Fictitious Play

POLITECNICO
MILANO 1863

Rationale

- Fictitious Play is a learning rule introduced by George W. Brown
- In Fictitious Play, each player presumes that the opponents are playing stationary (possibly mixed) strategies
  - At each round t, each player best responds to the empirical frequency of play of their opponent from round 0 to round $t - 1$
  - Such a method is of course adequate if the opponent indeed uses a stationary strategy, while it is flawed if the opponent's strategy is non-stationary
- He imagined that a player would "simulate" play of the game in their mind and update their future play based on this simulation; hence the name *fictitious play*

Assumptions

- The players do not observe the payoffs of the opponents
- The players observe the opponents' actions

Convergence

- In fictitious play strict Nash equilibria are absorbing states (i.e., if at any time period all the players play a Nash equilibrium, then they will do so for all subsequent rounds)

- The process converges for a 2-person game if:

  - Both players have only a finite number of strategies and the game is zero sum (Robinson 1951)

  - The game is solvable by iterated elimination of strictly dominated strategies (Nachbar 1990)

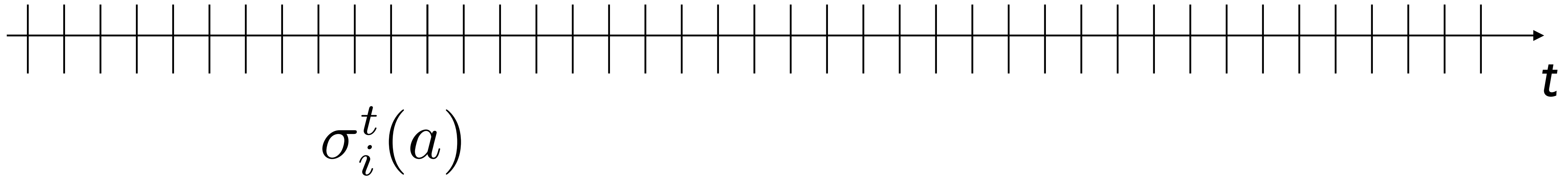
  - The game is a potential game (Monderer and Shapley 1996-a, 1996-b)

  - The game has generic payoffs and is $2 \times N$ (Berger 2005)

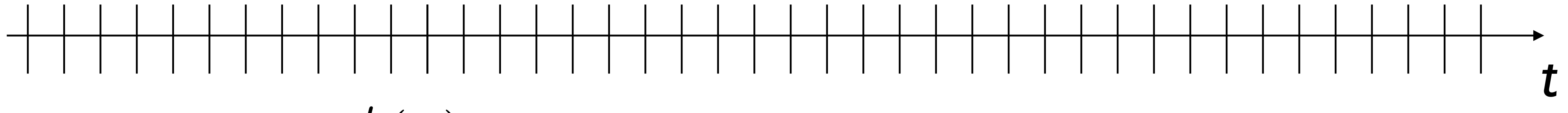
Coordination games (example of potential game)

		2	
	a_4	a_5	a_6
1: a_1	$2, 2$	$4, 4$	$6, 6$
1: a_2	$3, 3$	$0, 0$	$2, 2$
1: a_3	$1, 1$	$5, 5$	$3, 3$

Adaptive strategies



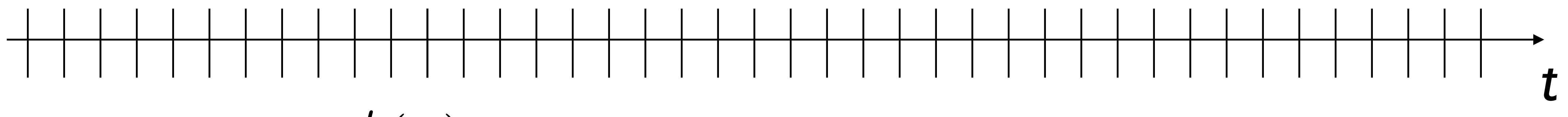
Adaptive strategies



$\sigma_i^t(a)$

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

Adaptive strategies



$\sigma_i^t(a)$

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

$$\sigma_1^1(a) = \begin{cases} \text{R} & 1/3 \\ \text{P} & 1/3 \\ \text{S} & 1/3 \end{cases} \qquad \sigma_2^1(a) = \begin{cases} \text{R} & 1/3 \\ \text{P} & 1/3 \\ \text{S} & 1/3 \end{cases}$$

Fictitious Play (FP) update

At every round, each player plays a best response to the average strategy of the opponent from the initial round to the current one

$$\sigma_1^{t+1} \in \arg\max_{\sigma_1 \in \Delta_1} \mathbb{E}\left[U_1\left(\sigma_1, \frac{1}{t}\sum_{\tau=1}^{t} \sigma_2^{\tau}\right)\right]$$

$$\sigma_2^{t+1} \in \arg\max_{\sigma_2 \in \Delta_2} \mathbb{E}\left[U_2\left(\sigma_2, \frac{1}{t}\sum_{\tau=1}^{t} \sigma_1^{\tau}\right)\right]$$

Example

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1					Player 2			
	Average strategy			BR		Average strategy			BR
	R	P	S			R	P	S	
1	1/3	1/3	1/3		1	1/3	1/3	1/3	
2					2				
3					3				

Example

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1					Player 2			
	Average strategy			BR		Average strategy			BR
	R	P	S			R	P	S	
1	1/3	1/3	1/3	P	1	1/3	1/3	1/3	S
2					2				
3					3				

Example

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1					Player 2			
	Average strategy			BR		Average strategy			BR
	R	P	S			R	P	S	
1	1/3	1/3	1/3	P	1	1/3	1/3	1/3	S
2	1/6	2/3	1/6		2	1/6	1/6	2/3	
3					3				

Example

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1					Player 2			
	Average strategy			BR		Average strategy			BR
	R	P	S			R	P	S	
1	1/3	1/3	1/3	P	1	1/3	1/3	1/3	S
2	1/6	2/3	1/6	P	2	1/6	1/6	2/3	P
3					3				

Example

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1					Player 2			
	Average strategy			BR		Average strategy			BR
	R	P	S			R	P	S	
1	1/3	1/3	1/3	P	1	1/3	1/3	1/3	S
2	1/6	2/3	1/6	P	2	1/6	1/6	2/3	P
3	1/9	7/9	1/9		3	1/9	4/9	4/9	

Example

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

<table>
<tr><th rowspan="3"></th><th colspan="4">Player 1</th><th rowspan="3"></th><th colspan="4">Player 2</th></tr>
<tr><th colspan="3">Average strategy</th><th rowspan="2">BR</th><th colspan="3">Average strategy</th><th rowspan="2">BR</th></tr>
<tr><th>R</th><th>P</th><th>S</th><th>R</th><th>P</th><th>S</th></tr>
<tr><td>1</td><td>1/3</td><td>1/3</td><td>1/3</td><td>P</td><td>1</td><td>1/3</td><td>1/3</td><td>1/3</td><td>S</td></tr>
<tr><td>2</td><td>1/6</td><td>2/3</td><td>1/6</td><td>P</td><td>2</td><td>1/6</td><td>1/6</td><td>2/3</td><td>P</td></tr>
<tr><td>3</td><td>1/9</td><td>7/9</td><td>1/9</td><td>P</td><td>3</td><td>1/9</td><td>4/9</td><td>4/9</td><td>P</td></tr>
</table>

Exploitability

- At every iteration, the strategy returned by FP is an epsilon-Nash
- The value of epsilon is given by the maximum regret (loss) of the player w.r.t. their best response (i.e., the difference between the utility a player gets by playing the best response and the utility given by playing the suggested strategy)

Exploitability

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1								Player 2						
	Average strategy			BR	utility	BR utility	epsilon		Average strategy			BR	utility	BR utility	epsilon
	R	P	S						R	P	S				
1	1/3	1/3	1/3	P	7/9	5/3		1	1/3	1/3	1/3	S	-7/9	0	
2	1/6	2/3	1/6	P				2	1/6	1/6	2/3	P			
3	1/9	7/9	1/9	P				3	1/9	4/9	4/9	P			

Exploitability

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1								Player 2						
	Average strategy			BR	utility	BR utility	epsilon		Average strategy			BR	utility	BR utility	epsilon
	R	P	S						R	P	S				
1	1/3	1/3	1/3	P	7/9	5/3	8/9	1	1/3	1/3	1/3	S	-7/9	0	7/9
2	1/6	2/3	1/6	P				2	1/6	1/6	2/3	P			
3	1/9	7/9	1/9	P				3	1/9	4/9	4/9	P			

Exploitability

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1								Player 2						
	Average strategy			BR	utility	BR utility	epsilon		Average strategy			BR	utility	BR utility	epsilon
	R	P	S						R	P	S				
1	1/3	1/3	1/3	P	7/9	5/3	8/9	1	1/3	1/3	1/3	S	-7/9	0	7/9
2	1/6	2/3	1/6	P	1.36	1.55	0.19	2	1/6	1/6	2/3	P	-1.36	-0.11	1.25
3	1/9	7/9	1/9	P				3	1/9	4/9	4/9	P			

Exploitability

	R	P	S
R	2 , -2	1 , -1	0 , 0
P	2 , -2	0 , 0	3 , -3
S	-1 , 1	3 , -3	-3 , 3

	Player 1								Player 2						
	Average strategy			BR	utility	BR utility	epsilon		Average strategy			BR	utility	BR utility	epsilon
	R	P	S						R	P	S				
1	1/3	1/3	1/3	P	7/9	5/3	8/9	1	1/3	1/3	1/3	S	-7/9	0	7/9
2	1/6	2/3	1/6	P	1.36	1.55	0.19	2	1/6	1/6	2/3	P	-1.36	-0.11	1.25
3	1/9	7/9	1/9	P	1.12	1.21	0.09	3	1/9	4/9	4/9	P	-1.12	-0.19	0.92

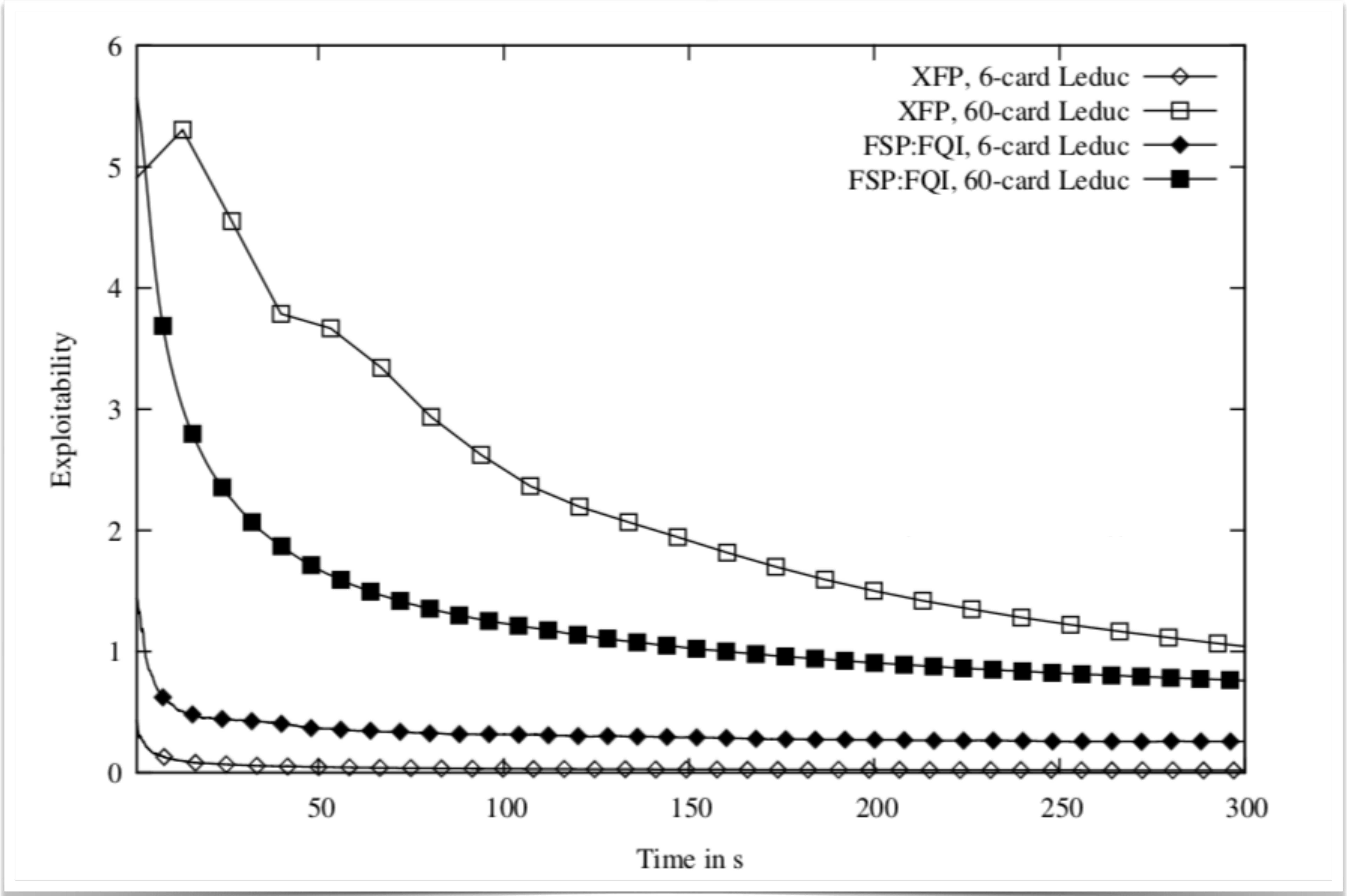
Convergence speed

Given every epsilon > 0, for every t larger than $(1/\text{epsilon})^{2m}$, the average strategies constitute an epsilon-Nash equilibrium, where m is the number of actions of each single player

FP practical motivation

FP may be much slower than linear programming techniques when returning an exact solution, but it may be much faster if we accept an approximate solution

Anytime exploitability

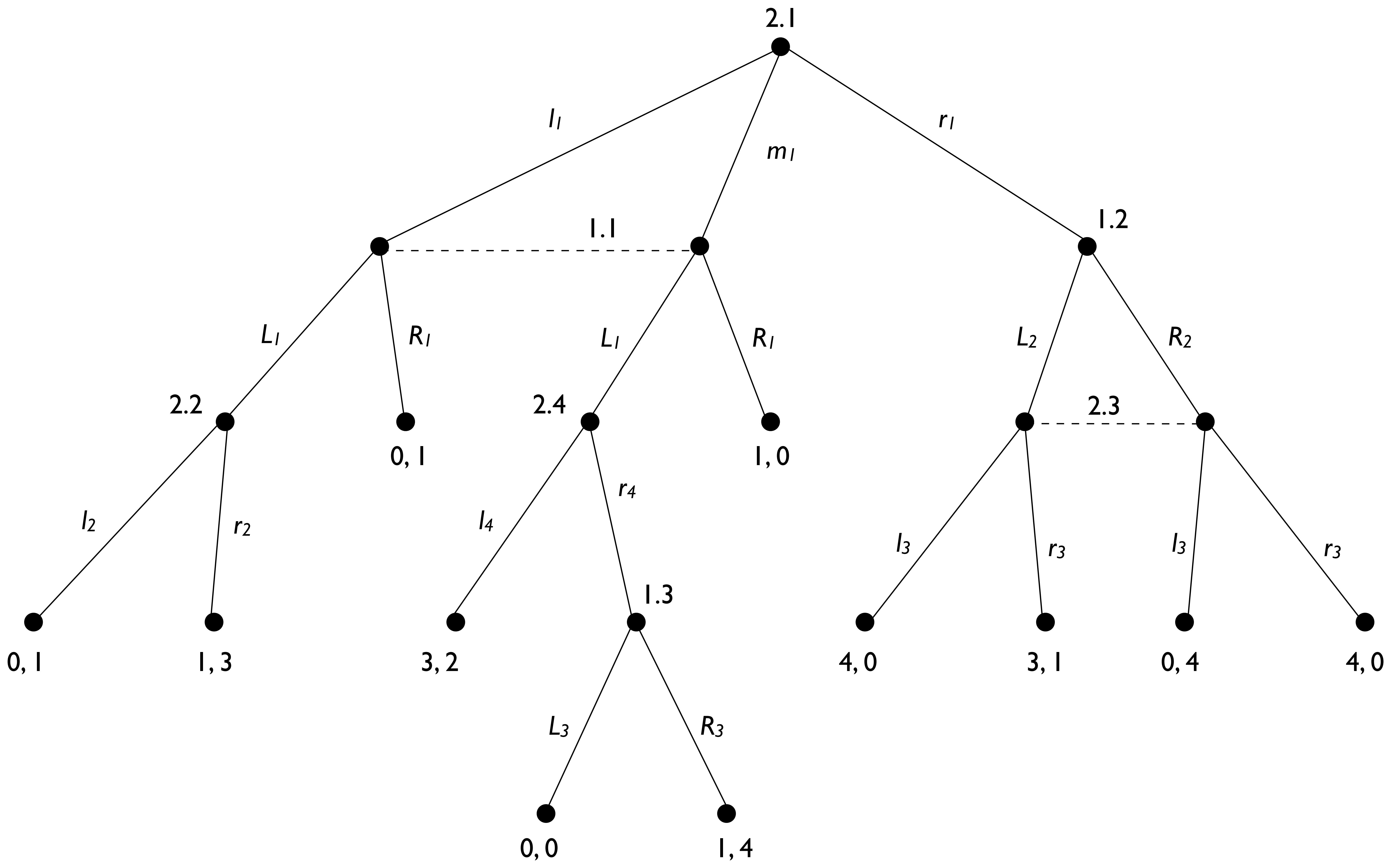


FP and extensive-form games

- The application of FP to the normal form of an extensive form game requires exponential space and time for every iteration of the algorithm

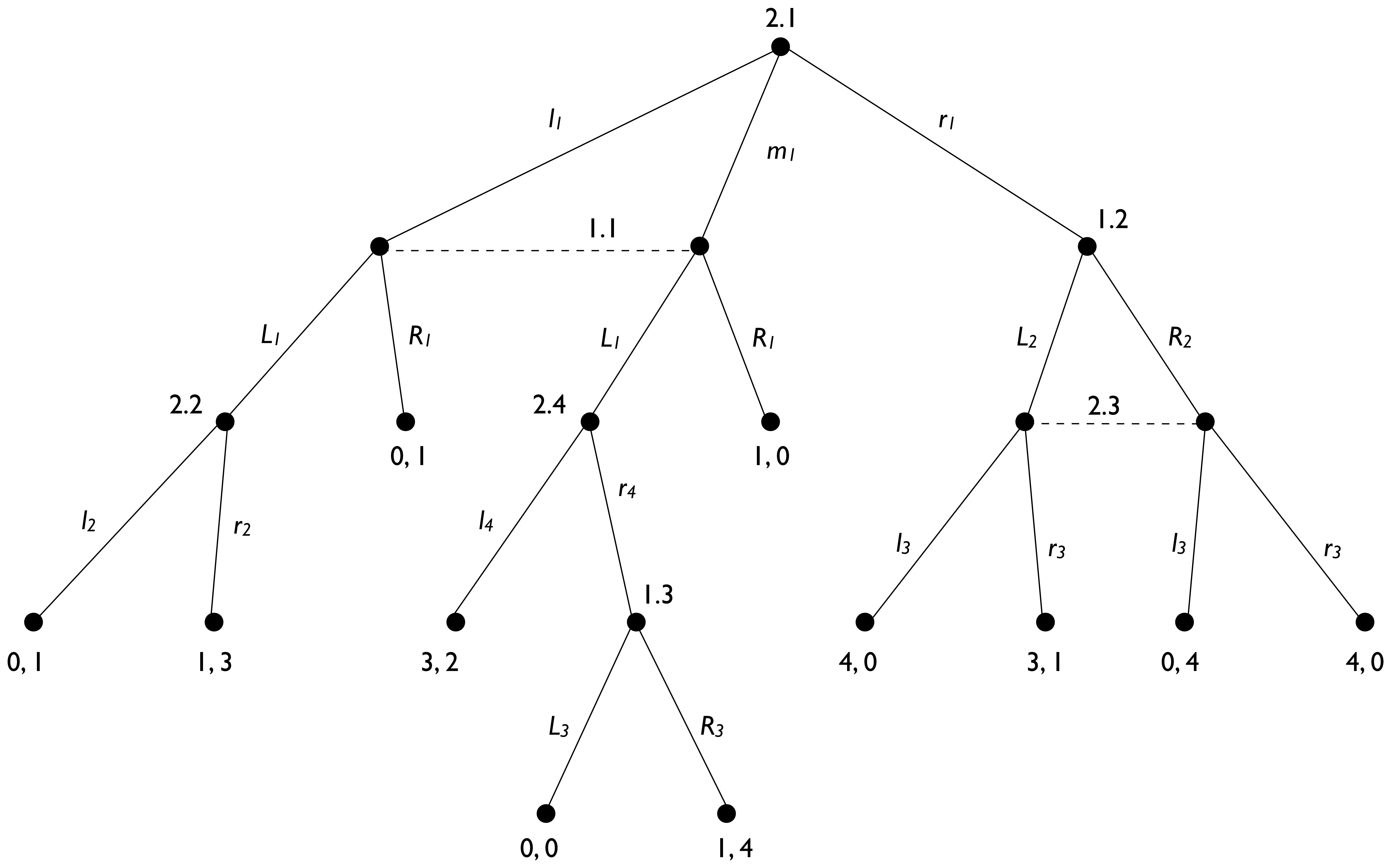
- In practice, a different representation can be used (sequence form) to avoid the exponential explosion of the strategies

Example



2.1
l_1
m_1
r_1
1.1
1.2
L_1
R_1
L_1
R_1
L_2
R_2
2.2
0, 1
2.4
1, 0
2.3
l_2
r_2
l_4
r_4
l_3
r_3
l_3
r_3
1.3
0, 1
1, 3
3, 2
4, 0
3, 1
0, 4
4, 0
L_3
R_3
0, 0
1, 4

Example



	l1 l2 * *	l1 r2 * *	m1 * * l4	m1 * * r4	r1 * l3 *	r1 * r3 *
L1 L2 L3						
L1 R2 L3						
L1 L2 R3						
L1 R2 R3						
R1 L2 *						
R1 R2 *						