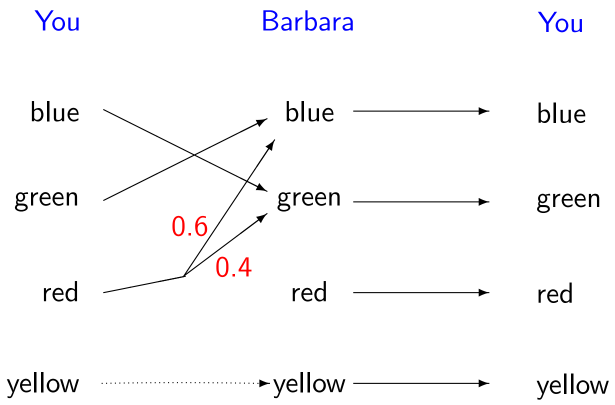
- A first-order belief is a belief about an opponent's choice.
- In order to judge whether a first-order belief about player $j$'s choice is reasonable, you must also hold
- a belief about what $j$ believes about his opponents' choices: second-order belief.
- In order to judge whether this second-order belief is reasonable, you must also hold
- a belief about what $j$ believes about what the others believe about their opponents' choices: third-order belief.
- And so on.
- This yields a belief hierarchy. Harsanyi (1962, 1967–1968).
- Belief hierarchies can be constructed from an extended beliefs diagram.

# Extended beliefs diagram



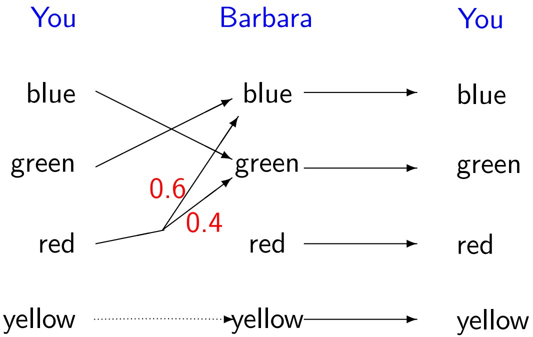
# Epistemic model

- Writing down a belief hierarchy explicitly is impossible. You must write down
- your belief about the opponents' choices
- your belief about what your opponents believe about their opponents' choices,
- a belief about what the opponents believe that their opponents believe about the other players' choices,
- and so on, ad infinitum.
- Is there an easy way to encode a belief hierarchy?

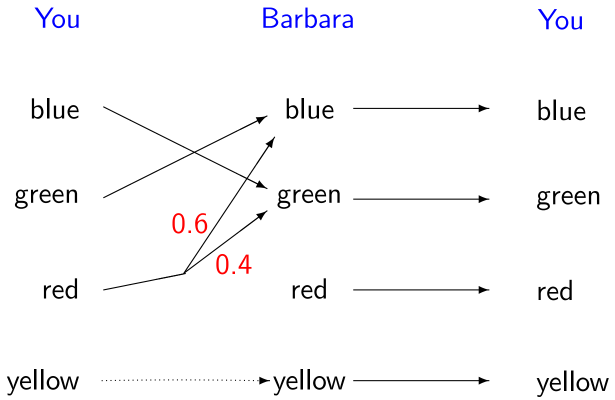



Even writing down the first three levels of the belief hierarchy that starts at your choice red is a nightmare!

- A belief hierarchy for you consists of a first-order belief, a second-order belief, a third-order belief, and so on.
- In a belief hierarchy, you hold a belief about
- the opponents' choices,
- the opponents' first-order beliefs,
- the opponents' second-order beliefs,
- and so on.
- Hence, in a belief hierarchy you hold a belief about
- the opponents' choices, and the opponents' belief hierarchies.
- Call a belief hierarchy a type.
- Then, a type holds a belief about the opponents' choices and the opponents' types.
- Idea goes back to Harsanyi (1967–68).

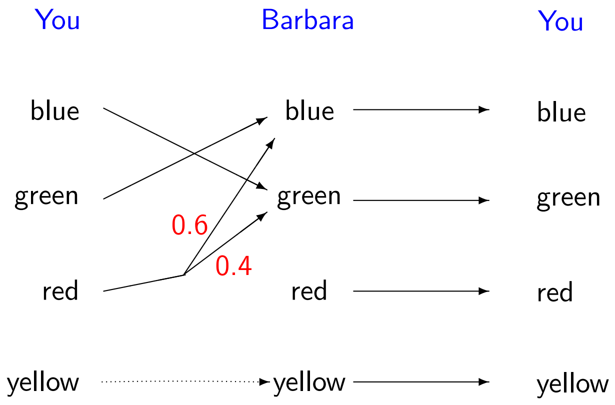



- Denote by $t_1^{red}$ your belief hierarchy that starts at your choice red.
- Denote by $t_2^{blue}$ and $t_2^{green}$ the belief hierarchies for Barbara that start at her choices blue and green.
- Then, $t_1^{red}$ believes that, with prob. 0.6, Barbara chooses blue and has belief hierarchy $t_2^{blue}$, and believes that, with prob. 0.4, Barbara chooses green and has belief hierarchy $t_2^{green}$.




- Formally: We call the belief hierarchies $t_1^{red}$, $t_2^{blue}$ and $t_2^{green}$ types.
- Type $t_1^{red}$ has belief

$$b_1(t_1^{red}) = (0.6) \cdot (blue, t_2^{blue}) + (0.4) \cdot (green, t_2^{green}).$$




- Also, $b_1(t_1^{blue}) = (green, t_2^{green})$ and $b_1(t_1^{green}) = (blue, t_2^{blue})$ and finally $b_1(t_1^{yellow}) = (yellow, t_2^{yellow})$.
- We can do the same for Barbara's belief hierarchies. This leads to an epistemic model.

# Epistemic model for "Going to a party"

| | |
|---|---|
| Types | $T_1 = \{t_1^{blue}, t_1^{green}, t_1^{red}, t_1^{yellow}\}$<br>$T_2 = \{t_2^{blue}, t_2^{green}, t_2^{red}, t_2^{yellow}\}$ |
| Beliefs for player 1 | $b_1(t_1^{blue}) = (green, \; t_2^{green})$<br>$b_1(t_1^{green}) = (blue, \; t_2^{blue})$<br>$b_1(t_1^{red}) = (0.6) \cdot (blue, \; t_2^{blue}) + (0.4) \cdot (green, \; t_2^{green})$<br>$b_1(t_1^{yellow}) = (yellow, \; t_2^{yellow})$ |
| Beliefs for player 2 | $b_2(t_2^{blue}) = (blue, \; t_1^{blue})$<br>$b_2(t_2^{green}) = (green, \; t_1^{green})$<br>$b_2(t_2^{red}) = (red, \; t_1^{red})$<br>$b_2(t_2^{yellow}) = (yellow, \; t_1^{yellow})$ |

- In an epistemic model, we can derive for every type the first-order belief, second-order belief, and so on.
- So, we can derive for every type the complete belief hierarchy .

| | |
|---|---|
| Types | $T_1 = \{t_1^{blue}, t_1^{green}, t_1^{red}, t_1^{yellow}\}$<br>$T_2 = \{t_2^{blue}, t_2^{green}, t_2^{red}, t_2^{yellow}\}$ |
| Beliefs for player 1 | $b_1(t_1^{blue}) = (green, t_2^{green})$<br>$b_1(t_1^{green}) = (blue, t_2^{blue})$<br>$b_1(t_1^{red}) = (0.6) \cdot (blue, t_2^{blue}) + (0.4) \cdot (green, t_2^{green})$<br>$b_1(t_1^{yellow}) = (yellow, t_2^{yellow})$ |
| Beliefs for player 2 | $b_2(t_2^{blue}) = (blue, t_1^{blue})$<br>$b_2(t_2^{green}) = (green, t_1^{green})$<br>$b_2(t_2^{red}) = (red, t_1^{red})$<br>$b_2(t_2^{yellow}) = (yellow, t_1^{yellow})$ |

## Definition (Epistemic model)

An epistemic model specifies for every player $i$ a set $T_i$ of possible types.

Moreover, for every type $t_i$ it specifies a probabilistic belief $b_i(t_i)$ over the set $C_{-i} \times T_{-i}$ of opponents' choice-type combinations.

- Here, $C_{-i} \times T_{-i}$ is the set of combinations

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

  of opponents' choices and opponents' types.

- For every such combination $(c_{-i}, t_{-i}) \in C_{-i} \times T_{-i}$, the probability

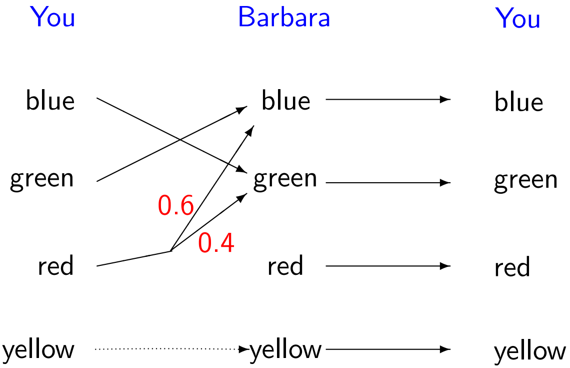
$$b_i(t_i)(c_{-i}, t_{-i})$$

  represents the probability that type $t_i$ assigns to the event that the opponents choose $c_{-i}$ and that the opponents' belief hierarchies are given by $t_{-i}$.

- Belief hierarchies can also be encoded by Kripke-structures (Kripke, 1963) and Aumann-structures (Aumann, 1974, 1976).

# Common belief in rationality

- Intuitively, common belief in rationality means that
- you believe that your opponents choose rationally,
- you believe that your opponents believe that their opponents choose rationally,
- and so on, ad infinitum.
- How can we state common belief in rationality formally, within an epistemic model?




- Your type $t_1^{red}$ has belief $b_1(t_1^{red}) = (0.6) \cdot (blue, t_2^{blue}) + (0.4) \cdot (green, t_2^{green})$.
- For Barbara, blue is optimal for type $t_2^{blue}$, and green is optimal for type $t_2^{green}$.
- So, type $t_1^{red}$ only assigns positive probability to choice-type pairs for Barbara where the choice is optimal for the type.
- We say that $t_1^{red}$ believes in Barbara's rationality.

## Definition (Belief in the opponents' rationality)

Type $t_i$ believes in the opponents' rationality if his belief $b_i(t_i)$ only assigns positive probability to choice-type combinations

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

where choice $c_1$ is optimal for type $t_1$, ..., choice $c_n$ is optimal for type $t_n$.

## 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 $k$.

- This definition is similar to Tan and Werlang (1988).
- In the literature, this concept is also known as correlated rationalizability. (Brandenburger and Dekel (1987))

## 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 $k$.

- Rationalizability (Bernheim (1984), Pearce (1984)) is obtained if in games with three players or more we impose the following additional condition:
- Player $i$'s belief about opponent $j$'s choice must be independent from his belief about opponent $k$'s choice.

## Definition

Player $i$ can rationally make choice $c_i$ under common belief in rationality if there is some epistemic model, and some type $t_i$ within this epistemic model, such that

type $t_i$ expresses common belief in rationality, and

choice $c_i$ is optimal for type $t_i$.

## Theorem (Sufficient condition for common belief in rationality)

*Consider an epistemic model in which all types believe in the opponents' rationality.*

*Then, all types in the epistemic model express common belief in rationality.*

- Proof: Show that every type expresses $k$-fold belief in rationality, for all $k$.
- Every type expresses 1-fold belief in rationality.
- Since a type can only assign positive probability to other types in the same model, every type expresses 2-fold belief in rationality.
- But then, every type also expresses 3-fold belief in rationality.
- And so on.
- Hence, all types express common belief in rationality. ■

| | blue | green | red | yellow | same color as friend |
|---|---|---|---|---|---|
| you | 4 | 3 | 2 | 1 | 0 |
| Barbara | 4 | 3 | 2 | 1 | 5 |

| | |
|---|---|
| Types | $T_1 = \{t_1^{blue}, t_1^{green}, t_1^{red}\}$ <br> $T_2 = \{t_2^{blue}, t_2^{green}, t_2^{red}\}$ |
| Beliefs for player 1 | $b_1(t_1^{blue}) = (green, \ t_2^{green})$ <br> $b_1(t_1^{green}) = (blue, \ t_2^{blue})$ <br> $b_1(t_1^{red}) = (0.6) \cdot (blue, \ t_2^{blue}) + (0.4) \cdot (green, \ t_2^{green})$ |
| Beliefs for player 2 | $b_2(t_2^{blue}) = (blue, \ t_1^{blue})$ <br> $b_2(t_2^{green}) = (green, \ t_1^{green})$ <br> $b_2(t_2^{red}) = (red, \ t_1^{red})$ |

- Every type believes in the opponent's rationality.
- Hence, every type expresses common belief in rationality.

# Algorithm

- We look for an algorithm that helps us find those choices you can rationally make under common belief in rationality.
- Start with more basic question: Can we characterize those choices that are rational – that is, optimal for some belief?

- Consider the example "Going to a party".

| blue | green | red | yellow | same color as Barbara |
|---|---|---|---|---|
| 4 | 3 | 2 | 1 | 0 |

- Only your choice yellow is irrational.
- Your choice yellow is strictly dominated by the randomized choice in which you choose blue and green with probability 0.5.

| | blue | green | red | yellow |
|---|---|---|---|---|
| yellow | 1 | 1 | 1 | 0 |
| randomized choice | 1.5 | 2 | 3.5 | 3.5 |

- In the example "Going to a party " we see the following:
- A choice is irrational precisely when it is strictly dominated by another choice, or strictly dominated by a randomized choice.
- In fact, this is always true!

**Theorem (Pearce's Lemma)**

*A choice is irrational, if and only if, it is strictly dominated by another choice, or strictly dominated by a randomized choice.*

- Due to Pearce (1984).
- Or, equivalently:

**Theorem (Pearce's Lemma)**

*A choice is rational, if and only if, it is not strictly dominated by another choice, nor strictly dominated by a randomized choice.*

- Formally, a choice $c_i$ is strictly dominated by a choice $c_i'$ if

$$u_i(c_i, c_{-i}) < u_i(c_i', c_{-i})$$

for every opponents' choice-combination $c_{-i}$.

- A randomized choice for player $i$ is a probability distribution $r_i$ over his set of choices $C_i$.

- A choice $c_i$ is strictly dominated by a randomized choice $r_i$ if

$$u_i(c_i, c_{-i}) < u_i(r_i, c_{-i})$$

for every opponents' choice-combination $c_{-i}$.

## Step 1: 1-fold belief in rationality

- Which choices are rational for a type that expresses 1-fold belief in rationality?
- If you believe in the opponents' rationality, then you assign positive probability only to opponents' choices that are rational.
- Remember: A choice is rational precisely when it is not strictly dominated.
- So, if you believe in the opponents' rationality, then you assign positive probability only to opponents' choices that are not strictly dominated.

## Step 1: 1-fold belief in rationality

- So, if you believe in the opponents' rationality, then you assign positive probability only to opponents' choices that are not strictly dominated.
- In a sense, you eliminate the opponents' strictly dominated choices from the game, and concentrate on the reduced game that remains.
- The choices that you can rationally make if you believe in your opponents' rationality, are exactly the choices that are optimal for you for some belief within this reduced game.
- But these are exactly the choices that are not strictly dominated for you within this reduced game.
- Hence, these are the choices that survive 2-fold elimination of strictly dominated choices.

## Step 2: Up to 2-fold belief in rationality

- Which choices are rational for a type that expresses up to 2-fold belief in rationality?
- Consider a type $t_i$ that expresses up to 2-fold belief in rationality. Then, $t_i$ only assigns positive probability to opponents' choice-type pairs $(c_j, t_j)$ where $c_j$ is optimal for $t_j$, and $t_j$ expresses 1-fold belief in rationality.
- So, type $t_i$ only assigns positive probability to opponents' choices $c_j$ which are optimal for a type that expresses 1-fold belief in rationality.
- Hence, type $t_i$ only assigns positive probability to opponents' choices $c_j$ which survive 2-fold elimination of strictly dominated choices.

## Step 2: Up to 2-fold belief in rationality

- Hence, type $t_i$ only assigns positive probability to opponents' choices $c_j$ which survive 2-fold elimination of strictly dominated choices.
- Then, every choice $c_i$ which is optimal for $t_i$ must be optimal for some belief within the reduced game obtained after 2-fold elimination of strictly dominated choices.
- So, every choice $c_i$ which is optimal for $t_i$ must not be strictly dominated within the reduced game obtained after 2-fold elimination of strictly dominated choices.
- Conclusion: Every choice that is optimal for a type that expresses up to 2-fold belief in rationality, must survive 3-fold elimination of strictly dominated choices.

## Algorithm (Iterated elimination of strictly dominated choices)

**Step 1.** *Within the original game, eliminate all choices that are strictly dominated.*

**Step 2.** *Within the reduced game obtained after step 1, eliminate all choices that are strictly dominated.*

**Step 3.** *Within the reduced game obtained after step 2, eliminate all choices that are strictly dominated.*

⋮

*Continue in this fashion until no further choices can be eliminated.*

**Theorem (Algorithm "works")**

*(1) For every $k \geq 1$, the choices that are optimal for a type that expresses up to k-fold belief in rationality are exactly those choices that survive $(k+1)$-fold elimination of strictly dominated choices.*

*(2) The choices that can rationally be made under common belief in rationality are exactly those choices that survive iterated elimination of strictly dominated choices.*

- Based on Theorems 5.1, 5.2 and 5.3 in Tan and Werlang (1988).
- Adam Brandenburger (2014) calls it the Fundamental Theorem of Epistemic Game Theory.

# Properties of the algorithm

**Algorithm (Iterated elimination of strictly dominated choices)**

**Step 1.** *Within the original game, eliminate all choices that are strictly dominated.*

**Step 2.** *Within the reduced game obtained after step 1, eliminate all choices that are strictly dominated.*

**Step 3.** *Within the reduced game obtained after step 2, eliminate all choices that are strictly dominated.*

⋮

*Continue in this fashion until no further choices can be eliminated.*

- This algorithm always stops after finitely many steps.
- It always yields a nonempty output for every player.
- The order and speed by which you eliminate choices is not relevant for the eventual output.

# Example: Going to a party

Barbara

| | | blue | green | red | yellow |
|---|---|---|---|---|---|
| You | blue | 0, 0 | 4, 1 | 4, 4 | 4, 3 |
| | green | 3, 2 | 0, 0 | 3, 4 | 3, 3 |
| | red | 2, 2 | 2, 1 | 0, 0 | 2, 3 |
| | yellow | 1, 2 | 1, 1 | 1, 4 | 0, 0 |

- **Step 1.** Your choice yellow is strictly dominated by randomized choice $(0.5) \cdot \textit{blue} + (0.5) \cdot \textit{green}$.
- Barbara's choice green is strictly dominated by randomized choice $(0.5) \cdot \textit{red} + (0.5) \cdot \textit{yellow}$.
- Eliminate your choice yellow and Barbara's choice green.

# Example: Going to a party

Barbara

| You | | blue | red | yellow |
|---|---|---|---|---|
| | blue | 0, 0 | 4, 4 | 4, 3 |
| | green | 3, 2 | 3, 4 | 3, 3 |
| | red | 2, 2 | 0, 0 | 2, 3 |

- **Step 2.** Your choice red is strictly dominated by green.
- Barbara's choice blue is strictly dominated by yellow.
- Eliminate your choice red and Barbara's choice blue.

# Example: Going to a party

Barbara

| | | red | yellow |
|---|---|---|---|
| You | blue | 4, 4 | 4, 3 |
| | green | 3, 4 | 3, 3 |

- **Step 3.** Your choice green is strictly dominated by blue.
- Barbara's choice yellow is strictly dominated by red.
- Eliminate your choice green and Barbara's choice yellow.

# Example: Going to a party

Barbara

| You | red |
|---|---|
| blue | 4, 4 |

- Procedure stops.
- Under common belief in rationality, you can only rationally wear blue, and Barbara can only rationally wear red.
- Under 1-fold belief in rationality, you can rationally wear blue or green, and Barbara can rationally wear red or yellow.

## Theorem (Algorithm "works")

*(1) For every $k \geq 1$, the choices that are optimal for a type that expresses up to k-fold belief in rationality are exactly those choices that survive $(k+1)$-fold elimination of strictly dominated choices.*

*(2) The choices that can rationally be made under common belief in rationality are exactly those choices that survive iterated elimination of strictly dominated choices.*

- Proof of part (2):
- We have shown: If a choice can rationally be made under common belief in rationality, then it must survive iterated elimination of strictly dominated choices.

- We now show the converse: If a choice survives iterated elimination of strictly dominated choices, then it can rationally be made under common belief in rationality.
- Assume two players. Suppose that the algorithm terminates after $K$ steps. Let $C_i^K$ be the set of surviving choices for player $i$.
- Then, every choice in $C_i^K$ is not strictly dominated within reduced game $\Gamma^K$. Hence, every choice $c_i$ in $C_i^K$ is optimal for some belief $b_i^{c_i} \in \Delta(C_j^K)$.
- Define set of types $T_i = \{t_i^{c_i} : c_i \in C_i^K\}$ for both players $i$.
- Every type $t_i^{c_i}$ only deems possible opponents' choice-type pairs $(c_j, t_j^{c_j})$, with $c_j \in C_j^K$, and

$$b_i(t_i^{c_i})(c_j, t_j^{c_j}) := b_i^{c_i}(c_j).$$

- Then, every type $t_i^{c_i}$ believes in the opponents' rationality.
- Hence, every type expresses common belief in rationality. ■

## Corollary (Common belief in rationality is always possible)

*We can always construct an epistemic model in which all types express common belief in rationality.*

# Other classes of games

- Common belief in rationality, and an associated algorithm, have also been defined for other classes of games:
- games with incomplete information: Battigalli and Siniscalchi (1999, 2002, 2007), Battigalli (2003), Battigalli, Di Tillio, Grillo and Penta (2011), Battigalli and Prestipino (2011), Bach and Perea (2016)
- games with unawareness: Perea (2017)
- psychological games: Battigalli and Dufwenberg (2009), Bjorndahl, Halpern and Pass (2013), Jagau and Perea (2017, 2019), Mourmans (2017, 2019)
- Research question: Simple algorithms for (certain classes of) psychological games?
- Research question: Application of common belief in rationality to models in economics?

# Example: Guessing two-thirds of the average

Story

- Is also known as the "Guessing Game", or "Beauty Contest".
- Has been studied experimentally in Nagel (1995).
- All students in this room must write a number on a piece of paper, between 1 and 100.
- The closer you are to two-thirds of the average of all numbers, the higher your prize money.

- What number(s) could you have rationally written down under common belief in rationality?
- Apply the algorithm of "iterated elimination of strictly dominated choices".
- Step 1: What numbers are strictly dominated?
- Two-thirds of the average can never be above 67.
- Hence, every number above 67 is strictly dominated by 67.
- Eliminate all numbers above 67.

- Step 2: Consider the reduced game $\Gamma^1$ in which only the numbers $1, \ldots, 67$ remain for all students.
- Which numbers are strictly dominated in $\Gamma^1$ ?
- Two-thirds of the average of all numbers in $\Gamma^1$ can never be above $\frac{2}{3} \cdot 67 \approx 45$.
- All numbers above 45 are strictly dominated in $\Gamma^1$.
- Eliminate all numbers above 45.
- And so on.
- Only the number 1 remains at the end.
- Under common belief in rationality, you must choose number 1.
- Would you really choose this number? Why?

- Armbruster, W. and W. Böge, (1979), Bayesian game theory, in O. Moeschlin and D. Pallaschke (eds.), *Game Theory and Related Topics*, North-Holland, Amsterdam.
- Aumann, R.J. (1974), Subjectivity and correlation in randomized strategies, *Journal of Mathematical Economics* **1**, 67–96.
- Aumann, R.J. (1976), Agreeing to disagree, *Annals of Statistics* **4**, 1236–1239.
- Aumann, R.J. (1987), Correlated equilibrium as an expression of Bayesian rationality, *Econometrica* **55**, 1–18.
- Bach, C.W. and A. Perea (2016), Incomplete information and generalized iterated strict dominance, *EPICENTER* Working Paper No. 7.
- Battigalli, P. (2003), Rationalizability in infinite, dynamic games of incomplete information, *Research in Economics* **57**, 1–38.

Battigalli, P. and A. Prestipino (2013), Transparent restrictions on beliefs and forward-induction reasoning in games with asymmetric information, *B. E. Journal of Theoretical Economics* **13**, 79–130.

Battigalli, P., Di Tillio, E., Grillo, E. and A. Penta (2011), Interactive epistemology and solution concepts for games with asymmetric information. *B. E. Journal of Theoretical Economics* **11,** Issue 1, Article 6.

Battigalli, P. and M. Siniscalchi (1999), Hierarchies of conditional beliefs and interactive epistemology in dynamic games, *Journal of Economic Theory* **88**, 188–230.

Battigalli, P. and M. Siniscalchi (2002), Strong belief and forward induction reasoning, *Journal of Economic Theory* **106,** 356–391.

Battigalli, P. and M. Siniscalchi (2007), Interactive epistemology in games with payoff uncertainty, *Research in Economics* **61**, 165–184.

Battigalli, P., and M. Dufwenberg (2009), Dynamic psychological games, *Journal of Economic Theory* **144**, 1–35.

- Bernheim, B.D. (1984), Rationalizable strategic behavior, *Econometrica* **52,** 1007–1028.
- Bjorndahl, A., Halpern, J.Y. and R. Pass (2013), Language-based games, *Proceedings of Fourteenth Conference on Theoretical Aspects of Rationality and Knowledge.*
- Böge, W. and T.H. Eisele (1979), On solutions of bayesian games, *International Journal of Game Theory* **8**, 193–215.
- Brandenburger, A. (2014), *The Language of Game Theory: Putting Epistemics into the Mathematics of Games,* World Scientific Series in Economic Theory, Volume 5.
- Brandenburger, A. and E. Dekel (1987), Rationalizability and correlated equilibria, *Econometrica* **55**, 1391–1402.
- Friedell, M.F. (1969), On the structure of shared awareness, *Behavioral Science* **14**, 28–39.

- Harsanyi, J.C. (1962), Bargaining in ignorance of the opponent's utility function, *Journal of Conflict Resolution* **6**, 29–38.
- Harsanyi, J.C. (1967–68), Games with incomplete information played by "bayesian" players, I–III, *Management Science* **14**, 159–182, 320–334, 486–502.
- 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.*
- Kripke, S. (1963), A semantical analysis of modal logic I: Normal modal propositional calculi, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **9**, 67–96.
- 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.
- Nagel, R. (1995), Unravelling in guessing games: An experimental study, *American Economic Review* **85**, 1313–1326.
- Pearce, D.G. (1984), Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52,** 1029–1050.
- Perea, A. (2014), From classical to epistemic game theory, *International Game Theory Review* **16,** No.1, 144001, 22 pages.
- Perea, A. (2017), Common belief in rationality in games with unawareness, *EPICENTER* Working Paper No. 13.
- Spohn, W. (1982), How to make sense of game theory, in W. Stegmüller, W. Balzer and W. Spohn (eds.), *Philosophy of Economics,* Springer Verlag, pp. 239–270.
- Tan, T. and S.R.C. Werlang (1988), The bayesian foundations of solution concepts of games, *Journal of Economic Theory* **45,** 370–391.