European Research Council

Established by the European Commission

Slide of the Seminar

<u>Learning to Flock, Flocking to Learn</u>

Dr. Mihir Durve

ERC Advanced Grant (N. 339032) “NewTURB”
(P.I. Prof. Luca Biferale)

Università degli Studi di Roma Tor Vergata
C.F. n. 80213750583 – Partita IVA n. 02133971008 - Via della Ricerca Scientifica, 1 – 00133 ROMA

Learning to Flock, Flocking to Learn

Mihir Durve

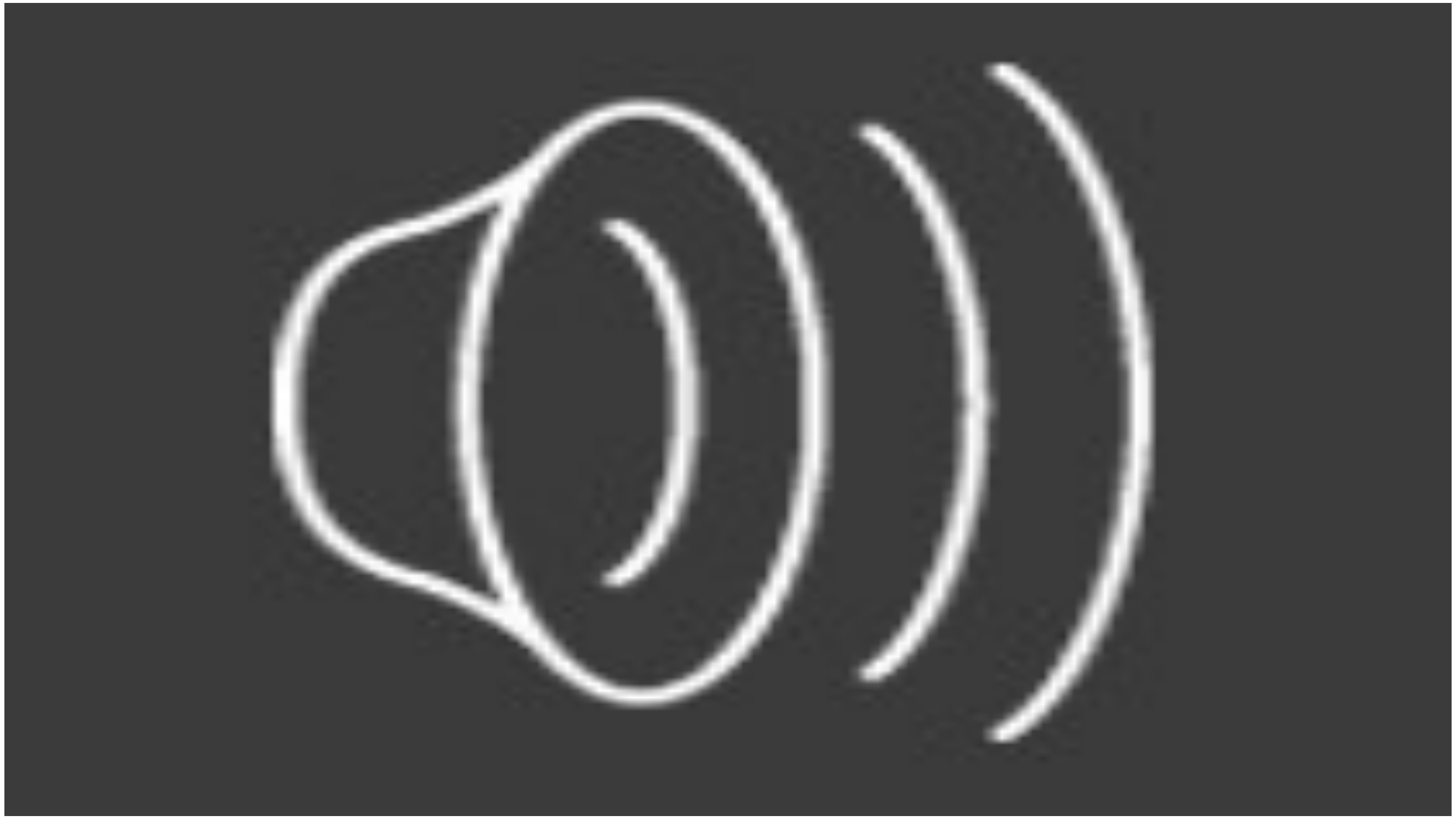
Department of Physics, University of Trieste
Quantitative Life Sciences, ICTP, Trieste

Supervisors

Prof. Antonio Celani

Prof. Edoardo Milotti

Flock of Starlings

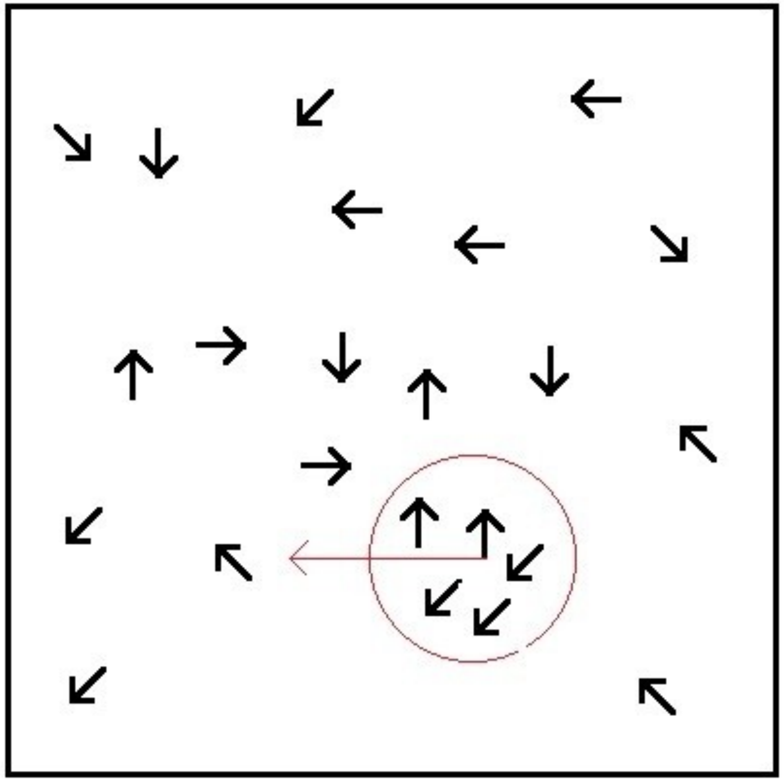


Examples

Credits
E Ben Jacob
O. Aburto

Vicsek Model

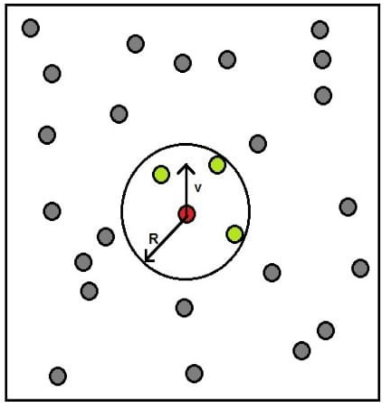
- Do as your neighbours are doing



Vicsek et al. Phys Rev Lett (1995)

Vicsek Model.

1. '*N*' particles are placed randomly and uniformly in a box of $L \times L$.
2. All the particles initially have random velocity.
3. All the particles move with a constant speed v_0
4. Neighborhood of interaction is a circle centered on the particle.



5. After every time interval all the particles adjust their direction to the average velocity of the particles in their neighborhood of interaction.
6. This adjustment is imperfect due to presence of noise.

Vicsek Model

Continue...

The noise is introduced in the model as;

$$\mathbf{v_i'} = v_0 \mathcal{R}(\theta)\hat{\mathbf{v}}(t) \tag{1}$$

here;

1. $\hat{\mathbf{v}}(t)$ is the unit velocity in the direction of the mean velocity of the particles in the neighborhood.
2. $\mathcal{R}(\theta)$ is the rotation operator which rotates the vector it acts upon (i.e., $\hat{\mathbf{v}}(t)$) by an angle θ. The angle θ is a random variable uniformly distributed over the interval $[-\eta\pi, \eta\pi]$.
3. η is strength of the noise in the range [0 to 1]

Vicsek Model

Order Parameter

Order parameter $\psi(t)$ is given by;

$$\psi(t) = \frac{1}{Nv_0} \left| \sum_{i=1}^{N} \mathbf{v}_i(t) \right| \tag{2}$$

$\psi(t) = 0$: Disordered state
$\psi(t) > 0$: Ordered state

Vicsek Model

Results

1. System undergoes second order phase transition as the noise is increased or particle density is decreased.

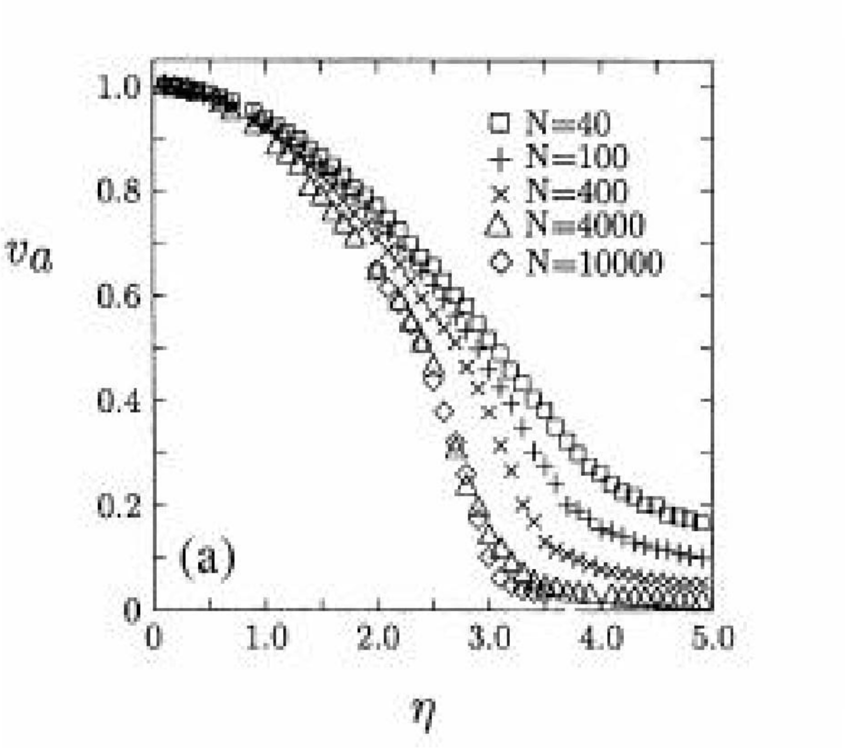


Figure : v_a Vs η

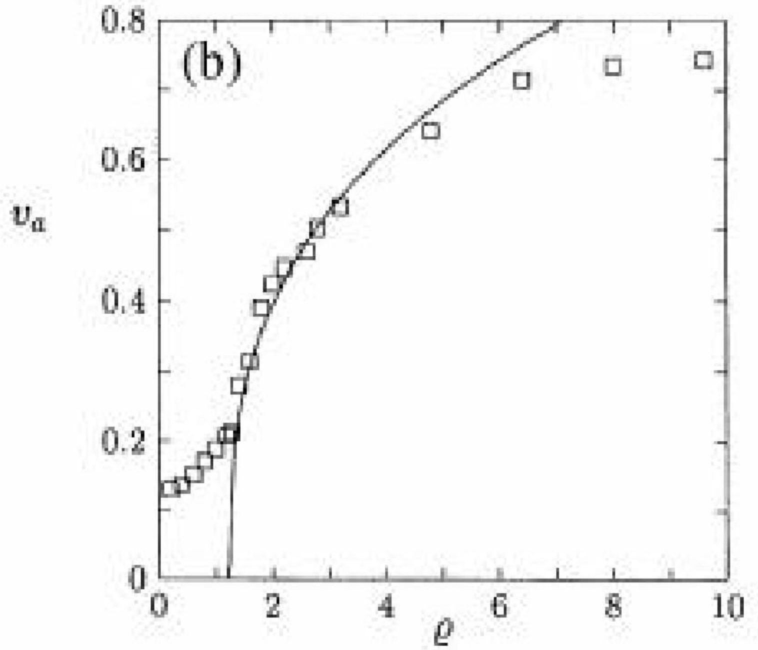


Figure : v_a Vs ρ

Figure Credits : Vicsek et al. [PRL 1995]

Vicsek Model

Random walkers

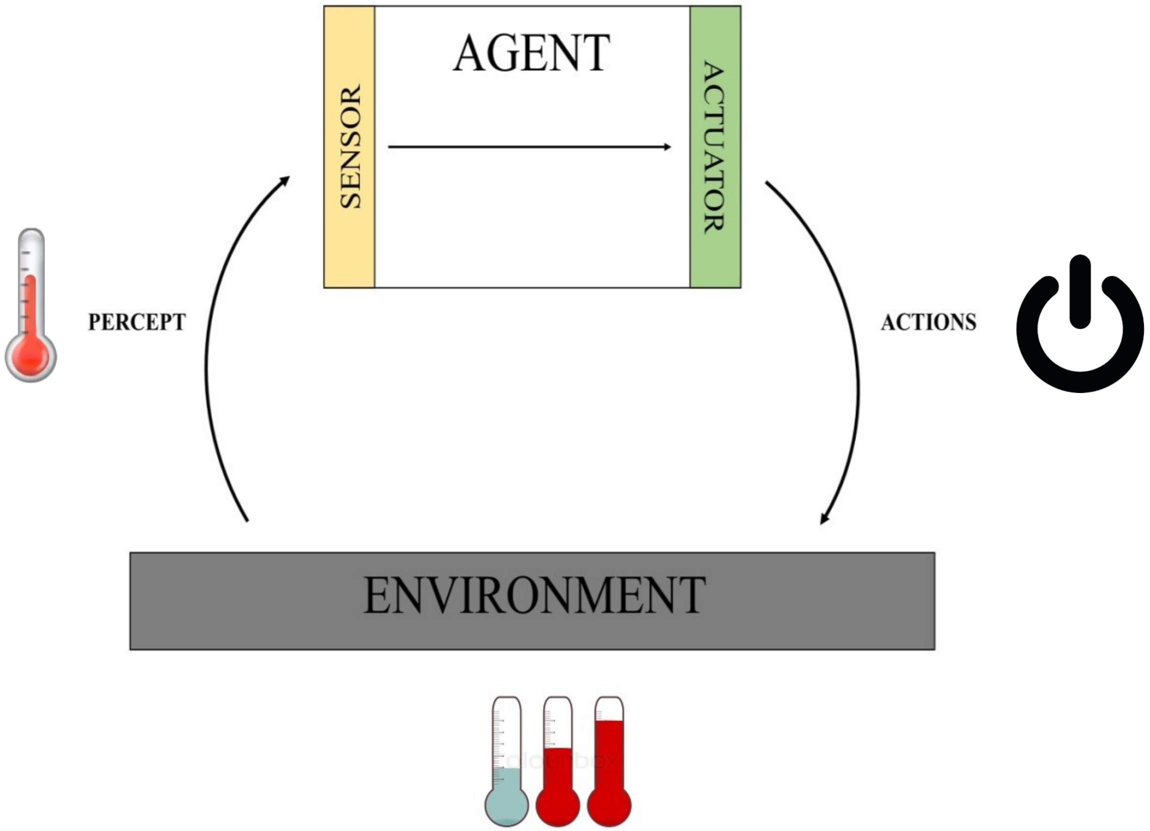
Flock

Vicsek et al. Phys Rev Lett (1995)

Goal of the study

- For Flocking quantity to be optimized : Number of neighbours
- We use stochastic optimization techniques

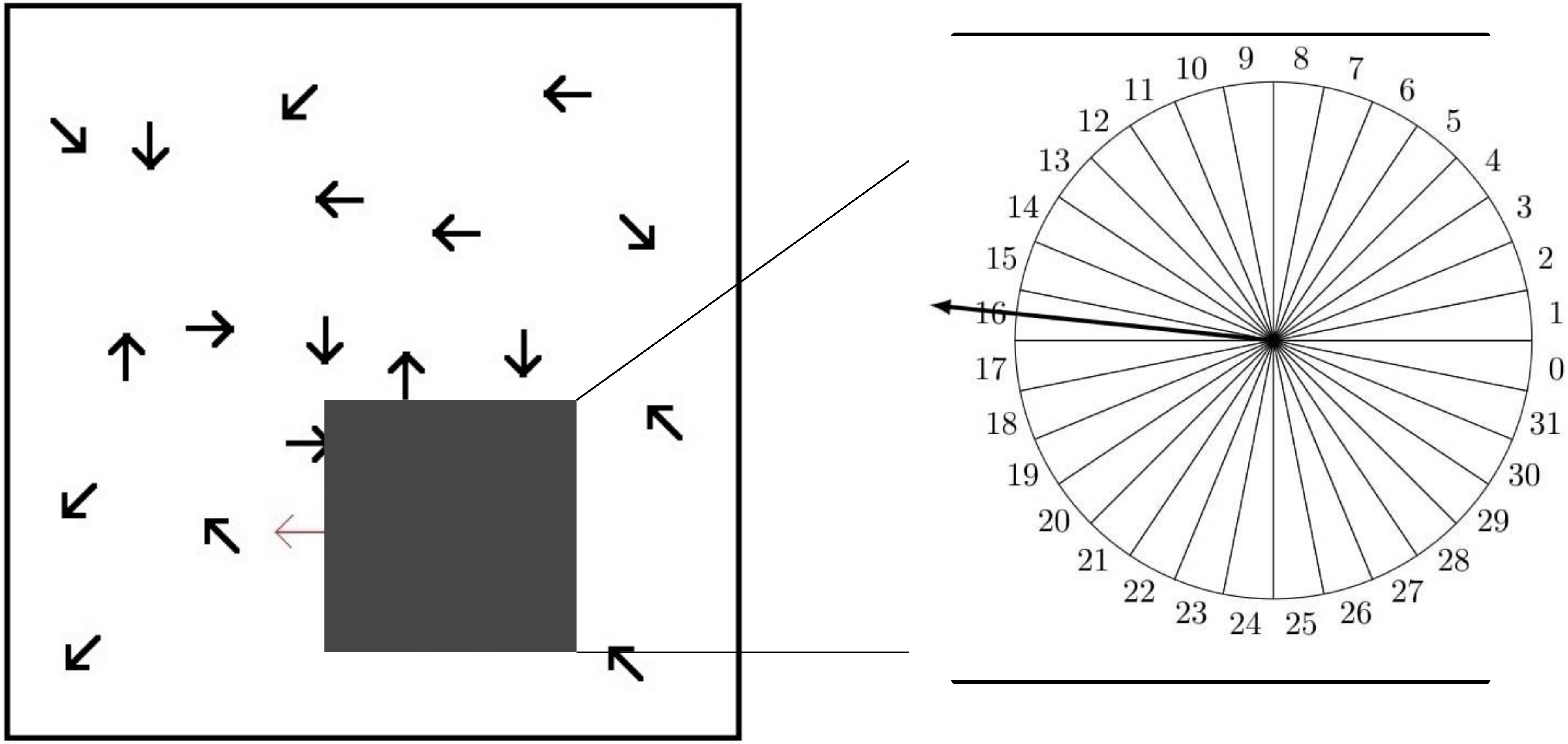
Reinforcement Learning



Sutton and Barto (1998)

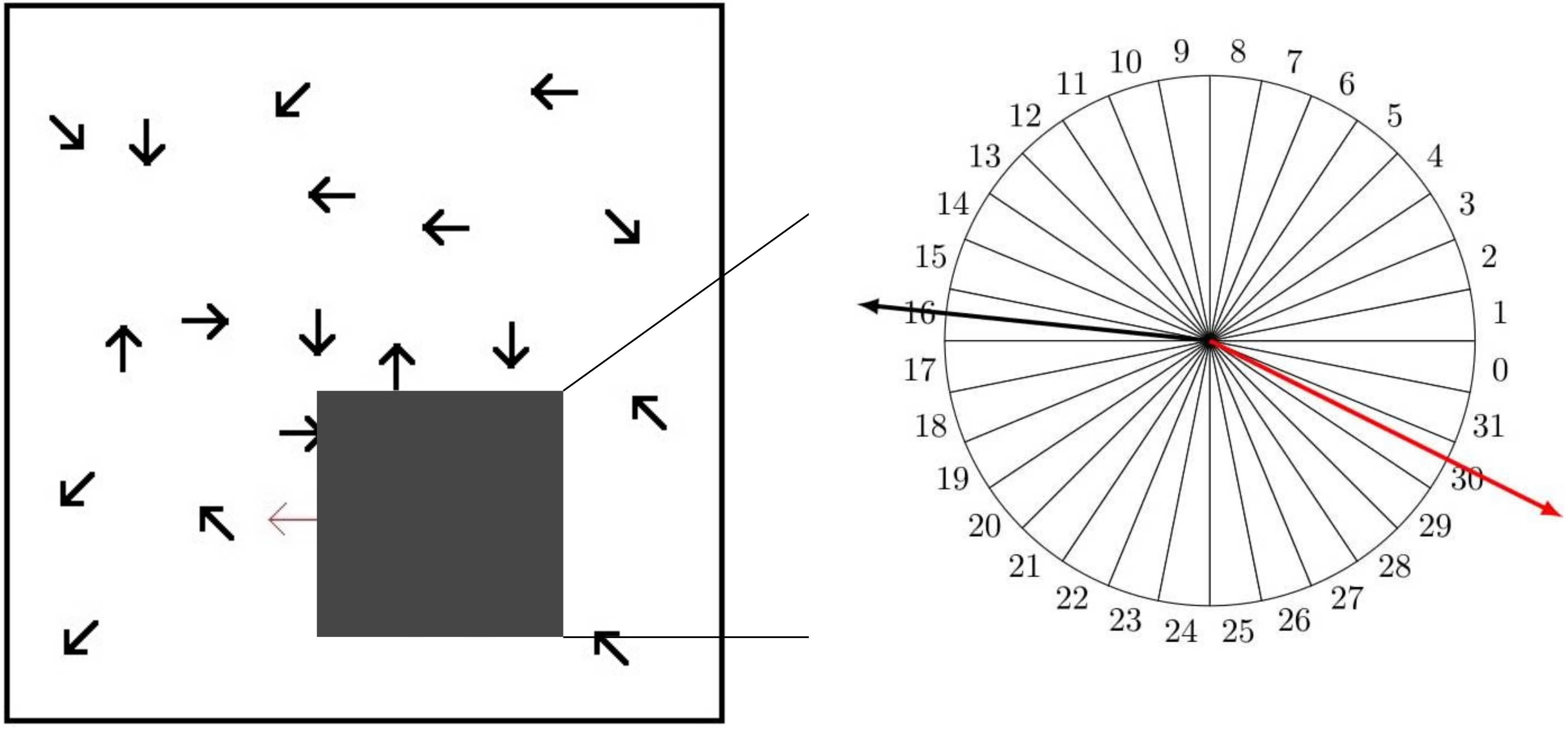
RL in multi agent system :

States, Actions



RL in multi agent system :

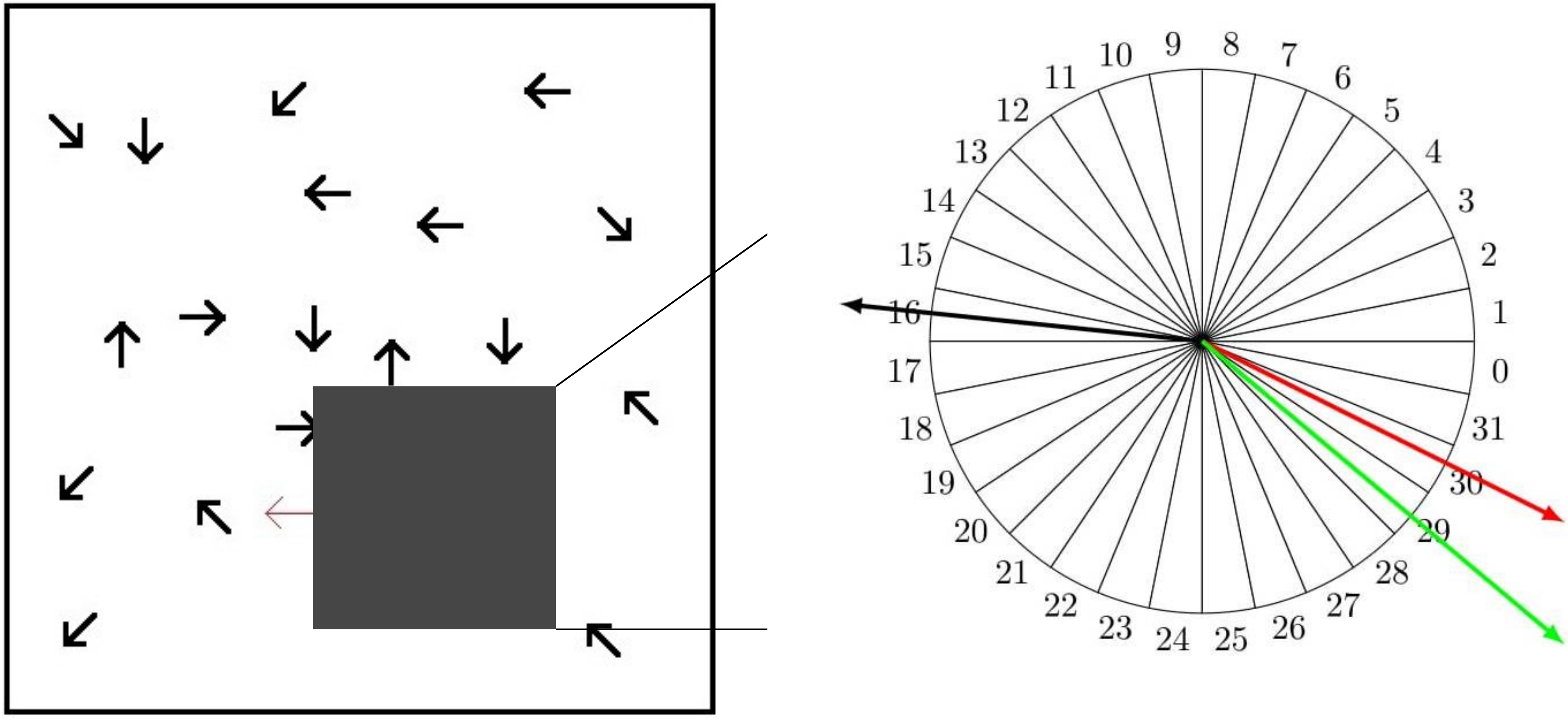
States, Actions



State label : 30

RL in multi agent system :

States, Actions



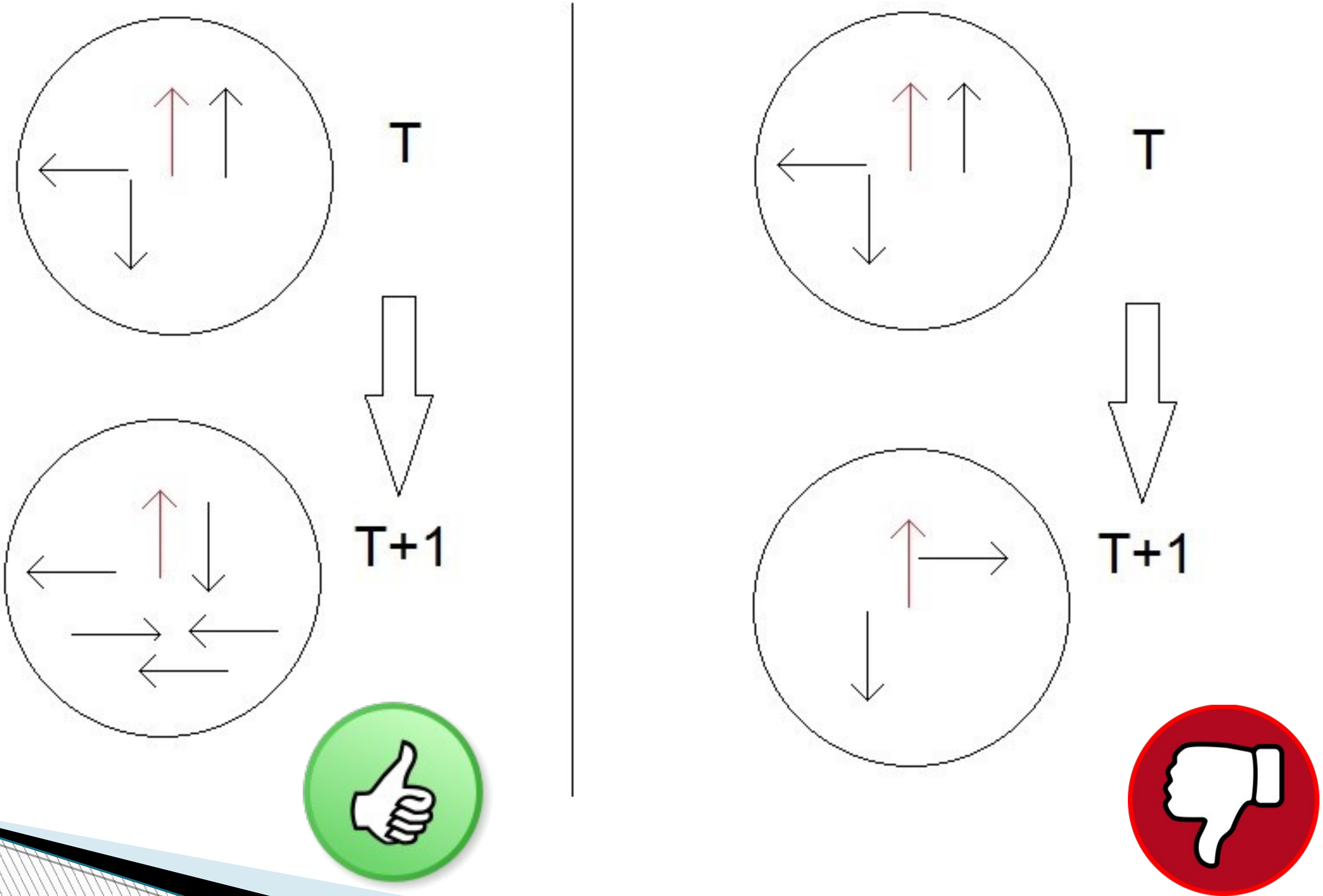
State label : 30
Possible Actions : 0-31

RL in multi agent system :

$$x_i(t+1) = x_i(t) + v_a(i) \times \Delta T$$

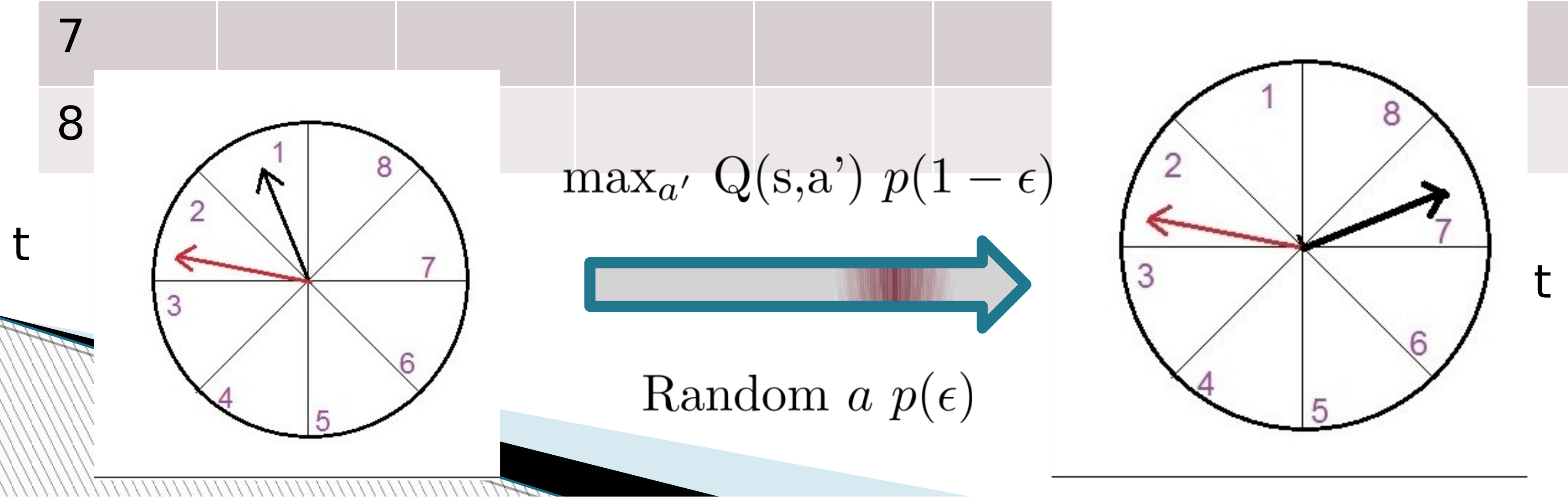
RL in multi agent system :

Reward for individual agents



Q-matrix

S	A	1	2	3	4	5	6	7	8
1		Q(1,1)	Q(1,2)	…					
2		Q(2,1)	…						
3		…							
4									
5									
6									
7									
8									



$\max_{a'}$ Q(s,a') $p(1-\epsilon)$

Random a $p(\epsilon)$

RL in multi agent system

Q-update rule

$$Q(s_n, a_n) \leftarrow Q(s_n, a_n) + \alpha[r_n - Q(s_n, a_n)]$$

👍 $r_n = +R_f$

👎 $r_n = -R_f$

S	A	1	2	3
1		Q(1,1)	Q(1,2)	…
2		Q(2,1)	…	
3		…		

C. Watkins (1992)

Q-learning
Episode

- Each agent begins in a box each with its own Q-matrix (initially flat)
- DO T=0, T=T_max
    - Observes the state s
    - Chooses action a
    - Updates position and orientation
    - Receives reward
    - Update Q-matrix
- End DO

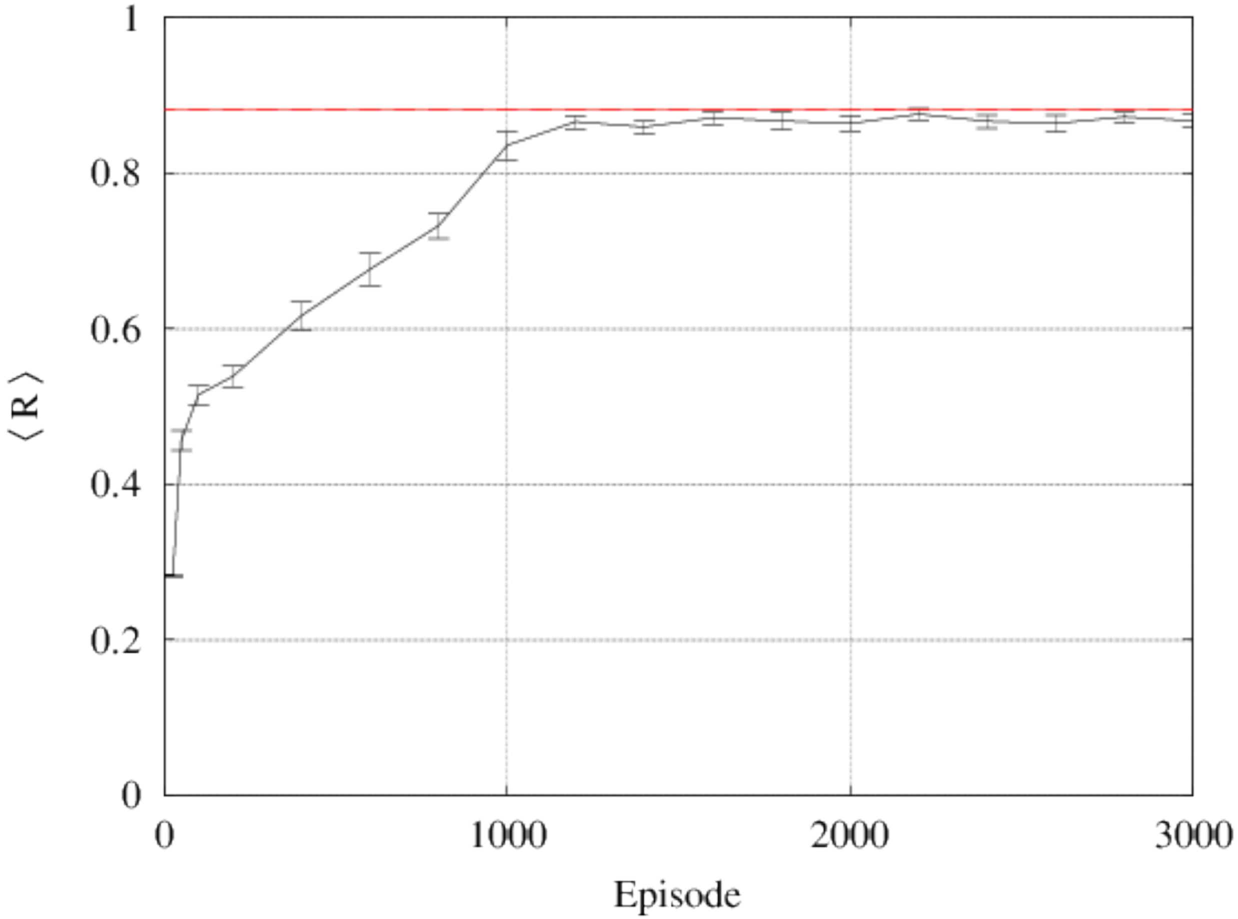
S	A	1	2	3
1		Q*(1,1)	Q*(1,2)	…
2		Q*(2,1)	…	
3		…		

Preliminary Results

In the beginning

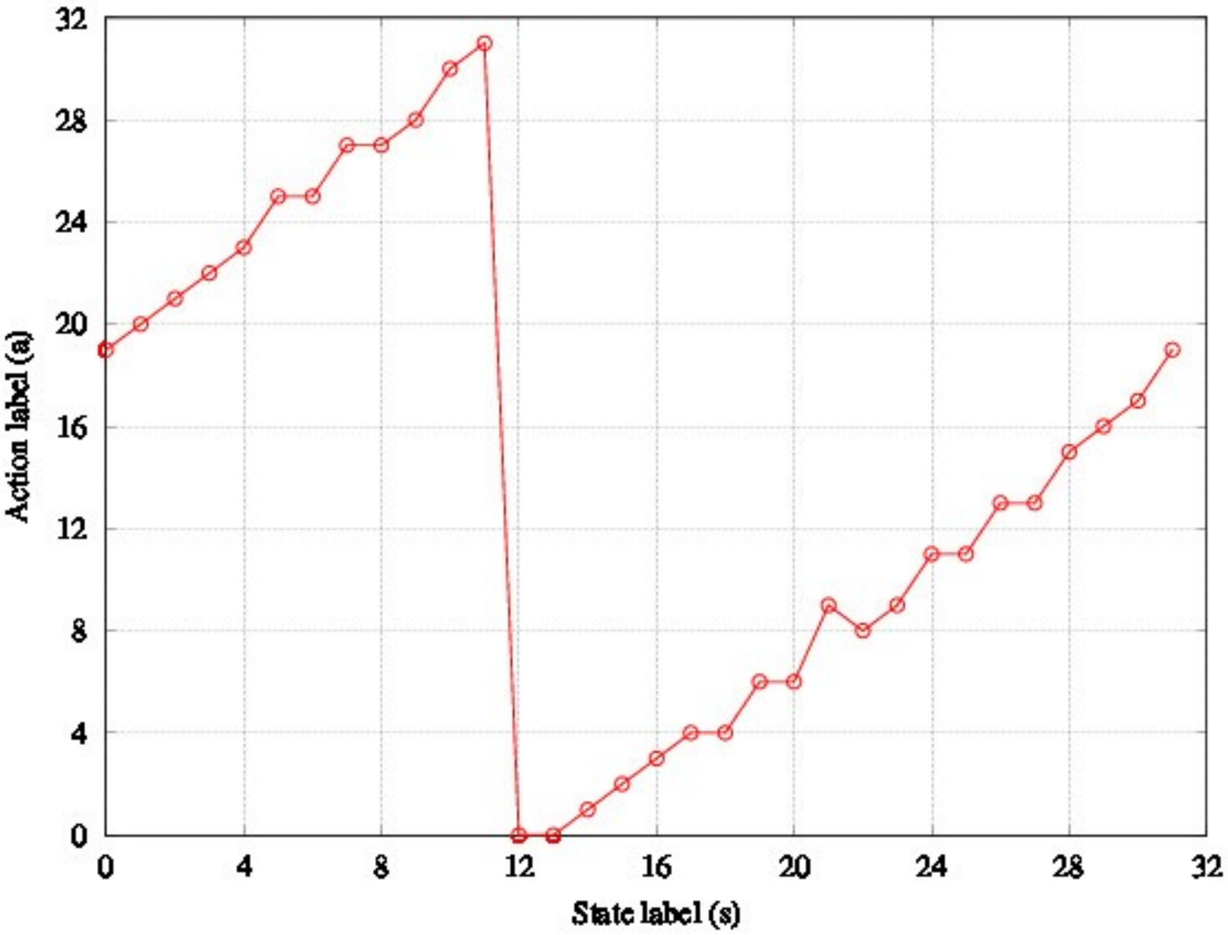
In the end

Preliminary Result
Average Reward with episodes



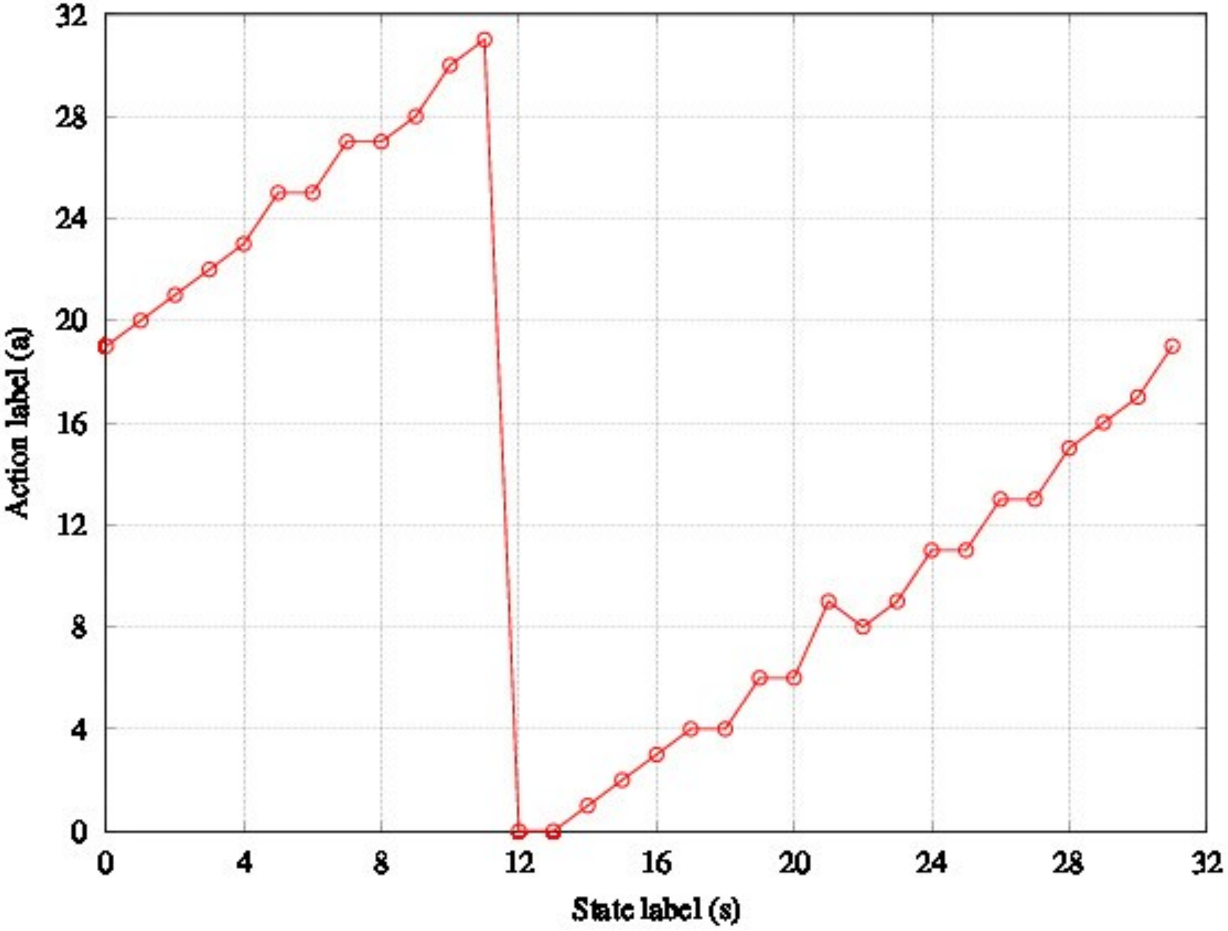
Preliminary Results

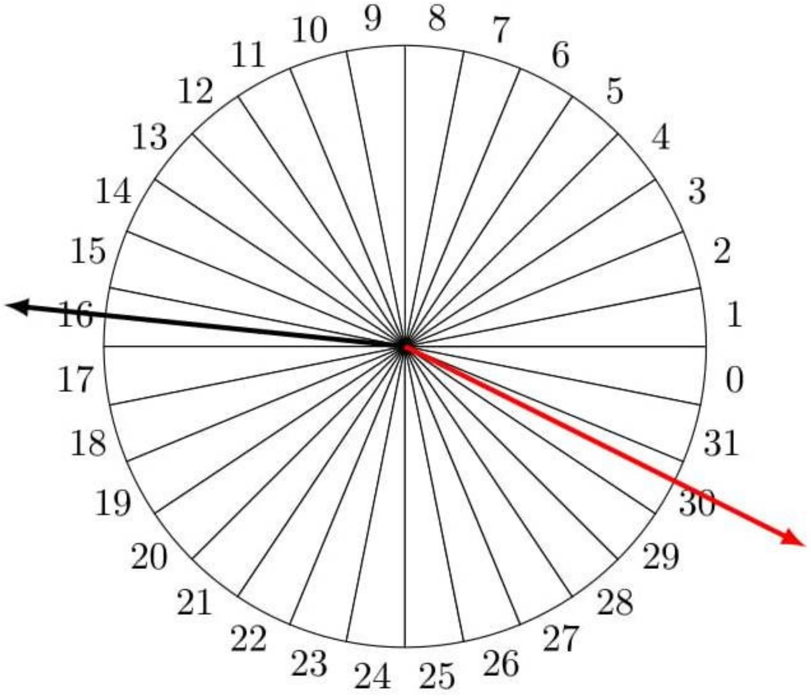
Best a for s in Q-matrix

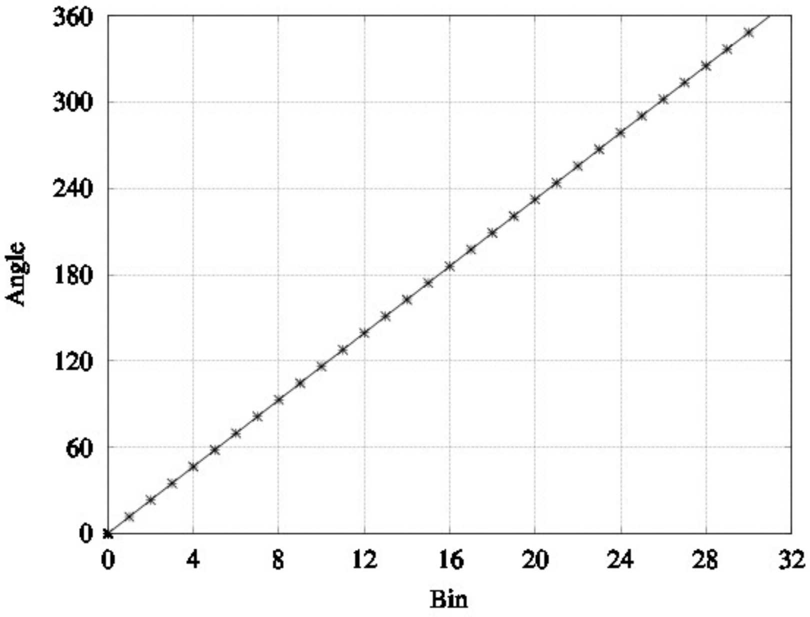


Preliminary Results

Best a for s in Q-matrix

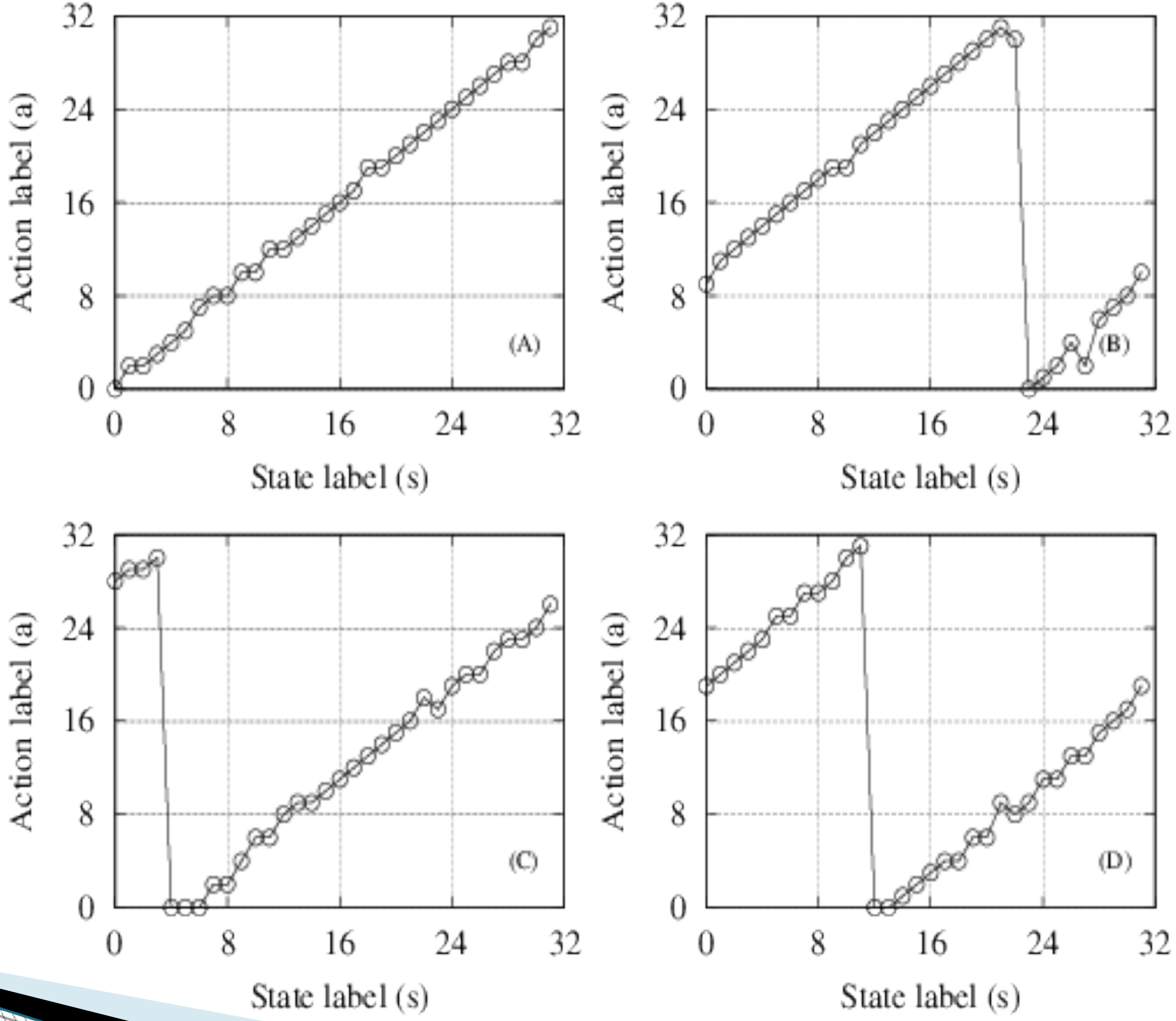






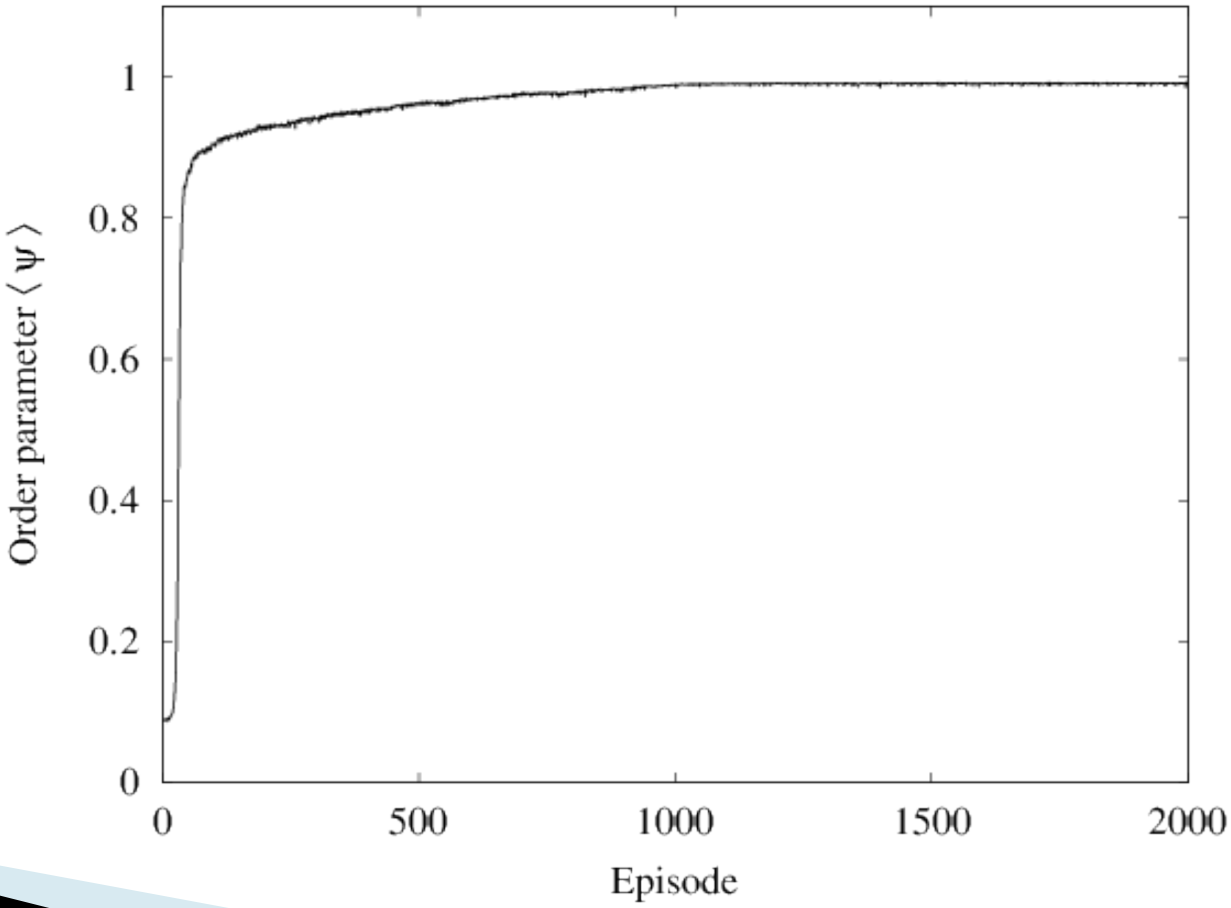
Preliminary Results

Max (Q(s,a))

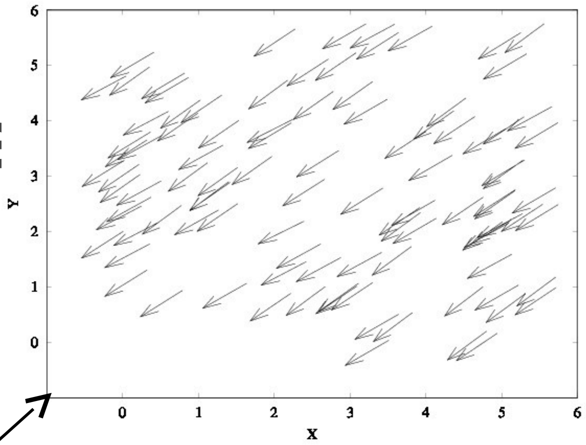


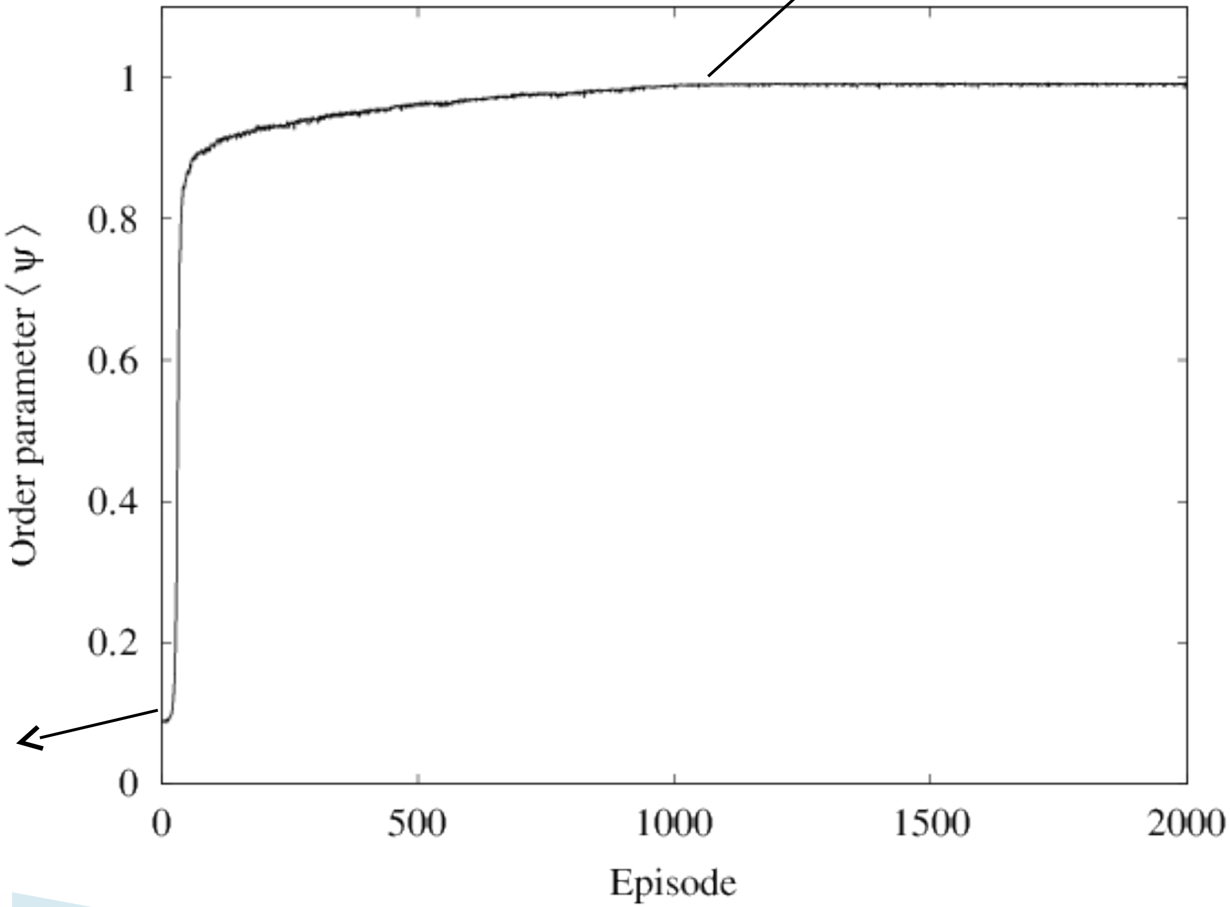
Preliminary Results
Order parameter with Episodes

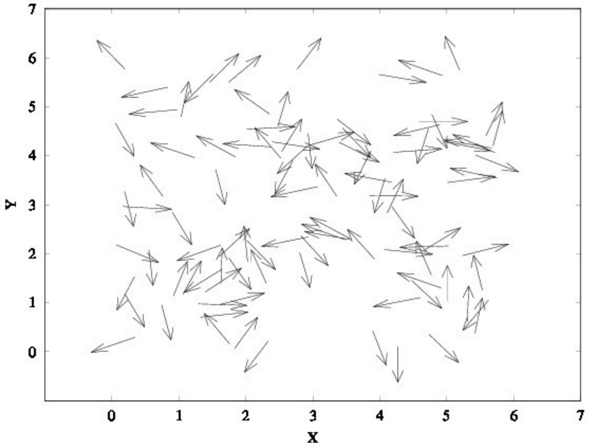
Order parameter : $\psi(t) = \frac{1}{N v_0} \left| \sum_{i=1}^{N} \mathbf{v}_i(t) \right|$



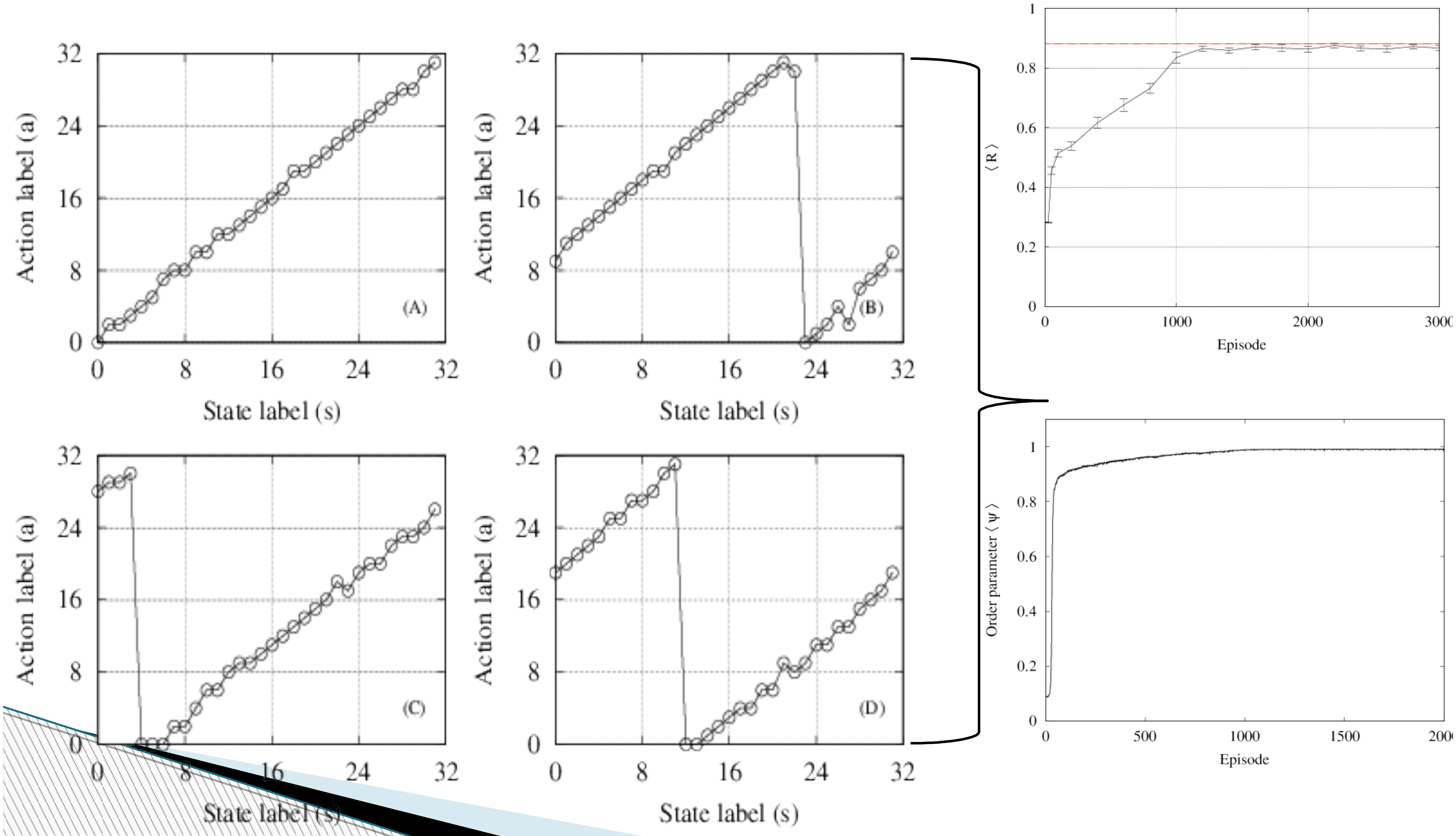
Preliminary Results
Order parameter with Episode







Preliminary Results
Policies



Conclusion

- Multi-agent system optimizing aggregation formed highly polar ordered state.

Next plans

- Restricting the set of actions
- Changing the reward schemes
- Changing the percept

THANK
YOU

$$Q(s_n, a_n) = Q(s_n, a_n) + \alpha(r_n + \gamma max_{a'} Q(s_{n+1}, a') - Q(s_n, a_n))$$

$$Q^*_\pi(s_n, a_n) = \langle r_{n+1} + \gamma max_a Q_\pi(s_{n+1}, a) \rangle$$