Research Project Proposal:

Regret Minimization for Non-Cooperative Games

Tommaso Bianchi
tommaso4.bianchi@mail.polimi.it
CSE Track

POLITECNICO MILANO 1863

HONOURS PROGRAMME HP-SR in Information Technology

Preliminaries

- Game representations
- Strategies
- Equilibria
- Regret
- Approachability

Normal-form games

Player 2

		L	R
Player 1	**T**	(4, 4)	(1, 5)
	B	(5, 1)	(2, 2)

Normal-form games - Players

Player 2

Player 1		L	R
	T	(4, 4)	(1, 5)
	B	(5, 1)	(2, 2)

Normal-form games - Actions

		Player 2	
		L	**R**
Player 1	**T**	(4, 4)	(1, 5)
	B	(5, 1)	(2, 2)

Normal-form games - Utilities

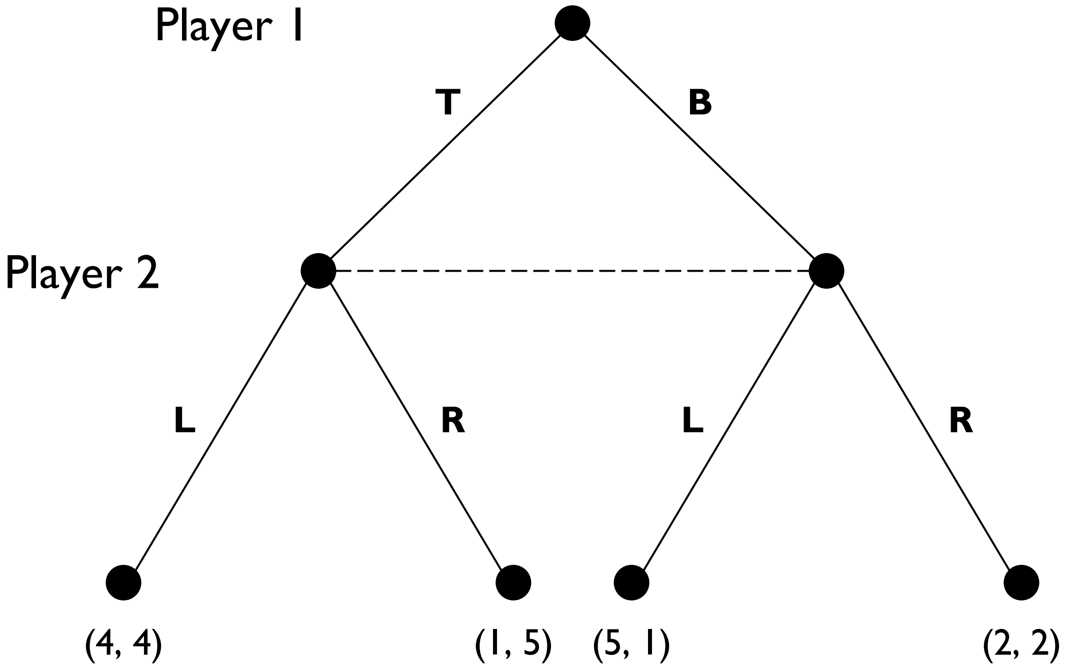
		Player 2	
		L	**R**
Player 1	**T**	(4, 4)	(1, 5)
	B	(5, 1)	(2, 2)

Normal-form games

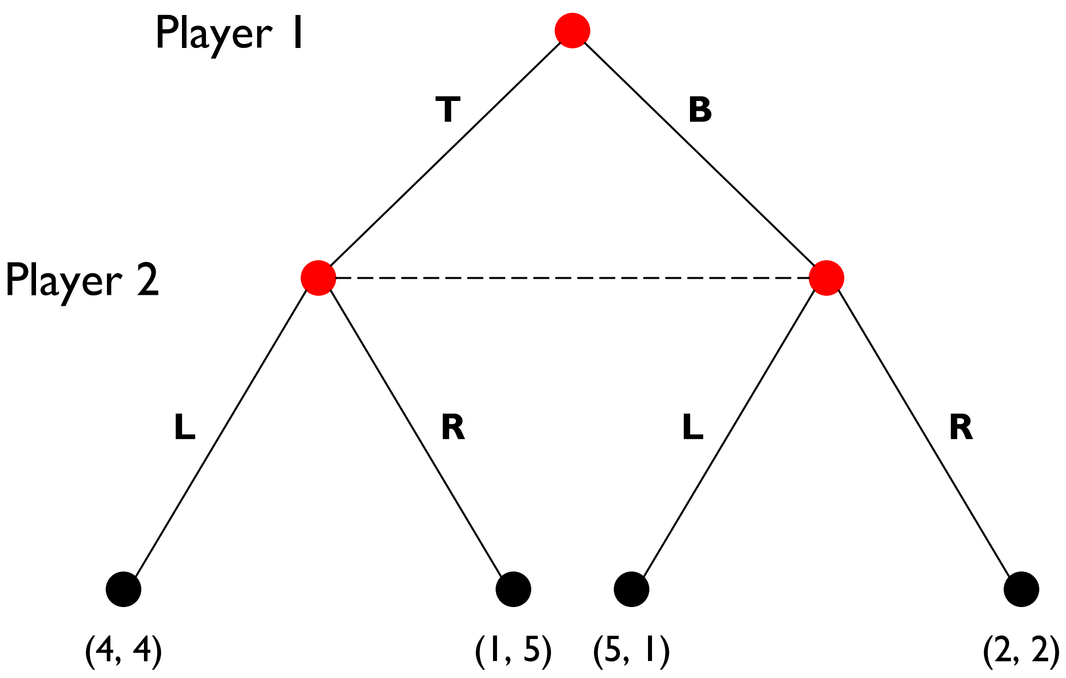
$G = (N, A, U)$

- set of *players* N
- set of *actions* $A = \times_{i \in N} A_i$, where A_i is the set of actions of player i
- set of *utility* functions $U = (U_1, U_2, \ldots, U_n)$, where $U_i : A \rightarrow \mathbb{R}$

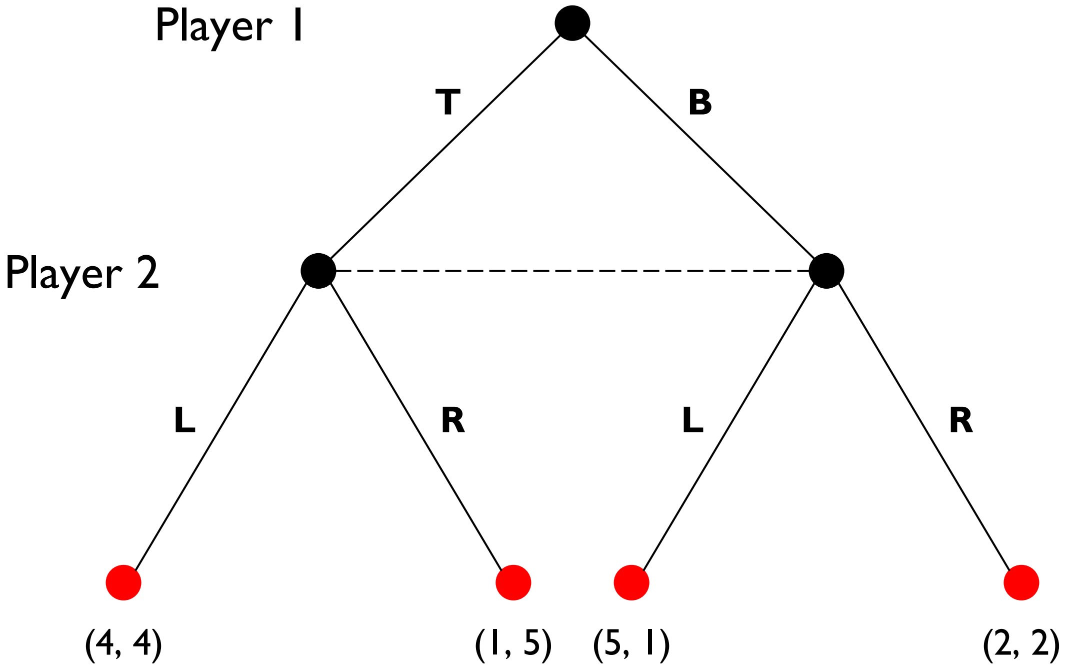
Extensive-form games



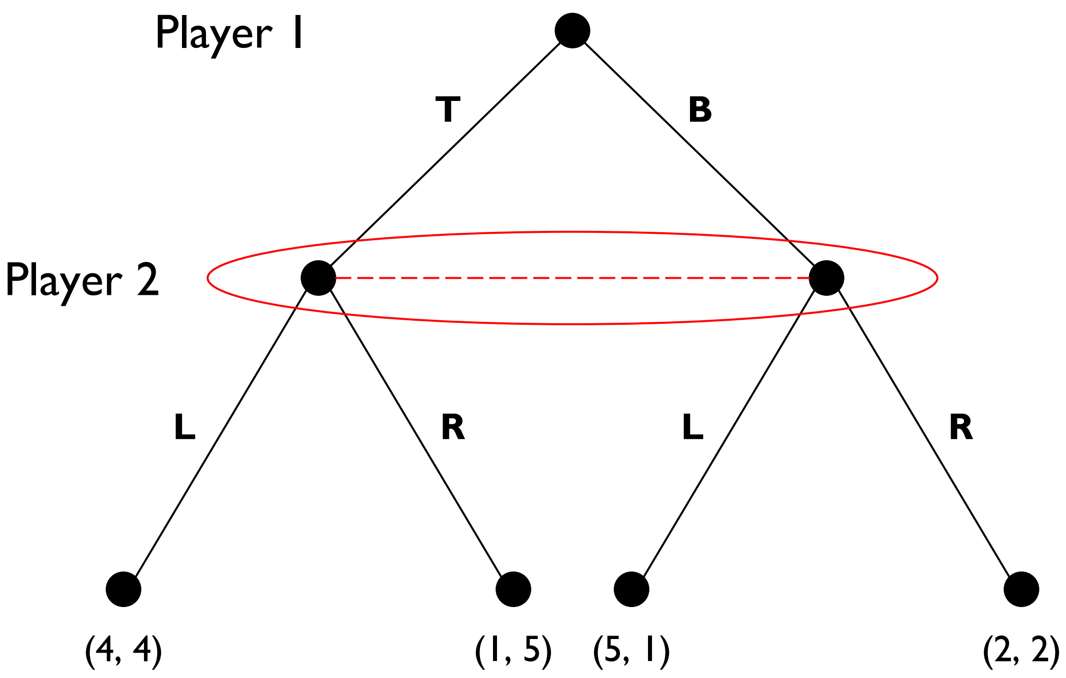
Extensive-form games - Decision nodes



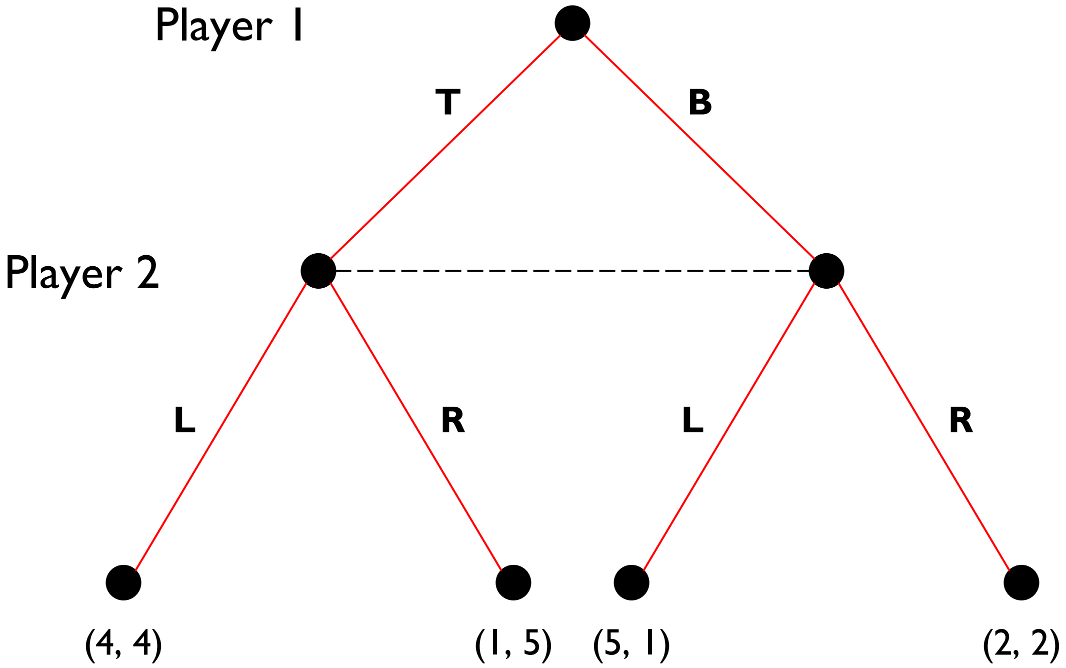
Extensive-form games - Terminal nodes



Extensive-form games - Information sets



Extensive-form games - Successor function



Extensive-form games

$\Gamma = (N, V, H, A, L, \chi, U)$

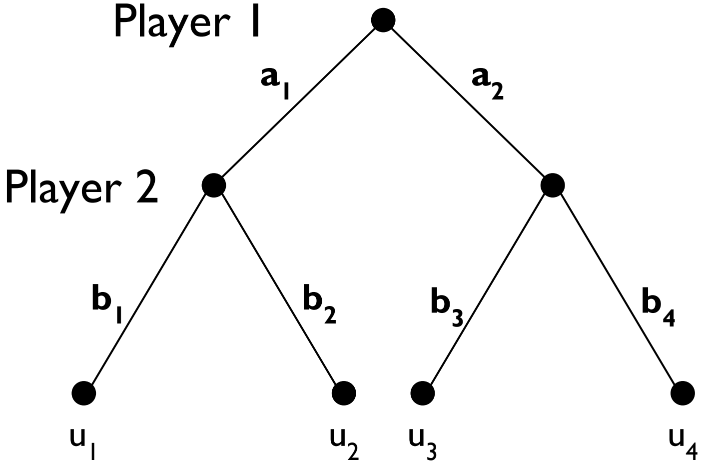
- set of nonterminal *decision nodes* V (each belonging to some player)
- set of *information sets* $H = (H_1, H_2, \ldots, H_m)$, such that H is a partition of V; each information set represent a set of indistinguishable nodes
- set of *terminal nodes* L (leaves)
- *successor function* $\chi : V \times A \rightarrow V \cup L$

Extensive-form games

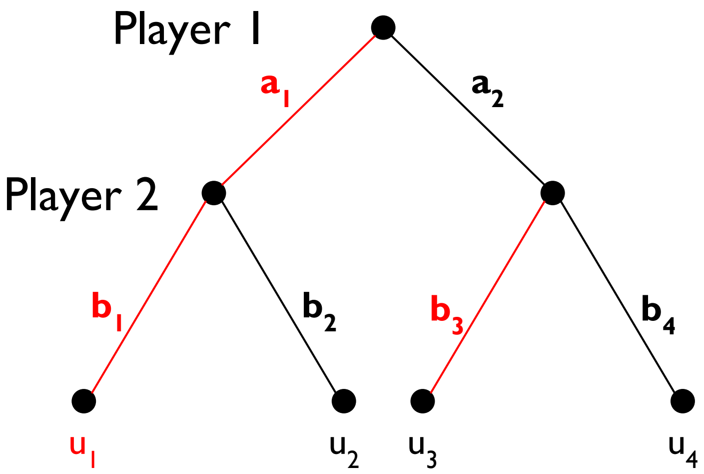
$\Gamma = (N, V, H, A, L, \chi, U)$

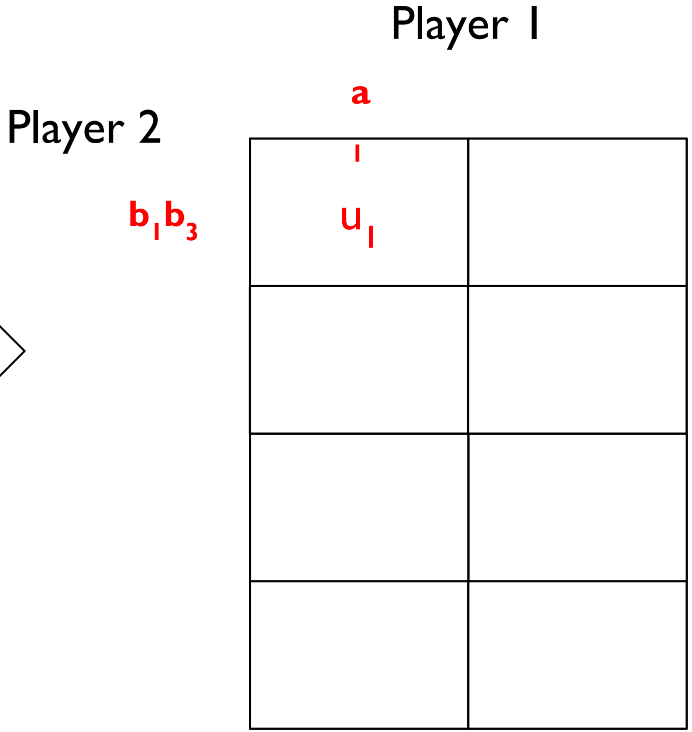
- set of *players* N
- set of *actions* $A = \times_{h \in H} C_h$, where C_h is the set of actions of information set h
- set of *utility* functions $U = (U_1, U_2, \ldots, U_n)$, where $U_i : L \rightarrow \mathbb{R}$

Equivalence between representations

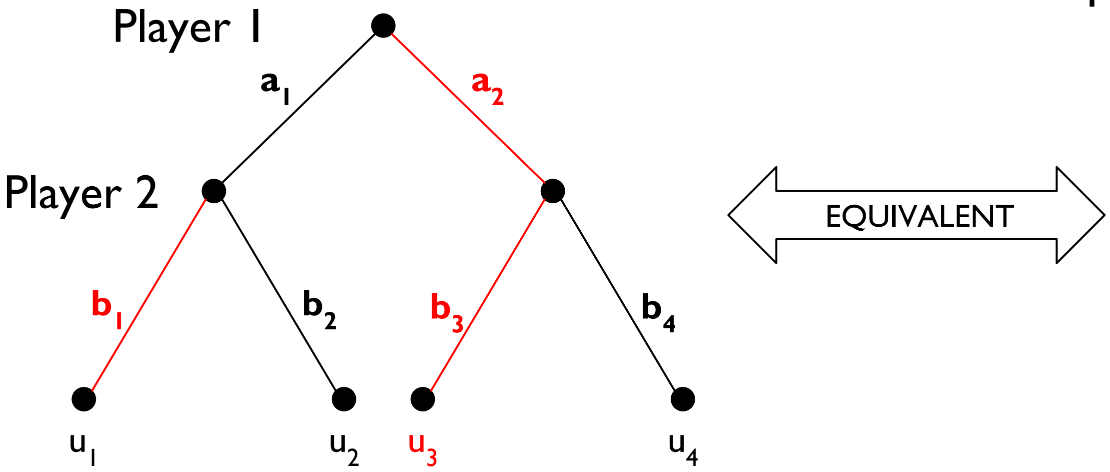


Equivalence between representations





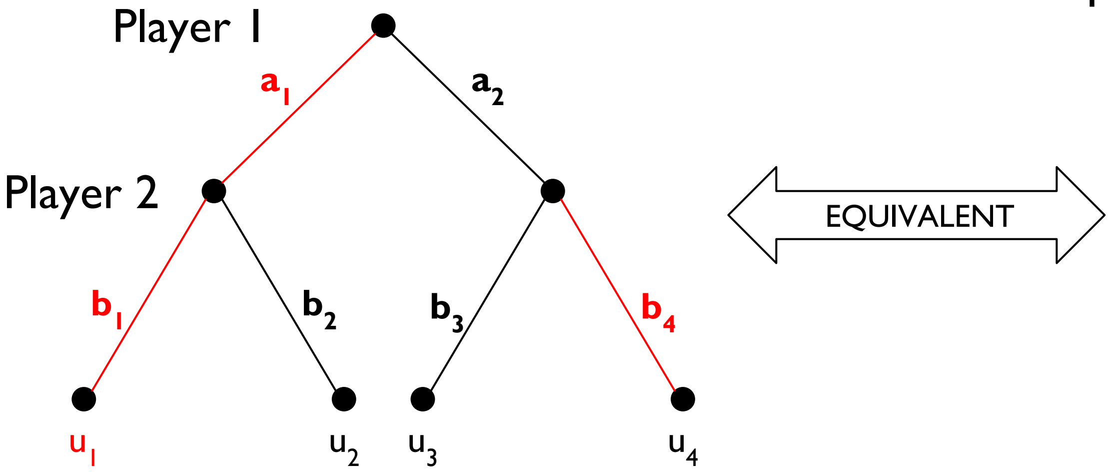
Equivalence between representations



Player 1

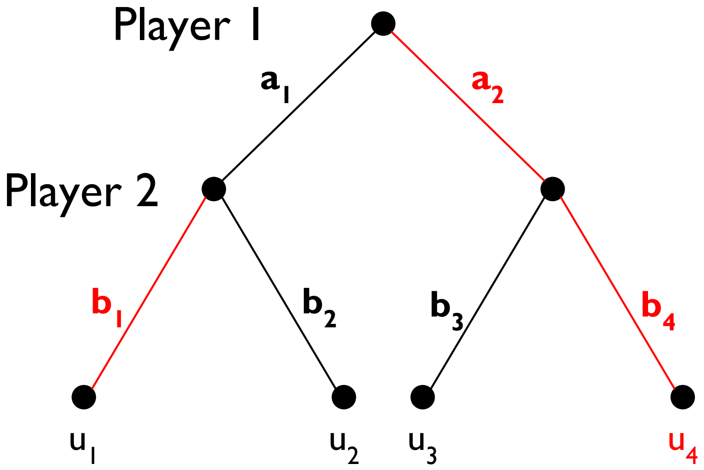
Player 2	a_1	a_2
b_1b_3	u_1	u_3

Equivalence between representations



Player 2 \ Player 1	a_1	a_2
b_1b_3	u_1	u_3
b_1b_4	u_1	

Equivalence between representations

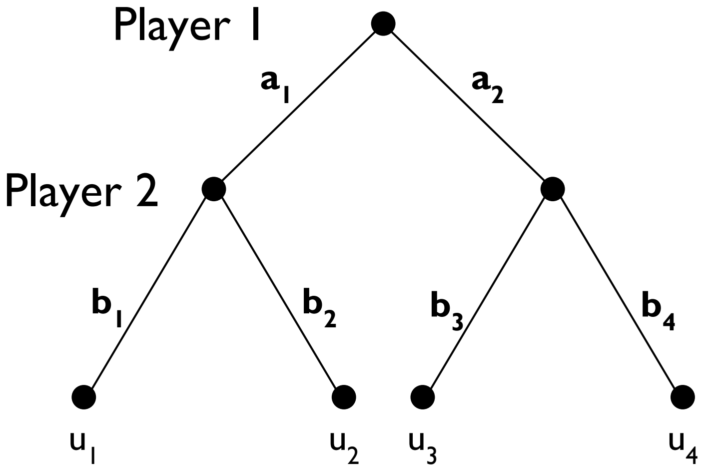


EQUIVALENT

Player 1

Player 2	a_1	a_2
b_1b_3	u_1	u_3
b_1b_4	u_1	u_4

Equivalence between representations

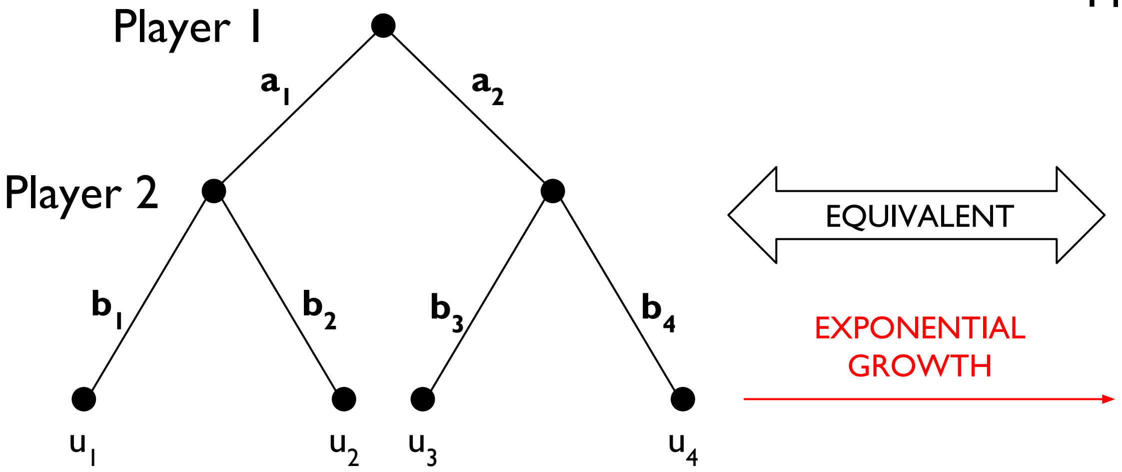


EQUIVALENT

Player 1

Player 2	a_1	a_2
b_1b_3	u_1	u_3
b_1b_4	u_1	u_4
b_2b_3	u_2	u_3
b_2b_4	u_2	u_4

Equivalence between representations



Player 2 \ Player 1	a_1	a_2
$b_1 b_3$	u_1	u_3
$b_1 b_4$	u_1	u_4
$b_2 b_3$	u_2	u_3
$b_2 b_4$	u_2	u_4

Equivalence between representations

The size of the normal-form equivalent game is, in general, **exponential** in the size of the extensive-form one

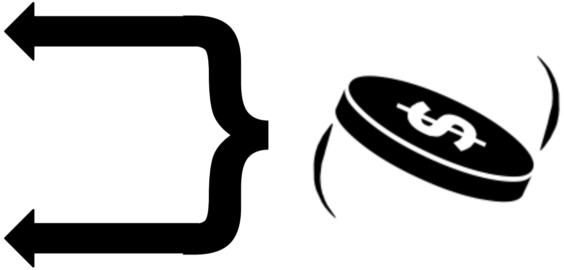
Definitions

- *zero-sum* game: for each outcome, the utilities of all the players sum to zero
- *imperfect-recall* game: players may forget what actions have they chosen at previous information sets

Normal-form (mixed) strategies

Player 2

		L	R
Player 1	T	(4, 4)	(1, 5)
	B	(5, 1)	(2, 2)



Normal-form (mixed) strategies

- a strategy for player i is a function $s_i : A_i \rightarrow \Delta^{|A_i|}$
- a strategy profile s = (s_1, s_2, …, s_n) is a collection of strategies for all the players
- S = (S_1, S_2, …, S_n) is the space of normal-form strategy profiles

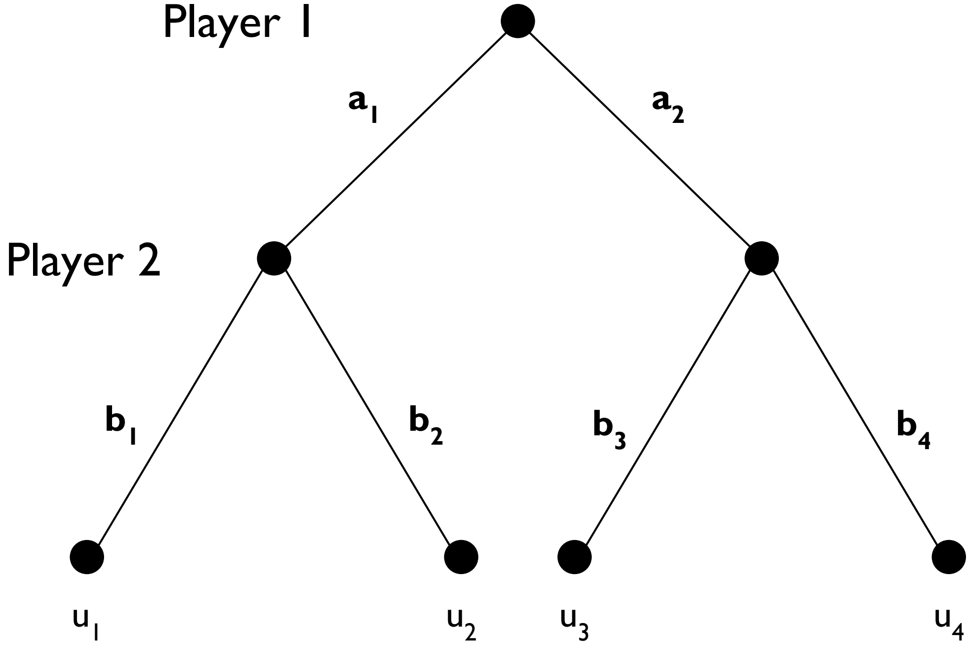
Joint (mixed) strategies

- a joint strategy is a function $x : A \to \Delta^{|A|}$
- a joint strategy enables player to *correlate* their decisions (actions)
- in general, joint strategies are not marginalizable

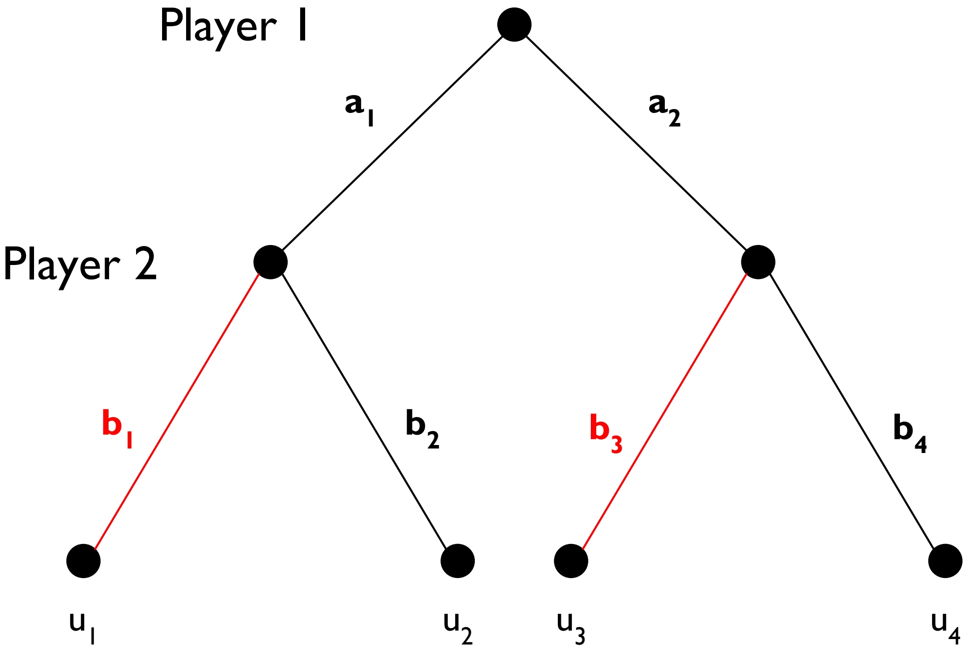
Behavioural strategies

- a strategy for player i is a function $\sigma_i : H_i \rightarrow \Delta^{|C_{H_i}|}$
- a behavioural strategy profile $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ is a collection of behavioural strategies for all the players
- $\Sigma = (\Sigma_1, \Sigma_2, \ldots, \Sigma_n)$ is the space of behavioural strategy profiles

Normal-form vs behavioural strategies

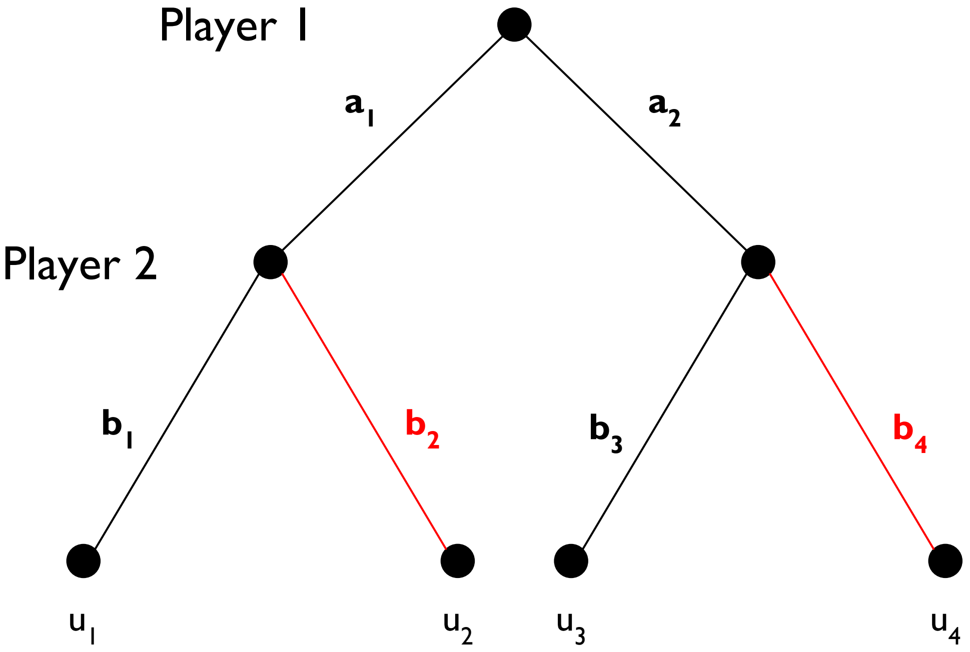


Normal-form vs behavioural strategies



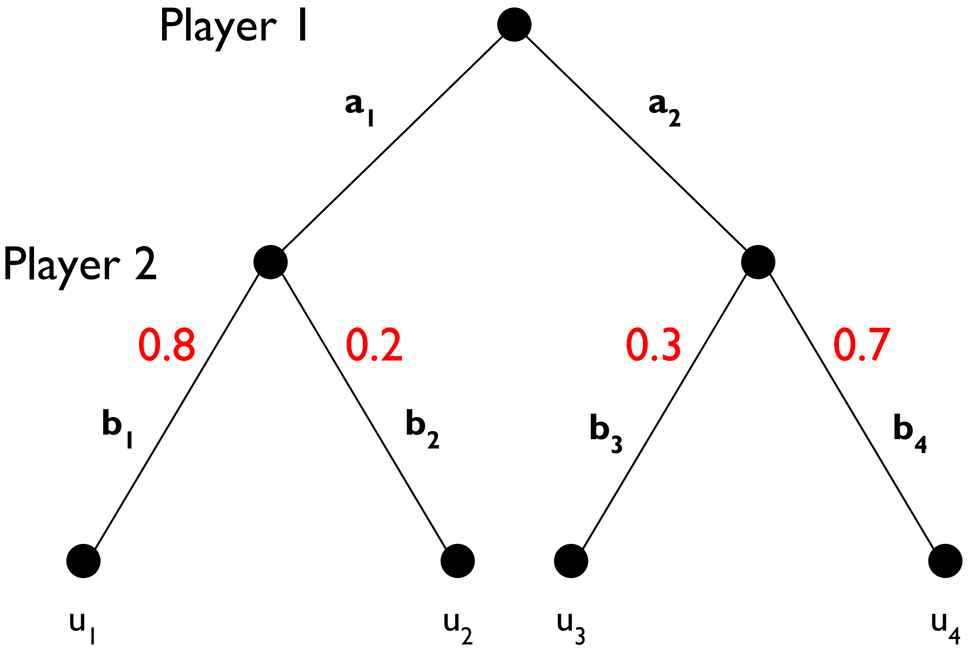
$s_1(b_1 b_3) = 0.8$

Normal-form vs behavioural strategies



$s_1(b_2 b_4) = 0.2$

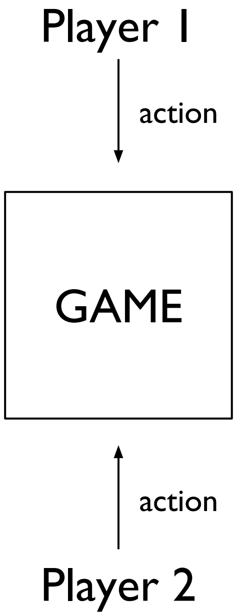
Normal-form vs behavioural strategies



Equilibria

- *Nash Equilibrium (NE)*: players have no incentive to deviate, given the strategies of the other players

Equilibria



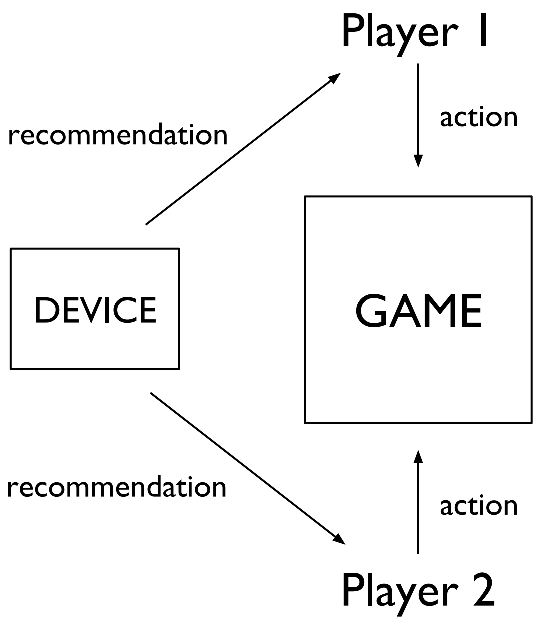
Equilibria

- *Nash Equilibrium (NE)*: players have no incentive to deviate, given the strategies of the other players

- *Correlated Equilibrium (CE)*: players have no incentive to deviate, given a recommendation received from a correlation device

Equilibria

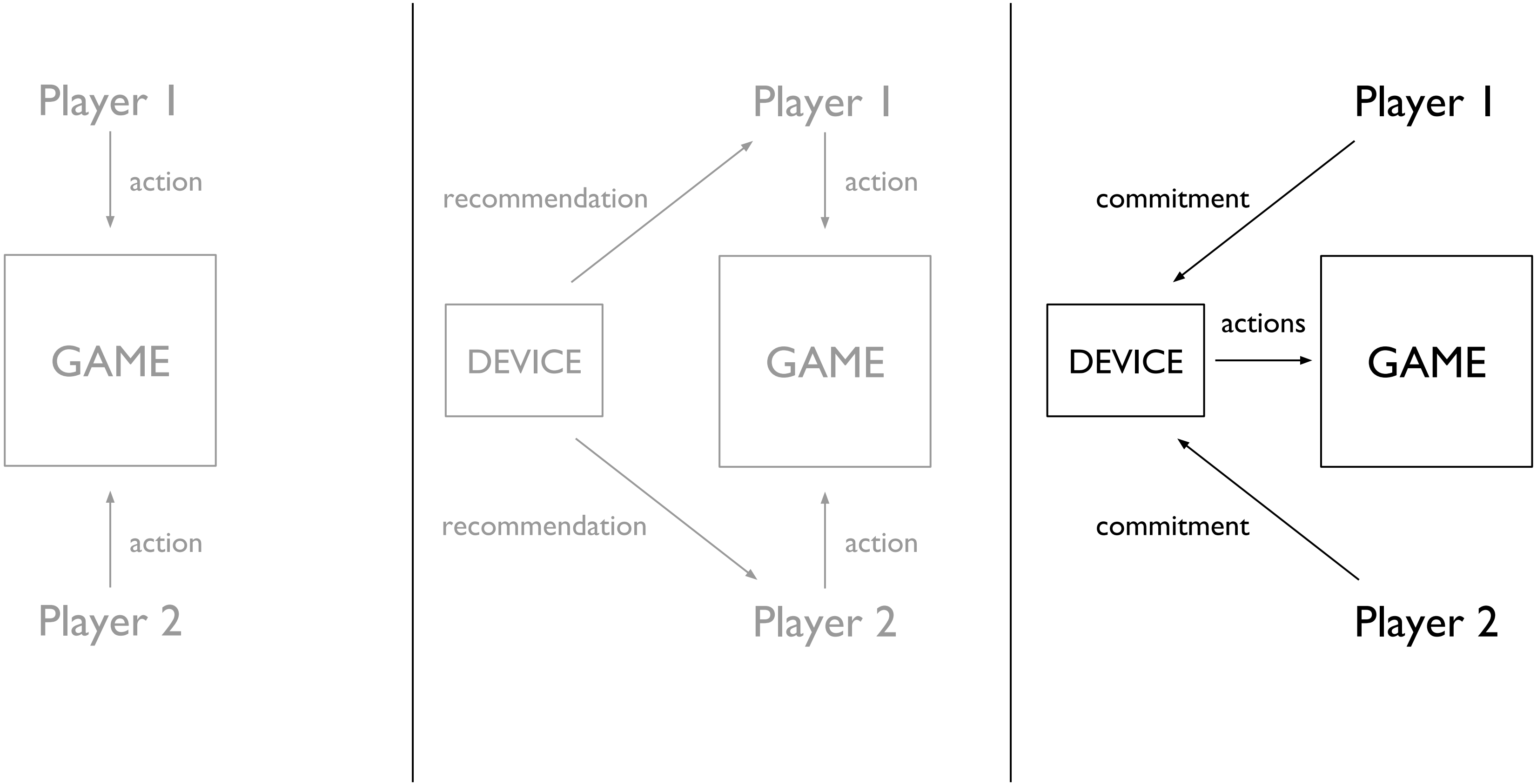

Player 1
action
GAME
action
Player 2


Player 1
recommendation
action
DEVICE
GAME
recommendation
action
Player 2

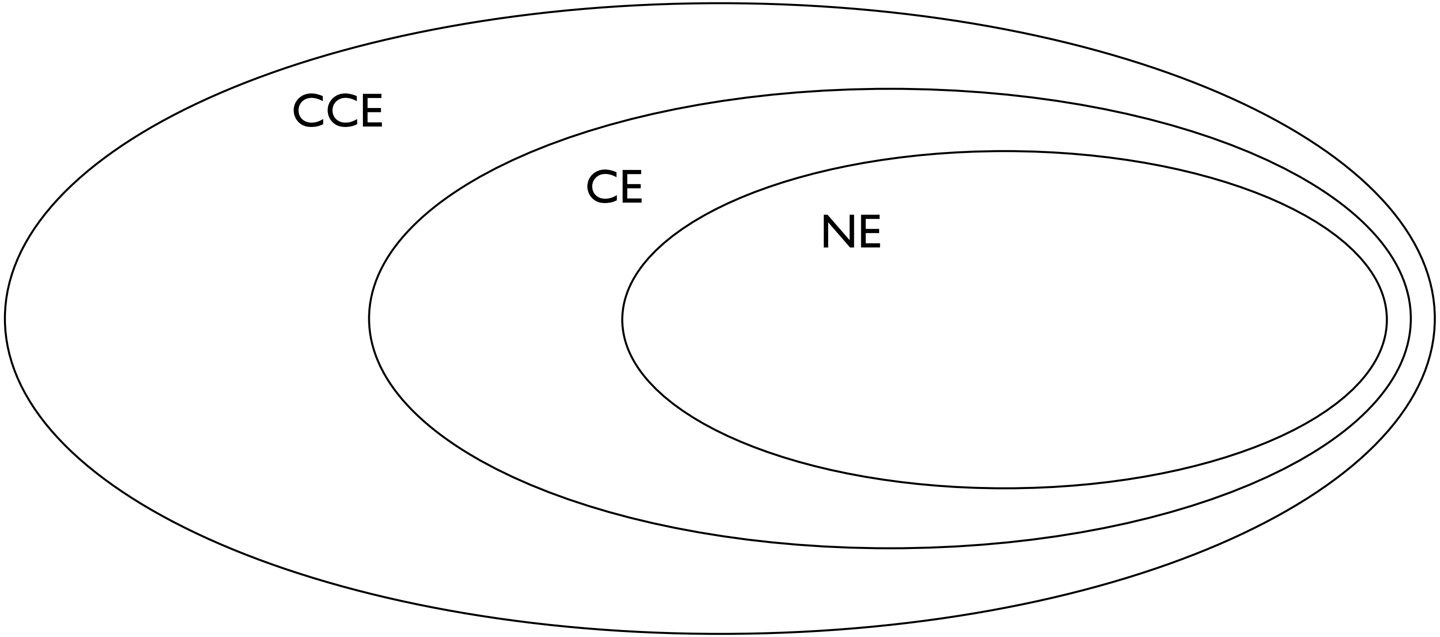
Equilibria

- *Nash Equilibrium (NE)*: players have no incentive to deviate, given the strategies of the other players
- *Correlated Equilibrium (CE)*: players have no incentive to deviate, given a recommendation received from a correlation device
- *Coarse Correlated Equilibrium (CCE)*: players have no incentive to deviate, given the distribution from which the correlation device will sample recommendations (i.e. before receiving the actual recommendation)

Equilibria


Player 1
action
GAME
action
Player 2
Player 1
recommendation
action
DEVICE
GAME
recommendation
action
Player 2
Player 1
commitment
actions
DEVICE
GAME
commitment
Player 2

Equilibria



NE ⊆ CE ⊆ CCE, and there are games for which those inclusions are strict

Nash equilibrium (NE)

equilibrium strategy profile

$$\underbrace{U_i(\hat{s})}_{\text{expected utility given a strategy profile}} \geq \max_{s_i \in S_i} U_i(s_i, \hat{s}_{-i})$$

unilateral deviation of player i

Correlated equilibrium (CE)

equilibrium joint strategy

$$\sum_{a_{-i} \in A_{-i}} x(a_i, a_{-i})(U(a_i, a_{-i}) - U(a'_i, a_{-i})) \geq 0$$

$$\forall a_i \in A_i \; \forall a'_i \in A_i$$

Correlated equilibrium (CE)

equilibrium joint strategy

$$\sum_{a_{-i} \in A_{-i}} x(a_i, a_{-i})(U(a_i, a_{-i}) - U(a_i', a_{-i})) \geq 0$$

$$\forall a_i \in A_i \; \forall a_i' \in A_i$$

recommendation

Correlated equilibrium (CE)

equilibrium joint strategy

$$\sum_{a_{-i} \in A_{-i}} x(a_i, a_{-i})(U(a_i, a_{-i}) - U(a'_i, a_{-i})) \geq 0$$

$$\forall a_i \in A_i \; \forall a'_i \in A_i$$

unilateral deviation
of player i

Coarse Correlated equilibrium (CCE)

equilibrium joint strategy

$$\sum_{a_i \in A_i} \sum_{a_{-i} \in A_{-i}} x(a_i, a_{-i})(U(a_i, a_{-i}) - U(a_i', a_{-i})) \geq 0$$

$$\forall a_i' \in A_i$$

Coarse Correlated equilibrium (CCE)

equilibrium joint strategy

$$\sum_{a_i \in A_i} \sum_{a_{-i} \in A_{-i}} x(a_i, a_{-i})(U(a_i, a_{-i}) - U(a_i', a_{-i})) \geq 0$$

$$\forall a_i' \in A_i$$

recommendations

Coarse Correlated equilibrium (CCE)

equilibrium joint strategy

$$\sum_{a_i \in A_i} \sum_{a_{-i} \in A_{-i}} x(a_i, a_{-i})(U(a_i, a_{-i}) - U(a'_i, a_{-i})) \geq 0$$

$$\forall a'_i \in A_i$$

unilateral deviation
of player i

How to build a correlation device

- Single centralized computation of an optimal joint strategy
- Single centralized computation of an approximate joint strategy
- Decentralized decisions, where players can use their own regret to autonomously organize to reach a good social outcome

Regret

		Player 2	
		L	R
Player 1	T	(4, 4)	(1, 5)
	B	(5, 1)	(2, 2)

Regret

Player 2

Player 1		L	R
	T	(4, 4)	(1, 5)
	B	(5, 1)	(2, 2)

Player 1 regrets not having played action B, which had an higher utility

Regret

- measure of how much a player would have preferred to play a different strategy
- *external regret* : regret with respect to the best constant-action strategy
- *internal regret* : regret of not having played action a_j each time action a_i was played

Hannan consistency

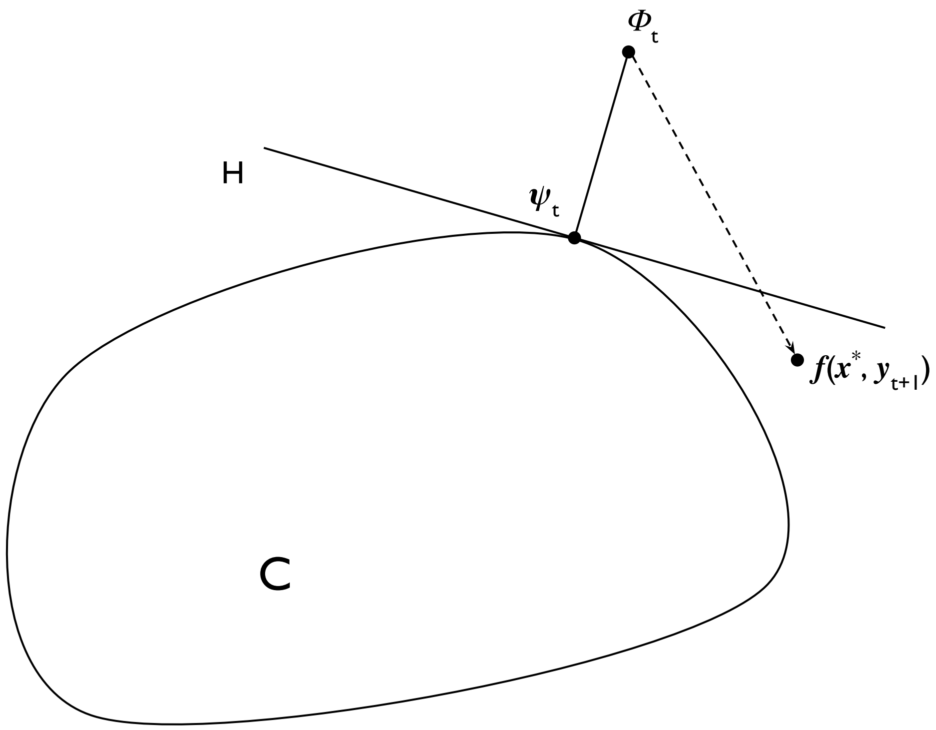
$$\lim_{T\to\infty} \sup \frac{1}{T}\left(\max_{a_i\in A_i} \sum_{t=1}^{T} U_i(a_i, s^t_{-i}) - \sum_{t=1}^{T} U_i(s^t)\right) = 0$$

- defined on the infinitely repeated game (each t ∈ [0, T] is a play of the game)
- equivalent to minimization of *external regret*
- equivalent to approaching the set of CCEs

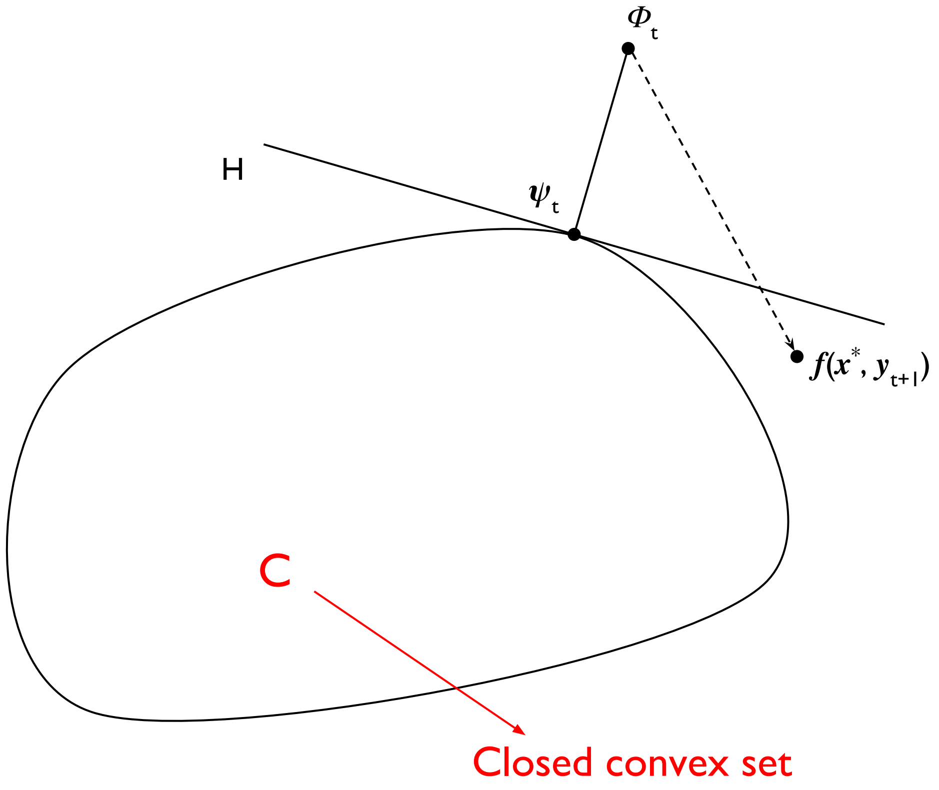
Blackwell Approachability

- 2-player vector-valued game, with utilities in $\mathbb{R}^L$
- Goal of player 1 is to make the average utility Φ as close as possible to a target convex set C
- Goal of player 2 is to prevent it
- The set C is said to be approachable if the player 1 has a strategy that, no matter what the player 2 does, guarantees him that the distance between its average utility and C uniformly goes to zero as t grows to infinity

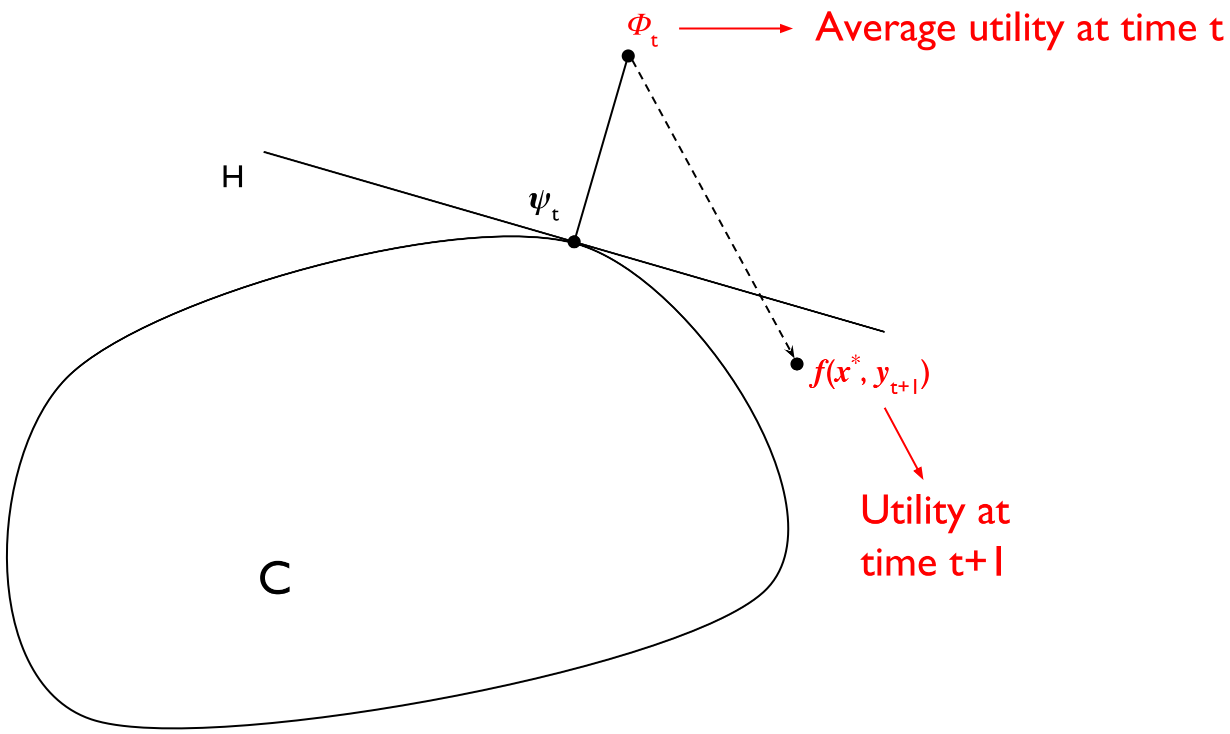
Blackwell Approachability



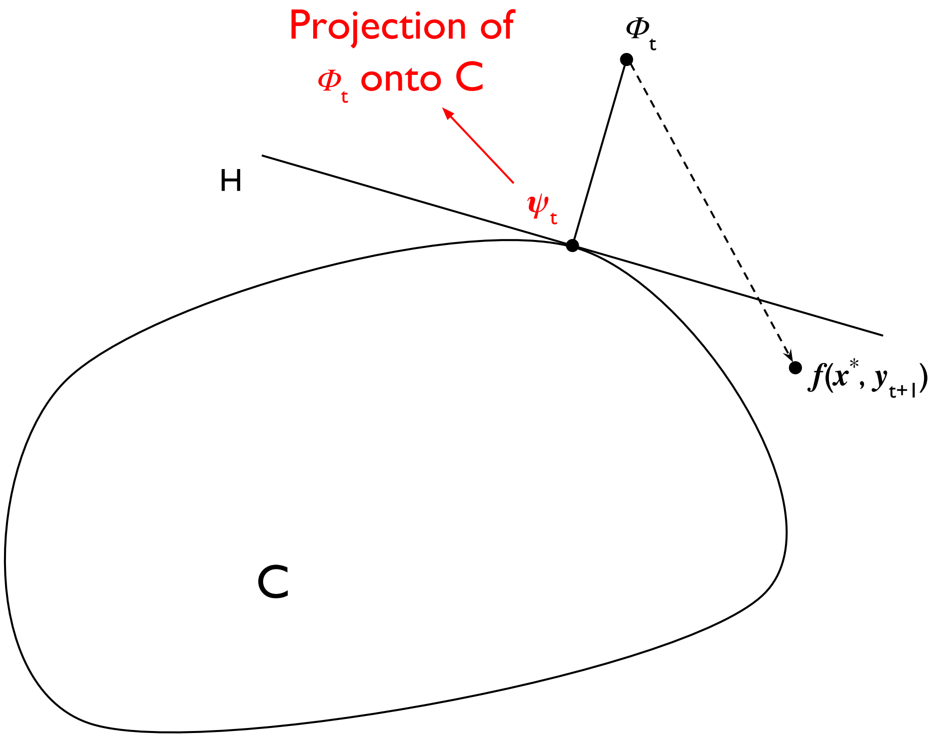
Blackwell Approachability



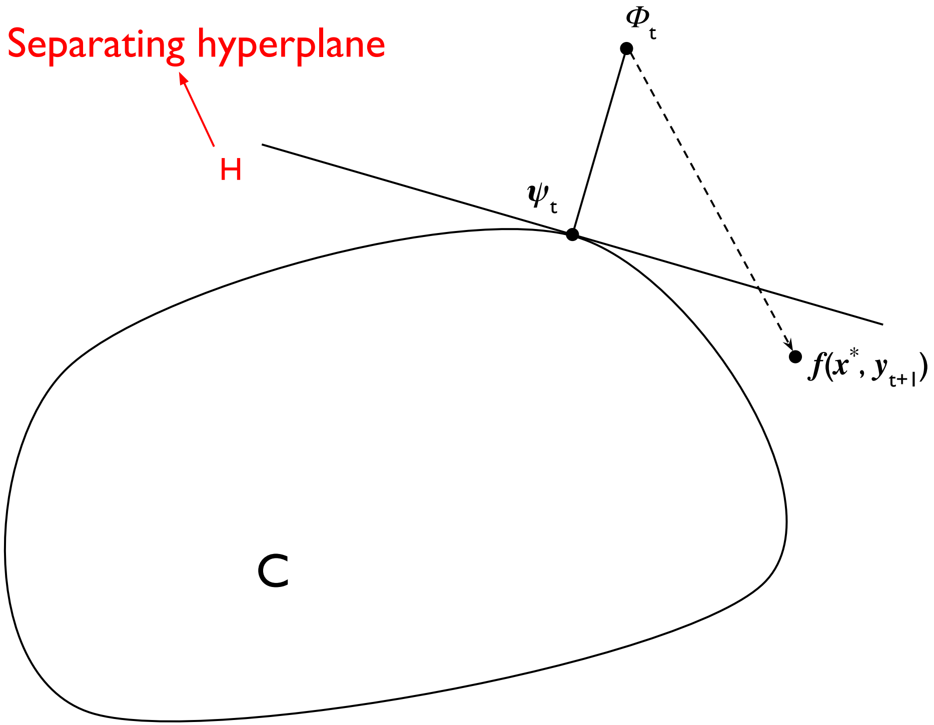
Blackwell Approachability



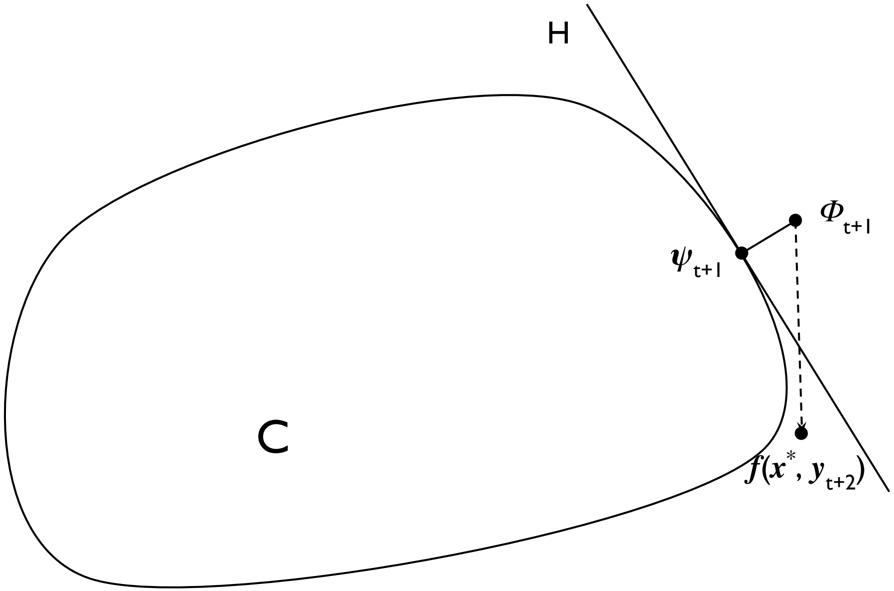
Blackwell Approachability



Blackwell Approachability



Blackwell Approachability



Blackwell Approachability Theorem

A closed convex set C is approachable if and only if every halfspace H $\supseteq$ C is forceable

A set H is said to be forceable if there exists an action $\boldsymbol{x}^*$ such that whatever the action $\boldsymbol{y}$ used by the opponent, $\boldsymbol{f}(\boldsymbol{x}^*, \boldsymbol{y}) \in$ H, where $\boldsymbol{f}$ is the vector-valued utility function

Approachability for regret minimizers

- The vector-valued utility of the game is defined by the vector of average regrets (one for each action in the case of external regret)

- The set to approach is the negative orthant $\mathbb{R}_-^L = \{ x \in \mathbb{R}^L : x \leq 0 \}$

- Approaching $\mathbb{R}_-^L$ is thus equivalent to minimize the average regret

State of the Art

- Research topic
- Main works
- Open problems

Regret Minimization for Non-Cooperative Games

- Decomposition of regret over a set of independent terms, that can be minimized separately (if the game representation allows it)
- Development of fast, scalable and computationally efficient algorithms
- Theoretical guarantees on convergence and rate of convergence (so-called regret bound)
- Characterization of the set of equilibria reached by the regret dynamics

Literature classification

	2 players zero-sum	N players zero-sum	N players general-sum
NE normal-form	[1], [2]		[10], [11]
NE extensive-form	[8], [9], [20], [22], [23], [30], [31]	[13], [18], [28]	[12]
CE normal-form	[16]		
CE extensive-form			
CCE normal-form	[17]		
CCE Bayesian	[19]		
CCE extensive-form			
Other	[4], [6], [14], [24], [25]		

Main research lines

- Solving large 2-player zero-sum extensive-form games (CFR algorithm and following refinements)

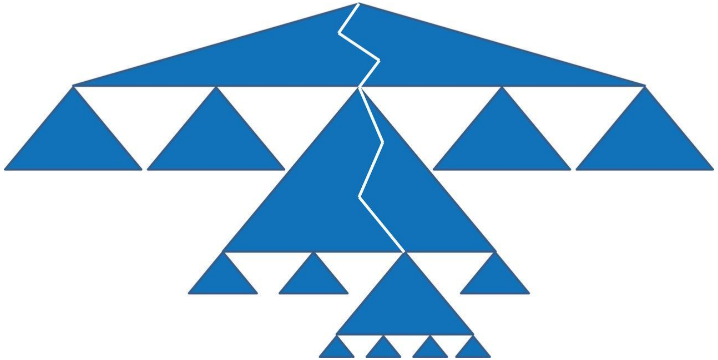
- Finding simple adaptive procedures leading to game theoretical equilibria (Hart and Mas Colell, dealing in particular with CE and CCE for general normal-form games)

- Application of regret-based approaches to find equilibria in more specialized game representations, such as Bayesian games or Security games

Main practical results

- In 2015, Heads-up limit Hold'em Poker is solved, using techniques based on CFR; it is the first imperfect-information game competitively played by humans to be solved

- In 2017 Libratus, an algorithm again based on CFR-like techniques, solved Heads-up no-limit Hold'em Poker beating by a large margin four world-class champions; this work won the 2018 *Marvin Minsky Medal* for Outstanding Achievements in Artificial Intelligence



Issues

- Most of the known results are for the 2-player zero-sum setting, which is the less general one
- Other compact game representations (like polymatrix games) totally lack regret-based algorithms
- For extensive-form games, no substantial research has been done for the imperfect-recall setting

Open problems

- How to approach Nash Equilibria for n-players general-sum extensive-form games
- How to approach CE and CCE in games other than normal-form ones, and in particular how to reach compactly representable equilibria (i.e. joint strategies with small support)
- How good (in terms of attained utility) are the equilibria found by regret minimization dynamics, with respect to the optimal ones

Project Proposal

- Project definition
- Tasks
- Challenges

Project objective

Efficiently find Coarse Correlated Equilibria (CCE) in compactly representable games

Why - Game representation

- Extensive form games are both more general and more natural than normal-form ones, as they allow the modelization of imperfect information

Why - Equilibrium

- CCE is per se a very natural solution concept, with an easy to implement correlation device

- CCE has very good properties in terms of computational complexity

- CCE can yield arbitrarily larger social utility with respect to both NE and CE

Why - Literature gap

- No one has addressed the problem of regret-based CCE computation for games other than normal-form ones

Main tasks

- CCE for 2-player extensive-form games
- CCE for n-players extensive-form games
- CCE for n-players polymatrix games

Main subtasks

For each of the main task:

- Develop the regret minimization algorithm
- Prove the regret bound
- Experimentally evaluate a prototypal implementation of the algorithm
- Debate on the quality of the equilibria reached by the algorithm, referring in particular to the concept of Price of Total Anarchy

Challenges

- Reaching “small” joint strategies
- Exploiting the structure of the game representation to decompose the regret
- Finding if the regret dynamics in compact game representations leads to the same subset of equilibria reached by the normal-form ones
- Finding how close the equilibria reached by regret minimization are with respect to the optimal one

Thank you

1. preliminari: cos’è un gioco, concetti di soluzione, … norma normale di un gioco in forma estesa
2. Regret intuizione (esempio), approachability con intuizione (esempio), combinare le cose — bound sul regret che va zero
3. stato dell’arte (tabella e discussione)
4. proposta di progetto

- definition del progetto
- struttura a task del progetto
- sfide aperte (decomposizione del regret) — far capire perchè non è ovvio il risultato
- …

No ggant.

Equivalence between representations: questa slide va cambiata perchè troppo scritta. Io riporterei a sinistra un albero piccolo, a destra una bimatrice con una doppia freccia e sotto scritto “strategic equivalence”, oppure solo “equivalence”, dicendo che ogni albero può essere rappresentato in modo equivalente con una forma normale e vice versa. A questo punto illumini una azione della forma normale e illumini il piano corrispondente sull’albero e dici che ogni azione della forma normale equivalente corrisponde a un piano di azione sull’albero. Chiudi dicendo che il numero di piani è esponenziale e quindi la forma normale è esponenzialmente grossa. **OK**

Magari cambierei il titolo delle slides "Normal-form strategies" in “Normal-form (mixed) strategies”. Dal punto di vista formale non serve, ma magari attira maggiormente l’attenzione degli ascoltatori. Stessa cosa per “Joint strategies” in “Joint (mixed) strategies”. **OK**

Quando introduci gli equilibri, io farei una slide che specifica le differenze tra questi, in modo che siano chiare le differenze concettuali. **OK**

equivalent to approaching the set of CCE -> CCEs **OK**

Una nota che potrebbe generare casino; quando definisci il regret (vedi external), lo fai di fatto sul singolo istante temporale, mentre nella Hanna consistency fai il cumulativo (poi, per carità, diventa medio perchè dividi per T). A mio avviso va chiarito esattamente cosa stai facendo e le relazioni. Qui il gioco è ripetuto e credo sia la prima volta che questa cosa emerge. Probabilmente nella slide della Hannan consistency va messa una riga in cui dice che il gioco viene ripetuto. **OK**

Nella slide di Blackwell scritta, manca qualcosa dopo “Goal of player 2 is to prevent that” **OK**

Nella slide “Blackwell Approachability Theorem”: il primo credo sia il teorema e lo metterei in corsivo. Toglierei poi i punti “.” Alla fine delle frasi. **OK**

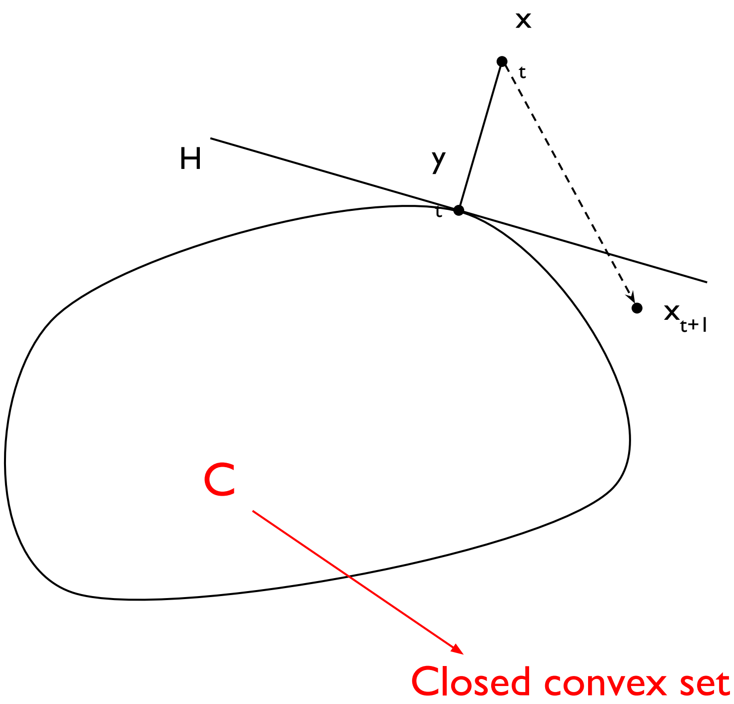
CCE Bayesian con la B maiuscola (tabella dello stato dell’arte) **OK**

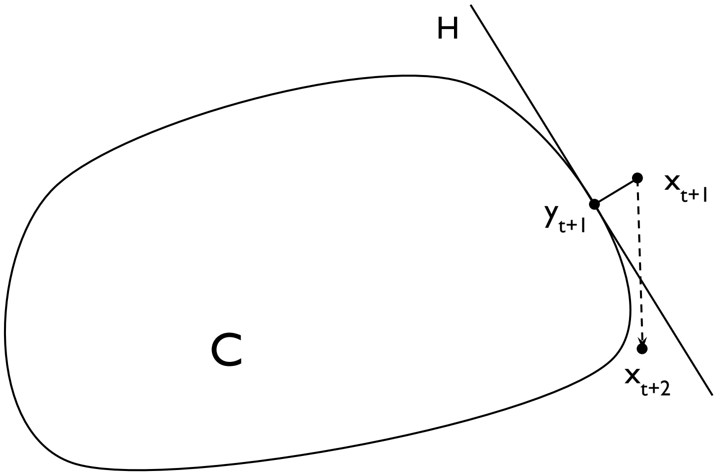
Nelle 4 slides che seguono la tabella dobbiamo trovare un modo per accorciare il testo, possiamo farlo assieme quando ci vediamo. Stessa cosa per slides successive con titolo “Why” e “Challenges”. **TODO**

Project objective -> via il punto alla fine della frase **OK**

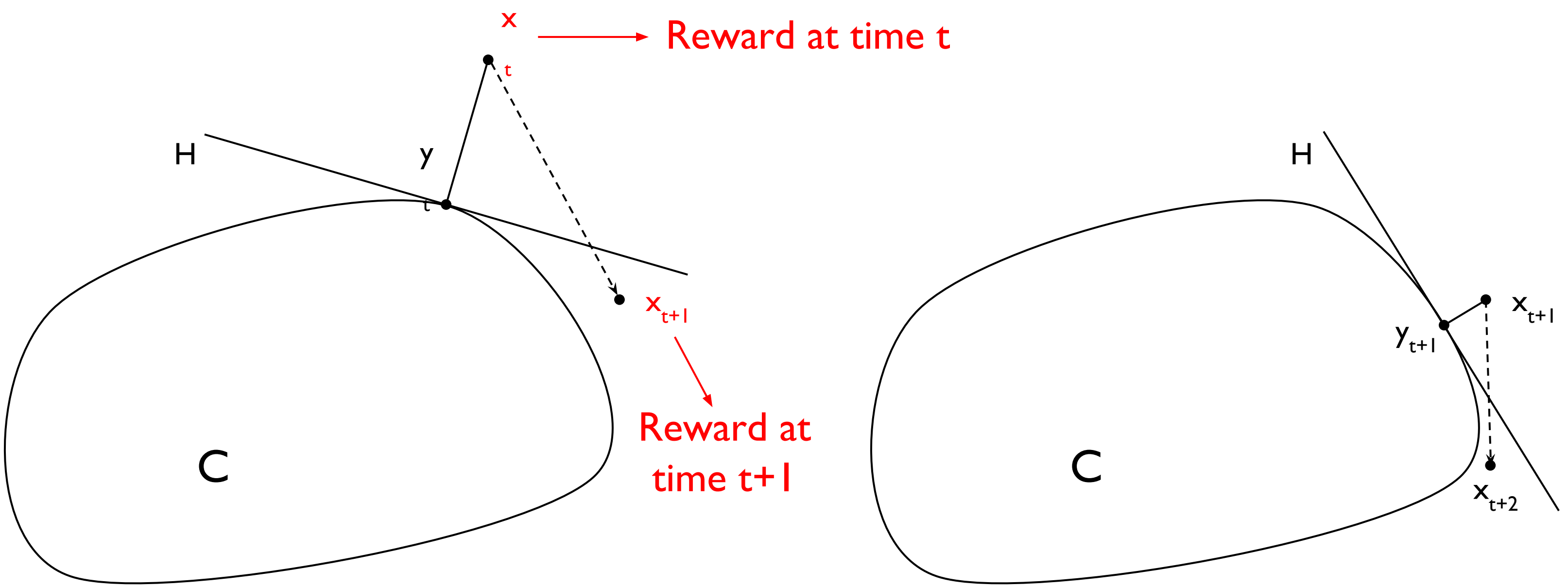
In tutta la presentazione non nomini mai Libratus, che è una cosa che piace sempre. Io lo fare, magari introducendo anche una foto e dicendo che questi algoritmi sono il cuore dell’ultima medaglia Minsky. **OK**

Blackwell Approachability

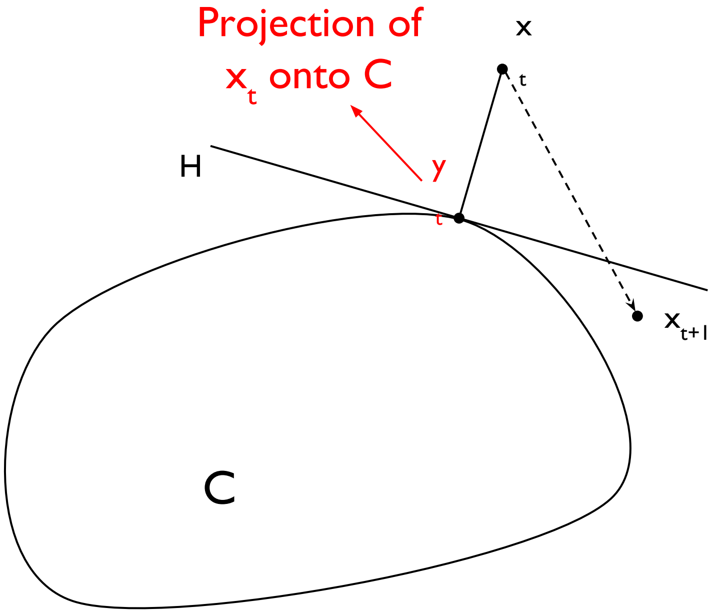


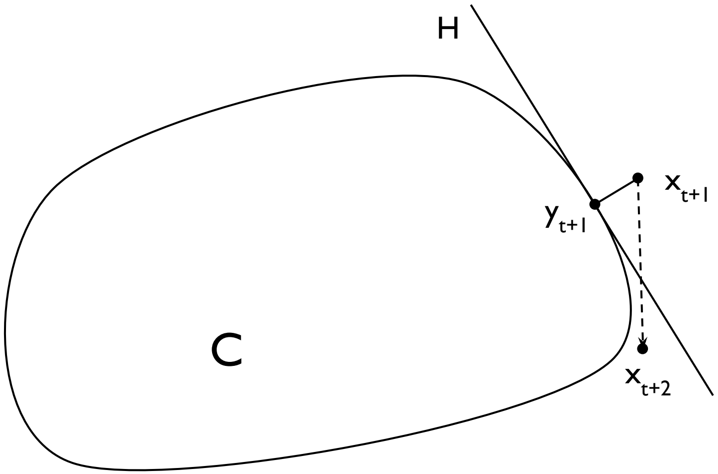


Blackwell Approachability

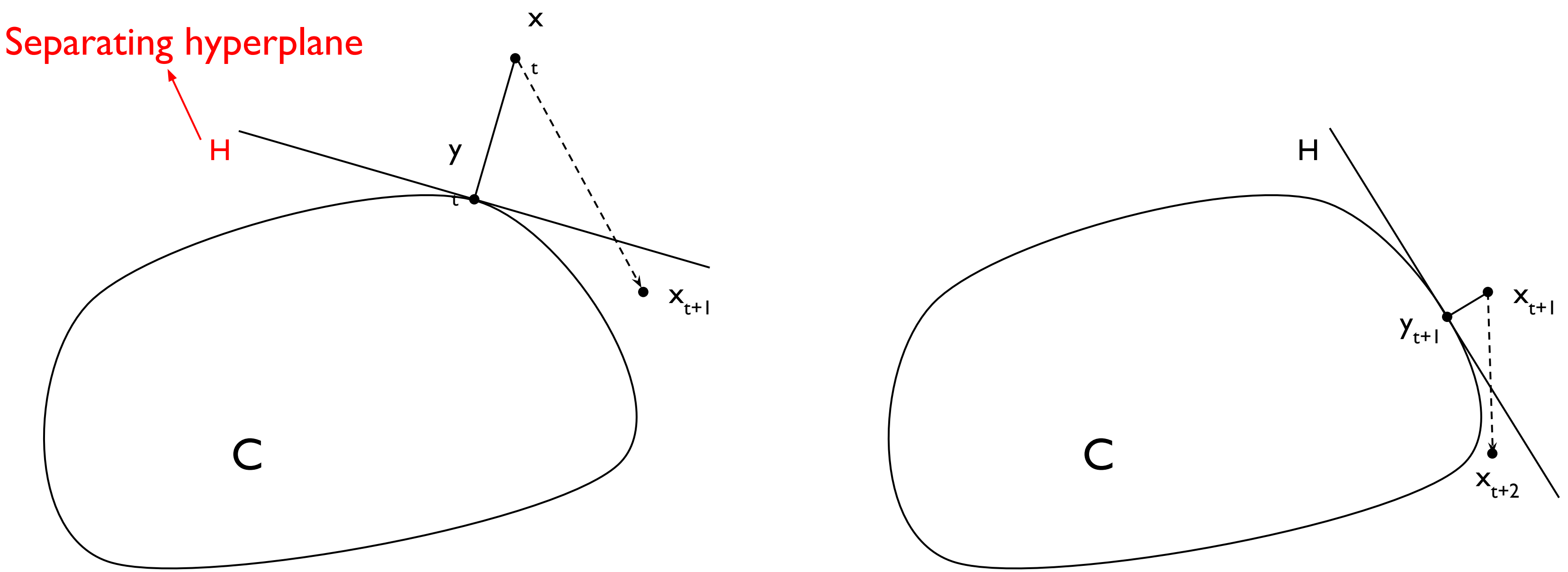


Blackwell Approachability

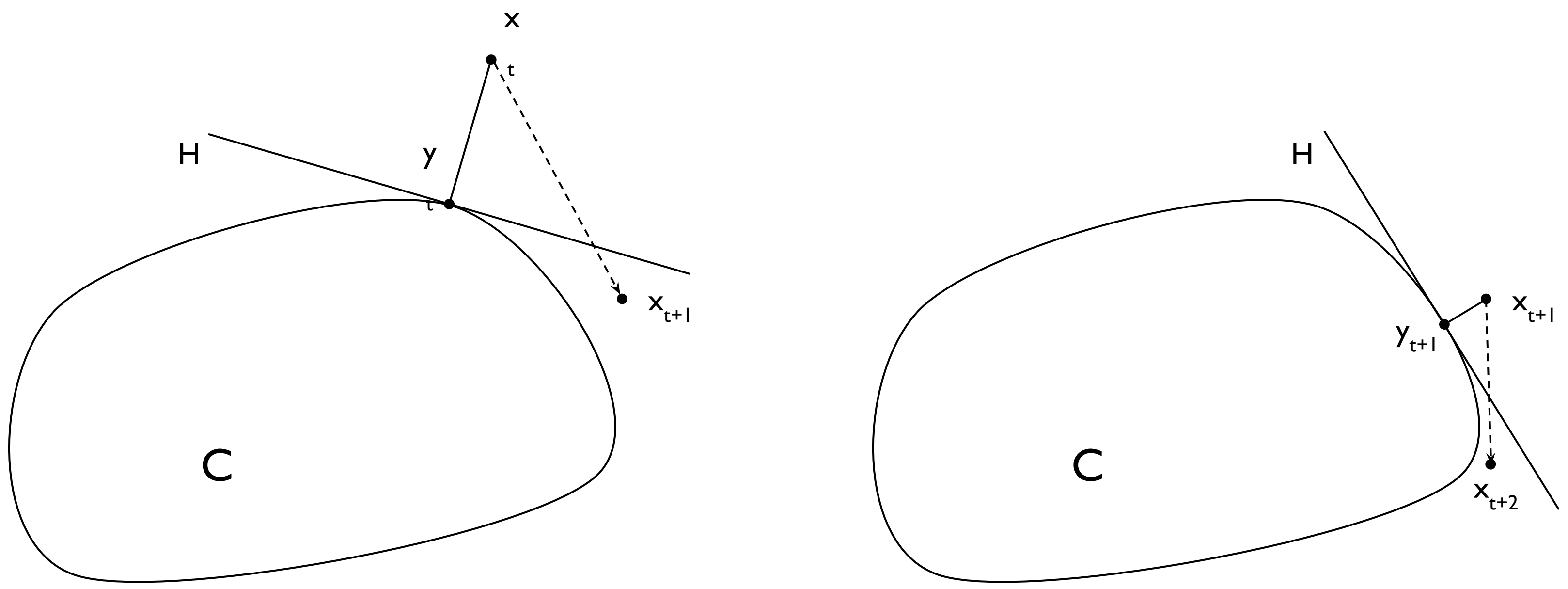




Blackwell Approachability



Blackwell Approachability



Why - semantica **OK**

CCE -> decentralized correlation device -> regret

Risponde alla domanda: come possono organizzarsi autonomamente al meglio i giocatori per raggiungere un outcome

Posso anche usare per costruire efficientemente un device centralizzato (approssimato) **OK**

Slide dopo behavioural con normal form sui piani e behavioural mostrate graficamente **OK**

Slide grafica sugli equilibri **OK**