Game Theory, Evolutionary Dynamics, and Multi-Agent Learning

Prof. Nicola Gatti
(nicola.gatti@polimi.it)

Game theory

Game theory: basics

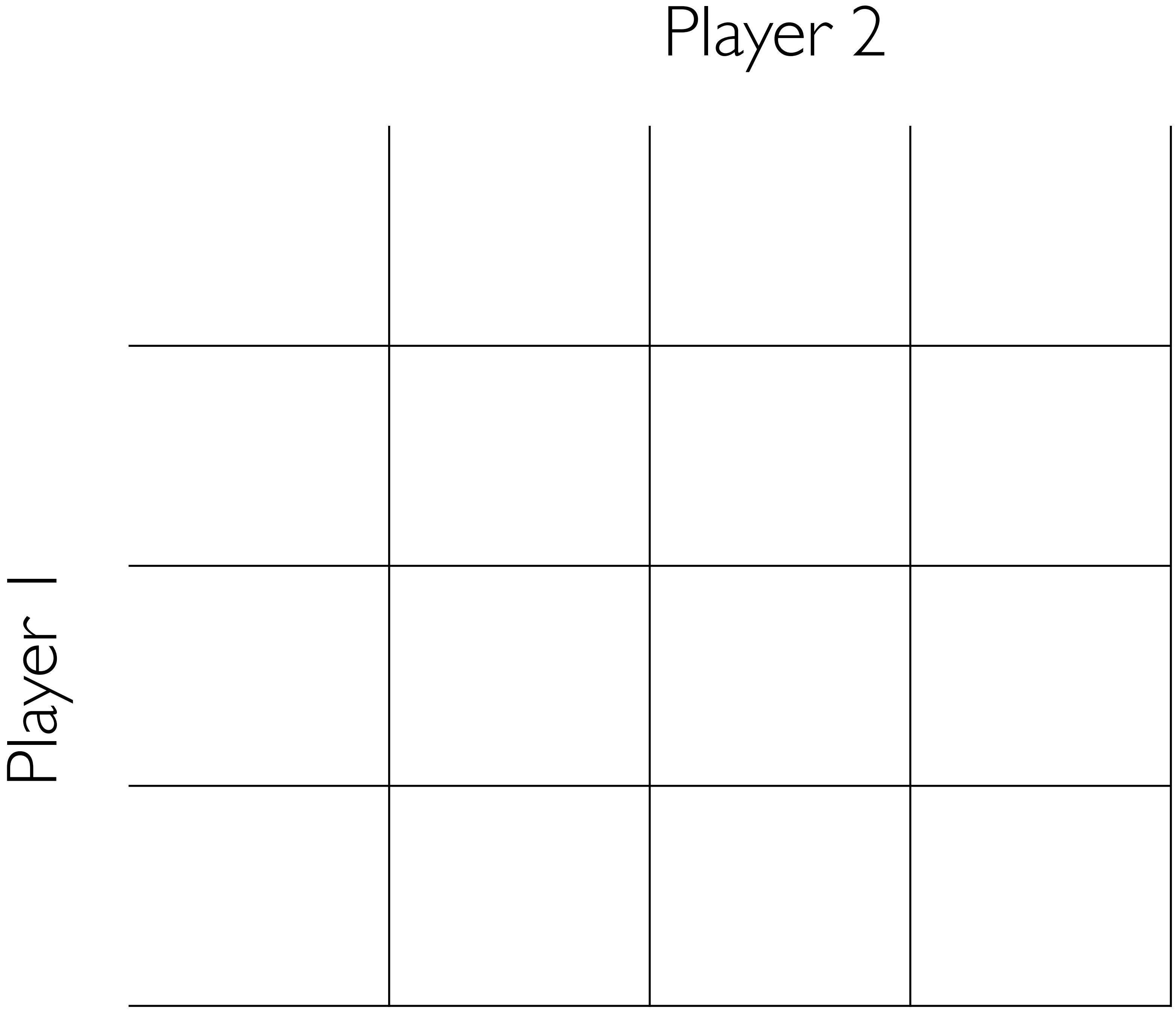
Normal form

- Players
- Actions
- Outcomes
- Utilities
- Strategies
- Solutions

Game theory: basics

Normal form

- Players
- Actions
- Outcomes
- Utilities
- Strategies
- Solutions



Player 2

Player 1

Game theory: basics

Normal form

- Players
- Actions
- Outcomes
- Utilities
- Strategies
- Solutions

Player 2

Player 1

	Rock	Paper	Scissors
Rock			
Paper			
Scissors			

Game theory: basics

Normal form

- Players
- Actions
- Outcomes
- Utilities
- Strategies
- Solutions

		Player 2		
		Rock	**Paper**	**Scissors**
Player 1	**Rock**	Tie	Player 2 wins	Player 1 wins
	Paper	Player 1 wins	Tie	Player 2 wins
	Scissors	Player 2 wins	Player 1 wins	Tie

Game theory: basics

Normal form

- Players
- Actions
- Outcomes
- Utilities
- Strategies
- Solutions

Player 2

Player 1

	Rock	Paper	Scissors
Rock	0,0	-1,1	1,-1
Paper	1,-1	0,0	-1,1
Scissors	-1,1	1,-1	0,0

Game theory: basics

Normal form

- Players
- Actions
- Outcomes
- Utilities
- Strategies
- Solutions

Player 2

		Rock	**Paper**	**Scissors**	
Player 1	**Rock**	0,0	-1,1	1,-1	$\sigma_1(\text{Rock})$
	Paper	1,-1	0,0	-1,1	$\sigma_1(\text{Paper})$
	Scissors	-1,1	1,-1	0,0	$\sigma_1(\text{Scissors})$
		$\sigma_2(\text{Rock})$	$\sigma_2(\text{Paper})$	$\sigma_2(\text{Scissors})$	

Game theory: basics

Normal form

- Players
- Actions
- Outcomes
- Utilities
- Strategies
- Solutions

Player 2

		Rock	Paper	Scissors	
Player 1	**Rock**	0,0	-1,1	1,-1	1/3
	Paper	1,-1	0,0	-1,1	1/3
	Scissors	-1,1	1,-1	0,0	1/3
		1/3	1/3	1/3	

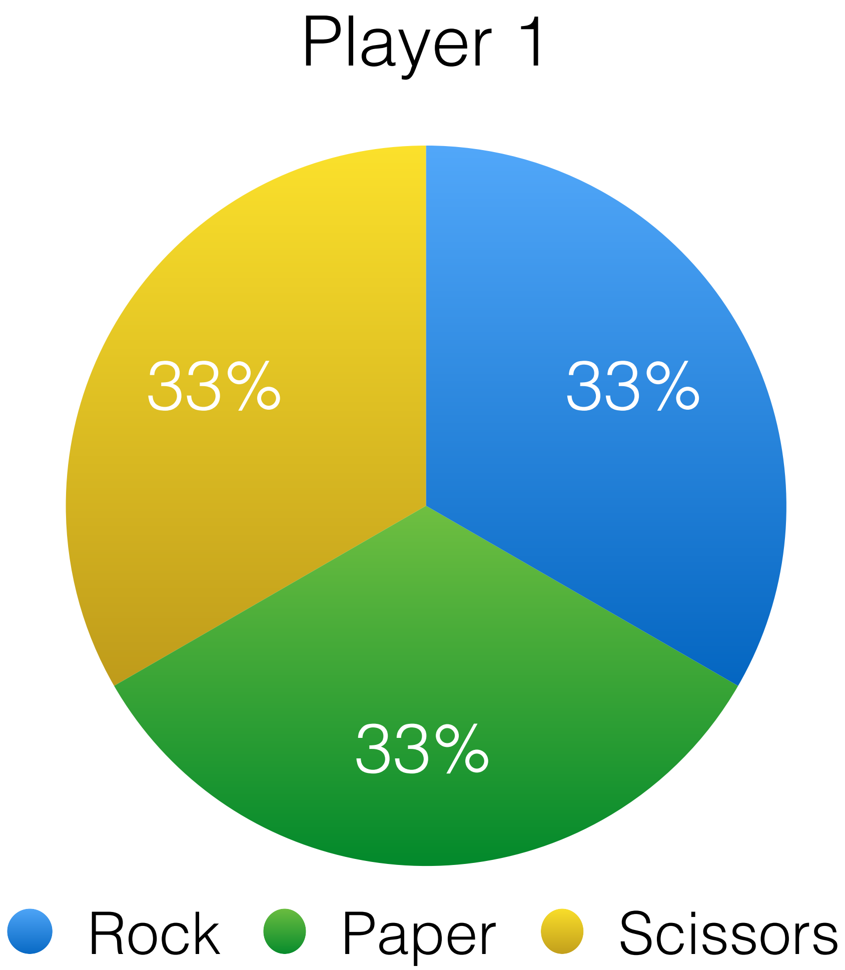
Nash Equilibrium

A strategy profile (σ_1^*, σ_2^*) is a Nash equilibrium if and if:

- $\sigma_1^* \in \arg\max_{\sigma_1} \left\{\sigma_1 U_1 \sigma_2^*\right\}$
- $\sigma_2^* \in \arg\max_{\sigma_2} \left\{\sigma_1^* U_2 \sigma_2\right\}$

Evolutionary dynamics

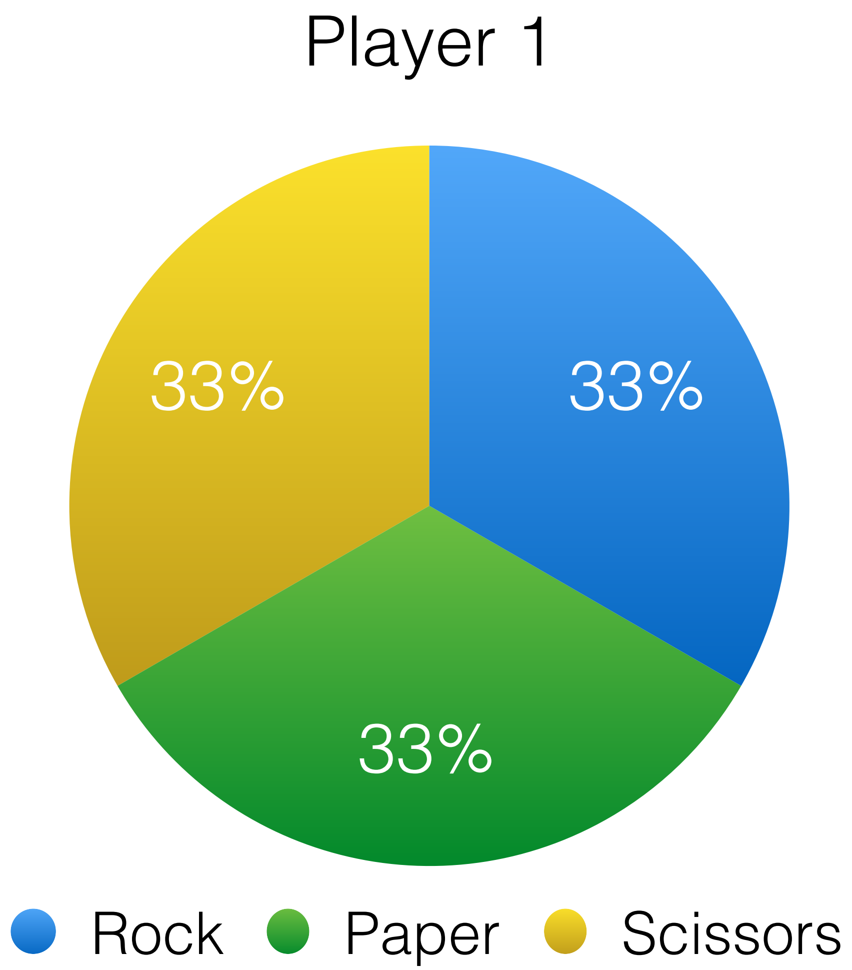
An evolutionary interpretation



Player 1

An evolutionary interpretation

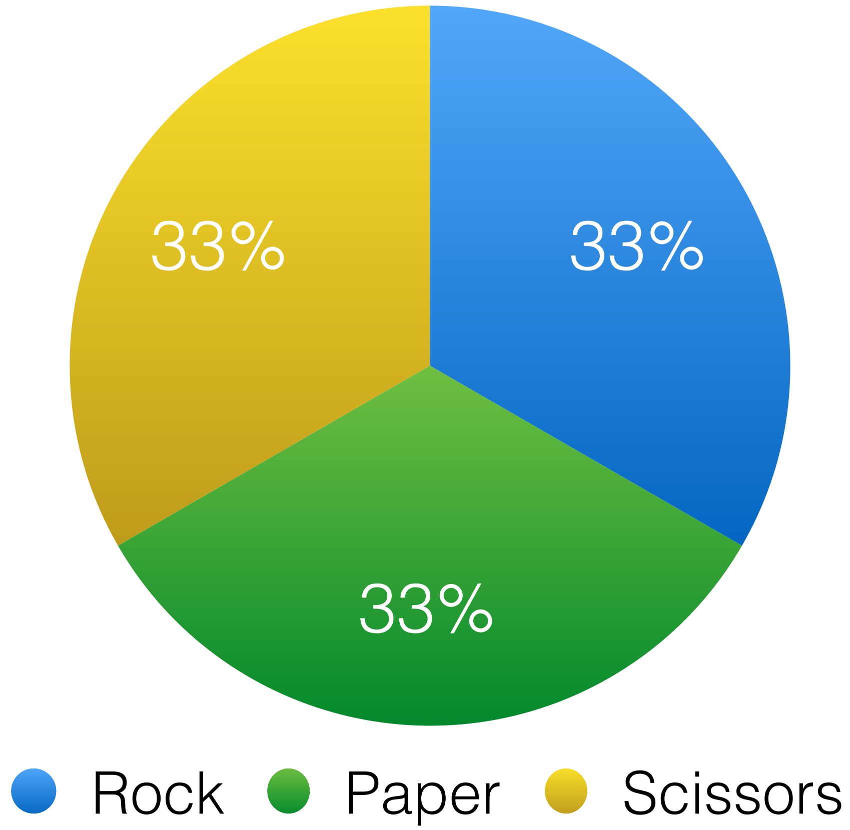
Infinite population of individuals



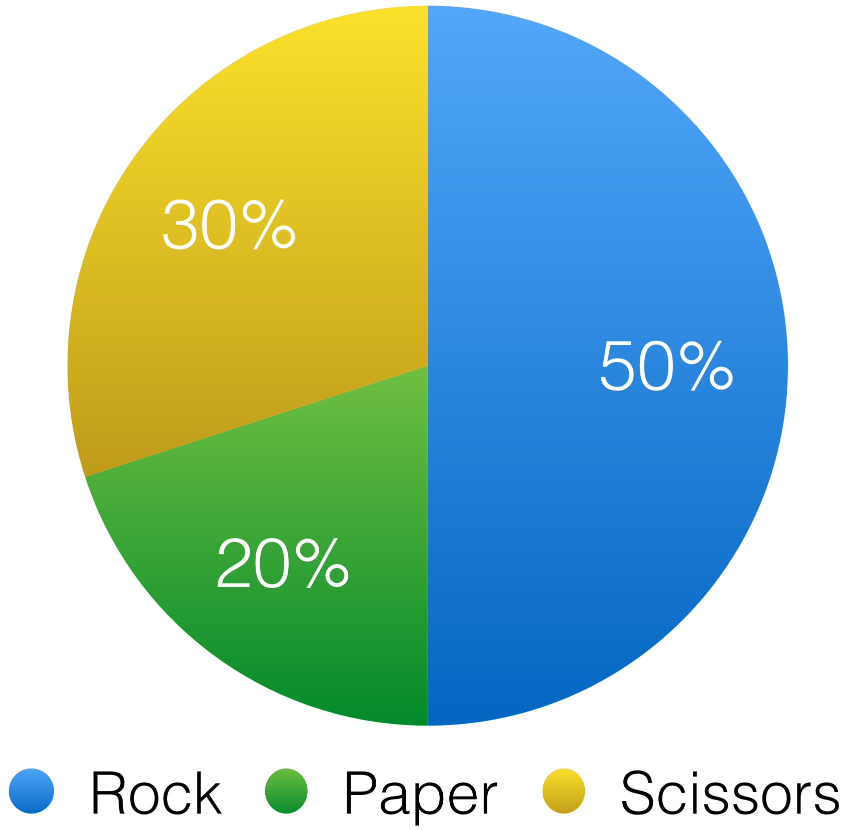
Bacterials

An evolutionary interpretation

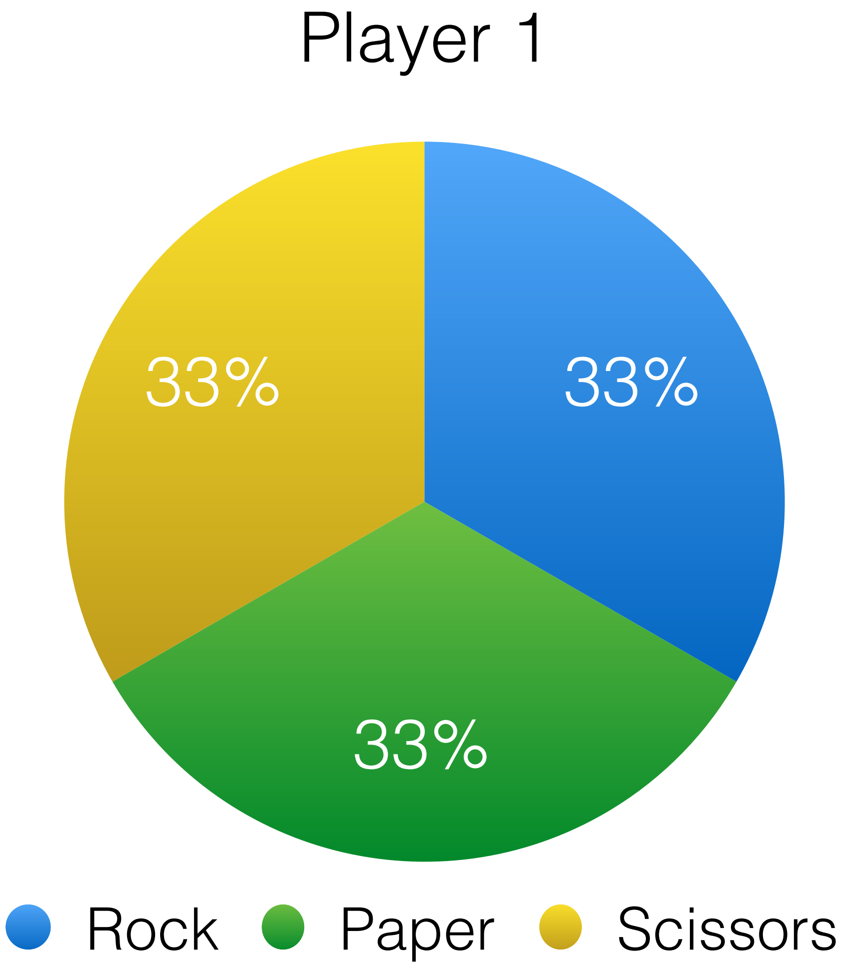
Player 1



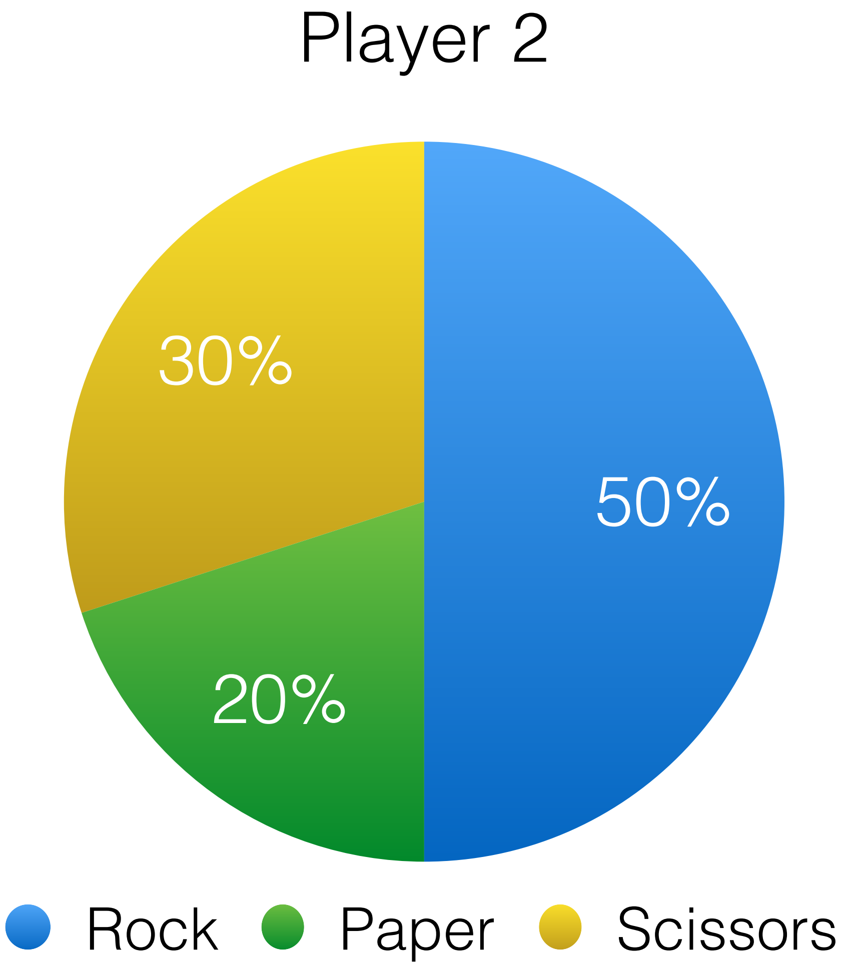
Player 2



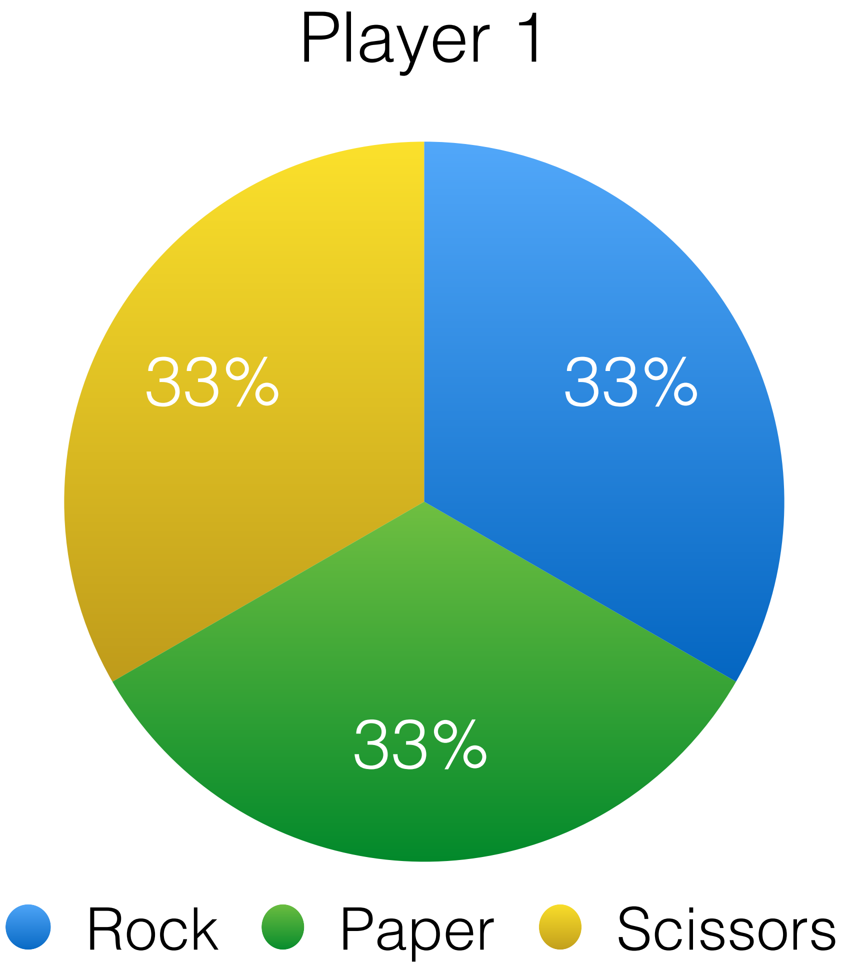
An evolutionary interpretation



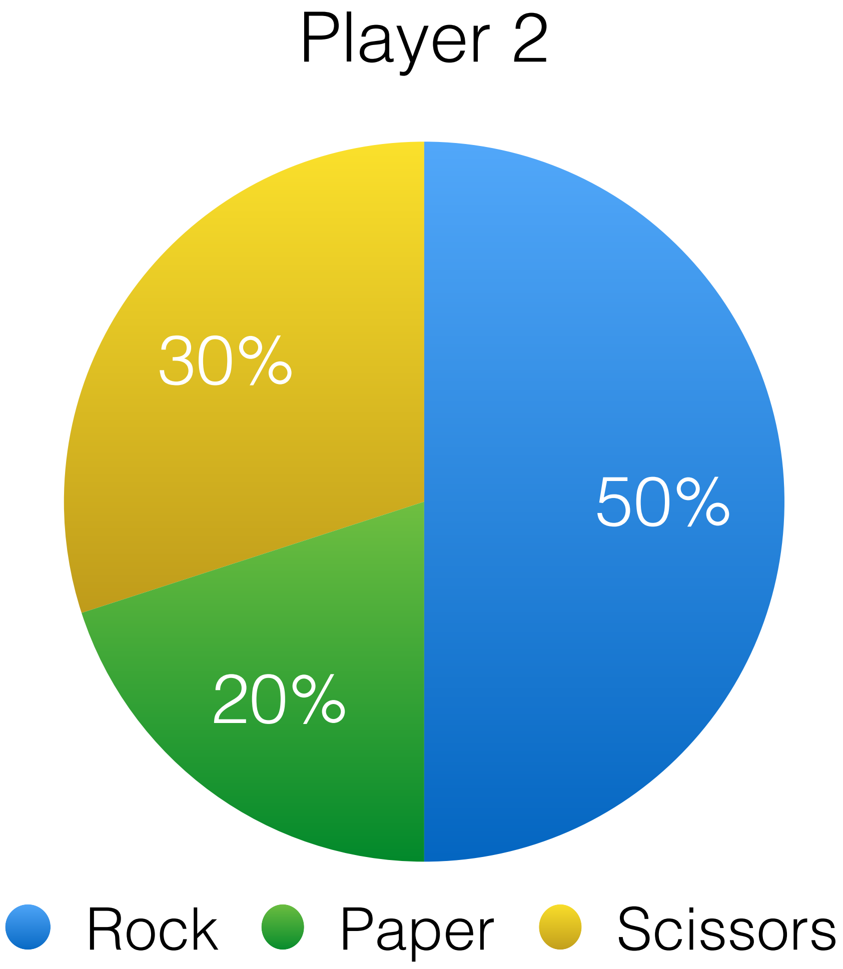
	Rock	Paper	Scissors
Rock	0,0	-1,1	1,-1
Paper	1,-1	0,0	-1,1
Scissors	-1,1	1,-1	0,0



An evolutionary interpretation



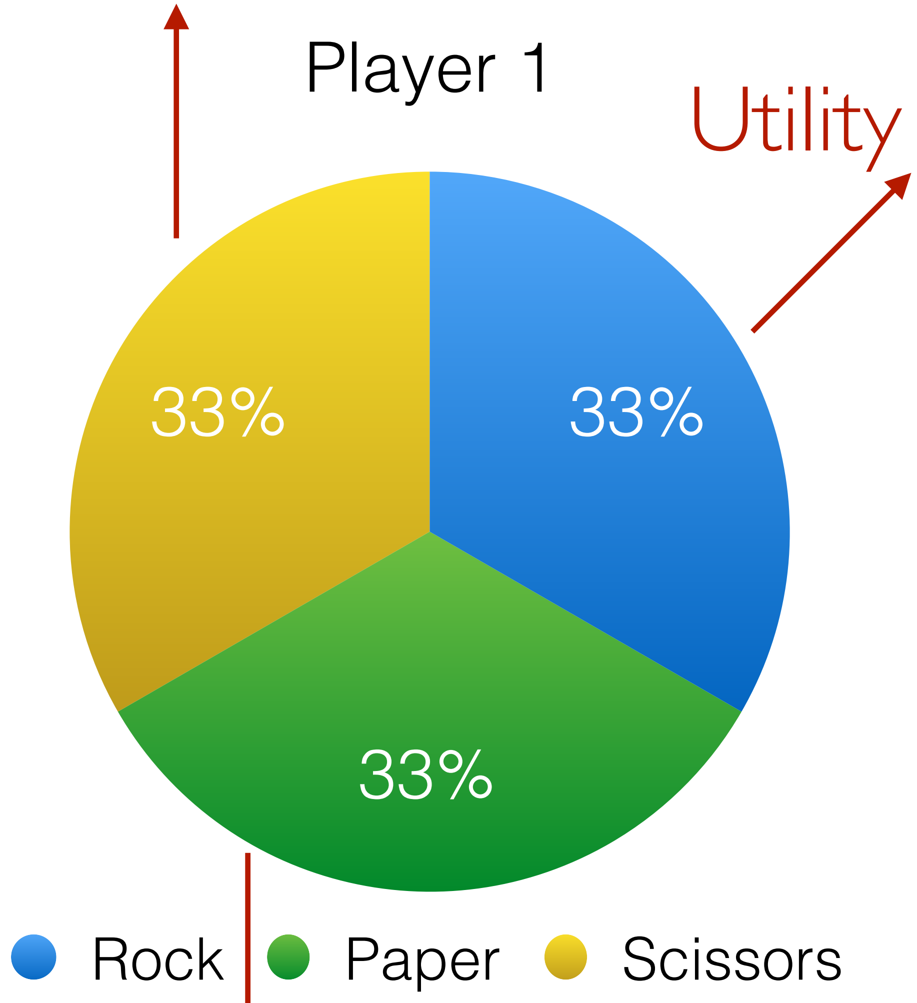
	Rock	Paper	Scissors
Rock	0,0	-1,1	1,-1
Paper	1,-1	0,0	-1,1
Scissors	-1,1	1,-1	0,0



At each round, each individual of a population plays against each individual of the opponent's population

An evolutionary interpretation

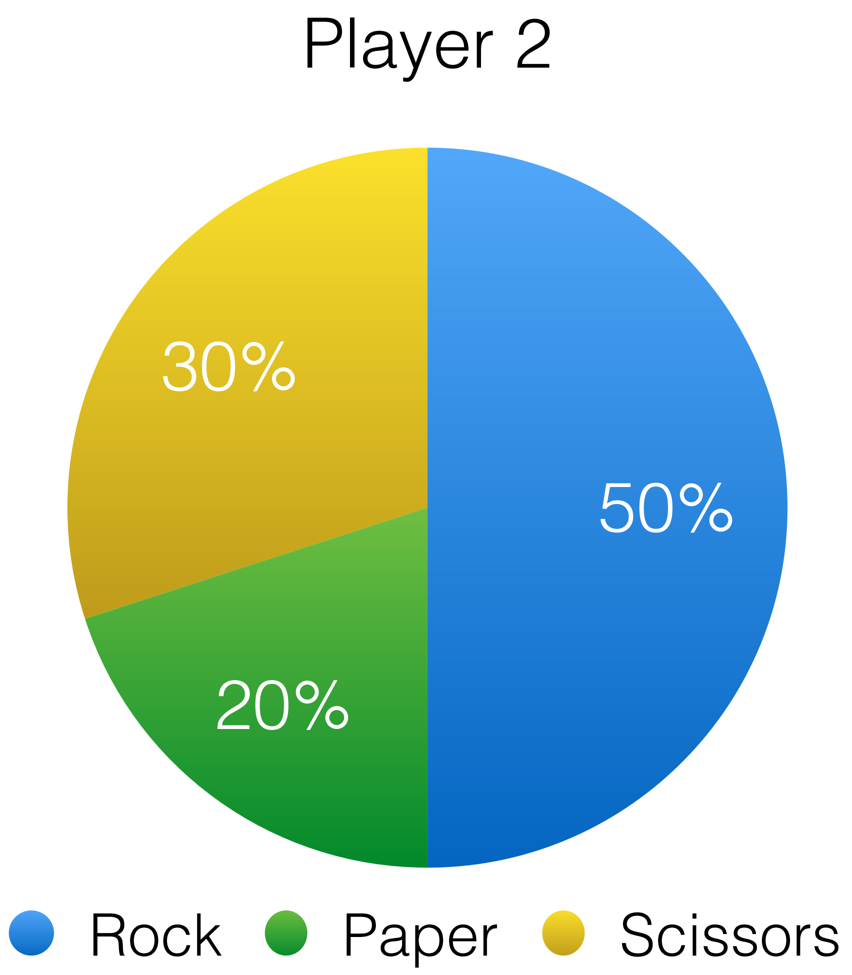
Utility = - 0.3



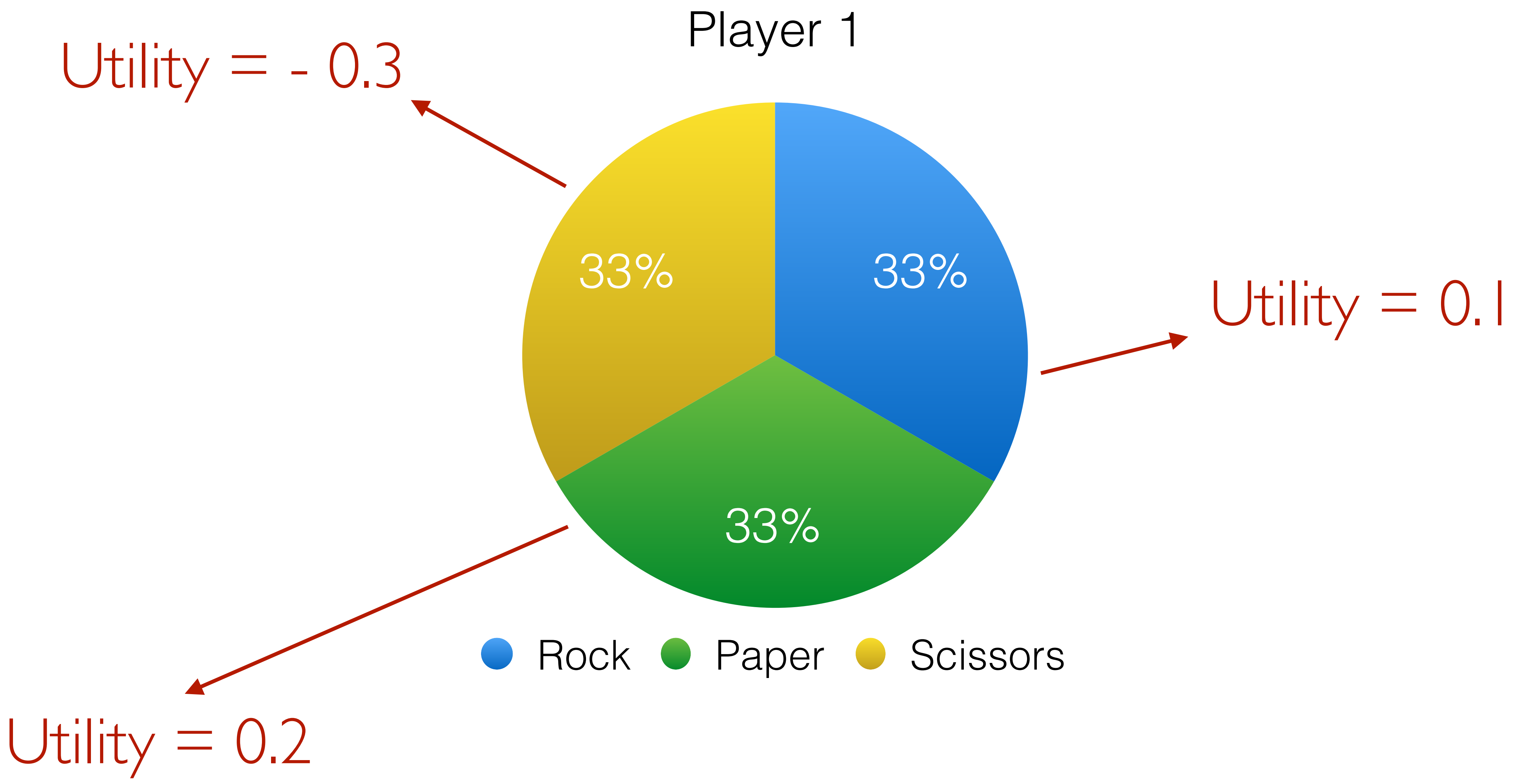
Utility = 0.1

Utility = 0.2

	Rock	Paper	Scissors
Rock	0,0	-1,1	1,-1
Paper	1,-1	0,0	-1,1
Scissors	-1,1	1,-1	0,0



An evolutionary interpretation

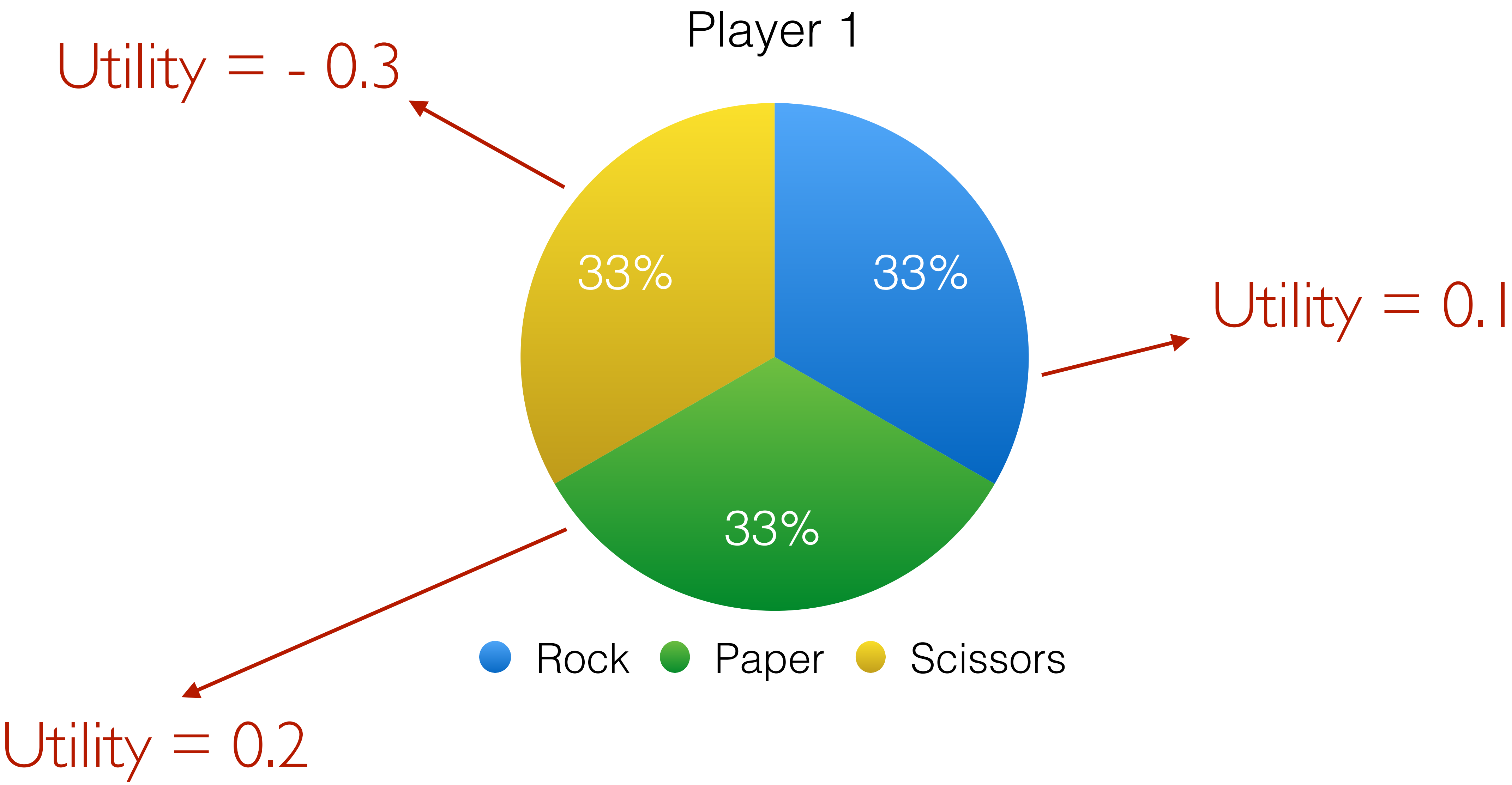


Utility = - 0.3

Utility = 0.1

Utility = 0.2

An evolutionary interpretation



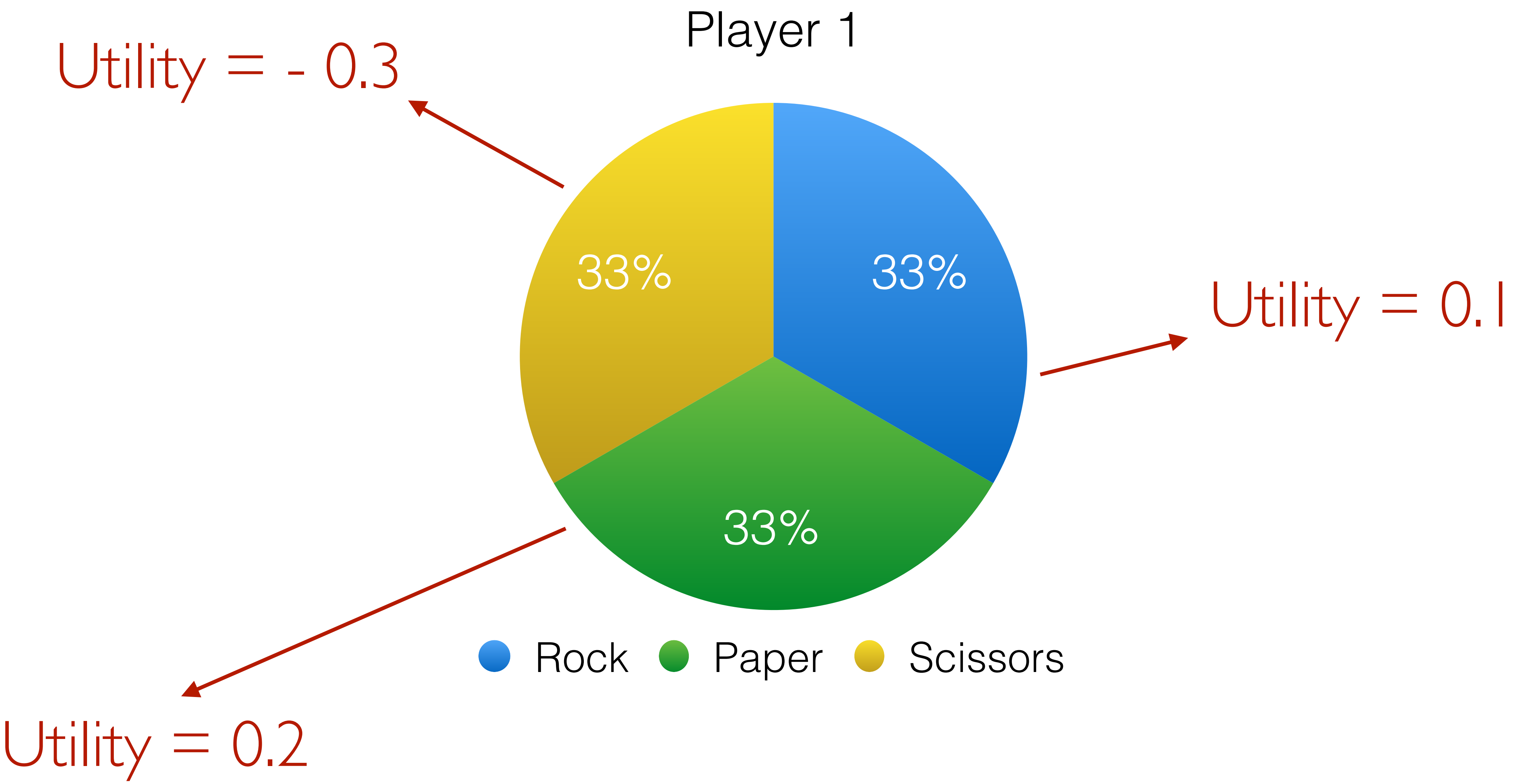
Utility = - 0.3

Utility = 0.1

Utility = 0.2

The average utility is 0

An evolutionary interpretation

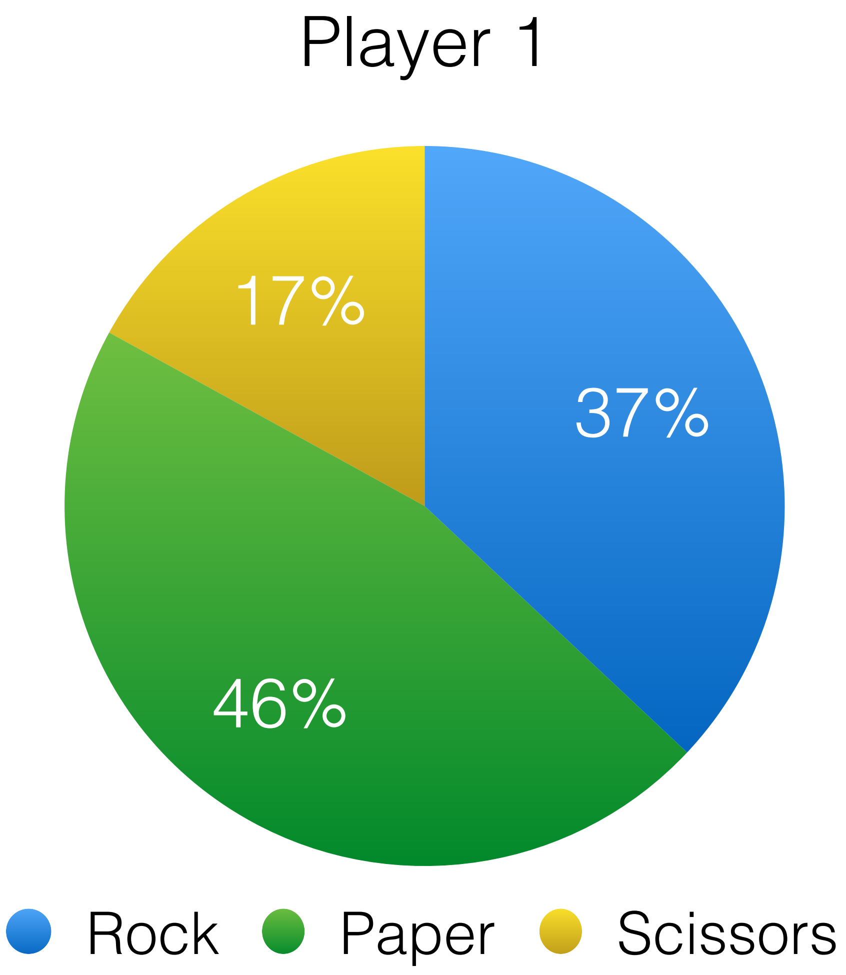


The average utility is 0

A positive (*utility - average utility*) leads to an increase of the population

A negative (*utility - average utility*) leads to a decrease of the population

An evolutionary interpretation



New population after the replication

Revision protocol

Question: how the populations change?

Replicator dynamics

$$\dot{\sigma}_1(a,t) = \sigma_1(a,t)\Big(e_a U_1 \sigma_2(t) - \sigma_1(t) U_1 \sigma_2(t)\Big)$$

Utility given by playing a with a probability of 1

Average population utility

Evolutionary Stable Strategies

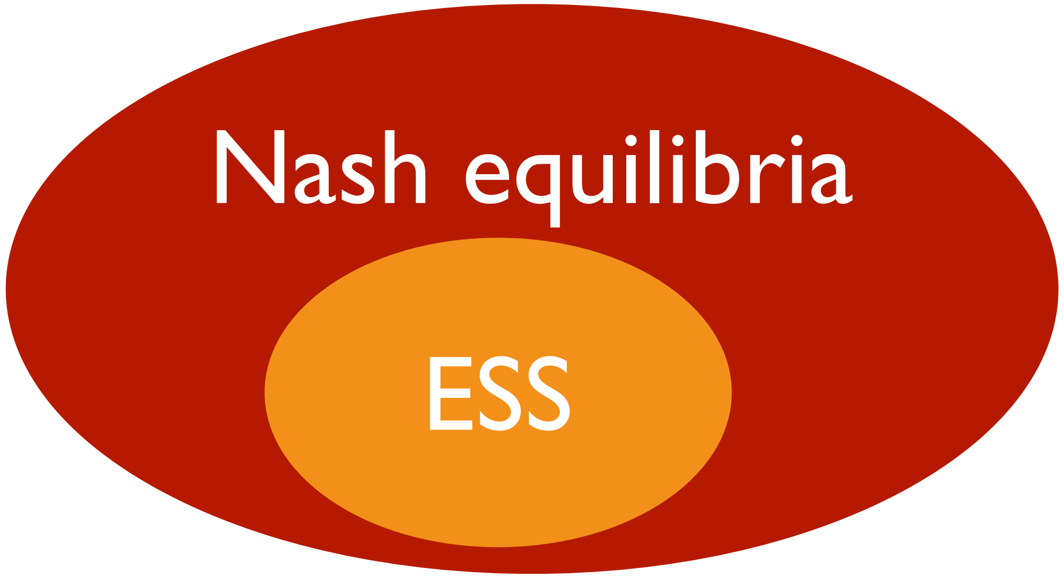
A strategy is an ESS if it is immune to invasion by mutant strategies, given that the mutants initially occupy a small fraction of population

Every ESS is an asymptotically stable fixed point of the replicator dynamics

Evolutionary Stable Strategies

A strategy is an ESS if it is immune to invasion by mutant strategies, given that the mutants initially occupy a small fraction of population

Every ESS is an asymptotically stable fixed point of the replicator dynamics



While a NE always exists,
an ESS may not exist

Prisoner’s dilemma

Player 2

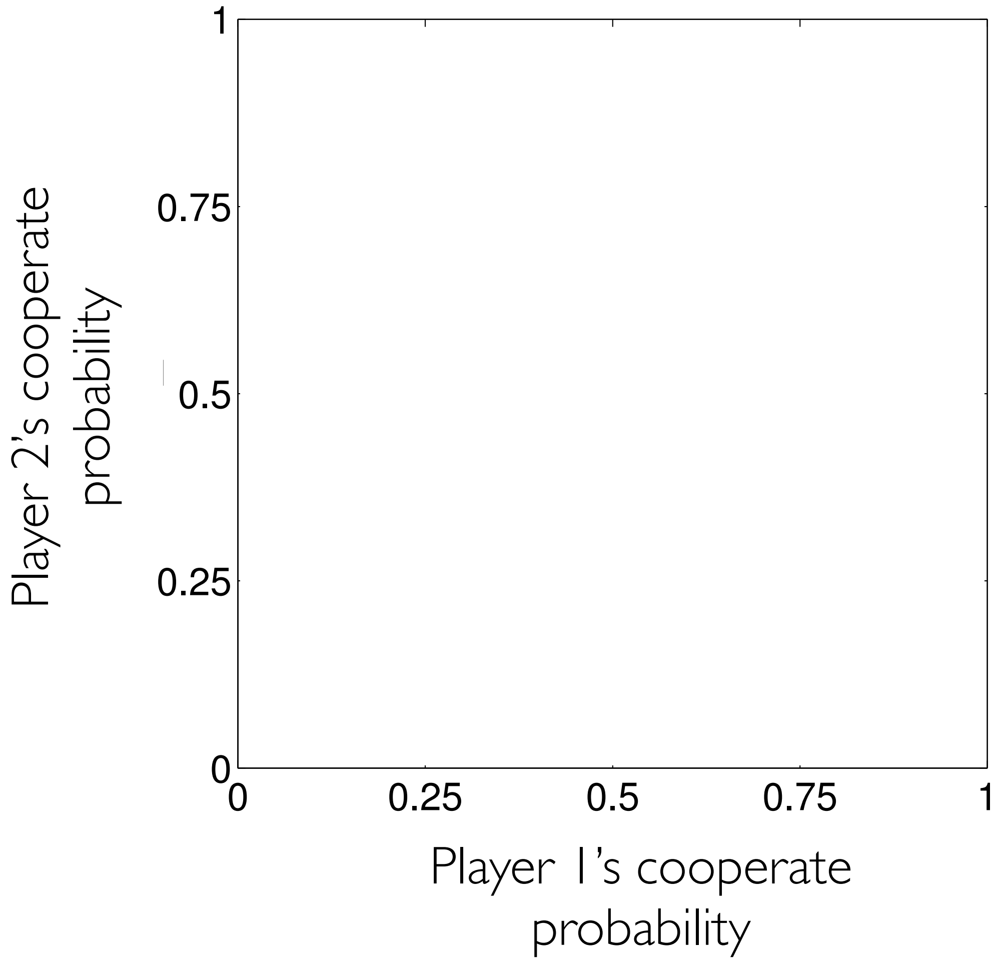
		Cooperate	**Defeat**
Player 1	**Cooperate**	3,3	0,5
	Defeat	5,0	1,1

Prisoner’s dilemma

		Player 2	
		Cooperate	**Defeat**
Player 1	**Cooperate**	3,3	0,5
	Defeat	5,0	● 1,1

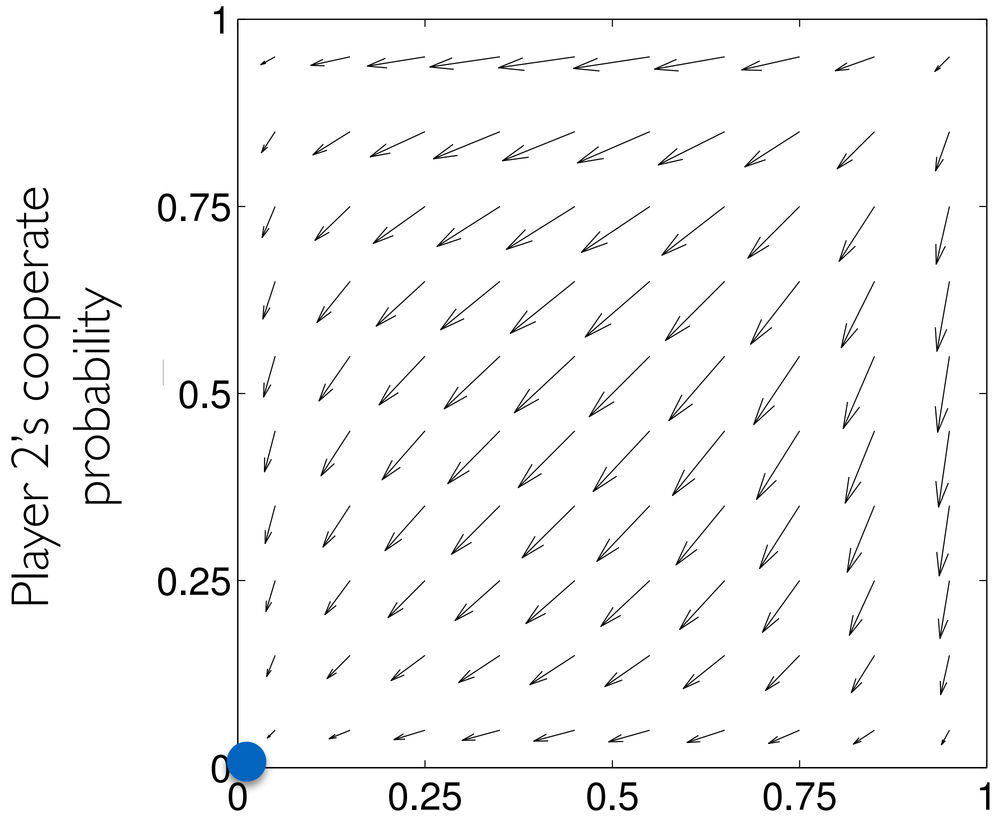
Prisoner’s dilemma

		Player 2	
		Cooperate	**Defeat**
Player 1	**Cooperate**	3,3	0,5
	Defeat	5,0	● 1,1



Prisoner's dilemma

		Player 2	
		Cooperate	Defeat
Player 1	Cooperate	3,3	0,5
	Defeat	5,0	● 1,1



asymptotically stable

Player 1's cooperate probability

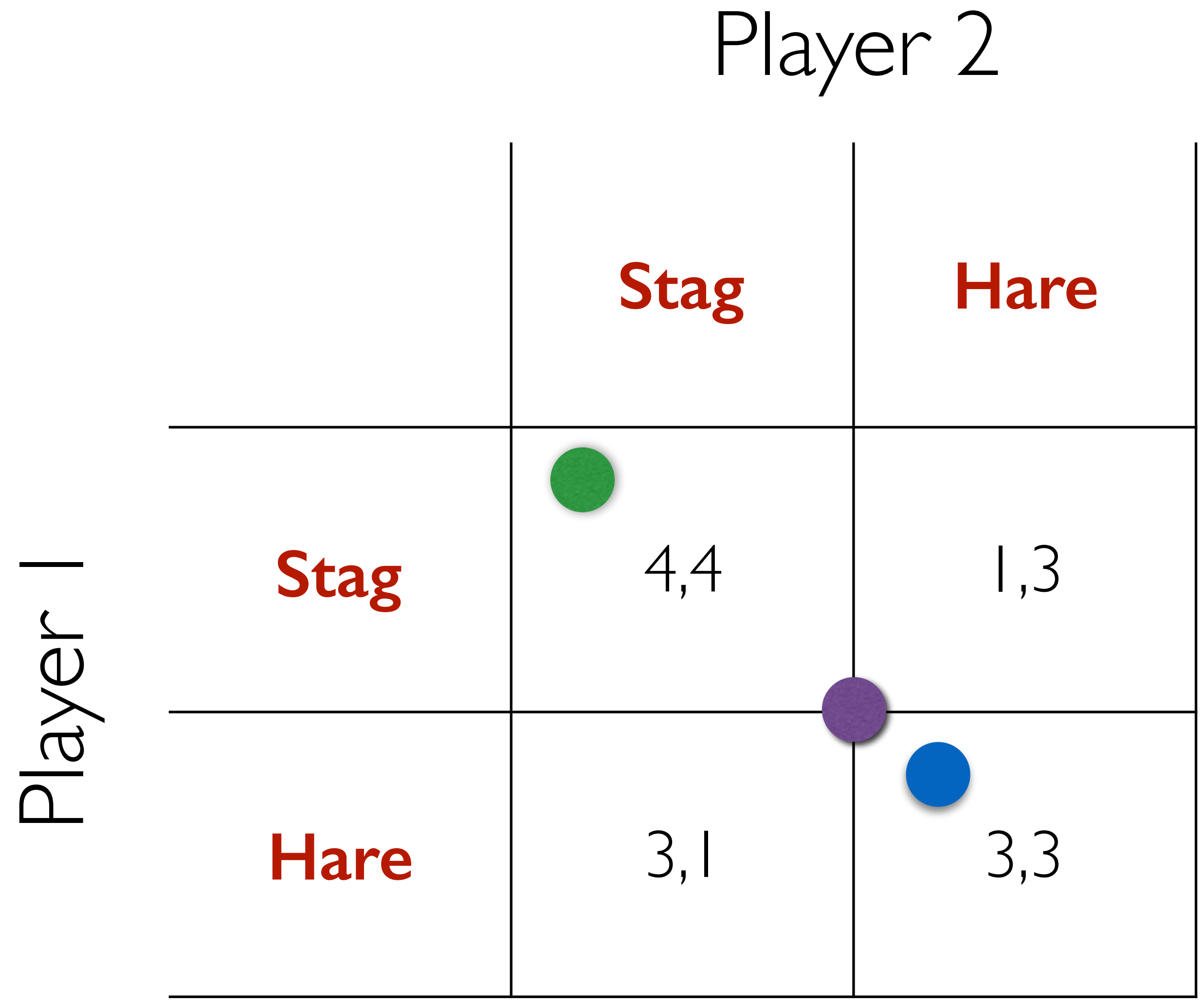
Stag hunt

Player 2

Player 1

	Stag	Hare
Stag	4,4	1,3
Hare	3,1	3,3

Stag hunt



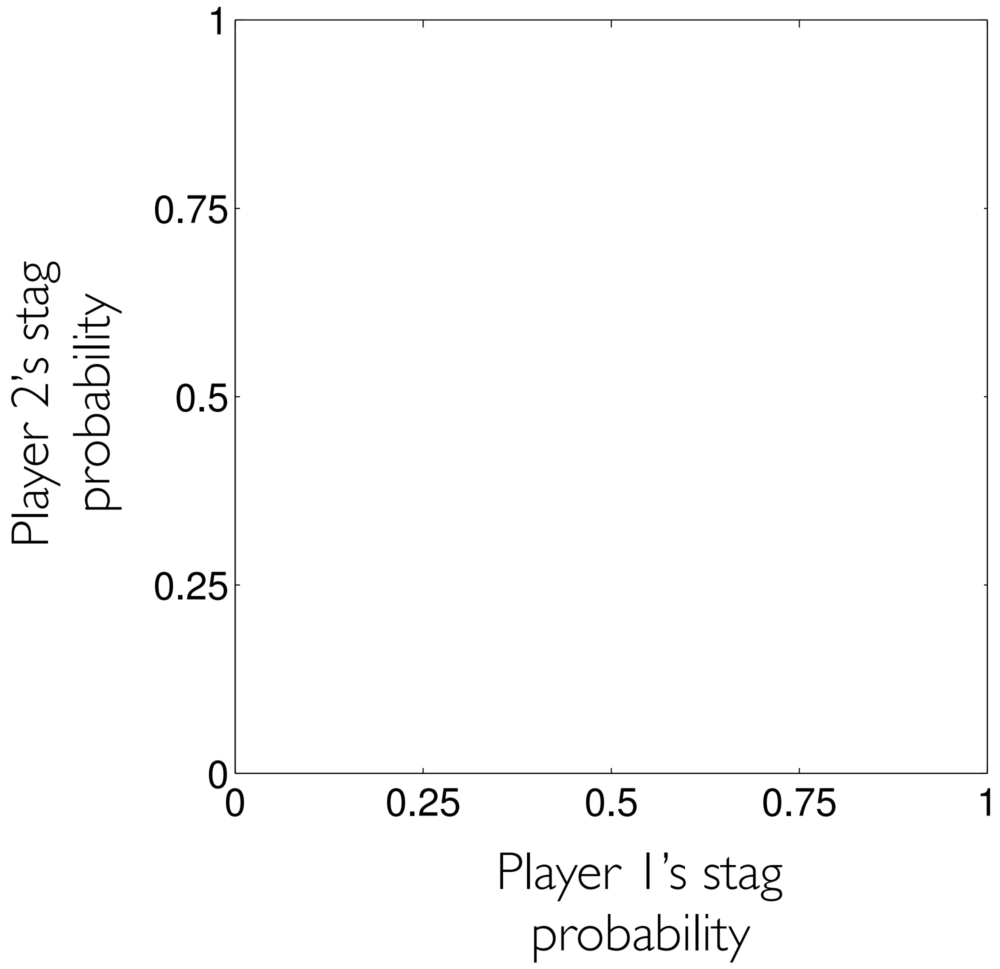
Player 2

Player 1

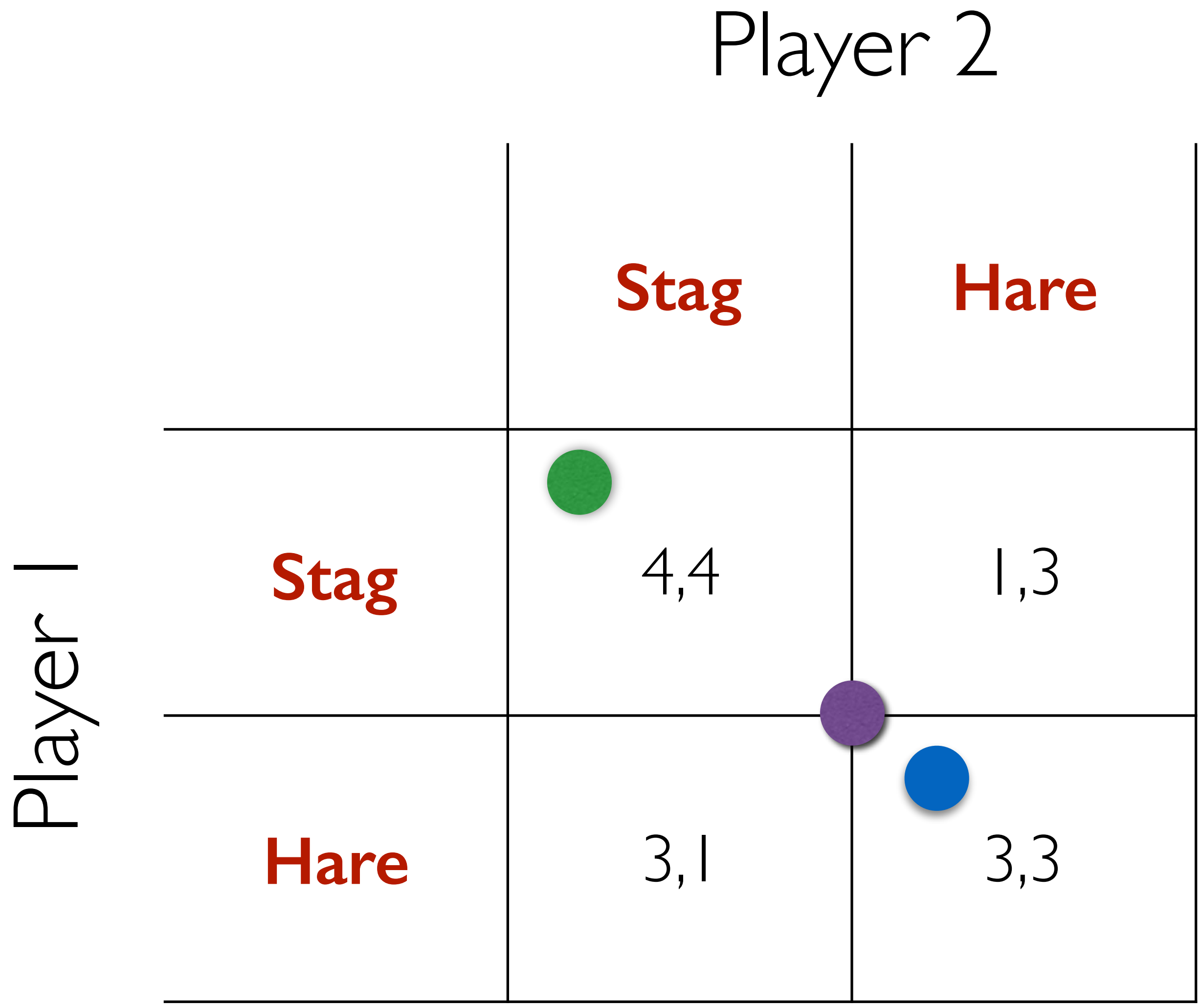
	Stag	Hare
Stag	4,4	1,3
Hare	3,1	3,3

Stag hunt

		Player 2	
		Stag	**Hare**
Player 1	**Stag**	4,4	1,3
	Hare	3,1	3,3

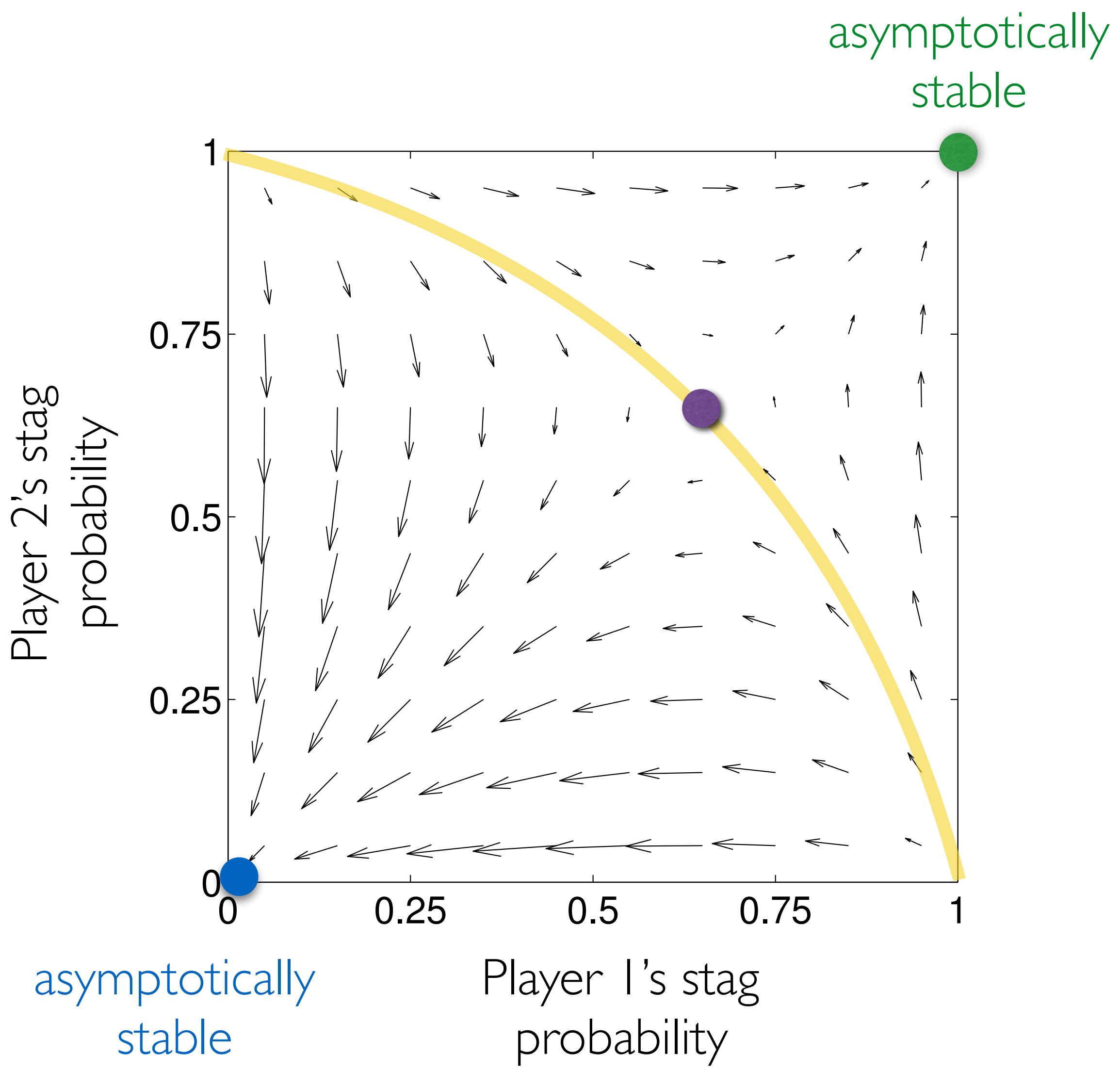


Stag hunt



Player 2

Player 1	Stag	Hare
Stag	4,4	1,3
Hare	3,1	3,3



Matching pennies

		Player 2	
		Head	**Tail**
Player 1	**Head**	0,1	1,0
	Tail	1,0	0,1

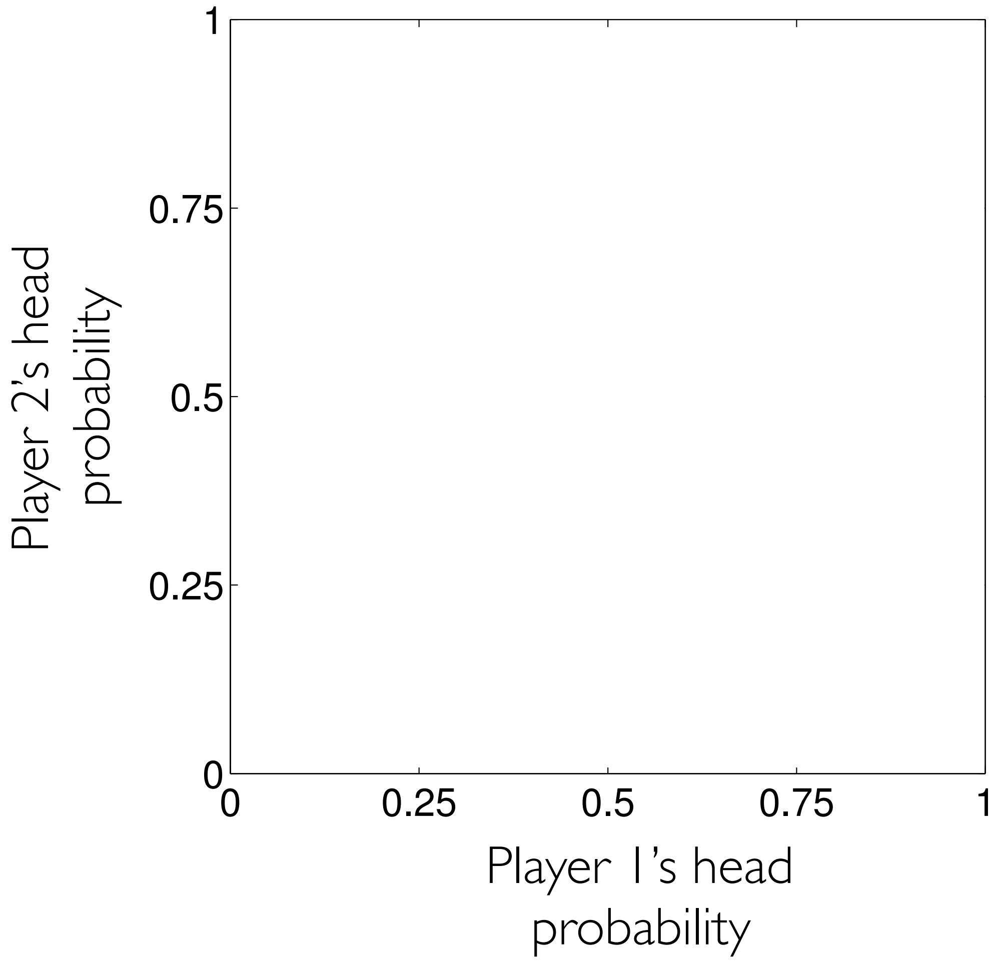
Matching pennies

		Player 2	
		Head	**Tail**
Player 1	**Head**	0,1	1,0
	Tail	1,0	0,1

Matching pennies

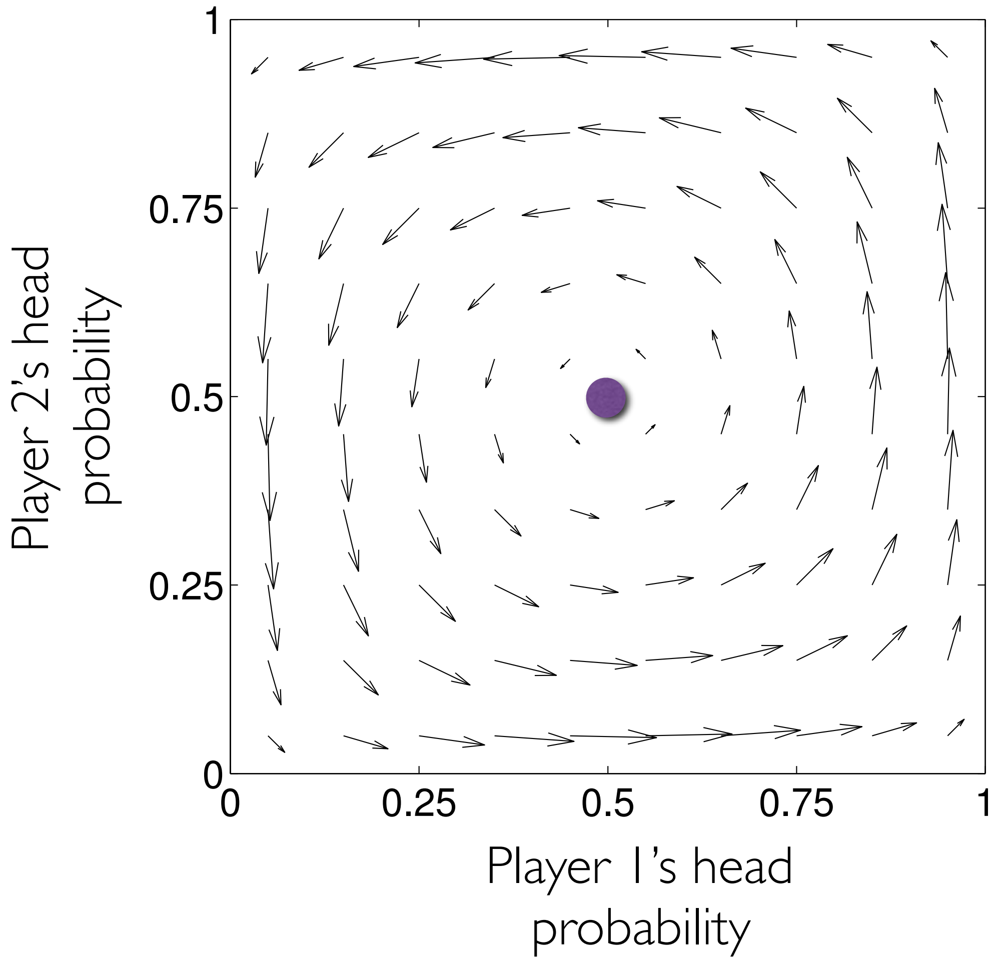
Player 2

		Head	Tail
Player 1	Head	0,1	1,0
	Tail	1,0	0,1



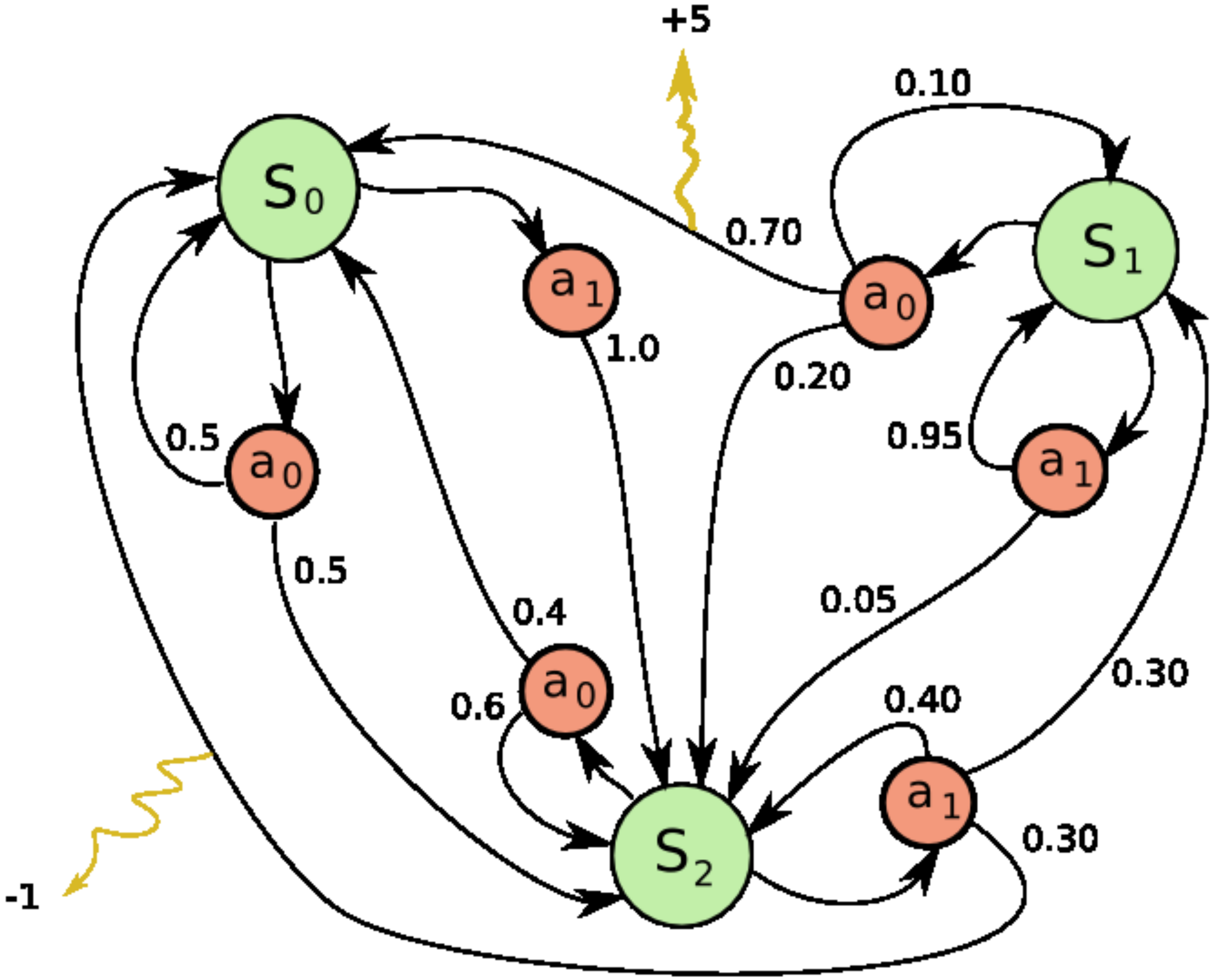
Matching pennies

		Player 2	
		Head	**Tail**
Player 1	**Head**	0,1	1,0
	Tail	1,0	0,1

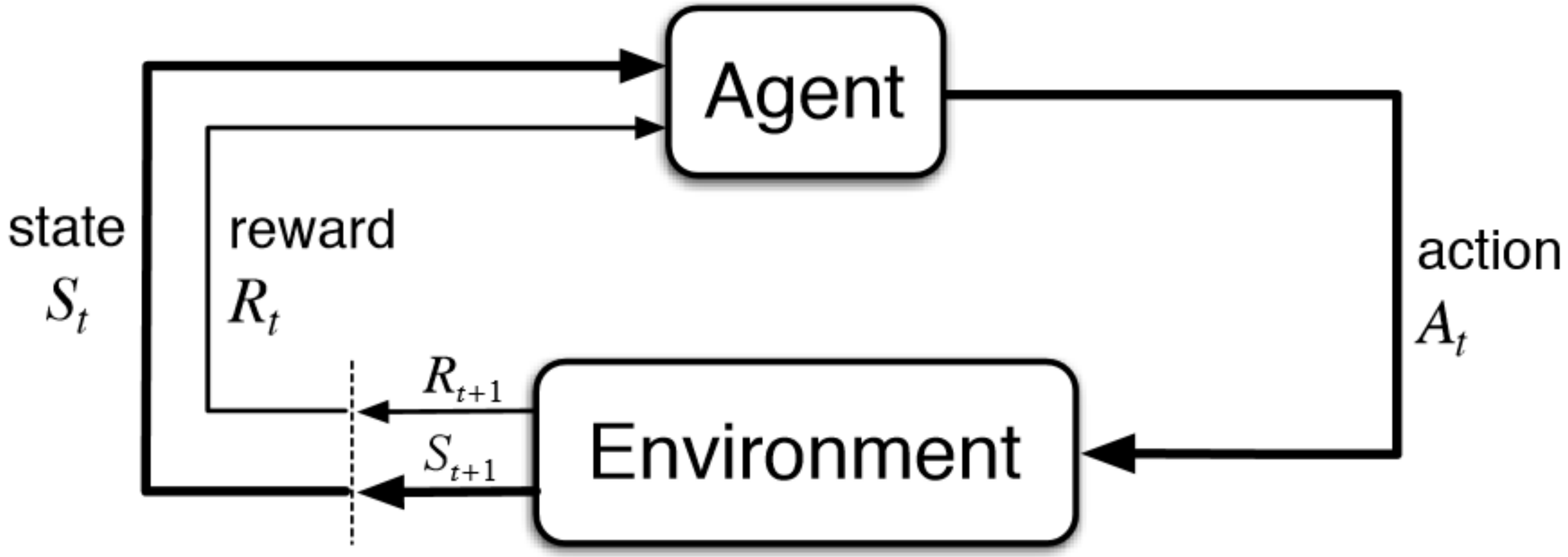


Multi-agent learning

Markov decision problem



Reinforcement learning



Q-learning (1)

For every pair state/action:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s,a') - Q(s,a) \right]$$

learning rate reward discount factor

Example: normal-form games

Player 1

- 1 state
- 2 actions (Cooperate, Defeat)

		Player 2	
		Cooperate	**Defeat**
Player 1	**Cooperate**	3,3	0,5
	Defeat	5,0	1,1

Example: normal-form games

$$\sigma_1(a) = \begin{cases} 1.0 & a = \text{Cooperate} \\ 0.0 & a = \text{Defeat} \end{cases}$$

$$Q(a) \leftarrow Q(a) + \alpha \Big(r - Q(a) \Big) \qquad \alpha = 0.2$$

$$\sigma_1(a) = \begin{cases} 0.2 & a = \text{Cooperate} \\ 0.8 & a = \text{Defeat} \end{cases}$$

Player 2

		Cooperate	**Defeat**
Player 1	**Cooperate**	3,3	0,5
	Defeat	5,0	1,1

Example: normal-form games

$$\sigma_1(a) = \begin{cases} 1.0 & a = \text{Cooperate} \\ 0.0 & a = \text{Defeat} \end{cases}$$

$$\sigma_1(a) = \begin{cases} 0.2 & a = \text{Cooperate} \\ 0.8 & a = \text{Defeat} \end{cases}$$

$$Q(a) \leftarrow Q(a) + \alpha \Big(r - Q(a) \Big) \qquad \alpha = 0.2$$

round	Player 2's action	Player 1's Q function
$t = 0$	—	$Q(\text{Cooperate}) = 0$
$t = 1$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.6$
$t = 2$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.48$
$t = 3$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.384$
$t = 4$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.3072$
$t = 5$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.24576$
$t = 6$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.496608$

Player 2

Player 1	**Cooperate**	**Defeat**
Cooperate	3,3	0,5
Defeat	5,0	1,1

Example: normal-form games

$$Q(a) \leftarrow Q(a) + \alpha \Big(r - Q(a) \Big) \qquad \alpha = 0.2$$

round	Player 2's action	Player 1's Q function
$t = 0$	—	$Q(\text{Cooperate}) = 0$
$t = 1$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.6$
$t = 2$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.48$
$t = 3$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.384$
$t = 4$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.3072$
$t = 5$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.24576$
$t = 6$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.496608$

$$\sigma_1(a) = \begin{cases} 1.0 & a = \text{Cooperate} \\ 0.0 & a = \text{Defeat} \end{cases}$$

$$\sigma_1(a) = \begin{cases} 0.2 & a = \text{Cooperate} \\ 0.8 & a = \text{Defeat} \end{cases}$$

Player 2

		Cooperate	Defeat
Player 1	Cooperate	3,3	0,5
	Defeat	5,0	1,1

Example: normal-form games

$$\sigma_1(a) = \begin{cases} 1.0 & a = \text{Cooperate} \\ 0.0 & a = \text{Defeat} \end{cases}$$

$$Q(a) \leftarrow Q(a) + \alpha \Big(r - Q(a) \Big) \qquad \alpha = 0.2$$

$$\sigma_1(a) = \begin{cases} 0.2 & a = \text{Cooperate} \\ 0.8 & a = \text{Defeat} \end{cases}$$

round	Player 2's action	Player 1's Q function
$t = 0$	—	$Q(\text{Cooperate}) = 0$
$t = 1$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.6$
$t = 2$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.48$
$t = 3$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.384$
$t = 4$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.3072$
$t = 5$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.24576$
$t = 6$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.496608$

Player 2

		Cooperate	**Defeat**
Player 1	**Cooperate**	3,3	0,5
	Defeat	5,0	1,1

Example: normal-form games

$$\sigma_1(a) = \begin{cases} 1.0 & a = \text{Cooperate} \\ 0.0 & a = \text{Defeat} \end{cases}$$

$$Q(a) \leftarrow Q(a) + \alpha \left(r - Q(a) \right) \qquad \alpha = 0.2$$

$$\sigma_1(a) = \begin{cases} 0.2 & a = \text{Cooperate} \\ 0.8 & a = \text{Defeat} \end{cases}$$

round	Player 2's action	Player 1's Q function
$t = 0$	—	$Q(\text{Cooperate}) = 0$
$t = 1$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.6$
$t = 2$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.48$
$t = 3$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.384$
$t = 4$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.3072$
$t = 5$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.24576$
$t = 6$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.496608$

		Player 2	
		Cooperate	**Defeat**
Player 1	**Cooperate**	3,3	0,5
	Defeat	5,0	1,1

Example: normal-form games

$$\sigma_1(a) = \begin{cases} 1.0 & a = \text{Cooperate} \\ 0.0 & a = \text{Defeat} \end{cases}$$

$$Q(a) \leftarrow Q(a) + \alpha\Big(r - Q(a)\Big) \qquad \alpha = 0.2$$

$$\sigma_1(a) = \begin{cases} 0.2 & a = \text{Cooperate} \\ 0.8 & a = \text{Defeat} \end{cases}$$

round	Player 2's action	Player 1's Q function
$t = 0$	—	$Q(\text{Cooperate}) = 0$
$t = 1$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.6$
$t = 2$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.48$
$t = 3$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.384$
$t = 4$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.3072$
$t = 5$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.24576$
$t = 6$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.496608$

		Player 2	
		Cooperate	**Defeat**
Player 1	**Cooperate**	3,3	0,5
	Defeat	5,0	1,1

Example: normal-form games

$$Q(a) \leftarrow Q(a) + \alpha \Big(r - Q(a) \Big) \qquad \alpha = 0.2$$

round	Player 2's action	Player 1's Q function
$t = 0$	—	$Q(\text{Cooperate}) = 0$
$t = 1$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.6$
$t = 2$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.48$
$t = 3$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.384$
$t = 4$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.3072$
$t = 5$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.24576$
$t = 6$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.496608$

$$\sigma_1(a) = \begin{cases} 1.0 & a = \text{Cooperate} \\ 0.0 & a = \text{Defeat} \end{cases}$$

$$\sigma_1(a) = \begin{cases} 0.2 & a = \text{Cooperate} \\ 0.8 & a = \text{Defeat} \end{cases}$$

Player 2

Player 1	**Cooperate**	**Defeat**
Cooperate	3,3	0,5
Defeat	5,0	1,1

Example: normal-form games

$$\sigma_1(a) = \begin{cases} 1.0 & a = \text{Cooperate} \\ 0.0 & a = \text{Defeat} \end{cases}$$

$$\sigma_1(a) = \begin{cases} 0.2 & a = \text{Cooperate} \\ 0.8 & a = \text{Defeat} \end{cases}$$

$$Q(a) \leftarrow Q(a) + \alpha \left(r - Q(a) \right) \qquad \alpha = 0.2$$

round	Player 2's action	Player 1's Q function
$t = 0$	—	$Q(\text{Cooperate}) = 0$
$t = 1$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.6$
$t = 2$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.48$
$t = 3$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.384$
$t = 4$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.3072$
$t = 5$	$a = \text{Defeat}$	$Q(\text{Cooperate}) = 0.24576$
$t = 6$	$a = \text{Cooperate}$	$Q(\text{Cooperate}) = 0.496608$

Player 2

Player 1	**Cooperate**	**Defeat**
Cooperate	3,3	0,5
Defeat	5,0	1,1

Q-learning (2)

Softmax (a.k.a. Boltzam exploration)

$$\sigma_i(a) = \frac{\exp(Q(s,a)/\tau)}{\sum_{a'} \exp(Q(s,a')/\tau)}$$

Q-learning (2)

Softmax (a.k.a. Boltzam exploration)

$$\sigma_i(a) = \frac{\exp(Q(s,a)/\tau)}{\sum_{a'} \exp(Q(s,a')/\tau)}$$

temperature

Q-learning (2)

Softmax (a.k.a. Boltzam exploration)

$$\sigma_i(a) = \frac{\exp(Q(s,a)/\tau)}{\sum_{a'} \exp(Q(s,a')/\tau)}$$

temperature

Every action is played with strictly positive probability

The larger the temperature, the smoother the function

If the temperature is 0, we would have a best response

Example: normal-form games

Q(Cooperate)	Q(Defeat)	σ_1(Cooperate)	σ_1(Cooperate)
0	0	0.5	0.5
1	0	0.731	0.269
5	0	0.99331	0.00669
10	0	0.999955	0.000045

Example: normal-form games

$Q(\text{Cooperate})$	$Q(\text{Defeat})$	$\sigma_1(\text{Cooperate})$	$\sigma_1(\text{Cooperate})$
0	0	0.5	0.5
1	0	0.731	0.269
5	0	0.99331	0.00669
10	0	0.999955	0.000045

Example: normal-form games

Q(Cooperate)	Q(Defeat)	σ_1(Cooperate)	σ_1(Cooperate)
0	0	0.5	0.5
1	0	0.731	0.269
5	0	0.99331	0.00669
10	0	0.999955	0.000045

Example: normal-form games

$Q(\text{Cooperate})$	$Q(\text{Defeat})$	$\sigma_1(\text{Cooperate})$	$\sigma_1(\text{Cooperate})$
0	0	0.5	0.5
1	0	0.731	0.269
5	0	0.99331	0.00669
10	0	0.999955	0.000045

Self-play Q-learning dynamics

Self-play learning

Q-learning algorithm

Player 2

Q-learning algorithm

Player 1

	Cooperate	**Defeat**
Cooperate	3,3	0,5
Defeat	5,0	1,1

Learning dynamics

Assumptions:

- Time is continuous
- All the actions can be selected simultaneously

Learning dynamics

Assumptions:

- Time is continuous
- All the actions can be selected simultaneously

$$\dot{\sigma}_1(a,t) = \frac{\alpha\,\sigma_1(a,t)}{\tau}\Big(e_a U_1 \sigma_2(t) - \sigma_1(t) U_1 \sigma_2(t)\Big) - \alpha\,\sigma_1(a,t)\,\Big(\log(\sigma_1(a)) - \sum_{a'} \sigma_1(a')\,\log(\sigma_1(a'))\Big)$$

Learning dynamics

Assumptions:

- Time is continuous
- All the actions can be selected simultaneously

$$\dot{\sigma}_1(a,t) = \frac{\alpha\,\sigma_1(a,t)}{\tau}\underbrace{\boxed{\Big(e_a U_1 \sigma_2(t) - \sigma_1(t) U_1 \sigma_2(t)\Big)}}_{\text{exploitation term}} - \alpha\,\sigma_1(a,t)\underbrace{\boxed{\Big(\log(\sigma_1(a)) - \sum_{a'} \sigma_1(a')\,\log(\sigma_1(a'))\Big)}}_{\text{exploration term}}$$

Learning dynamics

Assumptions:

- Time is continuous
- All the actions can be selected simultaneously

$$\dot{\sigma}_1(a,t) = \frac{\alpha\,\sigma_1(a,t)}{\tau}\underbrace{\boxed{\Big(e_a U_1 \sigma_2(t) - \sigma_1(t) U_1 \sigma_2(t)\Big)}}_{\text{exploitation term}} - \alpha\,\sigma_1(a,t)\underbrace{\boxed{\Big(\log(\sigma_1(a)) - \sum_{a'} \sigma_1(a')\,\log(\sigma_1(a'))\Big)}}_{\text{exploration term}}$$

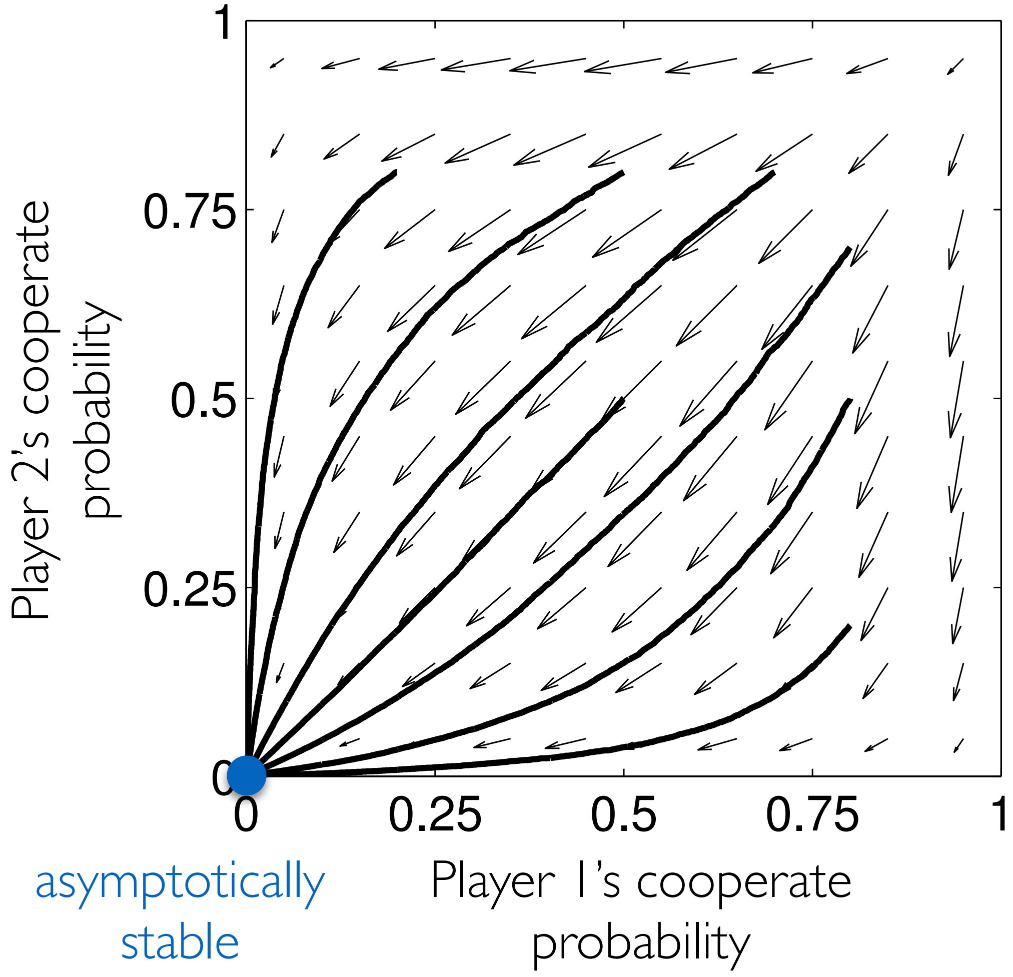
When the temperature is 0, the Q-learning behaves as the replicator dynamics

Prisoner’s dilemma

Player 2

Player 1

	Cooperate	Defeat
Cooperate	3,3	0,5
Defeat	5,0	● 1,1



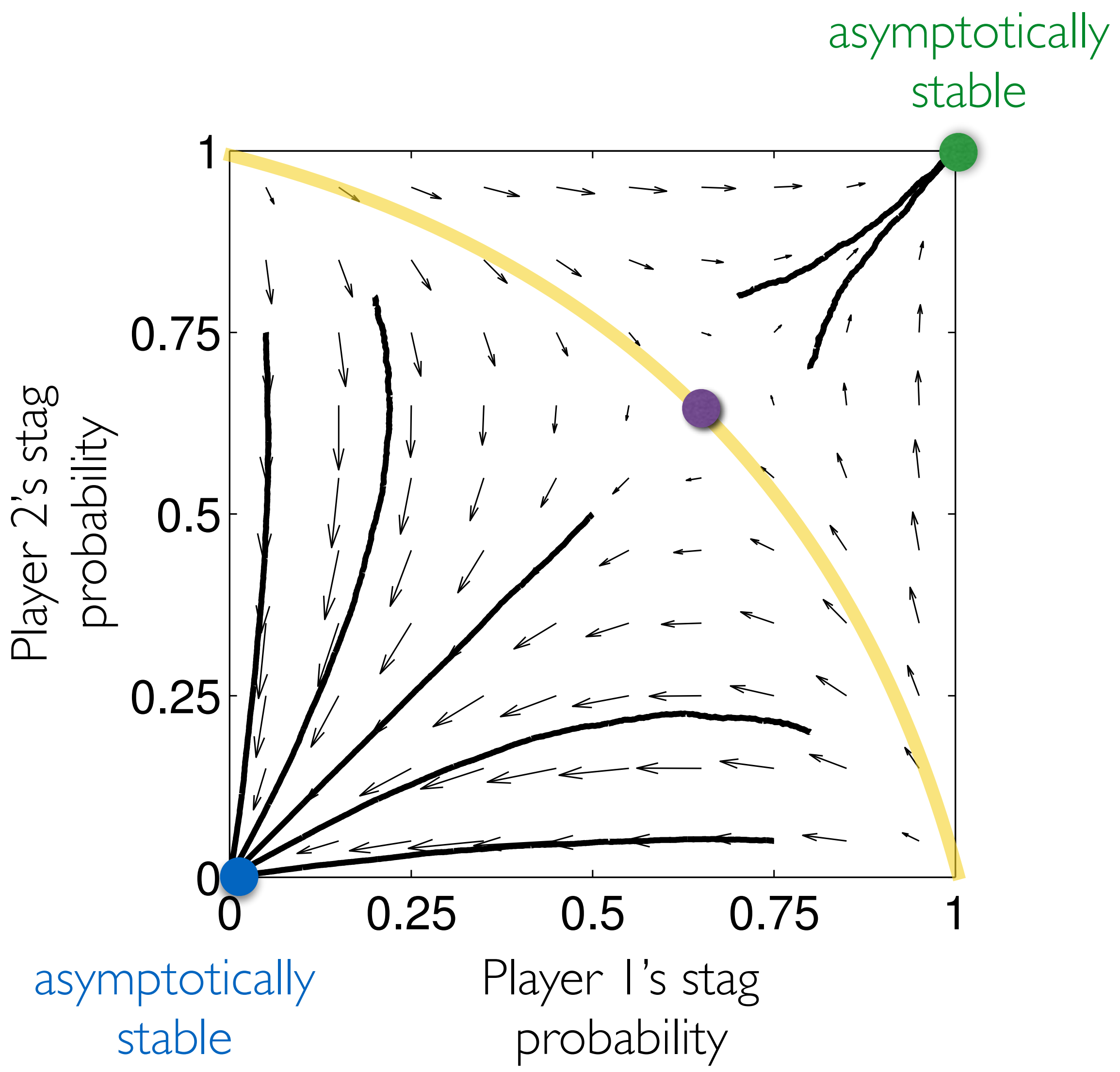
asymptotically stable

Stag hunt

Player 2

Player 1

	Stag	Hare
Stag	● 4,4	1,3
Hare	3,1	● 3,3

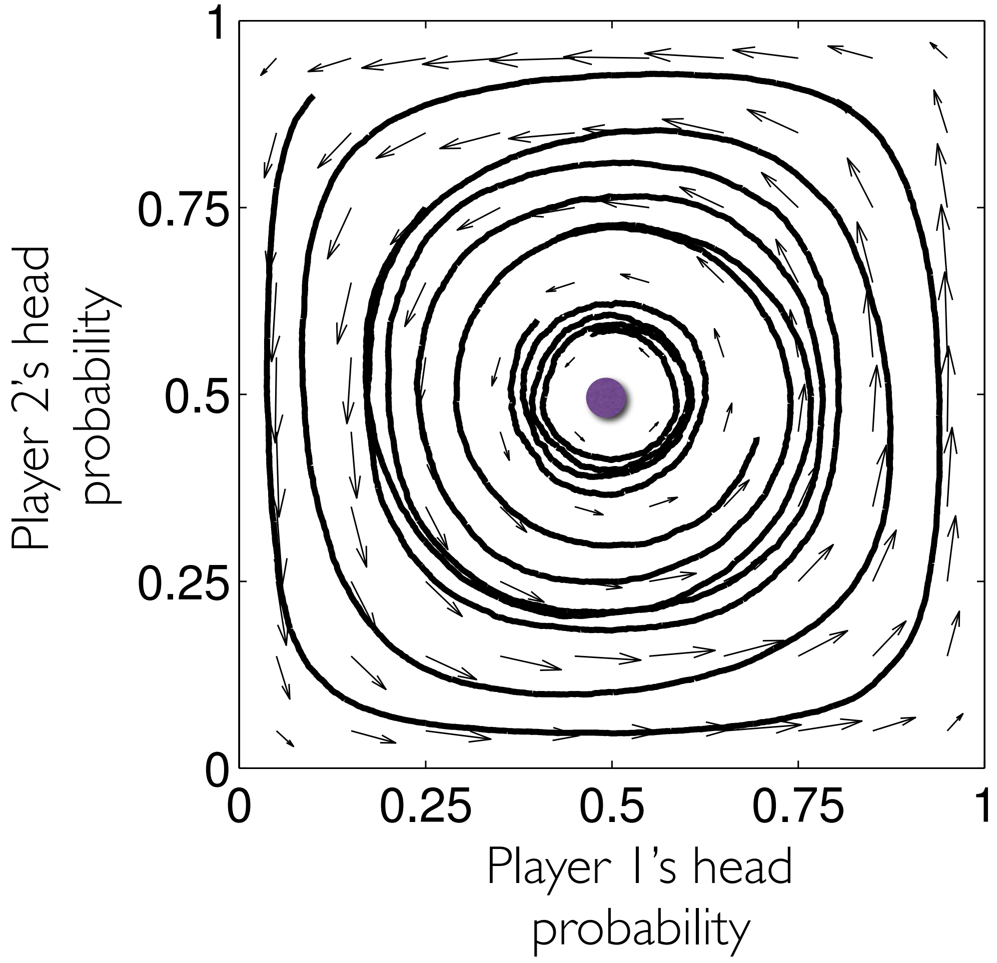


asymptotically stable

asymptotically stable

Matching pennies

		Player 2	
		Head	**Tail**
Player 1	**Head**	0,1	1,0
	Tail	1,0	0,1


Player 2's head probability
Player 1's head probability