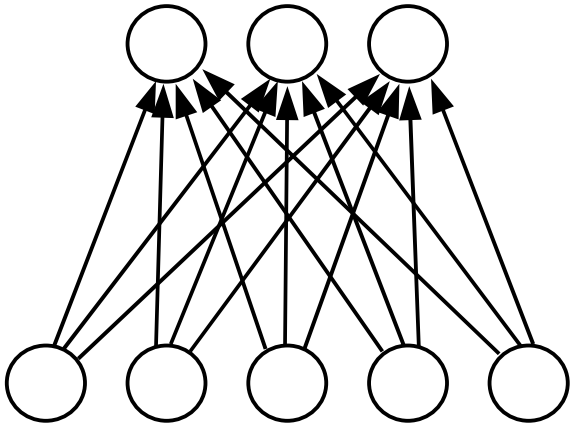
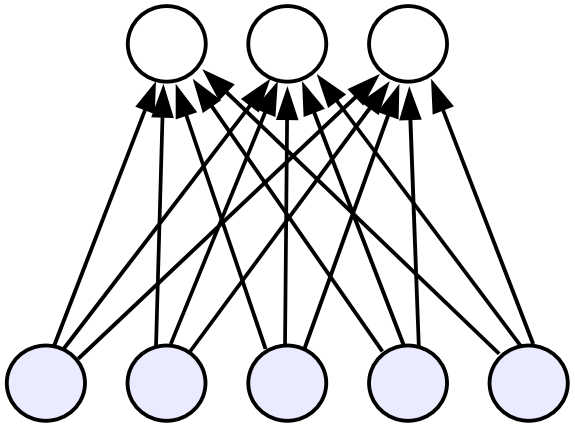
Evolutionary Reinforcement Learning

- Reference: David Ackley & Michael Littman, “Interactions between learning and evolution”, in *Artificial Life II*, edited by Langton, Taylor, Farmer & Rasmussen, Addison-Wesley, 1991.
- Studied the combined effects of evolution and learning within a simulated world
- Simulated world is a 2-dimensional grid containing agents, carnivores, food sources, and obstacles
- Agents must maintain their energy level by eating food and avoiding carnivores and obstacles, otherwise they will die
- Each agent is controlled by a pair of neural networks specified by its genome
  - Action network
  - Evaluation network

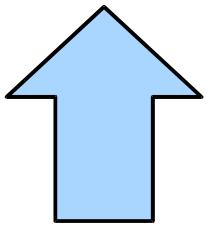
Action Network



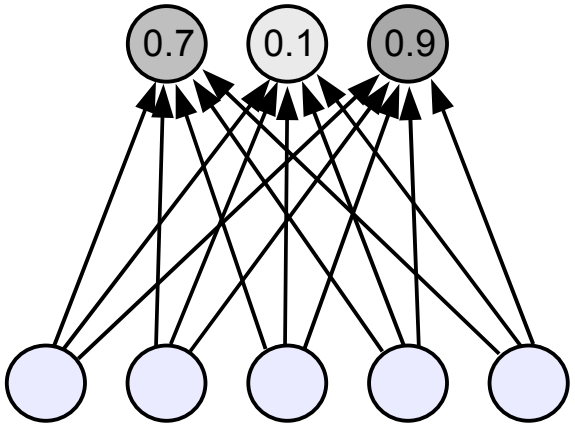
Action Network



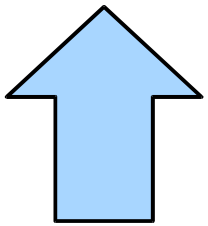
current sensory state



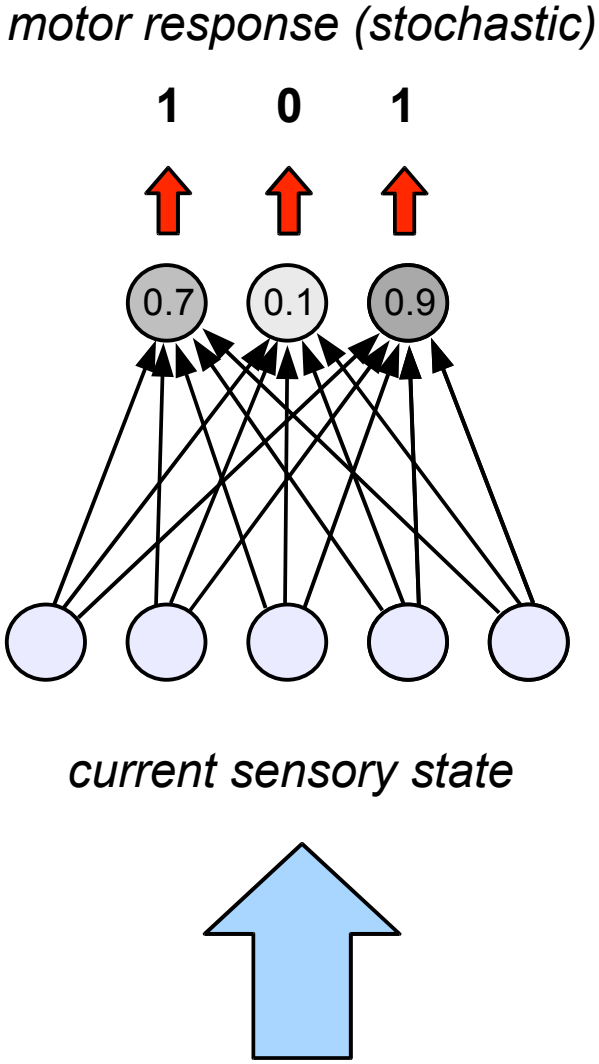
Action Network



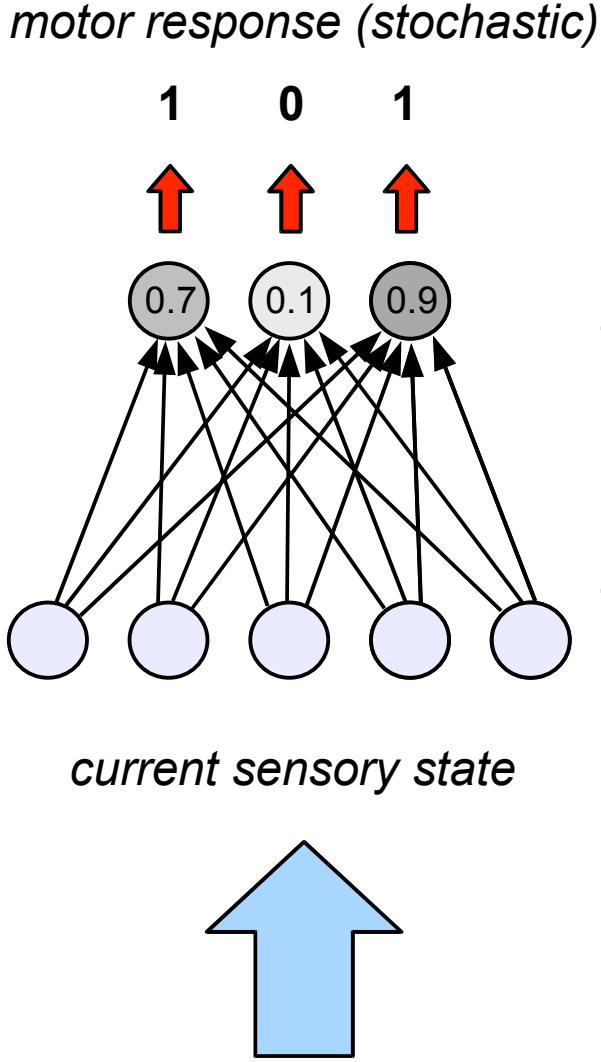
current sensory state



Action Network



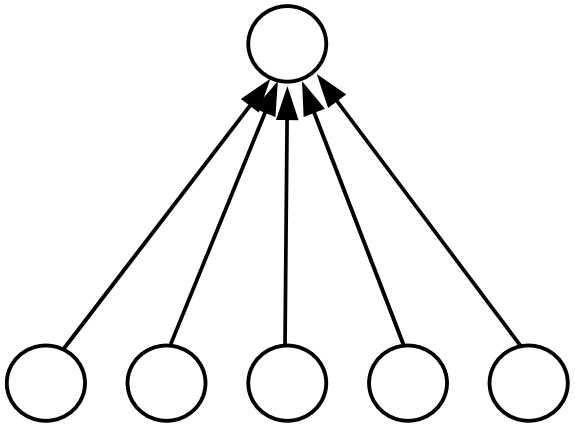
Action Network



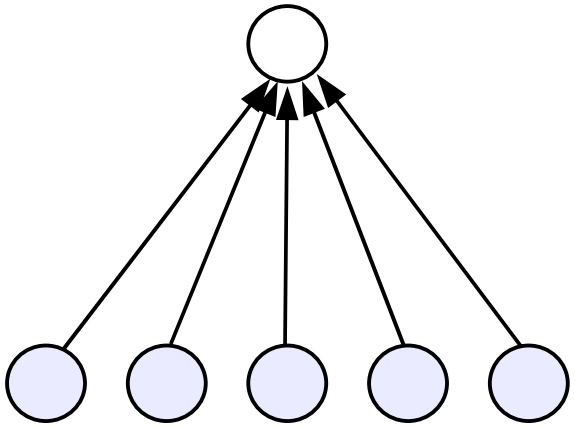
Initial weight values specified by agent's genome at “birth”

Modified by learning over agent's lifetime

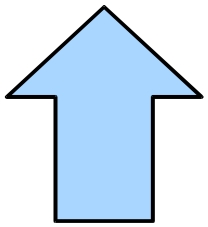
Evaluation Network



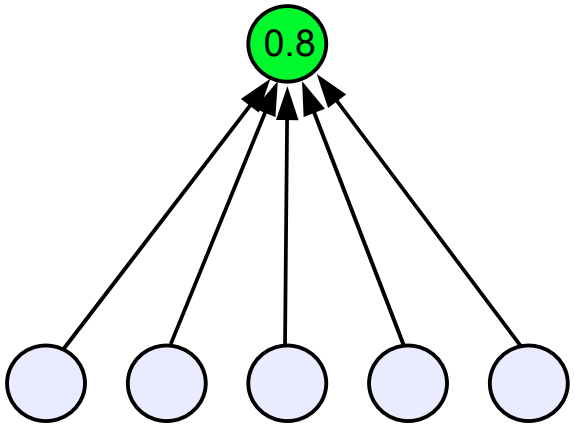
Evaluation Network



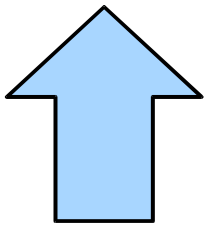
current sensory state



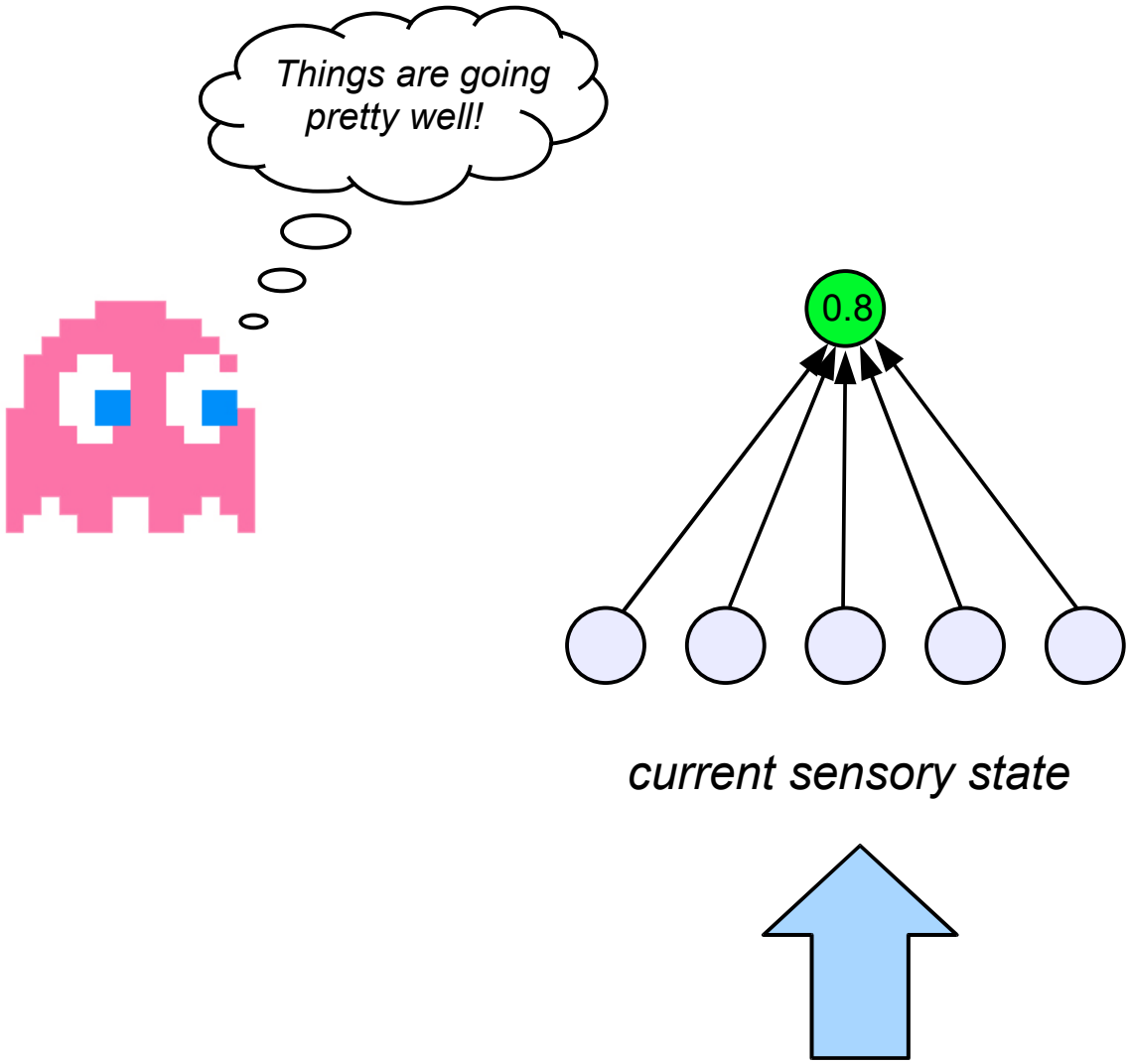
Evaluation Network



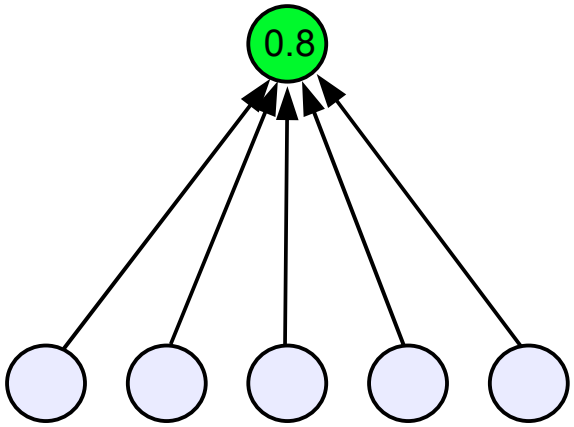
current sensory state



Evaluation Network

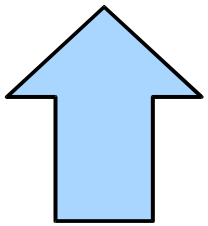


Evaluation Network

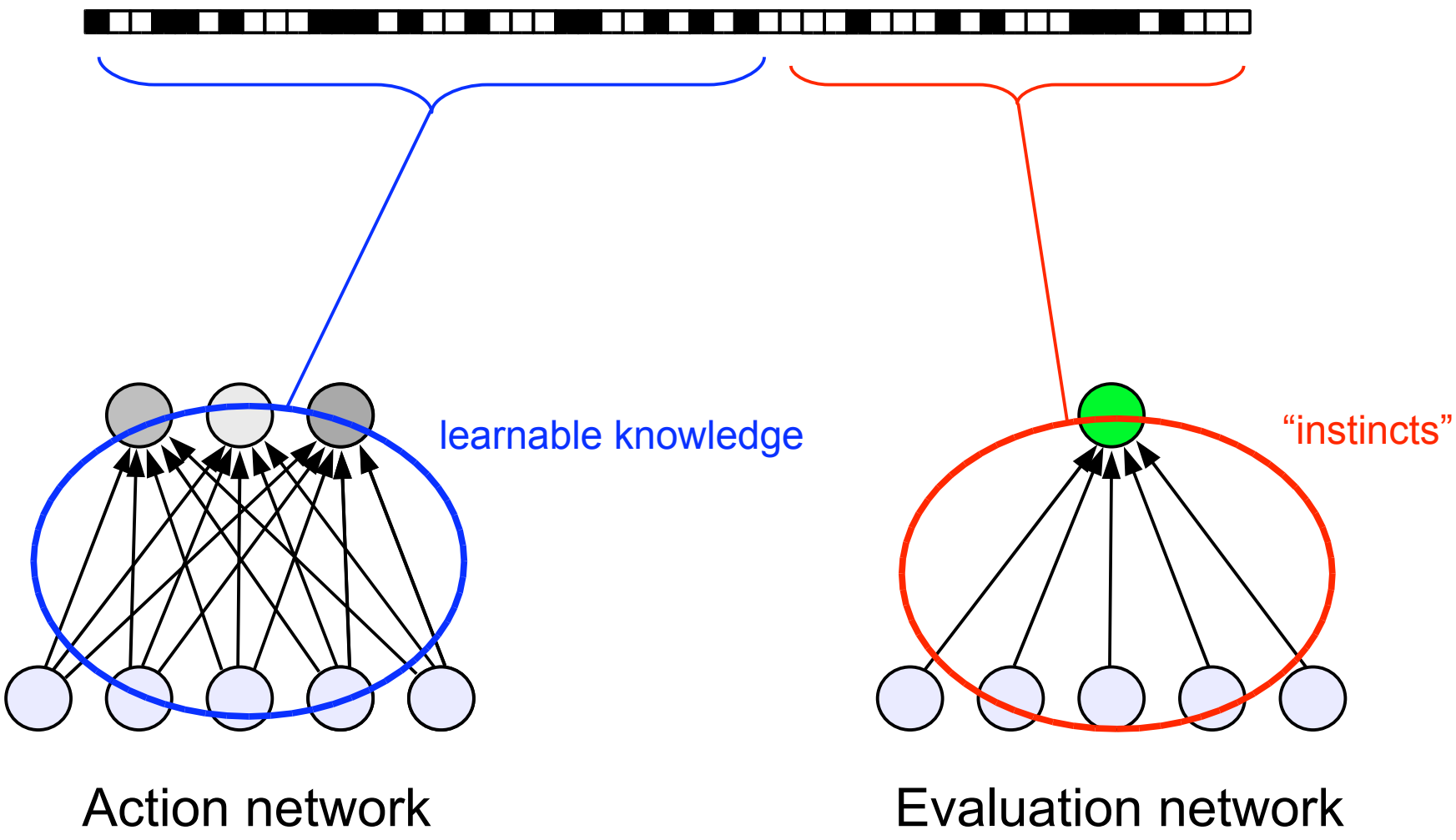


Weight values are fixed at “birth” and do not change over the agent's lifetime

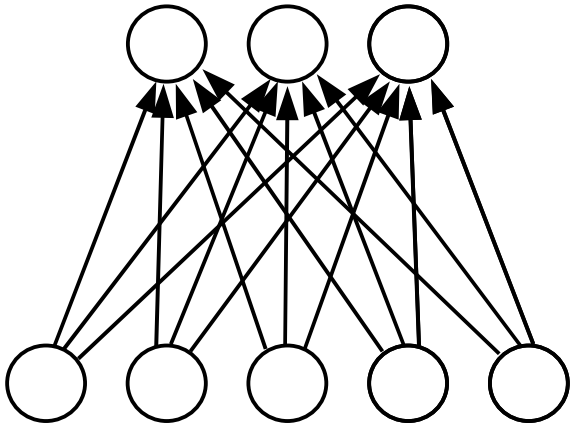
current sensory state

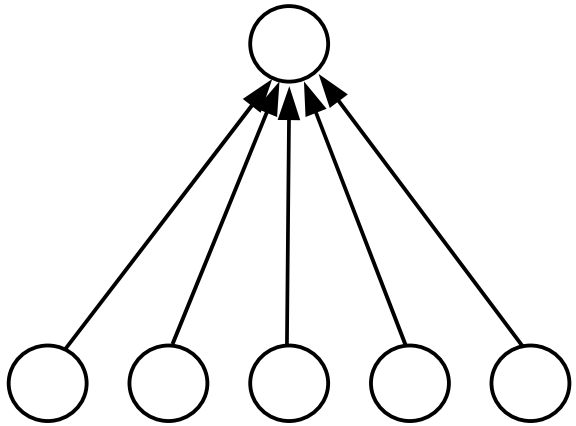


Genome

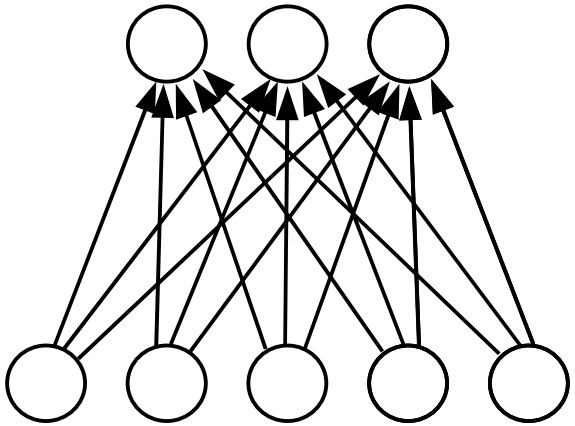

learnable knowledge
“instincts”
Action network
Evaluation network

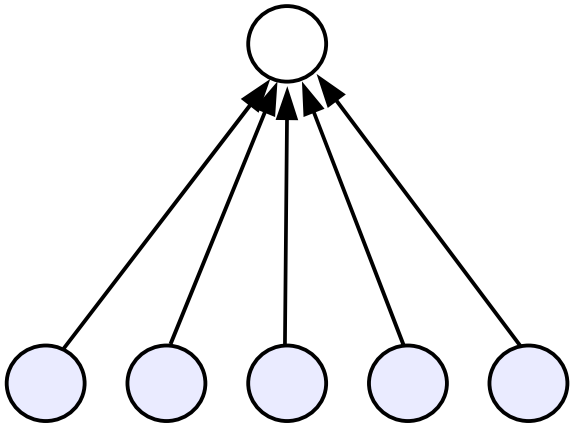
Reinforcement Learning





Reinforcement Learning

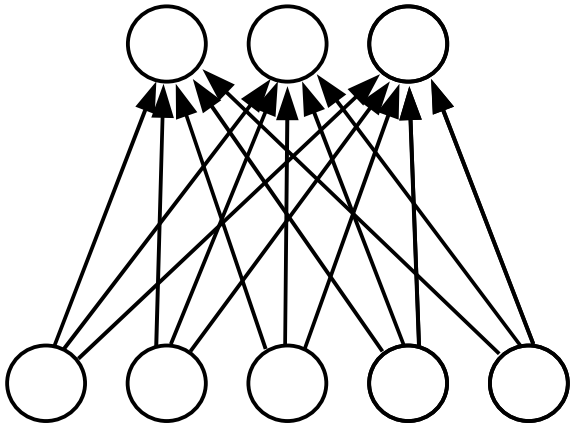


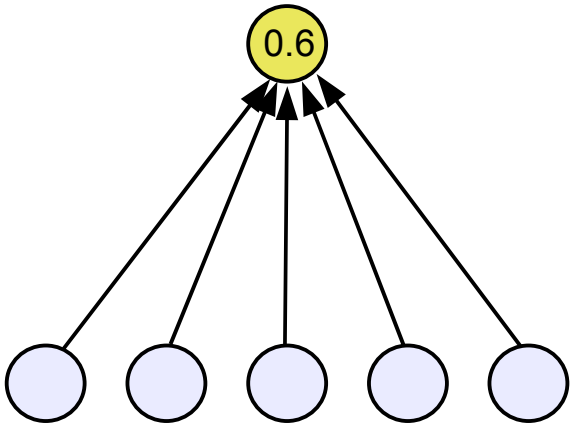


sensory state(t)

Reinforcement Learning

evaluation(t) = 0.6

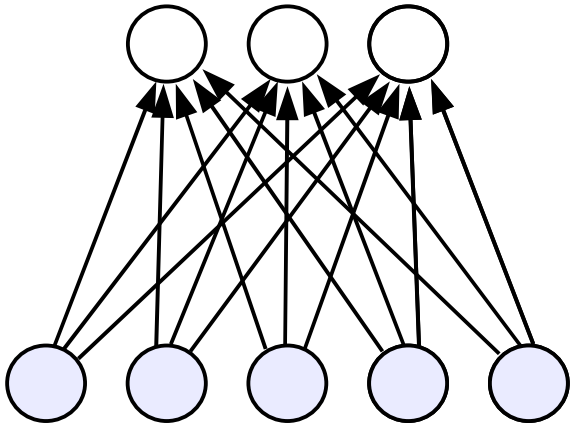




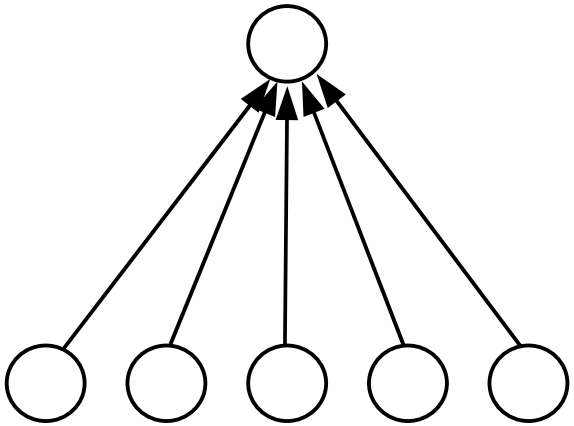
sensory state(t)

Reinforcement Learning

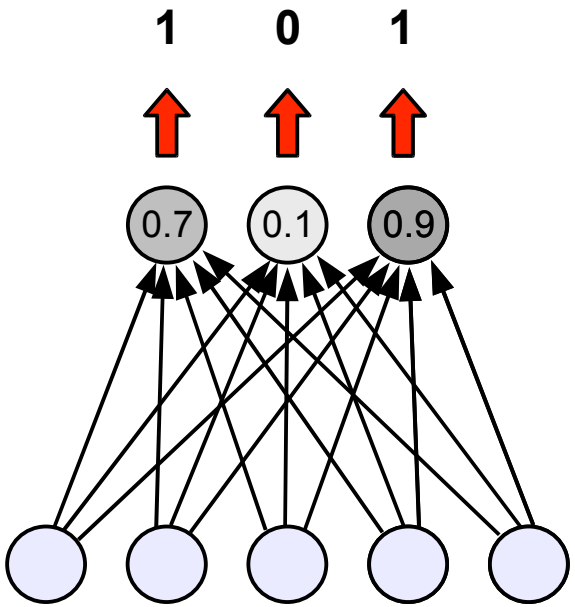
evaluation(t) = 0.6

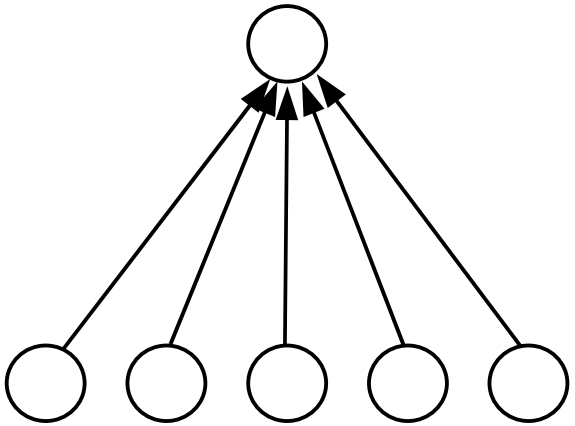


sensory state(t)



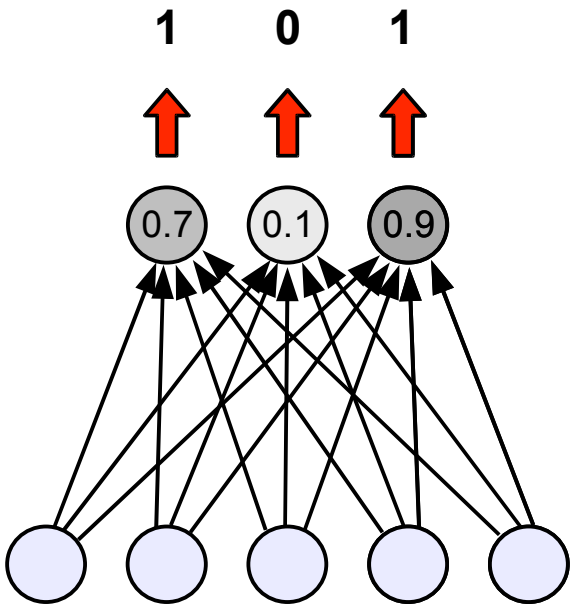
Reinforcement Learning





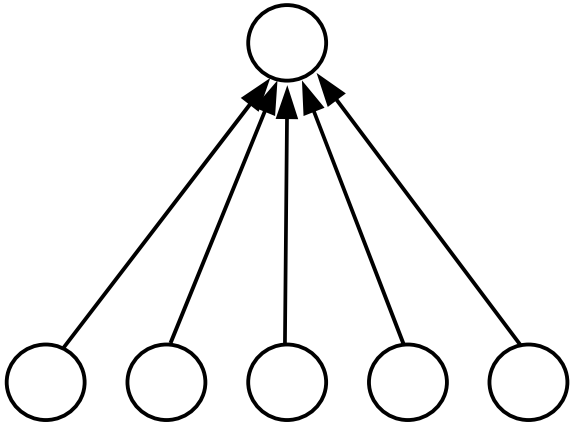
evaluation(t) = 0.6

Reinforcement Learning



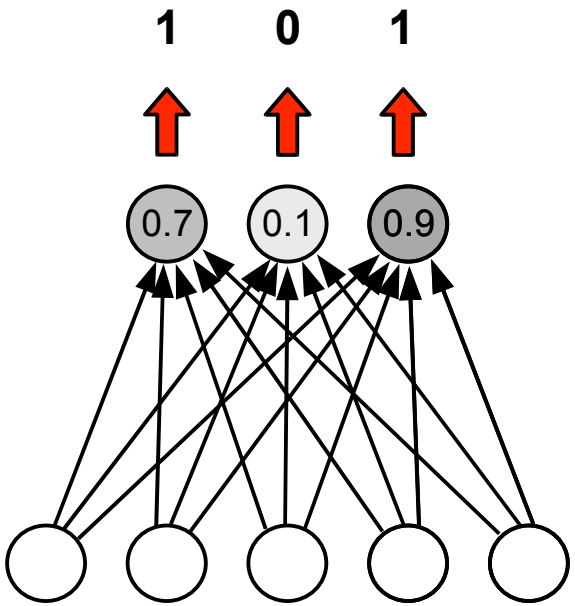
sensory state(t)

Perform action
1 0 1

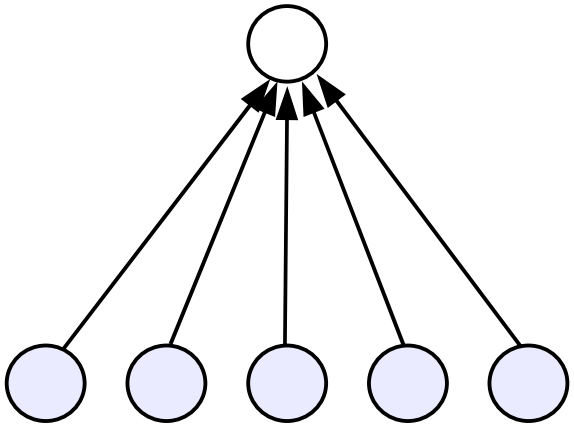


evaluation(t) = 0.6

Reinforcement Learning

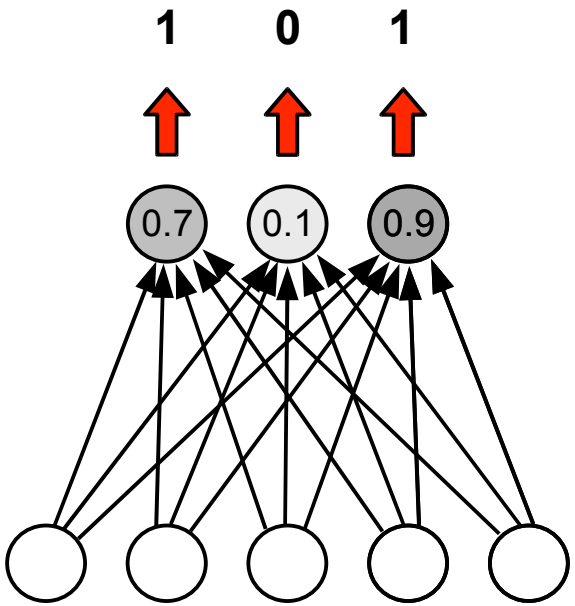


evaluation(t) = 0.6

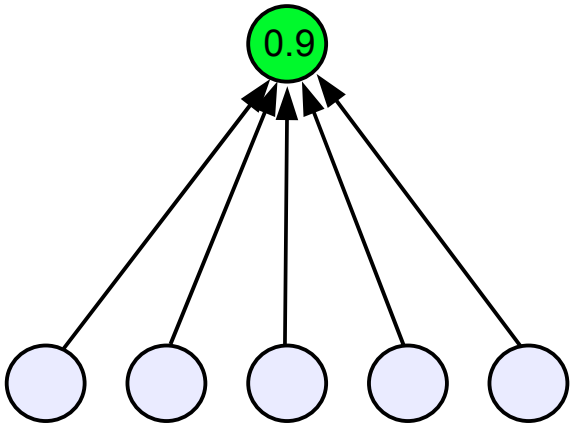


sensory state(t+1)

Reinforcement Learning



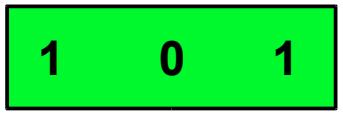
evaluation(t) = 0.6
evaluation(t+1) = 0.9



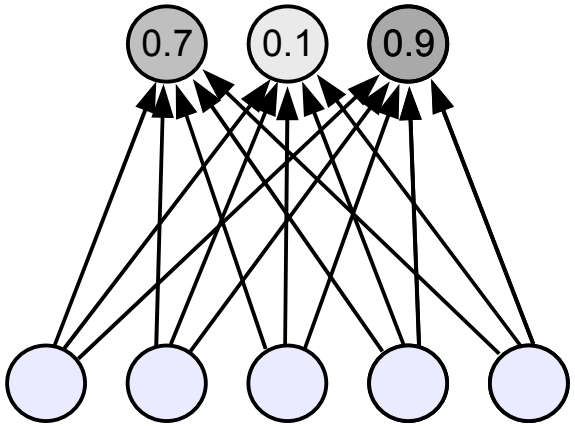
sensory state(t+1)

Reinforcement Learning

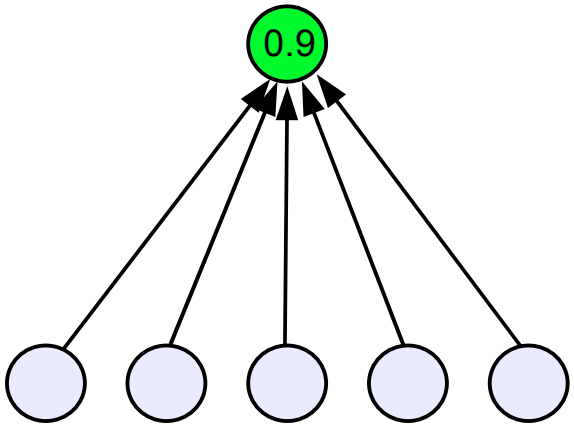
training target



evaluation(t) = 0.6
evaluation(t+1) = 0.9

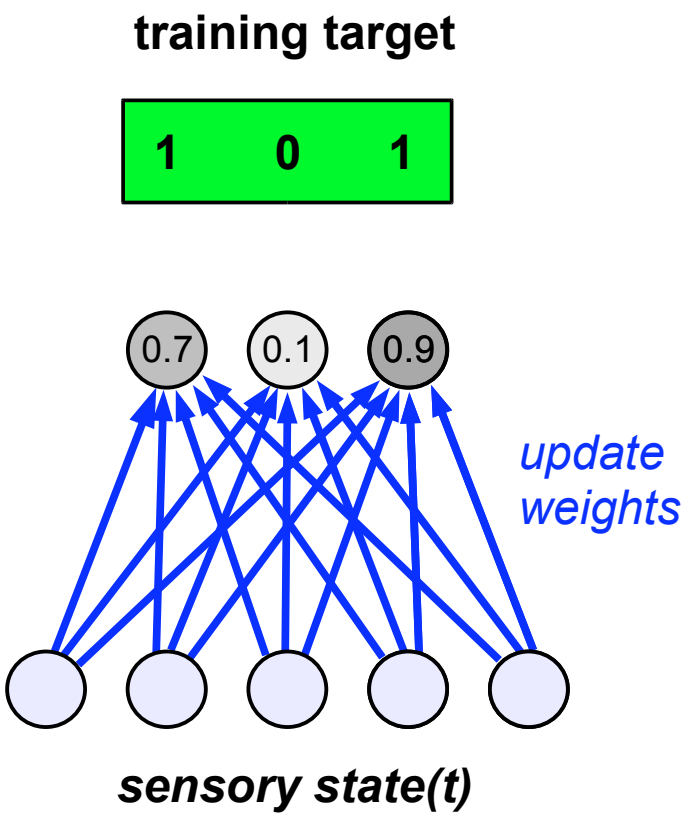


sensory state(t)

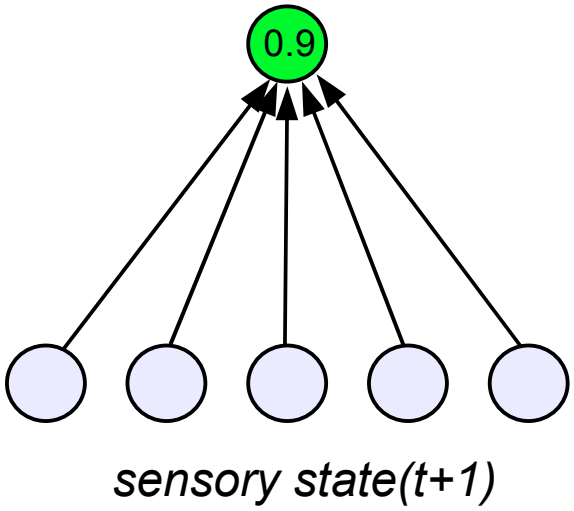


sensory state(t+1)

Reinforcement Learning

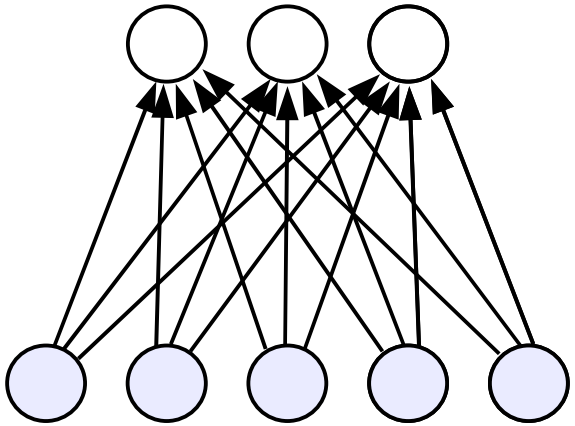


evaluation(t) = 0.6
evaluation(t+1) = 0.9

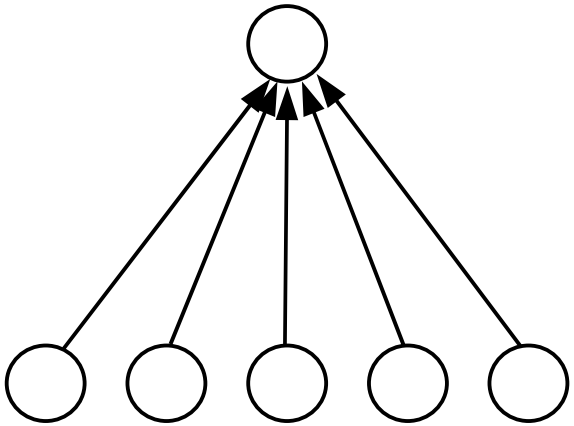


Reinforcement Learning

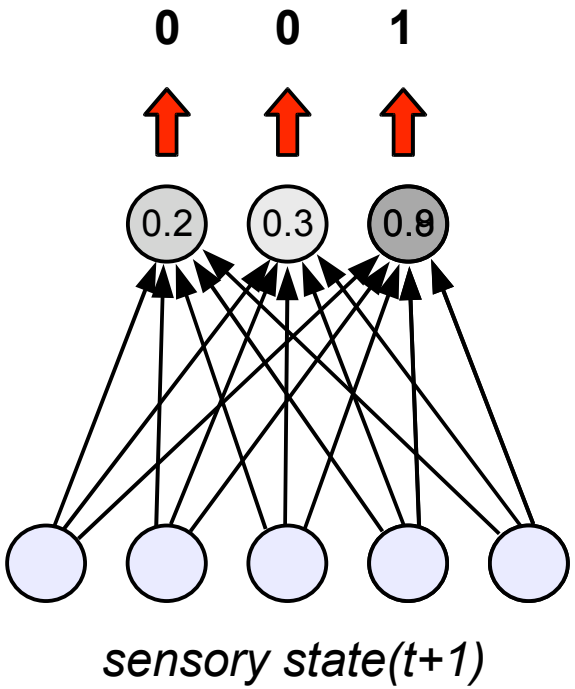
evaluation(t) = 0.6
evaluation(t+1) = 0.9



sensory state(t+1)

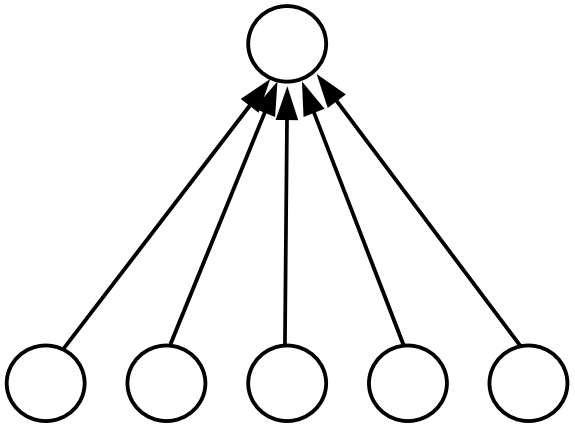


Reinforcement Learning

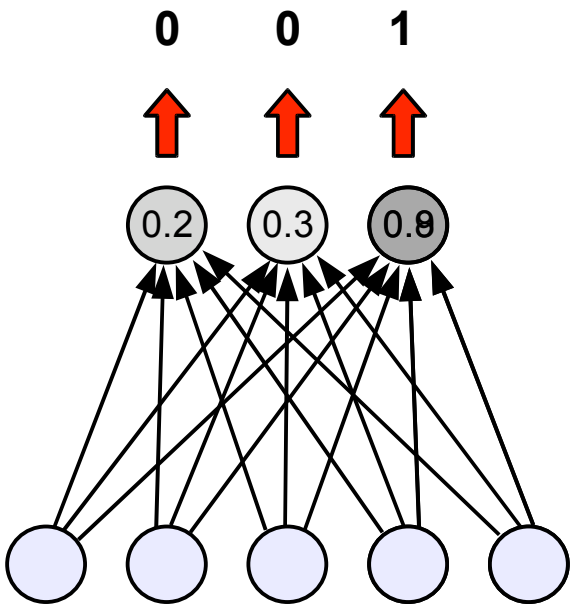


sensory state(t+1)

evaluation(t) = 0.6
evaluation(t+1) = 0.9

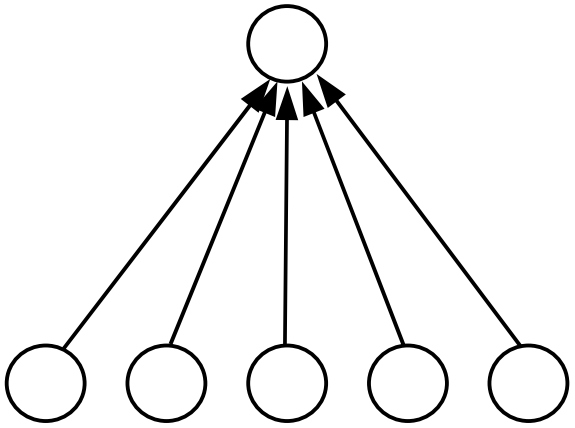


Reinforcement Learning



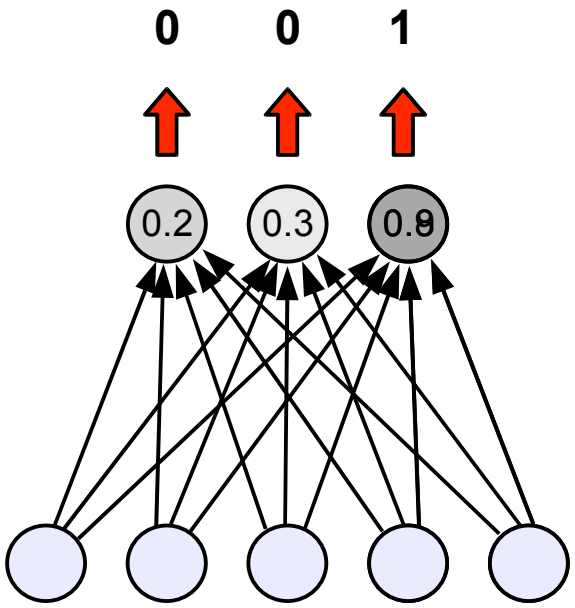
sensory state(t+1)

Perform action
0 0 1



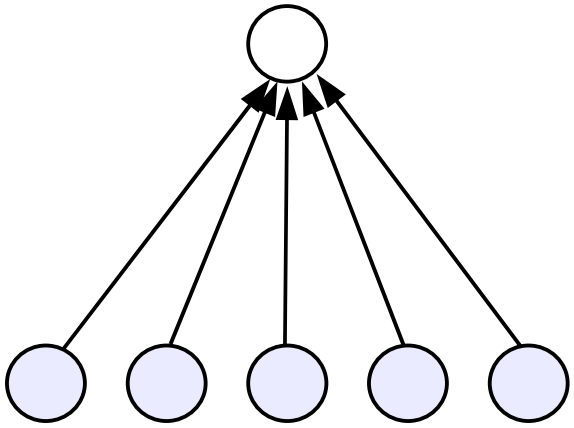
evaluation(t) = 0.6
evaluation(t+1) = 0.9

Reinforcement Learning



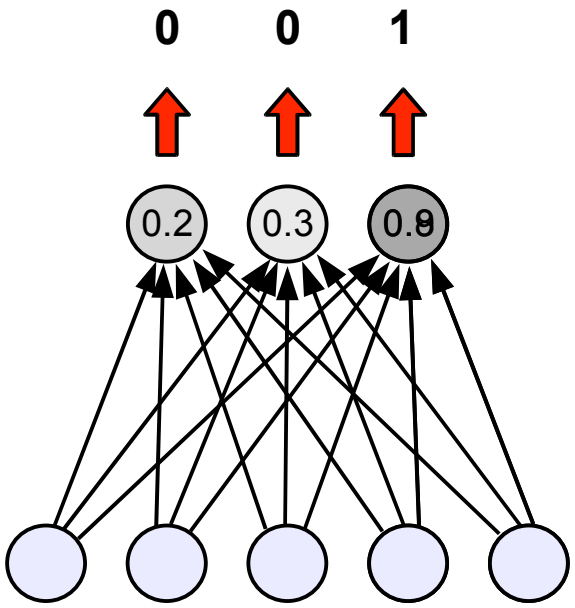
sensory state(t+1)

evaluation(t) = 0.6
evaluation(t+1) = 0.9



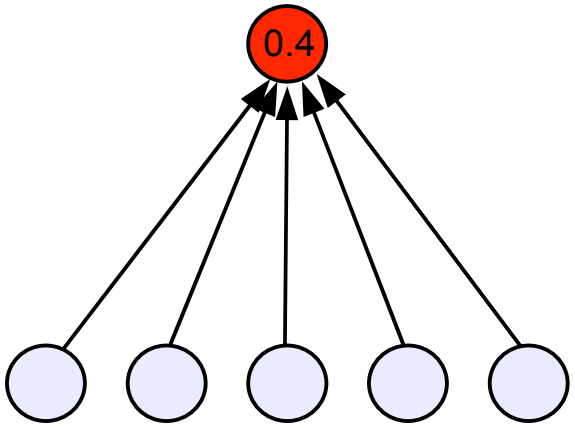
sensory state(t+2)

Reinforcement Learning



sensory state(t+1)

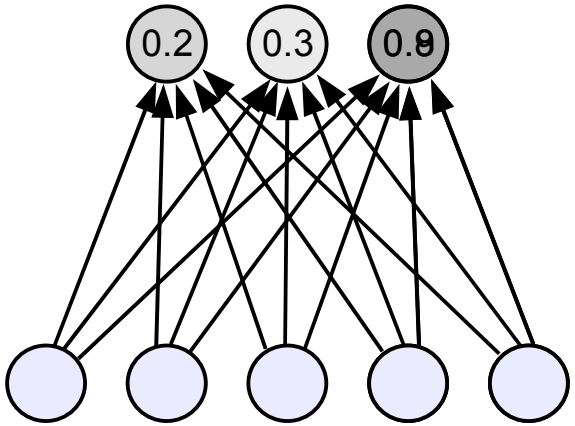
evaluation(t) = 0.6
evaluation(t+1) = 0.9
evaluation(t+2) = 0.4



sensory state(t+2)

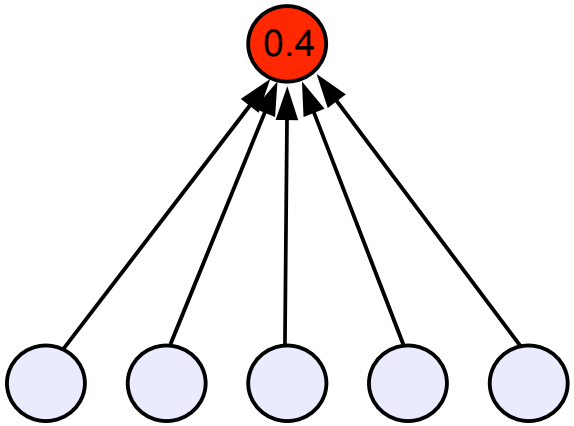
Reinforcement Learning

complement of training target



sensory state(t+1)

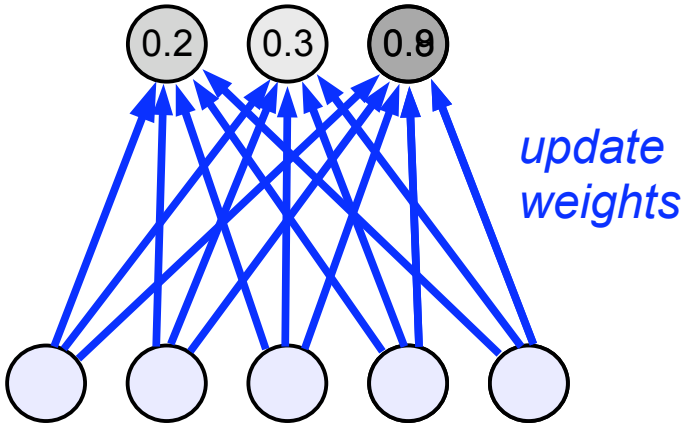
evaluation(t) = 0.6
evaluation(t+1) = 0.9
evaluation(t+2) = 0.4



sensory state(t+2)

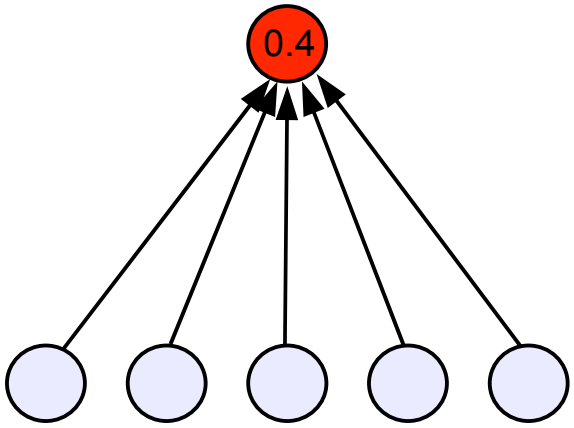
Reinforcement Learning

complement of training target



sensory state(t+1)

evaluation(t) = 0.6
evaluation(t+1) = 0.9
evaluation(t+2) = 0.4

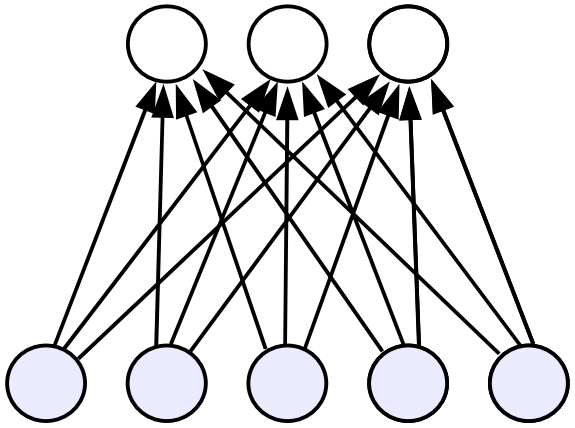


sensory state(t+2)

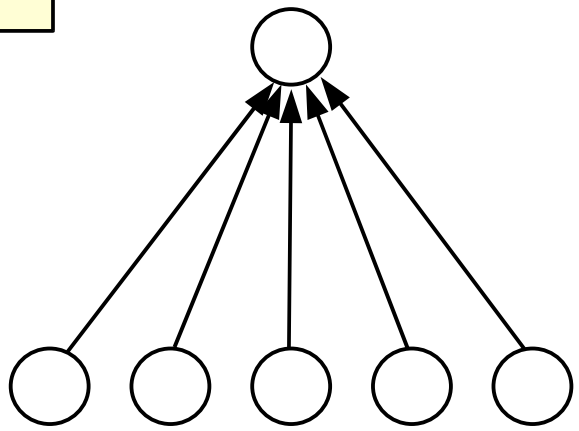
Reinforcement Learning

And so on...

evaluation(t) = 0.6
evaluation(t+1) = 0.9
evaluation(t+2) = 0.4



sensory state(t+2)

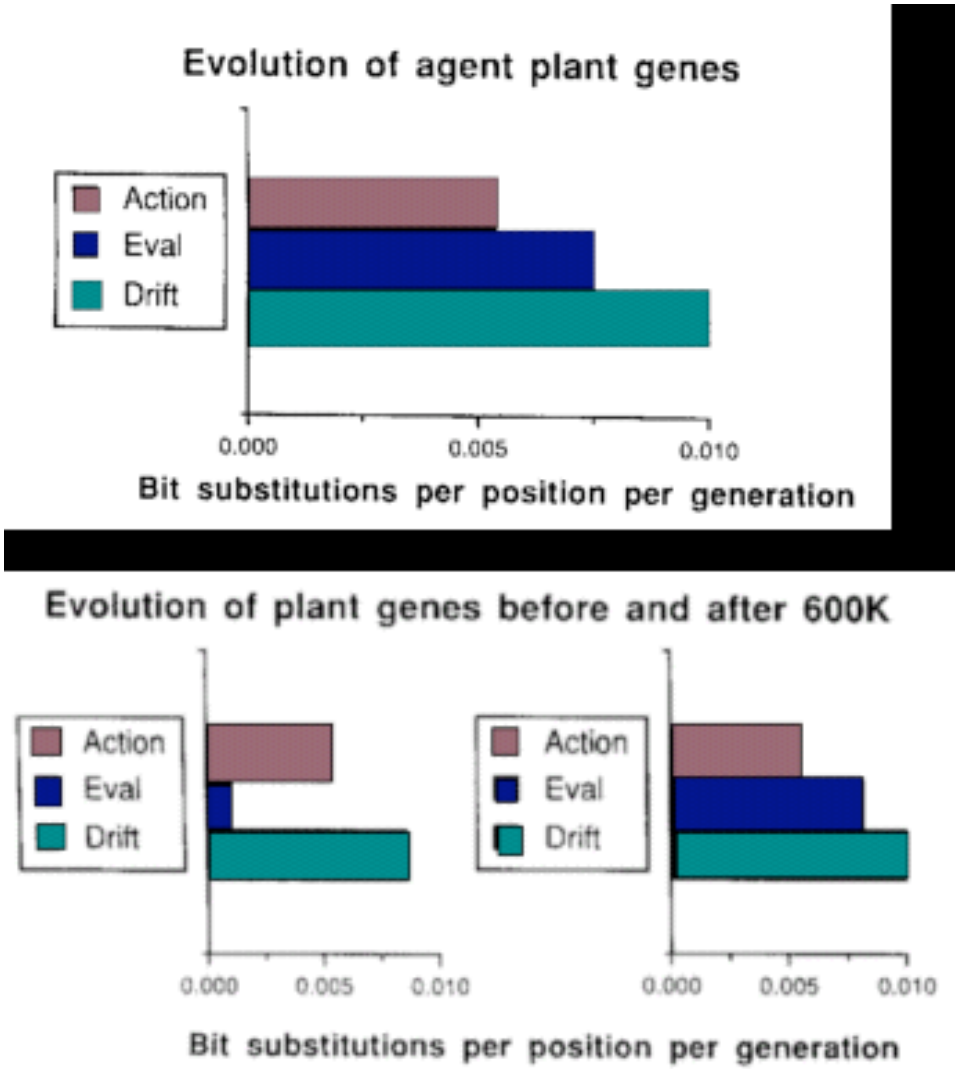


Experiments

- No explicit fitness function used
- Agents simply survived for as long as possible
- Ackley & Littman measured the **time to extinction** of various populations of agents
- Compared the performance of five types of agents
  - random movements (“Brownian” agents)
  - fixed random network weights
  - evolution only
  - learning only
  - evolution + learning (ERL)

Experiments

- Also examined the rate of change of different genes over time
- Analyzed genome via “functional constraints”
- Looked at genes associated with finding and evaluating food
- Early in a run, evaluation genes were more constrained
- Later in a run, action genes were more constrained
- Learned knowledge gradually replaced by innate knowledge
- Evidence for Baldwin Effect?




ERL Video