Mini-course on Epistemic Game Theory
Lecture 3: Backward Induction Reasoning

Andrés Perea
EpiCenter & Dept. of Quantitative Economics

EPICENTER
Research Center for
Epistemic Game Theory

Maastricht University

Toulouse, June/July 2015

Introduction

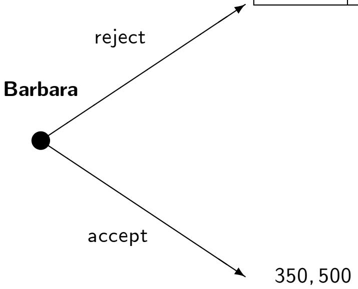
- In a **dynamic game,** players may choose one after the other.
- Before you make a choice, you may (partially) observe what your opponents have chosen so far.
- It may happen that your **initial belief** about the opponents' choices will be **contradicted** later on.
- Then you must **revise** your belief about the opponents' choices. **But how?**
- There may be **several plausible** ways to revise your belief.

Example: Painting Chris' house

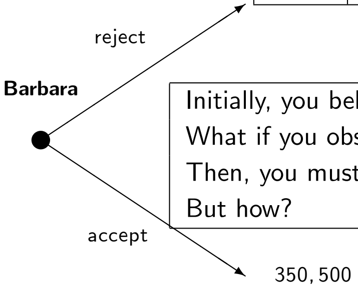
Story

- 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	$0, 200$	$0, 300$	$0, 400$	$250, 250$

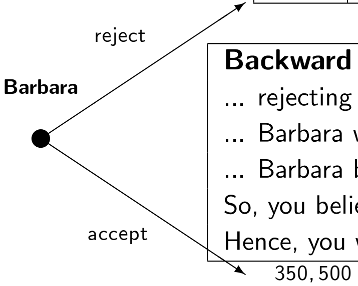


	200	300	400	500
200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250



Initially, you believe that Barbara accepts the offer.
What if you observe that she has rejected the offer?
Then, you must revise your belief.
But how?

	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



Backward induction: You believe that ...

... rejecting offer was a mistake by Barbara,

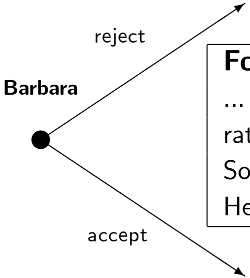
... Barbara will choose rationally in the future,

... Barbara believes that you will choose rationally.

So, you believe that Barbara chooses 200 or 300.

Hence, you will choose price 200.

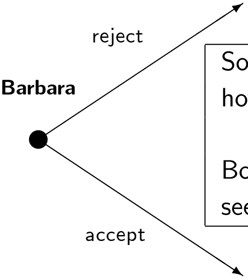
	200	300	400	500
200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250



350, 500

Forward induction: You believe that ...
... rejecting colleague's offer was a rational choice for Barbara.
So, you believe that Barbara chooses price 400.
Hence, you will choose price 300.

	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



So, your choice crucially depends on how you revise your belief about Barbara.

Both ways of revising your belief seem plausible.

350, 500

Conditional beliefs

We would like to model **hierarchies of conditional beliefs.**

That is, we want to model

- the conditional belief that player i has, at every information set $h \in H_i$, about his opponents' **strategy choices**,
- the conditional belief that player i has, at every information set $h \in H_i$, about the **conditional belief** that opponent j has, at every information set $h' \in H_j$, about the opponents' strategy choices,

and so on.

- So, at every information set, a player has a **conditional belief** about the opponents' **strategies** and the opponents' **conditional belief hierarchies.**
- Call a conditional belief hierarchy a **type.**

Strategies

An **information set** for player i

- is a situation where player i must make a choice,
- describes the **information** that player i has about the opponents' **past choices**.

H_i: collection of information sets for player i.

Definition (Strategy)

A **strategy** for player i is a function s_i that assigns to each of his information sets $h \in H_i$ some available choice $s_i(h)$, **unless** h cannot be reached due to some choice $s_i(h')$ at an earlier information set $h' \in H_i$.

In the latter case, no choice needs to be specified at h.

This is different from the **classical** definition of a strategy! It corresponds to **plan of action** in Rubinstein (1991).

Epistemic model

Definition (Epistemic model)

An **epistemic model** for a **dynamic game** specifies for every player i a set T_i of possible **types**.

Moreover, it specifies for every type $t_i \in T_i$, at every information set $h \in H_i$, a **conditional** probabilistic **belief** $b_i(t_i, h)$ over the set $S_{-i}(h) \times T_{-i}$ of opponents' **strategy-type** combinations.

- Here, $S_{-i}(h)$ is the set of opponents' strategy combinations that make reaching h possible.
- The epistemic model is based on Ben-Porath (1997) and Battigalli and Siniscalchi (1999).

Definition (Epistemic model)

An **epistemic model** for a **dynamic game** specifies for every player i a set T_i of possible **types**.

Moreover, it specifies for every type $t_i \in T_i$, at every information set $h \in H_i$, a **conditional** probabilistic **belief** $b_i(t_i, h)$ over the set $S_{-i}(h) \times T_{-i}$ of opponents' **strategy-type** combinations.

- From the epistemic model, we can **derive** the **complete belief hierarchy** for every type.
- A type may **revise his belief** about the opponents' **strategies** during the game.
- A type may also **revise his beliefs** about the opponents' **beliefs** during the game.

	g	h
e	1, 2	1, 1
f	3, 1	0, 2

h_1

	k	l
i	0, 2	1, 0
j	3, 0	0, 3

h_2

	c	d
a		
b	2, 2	3, 0

$\varnothing$

Types	$T_1 = \{t_1, \hat{t}_1\},\ T_2 = \{t_2, \hat{t}_2\}$
Beliefs for player 1	$b_1(t_1, \varnothing) = ((c,h), t_2)$ $b_1(t_1, h_1) = ((c,h), t_2)$ $b_1(t_1, h_2) = ((d,k), \hat{t}_2)$ $b_1(\hat{t}_1, \varnothing) = (0.3) \cdot ((c,g), t_2) + (0.7) \cdot ((d,l), \hat{t}_2)$ $b_1(\hat{t}_1, h_1) = ((c,g), t_2)$ $b_1(\hat{t}_1, h_2) = ((d,l), \hat{t}_2)$
Beliefs for player 2	$b_2(t_2, \varnothing) = (b, t_1)$ $b_2(t_2, h_1) = ((a,f,i), t_1)$ $b_2(t_2, h_2) = ((a,f,i), t_1)$ $b_2(\hat{t}_2, \varnothing) = ((a,e,j), \hat{t}_1)$ $b_2(\hat{t}_2, h_1) = ((a,e,j), \hat{t}_1)$ $b_2(\hat{t}_2, h_2) = ((a,e,j), \hat{t}_1)$

Common belief in future rationality

You **believe in the opponents' future rationality** if you always believe, throughout the game, that your opponents will make optimal choices at every **present** and **future** information set.

Definition (Belief in the opponents' rationality)

Type t_i believes at h that opponent j **chooses rationally at** h' if his conditional belief $b_i(t_i, h)$ only assigns positive probability to strategy-type pairs (s_j, t_j) for player j where strategy s_j is optimal for type t_j at information set h'.

Definition (Belief in the opponents' future rationality)

Type t_i believes at h in opponent j's **future** rationality if t_i believes at h that j chooses rationally at every information set h' for player j that weakly follows h.

Definition (Common belief in future rationality)

(Induction start) Type t_i expresses **1-fold** belief in future rationality if t_i believes in the opponents' future rationality.

(Inductive step) For every $k \geq 2$, type t_i expresses k-**fold** belief in future rationality if t_i assigns, at every information set $h \in H_i$, only positive probability to opponents' types that express $(k-1)$-fold belief in future rationality.

Type t_i expresses **common belief in future rationality** if t_i expresses k-fold belief in future rationality for all k.

- This concept has been presented in Perea (2014). See Baltag, Smets and Svesper (2009) and Penta (2009) for closely related conditions.
- It represents a **backward induction** type of reasoning: Players only think about the future.

Definition (Common belief in future rationality)

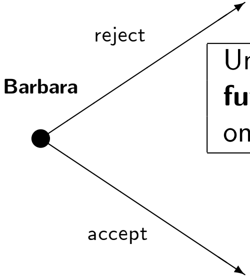
(Induction start) Type t_i expresses **1-fold** belief in future rationality if t_i believes in the opponents' future rationality.

(Inductive step) For every $k \geq 2$, type t_i expresses k-**fold** belief in future rationality if t_i assigns, at every information set $h \in H_i$, only positive probability to opponents' types that express $(k-1)$-fold belief in future rationality.

Type t_i expresses **common belief in future rationality** if t_i expresses k-fold belief in future rationality for all k.

- Is implicitly present in **subgame perfect equilibrium** (Selten (1965)) and **sequential equilibrium** (Kreps and Wilson (1982)).
- But these concepts, like Nash equilibrium, assume that a player always believes that his opponents are **correct about his beliefs**.

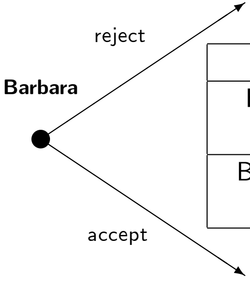
	200	300	400	500
200	100, 100	200, 0	200, 0	200, 0
300	0, 200	150, 150	300, 0	300, 0
400	0, 200	0, 300	200, 200	400, 0
500	0, 200	0, 300	0, 400	250, 250



350, 500

Under **common belief in future rationality** you can only rationally choose 200.

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	$b_2(t_2, h_1) = ((reject, 200), t_1)$

Both types express **common belief in future rationality.**

Algorithm

- We wish to find those strategies that you can rationally choose under **common belief in future rationality.**
- Can we construct an **algorithm** that helps us find these strategies?
- Yes! It will proceed by iteratedly removing strategies at the various information sets in the game.

Step 1: 1-fold belief in future rationality.

- Which strategies can player i rationally choose if he expresses **1-fold belief in future rationality?** That is, if he believes in the opponents' future rationality?
- Consider a type t_i that believes in the opponents' future rationality. Then, t_i believes at every information set $h \in H_i$ that opponent j chooses optimally at every information set $h' \in H_j$ that **weakly follows** h.
- A strategy s_j for player j is **optimal** at h' for some conditional belief at h', if and only if, s_j is **not strictly dominated** within the **full decision problem** $\Gamma^0(h') = (S_j(h'), S_{-j}(h'))$ at h'.
- So, t_i assigns at h only positive probability to j's strategies s_j that are not strictly dominated within any full decision problem $\Gamma^0(h')$ that weakly follows h, and at which j is active.

Step 1: 1-fold belief in future rationality.

- So, t_i assigns at h only positive probability to j's strategies s_j that are not strictly dominated within any full decision problem $\Gamma^0(h')$ that weakly follows h, and at which j is active.
- At every information set $h \in H_i$, delete from the full decision problem $\Gamma^0(h)$ those strategies s_j that are **strictly dominated** within some full decision problem $\Gamma^0(h')$ that **weakly follows** h, and at which j is active. This gives the **reduced decision problem** $\Gamma^1(h)$.
- Hence, type t_i assigns at every information set $h \in H_i$ only positive probability to opponents' strategies in $\Gamma^1(h)$.
- So, every strategy that is optimal for t_i at h, must **not** be **strictly dominated** within the **reduced decision problem** $\Gamma^1(h)$.

Step 1: 1-fold belief in future rationality.

- So, every strategy that is optimal for t_i at h, must **not** be **strictly dominated** within the **reduced decision problem** $\Gamma^1(h)$.
- 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.
- **Conclusion:** Every strategy s_i that is optimal for some type t_i which expresses **1-fold** belief in future rationality, must be in $\Gamma^2(\emptyset)$.

Step 2: Up to 2-fold belief in future rationality.

- Which strategies can player i rationally choose if he expresses **up to 2-fold belief** in future rationality?
- Consider a type t_i that expresses up to 2-fold belief in future rationality. Then, t_i assigns at every $h \in H_i$ only positive probability to opponents' strategy-type pairs (s_j, t_j) where s_j is optimal for t_j at every $h' \in H_j$ that weakly follows h, and t_j expresses 1-fold belief in future rationality.
- We know from Step 1 that every such type t_j assigns at every $h' \in H_j$ only positive probability to opponents' strategies in $\Gamma^1(h')$.
- So, every such strategy s_j above must at every $h' \in H_j$ weakly following h not be strictly dominated within $\Gamma^1(h')$.

Step 2: Up to 2-fold belief in future rationality.

- So, every such strategy s_j above must at every $h' \in H_j$ weakly following h not be strictly dominated within $\Gamma^1(h')$.
- Let $\Gamma^2(h)$ be the reduced decision problem at h which is obtained from $\Gamma^1(h)$ by removing all strategies s_j which are strictly dominated within some $\Gamma^1(h')$ weakly following h, at which j is active.
- Then, type t_i will assign at h only positive probability to strategies s_j in $\Gamma^2(h)$.
- So, every strategy s_i which is optimal for t_i at h must **not** be **strictly dominated** within $\Gamma^2(h)$.

Step 2: Up to 2-fold belief in future rationality.

- So, every strategy s_i which is optimal for t_i at h must **not** be **strictly dominated** within $\Gamma^2(h)$.
- 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.
- **Conclusion:** Every strategy s_i that is optimal for some type t_i which expresses **up to 2-fold** belief in future rationality, must be in $\Gamma^3(\varnothing)$.

Fix an information set h for player i.

- 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 make reaching h possible, and $S_{-i}(h)$ is the set of opponents' strategy combinations that make reaching h possible.

- 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 $D_{-i}(h) \subseteq S_{-i}(h)$.

Algorithm (Backward dominance procedure)

(Induction start) Let $\Gamma^0(h)$ *be the full decision problem at h for every information set h.*

(Inductive step) Let $k \geq 1$. *At every reduced decision problem* $\Gamma^{k-1}(h)$, *eliminate for every player i those strategies that are strictly dominated at some reduced decision problem* $\Gamma^{k-1}(h')$ *that weakly follows h and at which player i is active. This leads to new reduced decision problems* $\Gamma^k(h)$ *at every information set.*

- Algorithm is taken from Perea (2014). Similar procedures can be found in Penta (2009) and Chen and Micali (2011).
- The algorithm always **stops within finitely many steps**.
- At every information set, it yields a nonempty set of strategies for every player.

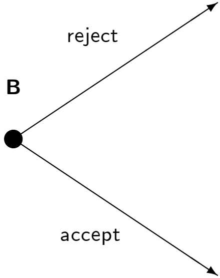
Theorem (Perea (2014))

(1) For every $k \geq 1$, the strategies that can rationally be chosen by a type that expresses **up to k-fold belief in future rationality** *are exactly the strategies that survive the* **first $k+1$ steps** *of the backward dominance procedure at $\emptyset$.*

(2) The strategies that can rationally be chosen by a type that expresses **common belief in future rationality** *are exactly the strategies that survive the* **full** *backward dominance procedure at $\emptyset$.*

- A strategy survives the first $k+1$ steps of the backward dominance procedure at $\emptyset$ if it is in the reduced decision problem $\Gamma^{k+1}(\emptyset)$.
- A strategy survives the full backward dominance procedure at $\emptyset$ if it is in the reduced decision problem $\Gamma^{k}(\emptyset)$ 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

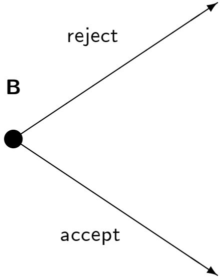


350, 500

$\Gamma^0(\varnothing)$	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

Step 1

$\Gamma^0(h_1)$	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

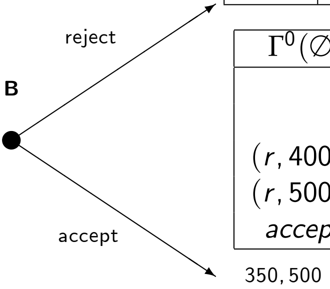


$\Gamma^0(\emptyset)$	200	300	400	500
$(r, 300)$	0, 200	150, 150	300, 0	300, 0
$(r, 400)$	0, 200	0, 300	200, 200	400, 0
$(r, 500)$	0, 200	0, 300	0, 400	250, 250
accept	350, 500	350, 500	350, 500	350, 500

350, 500

Step 1

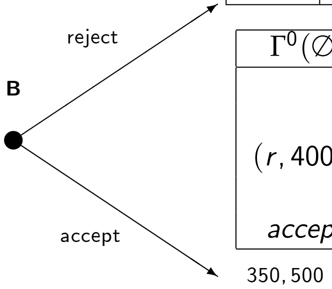
$\Gamma^0(h_1)$	200	300	400	500
$(r, 200)$	100, 100	200, 0	200, 0	200, 0
$(r, 300)$	0, 200	150, 150	300, 0	300, 0
$(r, 400)$	0, 200	0, 300	200, 200	400, 0
$(r, 500)$	0, 200	0, 300	0, 400	250, 250



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

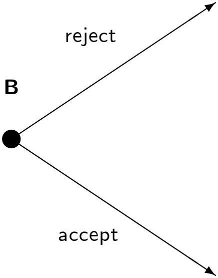
$\Gamma^0(h_1)$	200	300	400	500
$(r, 200)$	100, 100	200, 0	200, 0	200, 0
$(r, 300)$	0, 200	150, 150	300, 0	300, 0
$(r, 400)$	0, 200	0, 300	200, 200	400, 0
$(r, 500)$	0, 200	0, 300	0, 400	250, 250



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

Step 1

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

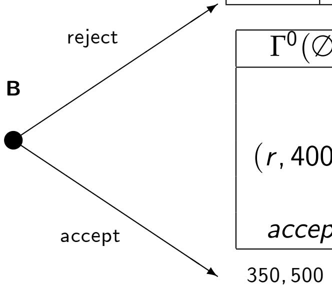


$\Gamma^0(\emptyset)$	200	300	400	500
$(r, 400)$	$0, 200$	$0, 300$	$200, 200$	$400, 0$
accept	$350, 500$	$350, 500$	$350, 500$	$350, 500$

$350, 500$

Step 1

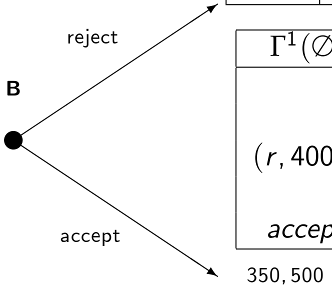
$\Gamma^0(h_1)$	200	300	400	
$(r, 200)$	$100, 100$	$200, 0$	$200, 0$	
$(r, 300)$	$0, 200$	$150, 150$	$300, 0$	
$(r, 400)$	$0, 200$	$0, 300$	$200, 200$	



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

Step 1

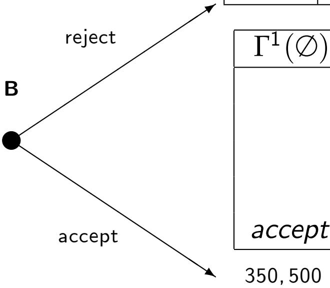
$\Gamma^1(h_1)$	200	300	400	
$(r, 200)$	100, 100	200, 0	200, 0	
$(r, 300)$	0, 200	150, 150	300, 0	
$(r, 400)$	0, 200	0, 300	200, 200	



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

End of Step 1

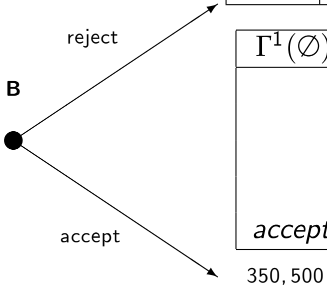
$\Gamma^1(h_1)$	200	300	400	
$(r, 200)$	100, 100	200, 0	200, 0	
$(r, 300)$	0, 200	150, 150	300, 0	
$(r, 400)$	0, 200	0, 300	200, 200	



$\Gamma^1(\emptyset)$	200	300	400
accept	350, 500	350, 500	350, 500

Step 2

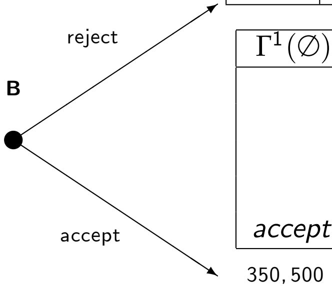
$\Gamma^1(h_1)$	200	300	400	
$(r, 200)$	$100, 100$	$200, 0$	$200, 0$	
$(r, 300)$	$0, 200$	$150, 150$	$300, 0$	



$\Gamma^1(\emptyset)$	200	300	400
accept	$350, 500$	$350, 500$	$350, 500$

Step 2

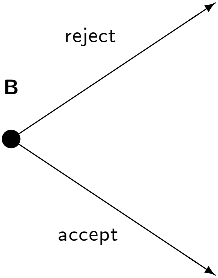
$\Gamma^1(h_1)$	200	300		
$(r, 200)$	$100, 100$	$200, 0$		
$(r, 300)$	$0, 200$	$150, 150$		



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

Step 2

$\Gamma^2(h_1)$	200	300		
$(r, 200)$	$100, 100$	$200, 0$		
$(r, 300)$	$0, 200$	$150, 150$		



$\Gamma^2(\emptyset)$	200	300
accept	$350, 500$	$350, 500$

$350, 500$

End of Step 2

$\Gamma^2(h_1)$	200	300		
$(r, 200)$	$100, 100$	$200, 0$		

reject

B

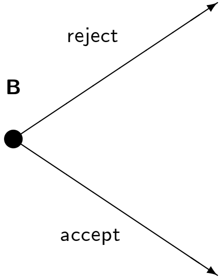
accept

$\Gamma^2(\varnothing)$	200	300
accept	$350, 500$	$350, 500$

$350, 500$

Step 3

$\Gamma^2(h_1)$	200			
$(r, 200)$	$100, 100$			

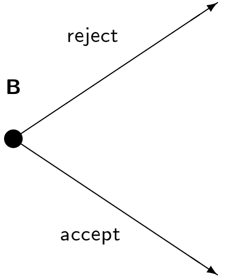


$\Gamma^2(\emptyset)$	200	300
accept	$350, 500$	$350, 500$

$350, 500$

Step 3

$\Gamma^3(h_1)$	200			
$(r, 200)$	$100, 100$			



$\Gamma^3(\emptyset)$	200
accept	$350, 500$

$350, 500$

End of algorithm

Games with perfect information

Algorithm (Backward dominance procedure)

(Induction start) Let $\Gamma^0(h)$ *be the full decision problem at* h *for every information set* h.

(Inductive step) Let $k \geq 1$. *At every reduced decision problem* $\Gamma^{k-1}(h)$, *eliminate for every player* i *those strategies that are strictly dominated at some reduced decision problem* $\Gamma^{k-1}(h')$ *that weakly follows* h *and at which player* i *is active. This leads to new reduced decision problems* $\Gamma^k(h)$ *at every information set.*

- The **order** in which we eliminate strategies – including the order in which we walk through the information sets – is **not important** for the final result!
- In dynamic games with **perfect information,** it coincides with **backward induction procedure** (**not** due to Zermelo (1913) !).

Theorem (Common belief in future rationality leads to backward induction)

Consider a dynamic game with **perfect information**.

Then, the strategies that can rationally be chosen under **common belief in future rationality** *are exactly the* **backward induction strategies.**

- If the game with perfect information is **generic** – that is, all utilities at the terminal histories are different – then there is a **unique** backward induction strategy for every player.
- In **non-generic** games with perfect information, there may be more than one backward induction strategy for a player.

Theorem (Common belief in future rationality leads to backward induction)

Consider a dynamic game with **perfect information**.

Then, the strategies that can rationally be chosen under **common belief in future rationality** *are exactly the* **backward induction strategies.**

- So, common belief in future rationality can be seen as an **epistemic foundation for backward induction.**
- **Other epistemic foundations for backward induction** can be found in Aumann (1995), Samet (1996), Balkenborg and Winter (1997), Stalnaker (1998), Asheim (2002), Quesada (2002, 2003), Clausing (2003, 2004), Asheim and Perea (2005), Feinberg (2005), Perea (2008), Baltag, Smets and Zvesper (2009) and Bach and Heilmann (2011).
- See Perea (2007) for an overview of these epistemic foundations.

G.B. Asheim, ‘On the epistemic foundation for backward induction’, *Mathematical Social Sciences*, 44 (2002), 121–144

G.B. Asheim and A. Perea, ‘Sequential and quasi-perfect rationalizability in extensive games’, *Games and Economic Behavior*, 53 (2005), 15–42

R.J. Aumann, ‘Backward induction and common knowledge of rationality’, *Games and Economic Behavior*, 8 (1995), 6–19

C.W. Bach and C. Heilmann, ‘Agent connectedness and backward induction’, *International Game Theory Review*, 13 (2011), 1–14

D. Balkenborg and E. Winter, ‘A necessary and sufficient epistemic condition for playing backward induction’, *Journal of Mathematical Economics*, 27 (1997), 325–345

A. Baltag, S. Smets and J.A. Zvesper, ‘Keep ‘hoping’ for rationality: a solution to the backward induction paradox’, *Synthese*, 169 (2009), 301–333 (*Knowledge, Rationality and Action*, 705–737)

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

E. Ben-Porath, ‘Rationality, Nash equilibrium and backwards induction in perfect-information games’, *Review of Economic Studies*, 64 (1997), 23–46

J. Chen and S. Micali, ‘The robustness of extensive-form rationalizability’, Working paper (2011)

T. Clausing, ‘Doxastic conditions for backward induction’, *Theory and Decision*, 54 (2003), 315–336

T. Clausing, ‘Belief revision in games of perfect information’, *Economics and Philosophy*, 20 (2004), 89–115

Y. Feinberg, ‘Subjective reasoning - dynamic games’, *Games and Economic Behavior*, 52 (2005), 54–93

- D.M. Kreps and R. Wilson, ‘Sequential equilibria’, *Econometrica*, 50 (1982), 863–894
- A. Penta, ‘Robust dynamic mechanism design’, Working paper (2009)
- A. Perea, ‘Epistemic foundations for backward induction: An overview’ in J. van Benthem, D. Gabbay and B. Löwe (eds.), *Interactive Logic Proceedings of the 7th Augustus de Morgan Workshop, London. Texts in Logic and Games 1* (Amsterdam University Press, 2007c), pp. 159–193
- A. Perea, ‘Minimal belief revision leads to backward induction’, *Mathematical Social Sciences*, 56 (2008), 1–26
- A. Perea, ‘Belief in the opponents’ future rationality’, *Games and Economic Behavior*, 83 (2014), 231–254
- A. Quesada, ‘Belief system foundations of backward induction’, *Theory and Decision*, 53 (2002), 393–403

- A. Quesada, 'From common knowledge of rationality to backward induction', *International Game Theory Review*, 2 (2003), 127–137
- A. Rubinstein, 'Comments on the interpretation of game theory', *Econometrica*, 59 (1991), 909–924
- D. Samet, 'Hypothetical knowledge and games with perfect information', *Games and Economic Behavior*, 17 (1996), 230–251
- R. Selten, 'Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragezeit', *Zeitschrift für die Gesammte Staatswissenschaft*, 121 (1965), 301–324, 667–689
- R. Stalnaker, 'Belief revision in games: forward and backward induction', *Mathematical Social Sciences*, 36 (1998), 31–56
- E. Zermelo, 'Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels', *Proceedings Fifth International Congress of Mathematicians*, 2 (1913), 501–504