# *EPICENTER* Spring Course on Epistemic Game Theory

## Chapter 9: Strong Belief in the Opponents' Rationality

Andrés Perea

Maastricht University

June 2017

# Strong belief in the opponents' rationality

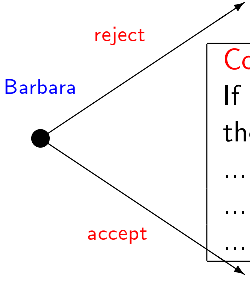
- In the previous chapter, we have discussed the concept of common belief in future rationality.
- Main idea: Whatever you observe in the game, you always believe that your opponents will choose rationally from now on.
- Common belief in this type of reasoning leads to common belief in future rationality.
- 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

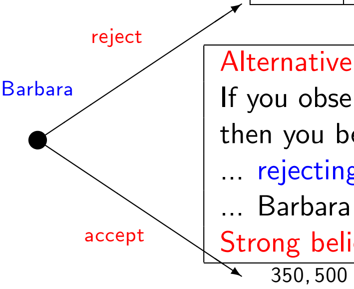
... rejecting offer was a mistake,

... 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,

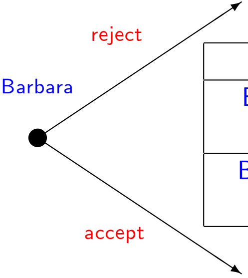
... Barbara will choose price 400.

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 |



350, 500

|  |  |
|---|---|
| Types | $T_1 = \{t_1\}, \; T_2 = \{t_2\}$ |
| Beliefs for Barbara | $b_1(t_1, \varnothing) \;=\; (200, t_2)$<br>$b_1(t_1, h_1) \;=\; (200, t_2)$ |
| 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,

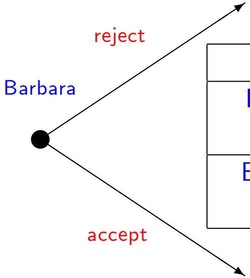
whenever we can find a combination of opponents' types in some epistemic model $M'$, for which there is a combination of optimal strategies leading to $h$,

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.

|  |  |  |  |  |
|---|---|---|---|---|
| 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 |

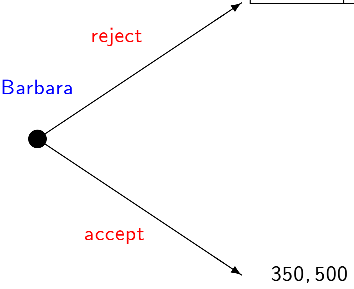


350, 500

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

Your type $t_2$ 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 |



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

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

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.

Barbara

You

| | B | D |
|---|---|---|
| B | 4, 1 | 0, 0 |
| D | 0, 0 | 1, 4 |

$h_1$

fight

Barbara

You

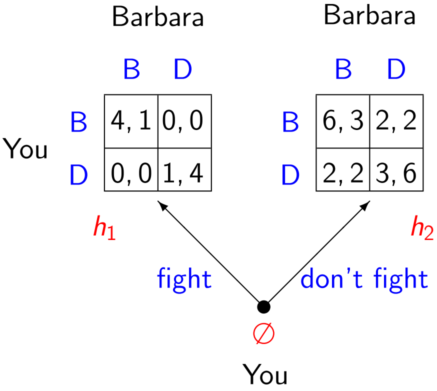
| | B | D |
|---|---|---|
| B | 6, 3 | 2, 2 |
| D | 2, 2 | 3, 6 |

$h_2$

don't fight

$\emptyset$

You



Suppose, Barbara strongly believes in your rationality.

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

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 $(\mathit{fight}, B)$ 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)

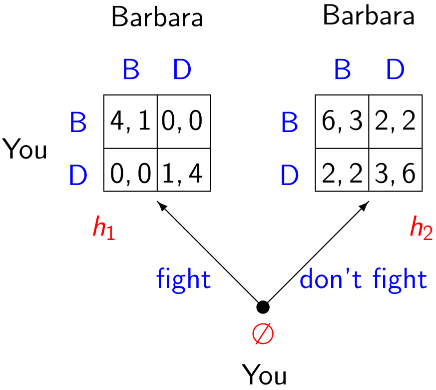
Type $t_i$ expresses 2-fold strong belief in rationality at $h$ if,

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

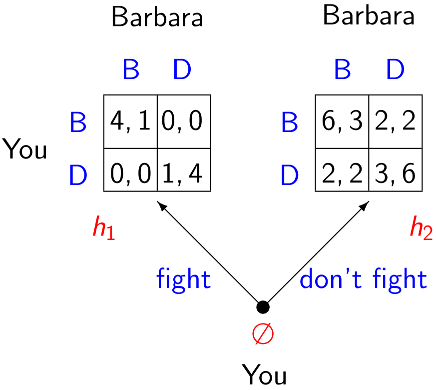
(2) type $t_i$ must at $h$ only assign positive probability to opponents' strategy-type combinations where the strategy combination leads to $h$, the types strongly believe in their opponents' rationality, and the strategies are optimal for the types.



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

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.



$$
\begin{aligned}
b_1(t_1^{fB}) &= ((B, D), t_2^{BD}) \\
b_1(t_1^{dB}) &= ((B, B), t_2^{BB}) \\
b_1(t_1^{dD}) &= ((D, D), t_2^{DD})
\end{aligned}
$$

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

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)

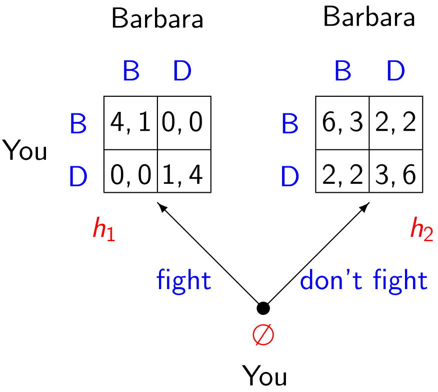
Type $t_i$ is said to express 1-fold strong belief in rationality if $t_i$ strongly believes in the opponents' rationality.

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.

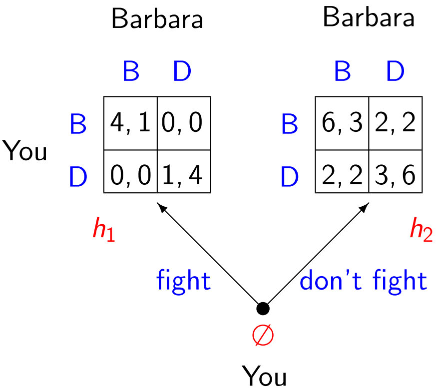
Type $t_i$ expresses common strong belief in rationality if it expresses $k$-fold strong belief in rationality for every $k$.



$$
\begin{aligned}
(2)\ b_1(t_1^{fB}) &= ((B, D), t_2^{BD}) \\
(2)\ b_1(t_1^{dB}) &= ((B, B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) &= ((D, D), t_2^{DD})
\end{aligned}
$$

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

- 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.



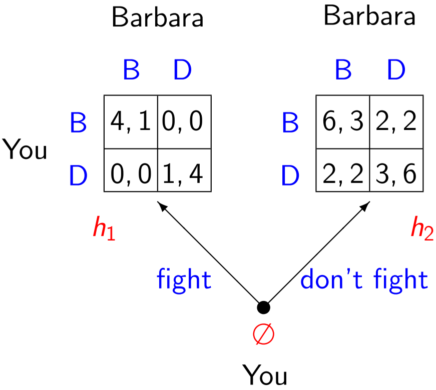
(2) $b_1(t_1^{fB}) = ((B, D), t_2^{BD})$
(2) $b_1(t_1^{dB}) = ((B, B), t_2^{BB})$
(1) $b_1(t_1^{dD}) = ((D, D), t_2^{DD})$

(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) $b_2(t_2^{DD}, h_1) = ((fight, D), t_1^{fB})$
$b_2(t_2^{DD}, h_2) = ((don't, D), t_1^{dD})$

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



$(2)\ b_1(t_1^{fB}) = ((B, D), t_2^{BD})$

$(2)\ b_1(t_1^{dB}) = ((B, B), t_2^{BB})$

$(1)\ b_1(t_1^{dD}) = ((D, D), t_2^{DD})$

$(2)\ b_2(t_2^{BB}, h_1) = ((fight, B), t_1^{fB})$

$b_2(t_2^{BB}, h_2) = ((don't, B), t_1^{dB})$

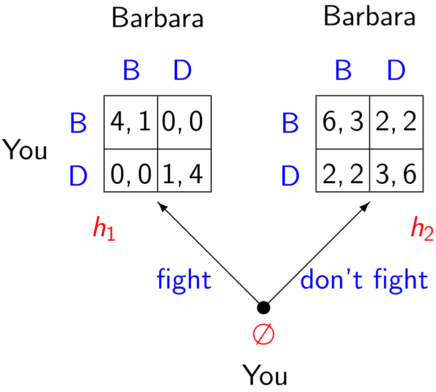
$(2)\ b_2(t_2^{BD}, h_1) = ((fight, B), t_1^{fB})$

$b_2(t_2^{BD}, h_2) = ((don't, D), t_1^{dD})$

$(0)\ b_2(t_2^{DD}, h_1) = ((fight, D), t_1^{fB})$

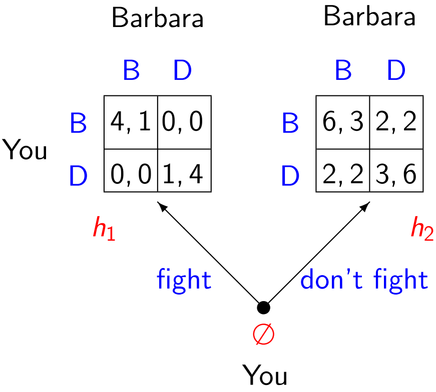
$b_2(t_2^{DD}, h_2) = ((don't, D), t_1^{dD})$

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



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

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



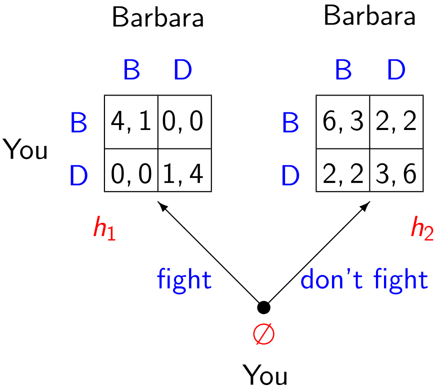
$(3)\ b_1(t_1^{fB}) = ((B, D), t_2^{BD})$
$(3)\ b_1(t_1^{dB}) = ((B, B), t_2^{BB})$
$(1)\ b_1(t_1^{dD}) = ((D, D), t_2^{DD})$

$(3)\ b_2(t_2^{BB}, h_1) = ((fight, B), t_1^{fB})$
$b_2(t_2^{BB}, h_2) = ((don't, B), t_1^{dB})$

$(2)\ b_2(t_2^{BD}, h_1) = ((fight, B), t_1^{fB})$
$b_2(t_2^{BD}, h_2) = ((don't, D), t_1^{dD})$

$(0)\ b_2(t_2^{DD}, h_1) = ((fight, D), t_1^{fB})$
$b_2(t_2^{DD}, h_2) = ((don't, D), t_1^{dD})$

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



Barbara

| | B | D |
|---|---|---|
| B | 4, 1 | 0, 0 |
| D | 0, 0 | 1, 4 |

Barbara

| | B | D |
|---|---|---|
| B | 6, 3 | 2, 2 |
| D | 2, 2 | 3, 6 |

You

$h_1$

$h_2$

fight

don't fight

$\emptyset$

You

(3) $b_1(t_1^{fB}) = ((B, D), t_2^{BD})$

(4) $b_1(t_1^{dB}) = ((B, B), t_2^{BB})$

(1) $b_1(t_1^{dD}) = ((D, D), t_2^{DD})$

(3) $b_2(t_2^{BB}, h_1) = ((fight, B), t_1^{fB})$

$b_2(t_2^{BB}, h_2) = ((don't, B), t_1^{dB})$

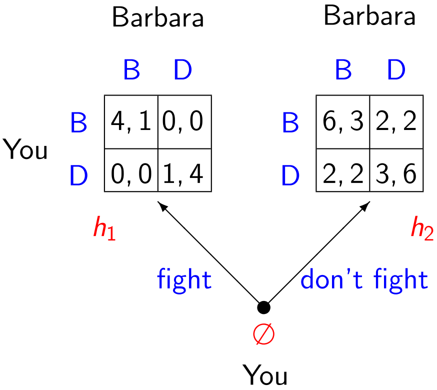
(2) $b_2(t_2^{BD}, h_1) = ((fight, B), t_1^{fB})$

$b_2(t_2^{BD}, h_2) = ((don't, D), t_1^{dD})$

(0) $b_2(t_2^{DD}, h_1) = ((fight, D), t_1^{fB})$

$b_2(t_2^{DD}, h_2) = ((don't, D), t_1^{dD})$

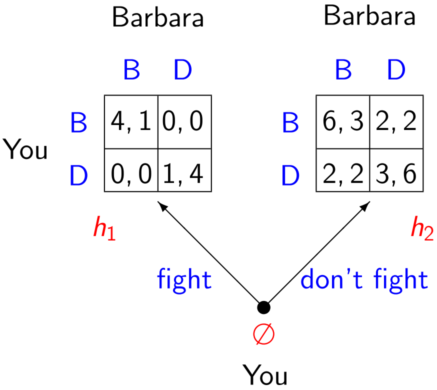
Show: Barbara's type $t_2^{BB}$ expresses 4-fold strong belief in rationality.



$$
\begin{aligned}
(3)\ b_1(t_1^{fB}) &= ((B, D), t_2^{BD}) \\
(4)\ b_1(t_1^{dB}) &= ((B, B), t_2^{BB}) \\
(1)\ b_1(t_1^{dD}) &= ((D, D), t_2^{DD})
\end{aligned}
$$

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

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



(3) $b_1(t_1^{fB}) = ((B, D), t_2^{BD})$

(5) $b_1(t_1^{dB}) = ((B, B), t_2^{BB})$

(1) $b_1(t_1^{dD}) = ((D, D), t_2^{DD})$

(4) $b_2(t_2^{BB}, h_1) = ((fight, B), t_1^{fB})$

$b_2(t_2^{BB}, h_2) = ((don't, B), t_1^{dB})$

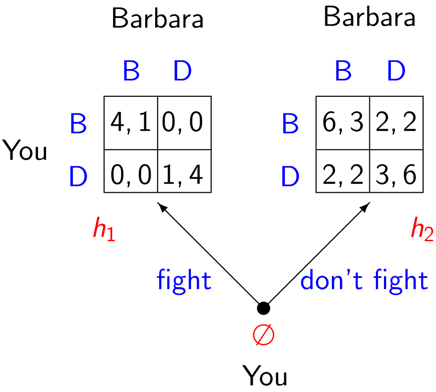
(2) $b_2(t_2^{BD}, h_1) = ((fight, B), t_1^{fB})$

$b_2(t_2^{BD}, h_2) = ((don't, D), t_1^{dD})$

(0) $b_2(t_2^{DD}, h_1) = ((fight, D), t_1^{fB})$

$b_2(t_2^{DD}, h_2) = ((don't, D), t_1^{dD})$

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

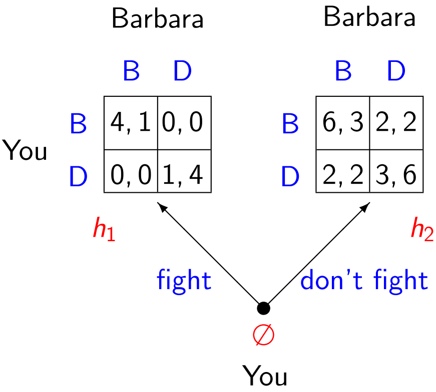


$$\begin{aligned}
&(3)\ b_1(t_1^{fB}) &=& ((B,D), t_2^{BD}) \\
&(5)\ b_1(t_1^{dB}) &=& ((B,B), t_2^{BB}) \\
&(1)\ b_1(t_1^{dD}) &=& ((D,D), t_2^{DD})
\end{aligned}$$

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

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

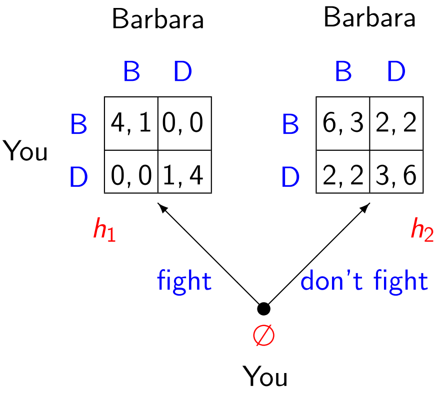
Hence, your type $t_1^{dB}$ and Barbara's type $t_2^{BB}$ express common strong belief in rationality.



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

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

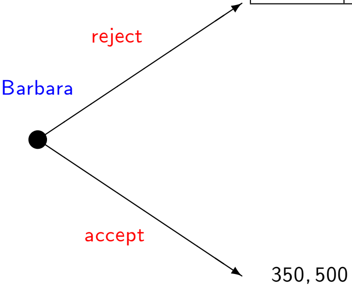
Hence, under common strong belief in rationality, you expect to be watching your favorite program together, without having to start a fight with Barbara.



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

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.

| | | | | |
|---|---|---|---|---|
| 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 |



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

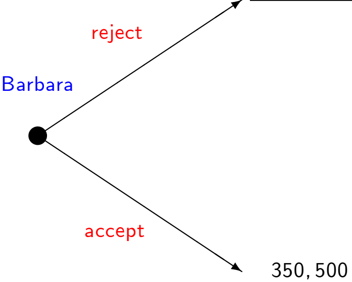
(c) $b_1(t_1^a, \emptyset) = (300, t_2)$
$b_1(t_1^a, h_1) = (300, t_2)$

(0) $b_1(t_1^r, \emptyset) = (500, t_2)$
$b_1(t_1^r, h_1) = (500, t_2)$

(c) $b_2(t_2, h_1) = ((reject, 400), t_1^r)$

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 |



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

(c) $b_1(t_1^a, \varnothing) = (300, t_2)$

$b_1(t_1^a, h_1) = (300, t_2)$

(0) $b_1(t_1^r, \varnothing) = (500, t_2)$

$b_1(t_1^r, h_1) = (500, t_2)$

(c) $b_2(t_2, h_1) = ((reject, 400), t_1^r)$

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.

# 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 $\Gamma^3(\emptyset)$ 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 $\Gamma^2(h)$ at which $i$ is active.
- Hence, every optimal strategy for $t_i$ must be in $\Gamma^3(\emptyset)$.
- Conclusion: Every strategy that is optimal for some type that expresses up to 2-fold strong belief in rationality, must be in $\Gamma^3(\emptyset)$.

## 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.*

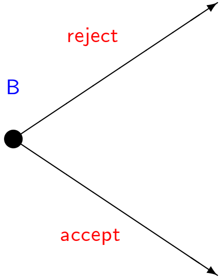
- The order of elimination is crucial for the strategies that survive this algorithm.

**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* $k$.

| $\Gamma^0(h_1)$ | 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 |

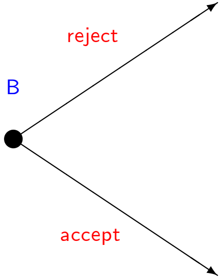


350, 500

| $\Gamma^0(\emptyset)$ | 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 |

**Step 1**

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

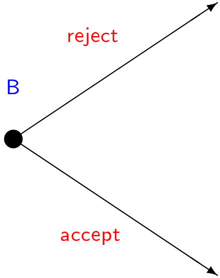


$350, 500$

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

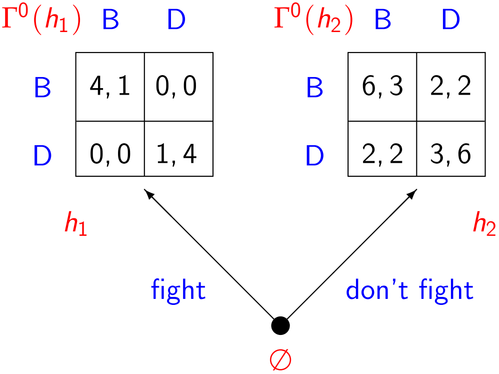
**Step 1**

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



| $\Gamma^2(\emptyset)$ | 300 |
|---|---|
| *accept* | $350, 500$ |

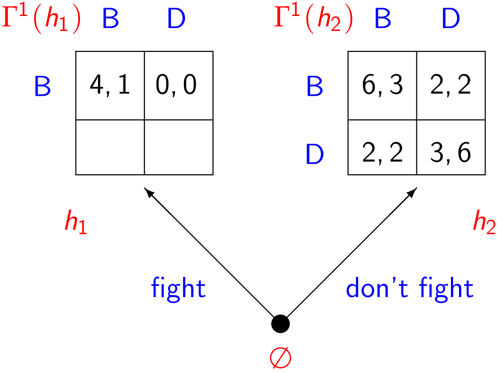
**Step 2: Algorithm stops**



$$\begin{array}{c|cc} \Gamma^0(h_1) & B & D \\ \hline B & 4,1 & 0,0 \\ D & 0,0 & 1,4 \end{array} \qquad \begin{array}{c|cc} \Gamma^0(h_2) & B & D \\ \hline B & 6,3 & 2,2 \\ D & 2,2 & 3,6 \end{array}$$

| $\Gamma^0(\varnothing)$ | $(B,B)$ | $(B,D)$ | $(D,B)$ | $(D,D)$ |
|---|---|---|---|---|
| $(fight, B)$ | 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**

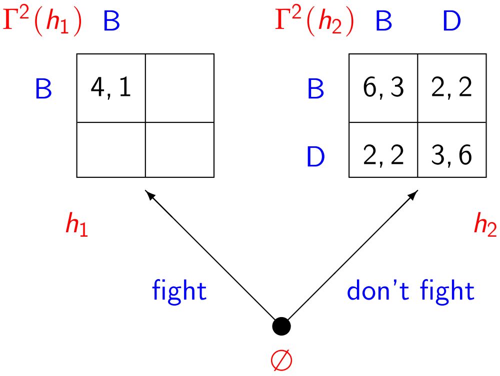


| $\Gamma^1(h_1)$ | B | D |
|---|---|---|
| B | 4, 1 | 0, 0 |
| | | |

| $\Gamma^1(h_2)$ | B | D |
|---|---|---|
| B | 6, 3 | 2, 2 |
| D | 2, 2 | 3, 6 |

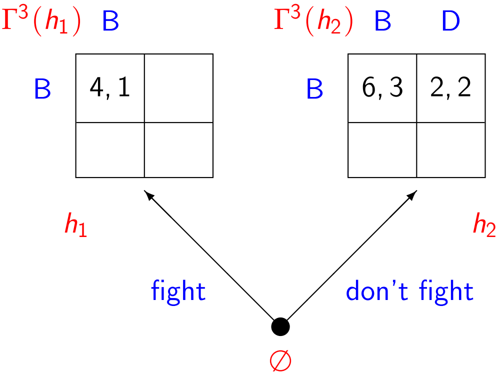
| $\Gamma^1(\varnothing)$ | $(B, B)$ | $(B, D)$ | $(D, B)$ | $(D, D)$ |
|---|---|---|---|---|
| $(fight, B)$ | 4, 1 | 4, 1 | 0, 0 | 0, 0 |
| $(don't, B)$ | 6, 3 | 2, 2 | 6, 3 | 2, 2 |
| $(don't, D)$ | 2, 2 | 3, 6 | 2, 2 | 3, 6 |

**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**

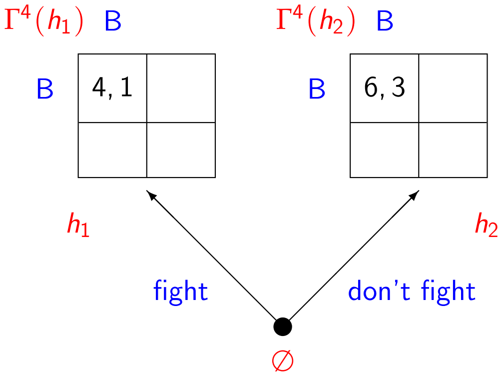


| $\Gamma^3(h_1)$ | B | |
|---|---|---|
| B | 4, 1 | |
| | | |

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

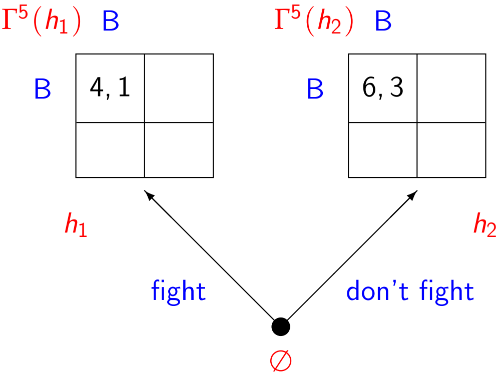
| $\Gamma^3(\emptyset)$ | $(B, B)$ | $(B, D)$ |
|---|---|---|
| $(fight, B)$ | 4, 1 | 4, 1 |
| $(don't, B)$ | 6, 3 | 2, 2 |

**Step 3**



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

**Step 4**



$\Gamma^5(h_1)$

|   | B |   |
|---|---|---|
| B | 4, 1 |   |
|   |   |   |

$\Gamma^5(h_2)$

|   | B |   |
|---|---|---|
| B | 6, 3 |   |
|   |   |   |

$h_1$ $h_2$

fight

don't fight

$\varnothing$

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

**Algorithm stops**

# Comparison with backward dominance procedure

- At step $k$ in the backward dominance procedure, we only eliminate a strategy at $h$ if it is strictly dominated at some reduced decision problem $\Gamma^{k-1}(h')$ weakly following $h$.

- In step $k$ of the iterated conditional dominance procedure, we eliminate a strategy at $h$ if it is strictly dominated at any reduced decision problem $\Gamma^{k-1}(h')$ – not necessarily weakly following $h$ – unless ...

- So, it seems at first sight that in the iterated conditional dominance procedure we eliminate more than in the backward dominance procedure.

- This is not true. Because of the unless above.

- In step $k$ of the iterated conditional dominance procedure, it may happen that by eliminating all strategies from $\Gamma^{k-1}(h)$ which are strictly dominated at some $\Gamma^{k-1}(h')$ we would eliminate all strategy combinations from $\Gamma^{k-1}(h)$.

- In that case, we eliminate no strategies at $h$ in the iterated conditional dominance procedure. But within the backward dominance procedure, we could still eliminate some strategies at $h$.

- In fact, the example "Painting Chris' house" has shown that in terms of strategies selected, there is no logical relationship between the two procedures. Both procedures lead to a unique, yet different, strategy choice for you.

- However, both procedures lead to the same outcome in that example, namely that Barbara accepts the colleague's offer at the beginning.

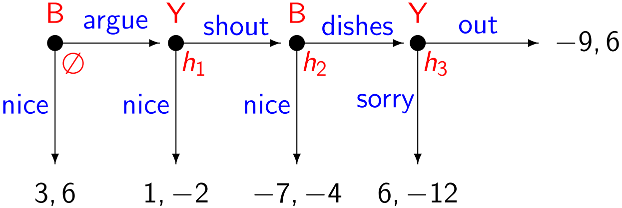
- The same phenomenon can also happen for dynamic games with perfect information.

# 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).

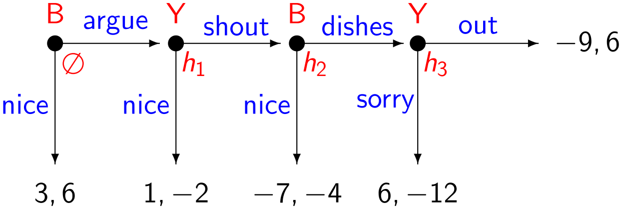
- At the beginning, Barbara can either be nice to you (let you watch your favorite program), or can start to argue with you.
- If she starts arguing, you can either be nice to her (let her watch her favorite program), or you can start shouting at her.
- If you start shouting, then Barbara can either 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.



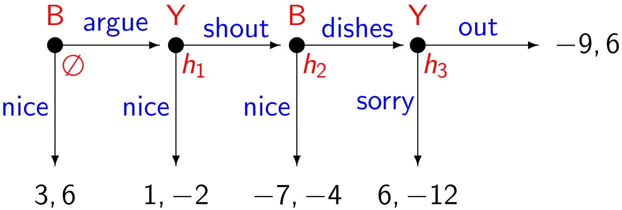
- Backward dominance procedure: Do backward induction.

- 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, the backward dominance procedure uniquely selects your strategy nice.
- You expect the outcome where Barbara is nice at the beginning.



- Iterated conditional dominance procedure:
- 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, the iterated dominance procedure uniquely selects your strategy (shout, out).
- You expect the outcome where Barbara is nice at the beginning.



- Hence, the backward dominance procedure and the iterated conditional dominance procedure lead to unique, yet different, strategy choices for you.

- However, both procedures lead to the same outcome, namely that Barbara will be nice at the beginning.

- The latter is true for every dynamic game with perfect information, where there is a unique backward induction outcome.

- 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)**

*Consider a game with perfect information. Then, every outcome that is possible under common strong belief in rationality, is also possible under common belief in future rationality.*

- Remember that in dynamic games with perfect information, common belief in future rationality selects exactly the backward induction strategies.
- Hence, in particular, common belief in future rationality leads to a backward induction outcome in such games.

**Corollary (Common strong belief in rationality leads to backward induction outcomes)**

*In a game with perfect information, every outcome that is possible under common strong belief in rationality must be a backward induction outcome.*

- However, common strong belief in rationality may not lead to the backward induction strategy for every player.

The End

Thank you for your attention