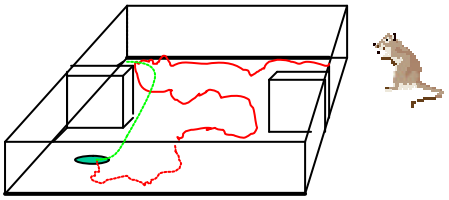
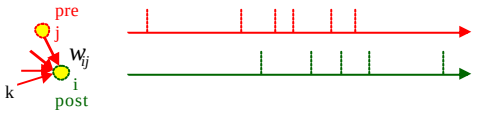
Unsupervised vs. reinforcement learning (via a model of rat navigation)



Hebbian Learning

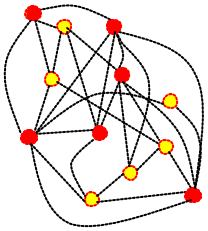


*When an axon of cell **j** repeatedly or persistently takes part in firing cell **i**, then j's efficiency as one of the cells firing i is increased*

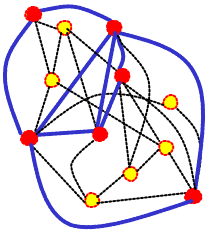
Hebb, 1949

- local rule
- simultaneously active (correlations)

Hebbian Learning



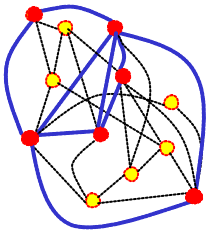
Hebbian Learning



item memorized

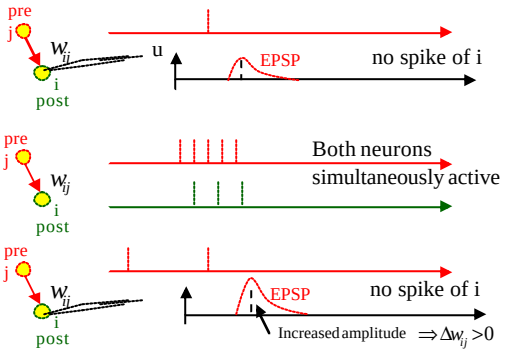
Hebbian Learning

Recall:
Partial info

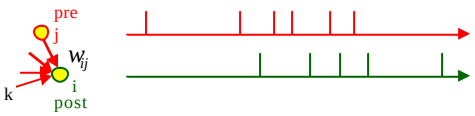


item recalled → *Associative memory*

Hebbian Learning in experiments (schematic)



Hebbian Learning



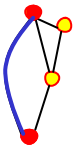
*When an axon of cell **j** repeatedly or persistently takes part in firing cell **i**, then j's efficiency as one of the cells firing i is increased*

Hebb, 1949

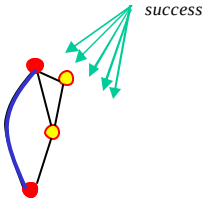
- local rule
- simultaneously active (correlations)

Rate model:
active = high rate = many spikes per second

Hebbian Learning
= unsupervised learning



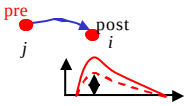
Reinforcement Learning
= reward + Hebb



Classification of plasticity: unsupervised vs reinforcement

LTP/LTD/Hebb

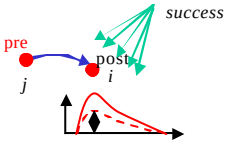
Theoretical concept
- *passive changes*
- *exploit statistical correlations*



Functionality
-useful for development
(wiring for receptive fields)

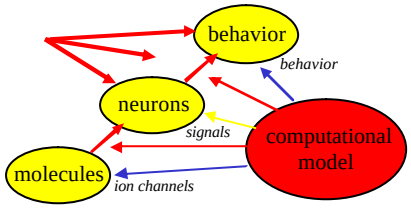
Reinforcement Learning

Theoretical concept
- *conditioned changes*
- *maximise reward*



Functionality
- useful for learning a new behavior

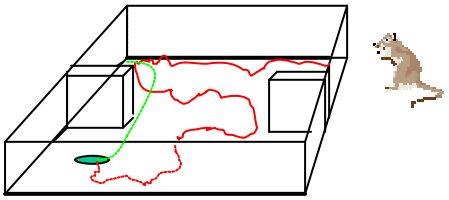
Introduction to reinforcement learning (via a model of rat navigation)



Introduction to reinforcement learning (via a model of rat navigation)

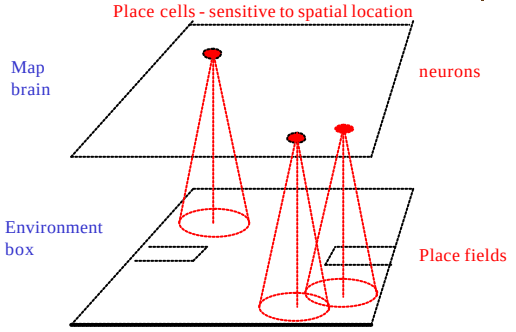
- -Basics or rat navigation
- -Place cells and Rat hippocampus
- -A model of spatial representation
- -Learning to find the goal location
- -Reward based learning
- -Reinforcement learning theory
- -Eligibitiy traces
- -the full model: behavioral experiments

Biological Principles of Learning: spatial learning

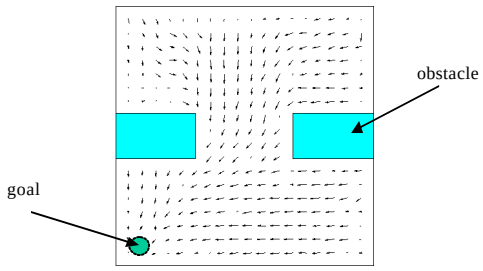


Spatial representation

Place cells - sensitive to spatial location



Use map



SPATIAL REPRESENTATION

- ***Model of place cells***

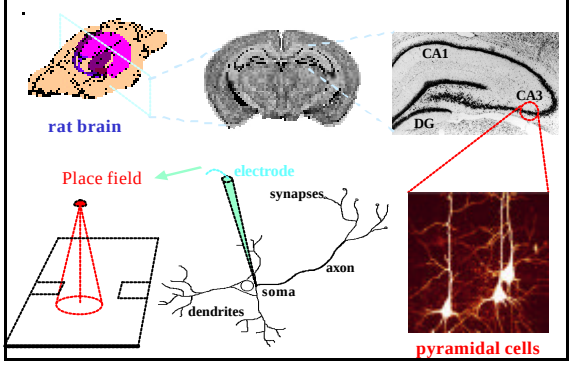
GOAL LEARNING

- ***Reward-based learning system***

Introduction to reinforcement learning (via a model of rat navigation)

- -Basics or rat navigation
- -Place cells and Rat hippocampus
- -A model of spatial representation
- -Learning to find the goal location
- -Reward based learning
- -Reinforcement learning theory
- -Eligibitiy traces
- -the full model: behavioral experiments

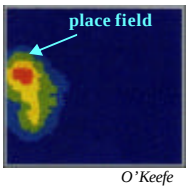
Neurophysiology of the Rat Hippocampus



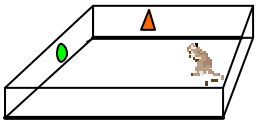
Hippocampal Place Cells

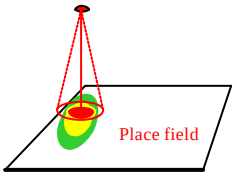
Depends on

- visual cues
- works also in the dark

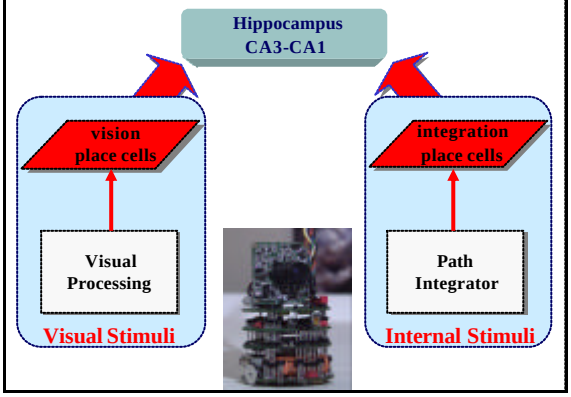


O'Keefe

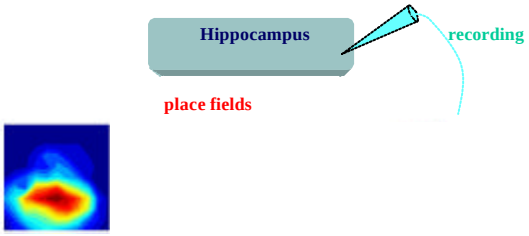




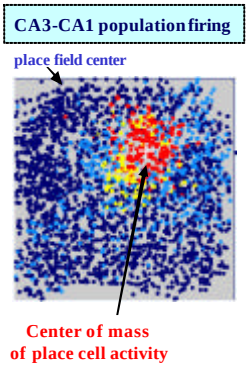
The Neural Model: Place Cells



The Neural Model: Place Cells



The Neural Model: Place Cells



√ ***SPATIAL REPRESENTATION***

- ***Model of place cells***

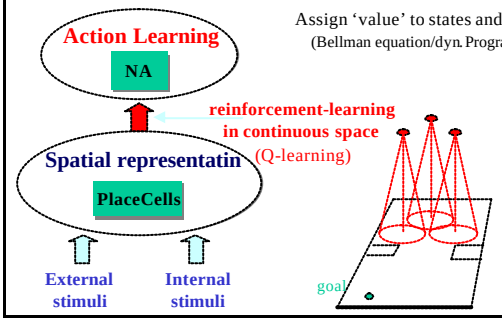
→ ***GOAL LEARNING***

- ***Reward-based learning system***

Reward-based Action Learning

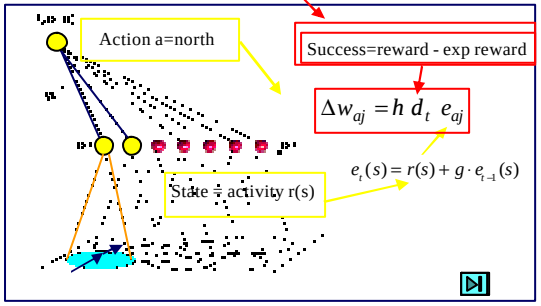
Hippocampal place cells = fuzzy discretisation of continuous space

Assign 'value' to states and actions
(Bellman equation/dyn. Programming)



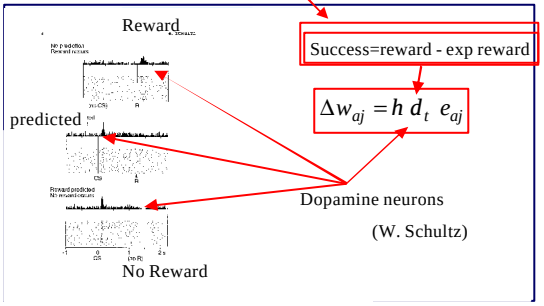
Reward-based Action Learning

Connection reinforced if
action a at state s successful



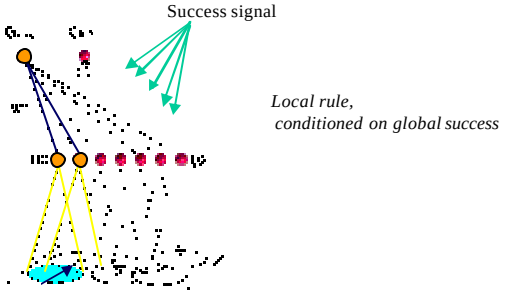
Reward-based Action Learning

Connection reinforced if
action a at state s successful



Reward-based Action Learning

Connection reinforced if
action a at state s successful



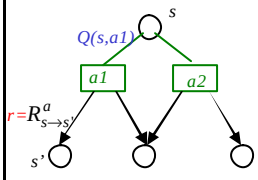
Local rule,
conditioned on global success

Introduction to reinforcement learning (via a model of rat navigation)

- -Basics or rat navigation
- -Place cells and Rat hippocampus
- -A model of spatial representation
- -Learning to find the goal location
- -Reward based learning (basic ideas)
- → -Reinforcement learning theory
- -Eligibitiy traces
- -the full model: behavioral experiments

Reward-based Action Learning

Connection reinforced if
action a at state s successful



s=state

a2 a=action

$R^a_{s \to s'}$ = reward

s'=new action

$Q(s,a) = \text{expected reward}$

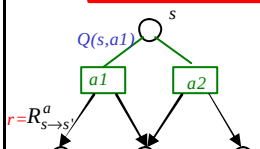
$Q(s,a) = \sum_{s'} P^a_{s \to s'} R^a_{s \to s'}$

Iterative Update

$DQ(s,a) = h\ [r - Q(a)]$

Blackboard :

Exercise now: Iterative update



s=state

a2 a=action

$R^a_{s \to s'}$ = reward

s'=new action

a) Show that an empirical evaluation of Q(s,a) by averaging the rewards for action a over k or (k+1) trials, leads to an iterative update rule of the form $DQ(s,a) = h\ [r - Q(a)]$

b) Calculate eta.

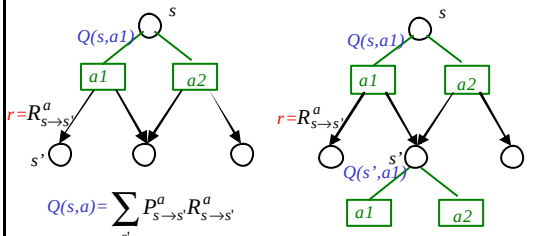
c) Give an intuitive explanation of the update rule

$Q(s,a) = \sum_{s'} P^a_{s \to s'} R^a_{s \to s'}$ *expected reward*

Reward-based Action Learning

Connection reinforced if action a at state s successful

Blackboard :



$$Q(s,a)=\sum_{s'} P^a_{s\to s'} R^a_{s\to s'}$$

Update

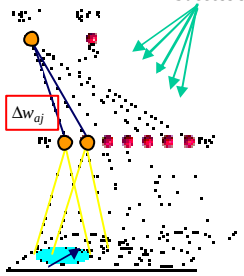
$DQ(s,a)=h\ [r-Q(s,a)]$

$DQ(s,a)=h\ [r-(Q(s,a)-Q(s',a'))]$

Reward-based Action Learning

Connection reinforced if action a at state s successful

Success signal



Function approximation

$$Q_w(s,a)=\sum_{i=1}^{n} w_{ai}\cdot r_i(s)$$

Update of Q = update of weights

$DQ(s,a)=h\ [r-(Q(s,a)-Q(s',a'))]$

d

Synaptic update of current action *a*

$\Delta w_{aj} = h\ d_t\ r_j$ *if a taken*

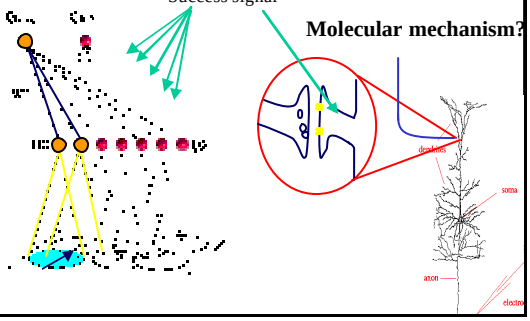
$\Delta w_{aj} = h\ d_t\ r_j\ r_a$

Reward-based Action Learning

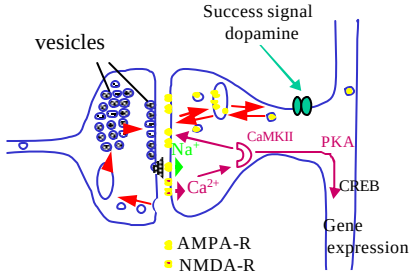
Connection reinforced if action a at state s successful

Success signal

Molecular mechanism?



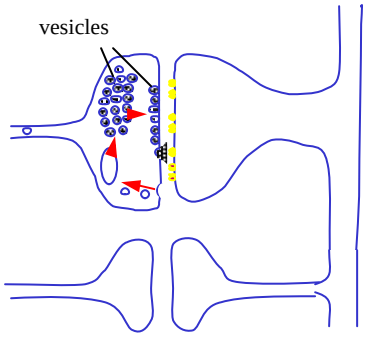
Changes in synaptic connections



CaMKII = Ca-Calmodulin dependent Protein Kinase2

PKA = cAMPdependent protein kinase

Changes in synaptic connections



Reward-based Learning TD( )

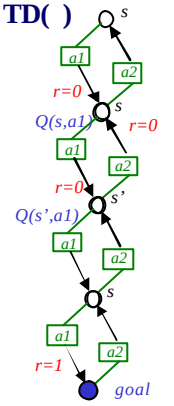
- ***SARSA***

$DQ(s,a)=h\ [r-(Q(s,a)-Q(s',a'))]$

- ***policy for action choice:***
  Pick most often action

$$a^*_t = \arg\max_a Q_a(s,a)$$

Blackboard :



Exercise now

- ***Update of Q values in SARSA***

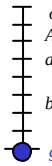
$DQ(s,a) = h\ [r-(Q(s,a)-Q(s',a'))]$

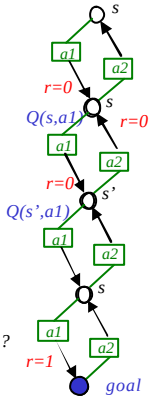
- ***policy for action choice:***
  Pick most often action

$a_t^* = \arg\max_a Q_a(s,a)$

Consider a linear sequence of states. Reward only at goal. Actions are up or down.

a) *Initialise Q values at 0. Start at top. How do Q values develop?*
b) *Q values after 3 complete trials?*





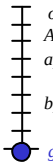
Problem: learning is slow

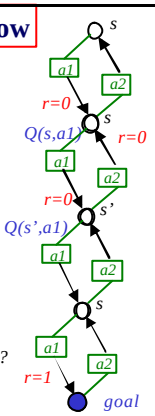
- Slow diffusion of information across several states

$a_t^* = \arg\max_a Q_a(s,a)$

Consider a linear sequence of states. Reward only at goal. Actions are up or down.

a) *Initialise Q values at 0. Start at top. How do Q values develop?*
b) *Q values after 3 complete trials?*





Introduction to reinforcement learning (via a model of rat navigation)

-Basics or rat navigation
-Place cells and Rat hippocampus
-A model of spatial representation
-Learning to find the goal location
-Reward based learning (basic ideas)
-Reinforcement learning theory
→ -Eligibitiy traces
-the full model: behavioral experiments

Reward-based Learning TD( )

- ***TD error in SARSA***

$d_t = R_{t+1} - [Q_a(s',a') - g \cdot Q_a(s,a)]$

- ***policy for action choice:***
  Pick most often action

$a_t^* = \arg\max_a Q_a(s,a)$

- ***Eligibility trace (memory at synapse):***

$e_{aj}(t) = r_j\, r_a + \begin{cases} g\, l\, e_{aj}(t) & \text{if exploiting} \\ 0 & \text{if exploring} \end{cases}$

pre post memory

- **Function approximation**

$Q_w(s,a) = \sum_{i=1}^{n} w_i^a \cdot r_i(s)$

- **Synaptic update**

$\Delta w_{aj} = h\, d_t\, e_{aj}$

pre post

Reward-based Action Learning

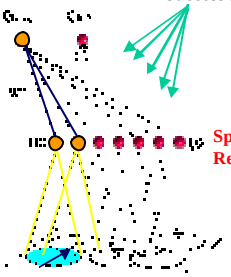
Connection reinforced if action a at state s successful

Success signal

Learning Rule

$\Delta w_{aj} = h\, d_t\, e_{aj}$

Spatial Representation

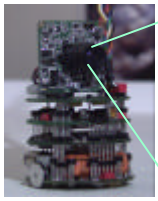


Introduction to reinforcement learning (via a model of rat navigation)

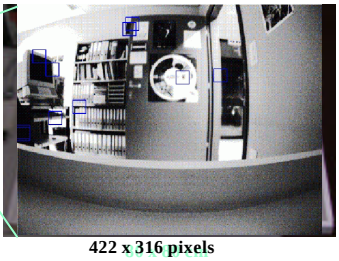
-Basics or rat navigation
-Place cells and Rat hippocampus
-A model of spatial representation
-Learning to find the goal location
-Reward based learning (basic ideas)
-Reinforcement learning theory
-Eligibitiy traces
→ -the full model: behavioral experiments

Validating the Model

The KHEPERA mobile miniature robot



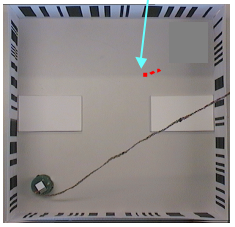
Experimental arena



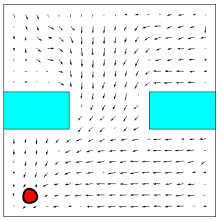
422 x 316 pixels

Open-field Navigation Experiments

robot trajectory



(Biol. Cybern., 2000)



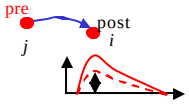
Navigation map after 20 training trials

Classification of plasticity: unsupervised vs reinforcement

LTP/LTD/Hebb

Theoretical concept

- *passive changes*
- *exploit statistical correlations*



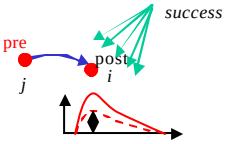
Functionality

-useful for development
(wiring for receptive fields)

Reinforcement Learning

Theoretical concept

- *conditioned changes*
- *maximise reward*

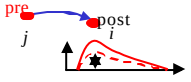


Functionality

- useful for learning a new behavior

Plasticity models: unsupervised vs reinforcement

STDP/Hebb



Model

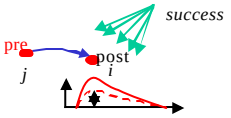
- STDP (see above)

$\Delta w_{ij} \propto pre \cdot post + other\ terms$

Reinforcement Learning

theoretical Protocol

- *maximise reward*

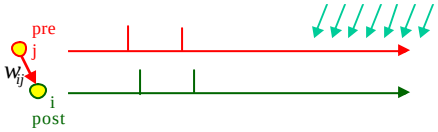


Model

- spike-based model?

$\Delta w_{ij} \propto success \cdot (pre \cdot post + other terms)$

Timing issues in Reinforcement Learning



Success signal is delayed

Spike time scale 1-10 ms
Reward delay 100-5000 ms

→Need memory trace (eligibility trace)