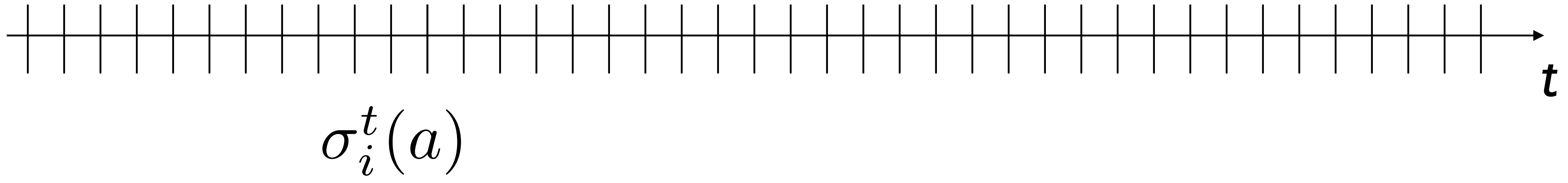
Regret Matching (RM) and Counter Factual Regret (CFR) minimization

POLITECNICO
MILANO 1863

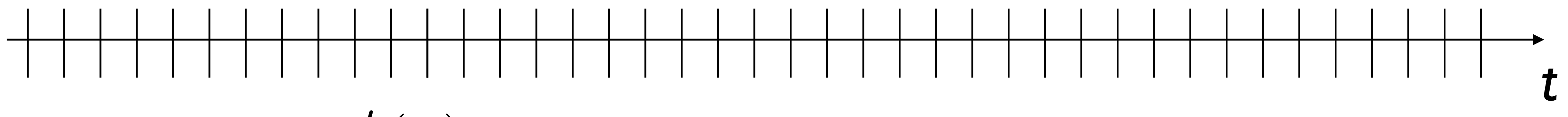
Assumptions

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

Adaptive strategies



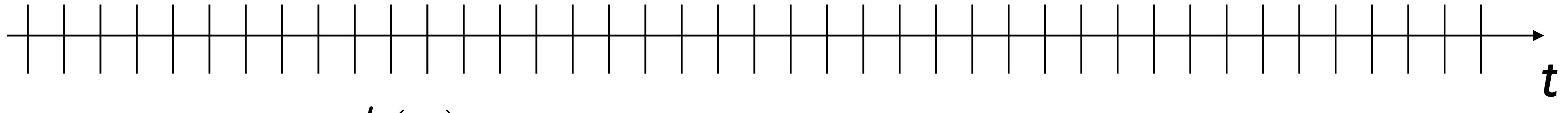
Adaptive strategies



$\sigma_i^t(a)$

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

Adaptive strategies



$$\sigma_i^t(a)$$

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

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

Algorithm

- At every iteration
  - Calculate the instantaneous regrets
  - Calculate the cumulative regrets
  - Calculate the positive cumulative regrets
  - Update the strategies accordingly

Regret

For every action a, the (instantaneous) ***regret*** at time t represents the difference between the expected utility provided by that action and the expected utility of the current strategy

$$r_i^t(a) = \mathbb{E}[U_i(a, \sigma_{-i}^t)] - \mathbb{E}[U_i(\sigma_i^t, \sigma_{-i}^t)]$$

Regret

For every action a, the (instantaneous) ***regret*** at time t represents the difference between the expected utility provided by that action and the expected utility of the current strategy

$$r_i^t(a) = \boxed{\mathbb{E}[U_i(a, \sigma_{-i}^t)]} - \mathbb{E}[U_i(\sigma_i^t, \sigma_{-i}^t)]$$

expected utility provided by action a

Regret

For every action a, the (instantaneous) ***regret*** at time t represents the difference between the expected utility provided by that action and the expected utility of the current strategy

$$r_i^t(a) = \boxed{\mathbb{E}[U_i(a, \sigma_{-i}^t)]} - \boxed{\mathbb{E}[U_i(\sigma_i^t, \sigma_{-i}^t)]}$$

expected utility provided by action a

expected utility provided by the current strategy of player i

Example

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

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

Example

$$r_1^1(\mathrm{R}) = \mathbb{E}[U_1(\mathrm{R}, \sigma_2^1)] - \mathbb{E}[U_1(\sigma_1^1, \sigma_2^1)] = 0$$

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

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

Example

$$r_1^1(\mathrm{R}) = \boxed{\mathbb{E}[U_1(\mathrm{R}, \sigma_2^1)]} - \mathbb{E}[U_1(\sigma_1^1, \sigma_2^1)] = 0$$

$$\left(0 \cdot \tfrac{1}{3} - 1 \cdot \tfrac{1}{3} + 1 \cdot \tfrac{1}{3}\right)$$

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

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

Example

$$\underbrace{r_1^1(\mathrm{R})}_{} = \boxed{\mathbb{E}[U_1(\mathrm{R}, \sigma_2^1)]} - \boxed{\mathbb{E}[U_1(\sigma_1^1, \sigma_2^1)]} = 0$$

$\left(0 \cdot \frac{1}{3} - 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}\right)$

0

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

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

Example

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

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

$$r_1^1(\mathrm{R}) = \underbrace{\boxed{\mathbb{E}[U_1(\mathrm{R}, \sigma_2^1)]}}_{\left(0 \cdot \frac{1}{3} - 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}\right)} - \underbrace{\boxed{\mathbb{E}[U_1(\sigma_1^1, \sigma_2^1)]}}_{0} = 0$$

$$r_1^1(\mathrm{P}) = \underbrace{\boxed{\mathbb{E}[U_1(\mathrm{P}, \sigma_2^1)]}}_{\left(1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} - 2 \cdot \frac{1}{3}\right)} - \underbrace{\boxed{\mathbb{E}[U_1(\sigma_1^1, \sigma_2^1)]}}_{0} = -\frac{1}{3}$$

Example

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

$$r_1^1(\mathrm{R}) = \underbrace{\boxed{\mathbb{E}[U_1(\mathrm{R}, \sigma_2^1)]}}_{(0 \cdot \frac{1}{3} - 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3})} - \underbrace{\boxed{\mathbb{E}[U_1(\sigma_1^1, \sigma_2^1)]}}_{0} = 0$$

$$r_1^1(\mathrm{P}) = \underbrace{\boxed{\mathbb{E}[U_1(\mathrm{P}, \sigma_2^1)]}}_{(1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} - 2 \cdot \frac{1}{3})} - \underbrace{\boxed{\mathbb{E}[U_1(\sigma_1^1, \sigma_2^1)]}}_{0} = -\frac{1}{3}$$

$$r_1^1(\mathrm{S}) = \underbrace{\boxed{\mathbb{E}[U_1(\mathrm{S}, \sigma_2^1)]}}_{(-2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3})} - \underbrace{\boxed{\mathbb{E}[U_1(\sigma_1^1, \sigma_2^1)]}}_{0} = \frac{1}{3}$$

Example

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

$$r_2^1(\mathrm{R}) = \underbrace{\boxed{\mathbb{E}[U_2(\mathrm{R}, \sigma_1^1)]}}_{\left(0 \cdot \frac{1}{3} - 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}\right)} - \underbrace{\boxed{\mathbb{E}[U_2(\sigma_2^1, \sigma_1^1)]}}_{0} = -\frac{1}{3}$$

$$r_2^1(\mathrm{P}) = \underbrace{\boxed{\mathbb{E}[U_2(\mathrm{P}, \sigma_1^1)]}}_{\left(1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} - 2 \cdot \frac{1}{3}\right)} - \underbrace{\boxed{\mathbb{E}[U_2(\sigma_2^1, \sigma_1^1)]}}_{0} = \frac{2}{3}$$

$$r_2^1(\mathrm{S}) = \underbrace{\boxed{\mathbb{E}[U_2(\mathrm{S}, \sigma_1^1)]}}_{\left(-2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}\right)} - \underbrace{\boxed{\mathbb{E}[U_2(\sigma_2^1, \sigma_1^1)]}}_{0} = -\frac{1}{3}$$

Cumulative regret

For every action a, the ***cumulative regret*** at time t represents the sum, for every time from 1 to t, of the difference between the expected utility provided by that action and the expected utility of the current strategy

$$R_i^t(a) = \sum_{\tau=1}^{t} r_i^\tau(a)$$

Example

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

$$R_1^1(\mathrm{R}) = \qquad r_1^1(\mathrm{R}) = \qquad 0$$

$$R_1^1(\mathrm{P}) = \qquad r_1^1(\mathrm{P}) = \qquad -\frac{1}{3}$$

$$R_1^1(\mathrm{S}) = \qquad r_1^1(\mathrm{S}) = \qquad \frac{1}{3}$$

$$R_2^1(\mathrm{R}) = \qquad r_2^1(\mathrm{R}) = \qquad -\frac{1}{3}$$

$$R_2^1(\mathrm{P}) = \qquad r_2^1(\mathrm{P}) = \qquad \frac{2}{3}$$

$$R_2^1(\mathrm{S}) = \qquad r_2^1(\mathrm{S}) = \qquad -\frac{1}{3}$$

Postive cumulative regret

We only take positive cumulative regrets

$$R_i^{t,+}(a) = \max\left\{R_i^t(a), 0\right\}$$

Example

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

$$R_1^{1,+}(\mathrm{R}) = \qquad R_1^1(\mathrm{R}) = \qquad 0$$

$$R_1^{1,+}(\mathrm{P}) = \qquad 0$$

$$R_1^{1,+}(\mathrm{S}) = \qquad R_1^1(\mathrm{S}) = \qquad \frac{1}{3}$$

$$R_2^{1,+}(\mathrm{R}) = \qquad 0$$

$$R_2^{1,+}(\mathrm{P}) = \qquad R_2^1(\mathrm{P}) = \qquad \frac{2}{3}$$

$$R_2^{1,+}(\mathrm{S}) = \qquad 0$$

Update rule: Regret Matching (RM)

For every action a, the new strategy is given by the ratio between the positive cumulative regret of that strategy and the sum of the positive cumulative regrets of all the actions of the player

$$\sigma_i^{t+1}(a) = \begin{cases} \dfrac{R_i^{t,+}(a)}{\sum_{a'} R_i^{t,+}(a')} & \text{if } \sum_{a'} R_i^{t,+}(a') > 0 \\ \dfrac{1}{|A_i|} & \text{if } \sum_{a'} R_i^{t,+}(a') = 0 \end{cases}$$

Example

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

$$\sigma_1^2(\mathrm{R}) = \frac{0}{\frac{1}{3}} = 0$$

$$\sigma_1^2(\mathrm{P}) = \frac{0}{\frac{1}{3}} = 0$$

$$\sigma_1^2(\mathrm{S}) = \frac{\frac{1}{3}}{\frac{1}{3}} = 1$$

$$\sigma_2^2(\mathrm{R}) = \frac{0}{\frac{2}{3}} = 0$$

$$\sigma_2^2(\mathrm{P}) = \frac{\frac{2}{3}}{\frac{2}{3}} = 1$$

$$\sigma_2^2(\mathrm{S}) = \frac{0}{\frac{2}{3}} = 0$$

Example

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

$$r_1^2(\mathrm{R}) = \underbrace{\boxed{\mathbb{E}[U_1(\mathrm{R}, \sigma_2^2)]}}_{(0\cdot 0 - 1\cdot 1 + 1\cdot 0)} - \underbrace{\boxed{\mathbb{E}[U_1(\sigma_1^2, \sigma_2^2)]}}_{3} = -4$$

$$r_1^2(\mathrm{P}) = \underbrace{\boxed{\mathbb{E}[U_1(\mathrm{P}, \sigma_2^2)]}}_{(1\cdot 0 + 0\cdot 1 - 2\cdot 0)} - \underbrace{\boxed{\mathbb{E}[U_1(\sigma_1^2, \sigma_2^2)]}}_{3} = -3$$

$$r_1^2(\mathrm{S}) = \underbrace{\boxed{\mathbb{E}[U_1(\mathrm{S}, \sigma_2^2)]}}_{(-2\cdot 0 + 3\cdot 1 + 0\cdot 0)} - \underbrace{\boxed{\mathbb{E}[U_1(\sigma_1^2, \sigma_2^2)]}}_{3} = 0$$

Example

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

$$r_2^2(\mathrm{R}) = \underbrace{\boxed{\mathbb{E}[U_2(\mathrm{R}, \sigma_1^2)]}}_{(0 \cdot 0 - 2 \cdot 0 + 1 \cdot 1)} - \underbrace{\boxed{\mathbb{E}[U_2(\sigma_2^2, \sigma_1^2)]}}_{-1} = 2$$

$$r_2^2(\mathrm{P}) = \underbrace{\boxed{\mathbb{E}[U_2(\mathrm{P}, \sigma_1^2)]}}_{(3 \cdot 0 + 0 \cdot 0 - 1 \cdot 1)} - \underbrace{\boxed{\mathbb{E}[U_2(\sigma_2^2, \sigma_1^2)]}}_{-1} = 0$$

$$r_2^2(\mathrm{S}) = \underbrace{\boxed{\mathbb{E}[U_2(\mathrm{S}, \sigma_1^2)]}}_{(-2 \cdot 0 + 1 \cdot 0 + 0 \cdot 1)} - \underbrace{\boxed{\mathbb{E}[U_2(\sigma_2^2, \sigma_1^2)]}}_{-1} = 1$$

Example

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

$$R_1^2(\mathrm{R}) = r_1^1(\mathrm{R}) + r_1^2(\mathrm{R}) = -4$$

$$R_1^2(\mathrm{P}) = r_1^1(\mathrm{P}) + r_1^2(\mathrm{P}) = -\frac{10}{3}$$

$$R_1^2(\mathrm{S}) = r_1^1(\mathrm{S}) + r_1^2(\mathrm{S}) = \frac{1}{3}$$

$$R_2^2(\mathrm{R}) = r_2^1(\mathrm{R}) + r_2^2(\mathrm{R}) = \frac{5}{3}$$

$$R_2^2(\mathrm{P}) = r_2^1(\mathrm{P}) + r_2^2(\mathrm{P}) = \frac{2}{3}$$

$$R_2^2(\mathrm{S}) = r_2^1(\mathrm{S}) + r_2^2(\mathrm{S}) = \frac{2}{3}$$

Example

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

$$R_1^{2,+}(\mathrm{R}) = 0$$
$$R_1^{2,+}(\mathrm{P}) = 0$$
$$R_1^{2,+}(\mathrm{S}) = \frac{1}{3}$$

$$R_2^{2,+}(\mathrm{R}) = \frac{5}{3}$$
$$R_2^{2,+}(\mathrm{P}) = \frac{2}{3}$$
$$R_2^{2,+}(\mathrm{S}) = \frac{2}{3}$$

Example

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

$$\sigma_1^3(\mathrm{R}) = \frac{0}{\frac{1}{3}} = 0$$

$$\sigma_1^3(\mathrm{P}) = \frac{0}{\frac{1}{3}} = 0$$

$$\sigma_1^3(\mathrm{S}) = \frac{\frac{1}{3}}{\frac{1}{3}} = 1$$

$$\sigma_2^3(\mathrm{R}) = \frac{\frac{5}{3}}{\frac{9}{3}} = \frac{5}{9}$$

$$\sigma_2^3(\mathrm{P}) = \frac{\frac{2}{3}}{\frac{9}{3}} = \frac{2}{9}$$

$$\sigma_2^3(\mathrm{S}) = \frac{\frac{2}{3}}{\frac{9}{3}} = \frac{2}{9}$$

Convergence

- As t increases, the average strategy from 1 to t returned by the Regret Matching algorithm converges to a Nash equilibrium in 2-player zero-sum games (against a fixed-strategy player it leads to a best response)

- The cumulative regret decreases as

$$\frac{R_i^t(a)}{t} \leq \frac{m}{\sqrt{t}} \Delta_{\max}$$

- The epsilon of the epsilon-Nash equilibrium decreases as

$$\epsilon \leq \frac{2m}{\sqrt{t}} \Delta_{\max}$$

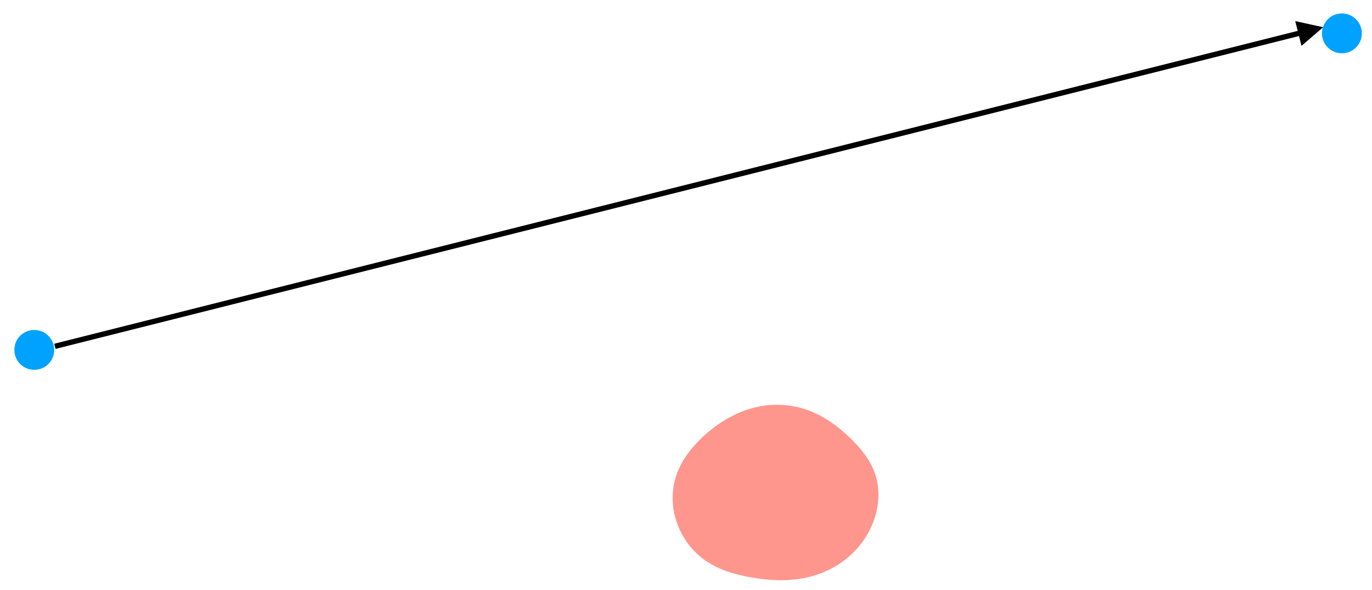
Outside the 2-player zero-sum realm

- The algorithm may not converge to a Nash equilibrium
- The algorithm is just guaranteed not to play dominated strategies
- Procedures based on RM converges to other form of equilibria (correlated equilibria and coarse correlated equilibria)

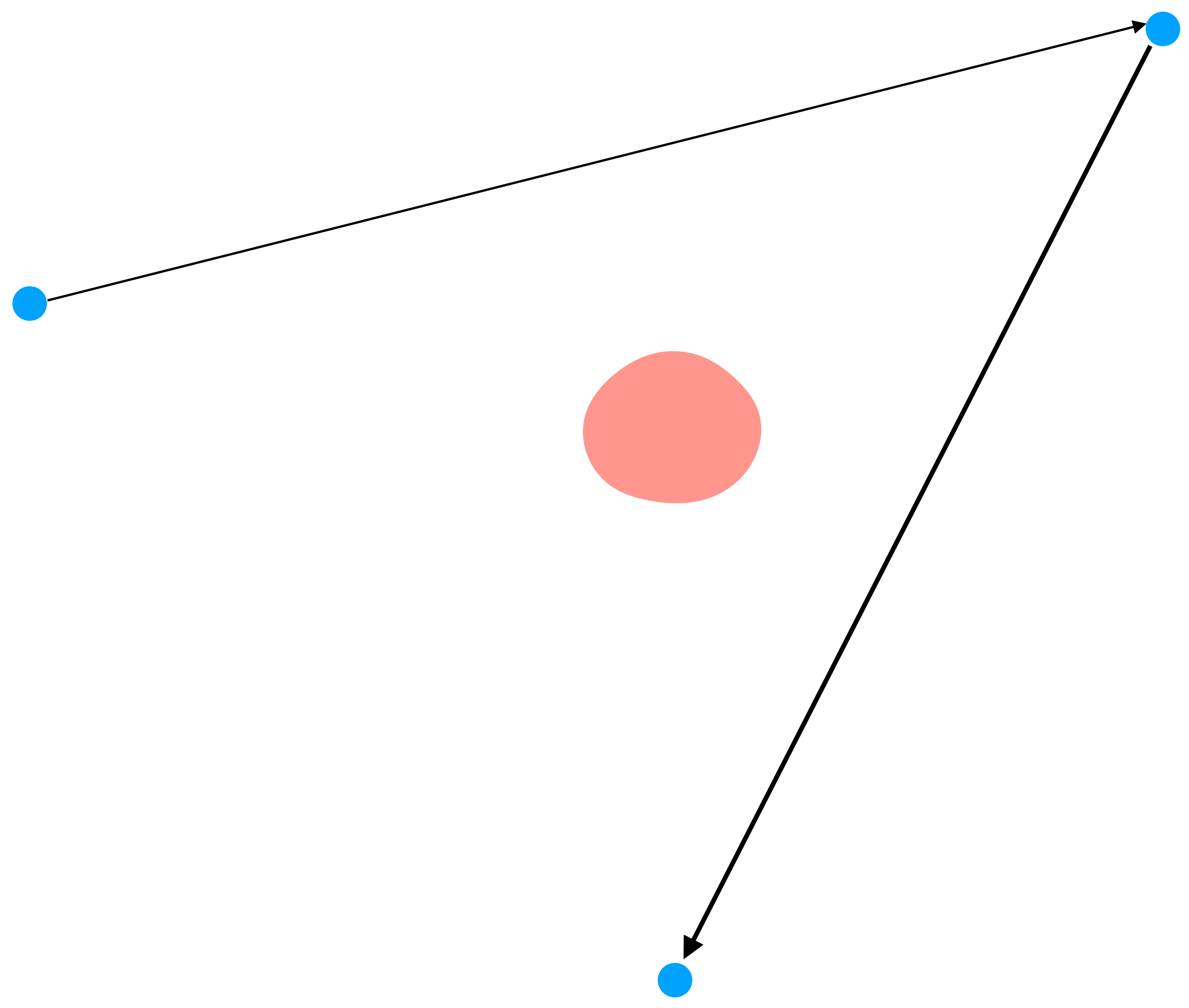
Approachability

Approachability

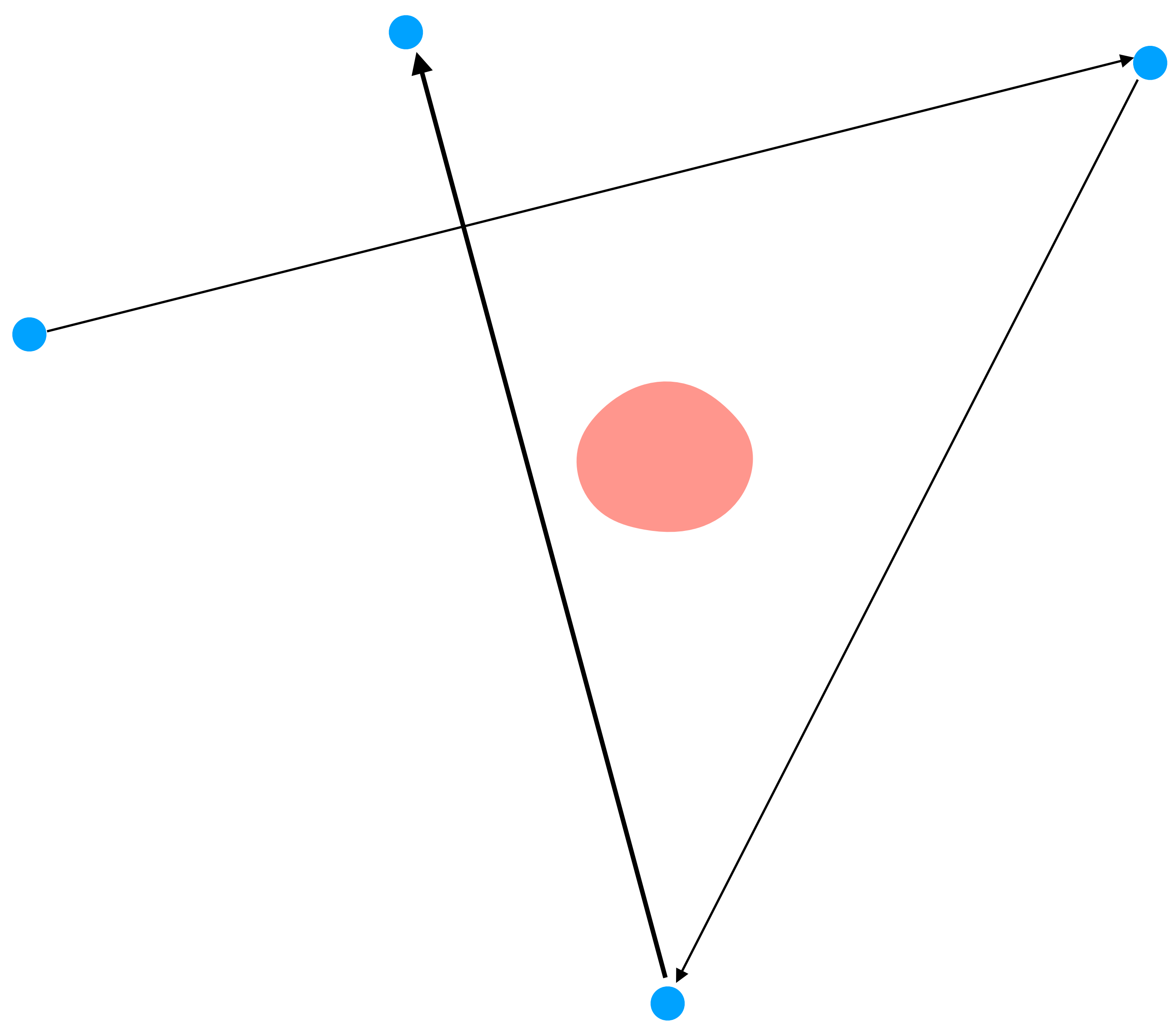
Approachability



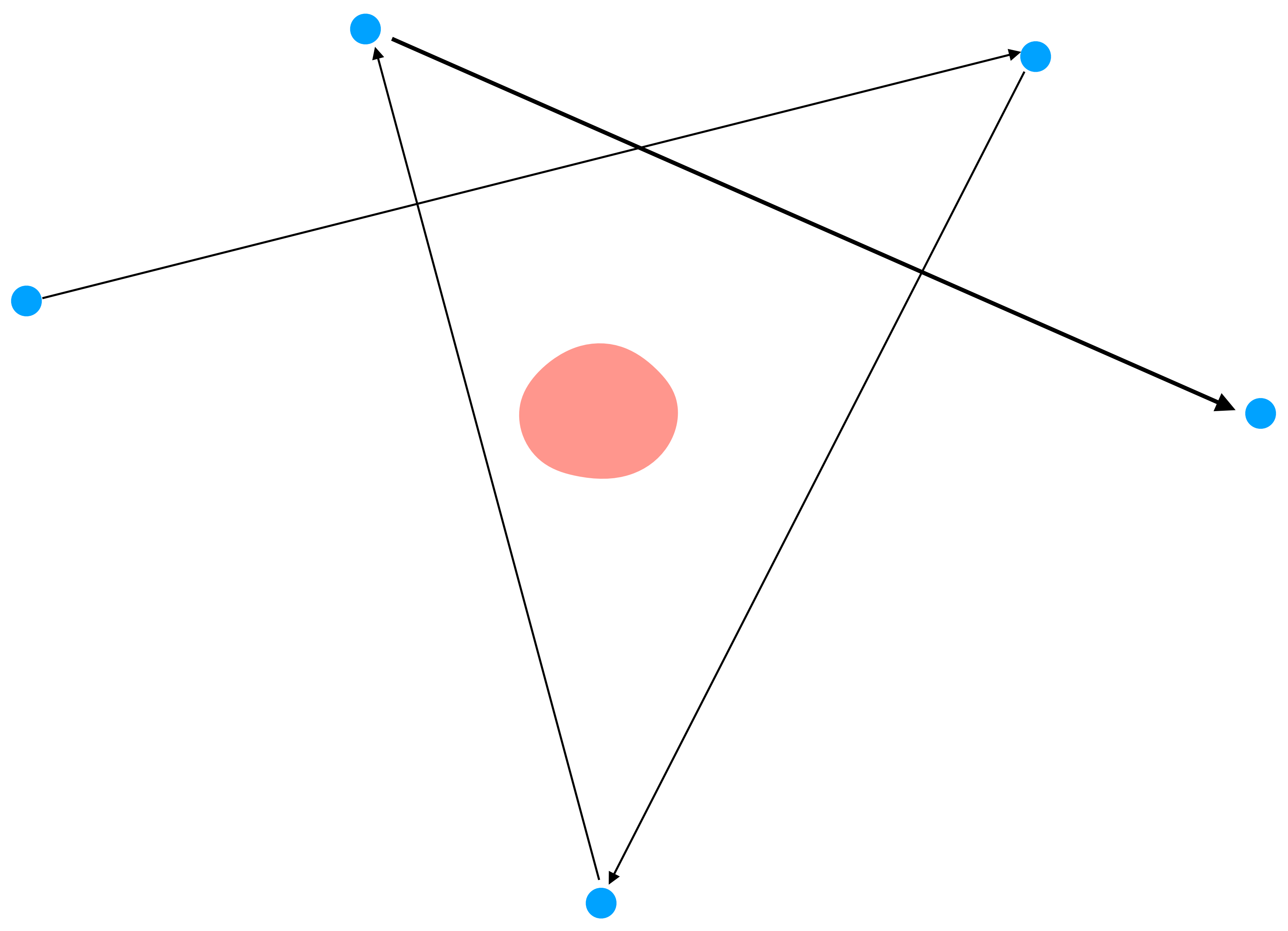
Approachability



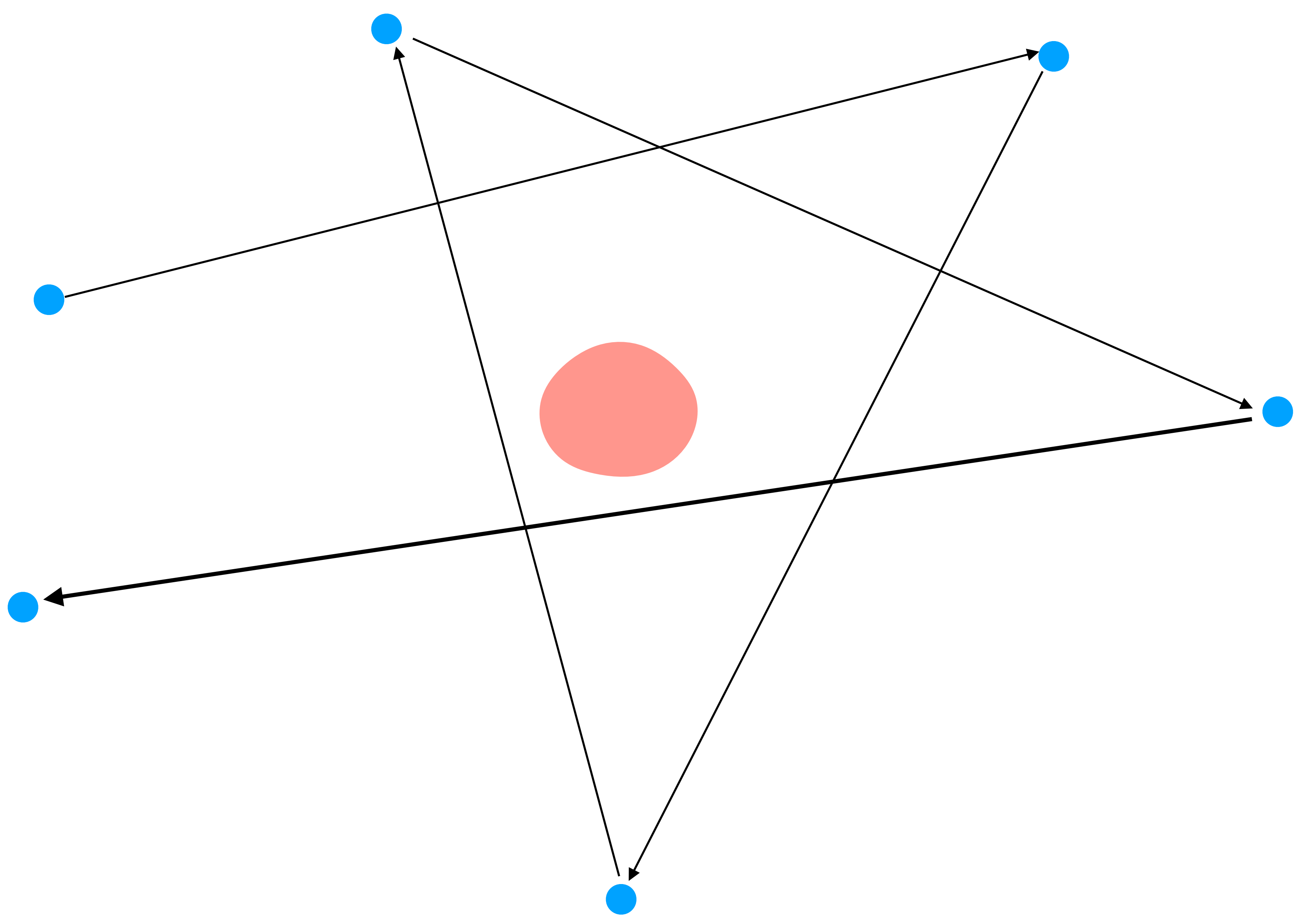
Approachability



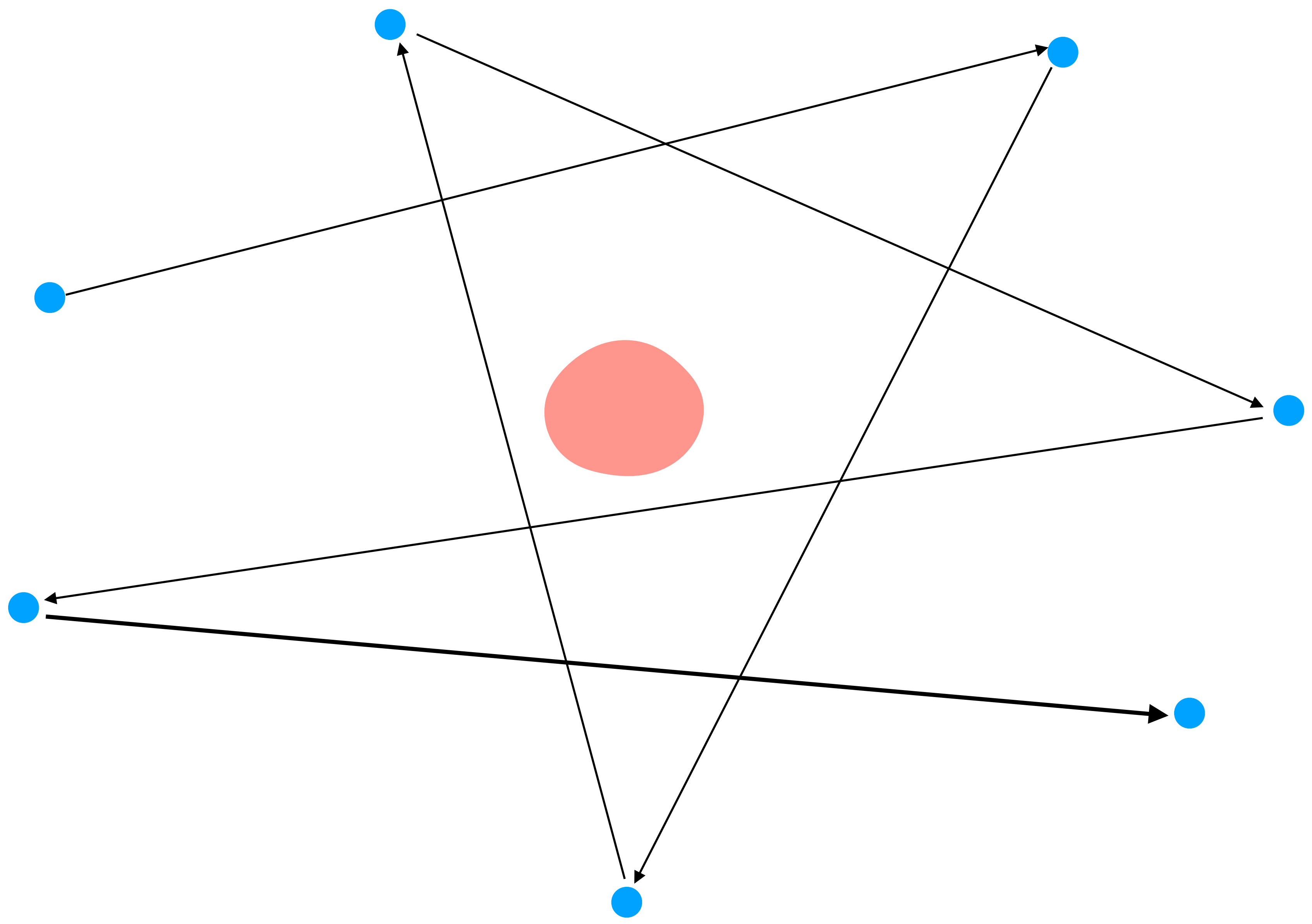
Approachability



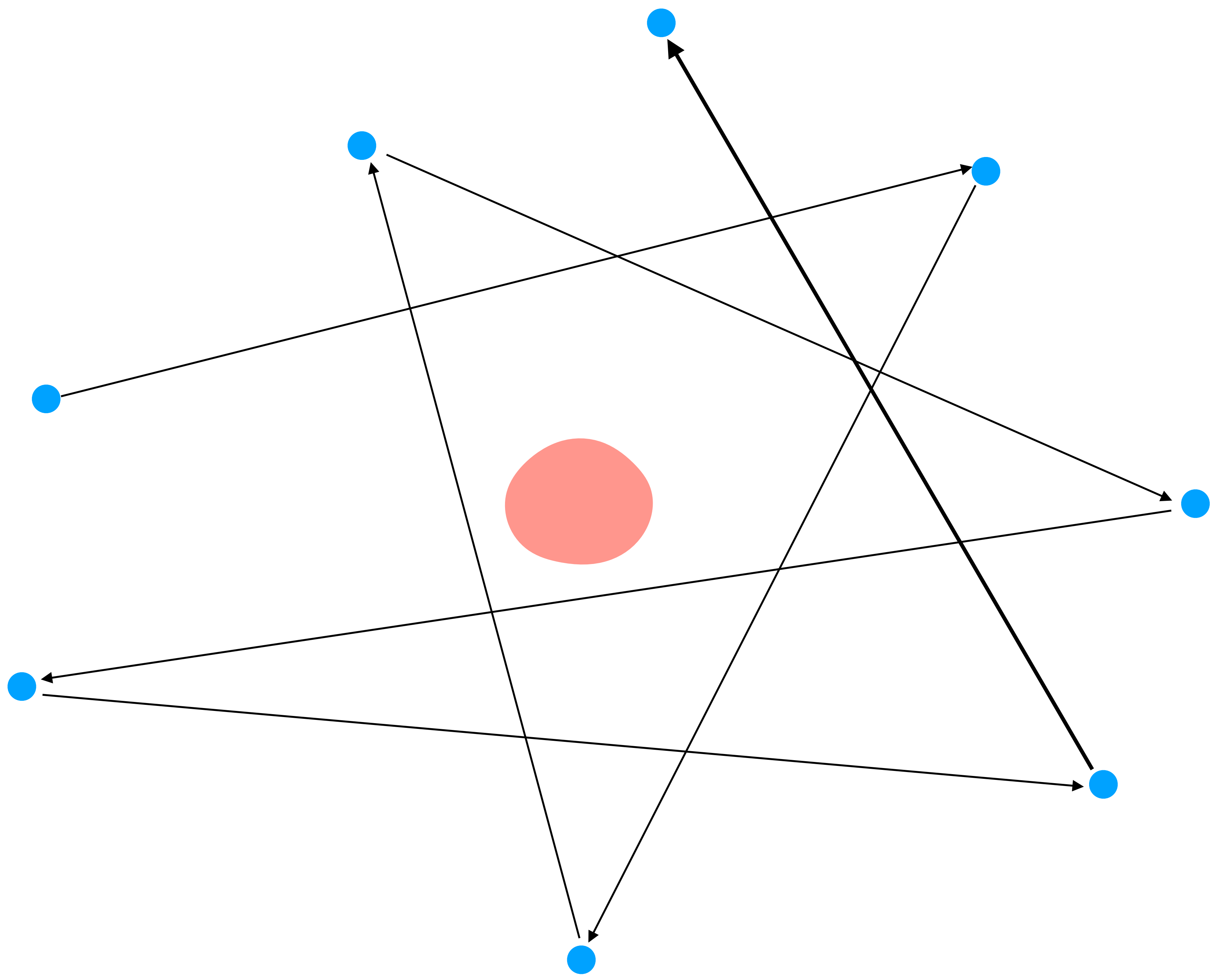
Approachability



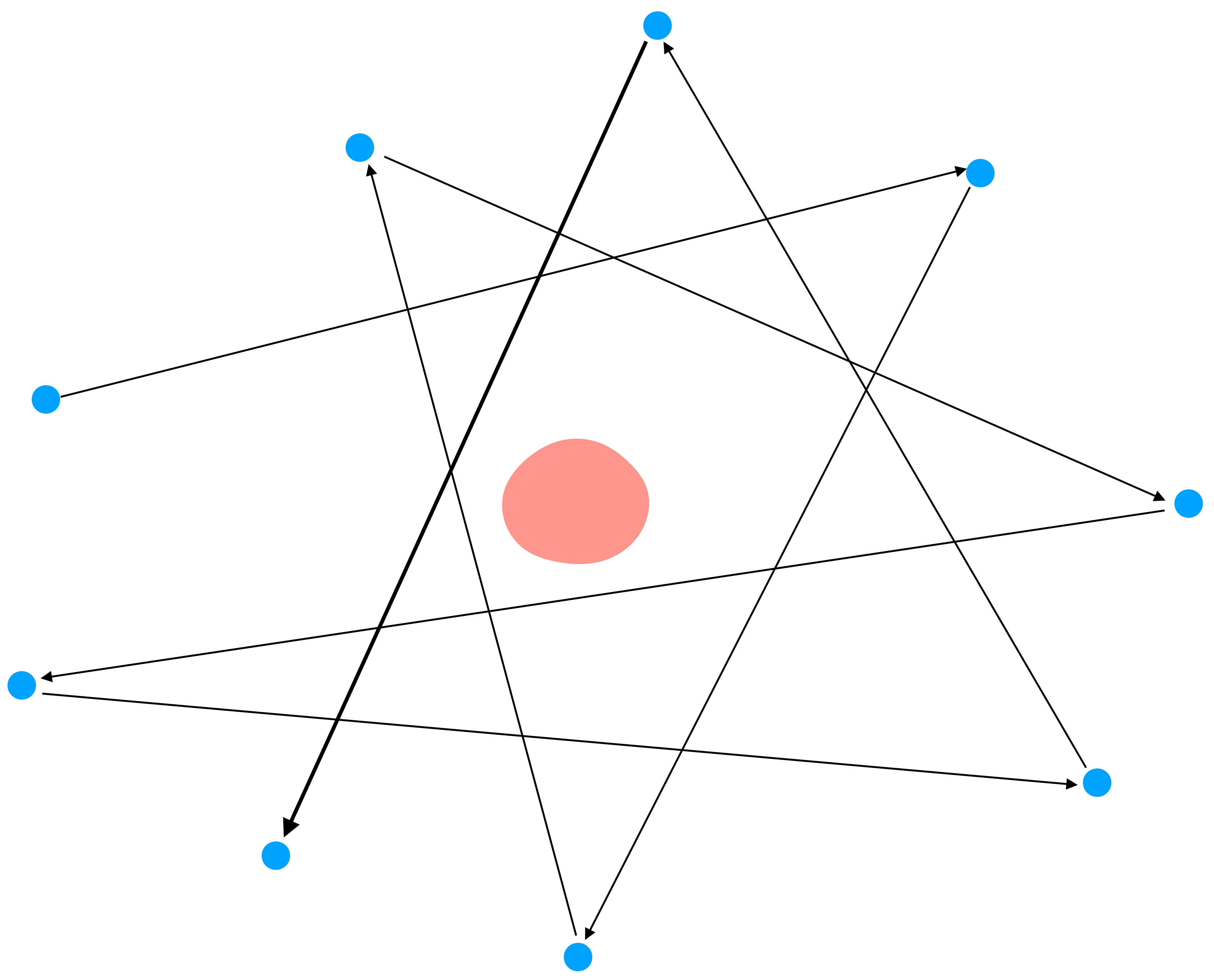
Approachability



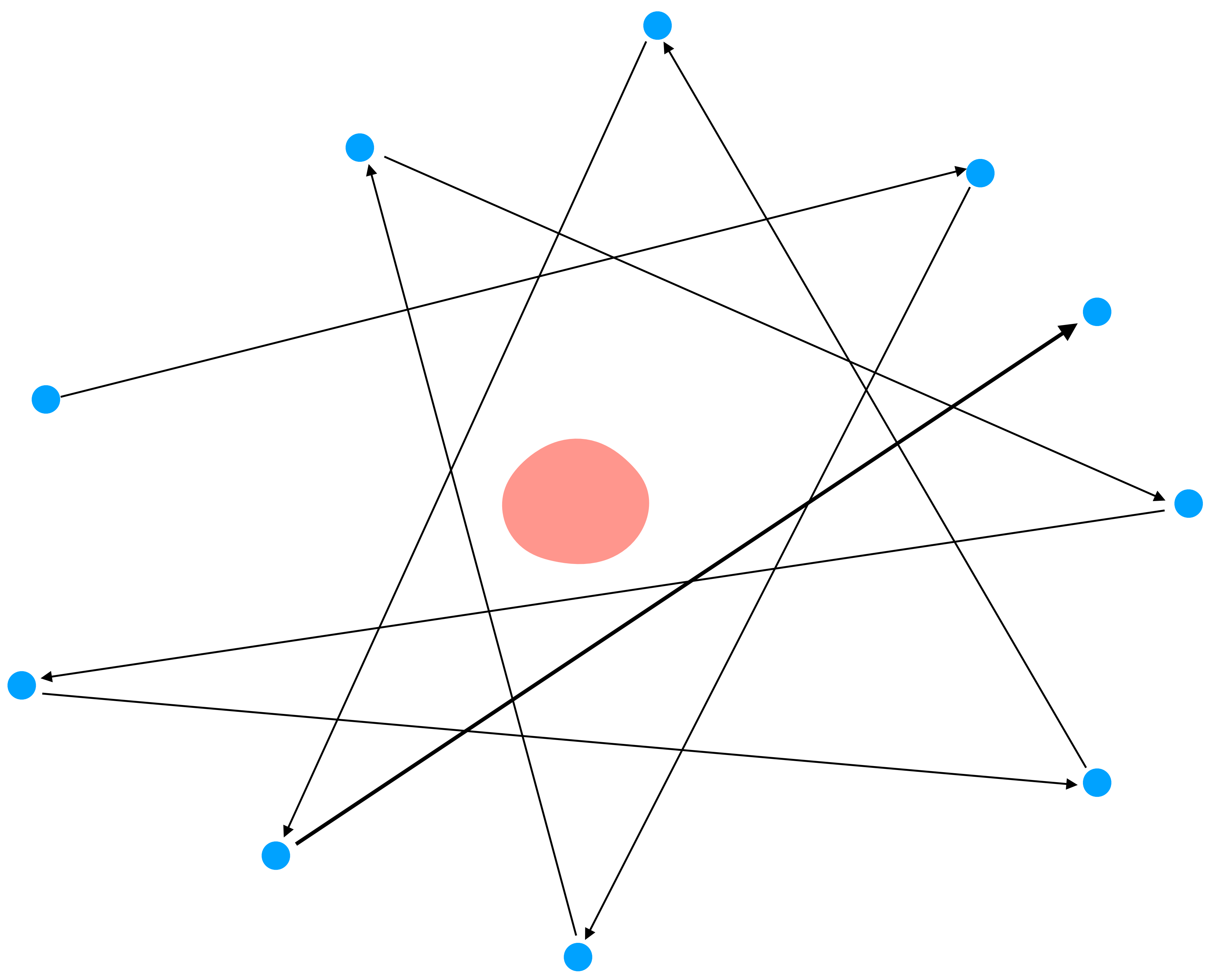
Approachability



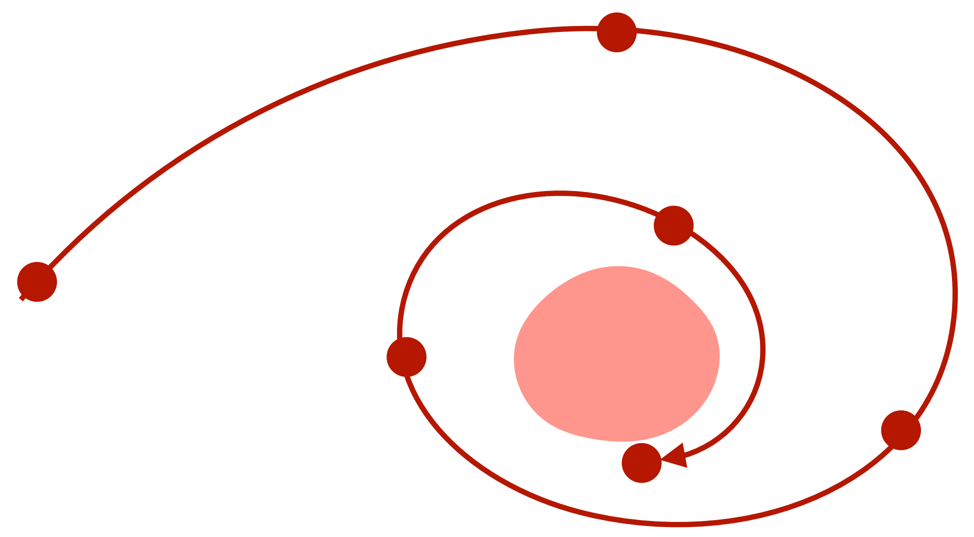
Approachability



Approachability



Approachability



Counter Factual Regret minimization (CFR)

The idea of CFR is the application of RM at every single information set

Regret at each infoset

For every infoset I and action a, the (instantaneous) ***regret*** at time t represents the difference between the expected utility provided by action a and that of the current strategy, **counterfactual on infoset *I* being reached**

$$r_i^t(I, a) = \mathbb{E}[U_{i,I}(a, \sigma_{-i}^t)] - \mathbb{E}[U_{i,I}(\sigma_i^t, \sigma_{-i}^t)]$$

Regret at each infoset

For every infoset I and action *a*, the (instantaneous) ***regret*** at time *t* represents the difference between the expected utility provided by action *a* and that of the current strategy, **counterfactual on infoset *I* being reached**

$$r_i^t(I, a) = \boxed{\mathbb{E}[U_{i,I}(a, \sigma_{-i}^t)]} - \boxed{\mathbb{E}[U_{i,I}(\sigma_i^t, \sigma_{-i}^t)]}$$

expected utility provided by action *a*

expected utility provided by the current strategy of player *i*

Regret at each infoset

For every infoset I and action *a*, the (instantaneous) ***regret*** at time t represents the difference between the expected utility provided by action *a* and that of the current strategy, **counterfactual on infoset *I* being reached**

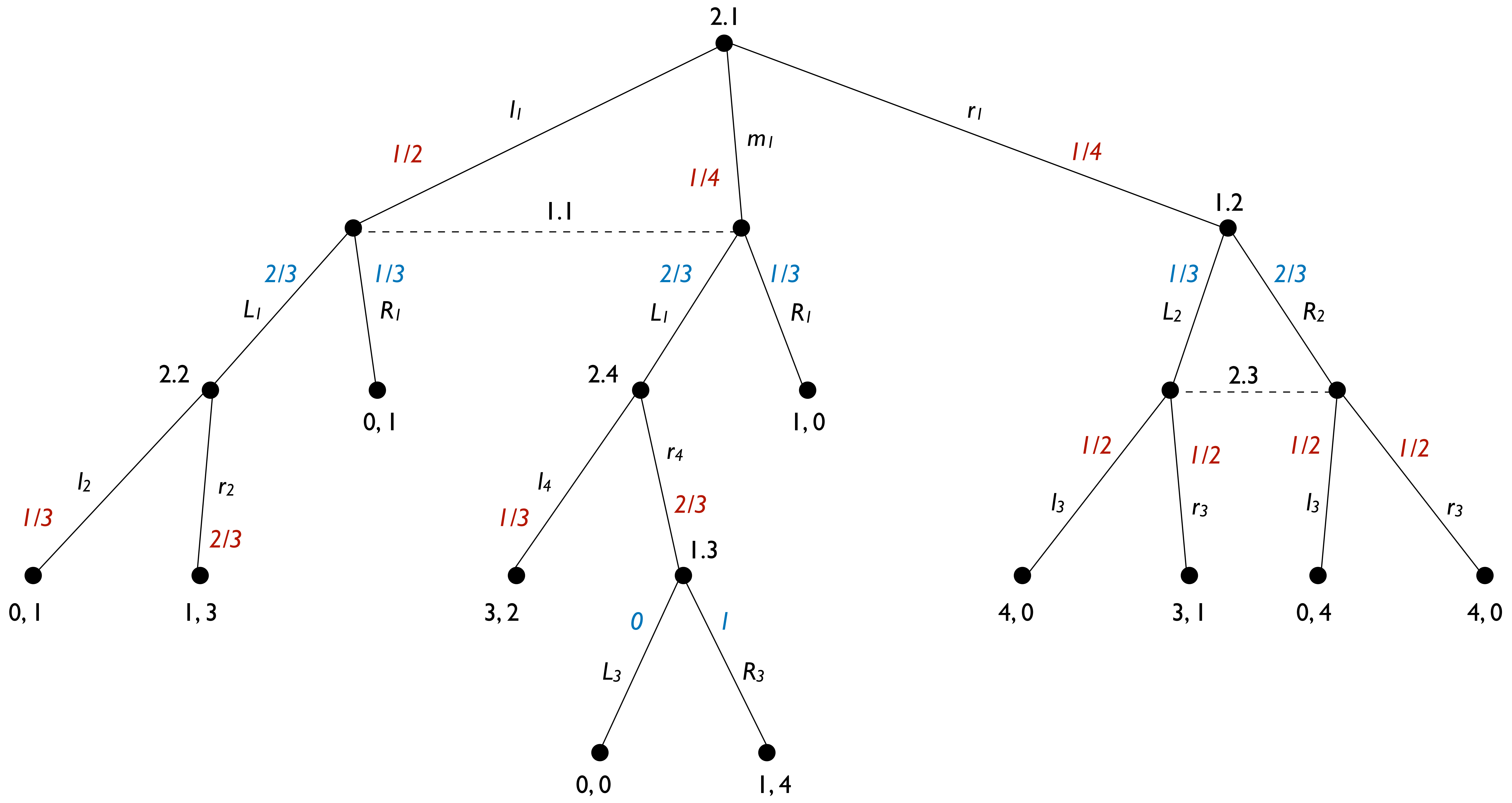
$$r_i^t(I, a) = \mathbb{E}[U_{i,I}(a, \sigma_{-i}^t)] - \mathbb{E}[U_{i,I}(\sigma_i^t, \sigma_{-i}^t)]$$

expected utility provided by action *a*

counterfactual utility at infoset I

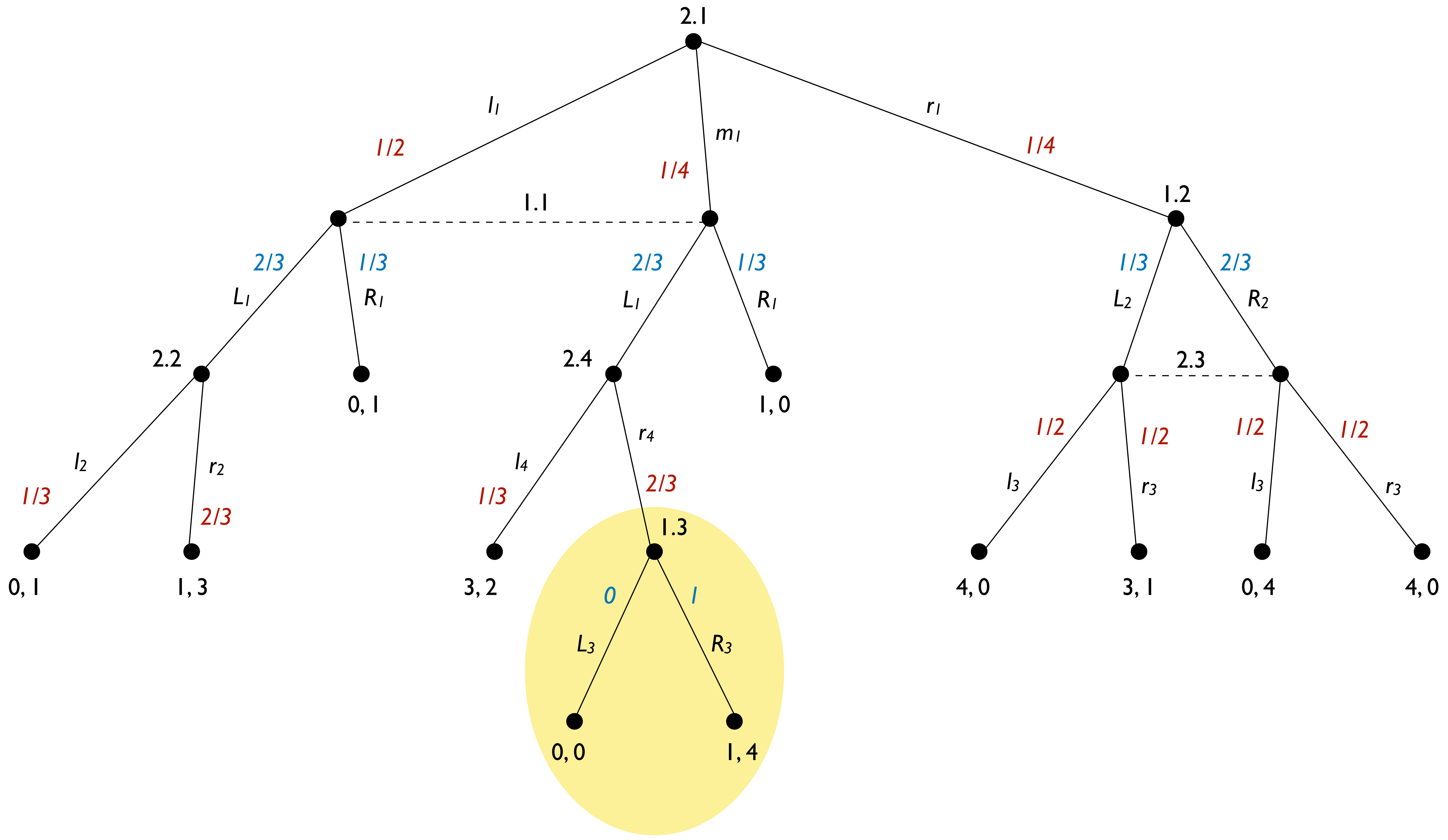
expected utility provided by the current strategy of player *i*

Example



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

Example

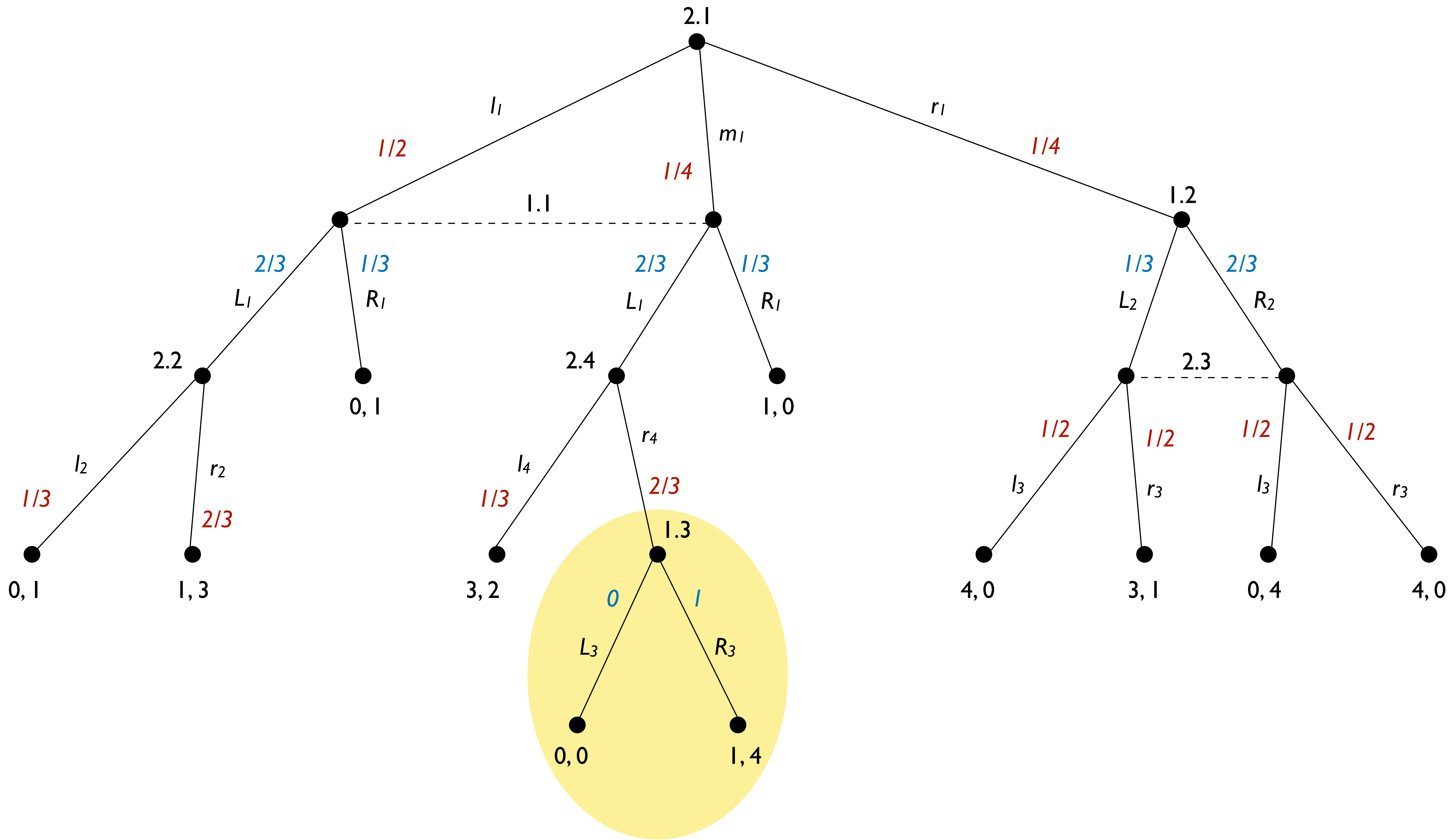


2.1
l_1
1/2
m_1
1/4
r_1
1/4
1.1
1.2
2/3
L_1
1/3
R_1
2/3
L_1
1/3
R_1
1/3
L_2
2/3
R_2
2.2
0, 1
2.4
1, 0
2.3
l_2
1/3
r_2
2/3
l_4
1/3
r_4
2/3
1/2
l_3
1/2
r_3
1/2
l_3
1/2
r_3
0, 1
1, 3
3, 2
1.3
4, 0
3, 1
0, 4
4, 0
0
L_3
1
R_3
0, 0
1, 4

Example

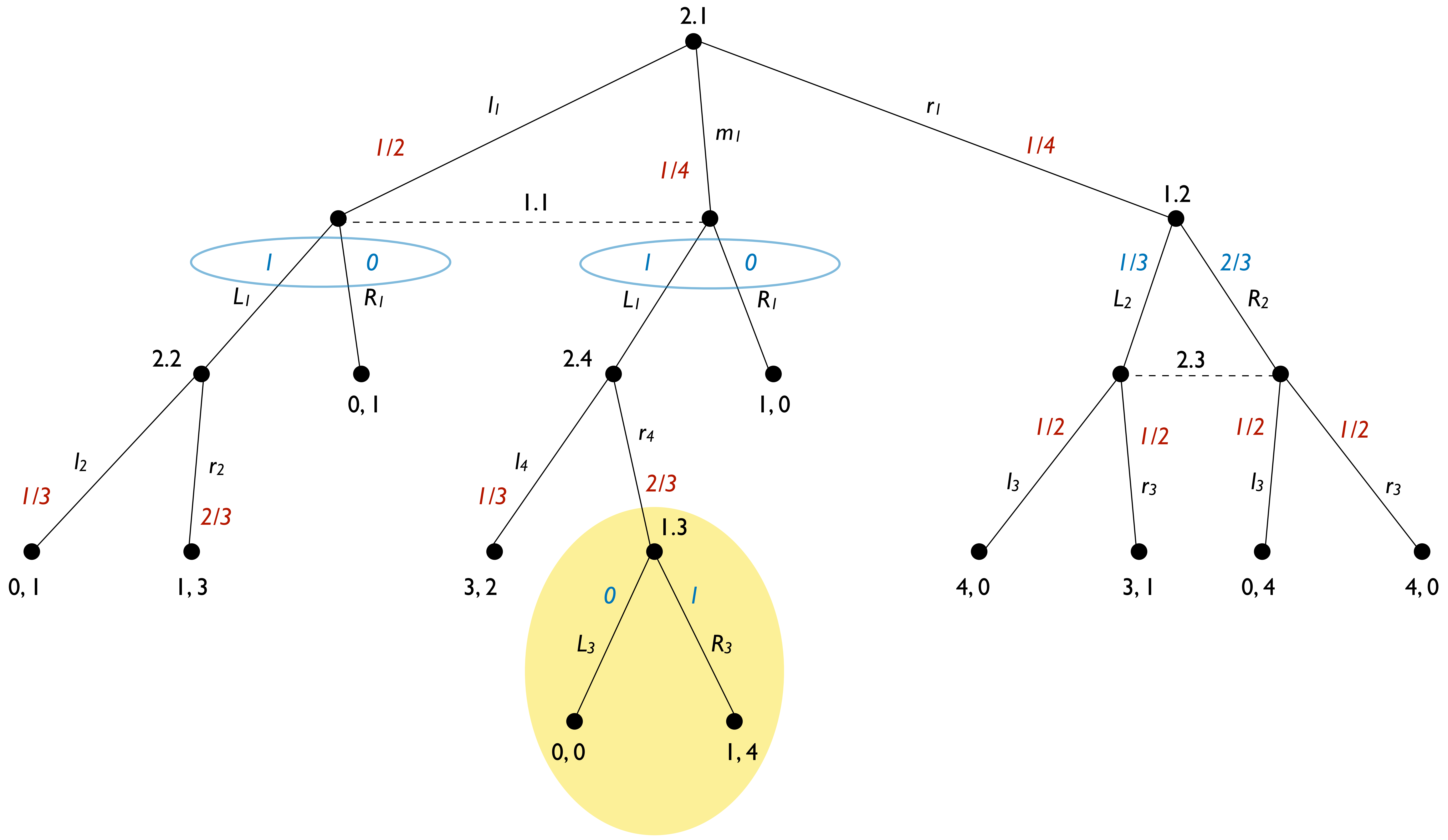
The strategy of player 1 before 1.3 is forced to reach 1.3

Example



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

Example

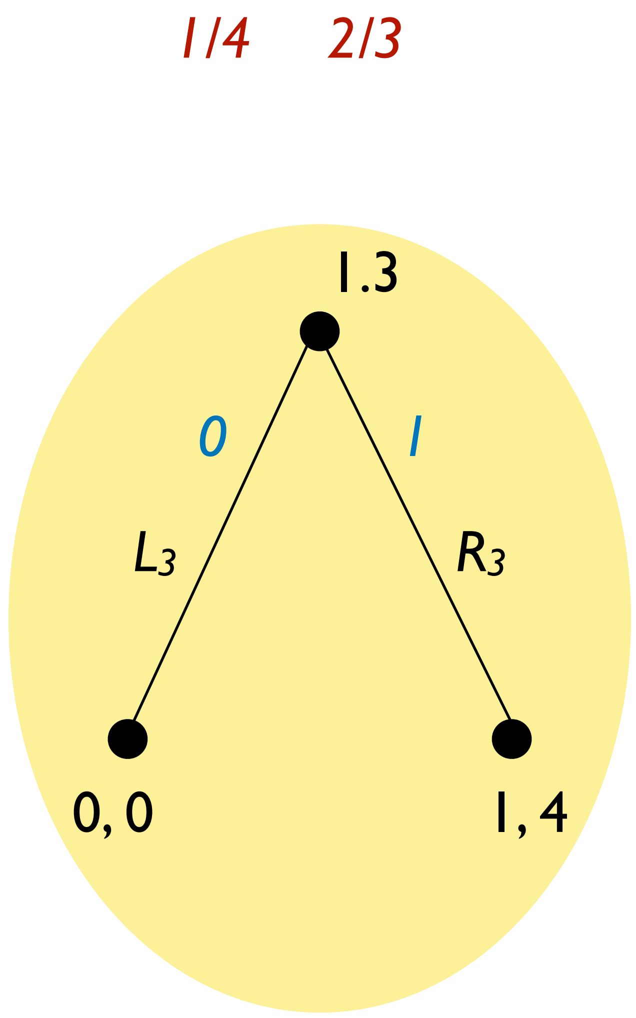


2.1
l_1
1/2
m_1
1/4
r_1
1/4
1.1
1.2
1
0
L_1
R_1
1
0
L_1
R_1
1/3
2/3
L_2
R_2
2.2
0, 1
2.4
1, 0
2.3
l_2
1/3
r_2
2/3
l_4
1/3
r_4
2/3
1/2
1/2
1/2
1/2
l_3
r_3
l_3
r_3
1.3
0, 1
1, 3
3, 2
4, 0
3, 1
0, 4
4, 0
0
1
L_3
R_3
0, 0
1, 4

Example

$$r_1^t(L_3) = 1/4 \cdot 2/3 \cdot (0 - 1) = -\frac{1}{6}$$
$$r_1^t(R_3) = 1/4 \cdot 2/3 \cdot (1 - 1) = 0$$

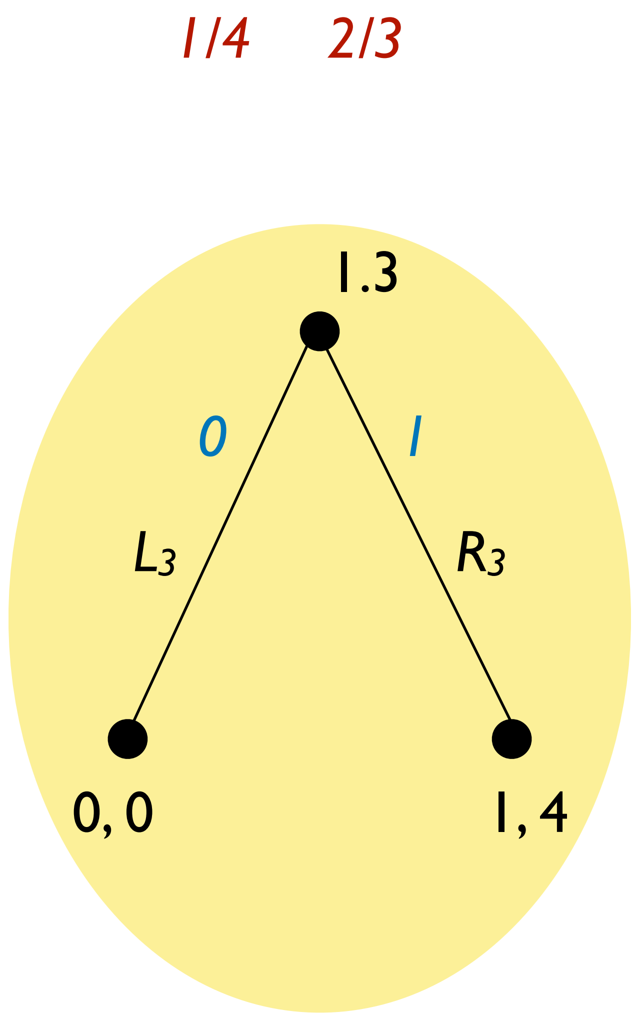
$$\sigma_1^{t+1}(L_3) = \frac{R_1^{t,+}(L_3)}{R_1^{t,+}(L_3) + R_1^{t,+}(R_3)}$$
$$\sigma_1^{t+1}(R_3) = \frac{R_1^{t,+}(R_3)}{R_1^{t,+}(L_3) + R_1^{t,+}(R_3)}$$



Example

$$r_1^t(L_3) = \boxed{1/4 \cdot 2/3} \cdot (0-1) = -\frac{1}{6}$$
$$r_1^t(R_3) = \boxed{1/4 \cdot 2/3} \cdot (1-1) = 0$$

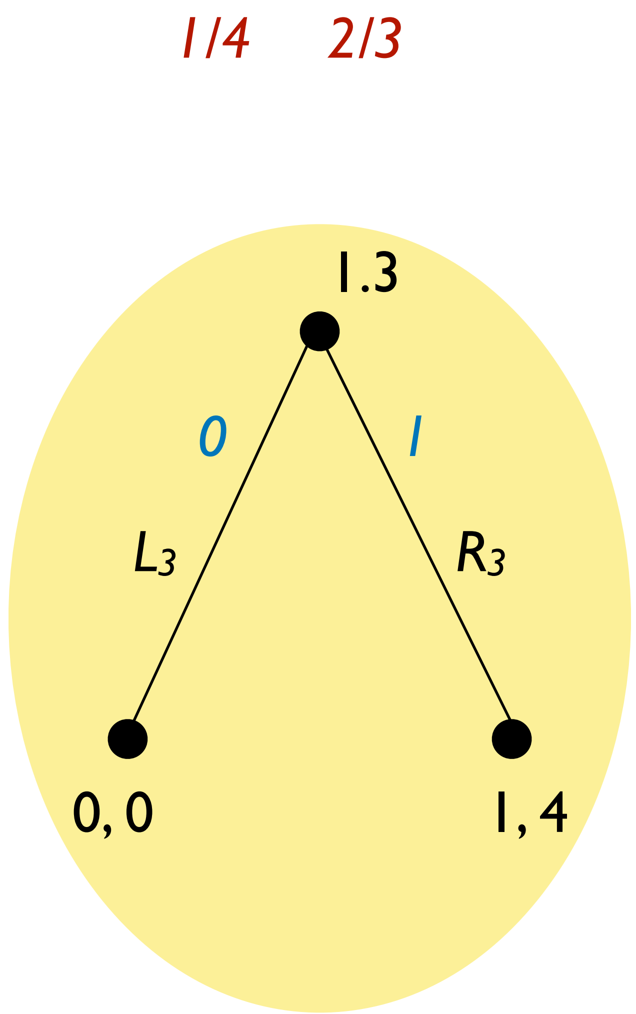
$$\sigma_1^{t+1}(L_3) = \frac{R_1^{t,+}(L_3)}{R_1^{t,+}(L_3) + R_1^{t,+}(R_3)}$$
$$\sigma_1^{t+1}(R_3) = \frac{R_1^{t,+}(R_3)}{R_1^{t,+}(L_3) + R_1^{t,+}(R_3)}$$



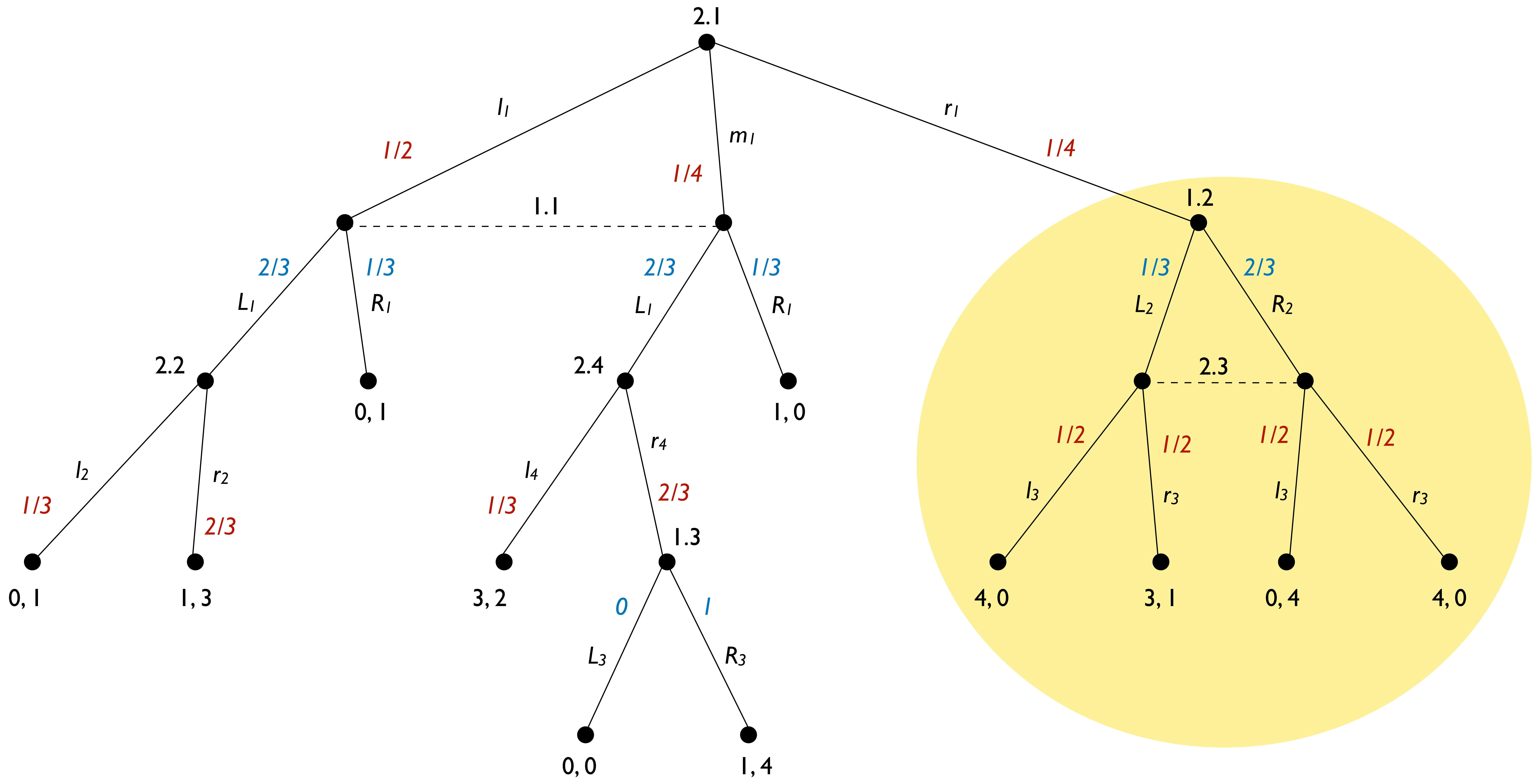
Example

$$r_1^t(L_3) = 1/4 \cdot 2/3 \cdot (0-1) = -\frac{1}{6}$$
$$r_1^t(R_3) = 1/4 \cdot 2/3 \cdot (1-1) = 0$$

$$\sigma_1^{t+1}(L_3) = \frac{R_1^{t,+}(L_3)}{R_1^{t,+}(L_3) + R_1^{t,+}(R_3)}$$
$$\sigma_1^{t+1}(R_3) = \frac{R_1^{t,+}(R_3)}{R_1^{t,+}(L_3) + R_1^{t,+}(R_3)}$$



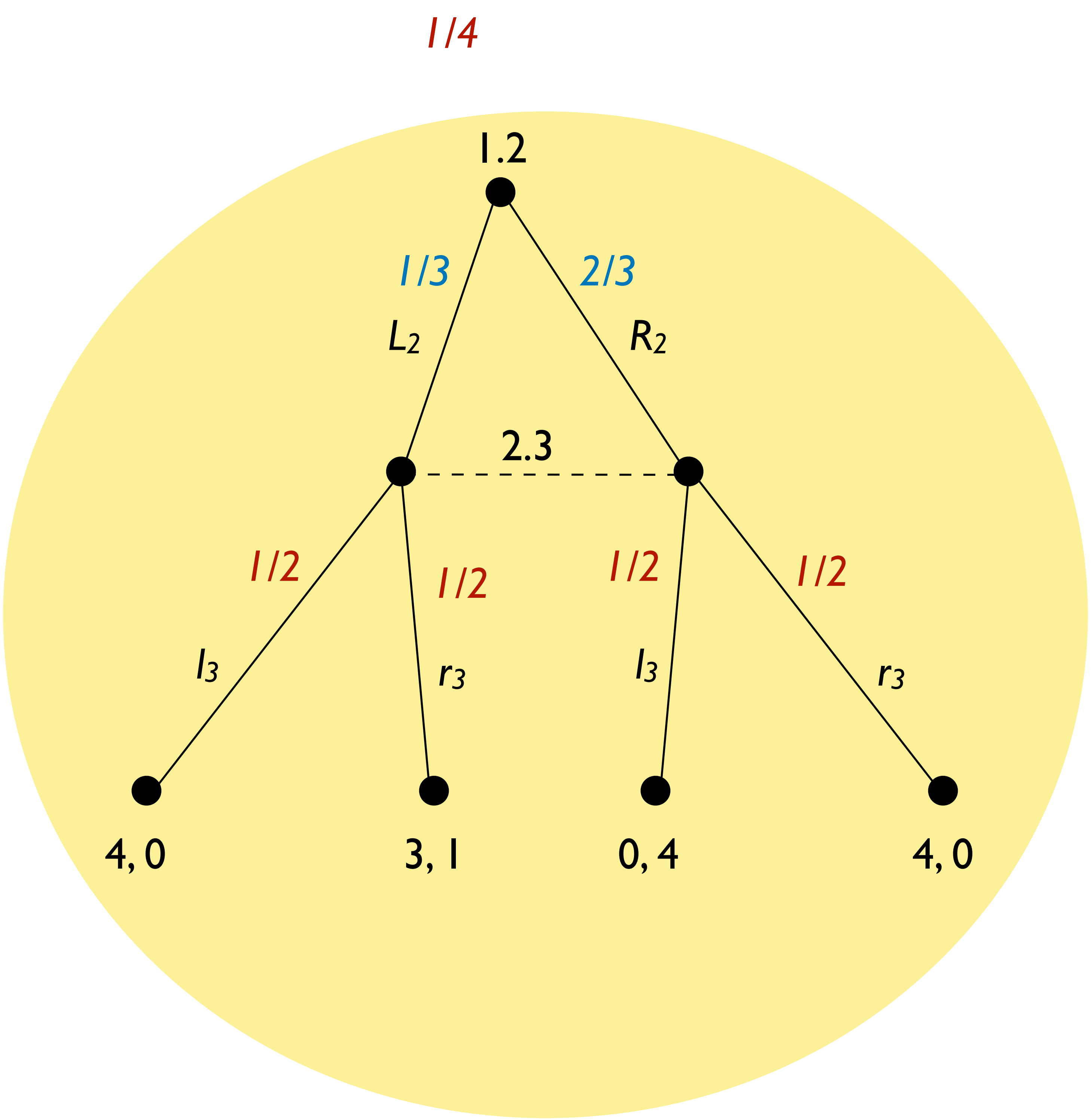
Example



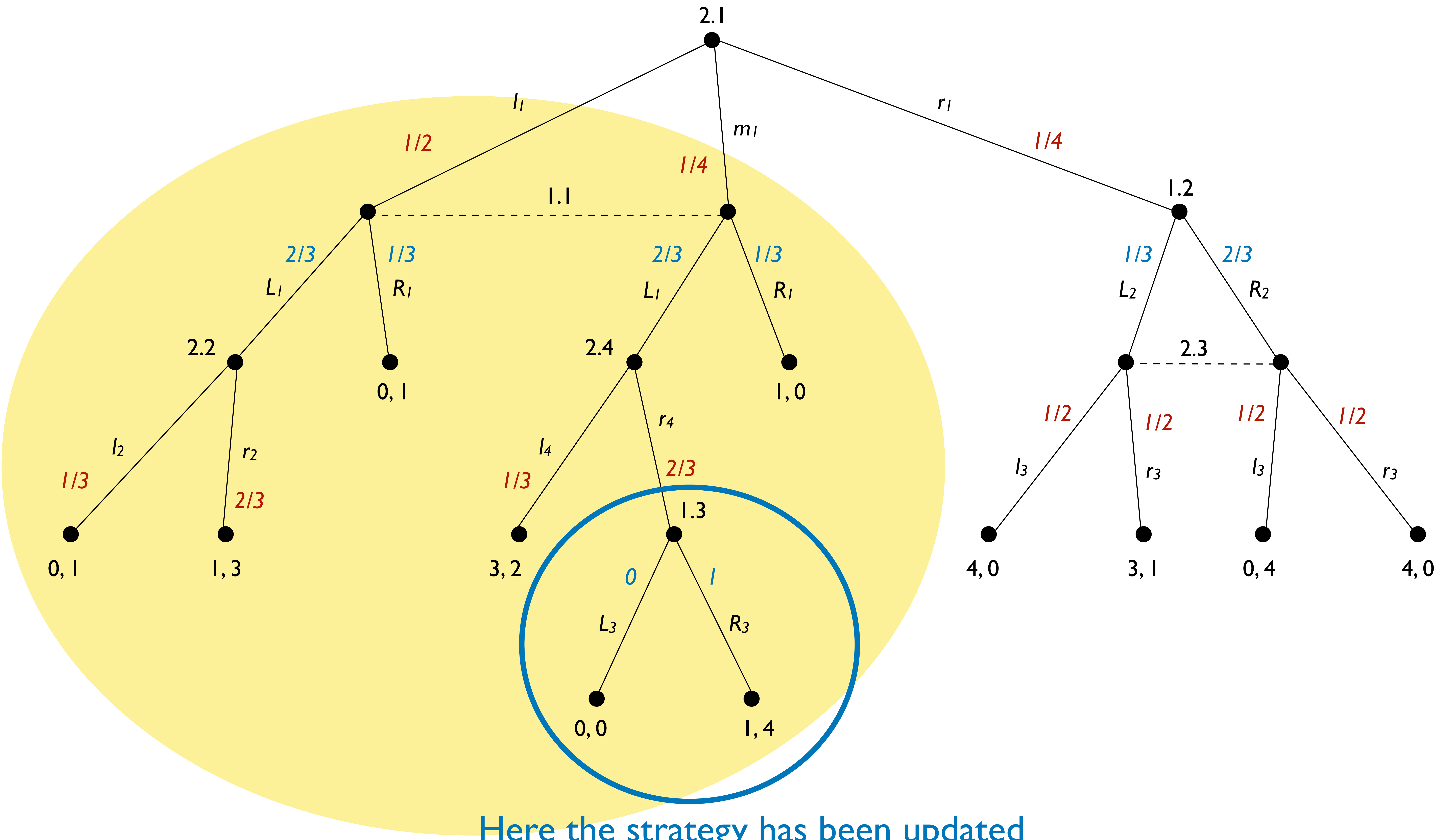
2.1
l1
1/2
m1
1/4
r1
1/4
1.1
1.2
2/3
L1
1/3
R1
2/3
L1
1/3
R1
1/3
L2
2/3
R2
2.2
0, 1
2.4
1, 0
2.3
l2
1/3
r2
2/3
l4
1/3
r4
2/3
1/2
l3
1/2
r3
1/2
l3
1/2
r3
0, 1
1, 3
3, 2
1.3
4, 0
3, 1
0, 4
4, 0
0
L3
1
R3
0, 0
1, 4

Example

As in a normal-form game, except that the instantaneous regrets are multiplied by the probability (i.e., 1/4) with which player 2 reaches 1.2

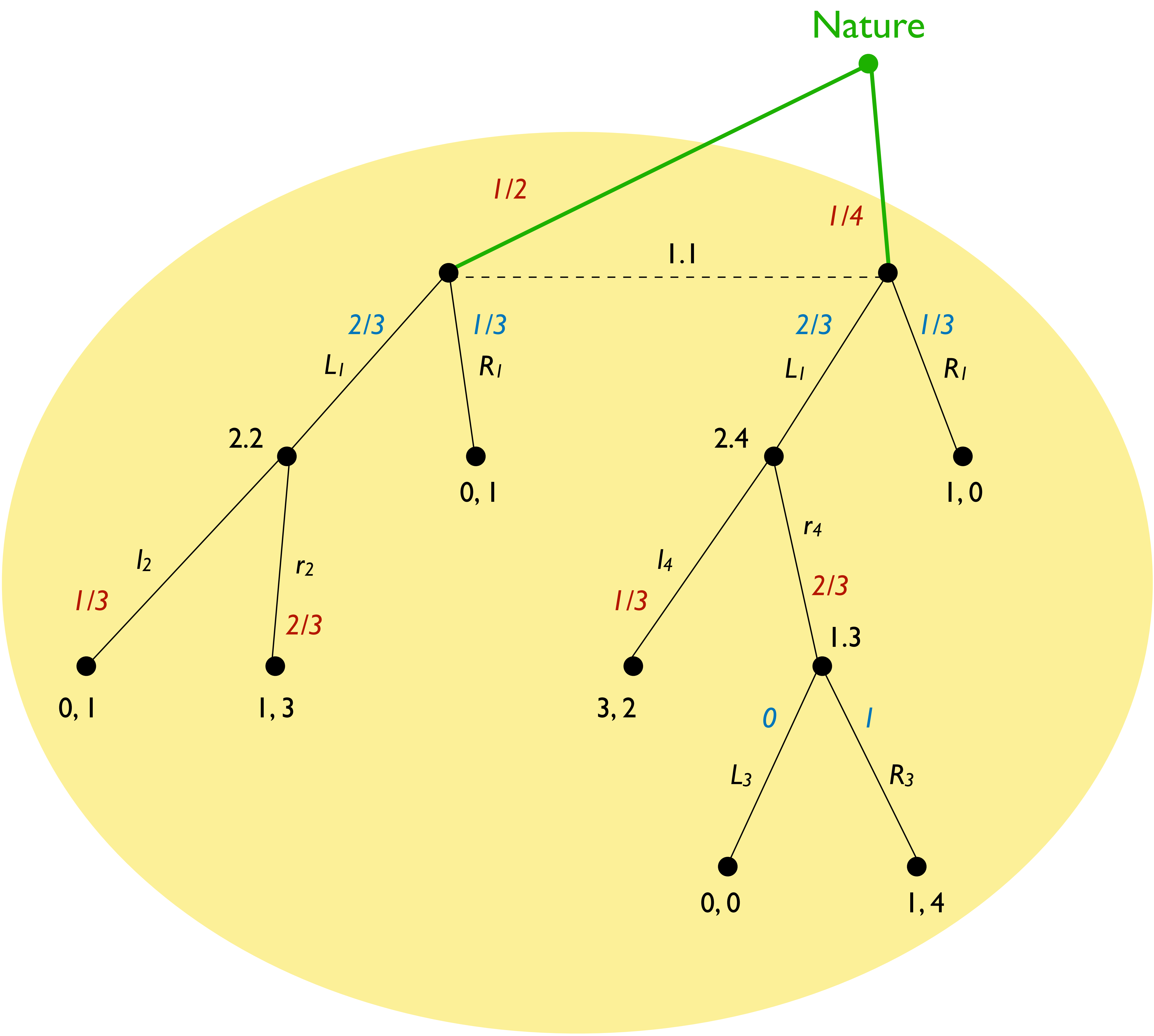


Example



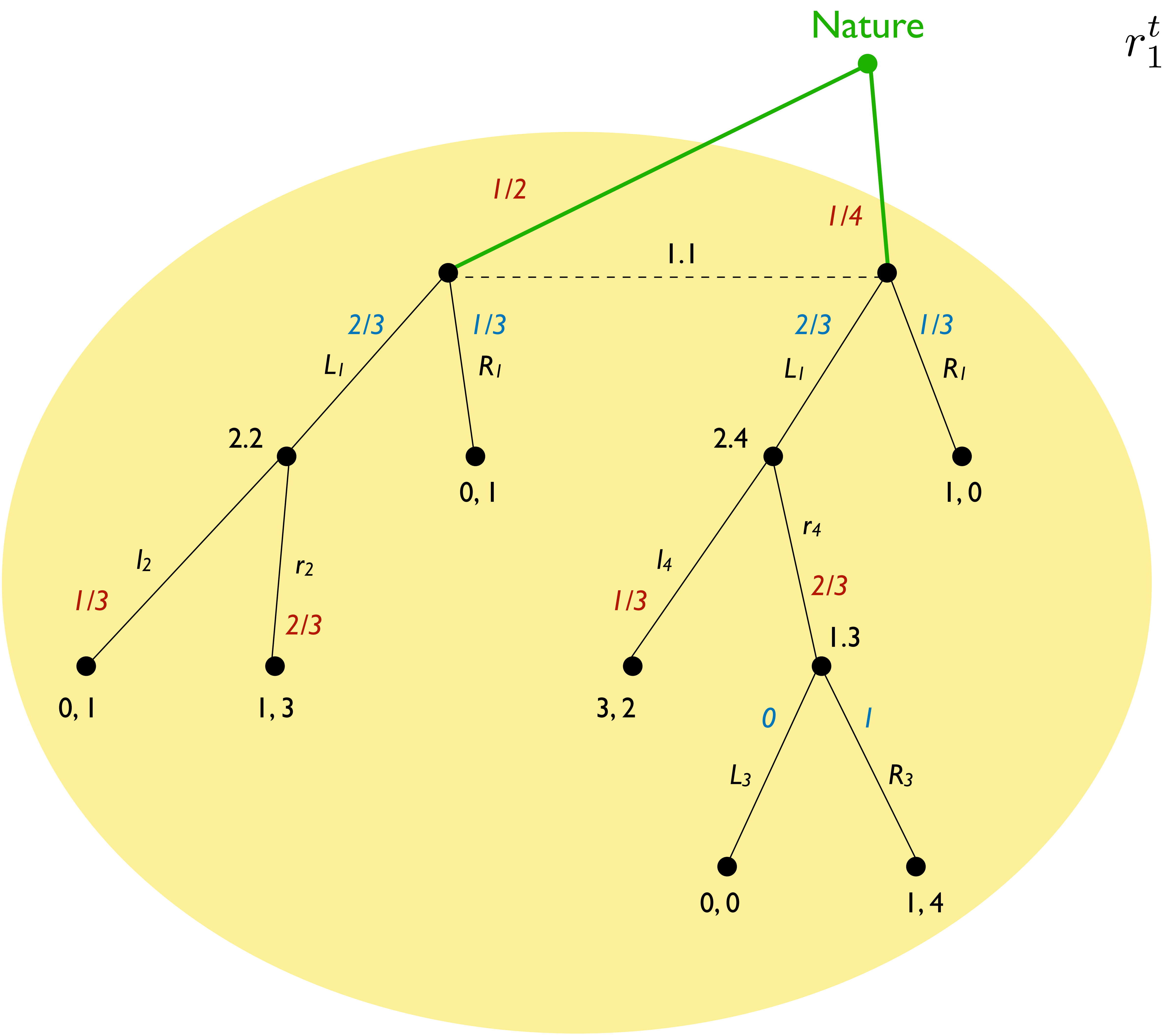
Here the strategy has been updated

Example



- the strategy of player 1 at 1.3 is fixed when calculating the regret at information set 1.1
- the nodes of information set 1.1 are reached with different probabilities; imagine that Nature would have played just before the infoset

Example



$$r_1^t(L_1) = [1/2\,(1/3 \cdot 0 + 2/3 \cdot 1) +$$

Expected utility of L1 from the left node

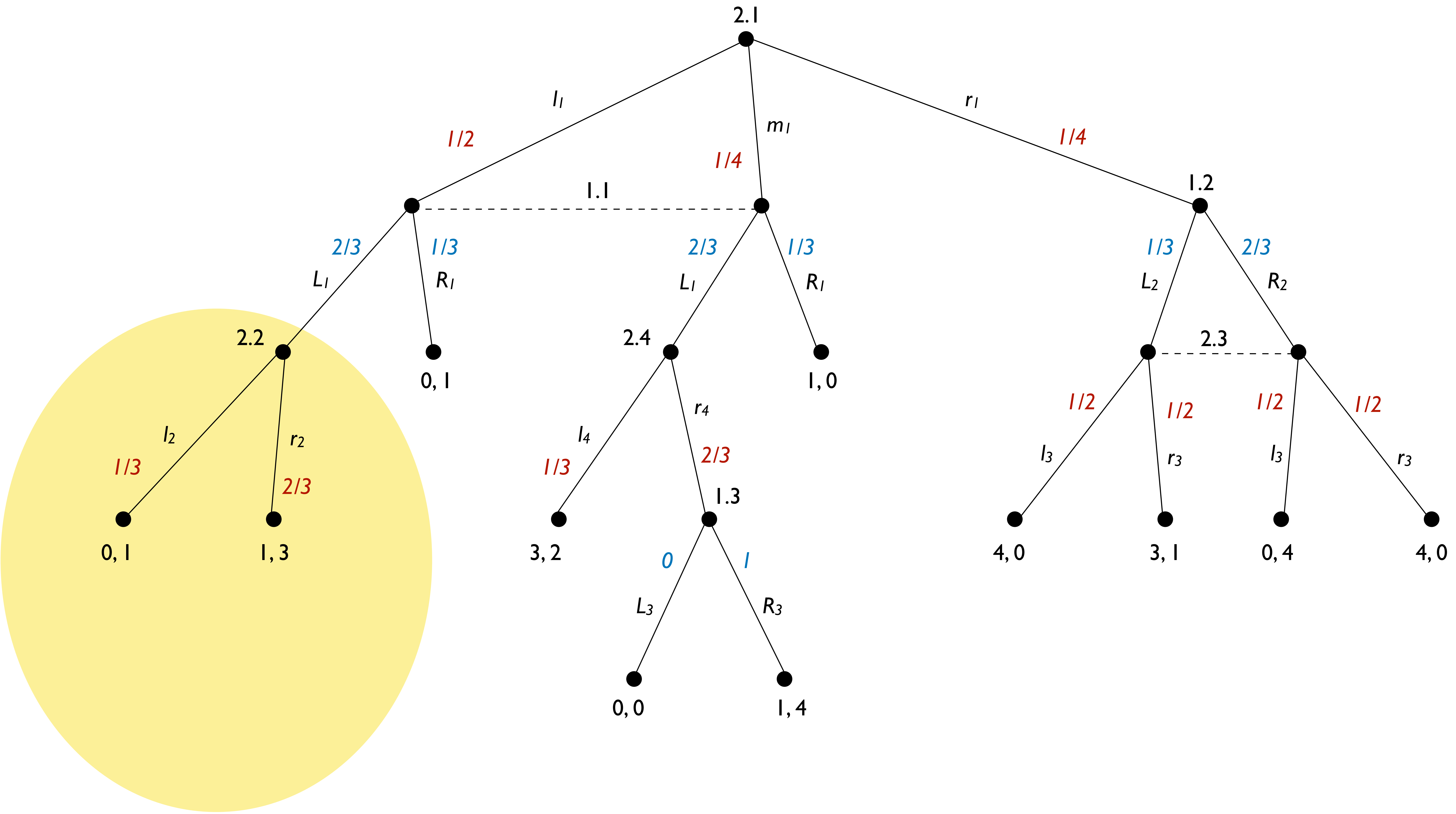
$$1/4(1/3 \cdot 3 + 2/3 \cdot 0 \cdot 0 + 2/3 \cdot 1 \cdot 1)] -$$

Expected utility of L1 from the right node

$$[1/2 \cdot 4/3 + 1/4 \cdot (2/3 + 4/9 + 1/3)]$$

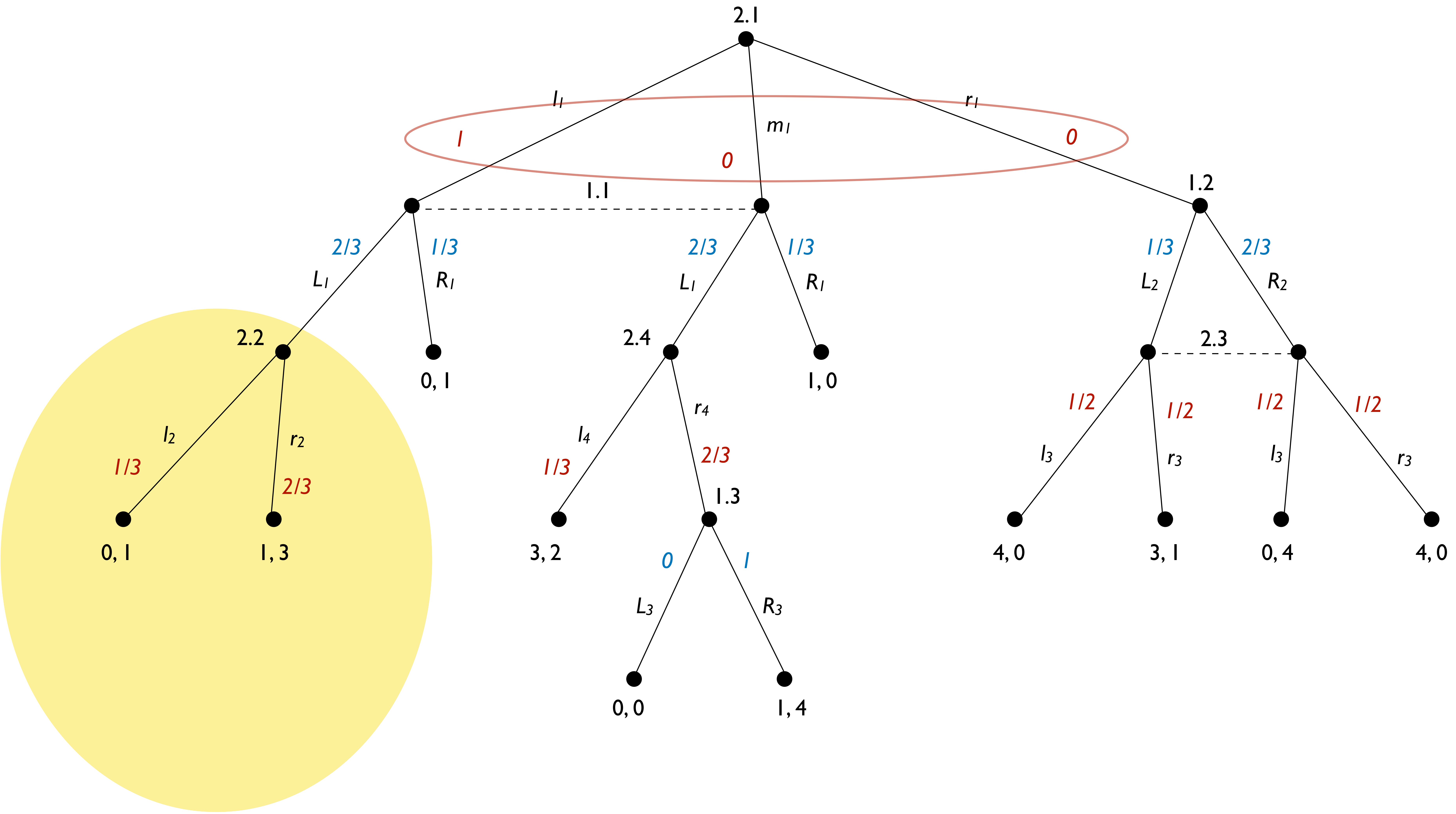
Expected utility of the entire strategy

Example



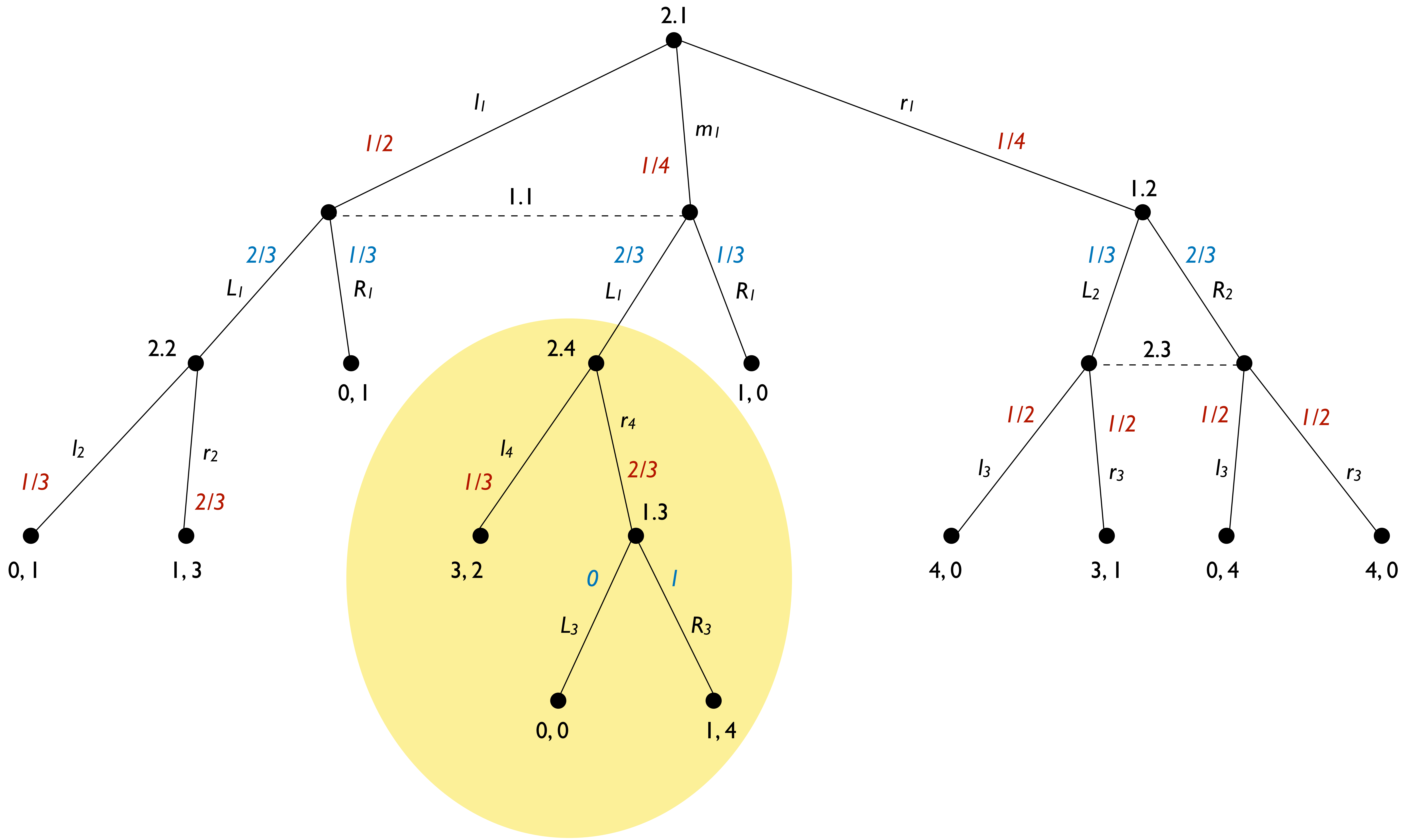
2.1
l1
1/2
m1
1/4
r1
1/4
1.1
1.2
2/3
L1
1/3
R1
2/3
L1
1/3
R1
1/3
L2
2/3
R2
2.2
0, 1
2.4
1, 0
2.3
l2
1/3
r2
2/3
l4
1/3
r4
2/3
1.3
1/2
l3
1/2
r3
1/2
l3
1/2
r3
0, 1
1, 3
3, 2
4, 0
3, 1
0, 4
4, 0
0
L3
1
R3
0, 0
1, 4

Example



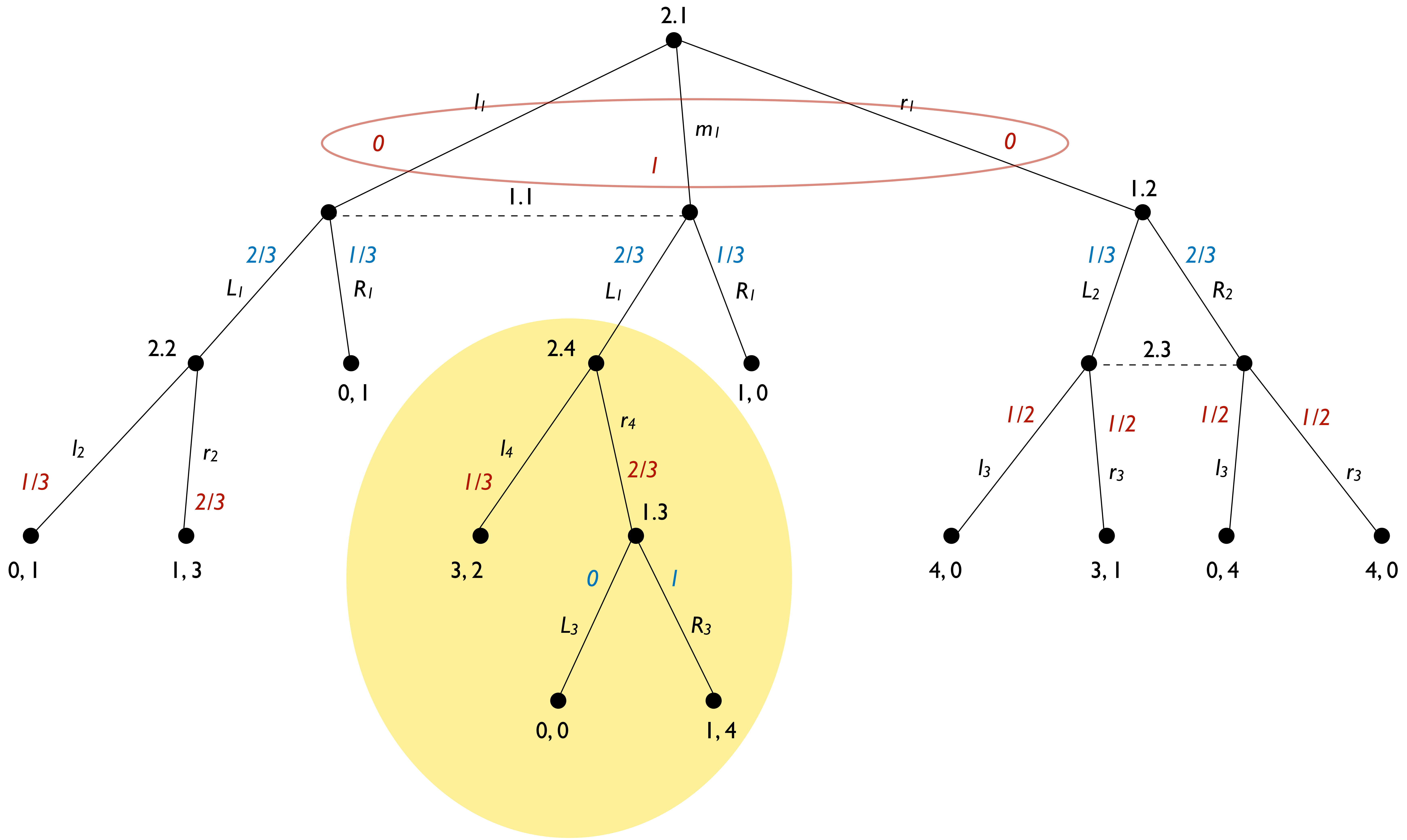
2.1
l1
m1
r1
1
0
0
1.1
1.2
2/3
1/3
L1
R1
2/3
1/3
L1
R1
1/3
2/3
L2
R2
2.2
0, 1
2.4
1, 0
2.3
l2
r2
1/3
2/3
0, 1
1, 3
l4
r4
1/3
2/3
3, 2
1.3
0
1
L3
R3
0, 0
1, 4
1/2
1/2
1/2
1/2
l3
r3
l3
r3
4, 0
3, 1
0, 4
4, 0

Example



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

Example

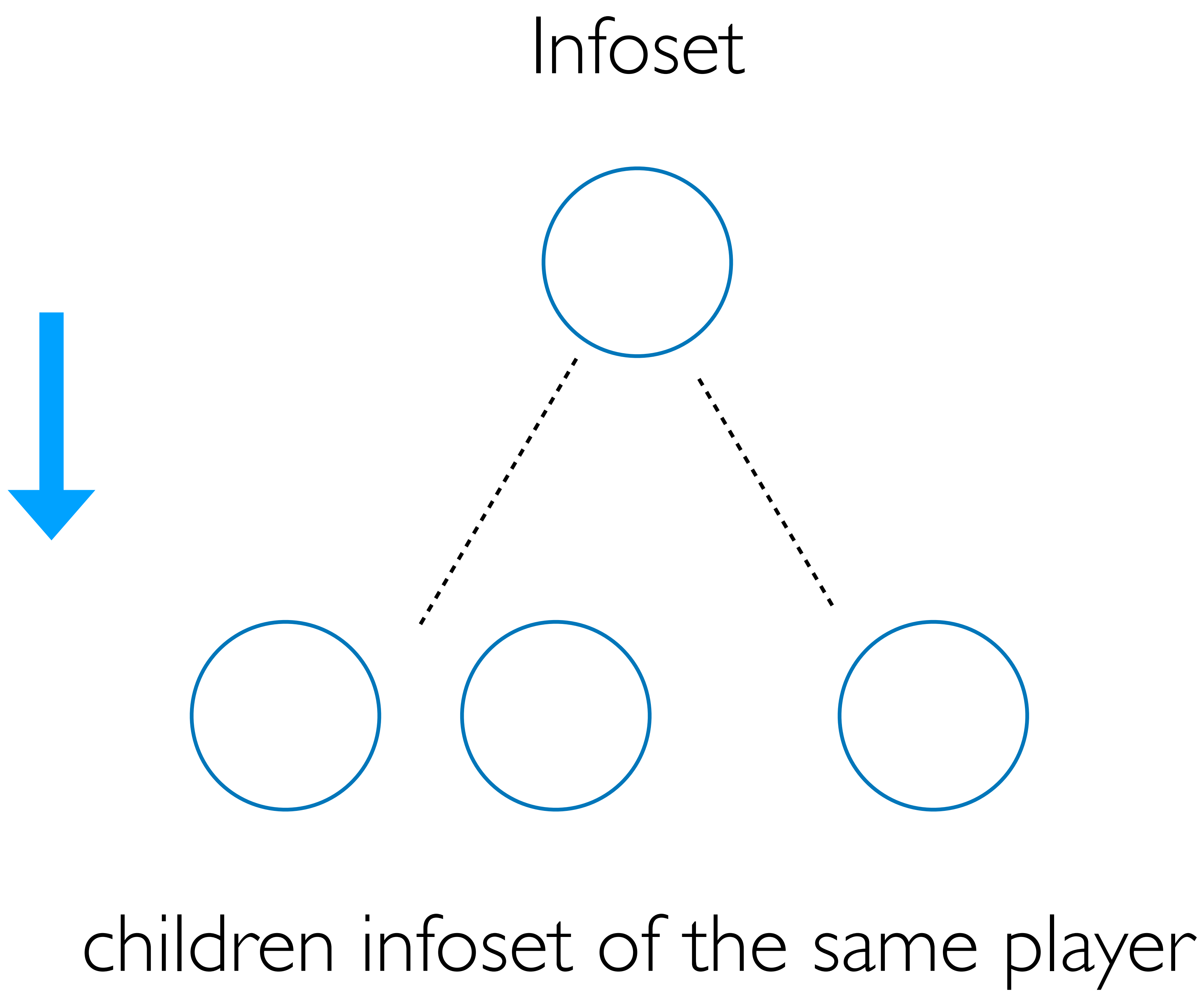


2.1
l1
m1
r1
0
1
0
1.1
1.2
2/3
1/3
2/3
1/3
1/3
2/3
L1
R1
L1
R1
L2
R2
2.2
0, 1
2.4
1, 0
2.3
l2
r2
1/3
2/3
l4
r4
1/3
2/3
1.3
1/2
1/2
1/2
1/2
l3
r3
l3
r3
0, 1
1, 3
3, 2
4, 0
3, 1
0, 4
4, 0
0
1
L3
R3
0, 0
1, 4

Example

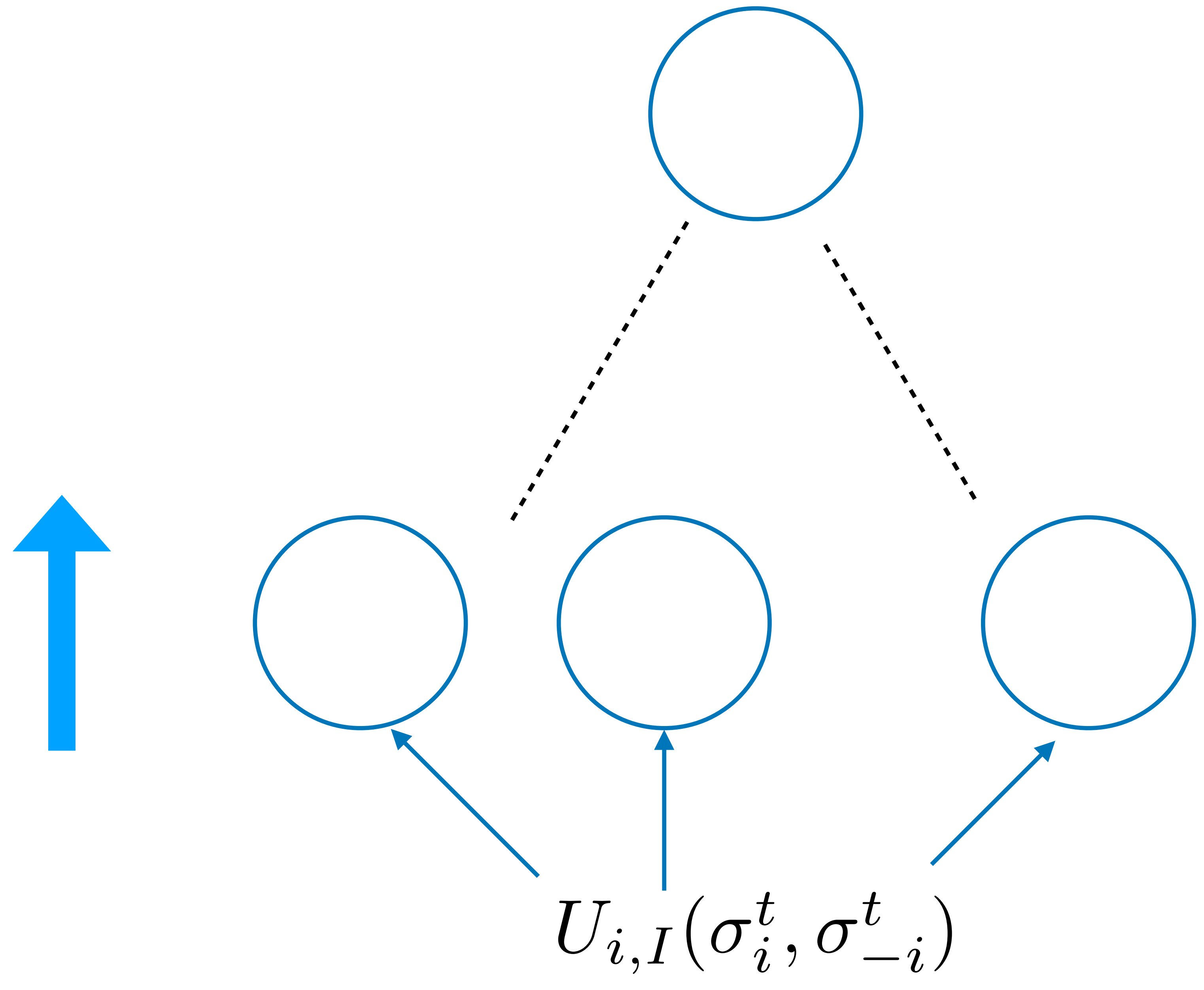
	Player 1								Player 2									
	0	L1	R1	L2	R2	L3	R3		0	l1	m1	r1	l2	r2	l3	r3	l4	r4
R								1										
strategy								1										
R								2										
strategy								2										

Algorithm



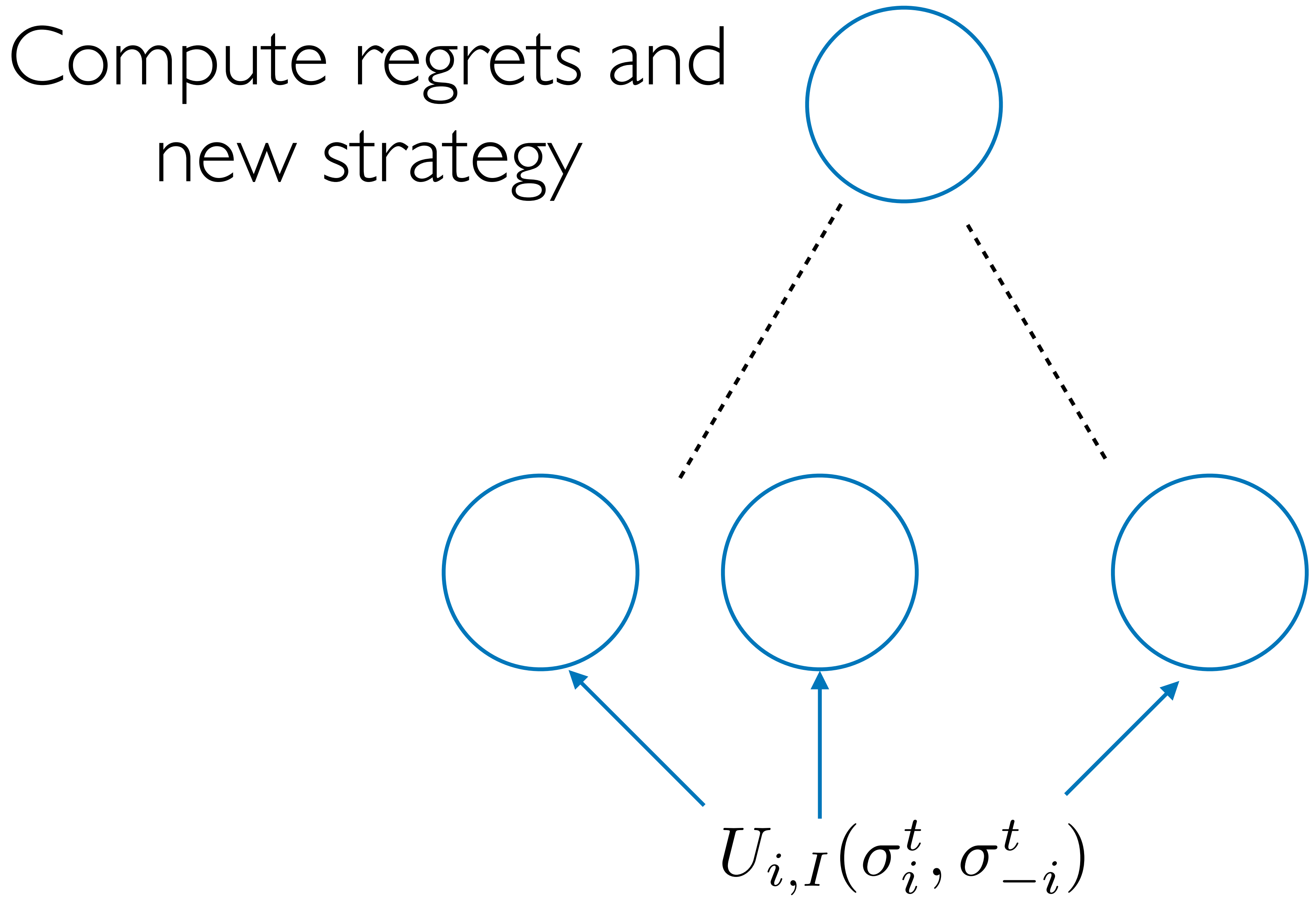
The game tree is traversed to compute regrets and new strategies, counterfactual utility values are propagated upwards

Algorithm



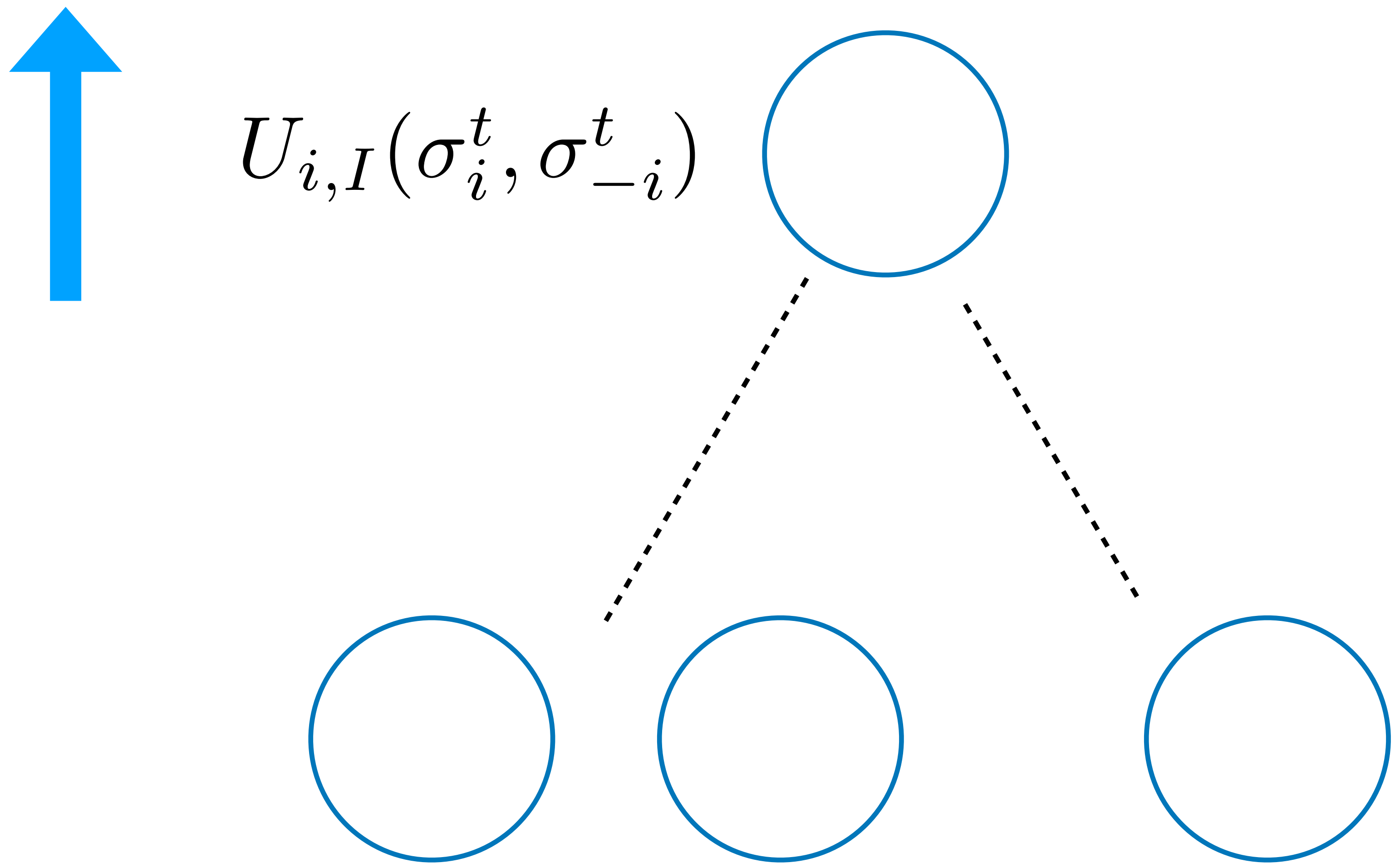
The game tree is traversed to compute regrets and new strategies, counterfactual utility values are propagated upwards

Algorithm



The game tree is traversed to compute regrets and new strategies, counterfactual utility values are propagated upwards

Algorithm



The game tree is traversed to compute regrets and new strategies, counterfactual utility values are propagated upwards

Advanced topics

Regret Matching + (RM+)

RM+ distinguishes from RM for

1. The cumulative regret plus is redefined as

$$R_i^{+,t+1}(a) = \max\left\{R_i^{+,t}(a) + r_i^t(a), 0\right\}$$

2. The calculation of the regrets and the update of the strategies are performed in an alternating fashion

3. The strategy returned by RM+ is obtained by linear weighted averaging

Properties

- RM+ has the same worst-case theoretical guarantees of RM
- RM+ empirically converges much faster than RM

Comparison (RM vs. RM+)

	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		
	R	P	S		R	P	S
R				1			
strategy				1			
R				2			
strategy				2			
R				3			
strategy				3			

Comparison (RM+)

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

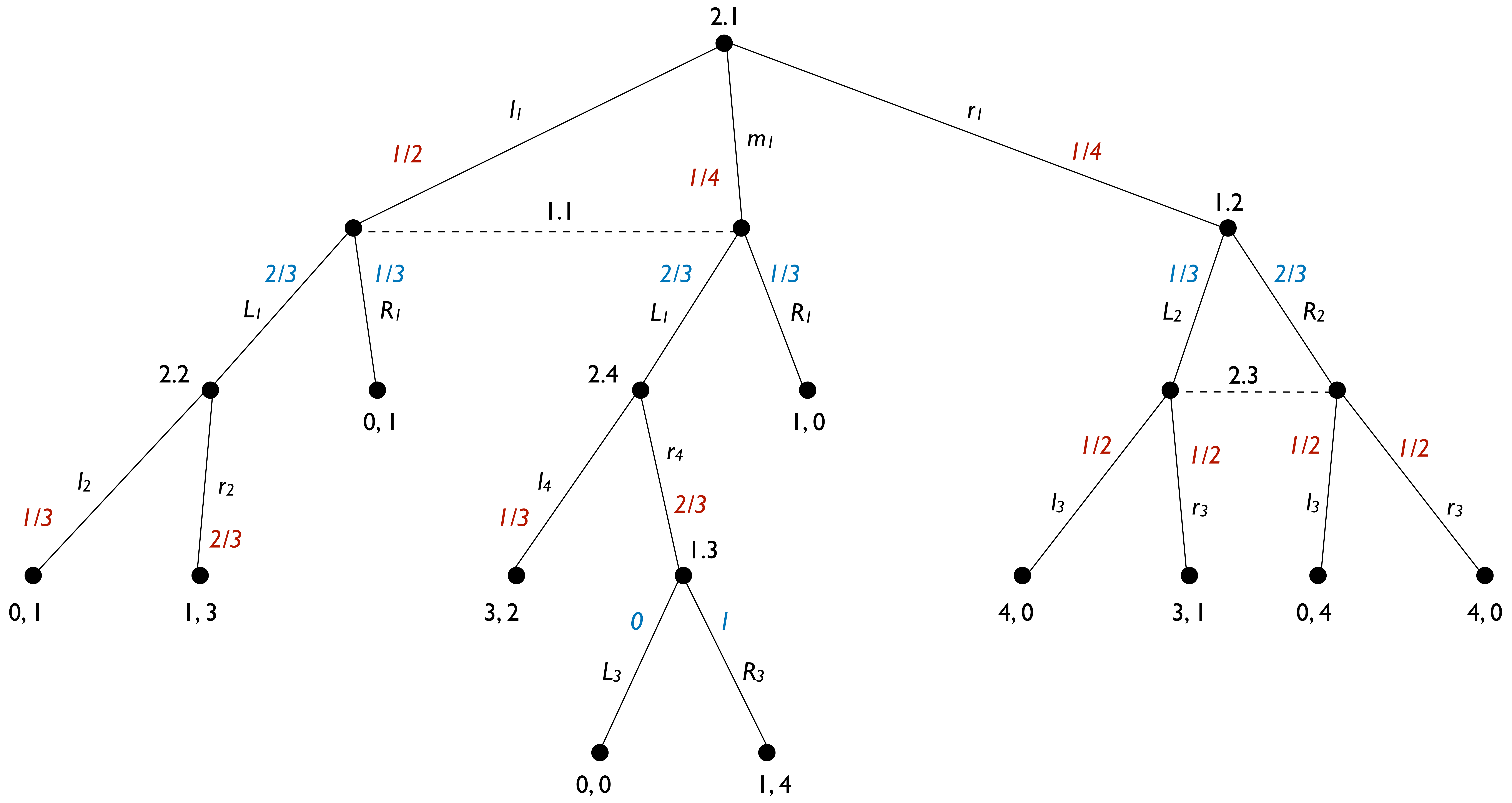
	Player 1				Player 2		
					Average strategy		
	R	P	S		R	P	S
R				1			
strategy				1			
R				2			
strategy				2			
R				3			
strategy				3			

Monte Carlo CFR/CFR+ (external sampling)

When calculating the regret and the strategy of player *i*:

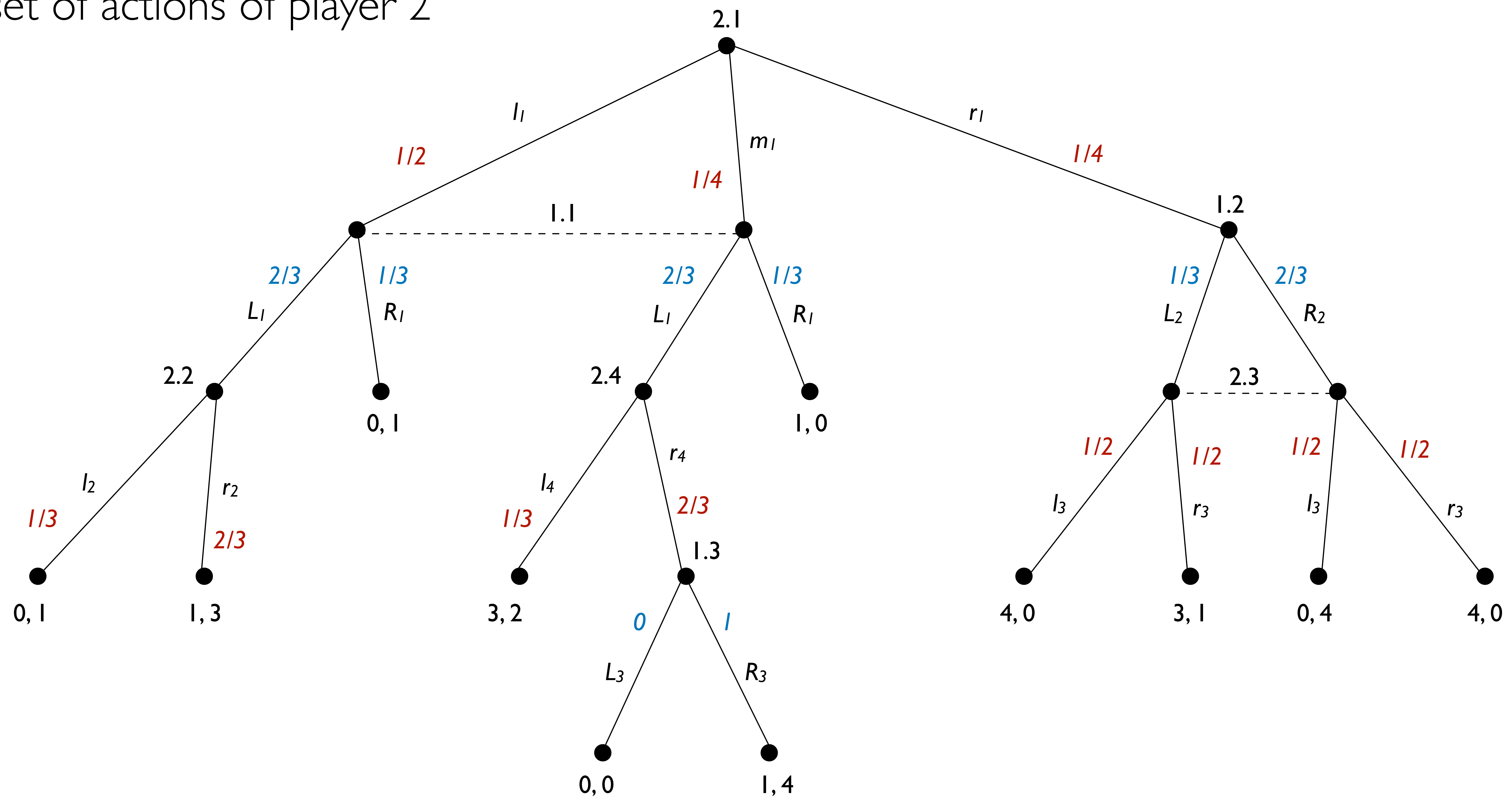
1. Sample a subset of actions of the opponent and of Nature
2. Multiply the payoffs by the inverse of the sampling probability of the corresponding terminal sequence of opponent and chance moves
3. Calculate the instantaneous regrets at every information set reached with strictly positive probability and update the cumulative regrets
4. Update the strategy of player *i* accordingly

Example (1)

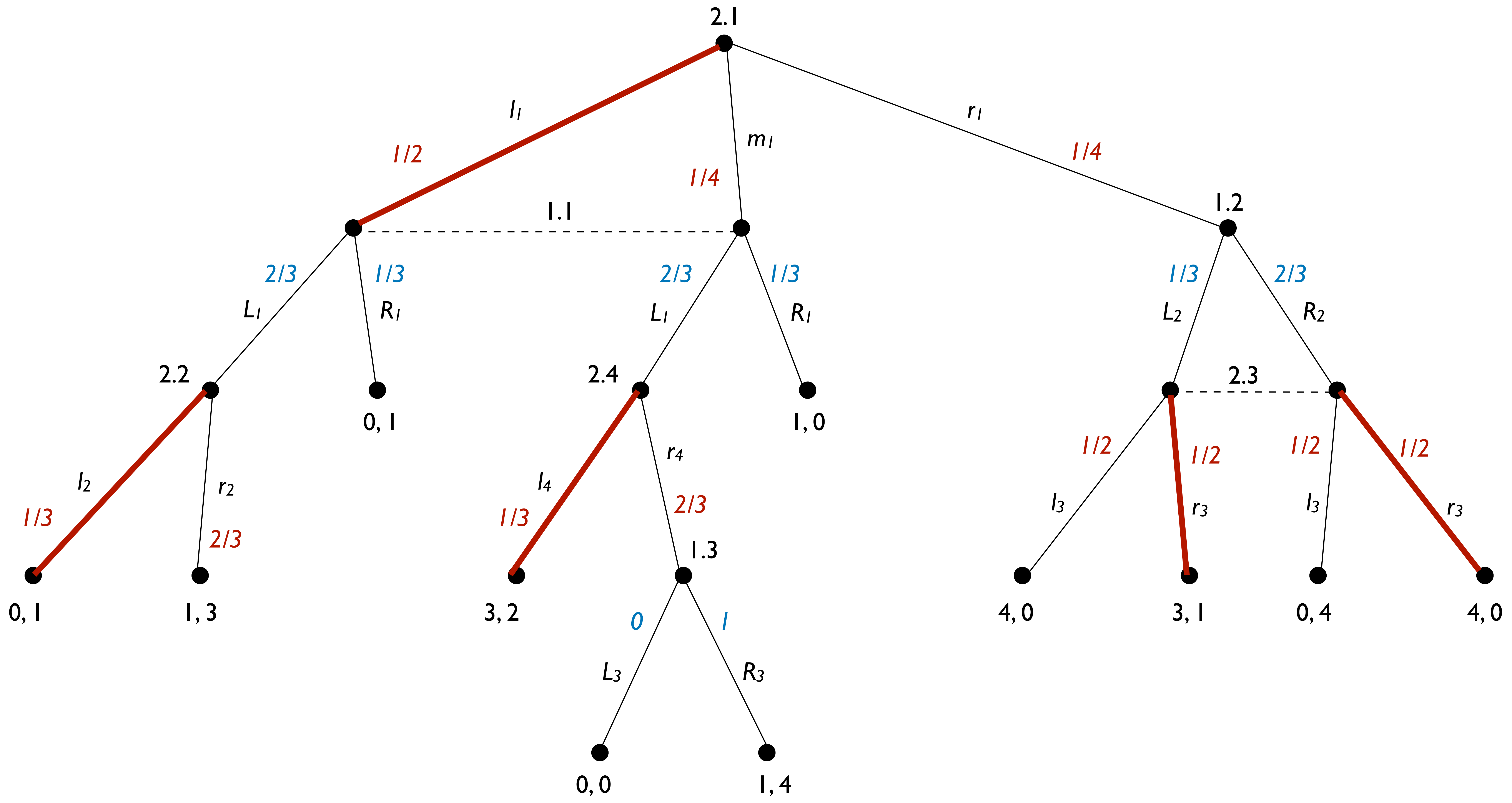

2.1
l1
1/2
m1
1/4
r1
1/4
1.1
1.2
2/3
L1
1/3
R1
2/3
L1
1/3
R1
1/3
L2
2/3
R2
2.2
0, 1
2.4
1, 0
2.3
l2
1/3
r2
2/3
l4
1/3
r4
2/3
1.3
1/2
l3
1/2
r3
1/2
l3
1/2
r3
0, 1
1, 3
3, 2
4, 0
3, 1
0, 4
4, 0
0
L3
1
R3
0, 0
1, 4

Example (1)

Sample a subset of actions of player 2



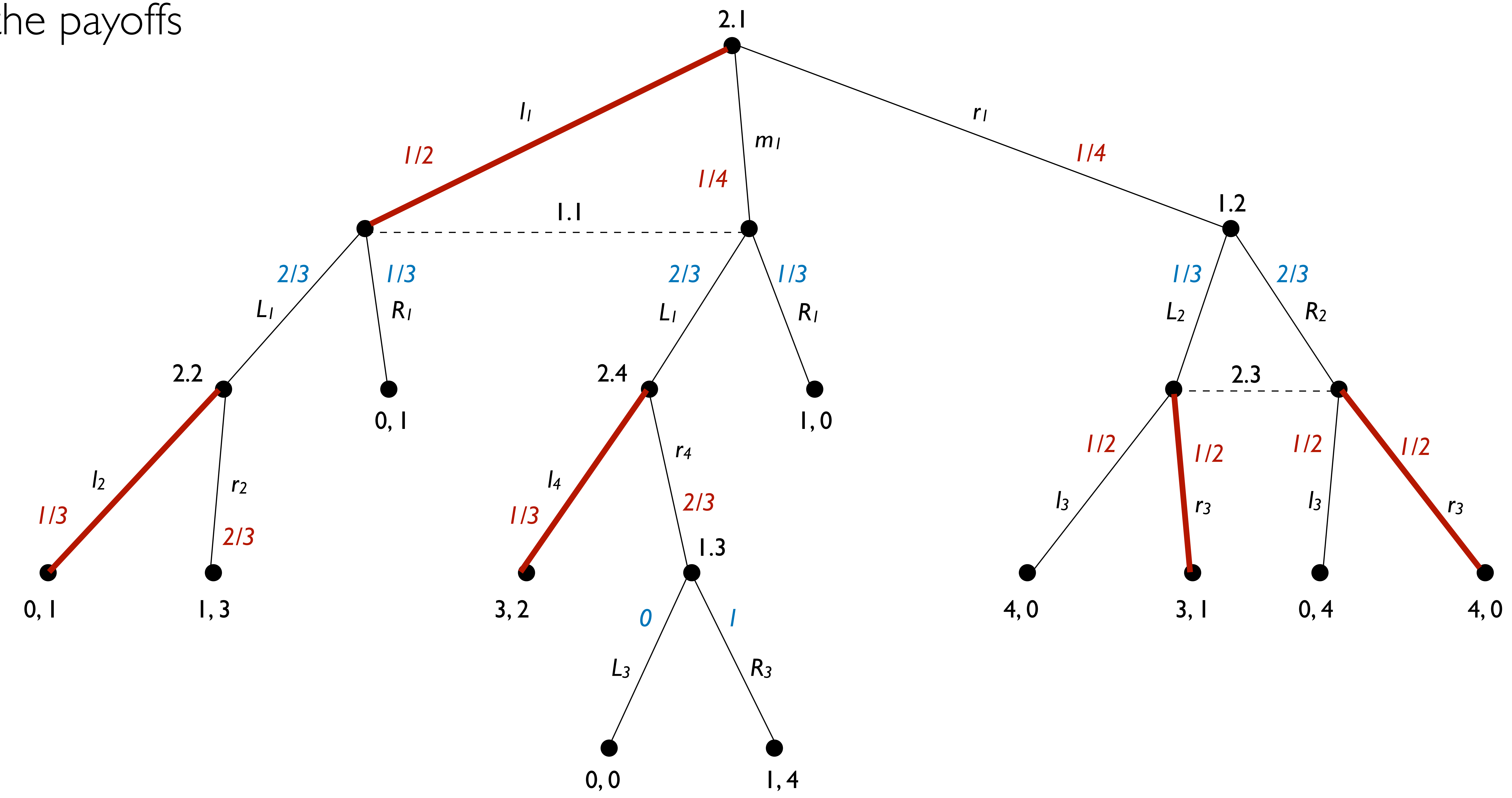
Example (1)



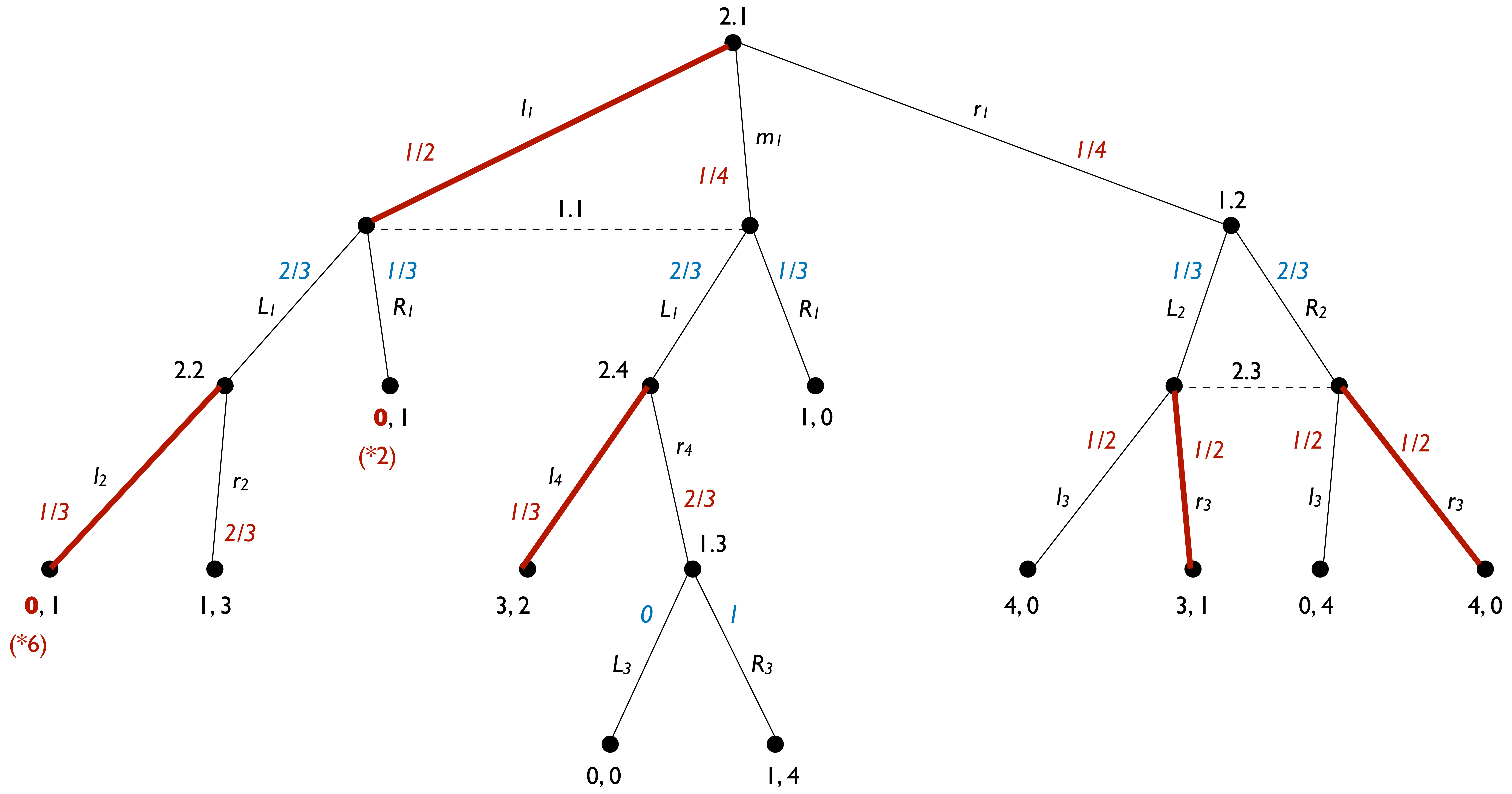
2.1
l_1
1/2
m_1
1/4
r_1
1/4
1.1
1.2
2/3
L_1
1/3
R_1
2/3
L_1
1/3
R_1
1/3
L_2
2/3
R_2
2.2
0, 1
2.4
1, 0
2.3
l_2
1/3
r_2
2/3
l_4
1/3
r_4
2/3
1.3
1/2
l_3
1/2
r_3
1/2
l_3
1/2
r_3
0, 1
1, 3
3, 2
4, 0
3, 1
0, 4
4, 0
0
L_3
1
R_3
0, 0
1, 4

Example (1)

Normalize the payoffs



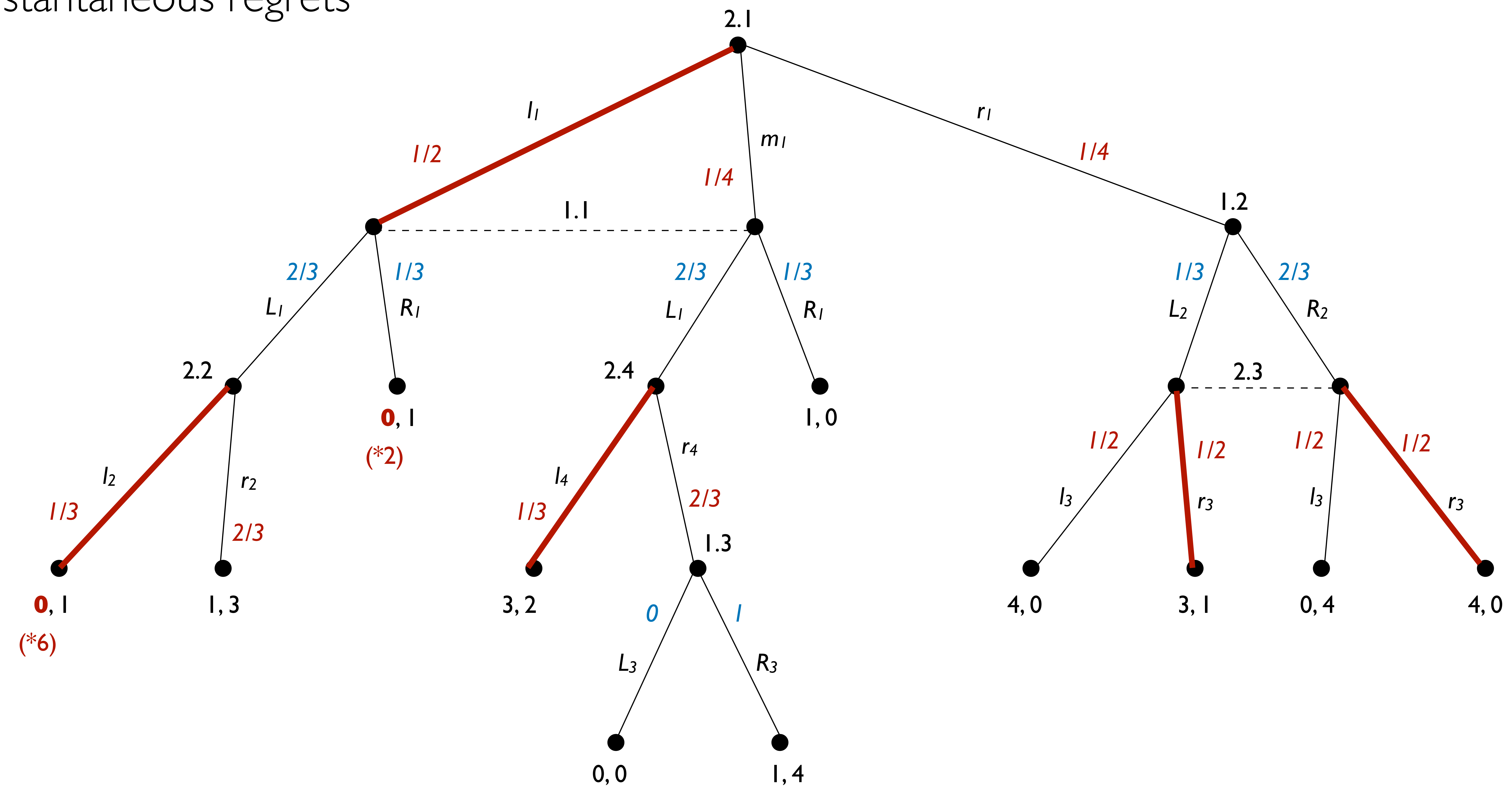
Example (1)



2.1
l1
1/2
m1
1/4
r1
1/4
1.1
1.2
2/3
1/3
L1
R1
2/3
1/3
L1
R1
1/3
2/3
L2
R2
2.2
0, 1
(*2)
2.4
1, 0
2.3
l2
1/3
r2
2/3
l4
1/3
r4
2/3
1/2
1/2
1/2
1/2
l3
r3
l3
r3
0, 1
(*6)
1, 3
3, 2
1.3
4, 0
3, 1
0, 4
4, 0
0
1
L3
R3
0, 0
1, 4

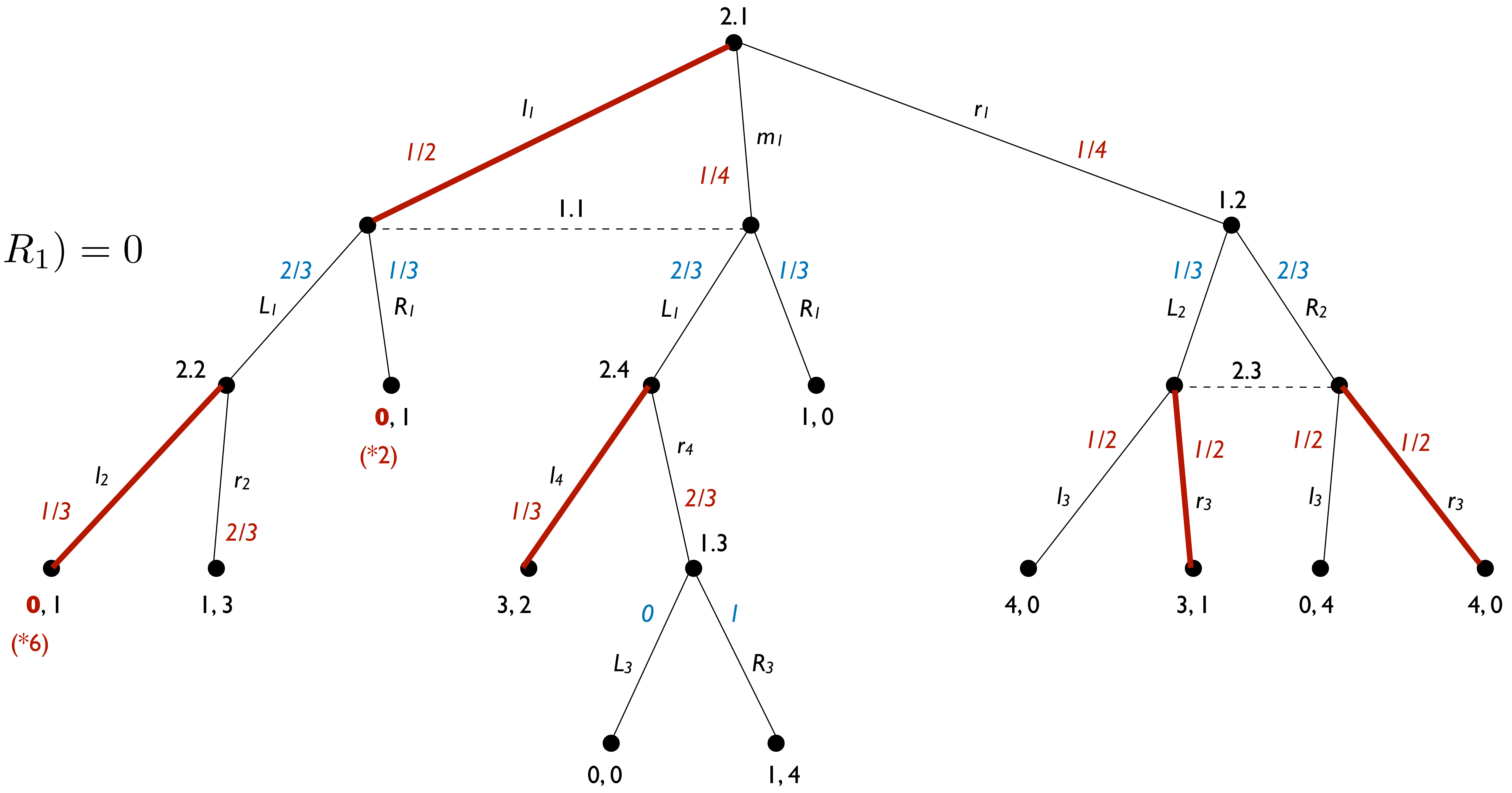
Example (1)

Calculate the instantaneous regrets

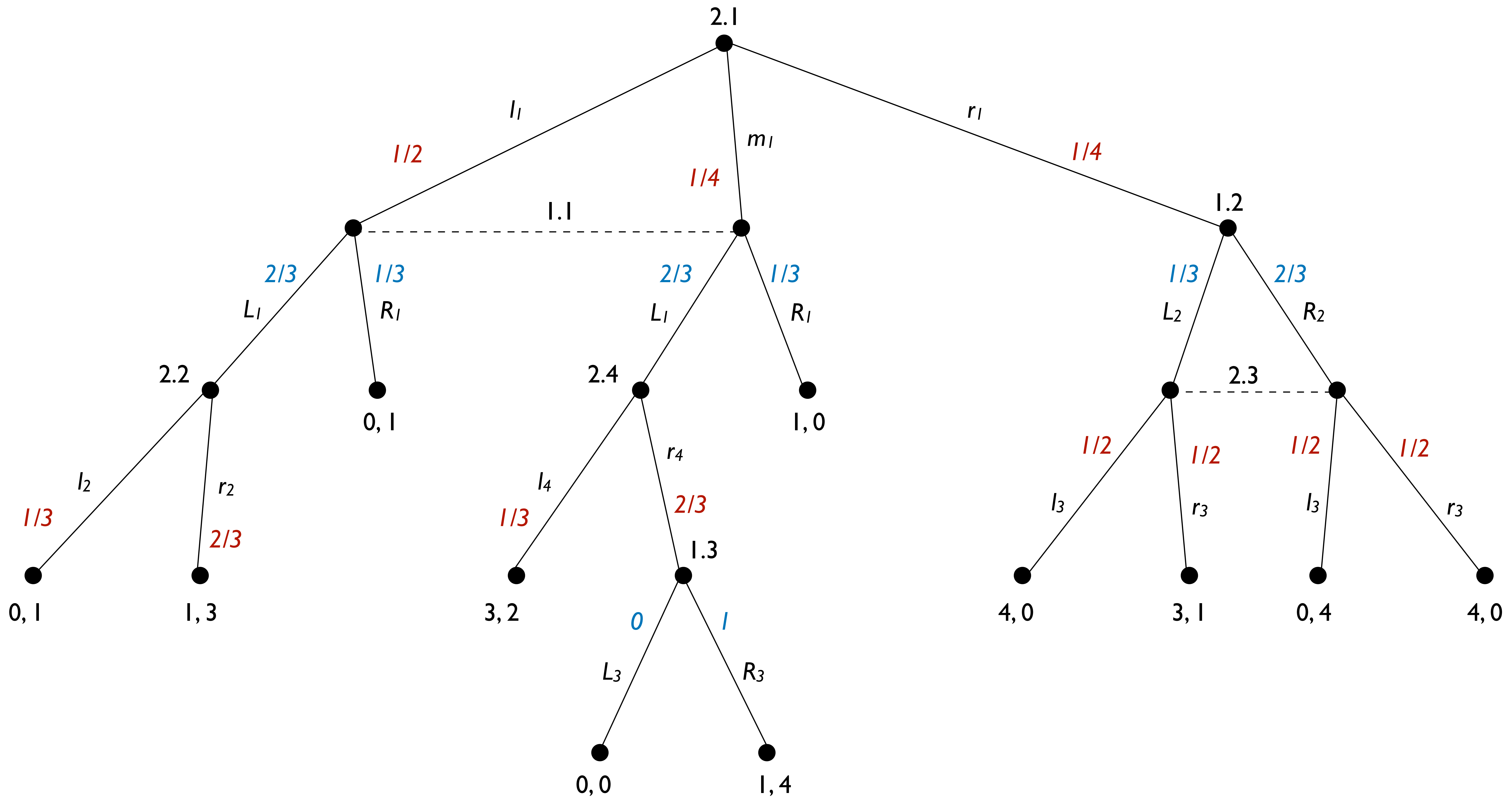


Example (1)

$$r_1^t(L_1) = r_1^t(R_1) = 0$$

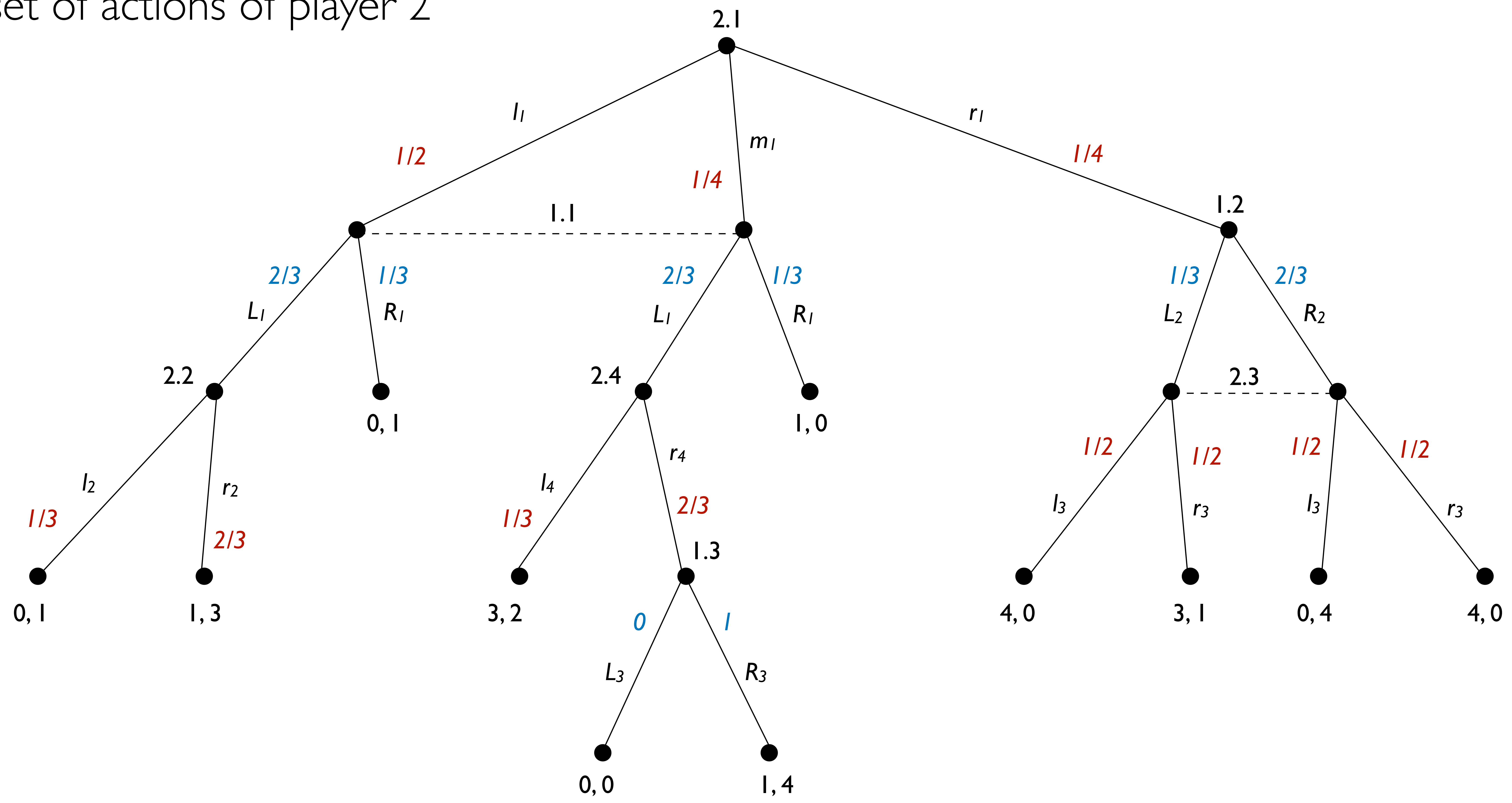


Example (2)

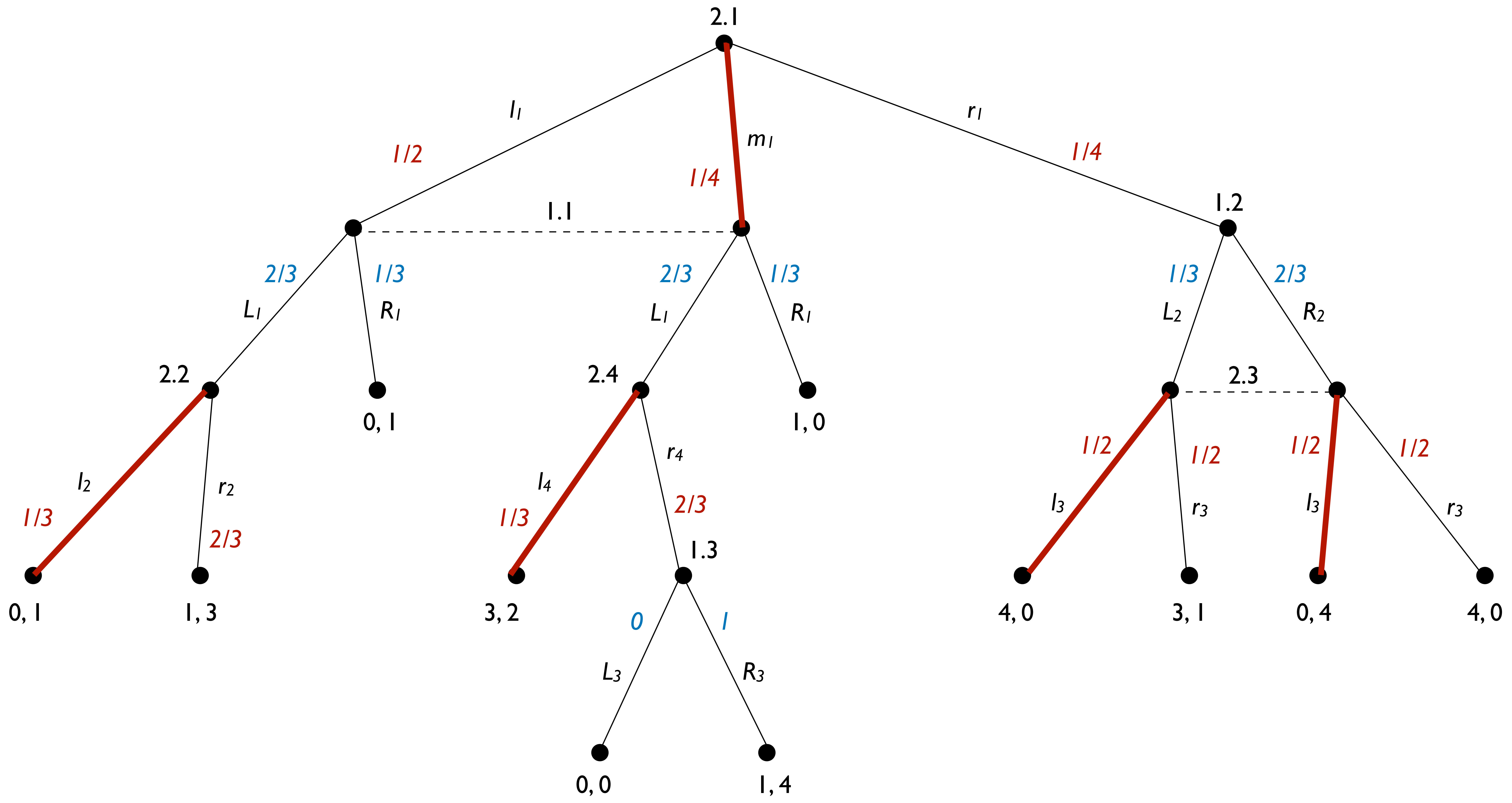

2.1
l_1
$1/2$
m_1
$1/4$
r_1
$1/4$
1.1
1.2
$2/3$
L_1
$1/3$
R_1
$2/3$
L_1
$1/3$
R_1
$1/3$
L_2
$2/3$
R_2
2.2
0, 1
2.4
1, 0
2.3
l_2
$1/3$
r_2
$2/3$
l_4
$1/3$
r_4
$2/3$
1.3
$1/2$
l_3
$1/2$
r_3
$1/2$
l_3
$1/2$
r_3
0, 1
1, 3
3, 2
4, 0
3, 1
0, 4
4, 0
0
L_3
1
R_3
0, 0
1, 4

Example (2)

Sample a subset of actions of player 2



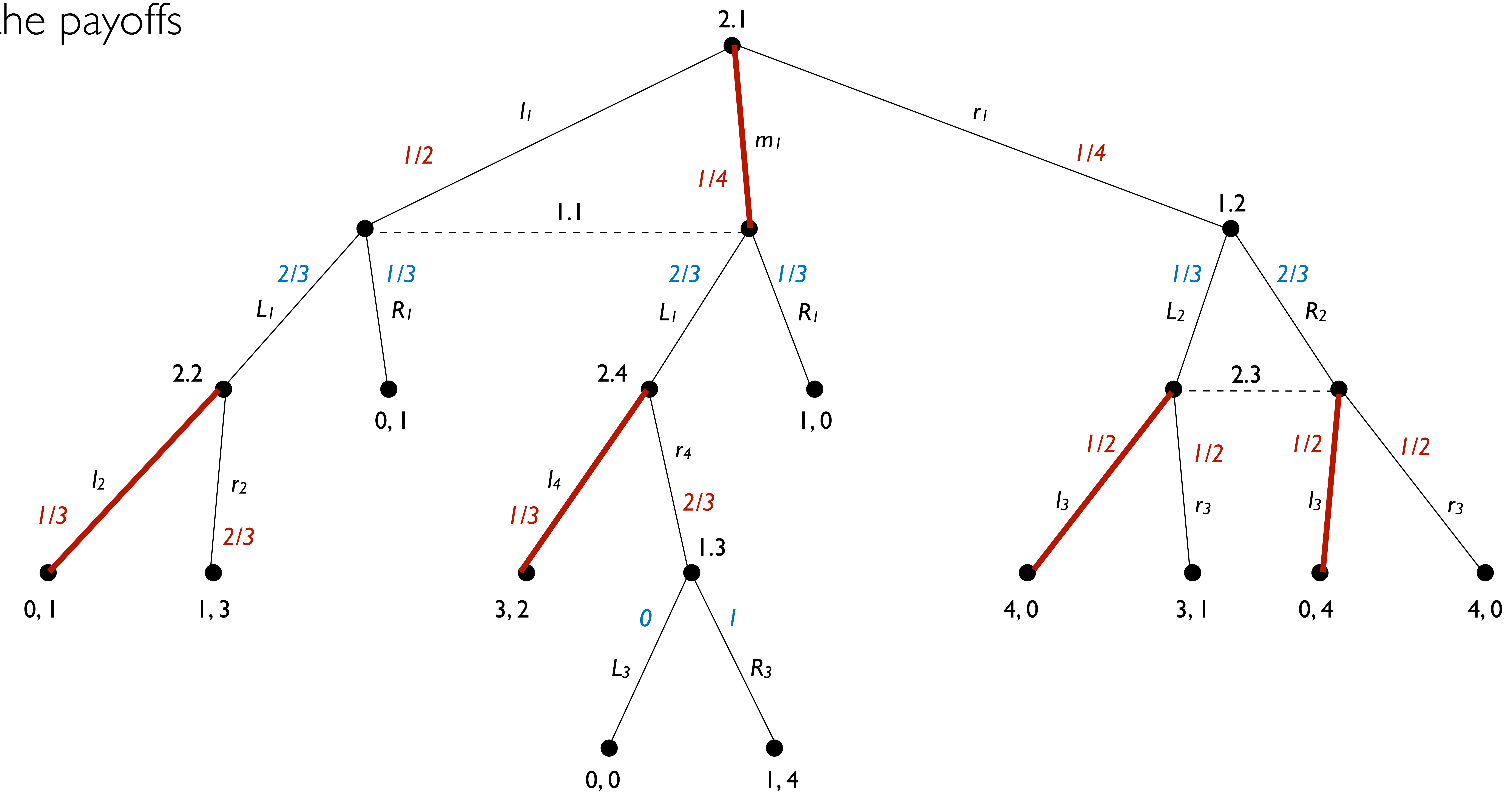
Example (2)



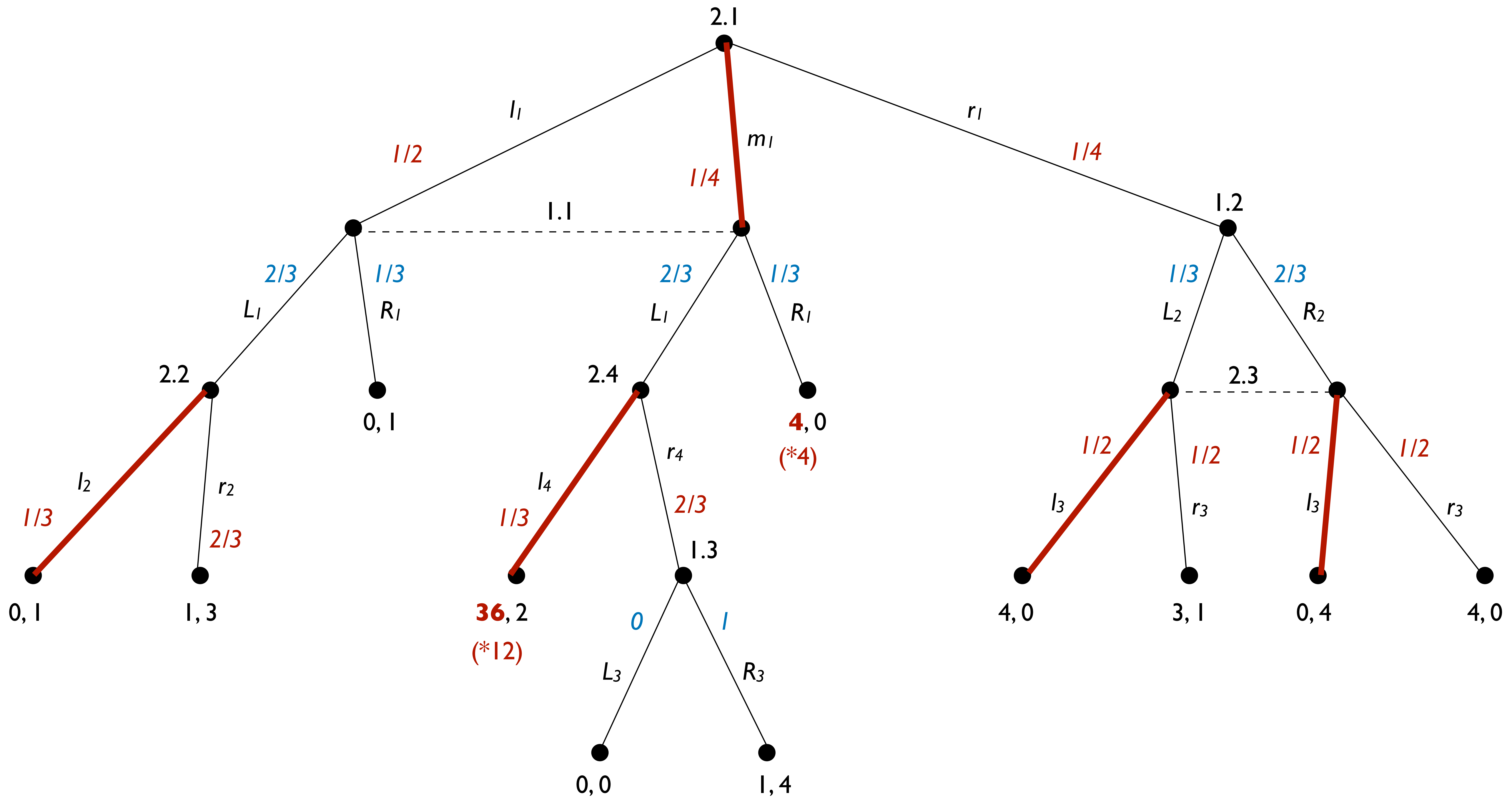
2.1
l1
1/2
m1
1/4
r1
1/4
1.1
1.2
2/3
L1
1/3
R1
2/3
L1
1/3
R1
1/3
L2
2/3
R2
2.2
0, 1
2.4
1, 0
2.3
l2
1/3
r2
2/3
l4
1/3
r4
2/3
1.3
1/2
l3
1/2
r3
1/2
l3
1/2
r3
0, 1
1, 3
3, 2
4, 0
3, 1
0, 4
4, 0
0
L3
1
R3
0, 0
1, 4

Example (2)

Normalize the payoffs

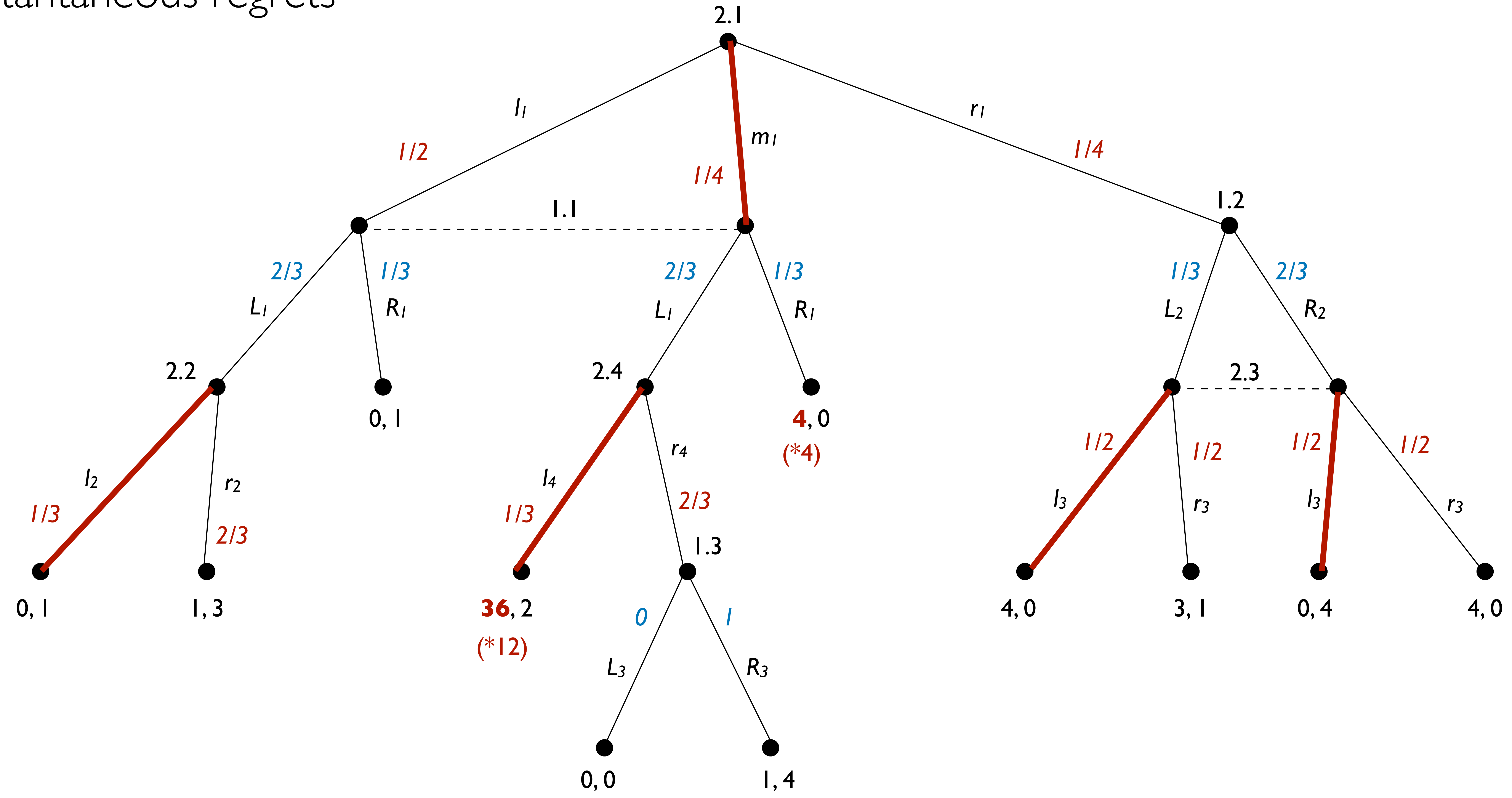


Example (2)


2.1
l_1
1/2
m_1
1/4
r_1
1/4
1.1
1.2
2/3
L_1
1/3
R_1
2/3
L_1
1/3
R_1
1/3
L_2
2/3
R_2
2.2
0, 1
2.4
4, 0
(*4)
2.3
l_2
1/3
r_2
2/3
l_4
1/3
r_4
2/3
1.3
1/2
l_3
1/2
r_3
1/2
l_3
1/2
r_3
0, 1
1, 3
36, 2
(*12)
4, 0
3, 1
0, 4
4, 0
0
L_3
1
R_3
0, 0
1, 4

Example (2)

Calculate the instantaneous regrets



Example (2)

$$r_1^t(L_1) = 3 - 7/3 = 2/3$$
$$r_1^t(R_1) = 1 - 7/3 = -4/3$$