EPICENTER Summer Course on Epistemic Game Theory
Chapter 9: Strong Belief in the Opponents' Rationality

Andrés Perea

EPICENTER
Research Center for Epistemic Game Theory

Maastricht University

July 13, 2022

Strong belief in the opponents' rationality

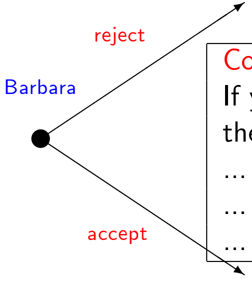
- In the previous chapter, we have discussed the concept of common belief in future rationality.
- Main idea: Whatever you observe in the game, you always believe that your opponents will choose rationally from now on.
- Common belief in this type of reasoning leads to common belief in future rationality.
- It may not be the only plausible way of reasoning in a dynamic game!

Example: Painting Chris' house

Story

- Chris is planning to paint his house tomorrow, and needs someone to help him.
- You and Barbara are both interested. This evening, both of you must come to Chris' house, and whisper a price in his ear. Price must be either 200, 300, 400 or 500 euros.
- Person with lowest price will get the job. In case of a tie, Chris will toss a coin.
- Before you leave for Chris' house, Barbara gets a phone call from a colleague, who asks her to repair his car tomorrow at a price of 350 euros.
- Barbara must decide whether or not to accept the colleague's offer.

	200	300	400	500
200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250



Common belief in future rationality:

If you observe that Barbara has rejected offer, then you believe that

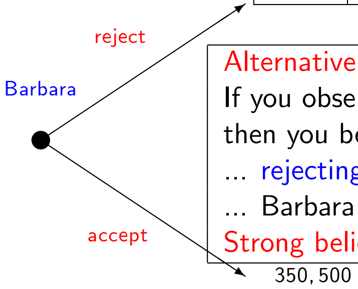
... rejecting offer was a mistake,

... Barbara chooses rationally from now on

... Barbara believes that you choose rationally.

You will choose price 200.

	200	300	400	500
200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250



Alternative way of reasoning:

If you observe that Barbara has rejected offer, then you believe that

... rejecting offer is part of a rational strategy,

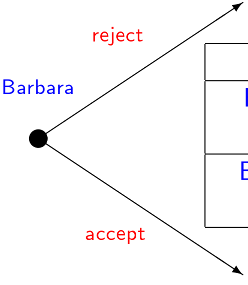
... Barbara will choose price 400.

Strong belief in Barbara's rationality.

You will choose price 300.

- Strong belief in the opponents' rationality:
- If at information set $h \in H_i$, it is possible for player i to believe that each of his opponents is implementing a rational strategy,
- then player i must believe at h that each of his opponents is implementing a rational strategy.
- How can we formalize this idea within an epistemic model?
- Attempt: Consider an epistemic model M, a type t_i and an information set $h \in H_i$.
- If for every opponent there is a type inside M for which there is an optimal strategy leading to h,
- then type t_i must at h only assign positive probability to strategy-type pairs where the strategy is optimal for the type.
- This will not work.

200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250



Types	$T_1 = \{t_1\},\ T_2 = \{t_2\}$
Beliefs for Barbara	$b_1(t_1, \varnothing) = (200, t_2)$ $b_1(t_1, h_1) = (200, t_2)$
Beliefs for you	$b_2(t_2, h_1) = ((reject, 200), t_1)$

Your type t_2 satisfies conditions, but does not strongly believe in Barbara's rationality.

Problem: Not sufficiently many types in epistemic model M.

- To make the definition of strong belief in the opponents' rationality work, we must require that the epistemic model M contains sufficiently many types.
- Consider an epistemic model M, and an information set $h \in H_i$:
- If for every opponent there is a type in some epistemic model M', for which there is an optimal strategy leading to h,
- then M must contain at least one such type for every opponent.

- A strategy s_i is optimal for type t_i is s_i is optimal for t_i at every information set $h \in H_i$ that s_i leads to.

Definition (Strong belief in the opponents' rationality)

Type t_i strongly believes in the opponents' rationality at h if,

whenever we can find a combination of opponents' types in some epistemic model M', for which there is a combination of optimal strategies leading to h,

then

(1) the epistemic model M must contain at least one such combination of opponents' types, and

(2) type t_i must at h only assign positive probability to opponents' strategy-type combinations where the strategy combination leads to h, and the strategies are optimal for the types.

Definition (Strong belief in the opponents' rationality)

Type t_i strongly believes in the opponents' rationality at h if,

whenever we can find a combination of opponents' types in some epistemic model M', for which there is a combination of optimal strategies leading to h,

then

(1) the epistemic model M must contain at least one such combination of opponents' types, and

(2) type t_i must at h only assign positive probability to opponents' strategy-type combinations where the strategy combination leads to h, and the strategies are optimal for the types.

- Based on Battigalli and Siniscalchi (2002).
- Difference: Battigalli and Siniscalchi (2002) assume that the epistemic model M contains all possible belief hierarchies.

Definition (Strong belief in the opponents' rationality)

Type t_i strongly believes in the opponents' rationality at h if,

whenever we can find a combination of opponents' types in some epistemic model M', for which there is a combination of optimal strategies leading to h,

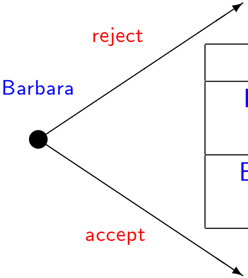
then

(1) the epistemic model M must contain at least one such combination of opponents' types, and

(2) type t_i must at h only assign positive probability to opponents' strategy-type combinations where the strategy combination leads to h, and the strategies are optimal for the types.

- In games with more than two players, if you conclude that player i has chosen irrationally in the past, you may believe that some other player j will choose irrationally in the future.
- Research question: Can you find a definition that does not suffer from this problem?

200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250

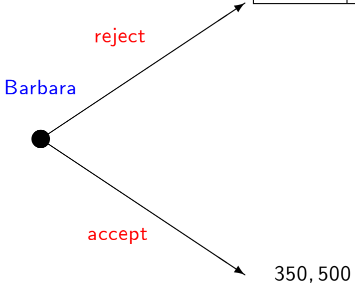


Types	$T_1 = \{t_1\},\ T_2 = \{t_2\}$
Beliefs for Barbara	$b_1(t_1, \varnothing) = (200, t_2)$ $b_1(t_1, h_1) = (200, t_2)$
Beliefs for you	$b_2(t_2, h_1) = ((\textit{reject}, 200), t_1)$

350, 500

Your type t_2 does not strongly believe in Barbara's rationality.

200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250



$$T_1 = \{t_1^a, t_1^r\}, \; T_2 = \{t_2\}$$

$$\begin{aligned}
b_1(t_1^a, \varnothing) &= (300, t_2) \\
b_1(t_1^a, h_1) &= (300, t_2) \\
b_1(t_1^r, \varnothing) &= (500, t_2) \\
b_1(t_1^r, h_1) &= (500, t_2) \\
b_2(t_2, h_1) &= ((\textit{reject}, 400), t_1^r)
\end{aligned}$$

Your type t_2 strongly believes in Barbara's rationality.

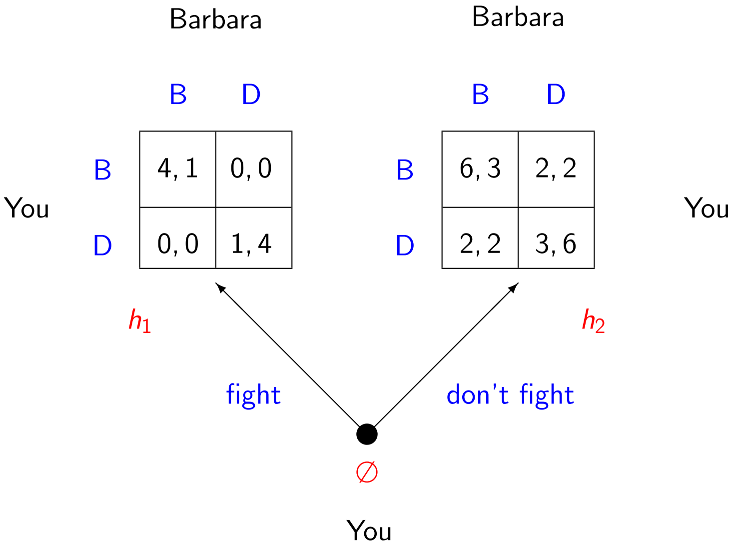
Two-fold strong belief in rationality

- Suppose player i is at information set h, and he reasons about two possible strategies s_j and s_j' for player j:
- strategy s_j is optimal for some type t_j, but not for any type that strongly believes in i's rationality,
- strategy s_j' is optimal for a type t_j that strongly believes in i's rationality.
- Then, according to the idea of strong belief in the opponents' rationality, s_j' seems the more plausible strategy.
- Hence, player i believes at h that player j has chosen s_j', and not s_j.
- Two-fold strong belief in rationality.

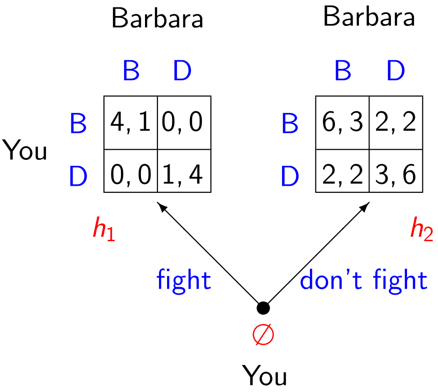
Example: Watching TV with Barbara

Story

- Barbara and you must decide with TV program to watch: Blackadder or Dallas.
- You prefer Blackadder (utility 6) to Dallas (utility 3).
- Barbara prefers Dallas (utility 6) to Blackadder (utility 3).
- You both must write down a program on a piece of paper. If you both write the same program, you will watch it together. Otherwise, you will play a game of cards (utility 2 for both).
- Before writing down a program, you have the option to start a fight with Barbara to convince her to watch your favorite program. This would reduce your utility and Barbara's utility by 2.



- This is a burning money game: See van Damme (1989), Ben-Porath and Dekel (1992) and Shimoji (2002).



Suppose, Barbara strongly believes in your rationality.

Then, at h_1 she will believe that you choose $(fight, B)$.

Hence, Barbara will choose B at h_1.

So, if Barbara strongly believes in your rationality, her only optimal strategies are (B, B) and (B, D).

So, if you express 2-fold strong belief in Barbara's rationality, then you believe that Barbara chooses (B, B) or (B, D).

Hence, you can only rationally choose $(fight, B)$ or $(don't, B)$.

- Two-fold strong belief in rationality:
- Consider an information set h for player i.
- If there is an opponents' strategy-type combination where (a) the opponents' strategy combination leads to h, (b) the strategies are optimal for the types, and (c) the types strongly believe in the opponents' rationality,
- then type t_i must at h only assign positive probability to opponents' strategy-type combinations that satisfy (a), (b) and (c).
- To make this definition work, we must require that the epistemic model M contains sufficiently many types:
- If we can find a combination of opponents' types, in some epistemic model M', that strongly believe in the opponents' rationality, and for which there is a combination of optimal strategies leading to h,
- then the epistemic model M must contain at least one such combination of opponents' types.

Definition (Two-fold strong belief in rationality)

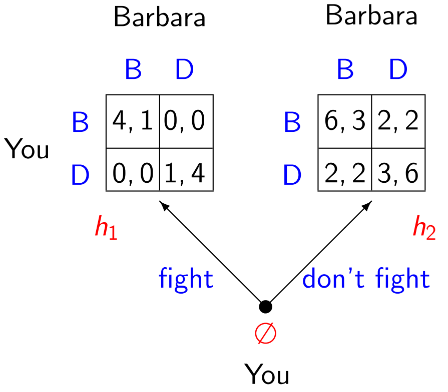
Type t_i expresses 2-fold strong belief in rationality at h if,

whenever we can find a combination of opponents' types, in some epistemic model M', that strongly believe in their opponents' rationality, and for which there is a combination of optimal strategies leading to h,

then

(1) the epistemic model M must contain at least one such combination of opponents' types, and

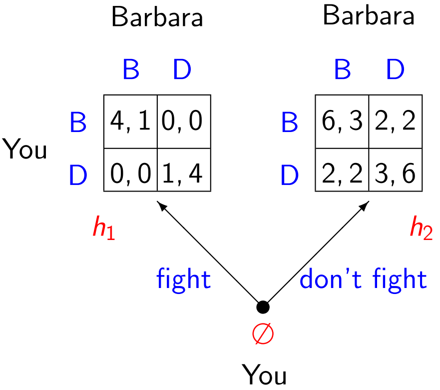
(2) type t_i must at h only assign positive probability to opponents' strategy-type combinations where the strategy combination leads to h, the types strongly believe in their opponents' rationality, and the strategies are optimal for the types.



$$
\begin{array}{lcl}
b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & \\
b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & \\
b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

Show: Your types t_1^{fB} and t_1^{dB} express 2-fold strong belief in rationality.

- Barbara's types t_2^{BB} and t_2^{BD} strongly believe in your rationality.



$$
\begin{array}{lcl}
b_1(t_1^{fB}) & = & ((B,D),\, t_2^{BD}) \\
b_1(t_1^{dB}) & = & ((B,B),\, t_2^{BB}) \\
b_1(t_1^{dD}) & = & ((D,D),\, t_2^{DD}) \\
\hline
b_2(t_2^{BB}, h_1) & = & ((fight, B),\, t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B),\, t_1^{dB}) \\
 & & \\
b_2(t_2^{BD}, h_1) & = & ((fight, B),\, t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D),\, t_1^{dD}) \\
 & & \\
b_2(t_2^{DD}, h_1) & = & ((fight, D),\, t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D),\, t_1^{dD})
\end{array}
$$

Show: Your type t_1^{dD} does not express 2-fold strong belief in rationality.

- Barbara's type t_2^{DD} does not strongly believe in your rationality.

Definition (Common strong belief in rationality)

Type t_i is said to express 1-fold strong belief in rationality if t_i strongly believes in the opponents' rationality.

Say that type t_i expresses k-fold strong belief in rationality at h if, whenever we can find a combination of opponents' types, in some epistemic model M', that express up to $(k-1)$-fold strong belief in rationality, and for which there is a combination of optimal strategies leading to h, then

(1) the epistemic model M must contain at least one such combination of opponents' types, and

(2) type t_i must at h only assign positive probability to opponents' strategy-type combinations where the strategy combination leads to h, the types express up to $(k-1)$-fold strong belief in rationality, and the strategies are optimal for the types.

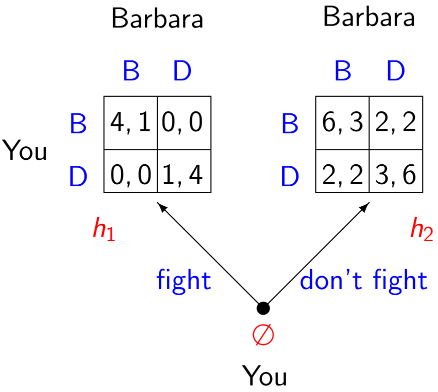
Type t_i expresses common strong belief in rationality if it expresses k-fold strong belief in rationality for every k.

Related literature

- Definition of common strong belief in rationality is based on Battigalli and Siniscalchi (2002).
- Again, difference is that Battigalli and Siniscalchi (2002) assume that the epistemic model M contains all possible belief hierarchies.
- This is a forward induction concept: Whenever possible, you try to explain the past choices made by your opponent.
- In contrast to common belief in future rationality, which is a backward induction concept: You ignore the opponent's past choices, and concentrate solely on the game that lies ahead.

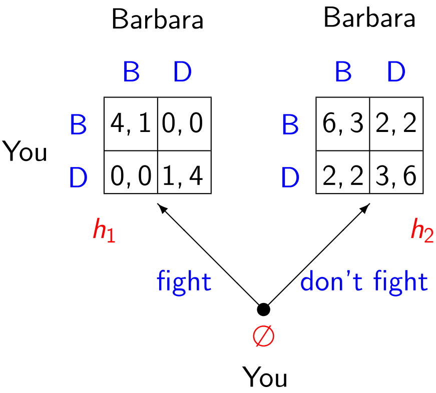
Related literature

- Battigalli and Siniscalchi (2002) show that common strong belief in rationality characterizes the concept of extensive-form rationalizability (Pearce (1984), Battigalli (1997)).
- Reny (1992) proposes a related forward induction concept: explicable equilibrium.
- Sometimes, iterated elimination of weakly dominated strategies is also used as a forward induction concept.
- In the 1980's and early 1990's, several forward induction refinements of sequential equilibrium have been proposed.
- For an overview, see Perea (2001, 2017a).
- Meier and Perea (2022) propose a concept that combines elements from common strong belief in rationality and common belief in future rationality.
- Research question: Application of common strong belief in rationality to models in economics?



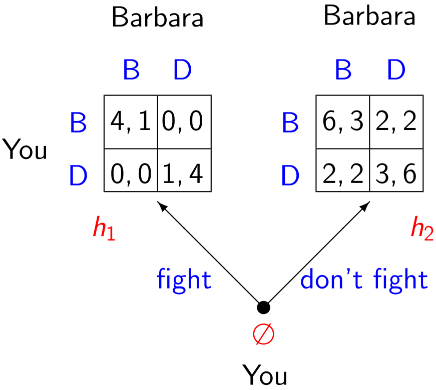
$$
\begin{array}{lcl}
(2)\ b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(2)\ b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(1)\ b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & \\
(1)\ b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

- We know:
- Your types t_1^{fB}, t_1^{dB} and t_1^{dD} express 1-fold strong belief in rationality.
- Your types t_1^{fB} and t_1^{dB} express 2-fold strong belief in rationality, but t_1^{dD} not.
- Barbara's types t_2^{BB} and t_2^{BD} express 1-fold strong belief in rationality, but t_2^{DD} not.



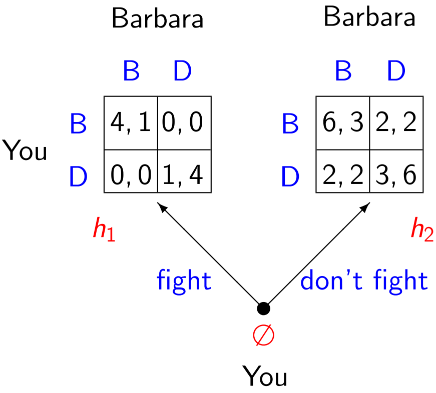
$$\begin{array}{lll}
(2)\ b_1(t_1^{fB}) & = & ((B,D),\, t_2^{BD}) \\
(2)\ b_1(t_1^{dB}) & = & ((B,B),\, t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D),\, t_2^{DD}) \\
\hline
(1)\ b_2(t_2^{BB}, h_1) & = & ((fight, B),\, t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B),\, t_1^{dB}) \\
 & & \\
(1)\ b_2(t_2^{BD}, h_1) & = & ((fight, B),\, t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D),\, t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D),\, t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D),\, t_1^{dD})
\end{array}$$

Show: Barbara's types t_2^{BB} and t_2^{BD} express 2-fold strong belief in rationality.



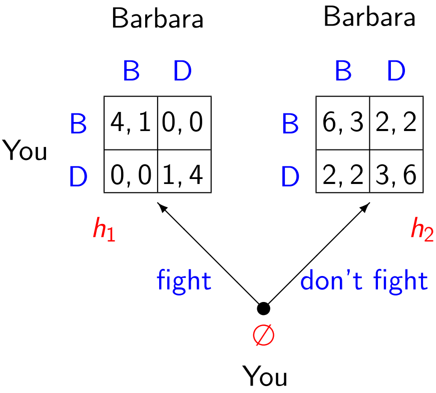
$$
\begin{array}{lcl}
(2)\ b_1(t_1^{fB}) & = & ((B,D),\ t_2^{BD}) \\
(2)\ b_1(t_1^{dB}) & = & ((B,B),\ t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D),\ t_2^{DD}) \\
\hline
(2)\ b_2(t_2^{BB}, h_1) & = & ((fight, B),\ t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B),\ t_1^{dB}) \\
 & & \\
(2)\ b_2(t_2^{BD}, h_1) & = & ((fight, B),\ t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D),\ t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D),\ t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D),\ t_1^{dD})
\end{array}
$$

Show: Your types t_1^{fB} and t_1^{dB} express 3-fold strong belief in rationality.



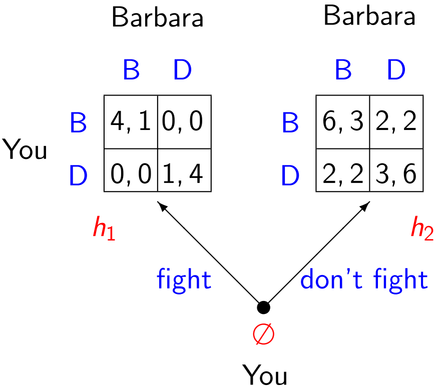
$$
\begin{array}{lcl}
(3)\ b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(3)\ b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(2)\ b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & \\
(2)\ b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dB})
\end{array}
$$

Show: Barbara's type t_2^{BB} expresses 3-fold strong belief in rationality, but her type t_2^{BD} not.



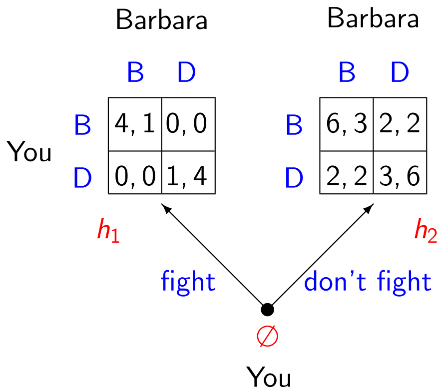
$$
\begin{array}{lcl}
(3)\ b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(3)\ b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(3)\ b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & \\
(2)\ b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

Show: Your type t_1^{dB} expresses 4-fold strong belief in rationality, but your type t_1^{fB} not.



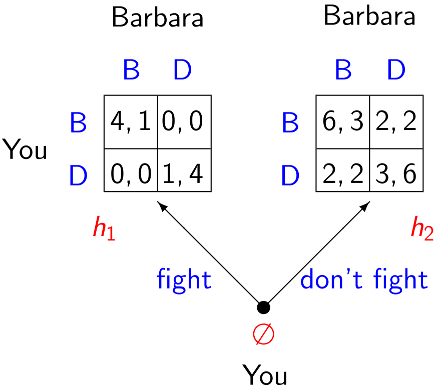
$$
\begin{array}{llll}
(3) & b_1(t_1^{fB}) & = & ((B, D), t_2^{BD}) \\
(4) & b_1(t_1^{dB}) & = & ((B, B), t_2^{BB}) \\
(1) & b_1(t_1^{dD}) & = & ((D, D), t_2^{DD}) \\
\hline
(3) & b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
 & b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & & \\
(2) & b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
 & b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & & \\
(0) & b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
 & b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

Show: Barbara's type t_2^{BB} expresses 4-fold strong belief in rationality.



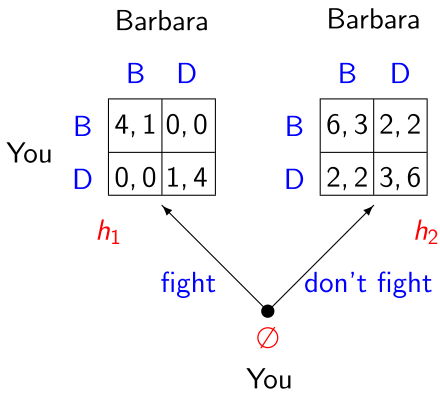
$$
\begin{array}{lll}
(3)\ b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(4)\ b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(4)\ b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & \\
(2)\ b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

Show: Your type t_1^{dB} expresses 5-fold strong belief in rationality.



$$
\begin{array}{llcl}
(3) & b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(5) & b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1) & b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(4) & b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
 & b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
(2) & b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
 & b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
(0) & b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
 & b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

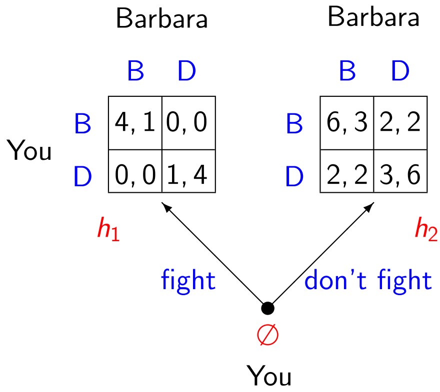
Show: Barbara's type t_2^{BB} expresses 5-fold strong belief in rationality.



$$
\begin{array}{llcl}
(3) & b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(5) & b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1) & b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(5) & b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
 & b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & & \\
(2) & b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
 & b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & & \\
(0) & b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
 & b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

Show: Your type t_1^{dB} and Barbara's type t_2^{BB} express k-fold strong belief in rationality, for every $k \geq 6$.

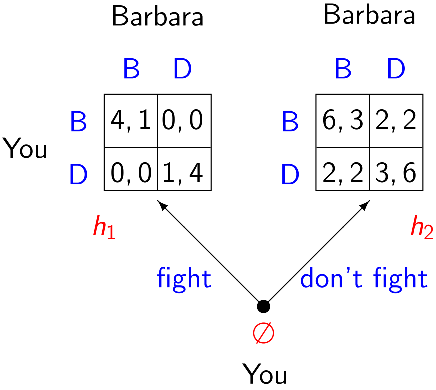
Hence, your type t_1^{dB} and Barbara's type t_2^{BB} express common strong belief in rationality.



$$
\begin{array}{lll}
(3)\ b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(c)\ b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(c)\ b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & \\
(2)\ b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dB})
\end{array}
$$

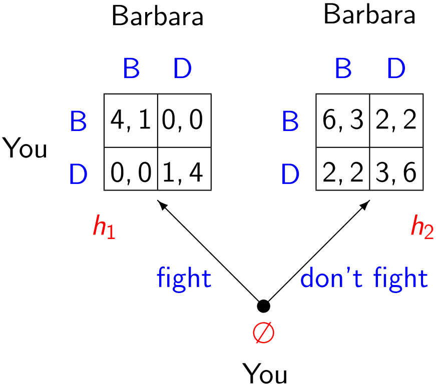
Conclusion: Under common strong belief in rationality, you can only rationally choose $(don't, B)$, and you expect Barbara to choose (B, B).

Hence, under common strong belief in rationality, you expect to be watching your favorite program together, without having to start a fight with Barbara.



$$
\begin{array}{lcl}
(3)\ b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(c)\ b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(c)\ b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & \\
(2)\ b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

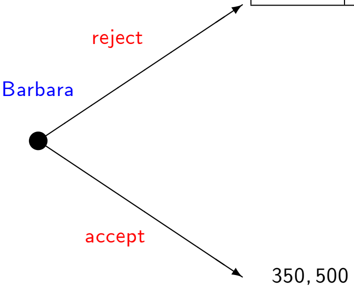
Note: In order to construct types that express common strong belief in rationality, we need to include types in the epistemic model that do not express common strong belief in rationality.



$$
\begin{array}{lcl}
(3)\ b_1(t_1^{fB}) & = & ((B,D), t_2^{BD}) \\
(c)\ b_1(t_1^{dB}) & = & ((B,B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) & = & ((D,D), t_2^{DD}) \\
\hline
(c)\ b_2(t_2^{BB}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BB}, h_2) & = & ((don't, B), t_1^{dB}) \\
 & & \\
(2)\ b_2(t_2^{BD}, h_1) & = & ((fight, B), t_1^{fB}) \\
b_2(t_2^{BD}, h_2) & = & ((don't, D), t_1^{dD}) \\
 & & \\
(0)\ b_2(t_2^{DD}, h_1) & = & ((fight, D), t_1^{fB}) \\
b_2(t_2^{DD}, h_2) & = & ((don't, D), t_1^{dD})
\end{array}
$$

- Your type t_1^{dB} believes that Barbara, at h_1, is wrong about your beliefs.
- Hence, in this game common strong belief in rationality is not compatible with equilibrium reasoning. Perea (2017a) shows that this is a structural fact.

200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250

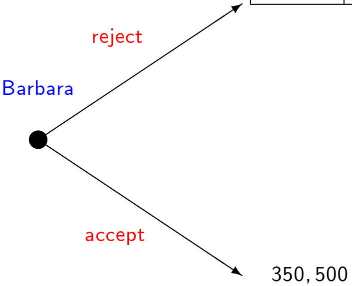


$$T_1 = \{t_1^a, t_1^r\}, \quad T_2 = \{t_2\}$$

$$\begin{array}{lrcl}
\text{(c)} & b_1(t_1^a, \varnothing) & = & (300, t_2) \\
 & b_1(t_1^a, h_1) & = & (300, t_2) \\
 & & & \\
\text{(0)} & b_1(t_1^r, \varnothing) & = & (500, t_2) \\
 & b_1(t_1^r, h_1) & = & (500, t_2) \\
\text{(c)} & b_2(t_2, h_1) & = & ((\textit{reject}, 400), t_1^r)
\end{array}$$

Exercise: Show that t_1^a and t_2 express common strong belief in rationality.

200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250

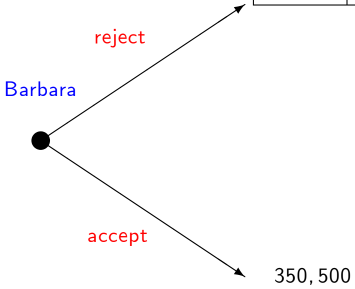


$T_1 = \{t_1^a, t_1^r\}, \ T_2 = \{t_2\}$

$$
\begin{aligned}
\text{(c)}\ b_1(t_1^a, \varnothing) &= (300, t_2) \\
b_1(t_1^a, h_1) &= (300, t_2) \\
\\
\text{(0)}\ b_1(t_1^r, \varnothing) &= (500, t_2) \\
b_1(t_1^r, h_1) &= (500, t_2) \\
\text{(c)}\ b_2(t_2, h_1) &= ((\textit{reject}, 400), t_1^r)
\end{aligned}
$$

Note: In order to construct types that express common strong belief in rationality, we need to include types in the epistemic model that do not express common strong belief in rationality.

200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250



$T_1 = \{t_1^a, t_1^r\}, \; T_2 = \{t_2\}$

$$
\begin{array}{lcl}
\text{(c)} \; b_1(t_1^a, \varnothing) & = & (300, t_2) \\
b_1(t_1^a, h_1) & = & (300, t_2) \\
& & \\
\text{(0)} \; b_1(t_1^r, \varnothing) & = & (500, t_2) \\
b_1(t_1^r, h_1) & = & (500, t_2) \\
\hline
\text{(c)} \; b_2(t_2, h_1) & = & ((\textit{reject}, 400), t_1^r)
\end{array}
$$

- Barbara's type t_1^a believes that you, at h_1, are wrong about her beliefs.

Algorithm

- We wish to find those strategies you can rationally choose under common strong belief in rationality.
- Is there an algorithm that helps us find these strategies?
- Yes. Algorithm is similar in flavor to the backward dominance procedure.

- Important ingredients:
- The full decision problem for player i at h is $\Gamma^0(h) = (S_i(h), S_{-i}(h))$, where $S_i(h)$ is the set of strategies for player i that lead to h, and $S_{-i}(h)$ is the set of opponents' strategy combinations that lead to h.
- A reduced decision problem for player i at h is $\Gamma(h) = (D_i(h), D_{-i}(h))$, where $D_i(h) \subseteq S_i(h)$ and $D_{-i}(h) \subseteq S_{-i}(h)$.

Step 1: 1-fold strong belief in rationality.

- Which strategies can player i rationally choose if he expresses 1-fold strong belief in rationality, that is, strongly believes in the opponents' rationality?
- Consider a type t_i that expresses 1-fold strong belief in rationality.
- Then, at every information set $h \in H_i$:
- if there is a combination of optimal opponents' strategies leading to h,
- then type t_i must at h only assign positive probability to combinations of optimal opponents' strategies leading to h.
- We know: an opponent's strategy s_j is optimal, if and only if, it is not strictly dominated at any full decision problem $\Gamma^0(h')$ where j is active.

- Hence, at every information set $h \in H_i$:
- if there is a combination of opponents' strategies s_j leading to h where s_j is not strictly dominated at any $\Gamma^0(h')$ where j is active,
- then type t_i must at h only assign positive probability to such opponents' strategy combinations.
- Let $\Gamma^1(h)$ be the reduced decision problem at h, obtained from $\Gamma^0(h)$ by eliminating all opponents' strategies s_j which are strictly dominated at some $\Gamma^0(h')$ where j is active,
- unless this would eliminate all opponents' strategy combinations from $\Gamma^0(h)$.
- In the latter case, $\Gamma^1(h) = \Gamma^0(h)$.
- Then, type t_i assigns at h only positive probability to opponents' strategy combinations in $\Gamma^1(h)$.

- Then, type t_i assigns at h only positive probability to opponents' strategy combinations in $\Gamma^1(h)$.
- So, every strategy that is optimal for t_i at h, must not be strictly dominated at $\Gamma^1(h)$.
- Let $\Gamma^2(\emptyset)$ be reduced decision problem at $\emptyset$ which is obtained by eliminating, for every player i, those strategies that are strictly dominated within some reduced decision problem $\Gamma^1(h)$ at which i is active.
- Hence, every optimal strategy for t_i must be in $\Gamma^2(\emptyset)$.
- Conclusion: Every strategy that is optimal for some type that expresses 1-fold strong belief in rationality, must be in $\Gamma^2(\emptyset)$.

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

- Which strategies can player i rationally choose if he expresses up to 2-fold strong belief in rationality?
- Consider a type t_i that expresses up to 2-fold strong belief in rationality. Then, at every information set $h \in H_i$:
- if there is an opponents' combination of strategies leading to h, where every opponents' strategy s_j is optimal for some type t_j that expresses 1-fold strong belief in rationality,
- then type t_i must at h only assign positive probability to such combinations of opponents' strategies.
- We know from Step 1, that every such opponent's strategy s_j is not strictly dominated within any reduced decision problem $\Gamma^1(h')$ where j is active.

- At every information set $h \in H_i$:
- if there is an opponents' combination of strategies leading to h, where every opponents' strategy s_j is optimal for some type t_j that expresses 1-fold strong belief in rationality,
- then type t_i must at h only assign positive probability to such combinations of opponents' strategies.
- We know from Step 1, that every such opponent's strategy s_j is not strictly dominated within any reduced decision problem $\Gamma^1(h')$ where j is active.
- Let $\Gamma^2(h)$ be the reduced decision problem at h, obtained from $\Gamma^1(h)$ by eliminating all opponents' strategies s_j which are strictly dominated at some $\Gamma^1(h')$ where j is active,
- unless this would eliminate all opponents' strategy combinations from $\Gamma^1(h)$.
- In the latter case, $\Gamma^2(h) = \Gamma^1(h)$.
- Then, type t_i assigns at h only positive probability to opponents' strategy combinations in $\Gamma^2(h)$.

- Then, type t_i assigns at h only positive probability to opponents' strategy combinations in $\Gamma^2(h)$.
- So, every strategy that is optimal for t_i at h, must not be strictly dominated at $\Gamma^2(h)$.
- Let $\Gamma^3(\varnothing)$ be reduced decision problem at $\varnothing$ which is obtained by eliminating, for every player i, those strategies that are strictly dominated within some reduced decision problem $\Gamma^2(h)$ at which i is active.
- Hence, every optimal strategy for t_i must be in $\Gamma^3(\varnothing)$.
- Conclusion: Every strategy that is optimal for some type that expresses up to 2-fold strong belief in rationality, must be in $\Gamma^3(\varnothing)$.

Algorithm (Iterated conditional dominance procedure)

Step 1. *At every full decision problem $\Gamma^0(h)$, eliminate for every player i those strategies that are strictly dominated at some full decision problem $\Gamma^0(h')$ at which player i is active, unless this would remove all strategy combinations that lead to h. In the latter case, we remove nothing from $\Gamma^0(h)$. This leads to reduced decision problems $\Gamma^1(h)$ at every information set h.*

Step 2. *At every reduced decision problem $\Gamma^1(h)$, eliminate for every player i those strategies that are strictly dominated at some reduced decision problem $\Gamma^1(h')$ at which player i is active, unless this would remove all strategy combinations that lead to h. In the latter case, we remove nothing from $\Gamma^1(h)$. This leads to new reduced decision problems $\Gamma^2(h)$ at every information set.*

And so on. Continue until no more strategies can be eliminated in this way.

- Algorithm is due to Shimoji and Watson (1998).

Algorithm (Iterated conditional dominance procedure)

Step 1. *At every full decision problem $\Gamma^0(h)$, eliminate for every player i those strategies that are strictly dominated at some full decision problem $\Gamma^0(h')$ at which player i is active, unless this would remove all strategy combinations that lead to h. In the latter case, we remove nothing from $\Gamma^0(h)$. This leads to reduced decision problems $\Gamma^1(h)$ at every information set h.*

Step 2. *At every reduced decision problem $\Gamma^1(h)$, eliminate for every player i those strategies that are strictly dominated at some reduced decision problem $\Gamma^1(h')$ at which player i is active, unless this would remove all strategy combinations that lead to h. In the latter case, we remove nothing from $\Gamma^1(h)$. This leads to new reduced decision problems $\Gamma^2(h)$ at every information set.*

And so on. Continue until no more strategies can be eliminated in this way.

- The order of elimination is crucial for the strategies that survive this algorithm.

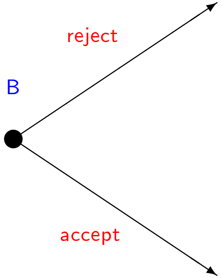
Theorem (Algorithm "works")

(1) For every $k \geq 1$, the strategies that can rationally be chosen by a type that expresses up to k-fold strong belief in rationality are precisely the strategies in $\Gamma^{k+1}(\emptyset)$.

(2) The strategies that can rationally be chosen by a type that expresses common strong belief in rationality are exactly the strategies that are in $\Gamma^k(\emptyset)$ for every k.

- Shimoji and Watson (1998) show that iterated conditional dominance procedure yields precisely the extensive-form rationalizable strategies (Pearce (1984), Battigalli (1997)).
- Battigalli and Siniscalchi (2002) show that common strong belief in rationality yields precisely the extensive-form rationalizable strategies.
- Proof follows from these two results.

$\Gamma^0(h_1)$	200	300	400	500
$(r, 200)$	100, 100	200, 0	200, 0	200, 0
$(r, 300)$	0, 200	150, 150	300, 0	300, 0
$(r, 400)$	0, 200	0, 300	200, 200	400, 0
$(r, 500)$	0, 200	0, 300	0, 400	250, 250

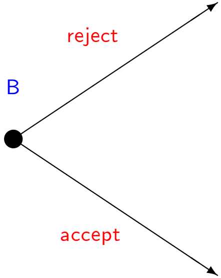


$\Gamma^0(\varnothing)$	200	300	400	500
$(r, 200)$	100, 100	200, 0	200, 0	200, 0
$(r, 300)$	0, 200	150, 150	300, 0	300, 0
$(r, 400)$	0, 200	0, 300	200, 200	400, 0
$(r, 500)$	0, 200	0, 300	0, 400	250, 250
accept	350, 500	350, 500	350, 500	350, 500

350, 500

Step 1

$\Gamma^1(h_1)$	200	300	400	
$(r, 400)$	$0, 200$	$0, 300$	$200, 200$	

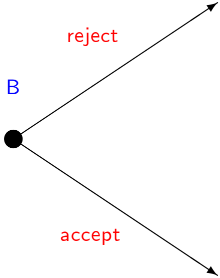


$350, 500$

$\Gamma^1(\varnothing)$	200	300	400
$(r, 400)$	$0, 200$	$0, 300$	$200, 200$
accept	$350, 500$	$350, 500$	$350, 500$

Step 1

$\Gamma^2(h_1)$		300		
$(r, 400)$		$0, 300$		



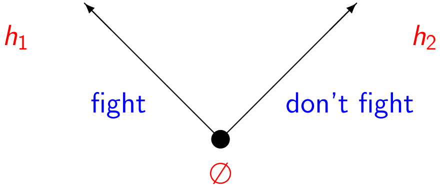
$\Gamma^2(\varnothing)$	300
accept	$350, 500$

$350, 500$

Step 2: Algorithm stops

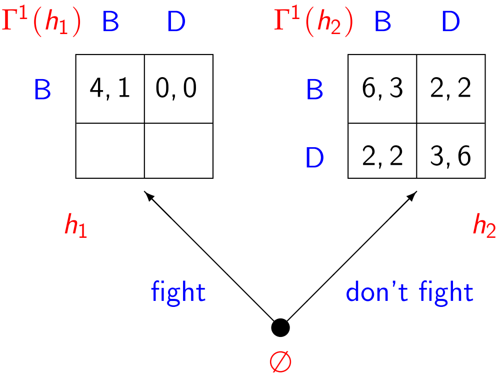
$\Gamma^0(h_1)$	B	D
B	4, 1	0, 0
D	0, 0	1, 4

$\Gamma^0(h_2)$	B	D
B	6, 3	2, 2
D	2, 2	3, 6



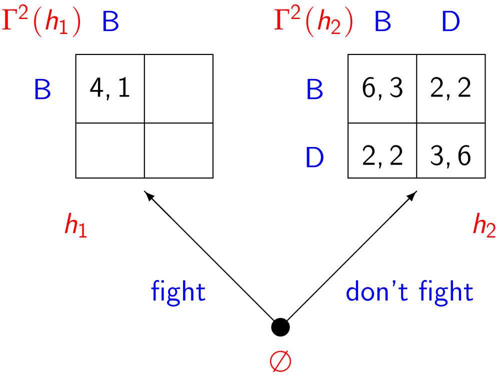
$\Gamma^0(\varnothing)$	(B, B)	(B, D)	(D, B)	(D, D)
$(fight, B)$	4, 1	4, 1	0, 0	0, 0
$(fight, D)$	0, 0	0, 0	1, 4	1, 4
$(don't, B)$	6, 3	2, 2	6, 3	2, 2
$(don't, D)$	2, 2	3, 6	2, 2	3, 6

Step 1



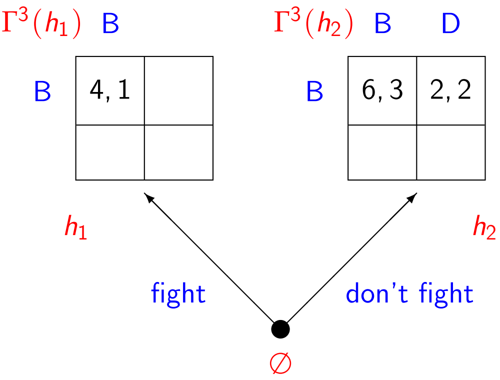
$\Gamma^1(\varnothing)$	(B, B)	(B, D)	(D, B)	(D, D)
$(fight, B)$	4, 1	4, 1	0, 0	0, 0
$(don't, B)$	6, 3	2, 2	6, 3	2, 2
$(don't, D)$	2, 2	3, 6	2, 2	3, 6

Step 1



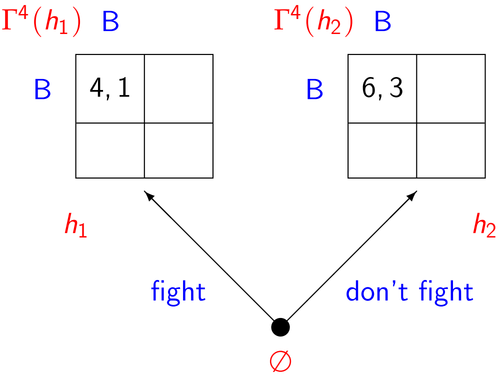
$\Gamma^2(\varnothing)$	(B, B)	(B, D)
$(fight, B)$	4, 1	4, 1
$(don't, B)$	6, 3	2, 2
$(don't, D)$	2, 2	3, 6

Step 2



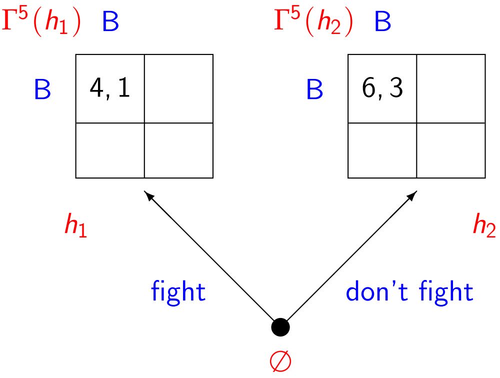
$\Gamma^3(\varnothing)$	(B, B)	(B, D)
$(fight, B)$	4, 1	4, 1
$(don't, B)$	6, 3	2, 2

Step 3



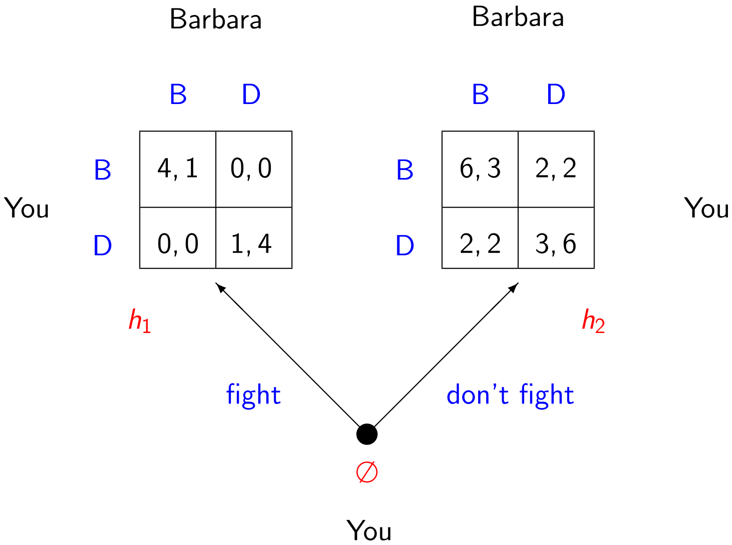
$\Gamma^4(\varnothing)$	(B, B)
$(fight, B)$	4, 1
$(don't, B)$	6, 3

Step 4



$\Gamma^5(\varnothing)$	(B, B)
$(don't, B)$	6, 3

Algorithm stops



- Common belief in future rationality only eliminates the strategy $(fight, D)$ for you.

Comparison with common belief in future rationality

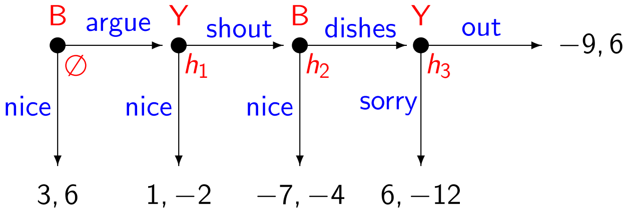
- Common strong belief in rationality and common belief in future rationality represent completely different lines of reasoning.
- The example "Painting Chris' house" has shown that in terms of strategies selected, there is no logical relationship between the two concepts. Both concepts lead to a unique, yet different, strategy choice for you.
- However, both concepts lead to the same outcome in that example, namely that Barbara accepts the colleague's offer at the beginning.
- In "Watching TV with Barbara", common strong belief in rationality leads to a unique outcome, whereas common belief in future rationality allows for many other outcomes as well.
- What about dynamic games with perfect information?

Example: The heat of the fight.

Story

- Barbara and you must decide with TV program to watch: Blackadder or Dallas.
- You prefer Blackadder (utility 6) to Dallas (utility 3).
- Barbara prefers Dallas (utility 6) to Blackadder (utility 3).

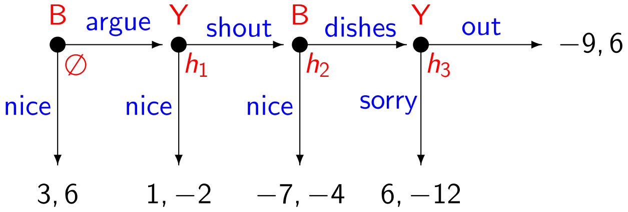
- At the beginning, Barbara can either be nice to you (let you watch your favorite program), or can start to argue with you.
- If she starts arguing, you can either be nice to her (let her watch her favorite program), or you can start shouting at her.
- If you start shouting, then Barbara can either be nice to you (let you watch your favorite program), or she can throw dishes on the floor, as a sign of her anger.
- If she starts throwing dishes on the floor, you can either apologize to her, and let her watch her favorite program, or you can walk out the door and watch Blackadder at Chris' freshly painted house.
- The utility for you and Barbara decreases by 5 every time the conflict escalates.
- If you apologize to Barbara, her utility would increase by 15.
- If you watch Blackadder at Chris' house, your utility would increase by 15.



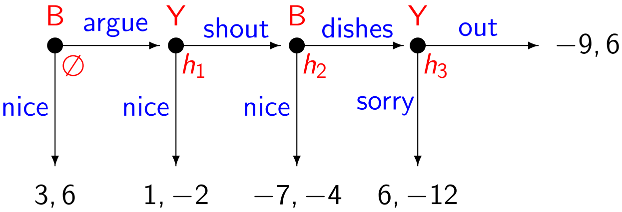
- Common belief in future rationality: Do backward induction.

- At h_3, your backward induction choice is out.
- At h_2, Barbara's backward induction choice is nice.
- At h_1, your backward induction choice is nice.
- At $\emptyset$, Barbara's backward induction choice is nice.

- Hence, common belief in future rationality uniquely selects your strategy nice.
- You expect the outcome where Barbara is nice at the beginning.



- Common strong belief in rationality:
- At h_1, you must believe that Barbara is choosing a rational strategy.
- Hence, at h_1 you must believe that Barbara is implementing the strategy (argue, dishes).
- But then, your unique optimal strategy is (shout, out).
- Hence, common strong belief in rationality uniquely selects your strategy (shout, out).
- You expect the outcome where Barbara is nice at the beginning.



- Hence, common belief in future rationality and common strong belief in rationality lead to unique, yet different, strategy choices for you.
- However, both concepts lead to the same outcome, namely that Barbara will be nice at the beginning.

- Outcome z is possible under common strong belief in rationality, if there is a strategy combination leading to z, where every strategy can rationally be chosen under common strong belief in rationality.
- Similarly for common belief in future rationality.

Theorem (Outcomes under common strong belief in rationality and common belief in future rationality)

Every outcome that is possible under common strong belief in rationality, is also possible under common belief in future rationality.

- A proof can be found in Perea (2017b) and Meier and Perea (2022).
- This result does not hold for strategies.
- Research question: Does the same result hold for explicable equilibrium (Reny, 1992) instead of common strong belief in rationality?
- Research question: Does the same result hold for common belief in future and restricted past rationality (Becerril and Perea, 2020)?

Theorem (Outcomes under common strong belief in rationality and common belief in future rationality)

Every outcome that is possible under common strong belief in rationality, is also possible under common belief in future rationality.

- Remember that in games with perfect information, common belief in future rationality leads to the backward induction strategies, and hence to the backward induction outcomes.
- In generic games with perfect information, the backward induction outcome is unique.

Corollary (Battigalli's Theorem)

Consider a generic dynamic game with perfect information. Then, the only outcome that is possible under common strong belief in rationality is the backward induction outcome.

- Result does not hold for strategies.

Corollary (Battigalli's Theorem)

Consider a generic dynamic game with perfect information. Then, the only outcome that is possible under common strong belief in rationality is the backward induction outcome.

- This result was first shown by Battigalli (1997).
- Other proofs can be found in Chen and Micali (2013), Heifetz and Perea (2015), Catonini (2020) and Perea (2018).

The End

Thank you for your attention

- Battigalli, P. (1997), On rationalizability in extensive games, *Journal of Economic Theory* **74,** 40–61.
- Battigalli, P. and M. Siniscalchi (2002), Strong belief and forward induction reasoning, *Journal of Economic Theory* **106,** 356–391.
- Becerril, R. and A. Perea (2020), Common belief in future and restricted past rationality, *International Journal of Game Theory* **49**, 711–747.
- Ben-Porath, E. and E. Dekel (1992), Signaling future actions and the potential for sacrifice, *Journal of Economic Theory* **57**, 36–51.
- Catonini, E. (2020), On non-monotonic strategic reasoning, *Games and Economic Behavior* **120,** 209–224.
- Chen, J. and S. Micali (2013), The order independence of iterated dominance in extensive games, *Theoretical Economics* **8,** 125–163.

- van Damme, E. (1989), Stable equilibria and forward induction, *Journal of Economic Theory* **48**, 476–496.
- Heifetz, A. and A. Perea (2015), On the outcome equivalence of backward induction and extensive form rationalizability, *International Journal of Game Theory* **44,** 37–59.
- Meier, M. and A. Perea (2022), Forward induction in a backward inductive manner, Manuscript.
- Pearce, D.G. (1984), Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52,** 1029–1050.
- Perea, A. (2001), *Rationality in Extensive Form Games,* Theory and Decision Library, Series C, Kluwer Academic Publishers, Boston / Dordrecht / London.
- Perea, A. (2017a), Forward induction and correct beliefs, *Journal of Economic Theory* **169**, 489–516.

Perea, A. (2017b), Order independence in dynamic games, *EPICENTER* Working Paper No. 8.

Perea, A. (2018), Why forward induction leads to the backward induction outcome: A new proof for Battigalli's theorem, *Games and Economic Behavior* **110,** 120–138.

Reny, P.J. (1992), Backward induction, normal form perfection and explicable equilibria, *Econometrica* **60,** 627–649.

Shimoji, M. (2002), On forward induction in money-burning games, *Economic Theory* **19**, 637–648.

Shimoji, M. and J. Watson (1998), Conditional dominance, rationalizability, and game forms, *Journal of Economic Theory* **83**, 161–195.