Day 8: correct and symmetric beliefs in psychological games

Niels Mourmans

July 3, 2024

EPICENTER
Research Center for
Epistemic Game Theory

Introduction

- Psychological games model situations where preferences directly depend on outcomes AND on (higher-order) beliefs
  - States comprise of combinations of choices AND expectations.
  - Examples: surprise (see e.g. Khalmetski et al. (2015)), guilt (se e.g. Dufwenbwerg and Charness (2006)), anger (see e.g. Aina et al. (2020)).

Introduction

- Psychological games model situations where preferences directly depend on outcomes AND on (higher-order) beliefs
  - States comprise of combinations of choices AND expectations.
  - Examples: surprise (see e.g. Khalmetski et al. (2015)), guilt (se e.g. Dufwenbwerg and Charness (2006)), anger (see e.g. Aina et al. (2020)).

Definition (Psychological Game)

A psychological game with two players specifies, for both players i a decision problem (C_i, S_i, u_i) where

1. the set of choices is C_i;
2. the set of states $S_i = C_j \times C_i$ consists of all choice-pairs (c_j, c_i) where $c_j \in C_j$ and $c_i \in C_i$; and
3. player i's conditional preference relation has an expected utility representation u_i assigning to every choice $c_i \in C_i$ and every state $(c_j, c_i') \in S_i$ some utility $u_i(c_i, (c_j, c_i'))$.

Introduction

- Yesterday and this morning: common belief in rationality (CBR) in psychological games
    - Same definition as in standard games.
    - Needed different procedure for characterization → more complexity.
- Simple belief hierarchies/Psychological Nash Equilibrium and symmetric belief hierarchies/Psychological Correlated Equilibrium

Outline

What we want to achieve is the following

- Explore the concepts of simple belief hierarchies and symmetric belief hierarchies in psychological games;
- Then see how these concept link to equilibrium concepts;
- Then characterize choices that can be made under belief hierarchies that (1) express common belief in rationality and (2) are simple/symmetric belief hierarchies.

Outline

What we want to achieve is the following

- Explore the concepts of simple belief hierarchies and symmetric belief hierarchies in psychological games;
- Then see how these concept link to equilibrium concepts;
- Then characterize choices that can be made under belief hierarchies that (1) express common belief in rationality and (2) are simple/symmetric belief hierarchies.
- **Note**: we will only look at 2-player psychological games throughout this Lecture.

Introducing Leading Example

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

Table 1: *Decision Problems for 'Barbara's Birthday'*

- You want to buy Barbara **surprising** present
  - necklace (3) better than ring (2) better than bracelet (1);
  - Above all: it must be a surprise (otherwise 0)
- Barbara wants to guess correctly

Introducing Leading Example

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

- Surprise: Your choice is different from what Barbara believes is your choice.
- Degree of Surprise: the probability that Barbara does NOT assign to your true choice $\tilde{c}_i$, that is: $1 - b_B^1(\tilde{c}_i)$

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

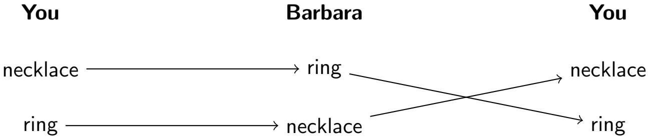
Barbara	$(n, \cdot)$	$(r, \cdot)$	$(b, \cdot)$
necklace	1	0	0
ring	0	1	0
bracelet	0	0	1

- Elimination of choices and states is enough: YOUR utility only depends on second order belief.
- Round 1: eliminate choice Bracelet for you
- Round 2: eliminate state $(\cdot, b)$ for YOU and $(b, \cdot)$ for Barbara and then choice Bracelet for Barbara.
- Round 3: Nothing can be eliminated further. Procedure terminates.
- You can choose necklace or ring under CBR.

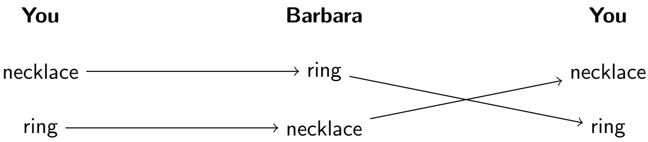
Introducing Leading Example

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

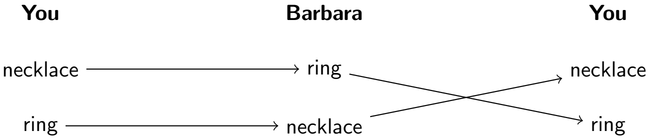


Simple belief hierarchies



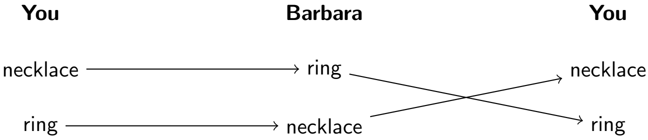
- Are these simple belief hierarchies?

Simple belief hierarchies



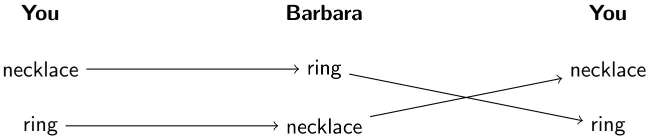
- Are these simple belief hierarchies?
  - No, they are not.
  - Barbara is **incorrect** about your beliefs
  - Simple belief hierarchy →

Simple belief hierarchies



- Are these simple belief hierarchies?
  - No, they are not.
  - Barbara is **incorrect** about your beliefs
  - Simple belief hierarchy → Chapter 4: all higher-order beliefs generated by single belief σ_1 about your choice and single belief σ_2 about Barbara's choice

Simple belief hierarchies



- Are these simple belief hierarchies?
  - No, they are not.
  - Barbara is **incorrect** about your beliefs
  - Simple belief hierarchy → Chapter 4: all higher-order beliefs generated by single belief σ_1 about your choice and single belief σ_2 about Barbara's choice
- What would change in prediction of behaviour in game if we assume simple belief hierarchy?
- Psychological game equivalent of Nash Equilibrium?

Simple belief hierarchies

- We can think of simple belief hierarchies in psychological games the same way as in standard games

Simple belief hierarchies

Definition (Simple belief hierarchy)

Let σ_1 be a probabilistic belief about player 1's choice and σ_2 be a probabilistic belief about player 2's choice. The belief hierarchy for player i generated by the belief (σ_1, σ_2) is defined as follows:

1. in the first-order belief, player i assigns to every opponent's choice c_j the probability $\sigma_j(c_j)$,
2. in the second-order belief, player i believes with probability 1 that opponent j assigns to every choice c_i for player i the probability $\sigma_i(c_i)$,
3. in the third-order belief, player i believes with probability 1 that player j believes with probability 1 that player i assigns to every opponent's choice c_j the probability $\sigma_j(c_j)$, and so on.

A belief hierarchy is called **simple** if it is generated by a pair of such beliefs (σ_1, σ_2).

Psychological Nash Equilibrium

- Combine common belief in rationality and simple belief hierarchy
- In standard games we get: **Nash Equilibrium** $\rightarrow (\sigma_1, \sigma_2)$ where σ_1 is best-response / optimal against σ_2 and vice versa
- In psychological games we get: **Psychological Nash Equilibrium (PNE)**:
  - Similar as NE in standard games;
  - only difference is that a choice is now optimal against a higher-order expectation / belief, not against just an expectation / belief about opponent's choice.

Psychological Nash Equilibrium

- Combine common belief in rationality and simple belief hierarchy
- In standard games we get: **Nash Equilibrium** $\rightarrow (\sigma_1, \sigma_2)$ where σ_1 is best-response / optimal against σ_2 and vice versa
- In psychological games we get: **Psychological Nash Equilibrium (PNE)**:
  - Similar as NE in standard games;
  - only difference is that a choice is now optimal against a higher-order expectation / belief, not against just an expectation / belief about opponent's choice.

We want to show here now that CBR + simple belief hierarchy implies PNE and vice versa

Psychological Nash Equilibrium

To do: common belief in rationality + simple belief hierarchy $\rightarrow$ Psychological Nash Equilibrium

- Let us start with a simple belief hierarchy for player i generated by (σ_1, σ_2).
- **Step 1**: player i expresses 1-fold belief in rationality $\rightarrow$ σ_2 must only assign positive probability $\sigma_2(c_j) > 0$ to choice c_j where c_j is optimal give a some second-order expectation e_j^2.

Psychological Nash Equilibrium

To do: common belief in rationality + simple belief hierarchy → Psychological Nash Equilibrium

- Let us start with a simple belief hierarchy for player i generated by (σ_1, σ_2).
- **Step 1**: player i expresses 1-fold belief in rationality → σ_2 must only assign positive probability $\sigma_2(c_j) > 0$ to choice c_j where c_j is optimal give a some second-order expectation e_j^2.
- What is e_j^2?

Psychological Nash Equilibrium

To do: common belief in rationality + simple belief hierarchy → Psychological Nash Equilibrium

- Let us start with a simple belief hierarchy for player i generated by (σ_1, σ_2).
- **Step 1**: player i expresses 1-fold belief in rationality → σ_2 must only assign positive probability $\sigma_2(c_j) > 0$ to choice c_j where c_j is optimal give a some second-order expectation e_j^2.
- What is e_j^2? second-order expectation e_j^2 is a probability distribution over $C_i \times C_j$, so $e_j^2 \in \Delta(C_i \times C_j)$.
- A second-order expectation e_j^2 specifically induced by (σ_1, σ_2) is as follows:
  - player j has belief σ_1 over player i's choices and believes that player i has belief σ_2 over player j's choices.
  - e_j^2 thus assigns to each pair $(c_i, c_j) \in C_i \times C_j$ probability $\sigma_1(c_i) \cdot \sigma_2(c_j)$.

Psychological Nash Equilibrium

Definition (induced second-order expectation)

Consider a pair of beliefs (σ_1, σ_2) where σ_1 is a probabilistic belief about 1's choice, and σ_2 a probabilistic belief about 2's choice. For player i, the **second-order expectation** $e_i[\sigma_1, \sigma_2]$ **induced by** (σ_1, σ_2) is the probability distribution that assigns to every pair of choice $(c_j, c_i) \in C_j \times C_i$ the probability $\sigma_j(c_j) \times \sigma_i(c_i)$.

Psychological Nash Equilibrium

- Example of an induced second-order expectation: say we have
  - $\sigma_1 = 0.2 \cdot \textit{necklace} + 0.5 \cdot \textit{ring} + 0.3 \cdot \textit{bracelet}$, and
  - $\sigma_2 = 0.6 \cdot \textit{necklace} + 0.4 \cdot \textit{bracelet}$
  - What is $e_j[\sigma_1, \sigma_2]$ induced by (σ_1, σ_2)?

$$\begin{aligned} e_j[\sigma_1, \sigma_2] = {} & \sigma_1(n)\sigma_2(n)(n, n) + \sigma_1(n)\sigma_2(r)(n, r) + \sigma_1(n)\sigma_2(b)(n, b) \\ & + \sigma_1(r)\sigma_2(n)(r, n) + \sigma_1(r)\sigma_2(r)(r, r) + \sigma_1(r)\sigma_2(b)(r, b) \\ & + \sigma_1(b)\sigma_2(n)(r, n) + \sigma_1(b)\sigma_2(r)(r, r) + \sigma_1(b)\sigma_2(b)(r, b) \end{aligned}$$

$$\begin{aligned} e_j[\sigma_1, \sigma_2] = {} & 0.2 \cdot 0.6 \cdot (n, n) + 0.2 \cdot 0 \cdot (n, r) + 0.2 \cdot 0.4 \cdot (n, b) \\ & + 0.5 \cdot 0.6 \cdot (r, n) + 0.5 \cdot 0 \cdot (r, r) + 0.5 \cdot 0.4 \cdot (r, b) \\ & + 0.3 \cdot 0.6 \cdot (r, n) + 0.3 \cdot 0 \cdot (r, r) + 0.3 \cdot 0.4 \cdot (r, b) \end{aligned}$$

Psychological Nash Equilibrium

- Example of an induced second-order expectation: say we have
  - $\sigma_1 = 0.2 \cdot \textit{necklace} + 0.5 \cdot \textit{ring} + 0.3 \cdot \textit{bracelet}$, and
  - $\sigma_2 = 0.6 \cdot \textit{necklace} + 0.4 \cdot \textit{bracelet}$
  - What is $e_j[\sigma_1, \sigma_2]$ induced by (σ_1, σ_2)?

$$
\begin{aligned}
e_j[\sigma_1,\sigma_2] = {} & \sigma_1(n)\sigma_2(n)(n,n) + \sigma_1(n)\sigma_2(r)(n,r) + \sigma_1(n)\sigma_2(b)(n,b) \\
& + \sigma_1(r)\sigma_2(n)(r,n) + \sigma_1(r)\sigma_2(r)(r,r) + \sigma_1(r)\sigma_2(b)(r,b) \\
& + \sigma_1(b)\sigma_2(n)(r,n) + \sigma_1(b)\sigma_2(r)(r,r) + \sigma_1(b)\sigma_2(b)(r,b)
\end{aligned}
$$

$$
\begin{aligned}
e_j[\sigma_1,\sigma_2] = {} & 0.2 \cdot 0.6 \cdot (n,n) + 0.2 \cdot 0 \cdot (n,r) + 0.2 \cdot 0.4 \cdot (n,b) \\
& + 0.5 \cdot 0.6 \cdot (r,n) + 0.5 \cdot 0 \cdot (r,r) + 0.5 \cdot 0.4 \cdot (r,b) \\
& + 0.3 \cdot 0.6 \cdot (r,n) + 0.3 \cdot 0 \cdot (r,r) + 0.3 \cdot 0.4 \cdot (r,b)
\end{aligned}
$$

Psychological Nash Equilibrium

$$
\begin{aligned}
e_j[\sigma_1,\sigma_2] = {} & \sigma_1(n)\sigma_2(n)(n,n) + \sigma_1(n)\sigma_2(r)(n,r) + \sigma_1(n)\sigma_2(b)(n,b) \\
& + \sigma_1(r)\sigma_2(n)(r,n) + \sigma_1(r)\sigma_2(r)(r,r) + \sigma_1(r)\sigma_2(b)(r,b) \\
& + \sigma_1(b)\sigma_2(n)(r,n) + \sigma_1(b)\sigma_2(r)(r,r) + \sigma_1(b)\sigma_2(b)(r,b)
\end{aligned}
$$

$$
\begin{aligned}
e_j[\sigma_1,\sigma_2] = {} & 0.2 \cdot 0.6 \cdot (n,n) + 0.2 \cdot 0 \cdot (n,r) + 0.2 \cdot 0.4 \cdot (n,b) \\
& + 0.5 \cdot 0.6 \cdot (r,n) + 0.5 \cdot 0 \cdot (r,r) + 0.5 \cdot 0.4 \cdot (r,b) \\
& + 0.3 \cdot 0.6 \cdot (r,n) + 0.3 \cdot 0 \cdot (r,r) + 0.3 \cdot 0.4 \cdot (r,b)
\end{aligned}
$$

$$
\begin{aligned}
e_j[\sigma_1,\sigma_2] = {} & 0.12(n,n) + 0.06(n,b) + 0.3(r,n) + 0.15(r,b) + 0.18(b,n) \\
& + 0.09(b,b)
\end{aligned}
$$

Psychological Nash Equilibrium

To do: common belief in rationality + simple belief hierarchy $\rightarrow$ Psychological Nash Equilibrium

Psychological Nash Equilibrium

To do: common belief in rationality + simple belief hierarchy $\rightarrow$ Psychological Nash Equilibrium

- player i expresses 1-fold belief in rationality when for simple belief hierarchy induced by (σ_1, σ_2) we have: $\sigma_2(c_j) > 0$ only when c_j is optimal for $e_j[\sigma_1, \sigma_2]$
- **Step 2:** player i expresses 2-fold belief in rationality: player i believes player j expresses 1-fold belief in rationality: $\sigma_1(c_i) > 0$ only when c_i is optimal for $e_i[\sigma_1, \sigma_2]$.

Psychological Nash Equilibrium

To do: common belief in rationality + simple belief hierarchy → Psychological Nash Equilibrium

- player i expresses 1-fold belief in rationality when for simple belief hierarchy induced by (σ_1, σ_2) we have: $\sigma_2(c_j) > 0$ only when c_j is optimal for $e_j[\sigma_1, \sigma_2]$
- **Step 2:** player i expresses 2-fold belief in rationality: player i believes player j expresses 1-fold belief in rationality: $\sigma_1(c_i) > 0$ only when c_i is optimal for $e_i[\sigma_1, \sigma_2]$.
- **Step 3:** in simple belief hierarchy induced by (σ_1, σ_2) the first-order belief and second-order beliefs "repeat". So if 1-fold and 2-fold are satisfied, so are 3-fold, 4-fold and so on.

Psychological Nash Equilibrium

To do: common belief in rationality + simple belief hierarchy $\rightarrow$ Psychological Nash Equilibrium

- player i expresses 1-fold belief in rationality when for simple belief hierarchy induced by (σ_1, σ_2) we have: $\sigma_2(c_j) > 0$ only when c_j is optimal for $e_j[\sigma_1, \sigma_2]$
- **Step 2:** player i expresses 2-fold belief in rationality: player i believes player j expresses 1-fold belief in rationality: $\sigma_1(c_i) > 0$ only when c_i is optimal for $e_i[\sigma_1, \sigma_2]$.
- **Step 3:** in simple belief hierarchy induced by (σ_1, σ_2) the first-order belief and second-order beliefs "repeat". So if 1-fold and 2-fold are satisfied, so are 3-fold, 4-fold and so on.

Simple belief hierarchy + 1-fold + 2-fold belief in rationality $\rightarrow$ Psychological Nash Equilibrium

Psychological Nash Equilibrium

Definition (Psychological Nash equilibrium)

Consider a probabilistic belief σ_1 about player 1's choice and a probabilistic belief σ_2 about player 2's choice. The pair of beliefs (σ_1, σ_2) is a **psychological Nash Equilibrium** if for both player i, and for every choice $c_i \in C_i$ we have that

$\sigma_i > 0$ only if c_i is optimal for second-order expectation $e_i[\sigma_1, \sigma_2]$

Psychological Nash Equilibrium

Definition (Psychological Nash equilibrium)

Consider a probabilistic belief σ_1 about player 1's choice and a probabilistic belief σ_2 about player 2's choice. The pair of beliefs (σ_1, σ_2) is a **psychological Nash Equilibrium** if for both player i, and for every choice $c_i \in C_i$ we have that

$$\sigma_i > 0 \text{ only if } c_i \text{ is optimal for second-order expectation } e_i[\sigma_1, \sigma_2]$$

- We have now shown by design that CBR + simple belief hierarchy implies psychological Nash equilibrium.
- The other direction is true as well: psychological Nash equilibrium implies a simple belief hierarchy that expresses common belief in rationality

Psychological Nash Equilibrium

Definition (Psychological Nash equilibrium)

Consider a probabilistic belief σ_1 about player 1's choice and a probabilistic belief σ_2 about player 2's choice. The pair of beliefs (σ_1, σ_2) is a **psychological Nash Equilibrium** if for both player i, and for every choice $c_i \in C_i$ we have that

$$\sigma_i > 0 \text{ only if } c_i \text{ is optimal for second-order expectation } e_i[\sigma_1, \sigma_2]$$

- We have now shown by design that CBR + simple belief hierarchy implies psychological Nash equilibrium.
- The other direction is true as well: psychological Nash equilibrium implies a simple belief hierarchy that expresses common belief in rationality
- Let us show this now

Psychological Nash Equilibrium

Overall goal: A psychological Nash equilibrium (PNE) defined by (σ_1, σ_2) implies a simple belief hierarchy generated by (σ_1, σ_2) that expresses common belief in rationality (CBR).

Psychological Nash Equilibrium

Overall goal: A psychological Nash equilibrium (PNE) defined by (σ_1, σ_2) implies a simple belief hierarchy generated by (σ_1, σ_2) that expresses common belief in rationality (CBR).

- **To show Step 1**: the simple belief hierarchy generated by (σ_1, σ_2) for player i expresses 1-fold belief in rationality
  - By definition of a PNE: each choice c_j where $\sigma_2(c_j) > 0$ must be optimal for second-order expectation $e_j[\sigma_1, \sigma_2]$
  - So simple belief hierarchy generated by (σ_1, σ_2) player i indeed only assign positive probability to choices c_j given those are optimal given player j's *believed* second-order expectation. So indeed 1-fold belief in rationality.

Psychological Nash Equilibrium

Overall goal: A psychological Nash equilibrium (PNE) defined by (σ_1, σ_2) implies a simple belief hierarchy generated by (σ_1, σ_2) that expresses common belief in rationality (CBR).

Psychological Nash Equilibrium

Overall goal: A psychological Nash equilibrium (PNE) defined by (σ_1, σ_2) implies a simple belief hierarchy generated by (σ_1, σ_2) that expresses common belief in rationality (CBR).

- **To show Step 2**: To show Step 2: the simple belief hierarchy generated by (σ_1, σ_2) for player i expresses 2-fold belief in rationality
  - By definition of a PNE: each choice c_i where $\sigma_1(c_i) > 0$ must be optimal for second-order expectation $e_i[\sigma_1, \sigma_2]$.
  - Simple belief hierarchy player i believes that player j believes that i has second-order expectation $e_i[\sigma_1, \sigma)2$ (first and second-order beliefs repeat!!).
  - Then i believes that j only assigns positive probability to choices gives those are optimal given i's *believed* second-order expectation. So indeed 2-fold belief in rationality.

Psychological Nash Equilibrium

Overall goal: A psychological Nash equilibrium (PNE) defined by (σ_1, σ_2) implies a simple belief hierarchy generated by (σ_1, σ_2) that expresses common belief in rationality (CBR).

- **To show Step 3**: Simple belief hierarchy → first-order and second-order beliefs repeat if *i* expresses 1-fold and 2-fold belief in rationality, *i* also expresses 3-fold, 4-fold, and so.

Psychological Nash Equilibrium

Overall goal: A psychological Nash equilibrium (PNE) defined by (σ_1, σ_2) implies a simple belief hierarchy generated by (σ_1, σ_2) that expresses common belief in rationality (CBR).

- **To show Step 3**: Simple belief hierarchy → first-order and second-order beliefs repeat if *i* expresses 1-fold and 2-fold belief in rationality, *i* also expresses 3-fold, 4-fold, and so.

Conclusion: if (σ_1, σ_2) is a PNE → the simple belief hierarchy induced by (σ_1, σ_2) expresses CBR.

Psychological Nash Equilibrium

Theorem (8.1: Relation with psychological Nash equilibrium)

Consider the simple belief hierarchy for player i generated by a belief pair (σ_1, σ_2). Then this belief hierarchy expresses common belief in rationality, if and only if, the belief pair (σ_1, σ_2) is a psychological Nash equilibrium.

Psychological Nash Equilibrium

Theorem (8.1: Relation with psychological Nash equilibrium)

Consider the simple belief hierarchy for player i generated by a belief pair (σ_1, σ_2). Then this belief hierarchy expresses common belief in rationality, if and only if, the belief pair (σ_1, σ_2) is a psychological Nash equilibrium.

- In the end, we want to describe/predict behaviour
- We want to characterize choices that are rational under (1) a simple belief hierarchy that (2) expresses CBR.
- With the above Theorem, the following holds:

Psychological Nash Equilibrium

Theorem (8.1: Relation with psychological Nash equilibrium)

Consider the simple belief hierarchy for player i generated by a belief pair (σ_1, σ_2). Then this belief hierarchy expresses common belief in rationality, if and only if, the belief pair (σ_1, σ_2) is a psychological Nash equilibrium.

- In the end, we want to describe/predict behaviour
- We want to characterize choices that are rational under (1) a simple belief hierarchy that (2) expresses CBR.
- With the above Theorem, the following holds:

Theorem (8.2:Relation with psychological Nash equilibrium choices)

A choice is optimal for a simple belief hierarchy that expresses common belief in rationality if and only if that choice is optimal for the second-order expectation induced by a psychological Nash equilibrium.

Literature on psychological Nash equilibrium

- First introduced by Geanakoplos, Pearce and Stacchetti (1989).
  - Static version: psychological Nash equilibrium
  - Dynamic equivalents: subgame perfect psychological equilibrium, sequential psychological equilibrium.
  - All with correct beliefs assumption, AND having beliefs fixed at start.
- Battigalli and Dufwenberg (2009) introduce own version of sequential equilibrium (allowing for endogeneous beliefs, not fixed).
- Battigalli, Corrao and Dufwenberg (2019) consider self-confirming equilibrium for psychological games: psychological Nash equilibrium in dynamic games purely for 'on-path' realizations.

Literature on psychological Nash equilibrium

- First introduced by Geanakoplos, Pearce and Stacchetti (1989).
  - Static version: psychological Nash equilibrium
  - Dynamic equivalents: subgame perfect psychological equilibrium, sequential psychological equilibrium.
  - All with correct beliefs assumption, AND having beliefs fixed at start.
- Battigalli and Dufwenberg (2009) introduce own version of sequential equilibrium (allowing for endogeneous beliefs, not fixed).
- Battigalli, Corrao and Dufwenberg (2019) consider self-confirming equilibrium for psychological games: psychological Nash equilibrium in dynamic games purely for 'on-path' realizations.
- Note 1: most developments of equilibrium concepts in dynamic games.
- Note 2: equilibrium concepts note without scrutiny in psychological games (discuss later).

Psychological Nash Equilibrium: Example

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

- **Goal 1: Find all simple belief hierarchies for you that express CBR**. How to do this?

Psychological Nash Equilibrium: Example

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

- **Goal 1: Find all simple belief hierarchies for you that express CBR**. How to do this?
- Theorem 8.1: these are exactly belief hierarchies generated by a psychological Nash equilibrium $(\sigma_1, \sigma_2) \rightarrow$ Find all psychological Nash equilibria.
- **Task 1:** Find all psychological Nash equilibria

Psychological Nash Equilibrium: Example

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

- **Task 1:** Find all psychological Nash equilibria (PNE)

Psychological Nash Equilibrium: Example

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

- **Task 1:** Find all psychological Nash equilibria (PNE)
- First note: *bracelet* is strictly dominated for you $\rightarrow \sigma_1(\mathit{bracelet}) = 0$ in any PNE.
- Second note: since $\sigma_1(\mathit{bracelet}) = 0$, we have $e_B^2[\sigma_1, \sigma_2](\mathit{bracelet}, \cdot) = 0$.

Psychological Nash Equilibrium: Example

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

- **Task 1:** Find all psychological Nash equilibria (PNE)
- First note: *bracelet* is strictly dominated for you → $\sigma_1(bracelet) = 0$ in any PNE.
- Second note: since $\sigma_1(bracelet) = 0$, we have $e_B^2[\sigma_1, \sigma_2](bracelet, \cdot) = 0$. Then bracelet is not optimal for Barbara in a PNE. So $\sigma_2(bracelet) = 0$.

Psychological Nash Equilibrium: Example

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

- **Task 1:** Find all psychological Nash equilibria (PNE)
- First note: *bracelet* is strictly dominated for you $\rightarrow \sigma_1(\textit{bracelet}) = 0$ in any PNE.
- Second note: since $\sigma_1(\textit{bracelet}) = 0$, we have $e_B^2[\sigma_1, \sigma_2](\textit{bracelet}, \cdot) = 0$. Then bracelet is not optimal for Barbara in a PNE. So $\sigma_2(\textit{bracelet}) = 0$.
- We now look at two cases (depend on game which cases you want to make).
  - Case 1: Start reasoning from assumption that you play necklace in PNE;see if that is possible: Assume $\sigma_1(\textit{necklace}) > 0$.
  - Case 2: Start reasoning from assumption that you play ring in PNE; see if that is possible: Assume $\sigma_1(\textit{ring}) > 0$

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(\textit{necklace}) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(\textit{necklace}) > 0$ with mutual best-responses.

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(necklace) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(necklace) > 0$ with mutual best-responses.

- $\sigma_1(necklace) > 0 \rightarrow$ necklace optimal for induced second-order expectation $e_i[\sigma_1, \sigma_2]$.

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(necklace) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(necklace) > 0$ with mutual best-responses.

- $\sigma_1(necklace) > 0 \rightarrow$ necklace optimal for induced second-order expectation $e_i[\sigma_1, \sigma_2]$.
- Above is only possible when $\sigma_1(ring) > 0$. If $\sigma_1(ring) = 0$, then $u_i(necklace, e_i[\sigma_1, \sigma_2]) < 2 = u_i(ring, e_i[\sigma_1, \sigma_2])$.

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(\textit{necklace}) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(\textit{necklace}) > 0$ with mutual best-responses.

- $\sigma_1(\textit{necklace}) > 0 \rightarrow$ necklace optimal for induced second-order expectation $e_i[\sigma_1, \sigma_2]$.
- Above is only possible when $\sigma_1(\textit{ring}) > 0$. If $\sigma_1(\textit{ring}) = 0$, then $u_i(\textit{necklace}, e_i[\sigma_1, \sigma_2]) < 2 = u_i(\textit{ring}, e_i[\sigma_1, \sigma_2])$.
- So $\sigma_1(\textit{necklace}) > 0$ and $\sigma_1(\textit{ring}) > 0 \rightarrow$ necklace and ring are optimal choice in the same PNE. So they must be optimal under the same second-order expectation $e_i[\sigma_1, \sigma_2]$.
- $u_i(\textit{necklace}, e_i[\sigma_1, \sigma_2]) = u_i(\textit{ring}, e_i[\sigma_1, \sigma_2])$.

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(\textit{necklace}) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(\textit{necklace}) > 0$ with mutual best-responses.

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(necklace) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(necklace) > 0$ with mutual best-responses.

- $u_i(necklace, e_i[\sigma_1, \sigma_2]) = u_i(ring, e_i[\sigma_1, \sigma_2])$.
- $\sigma_1(necklace) \cdot 0 + \sigma_1(ring) \cdot 3 = \sigma_1(necklace) \cdot 2 + \sigma_1(ring) \cdot 2$.

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(necklace) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(necklace) > 0$ with mutual best-responses.

- $u_i(necklace, e_i[\sigma_1, \sigma_2]) = u_i(ring, e_i[\sigma_1, \sigma_2])$.
- $\sigma_1(necklace) \cdot 0 + \sigma_1(ring) \cdot 3 = \sigma_1(necklace) \cdot 2 + \sigma_1(ring) \cdot 2$.
- $3\sigma_1(ring) = 2\sigma_1(necklace)$.

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(necklace) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(necklace) > 0$ with mutual best-responses.

- $u_i(necklace, e_i[\sigma_1, \sigma_2]) = u_i(ring, e_i[\sigma_1, \sigma_2])$.
- $\sigma_1(necklace) \cdot 0 + \sigma_1(ring) \cdot 3 = \sigma_1(necklace) \cdot 2 + \sigma_1(ring) \cdot 2$.
- $3\sigma_1(ring) = 2\sigma_1(necklace)$.
- $3\sigma_1(ring) = 2(1 - \sigma_1(ring))$.

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(necklace) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(necklace) > 0$ with mutual best-responses.

- $u_i(necklace, e_i[\sigma_1, \sigma_2]) = u_i(ring, e_i[\sigma_1, \sigma_2])$.
- $\sigma_1(necklace) \cdot 0 + \sigma_1(ring) \cdot 3 = \sigma_1(necklace) \cdot 2 + \sigma_1(ring) \cdot 2$.
- $3\sigma_1(ring) = 2\sigma_1(necklace)$.
- $3\sigma_1(ring) = 2(1 - \sigma_1(ring))$.
- $5\sigma_1(ring) = 2 \rightarrow \sigma_1(ring) = 0.4$ and $\sigma_1(necklace) = 0.6$

Psychological Nash Equilibrium: Example

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

- Case 1: Assume that $\sigma_1(necklace) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(necklace) > 0$ with mutual best-responses.

- $u_i(necklace, e_i[\sigma_1, \sigma_2]) = u_i(ring, e_i[\sigma_1, \sigma_2])$.
- $\sigma_1(necklace) \cdot 0 + \sigma_1(ring) \cdot 3 = \sigma_1(necklace) \cdot 2 + \sigma_1(ring) \cdot 2$.
- $3\sigma_1(ring) = 2\sigma_1(necklace)$.
- $3\sigma_1(ring) = 2(1 - \sigma_1(ring))$.
- $5\sigma_1(ring) = 2 \rightarrow \sigma_1(ring) = 0.4$ and $\sigma_1(necklace) = 0.6$
- Since $\sigma_1 = 0.6 \cdot necklace + 0.4 \cdot ring$: necklace is preferred over ring by Barbara $\rightarrow \sigma_2 = 1 \cdot necklace$

Psychological Nash Equilibrium: Example

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

- Case 2: Assume that $\sigma_1(ring) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(ring) > 0$ with mutual best-responses.

Psychological Nash Equilibrium: Example

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

- Case 2: Assume that $\sigma_1(ring) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(ring) > 0$ with mutual best-responses.

- $\sigma_1(ring) > 0 \rightarrow$ ring optimal for induced second-order expectation $e_i[\sigma_1, \sigma_2]$.

Psychological Nash Equilibrium: Example

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

- Case 2: Assume that $\sigma_1(ring) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(ring) > 0$ with mutual best-responses.

- $\sigma_1(ring) > 0 \rightarrow$ ring optimal for induced second-order expectation $e_i[\sigma_1, \sigma_2]$.
- Above is only possible when $\sigma_1(necklace) > 0$. If $\sigma_1(necklace) = 0$, then $u_i(ring, e_i[\sigma_1, \sigma_2]) = 0 < 3 = u_i(necklace, e_i[\sigma_1, \sigma_2])$.

Psychological Nash Equilibrium: Example

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

- Case 2: Assume that $\sigma_1(ring) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(ring) > 0$ with mutual best-responses.

- $\sigma_1(ring) > 0 \rightarrow$ ring optimal for induced second-order expectation $e_i[\sigma_1, \sigma_2]$.
- Above is only possible when $\sigma_1(necklace) > 0$. If $\sigma_1(necklace) = 0$, then $u_i(ring, e_i[\sigma_1, \sigma_2]) = 0 < 3 = u_i(necklace, e_i[\sigma_1, \sigma_2])$.
- So $\sigma_1(ring) > 0$ and $\sigma_1(necklace) > 0$. Exactly like Case 1

Psychological Nash Equilibrium: Example

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

- Case 2: Assume that $\sigma_1(ring) > 0$.
- **We want to show**: that there exists (σ_1, σ_2) with $\sigma_1(ring) > 0$ with mutual best-responses.

- $\sigma_1(ring) > 0 \rightarrow$ ring optimal for induced second-order expectation $e_i[\sigma_1, \sigma_2]$.
- Above is only possible when $\sigma_1(necklace) > 0$. If $\sigma_1(necklace) = 0$, then $u_i(ring, e_i[\sigma_1, \sigma_2]) = 0 < 3 = u_i(necklace, e_i[\sigma_1, \sigma_2])$.
- So $\sigma_1(ring) > 0$ and $\sigma_1(necklace) > 0$. Exactly like Case 1
- **Conclusion**: $\sigma_1 = 0.6 \cdot necklace + 0.4 \cdot ring$ and $\sigma_2 = 1 \cdot necklace$ is unique PNE.

Psychological Nash Equilibrium: Example

- See drawing on board for beliefs diagram belonging to the simple belief hierarchy generated by (σ_1, σ_2) that expresses CBR.
- This beliefs diagram is the unique one representing simple belief hierarchies that express CBR in this game.

Psychological Nash Equilibrium: Example

- See drawing on board for beliefs diagram belonging to the simple belief hierarchy generated by (σ_1, σ_2) that expresses CBR.
- This beliefs diagram is the unique one representing simple belief hierarchies that express CBR in this game.
- Note: surprise of degree 0.6 is maximum possible. This happens when you choose ring.
- Under non-simple belief hierarchies surprise of degree 1 is possible.
- Difference due to correct beliefs assumption.

Is psychological Nash equilibrium reasonable?

- Correct beliefs assumption has its critics
- Justification in standard games: learning from repeated interactions → choices and payoffs observable → convergence to equilibrium.

Is psychological Nash equilibrium reasonable?

- Correct beliefs assumption has its critics
- Justification in standard games: learning from repeated interactions → choices and payoffs observable → convergence to equilibrium.
- Issue 1: beliefs of opponent's are not observable → not all that is relevant can be learned → convergence may never happen. (see e.g. Aina et al. (2020) and Dhami and Wei (2023)).

Is psychological Nash equilibrium reasonable?

- Correct beliefs assumption has its critics
- Justification in standard games: learning from repeated interactions → choices and payoffs observable → convergence to equilibrium.
- Issue 1: beliefs of opponent's are not observable → not all that is relevant can be learned → convergence may never happen. (see e.g. Aina et al. (2020) and Dhami and Wei (2023)).
- Issue 2: in psychological games belief are part of the structure of the game (see definition).
  - Impose restrictions on beliefs → impose restrictions on which structures game can represent (Mourmans, 2017)

Is psychological Nash equilibrium reasonable?

- Consider a player i in a generic two-player *surprise game*, which we define here as follows:
  - player i only has **two choices**: $C_i := \{c_i^1, c_i^2\}$,
  - **Surprise motivation**: The preference for choice a choice c_i decreases for player i if it is believed that player j has a higher belief that c_i will be chosen: $u_i(c_i, c_j, e_i^2) = 1 - e_i^2(\cdot, c_i) - \alpha_{c_i} e_i^2(\cdot, c_i)$ where $\alpha_{c_i} > 0$, and
  - **Preference reversal (non-triviality)**: for each c_i for player i we have the following: there is a $\hat{p} \in (0,1)$ such that when $e_i^2(\cdot, c_i) > \hat{p}$ we have that $u_i(c_i, c_j, e_i^2) < u_i(c_i', c_j, e_i^2)$, and when $e_i^2(\cdot, c_i) < \hat{p}$ we have that $u_i(c_i, e_i^2) > u_i(c_i', e_i^2)$.

Is psychological Nash equilibrium reasonable?

- Consider a player i in a generic two-player *surprise game*, which we define here as follows:
  - player i only has **two choices**: $C_i := \{c_i^1, c_i^2\}$,
  - **Surprise motivation**: The preference for choice a choice c_i decreases for player i if it is believed that player j has a higher belief that c_i will be chosen:
    $u_i(c_i, c_j, e_i^2) = 1 - e_i^2(\cdot, c_i) - \alpha_{c_i} e_i^2(\cdot, c_i)$ where $\alpha_{c_i} > 0$, and
  - **Preference reversal (non-triviality)**: for each c_i for player i we have the following: there is a $\hat{p} \in (0,1)$ such that when $e_i^2(\cdot, c_i) > \hat{p}$ we have that $u_i(c_i, c_j, e_i^2) < u_i(c_i', c_j, e_i^2)$, and when $e_i^2(\cdot, c_i) < \hat{p}$ we have that $u_i(c_i, e_i^2) > u_i(c_i', e_i^2)$.

Proposition: [Full surprise in a psychological Nash equilibrium is not possible] There is no Psychological Nash equilibrium such that: a choice c_i is optimal for player i while $e_i[\sigma_1, \sigma_2](\cdot, c_i) = 0$.

Is psychological Nash equilibrium reasonable?

Proposition: [Full surprise in a psychological Nash equilibrium is not possible] In any psychological Nash equilibrium of a two-player surprise game, it is never the case that a choice c_i is optimal for player i while $e_i[\sigma_1, \sigma_2](\cdot, c_i) = 0$.

- Proof by contradiction
- Consider a PNE characterized by a pair of beliefs (σ_1, σ_2) where $\sigma_1(c_i) = 0$ and where choice c_i is optimal.
- Then $e_i[\sigma_1, \sigma_2](\cdot, c_i) = 0$, and $e_i[\sigma_1, \sigma_2](\cdot, c_i') = 1$.
- If $e_i[\sigma_1, \sigma_2](\cdot, c_i) = 0$, then $e_i[\sigma_1, \sigma_2](\cdot, c_i) < \hat{p}$.
- Then $u_i(c_i, e_i^2) > u_i(c_i', e_i^2)$. This means that c_i' is not optimal for $e_i[\sigma_1, \sigma_2]$.
- But then $e_i[\sigma_1, \sigma_2](\cdot, c_i') = 0 \neq 1$. Contradiction.

Symmetric belief hierarchies: recap

Symmetric belief hierarchies: recap

- The idea of a belief hierarchy does not change from standard games to psychological games:
- Beliefs about choices, beliefs about beliefs about choices, beliefs about beliefs about beliefs about choices, and so.
- Therefore, the idea of simple belief hierarchies or **symmetric belief hierarchies** also do not change.

Symmetric belief hierarchies: recap

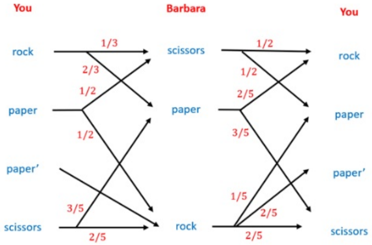
- The idea of a belief hierarchy does not change from standard games to psychological games:
- Beliefs about choices, beliefs about beliefs about choices, beliefs about beliefs about beliefs about choices, and so.
- Therefore, the idea of simple belief hierarchies or **symmetric belief hierarchies** also do not change.
- Symmetric beliefs: certain symmetry between beliefs you have about your opponent's choices and the belief you have about your opponent's belief about your choices.
- Key words: weighted beliefs diagram, symmetric counterpart and symmetric weighted beliefs diagram

Symmetric belief hierarchies: recap

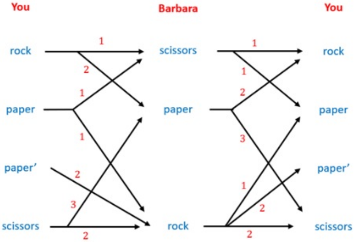
Reminder of Day 3

You	*rock*	*paper*	*scissors*	*diamond*
rock	1	3	4	1
paper	4	1	3	4
scissors	3	4	1	3
bomb	4	0	1	1

Barbara	*rock*	*paper*	*scissors*	*bomb*
rock	1	3	4	0
paper	4	1	3	4
scissors	3	4	1	1
diamond	1	3	4	1



beliefs diagram induced by



symmetric weighted beliefs diagram

All belief hierarchies are symmetric

Symmetric belief hierarchies: definition

Definition (Symmetric belief hierarchy)

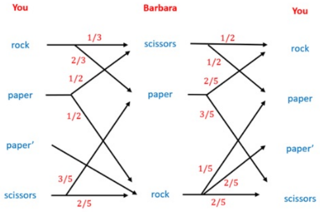
(a) A **weighted beliefs diagram** starts from a beliefs diagram, removes the probabilities at the forked arrows (if there are any), and assigns to every arrow a from a choice c_i to an opponent's choice c_j some positive weight, which we call $w(a)$.
(b) Consider an arrow a from a choice c_i to an opponent's choice c_j. The **symmetric counterpart** to arrow a is the arrow from choice c_j to c_i.
(c) A weighted beliefs diagram is **symmetric** if for every a, the symmetric counterpart is also part of the diagram and carries the same weight as a.
(d) The weighted beliefs diagram induces a (normal) beliefs diagram in which the probability of an arrow a leaving choice c_i is equal to

$$p(a) = \frac{w(a)}{\sum_{\text{arrows a' leaving} c_i'} w(a')}.$$

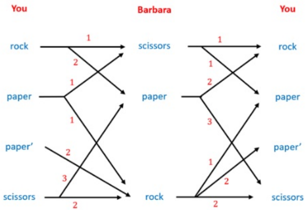
(e) A belief hierarchy is **symmetric** if it is part of a beliefs diagram that is induced by a symmetric weighted beliefs diagram.

Common prior: recap

Reminder of Day 3



beliefs diagram induced by **symmetric weighted beliefs diagram**



All belief hierarchies are induced by the following common prior on choice-type combinations:

π	$(rock,\ t_2^r)$	$(paper,\ t_2^p)$	$(scissors,\ t_2^s)$
$(rock,\ t_1^r)$	0	2/12	1/12
$(paper,\ t_1^p)$	1/12	0	1/12
$(paper,\ \hat{t}_1^p)$	2/12	0	0
$(scissors,\ t_1^s)$	2/12	3/12	0

Common prior: recap

Definition (Common prior)

Consider a beliefs diagram in choice-type representation, with associated sets of types T_i for every player i. Let $C \times T$ be the corresponding set of all choice-type combinations.

(a) A **common prior on choice-type combinations** is a probability distribution π that assigns to every choice-type combination (c, t) in $C \times T$ a probability $\pi(c, t)$

(b) The beliefs diagram is **induced by a common prior** π on $C \times T$, if for every combination $((c_i, t_i), (c_j, t_j))$ and every player i, the corresponding arrow a from (c_i, t_i) to (c_j, t_j) is present exactly when $\pi((c_i, t_i), (c_j, t_j)) > 0$ and the probability of the arrow is equal to

$$p(a) = \frac{\pi((c_i, t_i), (c_j, t_j))}{\pi(c_i, t_i)}.$$

(c) A belief hierarchy is **induced by a common prior** π on choice-type combinations if its is part of a beliefs diagram that is induced by π.

Common prior: recap

- In standard games: symmetric belief hierarchies are exactly those belief hierarchies induced by a common prior.
- In psychological games: ideas of belief hierarchies, symmetric and common priors are exactly the same →
- Also in psychological games: symmetric belief hierarchies are exactly those belief hierarchies induced by a common prior.

Common prior: recap

- In standard games: symmetric belief hierarchies are exactly those belief hierarchies induced by a common prior.
- In psychological games: ideas of belief hierarchies, symmetric and common priors are exactly the same →
- Also in psychological games: symmetric belief hierarchies are exactly those belief hierarchies induced by a common prior.
- We will show: a belief hierarchy is symmetric and expresses common belief in rationality if and only if the belief hierarchy is induced by a psychological correlated equilibrium.

Leading example: Dinner with a huge preference for surprise

You	(b, b)	(b, w)	(w, b)	(w, w)
black	0	0	0	8
white	2	2	2	2

Bar-bara	(b, b)	(b, w)	(w, b)	(w, w)
black	2	2	2	2
white	8	0	0	0

Table 2: *Decision Problems for 'Dinner with huge preference for surprise'*

- 'Black and White' dinner party.
- You prefer to wear *white*, Barbara prefers to wear *black*.
- Only exception: you have a huge preference to wear *black* if you believe to surprise Barbara with that choice; Barbara has huge to wear *white* if she believes to surprise you with that choice.

Psychological correlated equilibrium

You	(b,b)	(b,w)	(w,b)	(w,w)
black	0	0	0	8
white	2	2	2	2

Bar-bara	(b,b)	(b,w)	(w,b)	(w,w)
black	2	2	2	2
white	8	0	0	0

Table 3: *Decision Problems for 'Dinner with huge preference for surprise'*

Goal: what do we impose on common prior π if we assume symmetric belief hierarchy + CBR? We show by leading example

- Consider symmetric belief hierarchy β_i induced by common prior π on choice-type combinations $C \times T$.
- Assume that in the beliefs diagram in choice-type combinations β_i starts at some pair (c_i^*, t_i^*).
- Assume β_i expresses CBR.

Psychological correlated equilibrium

You	(b,b)	(b,w)	(w,b)	(w,w)
black	0	0	0	8
white	2	2	2	2

Barbara	(b,b)	(b,w)	(w,b)	(w,w)
black	2	2	2	2
white	8	0	0	0

Table 4: *Decision Problems for 'Dinner with huge preference for surprise'*

Goal: what do we impose on common prior π if we assume symmetric belief hierarchy + CBR?

- **Step 1:** If β_i expresses CBR, it expresses 1-fold: if β_i in first-order belief assigns positive prob to a pair (c_j^*, t_j^*) then c_j^* must be optimal given what player i believes is player j's second-order expectation conditional on (c_j^*, t_j^*): $e_j^{2,*}$

Psychological correlated equilibrium

You	(b,b)	(b,w)	(w,b)	(w,w)
black	0	0	0	8
white	2	2	2	2

Bar-bara	(b,b)	(b,w)	(w,b)	(w,w)
black	2	2	2	2
white	8	0	0	0

Table 4: *Decision Problems for 'Dinner with huge preference for surprise'*

Goal: what do we impose on common prior π if we assume symmetric belief hierarchy + CBR?

- **Step 1:** If β_i expresses CBR, it expresses 1-fold: if β_i in first-order belief assigns positive prob to a pair (c_j^*, t_j^*) then c_j^* must be optimal given what player i believes is player j's second-order expectation conditional on (c_j^*, t_j^*): $e_j^{2,*}$
- What is second-order expectation $e_j^{2,*}$?

Psychological correlated equilibrium

You	(b,b)	(b,w)	(w,b)	(w,w)
black	0	0	0	8
white	2	2	2	2

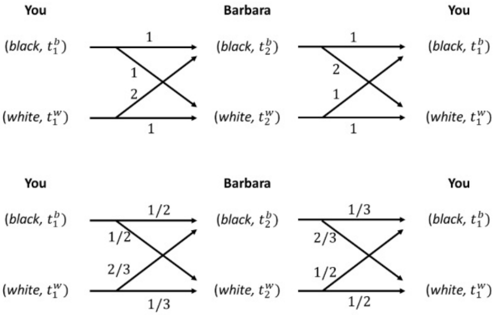
Bar-bara	(b,b)	(b,w)	(w,b)	(w,w)
black	2	2	2	2
white	8	0	0	0

Table 4: *Decision Problems for 'Dinner with huge preference for surprise'*

Goal: what do we impose on common prior π if we assume symmetric belief hierarchy + CBR?

- **Step 1:** If β_i expresses CBR, it expresses 1-fold: if β_i in first-order belief assigns positive prob to a pair (c_j^*, t_j^*) then c_j^* must be optimal given what player i believes is player j's second-order expectation conditional on (c_j^*, t_j^*): $e_j^{2,*}$
- What is second-order expectation $e_j^{2,*}$? Let's explore by example.

Psychological correlated equilibrium



Common prior π below induces weighted symmetric beliefs diagram above

	$(black, t_2^b)$	$(white, t_w^2)$
$(black, t_1^b)$	0.2	0.2
$(white, t_1^w)$	0.4	0.2

Psychological correlated equilibrium

- **Step 1:** If β_i expresses CBR, it expresses 1-fold: if β_i in first-order belief assigns positive prob to a pair (c_j^*, t_j^*) then c_j^* must be optimal given what player i believes is player j's second-order expectation conditional on (c_j^*, t_j^*): $e_j^{2,*}$.

- Let (c_i^*, t_i^*) be $(black, t_1^b)$ be the starting point.
- You assign prob 1/2 to $(black, t_2^b)$.
- Then You must believe that Barbara's choice *black* is optimal given her second-order expectation conditional on $(black, t_2^b)$.
- What is this second-order expectation? $\rightarrow e_2(\cdot|\pi, (black_2, t_2^b))$

Psychological correlated equilibrium

The second order expectation conditional on pair $(black, t_2^b)$ is $e_2(\cdot|\pi, (black_2, t_2^b))$, where:

- $e_2((black_1, t_1^b), (black_2, t_2^b)|\pi, (black_2, t_2^b)) = 1/3 \cdot 1/2 = 1/6,$
- $e_2((black_1, t_1^b), (white_2, t_2^w)|\pi, (black_2, t_2^b)) = 1/3 \cdot 1/2 = 1/6,$
- $e_2((white_1, t_1^b), (black_2, t_2^b)|\pi, (black_2, t_2^b)) = 2/3 \cdot 2/3 = 4/9,$
- $e_2((white_1, t_1^b), (white_2, t_2^w)|\pi, (black_2, t_2^b)) = 2/3 \cdot 1/3 = 2/9.$

- Expected utility Barbara from choosing *white*: $1/6 \cdot 8 + 1/6 \cdot 0 + 4/9 \cdot 0 + 2/9 \cdot 0 = 1/6 \cdot 8 = 8/6 < 2.$
- *black* for Barbara is indeed optimal

Psychological correlated equilibrium

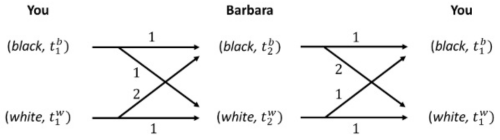
The second order expectation conditional on pair $(black, t_2^b)$ is $e_2(\cdot|\pi, (black_2, t_2^b))$, where:

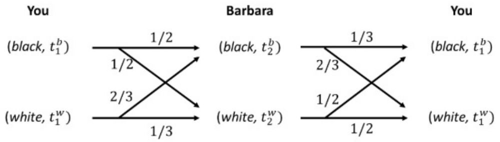
- $e_2((black_1, t_1^b), (black_2, t_2^b)|\pi, (black_2, t_2^b)) = 1/3 \cdot 1/2 = 1/6$,
- $e_2((black_1, t_1^b), (white_2, t_2^w)|\pi, (black_2, t_2^b)) = 1/3 \cdot 1/2 = 1/6$,
- $e_2((white_1, t_1^b), (black_2, t_2^b)|\pi, (black_2, t_2^b)) = 2/3 \cdot 2/3 = 4/9$,
- $e_2((white_1, t_1^b), (white_2, t_2^w)|\pi, (black_2, t_2^b)) = 2/3 \cdot 1/3 = 2/9$.

- Expected utility Barbara from choosing *white*: $1/6 \cdot 8 + 1/6 \cdot 0 + 4/9 \cdot 0 + 2/9 \cdot 0 = 1/6 \cdot 8 = 8/6 < 2$.
- *black* for Barbara is indeed optimal

- How to generalize this using only the common prior?
- How do we get $e_j((c_j, t_j), (c_i, t_i)|\pi, (c_i^*, t_i^*))$ in general?

Psychological correlated equilibrium

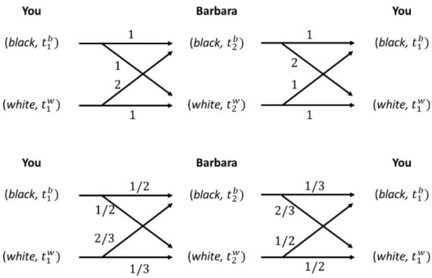




Common prior π below induces weighted symmetric beliefs diagram above

	$(black, t_2^b)$	$(white, t_w^2)$
$(black, t_1^b)$	0.2	0.2
$(white, t_1^w)$	0.4	0.2

Psychological correlated equilibrium



- Focus on: $e_2((white_1, t_1^b), (white_2, t_2^w)|\pi, (black_2, t_2^b)) = 2/3 \cdot 1/3 = 2/9$.
- We have $2/3 = \pi((white_1, t_1^w)|black_2, t_2^b) = \frac{0.4}{0.2+0.4}$, and
- $1/3 = \pi((white_2, t_2^w)|(white_1, t_1^w)) = \frac{0.2}{0.4+0.2}$.

So: $e_2((white_1, t_1^w), (white_2, t_2^w)|\pi, (black_2, t_2^b)) = \pi((white_1, t_1^w)|(black_2, t_2^b)) \cdot \pi((white_2, t_2^w)|(white_1, t_1^w))$

Psychological correlated equilibrium

- In general, assume we have symmetric belief hierarchy generated by common prior π.
- The second-order expectation conditional on (c_i^*, t_i^*) is given by $e_i(\cdot|\pi, (c_i^*, t_i^*))$
- $e_i((c_j, t_j), (c_i, t_i)|\pi, (c_i^*, t_i^*)) := \pi((c_j, t_j)|(c_i^*, t_i^*)) \cdot \pi((c_i, t_i)|(c_j, t_j))$,for every pair (c_i, t_i) for i and every pair (c_j, t_j) for j.
- We have now defined what the second-order expectation conditional on (c_j^*, t_j^*) is.
- Let us go back to **Step 1**

Psychological correlated equilibrium

- **Step 1:** If β_i expresses CBR, it expresses 1-fold: if β_i in first-order belief assigns positive prob to a pair (c_j^*, t_j^*) then c_j^* must be optimal given what player i believes is player j's conditional second-order expectation $e_j(\cdot|\pi, (c_j^*, t_j^*))$.

Psychological correlated equilibrium

- **Step 1:** If β_i expresses CBR, it expresses 1-fold: if β_i in first-order belief assigns positive prob to a pair (c_j^*, t_j^*) then c_j^* must be optimal given what player i believes is player j's conditional second-order expectation $e_j(\cdot|\pi, (c_j^*, t_j^*))$.
- **Step 2:** If β_i expresses CBR, it expresses 2-fold: i believes that j believes in i's rationality.
- Suppose that in the belief hierarchy β_i player i believes that j assigns positive probability to the pair (c_i, t_i).
- Then c_i is optimal for induced second-order expectation $e_i(\cdot|\pi, (c_i, t_i))$.

Psychological correlated equilibrium

- **Step 1:** If β_i expresses CBR, it expresses 1-fold: if β_i in first-order belief assigns positive prob to a pair (c_j^*, t_j^*) then c_j^* must be optimal given what player i believes is player j's conditional second-order expectation $e_j(\cdot|\pi, (c_j^*, t_j^*))$.
- **Step 2:** If β_i expresses CBR, it expresses 2-fold: i believes that j believes in i's rationality.
- Suppose that in the belief hierarchy β_i player i believes that j assigns positive probability to the pair (c_i, t_i).
- Then c_i is optimal for induced second-order expectation $e_i(\cdot|\pi, (c_i, t_i))$.

- Repeat Step 1 and Step 2 for every starting point (c_i^*, t_j^*) in common prior.
- A common prior π on choice-type combinations with the above properties we call psychological correlated equilibrium.

Psychological correlated equilibrium

Definition (Psychological correlated equilibrium)

A common prior π on choice-type combinations is a **psychological correlated equilibrium** if for every player i, and every choice-type pair (c_i, t_i) with $\pi(c_i, t_i) > 0$, the choice c_i is optimal for the induced second-order expectation $e_i(\cdot|\pi, (c_i, t_i))$ of player i conditional on his choice-type pair (c_i, t_i).

Psychological correlated equilibrium

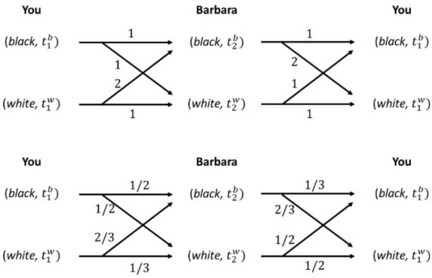
Definition (Psychological correlated equilibrium)

A common prior π on choice-type combinations is a **psychological correlated equilibrium** if for every player i, and every choice-type pair (c_i, t_i) with $\pi(c_i, t_i) > 0$, the choice c_i is optimal for the induced second-order expectation $e_i(\cdot|\pi, (c_i, t_i))$ of player i conditional on his choice-type pair (c_i, t_i).

In easy terms

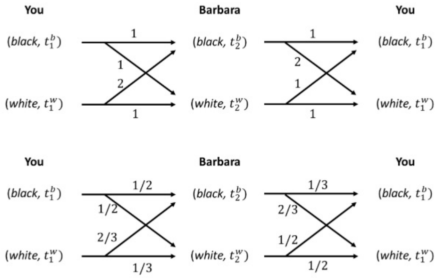
- From π one can derive conditional second-order expectations for every choice-type for a player that appears in the common prior by looking at the conditional beliefs $\pi((c_j, t_j)|(c_i, t_i))$.
- If for every choice-type pair assigned positive probability in the common prior, the choice is optimal for the induced second-order expectation, then we have a (psychological) correlated equilibrium.
- Only difference with standard games: optimal against *second*-order expectations.

Psychological correlated equilibrium: Example



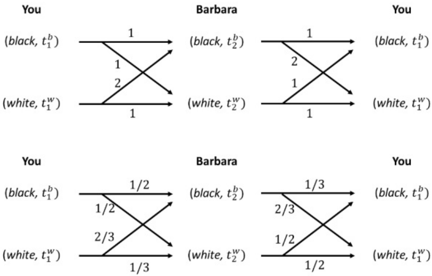
- Earlier: choice $black_2$ for Barbara optimal given induced second-order expectation $e_2(\cdot|black_2, t_2^b)$.

Psychological correlated equilibrium: Example



- Earlier: choice $black_2$ for Barbara optimal given induced second-order expectation $e_2(\cdot|black_2, t_2^b)$.
- PCE: for *all* choice-type combinations in common prior assigned positive probability, we have that choice is optimal for the induced second-order expectation.
- Let's check this now.

Psychological correlated equilibrium: Example



	$(black, t_2^b)$	$(white, t_w^2)$
$(black, t_1^b)$	0.2	0.2
$(white, t_1^w)$	0.4	0.2

Reminder: $e_i((c_j, t_j), (c_i, t_i)|\pi, (c_i^*, t_i^*)) :=$
$\pi((c_j, t_j)|(c_i^*, t_i^*)) \cdot \pi((c_i, t_i)|(c_j, t_j))$

Psychological correlated equilibrium: Example

The second order expectation conditional on pair $(white_2, t_2^w)$ is $e_2(\cdot|\pi, (white_2, t_2^b))$, where:

- $e_2((black_1, t_1^b), (black_2, t_2^b)|\pi, (white_2, t_2^w)) = 1/2 \cdot 1/2 = 1/4$,
- $e_2((black_1, t_1^b), (white_2, t_2^w)|\pi, (white_2, t_2^w)) = 1/2 \cdot 1/2 = 1/4$,
- $e_2((white_1, t_1^w), (black_2, t_2^b)|\pi, (white_2, t_2^w)) = 1/2 \cdot 2/3 = 2/6$,
- $e_2((white_1, t_1^w), (white_2, t_2^w)|\pi, (white_2, t_2^w)) = 1/2 \cdot 1/3 = 1/6$.

We have:

- $u_2(white_2, e_2(\cdot|\pi, (white_2, t_2^b))) =$ $[1/4] \cdot 8 + [1/4] \cdot 0 + [2/6] \cdot 0 + [1/6] \cdot 0 = 2$
- We have $u_2(black_2, e_2(\cdot|\pi, (white_2, t_2^b))) = 2$
- $white_2$ indeed optimal given $e_2(\cdot|\pi, (white_2, t_2^b))$

Psychological correlated equilibrium: Example

The second order expectation conditional on pair $(black_1, t_1^b)$ is $e_1(\cdot|\pi, (black_1, t_1^b))$, where:

- $e_2((black_2, t_2^b), (black_1, t_1^b)|\pi, (black_1, t_1^b)) = 1/2 \cdot 2/3 = 2/6$,
- $e_2((black_2, t_2^b), (white_1, t_1^w)|\pi, (black_1, t_1^b)) = 1/2 \cdot 1/3 = 1/6$,
- $e_2((white_2, t_2^w), (black_1, t_1^b)|\pi, (black_1, t_1^b)) = 1/2 \cdot 1/2 = 1/4$,
- $e_2((white_2, t_2^w), (white_1, t_1^w)|\pi, (black_1, t_1^b)) = 1/2 \cdot 1/2 = 1/4$.

We have:

- $u_1(black_1, e_1(\cdot|\pi, (black_1, t_1^b))) =$
  $[2/6] \cdot 0 + [1/6] \cdot 0 + [1/4] \cdot 0 + [1/4] \cdot 8 = 2$
- We have $u_1(white_1, e_1(\cdot|\pi, (black_1, t_1^b))) = 2$
- $black_1$ indeed optimal given $e_1(\cdot|\pi, (black_1, t_1^b))$

Psychological correlated equilibrium: Example

The second order expectation conditional on pair $(white_1, t_1^w)$ is $e_1(\cdot|\pi, (white_1, t_1^w))$, where:

- $e_2((black_2, t_2^b), (black_1, t_1^b)|\pi, (white_1, t_1^w)) = 2/3 \cdot 1/3 = 2/9$,
- $e_2((black_2, t_2^b), (white_1, t_1^w)|\pi, (white_1, t_1^w)) = 2/3 \cdot 2/3 = 4/9$,
- $e_2((white_2, t_2^w), (black_1, t_1^b)|\pi, (white_1, t_1^w)) = 1/3 \cdot 1/2 = 1/6$,
- $e_2((white_2, t_2^w), (white_1, t_1^w)|\pi, (white_1, t_1^w)) = 1/3 \cdot 1/2 = 1/6$.

We have:

- $u_1(black_1, e_1(\cdot|\pi, (white_1, t_1^w))) =$ $[2/9] \cdot 0 + [4/9] \cdot 0 + [1/6] \cdot 0 + [1/6] \cdot 8 = 8/6$
- We have $u_1(white_1, e_1(\cdot|\pi, (white_1, t_1^w))) = 2$
- $white_1$ indeed optimal given $e_1(\cdot|\pi, (white_1, t_1^w))$

Psychological correlated equilibrium: Example

The second order expectation conditional on pair $(white_1, t_1^w)$ is $e_1(\cdot|\pi, (white_1, t_1^w))$, where:

- $e_2((black_2, t_2^b), (black_1, t_1^b)|\pi, (white_1, t_1^w)) = 2/3 \cdot 1/3 = 2/9$,
- $e_2((black_2, t_2^b), (white_1, t_1^w)|\pi, (white_1, t_1^w)) = 2/3 \cdot 2/3 = 4/9$,
- $e_2((white_2, t_2^w), (black_1, t_1^b)|\pi, (white_1, t_1^w)) = 1/3 \cdot 1/2 = 1/6$,
- $e_2((white_2, t_2^w), (white_1, t_1^w)|\pi, (white_1, t_1^w)) = 1/3 \cdot 1/2 = 1/6$.

We have:

- $u_1(black_1, e_1(\cdot|\pi, (white_1, t_1^w))) =$
  $[2/9] \cdot 0 + [4/9] \cdot 0 + [1/6] \cdot 0 + [1/6] \cdot 8 = 8/6$
- We have $u_1(white_1, e_1(\cdot|\pi, (white_1, t_1^w))) = 2$
- $white_1$ indeed optimal given $e_1(\cdot|\pi, (white_1, t_1^w))$
- **Common prior π is a psychological correlated equilibrium**

PCE and symmetric belief hierarchies + CBR

- We have shown: Symmetric belief hierarchy generated by π + CBR → PCE
- **Opposite way is also true**:
  - In a PCE, all choice-type pairs (c_i, t_i) assigned positive probability to in common prior π are such that c_i is optimal for the induced second-order expectation $e_i^2(\cdot|\pi, (c_i, t_i))$.
  - Then c_i is optimal for type t_i.
  - Belief hierarchy β_i derived from the beliefs diagram induced by π (so symmetric) only has solid arrows going out /assigned positive probability to choice-type pairs that also receive positive probability in common prior π.
  - Then for each choice-type pair (c_i', t_i') assigned positive probability to in the belief hierarchy: c_i optimal for t_i.
  - β_i expresses CBR

PCE and symmetric belief hierarchies + CBR

Theorem (Relation with psychological correlated equilibrium)

A belief hierarchy is symmetric and expresses common belief in rationality, if and only if, the belief hierarchy is induced by a psychological correlated equilibrium.

- In the end, we want to describe/predict behaviour
- We want to characterize choices that are rational under (1) a symmetric belief hierarchy that (2) expresses CBR.

Theorem (Relation with psychological correlated choices)

A choice is optimal for a symmetric belief hierarchy that expresses common belief in rationality, if and only if, the choice is optimal in a psychological correlated equilibrium.

PCE and symmetric belief hierarchies + CBR

Theorem (Relation with psychological correlated equilibrium)

A belief hierarchy is symmetric and expresses common belief in rationality, if and only if, the belief hierarchy is induced by a psychological correlated equilibrium.

- In the end, we want to describe/predict behaviour
- We want to characterize choices that are rational under (1) a symmetric belief hierarchy that (2) expresses CBR.

Theorem (Relation with psychological correlated choices)

A choice is optimal for a symmetric belief hierarchy that expresses common belief in rationality, if and only if, the choice is optimal in a psychological correlated equilibrium.

- Note: simple belief hierarchy expressing CBR always exists → symmetric belief hierarchy expressing CBR always exists.

Canonical psychological correlated equilibrium

Canonical psychological correlated equilibrium

- One theory per choice: if c_i appears in a belief hierarchy, it is only coupled to one type, say $t_i^{c_i}$.
- Theorem 4.3.2 (Book): a symmetric belief hierarchy uses one theory per choice if and only if it is generated by a common prior on choices.
- A common prior π is psychological correlated equilibrium is a canonical psychological correlated equilibrium if it is a common prior on choices.
- Same relation (symmetric belief hierarchy, CBR) - canonical PCE as in standard games:

Canonical psychological correlated equilibrium

Theorem (Relation with canonical PCE)

A belief hierarchy is symmetric, uses one theory per choice and expresses common belief in rationality, if and only if the belief hierarchy is induced by a canonical psychological correlated equilibrium.

- Intuition: same as with regular PCE, just common prior on choices → fix on type $t_i^{c_i}$ per choice c_i

Canonical psychological correlated equilibrium

Theorem (Relation with canonical PCE)

A belief hierarchy is symmetric, uses one theory per choice and expresses common belief in rationality, if and only if the belief hierarchy is induced by a canonical psychological correlated equilibrium.

- Intuition: same as with regular PCE, just common prior on choices → fix on type $t_i^{c_i}$ per choice c_i
- Note: simple belief hierarchy is a symmetric belief hierarchy that uses on theory per choice. And a simple belief hierarchy that expresses CBR always exists (as PNE always exists)
- → a symmetric belief hierarchy that uses on theory per choice and expresses CBR always exists (and thus canonical PCE too).

Possibility of surprise with symmetric belief hierarchies?

- Recall the simple surprise game.
- Let's have a brief look at the board.

Possibility of surprise with symmetric belief hierarchies?

- Recall the simple surprise game.
- Let's have a brief look at the board.

- Say you want to surprise with a choice c_i.
- Say you have a belief hierarchy t_i under which you try to reason for you choice c_i.
- Symmetric belief hierarchy implies that you believe your opponent will mirror you in some sense $\rightarrow$
- You believe your opponent believes at least with some positive probability in choice-type pair (c_i, t_i), otherwise belief-hierarchy is not symmetric.
- You believe your opponent believes with at least some probability you will choose c_i $\rightarrow$ **Full surprise is not possible!**

Comparison of concepts

CBR with ...	**Optimal choices ..**
...	survive I.E. of choices and 2nd-order expectations
symmetric belief hierarchy	are ones optimal in PCE
symmetric belief hierarchy using one theory per choice	are ones optimal in canonical PCE
simple belief hierarchy	are ones optimal in PNE

Table 5: Comparison of concepts

References

- Aina, C., Battigalli, P., and Gamba, A. (2020). Frustration and anger in the ultimatum game: an experiment. *Games and Economic Behavior*, 122, 150–167.
- Battigalli, P., Corrao, R., and Dufwenberg, M. (2019a). Incorporating belief-dependent motivation in games. *Journal of Economic Behaviour and Organization*, 167, 185–218.
- Battigalli, P. and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144, 1–35.
- Dhami, S. and Wei, M. (2023). Norms, Emotions, and Culture in Human Cooperation and Punishment: Theory and Evidence. *CESifo Working Paper No. 10220*.

References

- Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behaviour*, 1(1), 60–79.
- Mourmans, N. (2017). Reasoning about the surprise exam paradox: An application of psychological game theory. *EPICENTER Working Paper No. 20.*