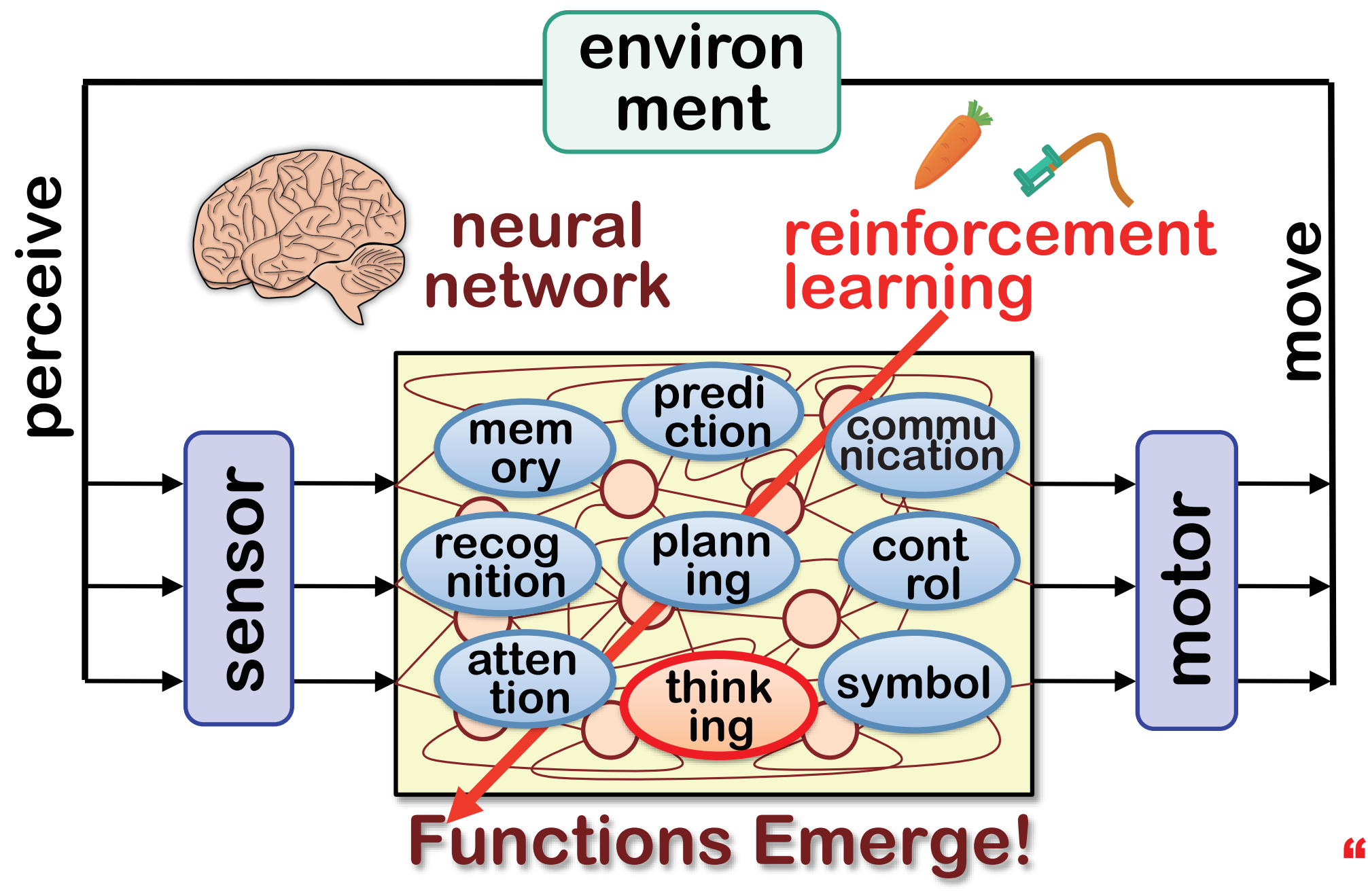
New Reinforcement Learning Using a Chaotic Neural Network for Emergence of “Thinking”
— “Exploration” Grows into “Thinking” through Learning —

RLDM 2017

Katsunari SHIBATA & Yuki GOTO (Oita University, JAPAN)　katsunarishibata@gmail.com　http://shws.cc.oita-u.ac.jp/shibata/home.html

☆ Function Emergence through End-to-End Reinforcement Learning

(K. Shibata et al. 1997--, D. Hassabis et al. 2013--)



We have already confirmed that various functions including memory-required functions emerge in a recurrent neural network (RNN) through RL in simple tasks as shown in the yesterday s poster. Feel free to ask some original papers.

but

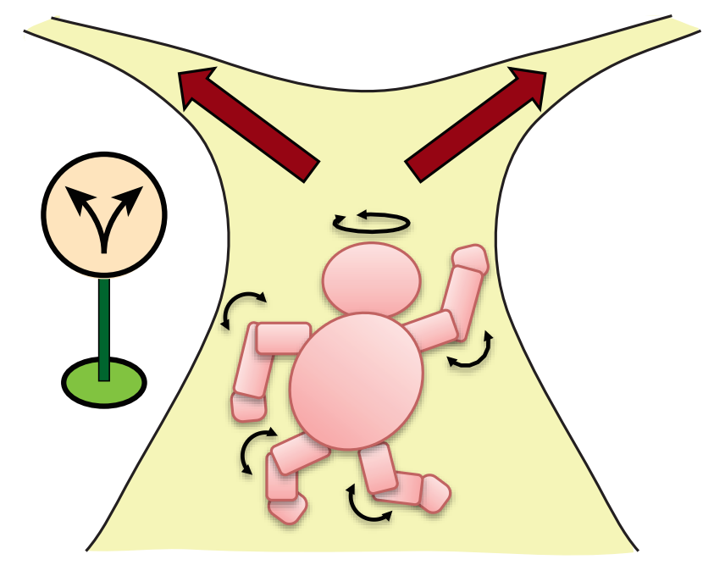
“Thikning” has hardly emerged.

Memory: convergence dynamics

Thinking: dynamics with autonomous state transitions

☆ Exploration is not only stochastic action selection but one of our important functions!

At a fork road



* We don’t work our muscle randomly.

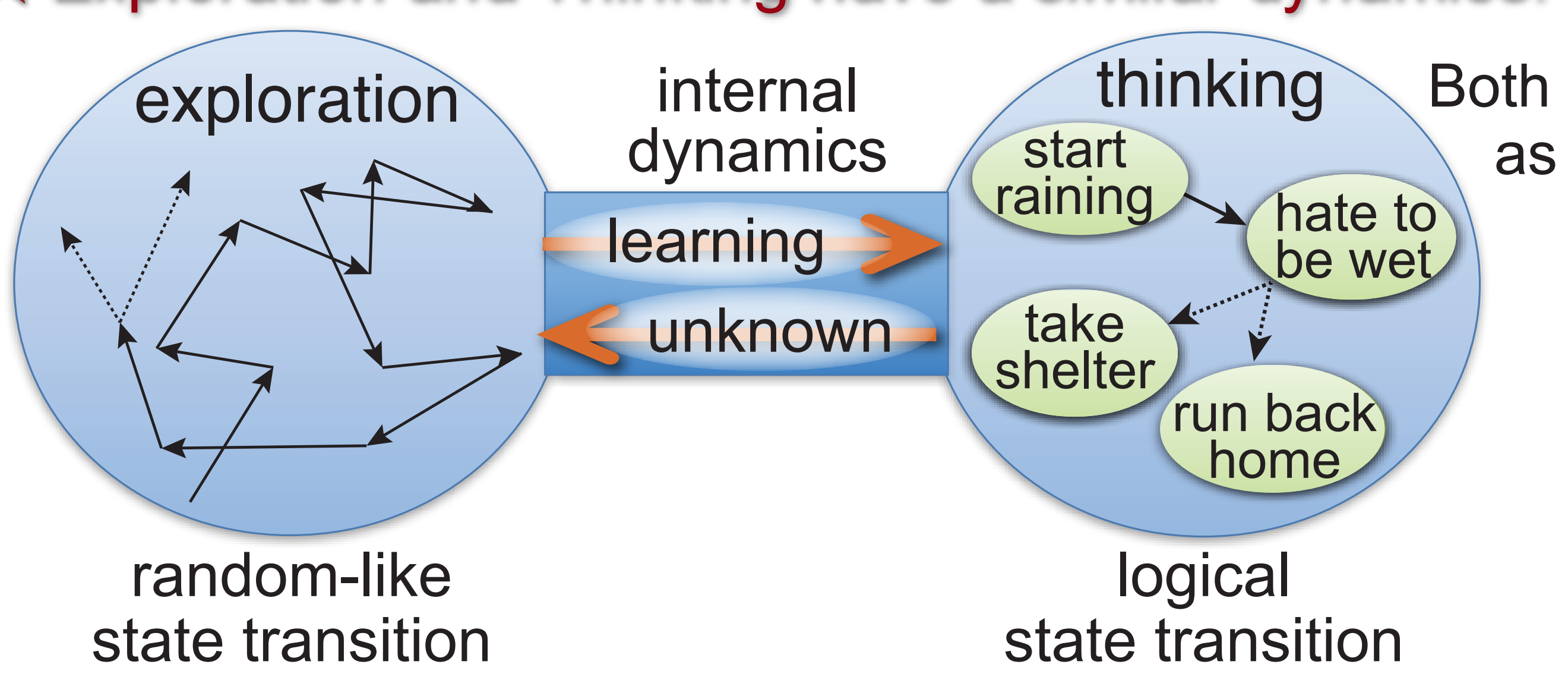
* We usually don’t go into a off-road place.

* We think many things
How is the road condition?
Which path is approaching to the destination?
Is there any information sign?

Our exploration is usually very intelligent reflecting past learning.

We call it Higher exploration

☆ Exploration and Thinking have a similar dynamics!

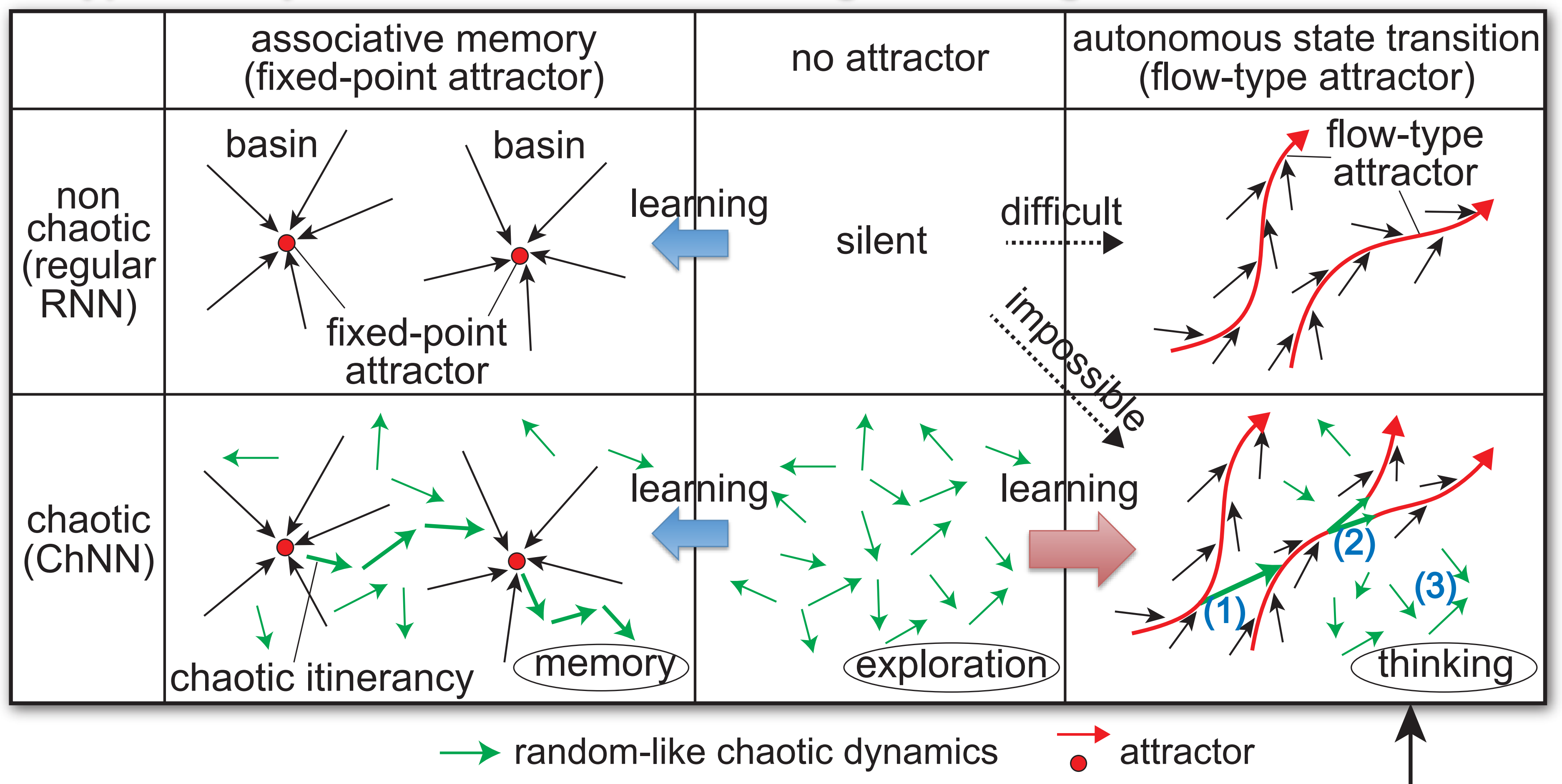


Both are generated as internal dynamics!

☆ What is the essential property?

- * Non-fixed-point convergence Dynamics
  (Autonomous state transition is necessary.
  Even though you close your eyes and ears, you can think)
- * Learning is reflected
- * Higher exploration
- * Inspiration or discovery (unexpected but rational transition)
- * Exploration in unknown situations.

☆ Types of Dynamics and Growth through Learning



Hypothesis 1

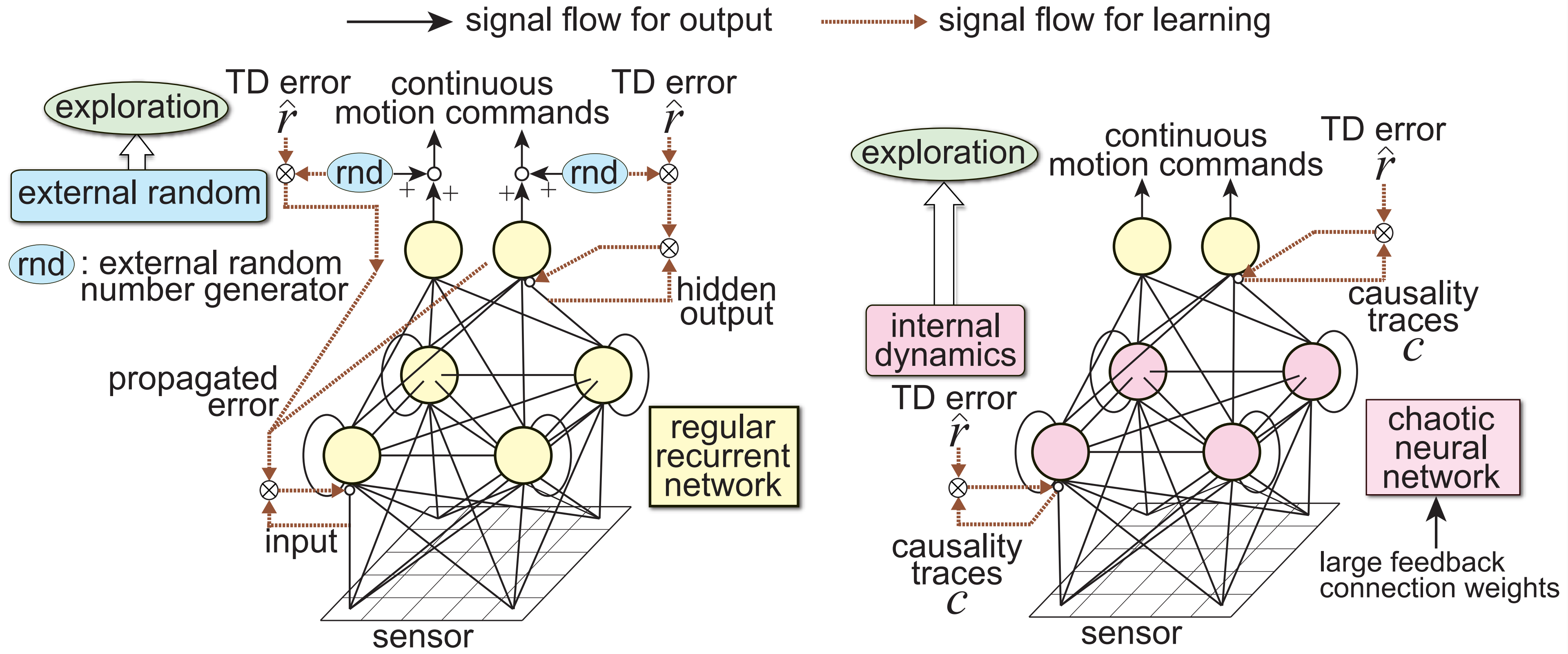
Exploration Grows into Thinking through Learning

Hypothesis 2

The growth is done on a chaotic NN (ChNN)! The ChNN produces
(early phase) exploration
(through learning) (1) inspiration and discovery,
(2) higher exploration and
(3) exploratory behaviors in unknown situations.

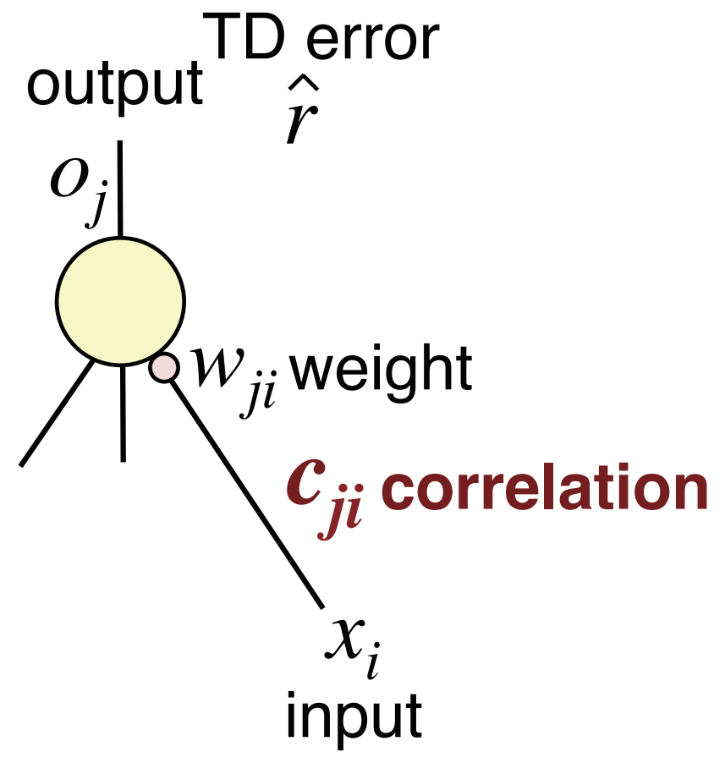
☆ New Reinforcement Learning Using a chaotic NN

Comparison with Conventional RL (Actor network)



(a) conventional RL　(b) proposed RL

Causality traces and Learning



Update of Causality Traces

$$c_{ji,t} = (1-|\Delta o_{j,t}|)c_{ji,t-1} + \Delta o_{j,t}\, x_{i,t}$$

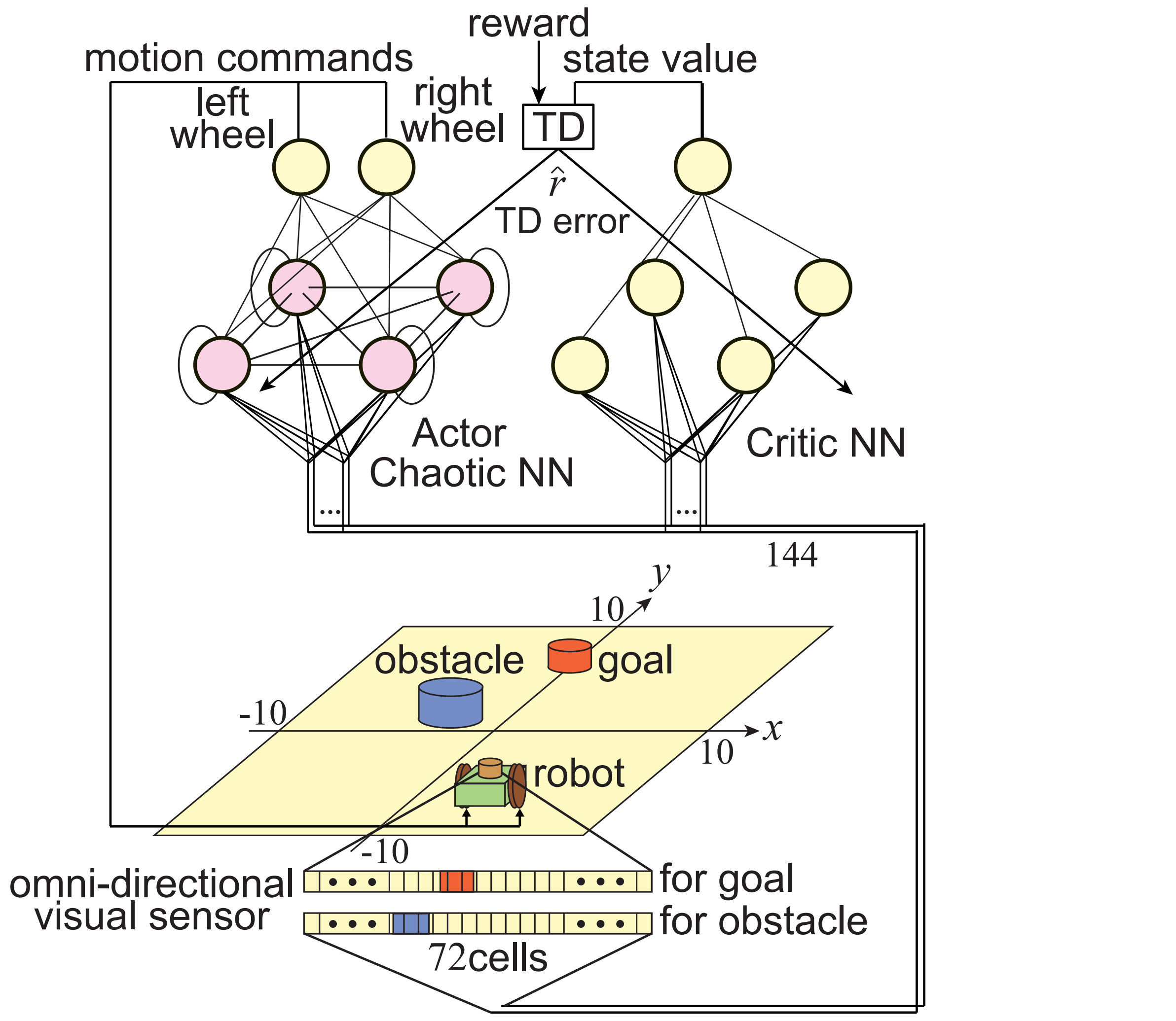
Representing how the input has contributed to the present output

Update of connection weights

$$\underbrace{\Delta w_{ji} = \eta \hat{r}_t\, c_{ji,t}}$$

learning rate　TD error

☆ Obtacle Avoidance Task and Signal Flow

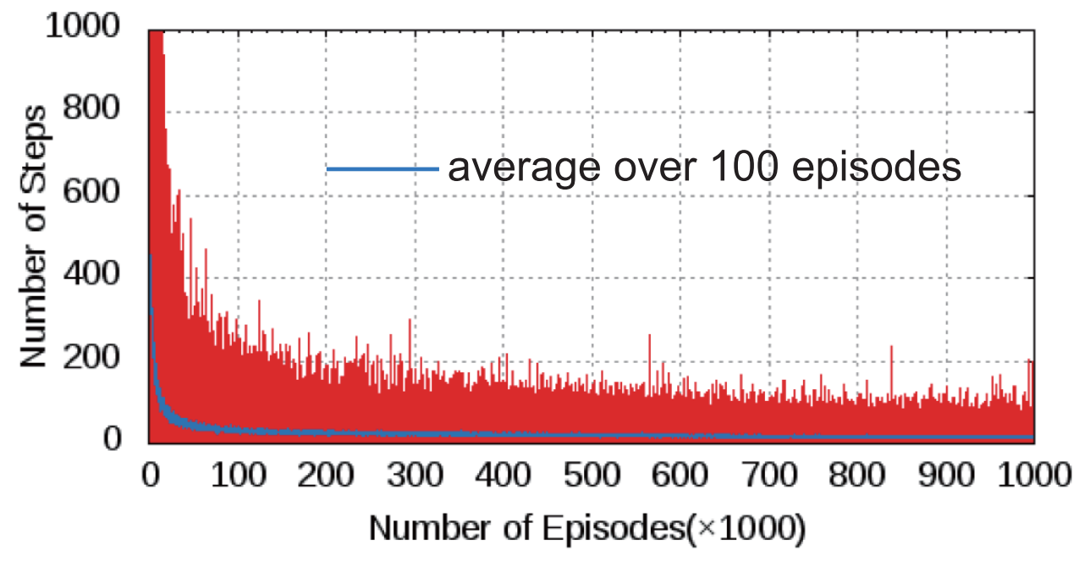


* The robot and obstacle are located randomly at each trial.

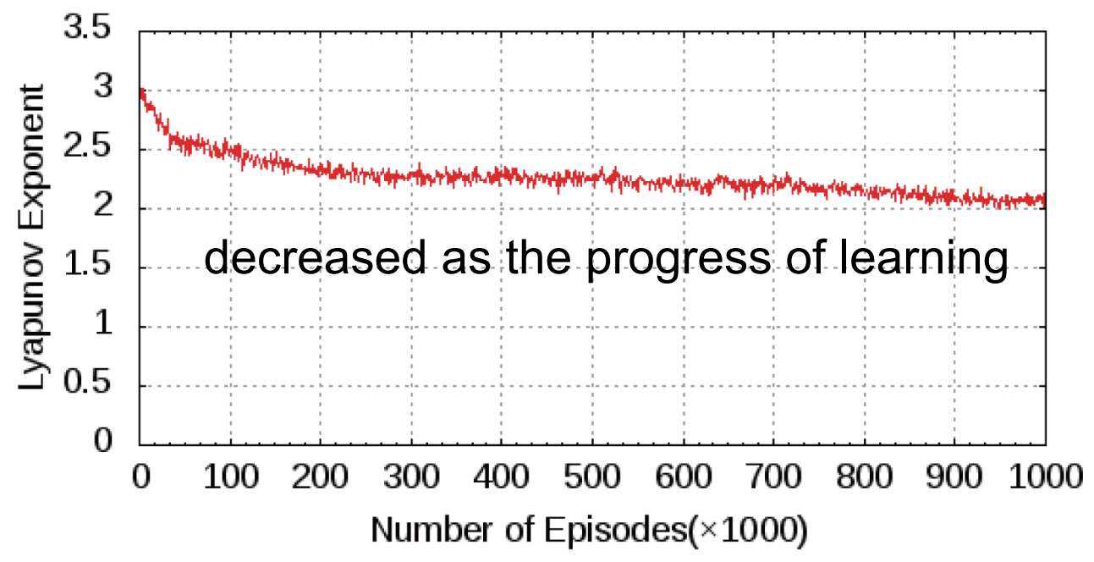
* Reward: when reaching to the goal
Small penalty: when colliding with the obstacle or a wall.

☆ Learning Results

Learning curve (Number of steps)



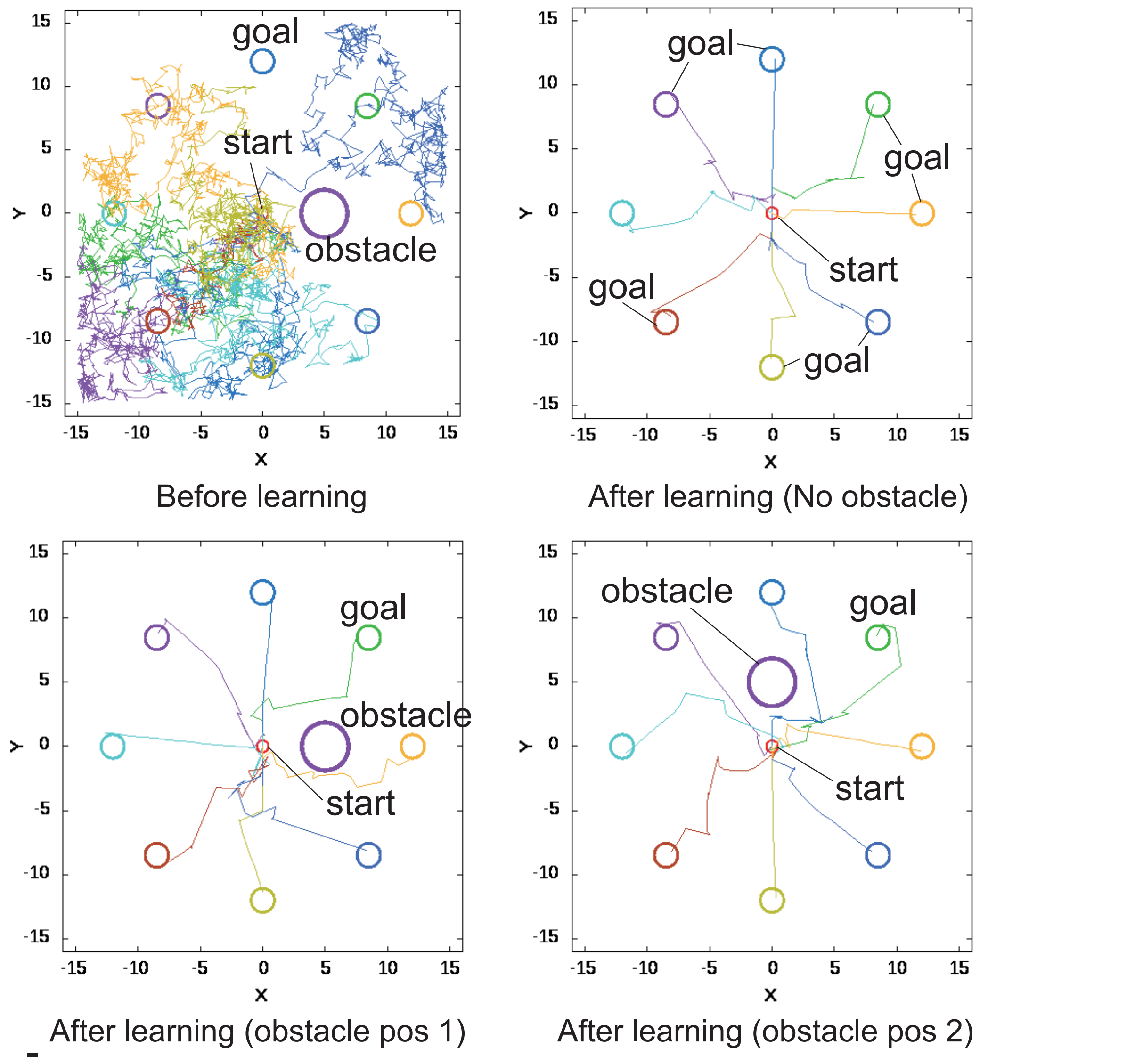
Lyapunov exponent during learning



The neurons in the chaotic NN were observed.
This exponent indicates
positive : diverging chaotic dynamics
negative: converging dynamics

Trajectory samples

Actually the goal location is fixed, but this figure is drawn such that the robot is located at the center and its orientation is the same as y-axis.



Before learning　After learning (No obstacle)

After learning (obstacle pos 1)　After learning (obstacle pos 2)

Some irregularities are seen in the trajectories --> A future work