OIST Course B26

Introduction to Neuroscience

2019.12.12

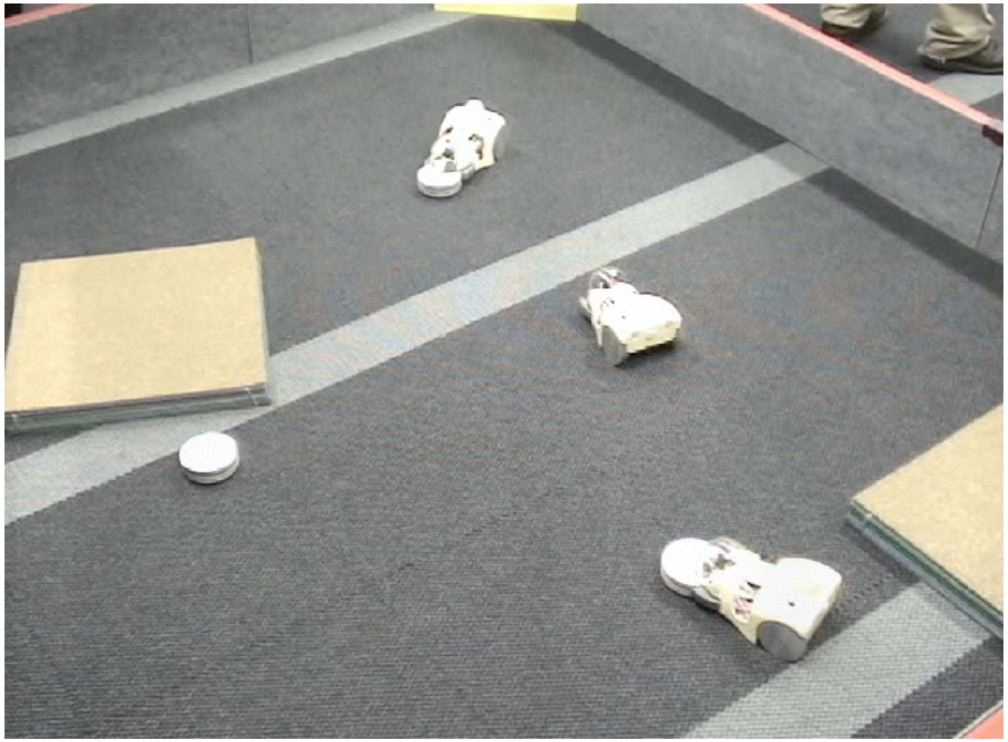
Reinforcement Learning and Unsupervised Learning

Kenji Doya
doya@oist.jp
Neural Computation Unit
Okinawa Institute of Science and Technology

OIST Neural Computation Unit

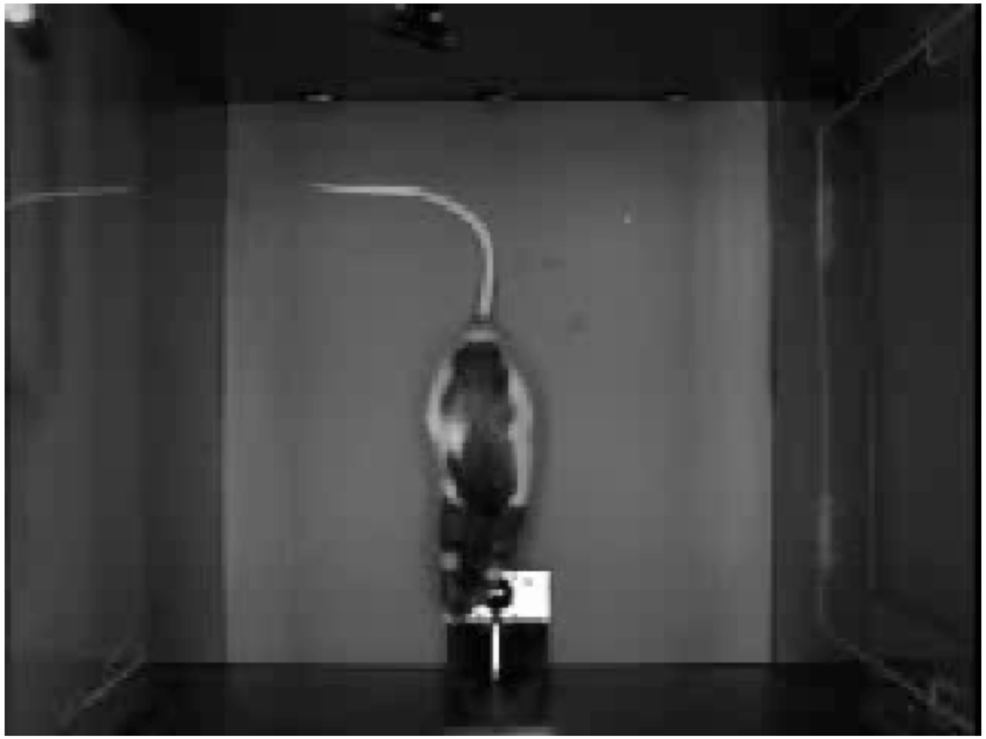
How to build adaptive, autonomous systems

- robot experiments

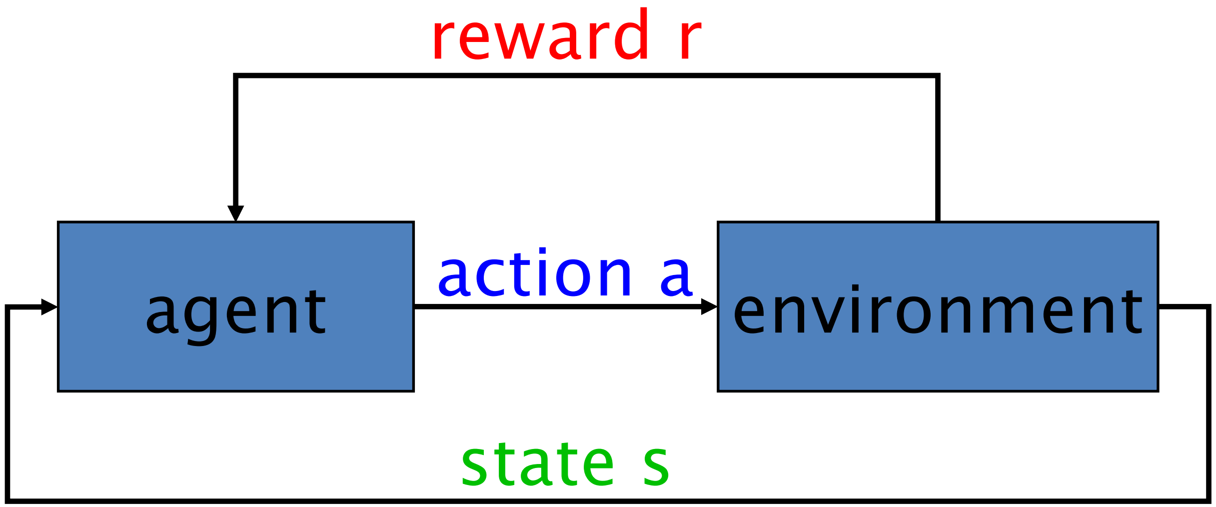


How the brain realizes robust, flexible learning

- neurobiology



Reinforcement Learning



Learn action policy: s → a to maximize rewards

- Efficient algorithms for artificial agents
- Circuit and molecular mechanisms in the brain

Outline

AI and Brain Science

Reinforcement Learning

Basal Ganglia

Unsupervised Learning

Mental Simulation

AI and Brain Science

To make intelligent machines by electronics,
we should not bother biological constraints.

As there's a superb implementation
in the brain, we should learn from that.

AI in 20th century: program human expertise

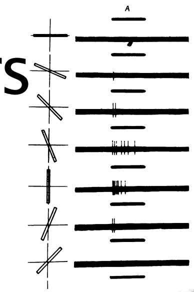
AI in 21st century: learn from big data

Brain-like implementation like *Deep Learning*
gives the best performance.

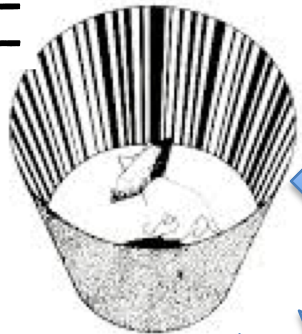
Coevolution in Pattern Recognition

Brain Science

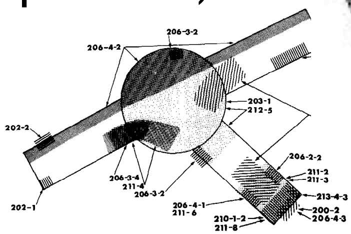
- Feature detectors
  (Hubel & Wiesel 1959)



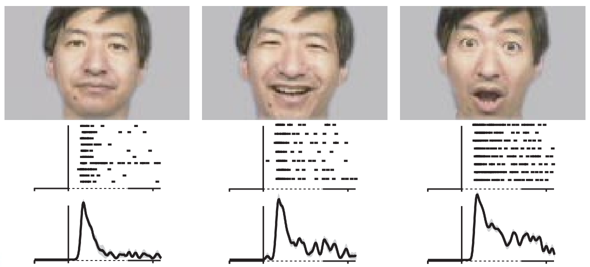
- Experience dependence
  (Blakemore & Cooper 1970)



- Place cell
  (O'Keefe 1976)

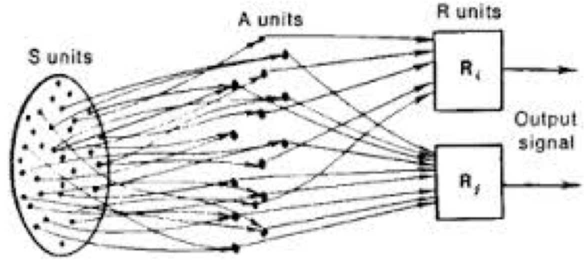


- Face cell (Bruce, Desimone, Gross 1981)

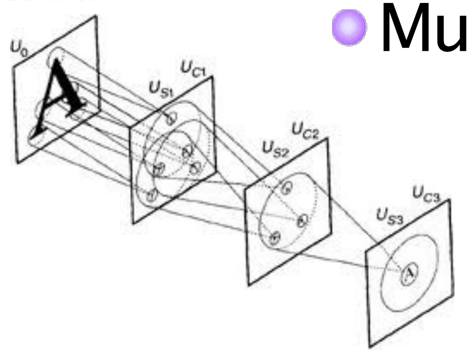


(Sugase et al. 1999)

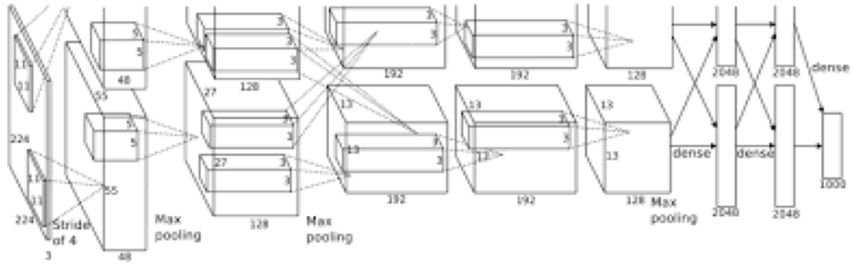
Artificial Intelligence



- Perceptron
  (Rosenblatt 1962)



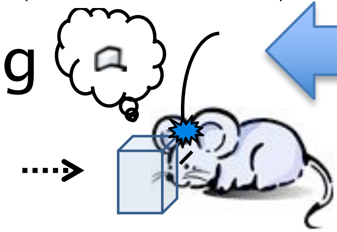
- Multi-layer learning
  (Amari, 1967)
- Neocognitron
  (Fukushima 1980)
- ConvNet (Krizhevsky, Sutskever, Hinton, 2012)
- GoogleBrain (2012)



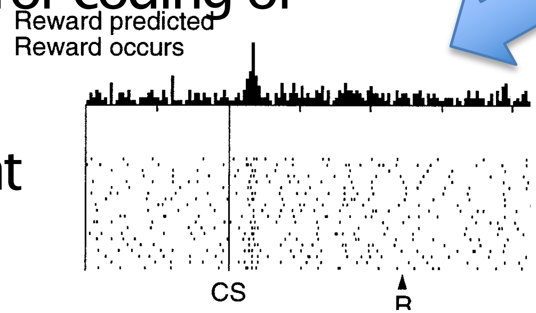
Co-Evolution: Reinforcement Learning

Brain Science

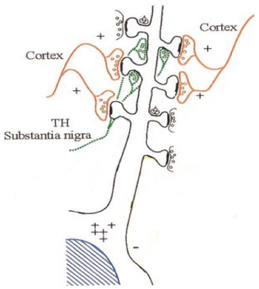
- Classic conditioning (Pavlov 1903)
- Operant conditioning
  (Thorndike 1898,
  Skinner 1938)

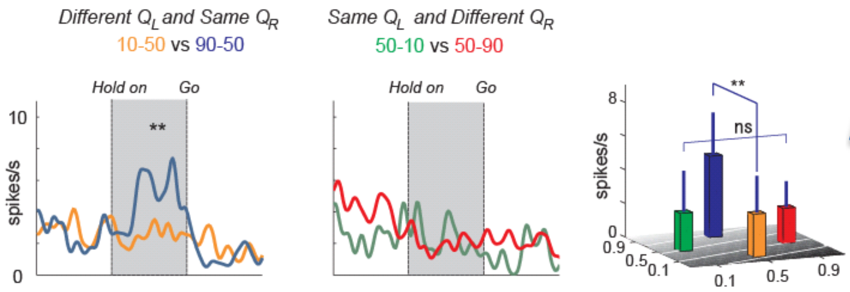


- Reward prediction error coding of dopamine neurons
  (Schultz et al. 1993, 1997)



- Dopamine-dependent synaptic plasticity
  (Wickens et al. 2000)

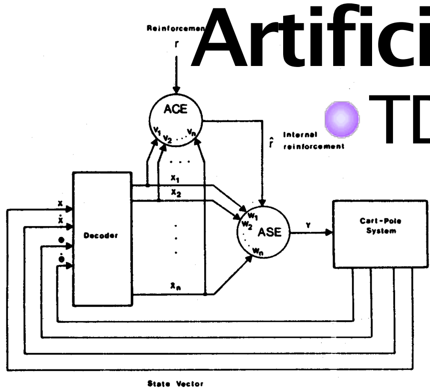




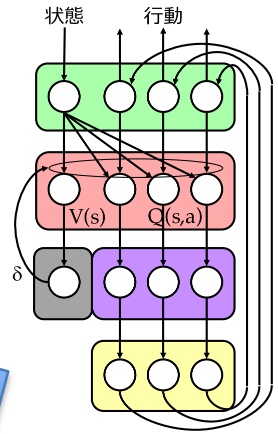
- Value coding in striatum (Samejima et al. 2005)

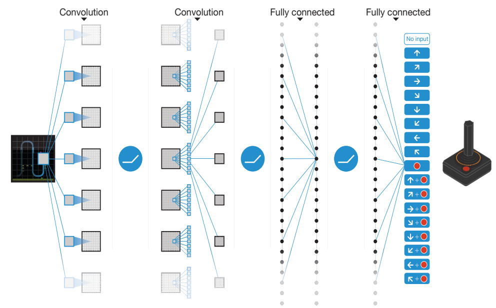
Artificial Intelligence

- TD learning (Barto et al. 1983)



- Dopamine TD learning hypothesis
  (Barto et al. 1995,
  Montague et al. 1996)

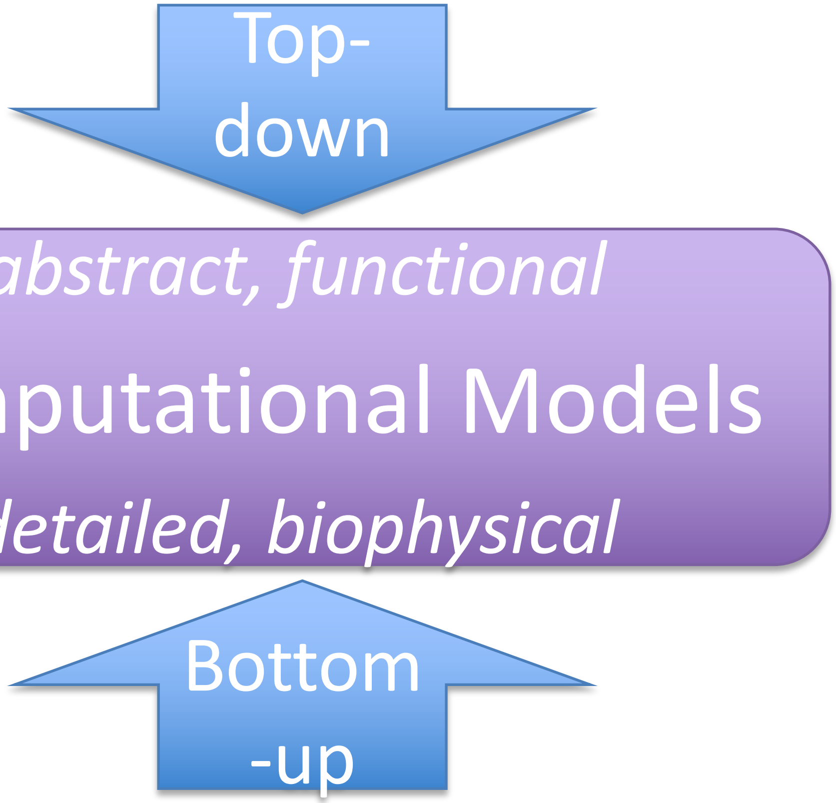




- Deep Q network (Mnih et al. 2015)

Brain Theory Meets Big Data

How the Brain Should Be Working



What the Brain Really Is

Machine Learning is Essential for Both Directions

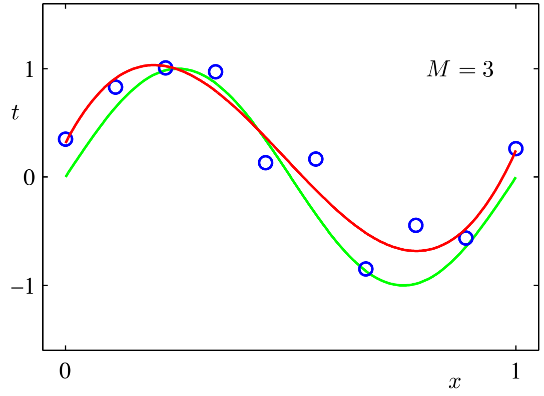
Three Classes of Machine Learning

Supervised Learning

- Input-output pairs $\{(x_1,y_1), (x_2,y_2),\ldots\}$

  $\rightarrow$ input-output model $y = f(x) + \varepsilon$

  - for new input x, predict output y



Reinforcement Learning

- state-action-reward triplets $\{(x_1,y_1,r_1), (x_2,y_2,r_2),\ldots\}$

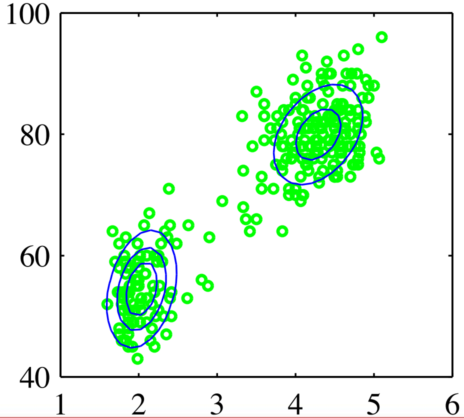
  $\rightarrow$ action policy $y = f(x)$ to maximize reward

(unsupervised) Representation Learning

- Input data $\{ x_1, x_2, x_3,\ldots\}$

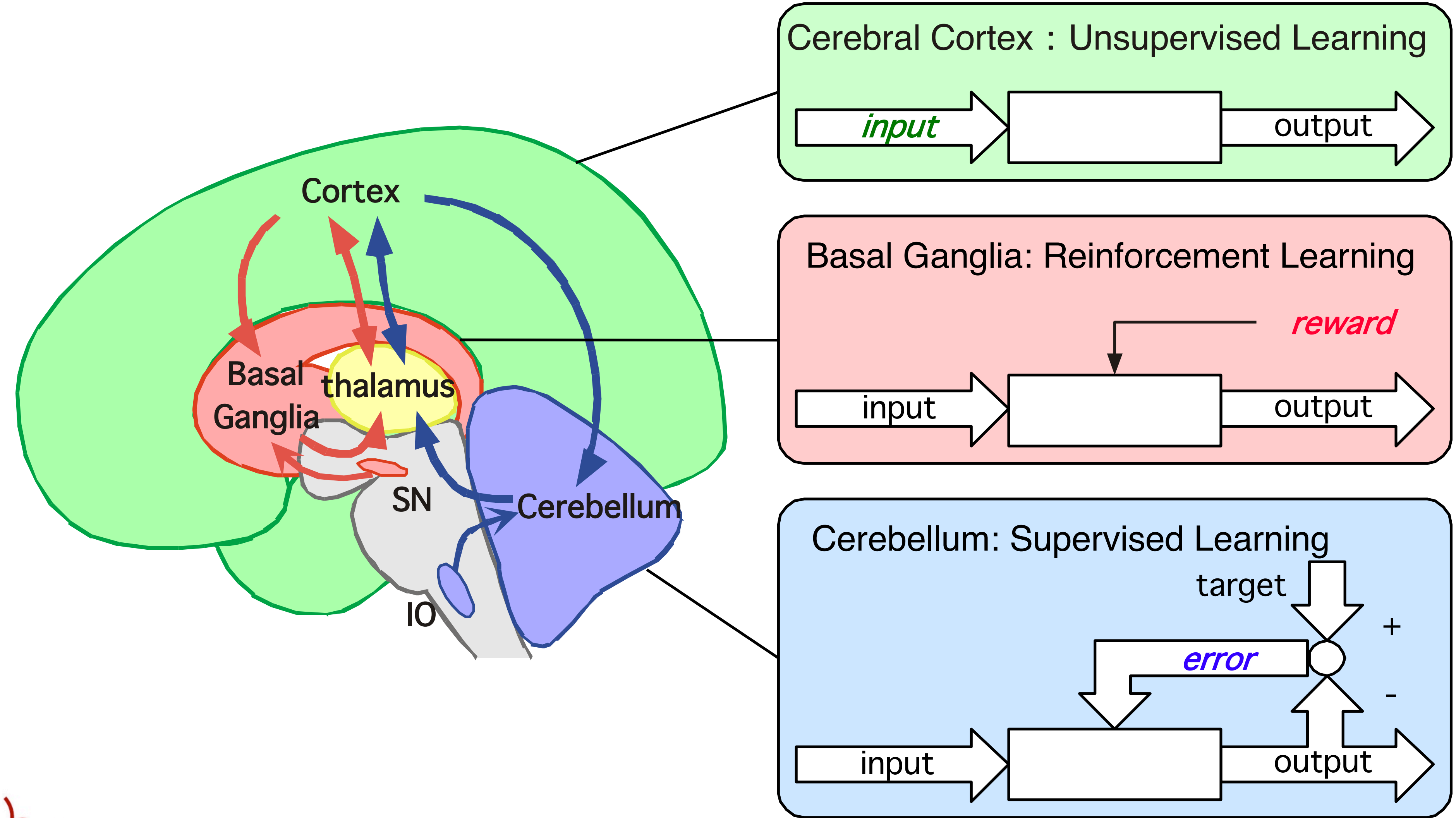
  $\rightarrow$ statistical model of $P(x)$

  - discover structure behind data



Specialization by Learning Algorithms

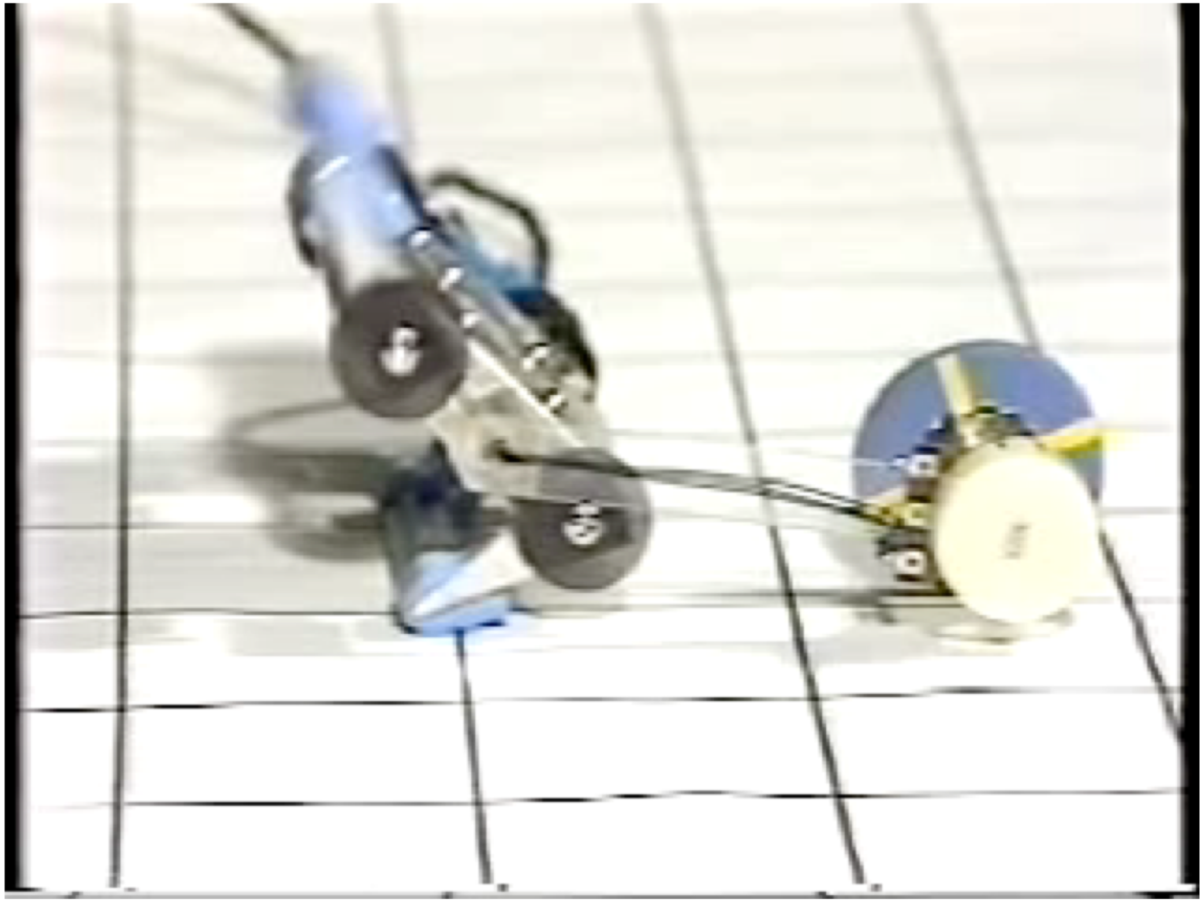
(Doya, 1999)



Learning to Walk

(Doya & Nakano, 1985)

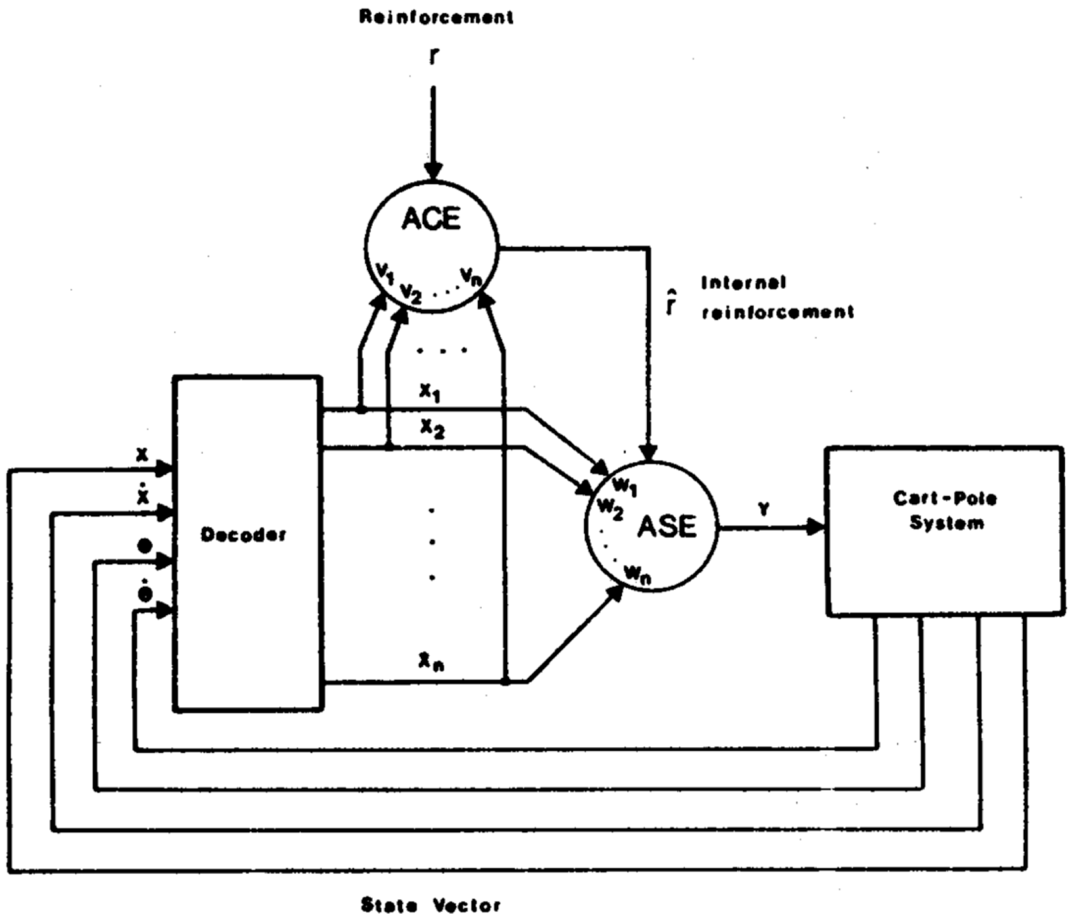
- Explore actions (cycle of 4 postures)
- Learn from performance feedback (speed sensor)

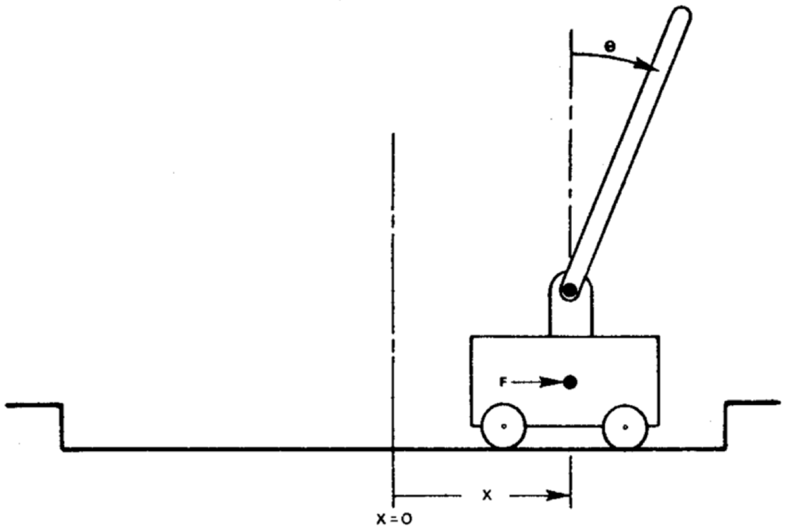


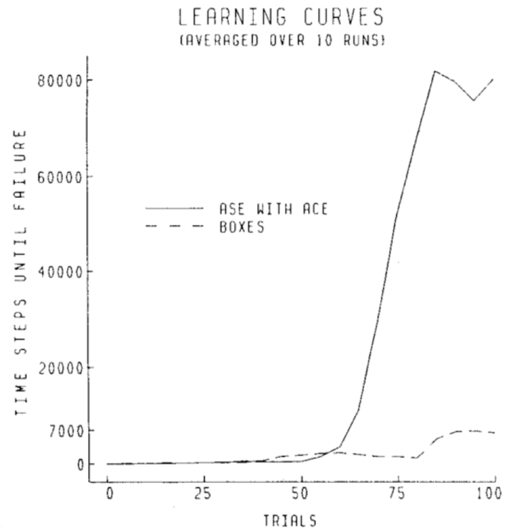
Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems

ANDREW G. BARTO, MEMBER, IEEE, RICHARD S. SUTTON, AND CHARLES W. ANDERSON

(1983)

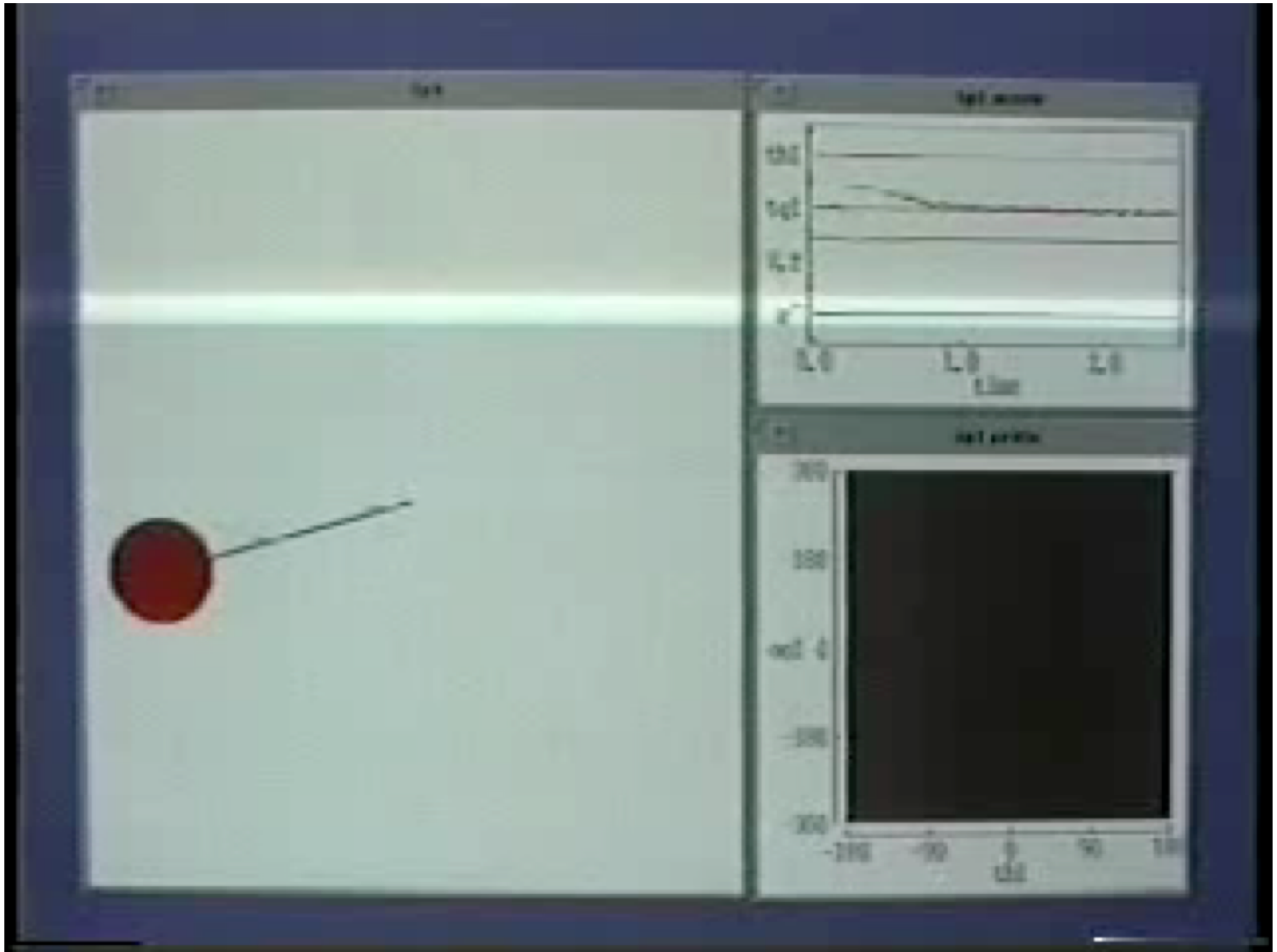






Pendulum Swing-Up

- reward function: potential energy

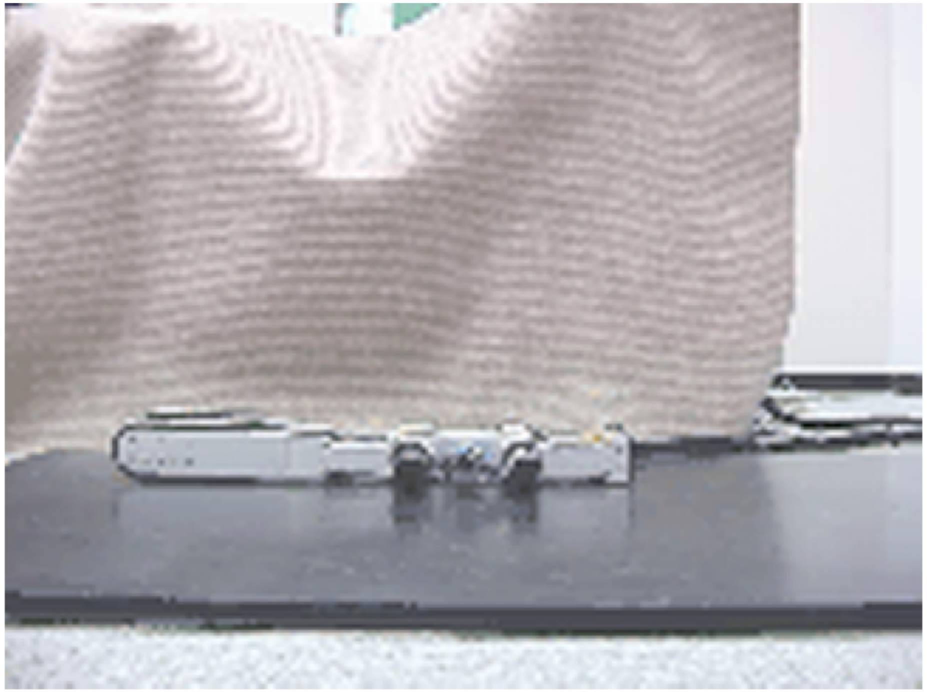


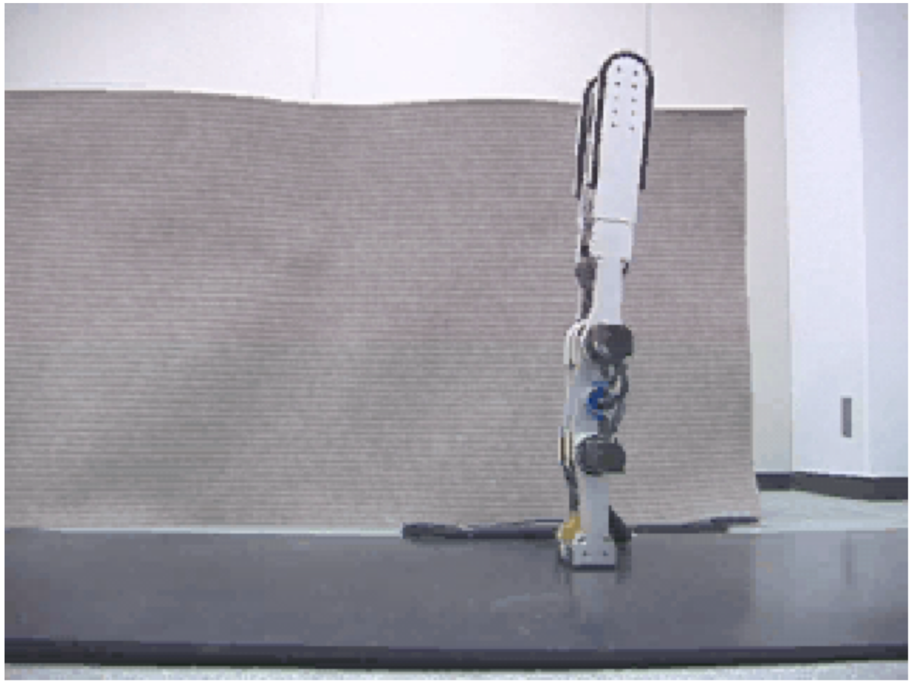
value function V(s)

s =(angle,angular velocity)

Learning to Stand Up

(Morimoto & Doya, 2001)

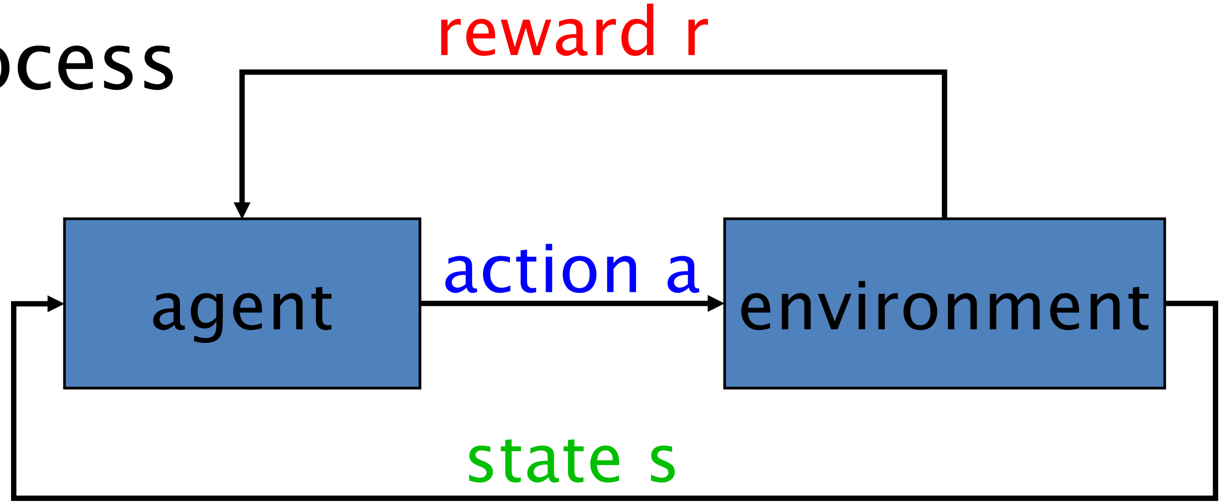




- Learning from reward and punishment
  - reward: height of the head
  - punishment: bump on the floor

Markov Decision Process (MDP)

- Markov decision process
  - state $s \in S$
  - action $a \in A$
  - policy $p(a|s)$
  - reward $p(r|s,a)$
  - dynamics $p(s'|s,a)$



- Optimal policy: maximize cumulative reward
  - finite horizon: $E[\, r(1) + r(2) + r(3) + \ldots + r(T)]$
  - infinite horizon: $E[\, r(1) + \gamma r(2) + \gamma^2 r(3) + \ldots\,]$

    $0 \le \gamma \le 1$: temporal discount factor
  - average reward: $E[\, r(1) + r(2) + \ldots + r(T)]/T,\ T \to \infty$

Actor-Critic and TD learning

- Actor: policy with parameter w

  e.g., $a(t) = \Sigma_j w_j s_j(t) + \sigma n(t)$

- Critic: learn state value function
  - $V(s(t)) = E[\ r(t) + \gamma r(t+1) + \gamma^2 r(t+2) + \ldots]$

  e.g., $V(s(t);v) = \Sigma_j v_j s_j(t)$

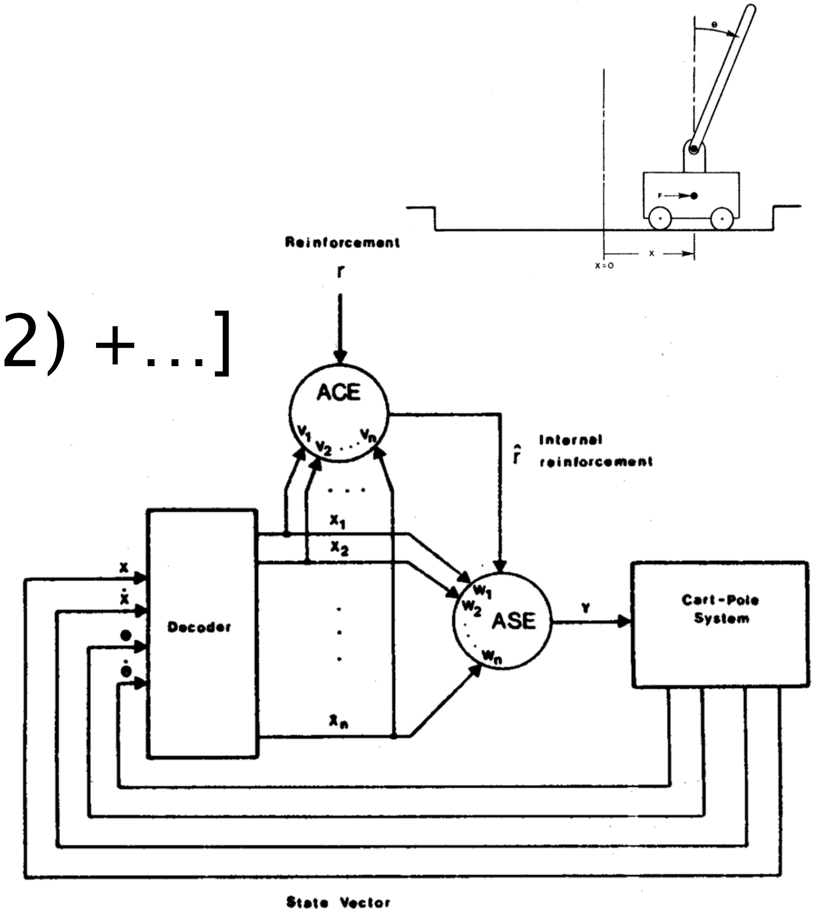
- Temporal Difference (TD) error:
  - $\delta(t) = r(t) + \gamma V(s(t+1)) - V(s(t))$

- Critic learning: $\Delta V(s(t)) \propto \delta(t)$

  $\Delta v_j = \alpha\ \delta(t)\ s_j(t)$

- Actor learning: $\Delta w \propto \delta(t)\ \partial \log P(a(t)|s(t);w)/\partial w$

  $\Delta w_j = \alpha_a\ \delta(t)\ \{a(t) - \Sigma_j w_j s_j(t)\}\ s_j(t)$ … weighted Hebb



SARSA and Q Learning

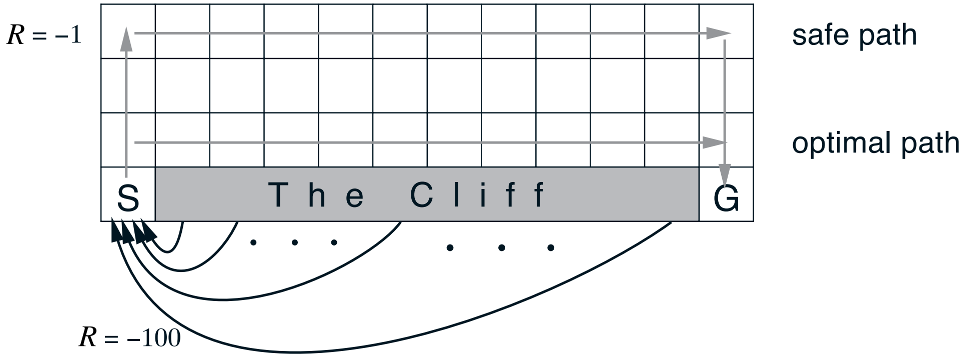
- Action value function
  - $Q(s,a) = E[\ r(t) + \gamma r(t+1) + \gamma^2 r(t+2)\ \ldots |\ s(t)=s, a(t)=a]$
- Action selection
  - ε-greedy: $a = \mathrm{argmax}_a\ Q(s,a)$ with prob $1-\varepsilon$
  - Boltzman: $P(a_i|s) = \exp[\beta Q(s,a_i)]\ /\ \Sigma_j \exp[\beta Q(s,a_j)]$
- Update by temporal difference (TD) error
  - $\Delta Q(s(t),a(t)) = \alpha\ \delta(t)$
  - SARSA: on-policy

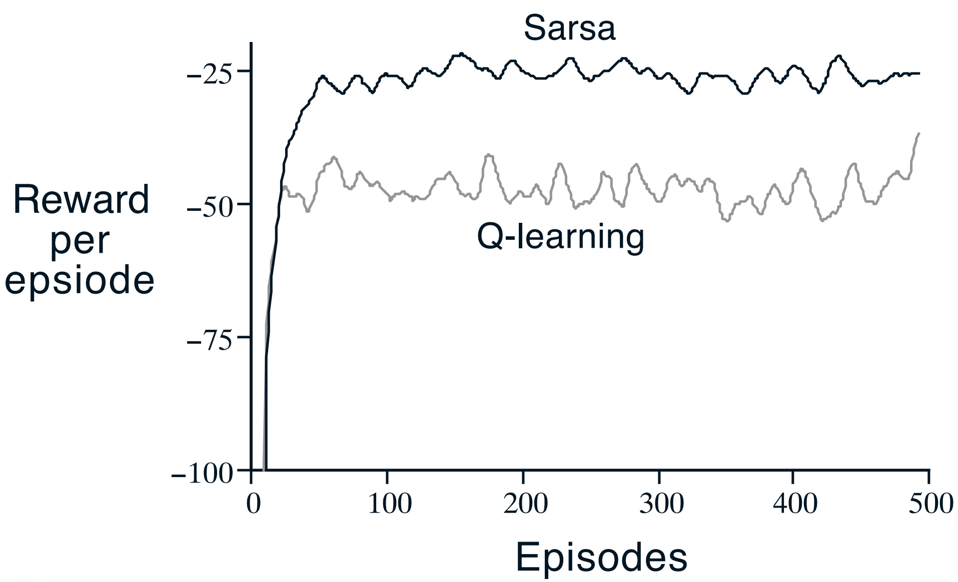
    $\delta(t) = r(t) + \gamma Q(s(t+1),a(t+1)) - Q(s(t),a(t))$
  - Q learning: off-policy

    $\delta(t) = r(t) + \gamma \max_{a'} Q(s(t+1),a') - Q(s(t),a(t))$

SARSA and Q Learning

- Cliff walking task (Sutton & Barto, 1998)



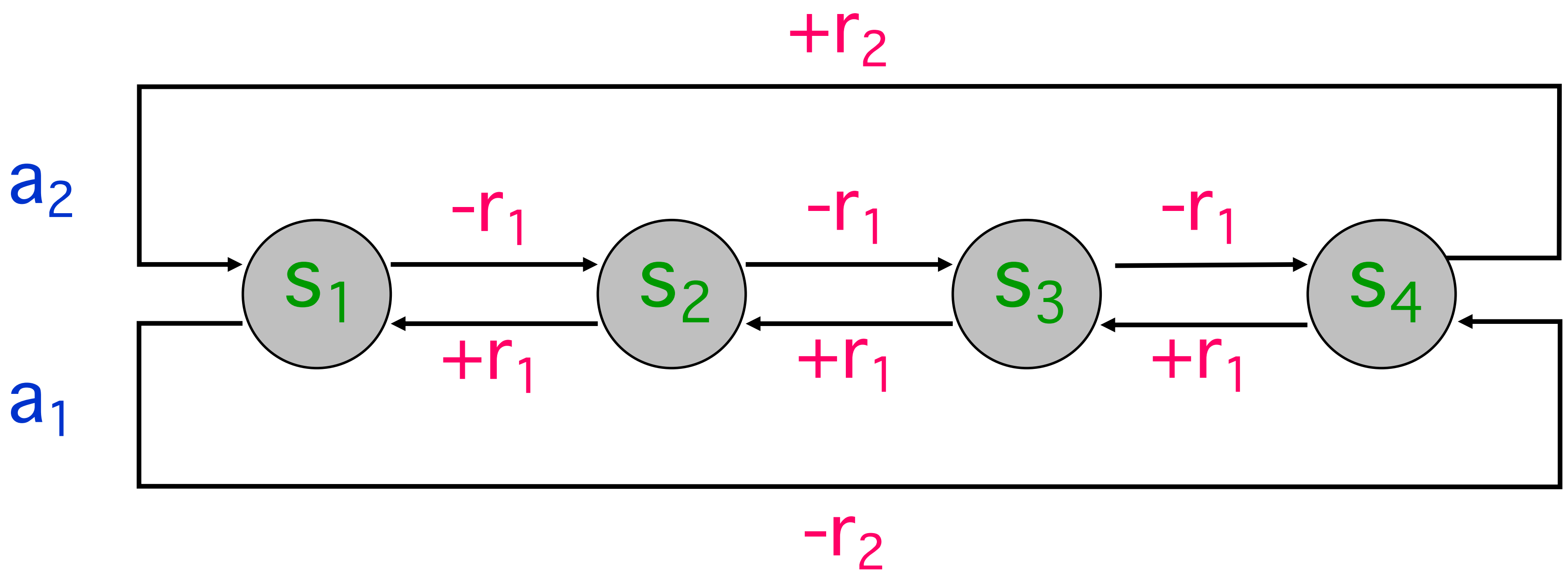


consider exploration

optimal greedy policy

"Pain-Gain" Task

- N states, 2 actions

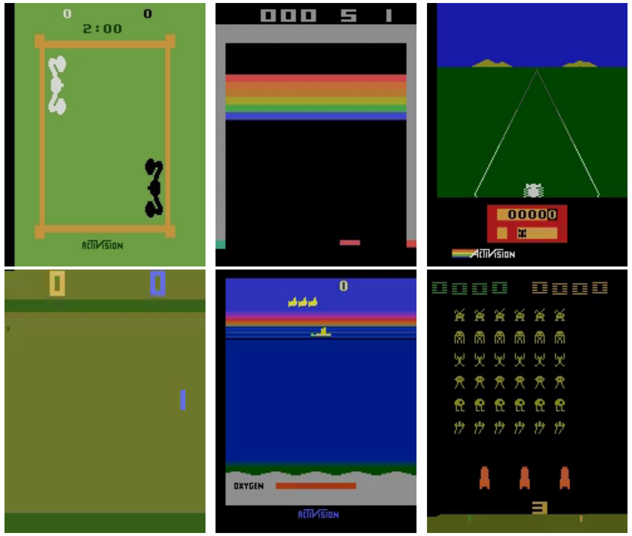


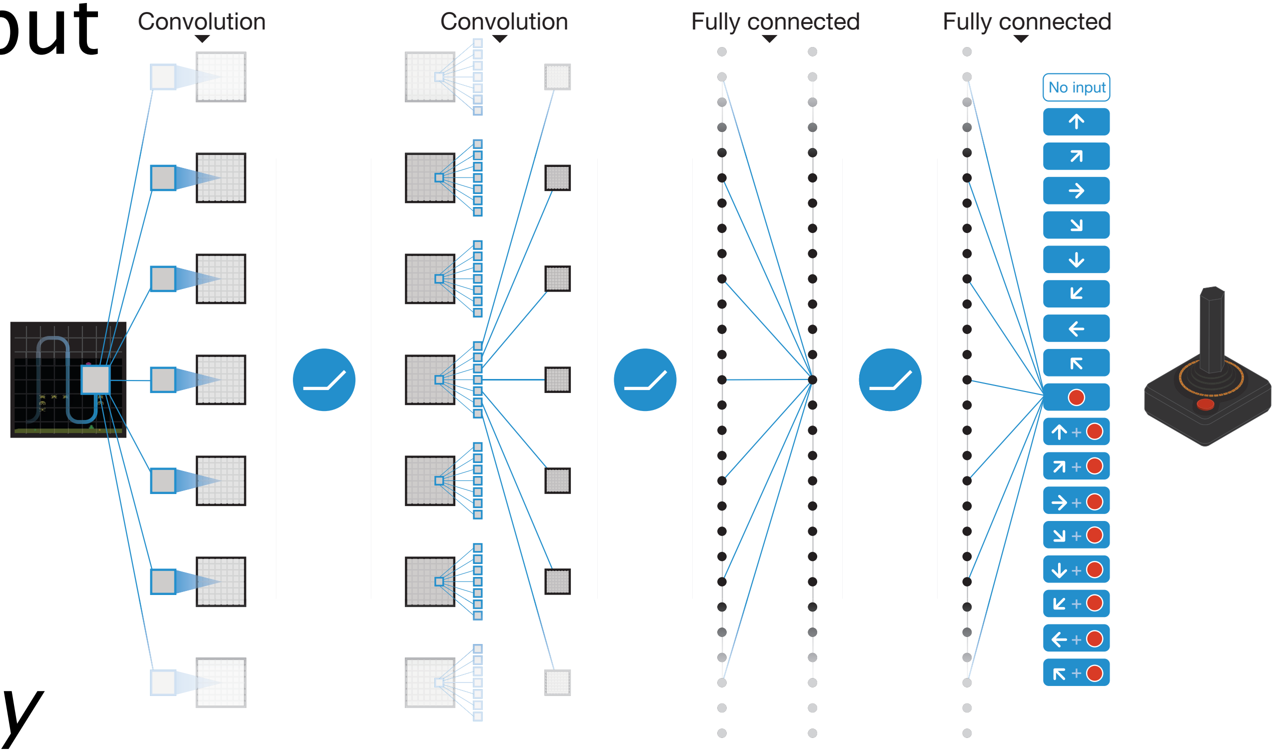
- if $r_2 \gg r_1$, then better take a_2

Deep Q-Network

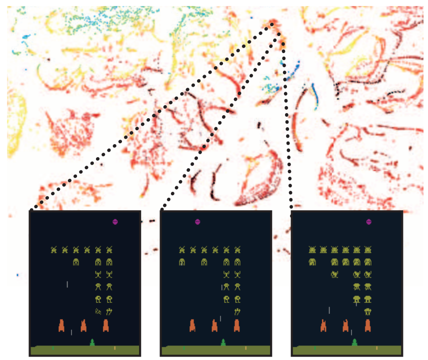
(Mnih et al., 2015)

- Game screen as input





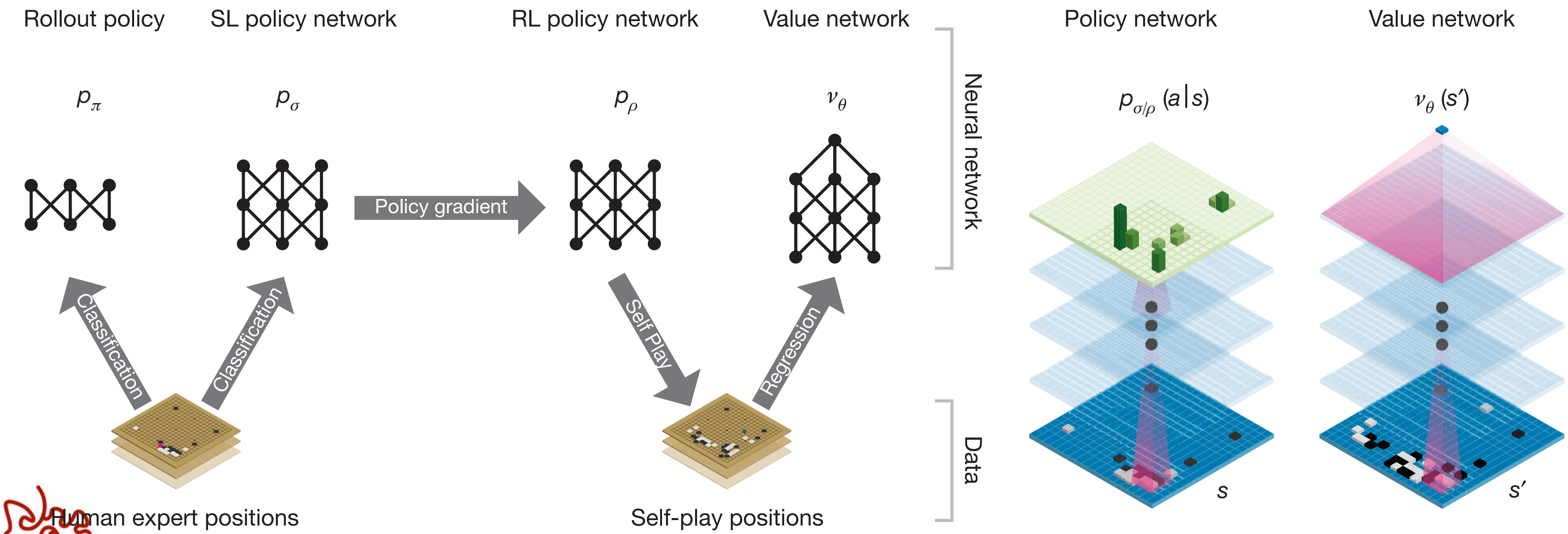
  - *Experience replay*
  - Fixing the *target network*
- DNN captures important features



  - human level in 29/49 Atari games

AlphaGo (Silver et al., 2016)

- *Supervised learning* from play data
- *Reinforcement learning* by self-play
- *Representation learning* by deep neural networks
- Shallow, narrow tree search



Model-free/Model-based Strategies

Model-free

- No prior knowledge

- Learn from experience
  - state–action–reward
  - values of states/actions

Simple, but slow learning

Model-based

- Internal model of the world
  - state, action $\rightarrow$ new state
  - state, action $\rightarrow$ reward
- Mental simulation
  - hidden state estimation

    cope with noisy observation
  - action planning

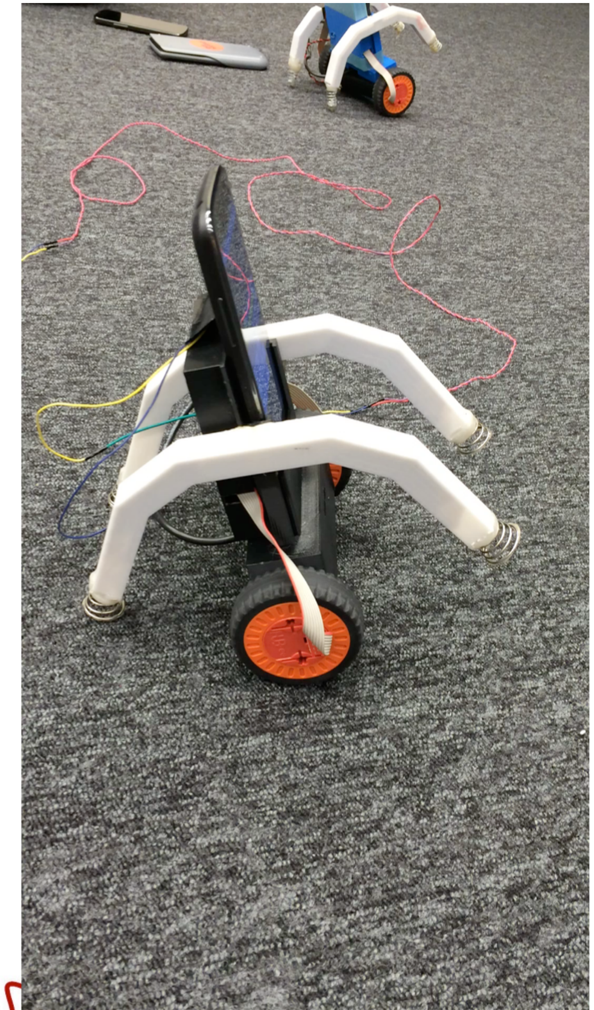
    find the best action sequence

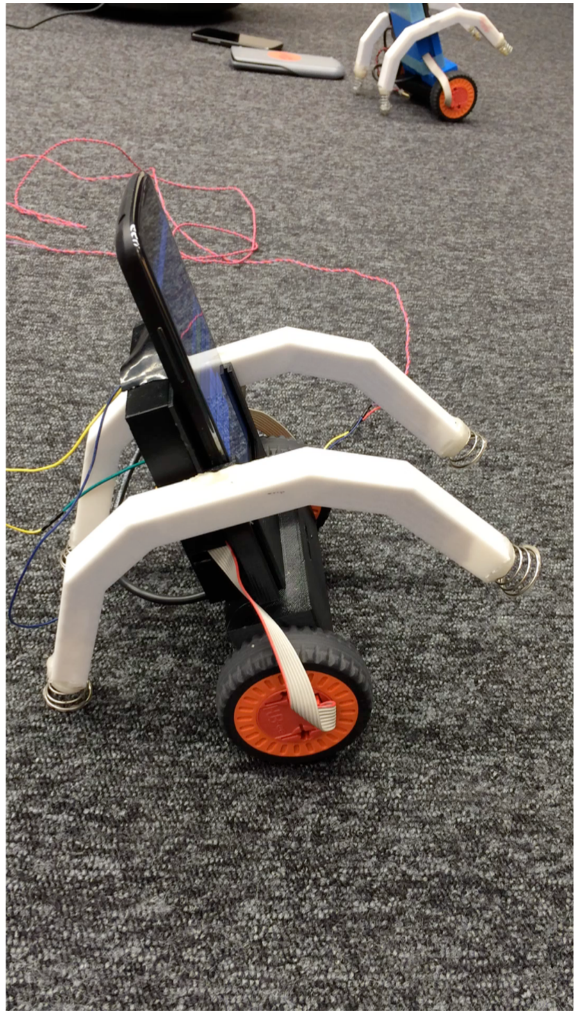
Flexible, but heavy load

Bounce Up and Balance by PILCO

(Paavo Parmas)

1st try	2nd try	8th try





Reinforcement Learning

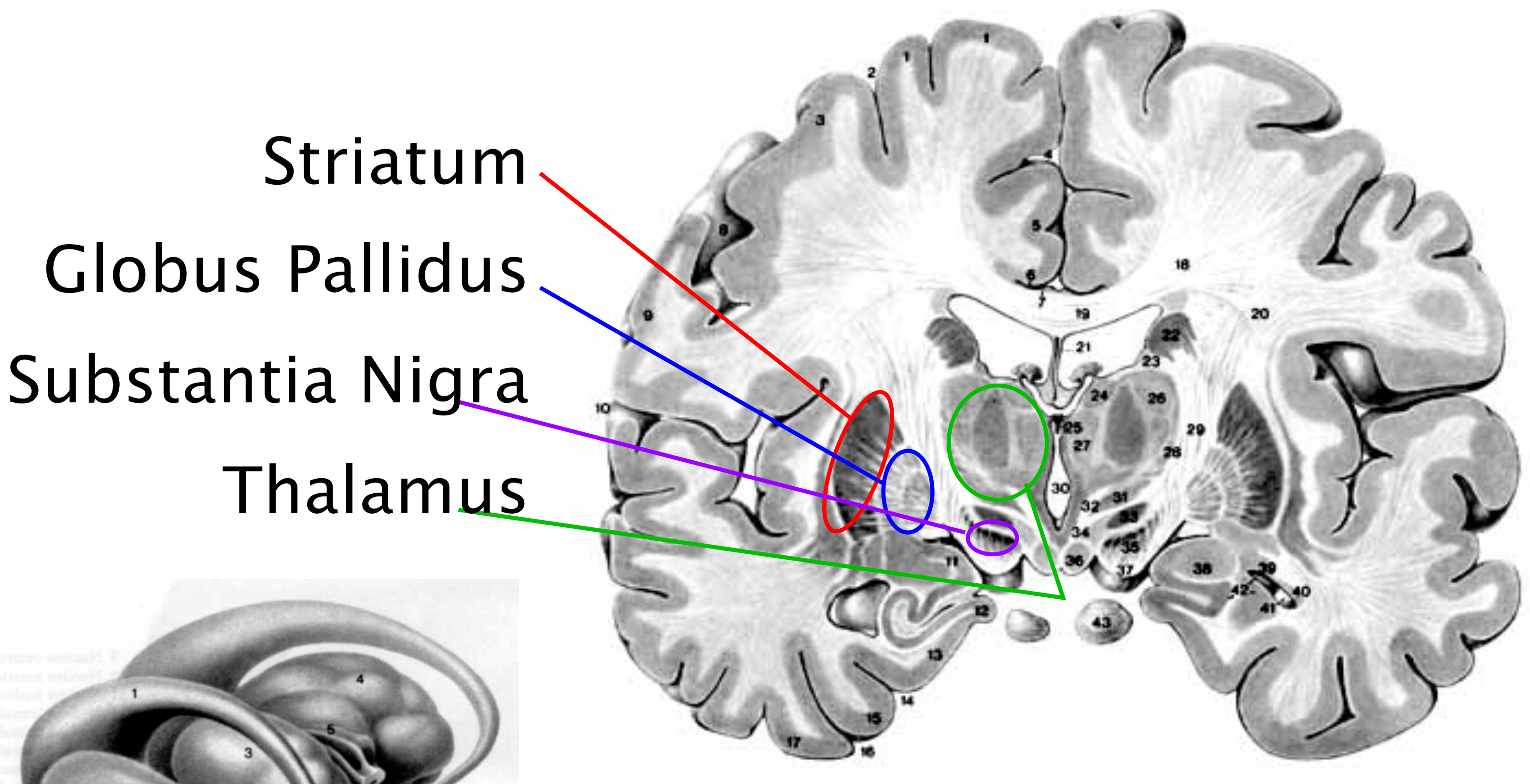
- Predict reward: *value function*
  - $V(s) = E[\, r(t) + \gamma r(t+1) + \gamma^2 r(t+2) \ldots |\, s(t)=s]$
  - $Q(s,a) = E[\, r(t) + \gamma r(t+1) + \gamma^2 r(t+2) \ldots |\, s(t)=s,\, a(t)=a]$
- Select action

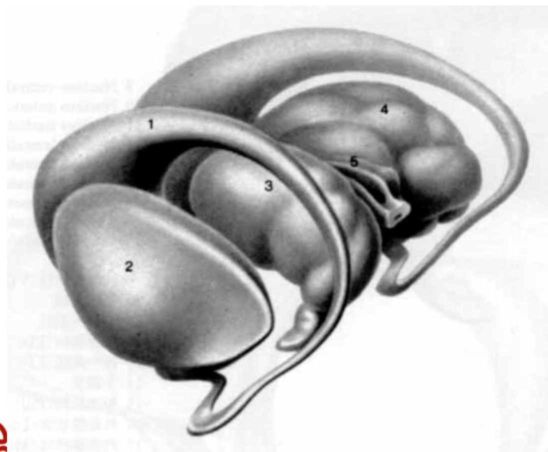
  How to implement these steps?
  - *greedy*: $a = \operatorname{argmax} Q(s,a)$
  - *Boltzmann*: $P(a|s) \propto \exp[\, \beta\, Q(s,a)]$
- Update prediction: *TD error*
  - $\delta(t) = r(t) + \gamma V(s(t+1)) - V(s(t))$
  - $\Delta V(s(t)) = \alpha\, \delta(t)$

    How to tune these parameters?
  - $\Delta Q(s(t),a(t)) = \alpha\, \delta(t)$

Basal Ganglia

- Locus of Parkinson's and Huntington's diseases

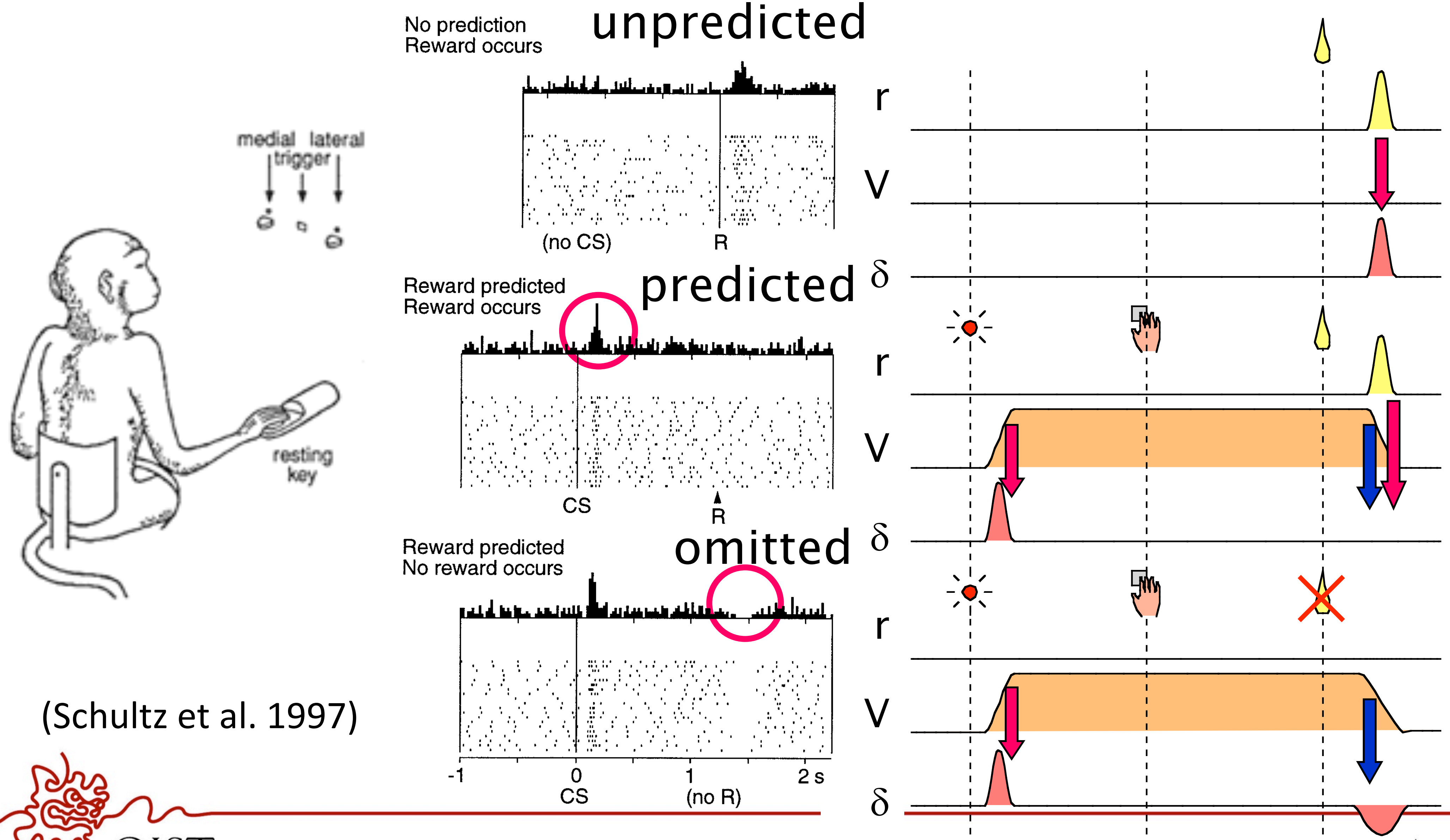




- What is their normal function??

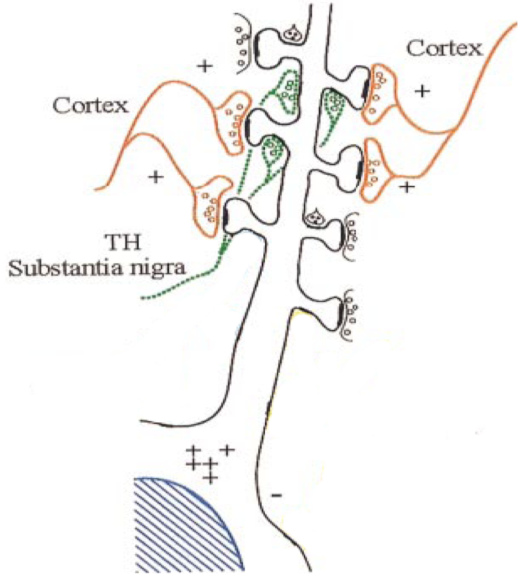
Dopamine Neurons Code TD Error

$$\delta(t) = r(t) + \gamma V(s(t+1)) - V(s(t))$$

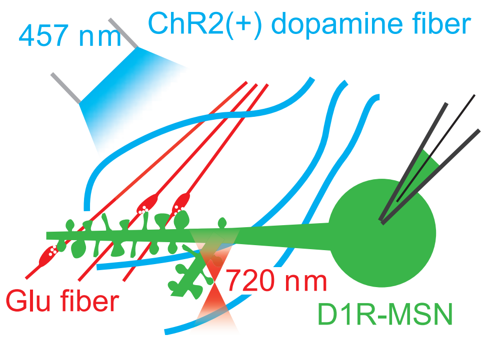


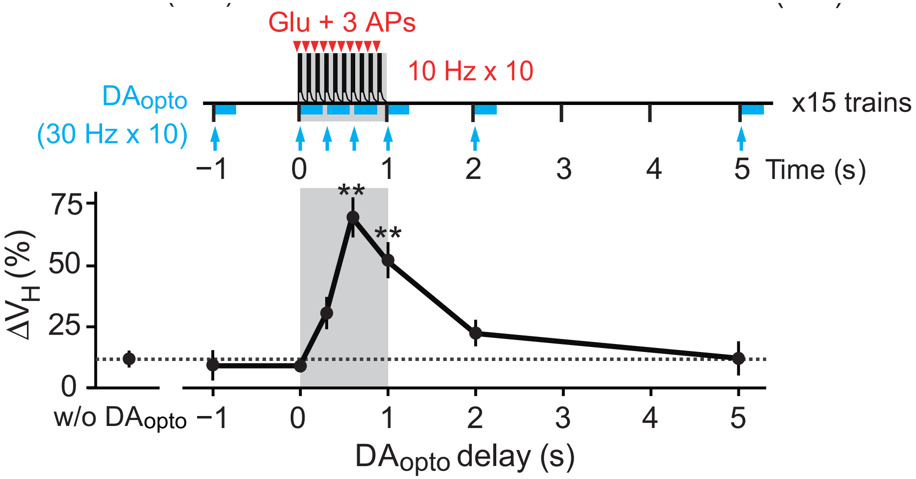
(Schultz et al. 1997)

Dopamine-dependent Plasticity



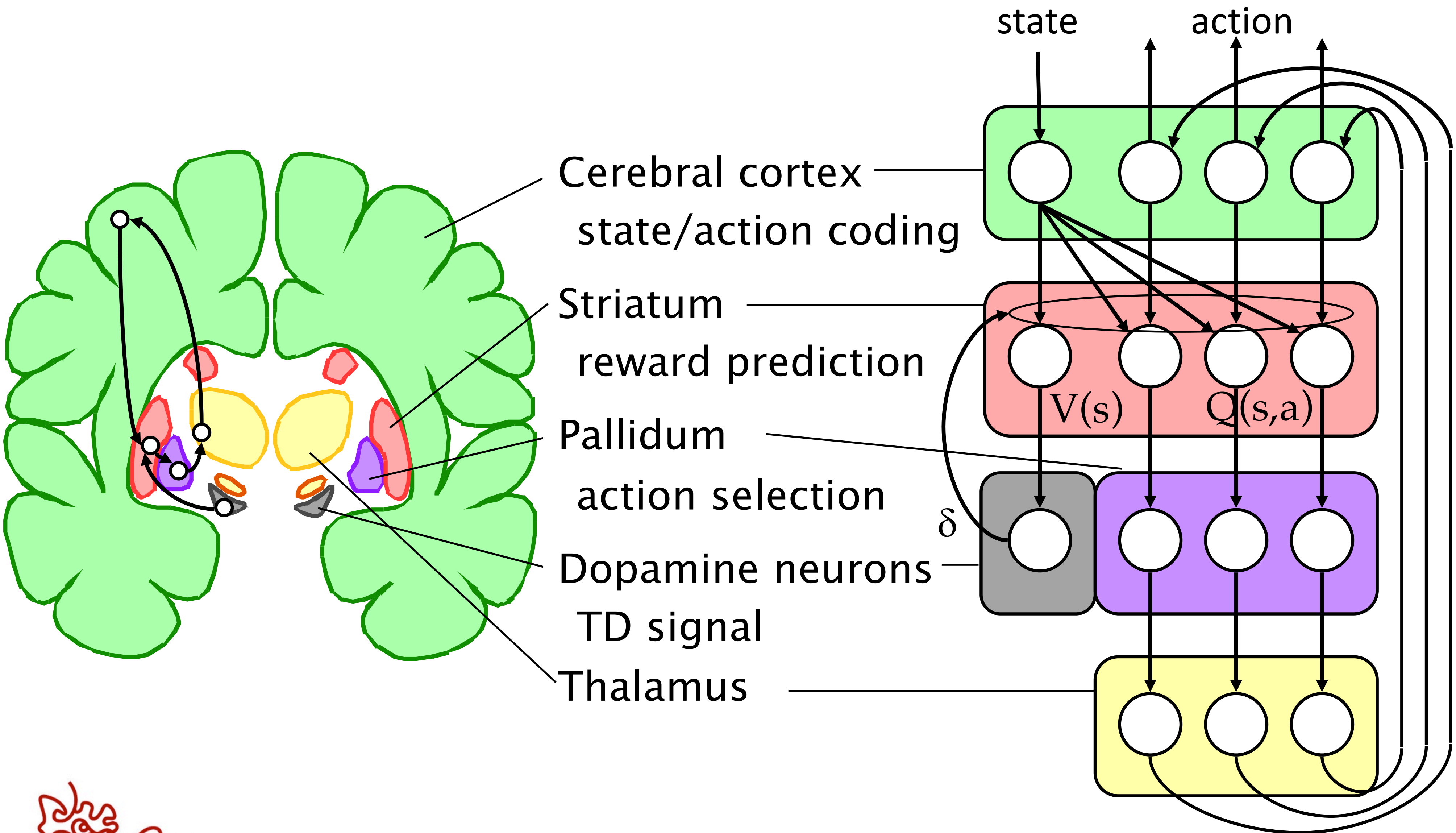
- Medium spiny neurons in striatum
  - glutamate from cortex
  - dopamine from midbrain
- Three-factor learning rule (Wickens et al.)
  - cortical input + spike → LTD
  - cortical input + spike + dopamine → LTP
  - input x output x reward
- Time window of plasticity (Yagishita et al., 2014)





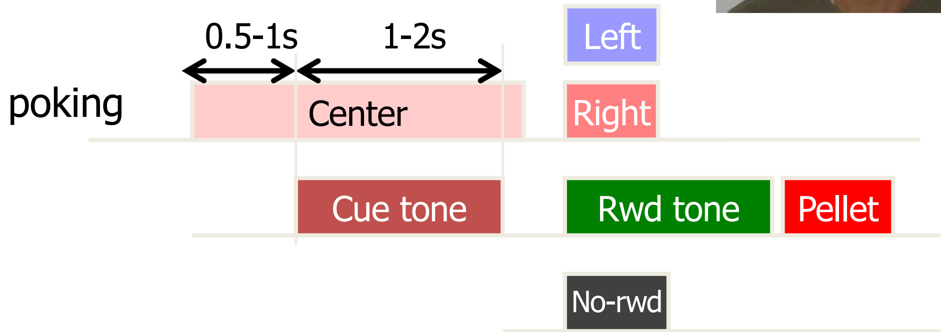
Basal Ganglia for Reinforcement Learning?

(Doya 2000, 2007)



Fixed and Free Choice Task

(Ito & Doya, 2015, J Neuroscience)

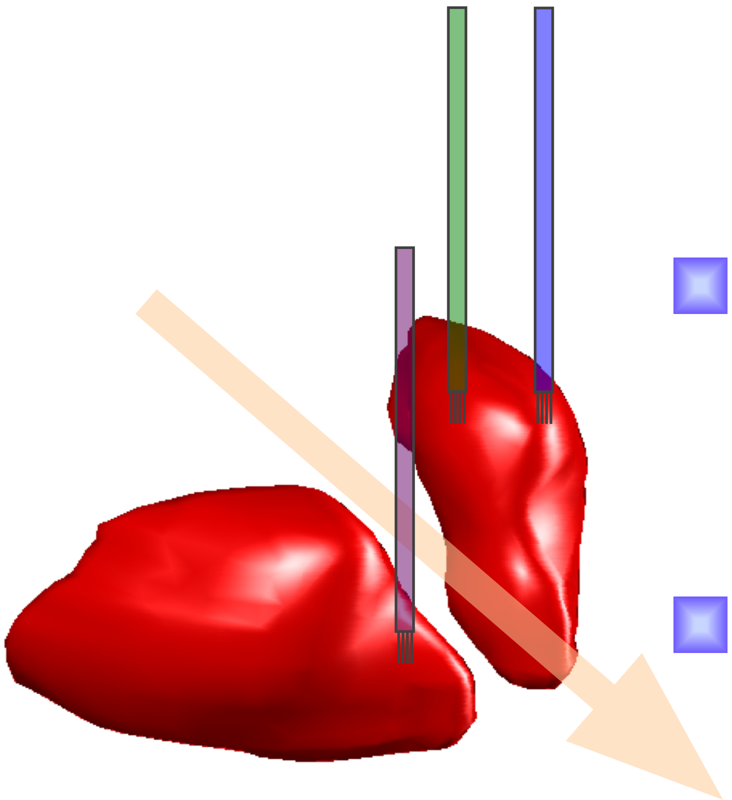


Left Center Right

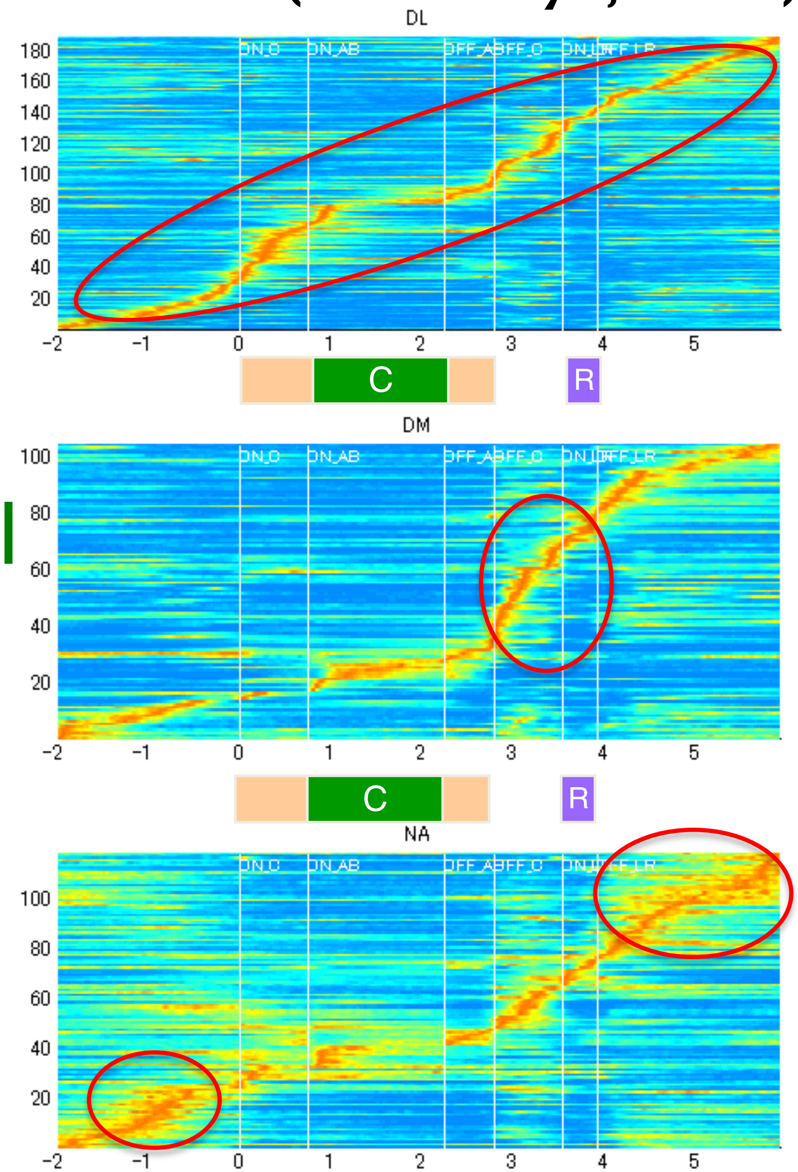
Cue tone	Reward prob. (L, R)
Left tone (900Hz)	Fixed (50%,0%)
Right tone (6500Hz)	Fixed (0%, 50%)
Free-choice tone (White noise)	Varied (90%, 50%) (50%, 90%) (50%, 10%) (10%, 50%)

Neural Activity in the Striatum

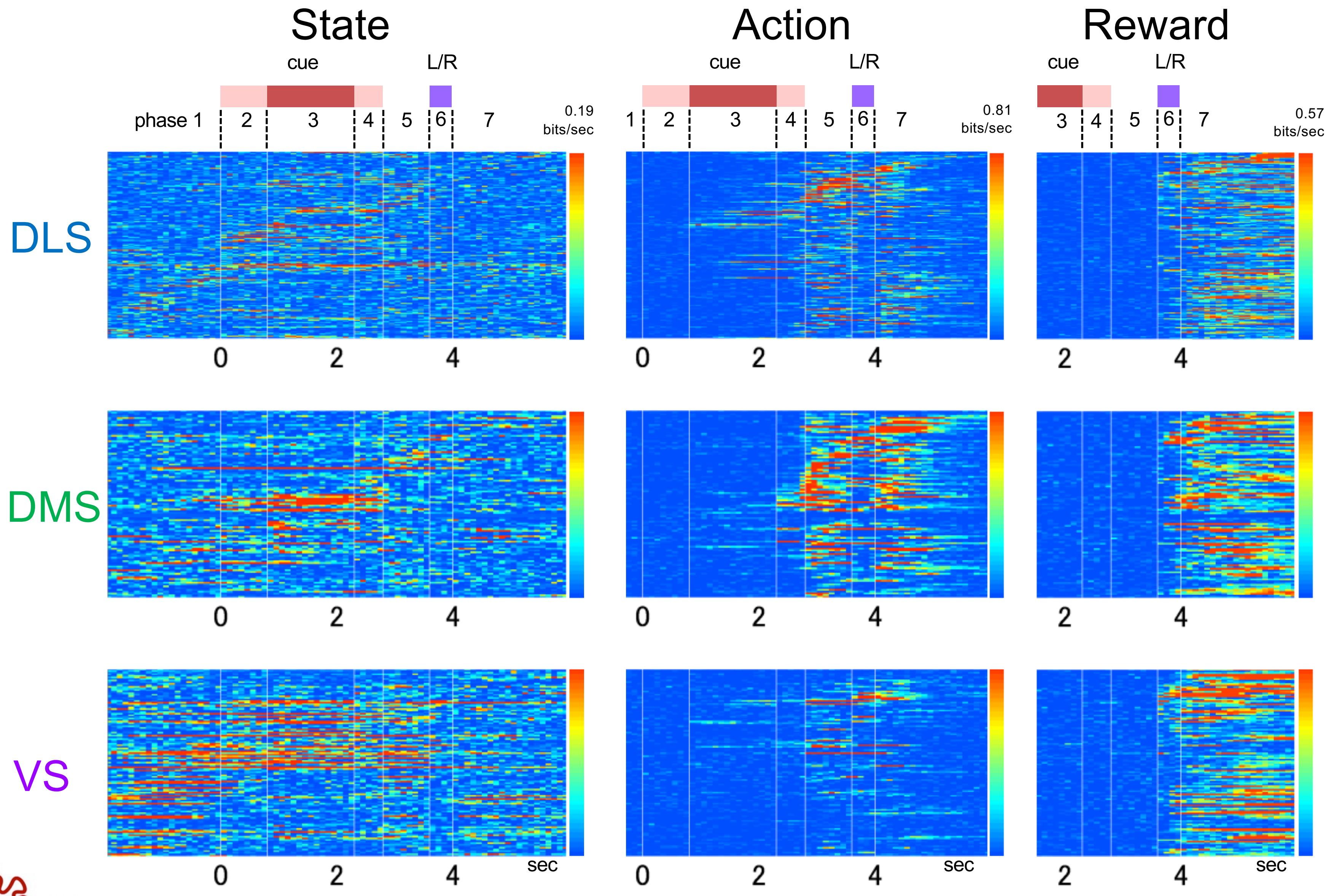
(Ito & Doya, 2015)



- Dorsolateral
- Dorsomedial
- Ventral



State/Action/Reward Coding



Generalized Q-learning Model

(Ito & Doya, 2009)

- Action selection

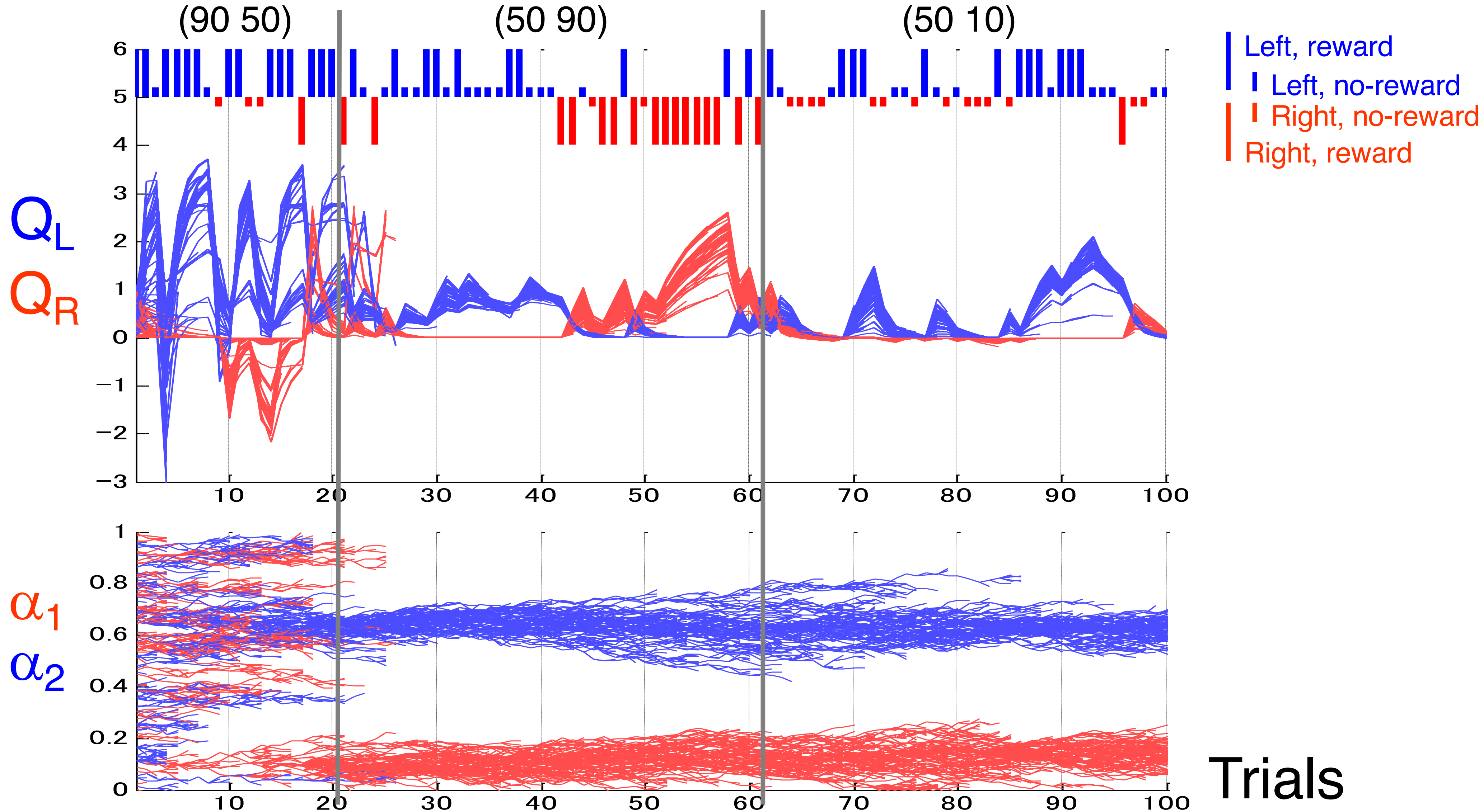
$$P(a(t)=L) = \exp Q_L(t)/(\exp Q_L(t)+\exp Q_R(t))$$

- Action value update: $i \in \{L,R\}$

$$Q_i(t+1) = \begin{cases} (1-\alpha_1)Q_i(t) + \alpha_1\kappa_1 & \text{if } a(t)=i,\ r(t)=1 \\ (1-\alpha_1)Q_i(t) - \alpha_1\kappa_2 & \text{if } a(t)=i,\ r(t)=0 \\ (1-\alpha_2)Q_i(t) & \text{if } a(t)\neq i,\ r(t)=1 \\ (1-\alpha_2)Q_i(t) & \text{if } a(t)\neq i,\ r(t)=0 \end{cases}$$

- Parameters
  - α_1: learning rate
  - α_2: forgetting rate
  - κ_1: reward reinforcement
  - κ_2: no-reward aversion

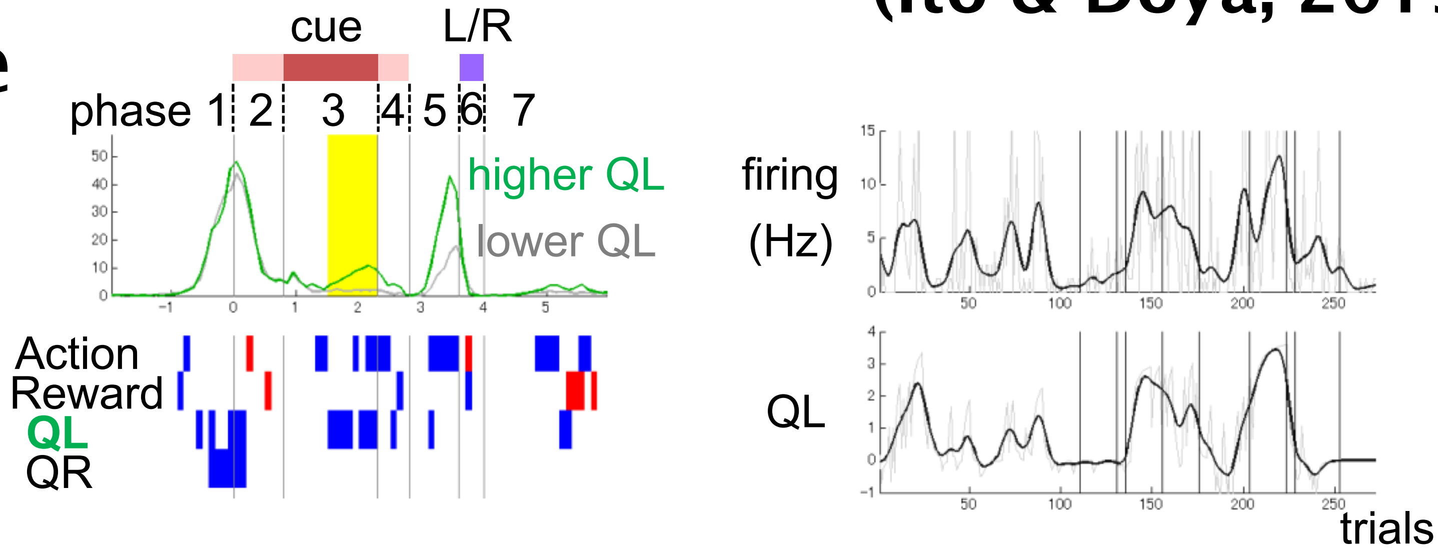
Model Fitting by Particle Filter



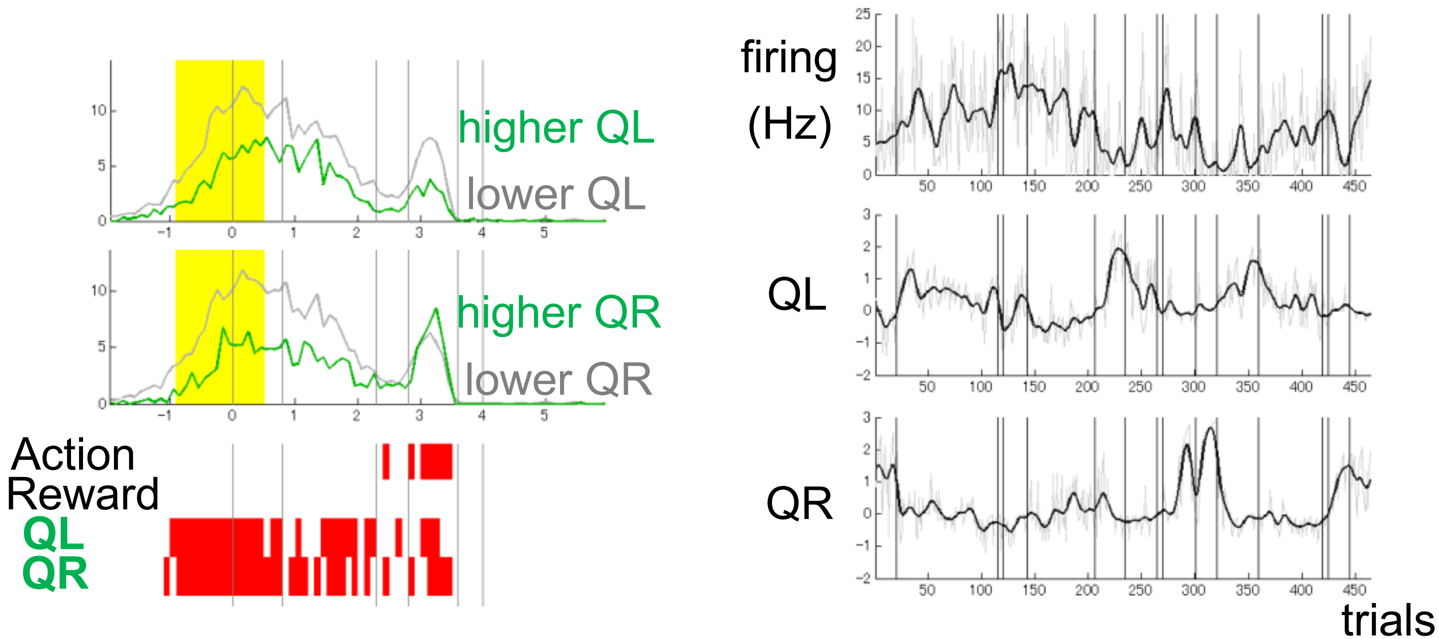
Action/State Values in Striatum

(Ito & Doya, 2015)

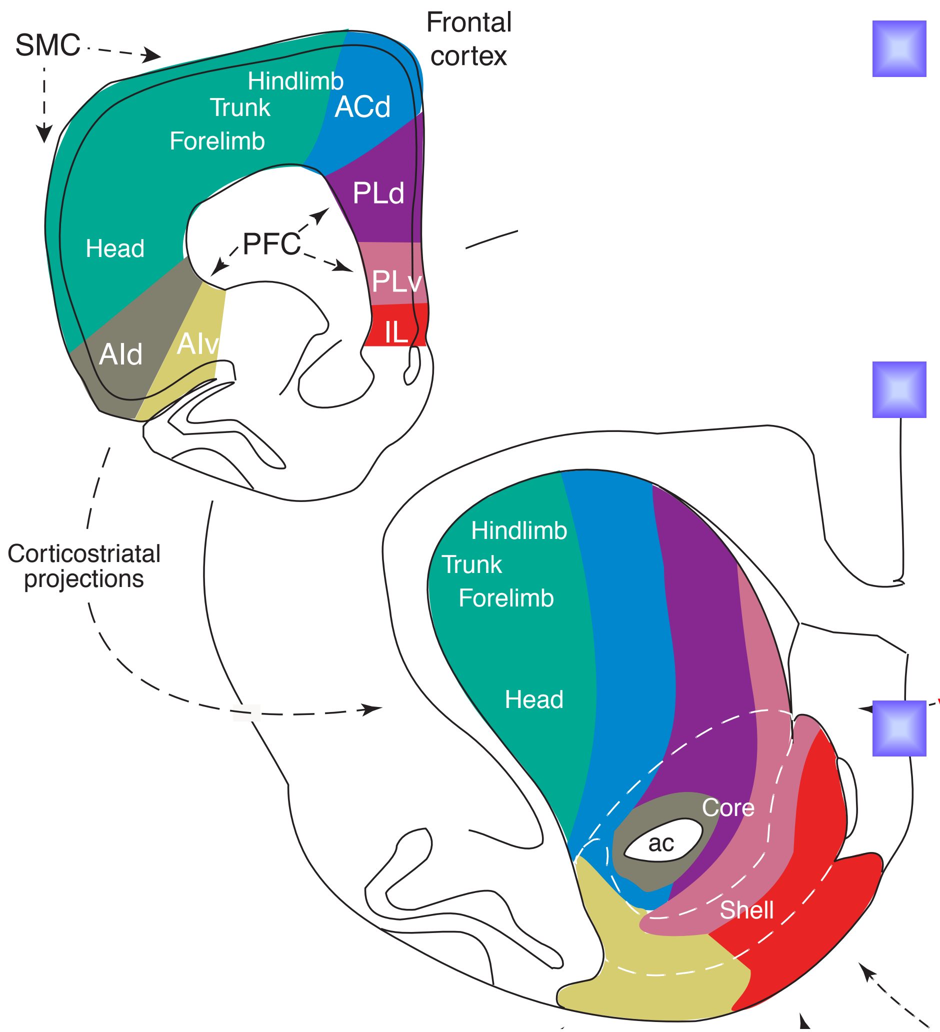
- Action value
  - DLS
  - DMS



- State value
  - VS



Hierarchy in Cortico-Striatal Network



(Voorn et al., 2004)

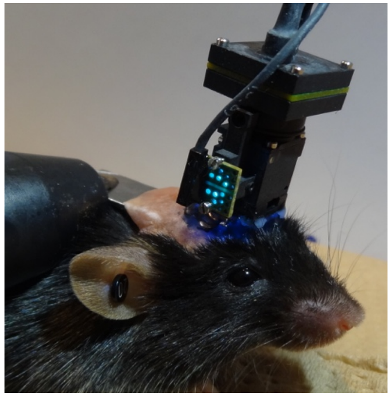
- **Dorsolateral striatum: motor**
  - early action coding
  - what motor action?
- **Dorsomedial striatum: cognitive**
  - choice action value
  - which goal?
- **Ventral striatum: motivational?**
  - state value
  - whether worth doing?

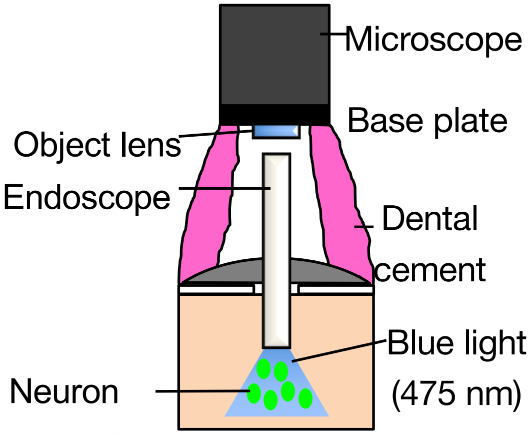
eNeuro (2018)

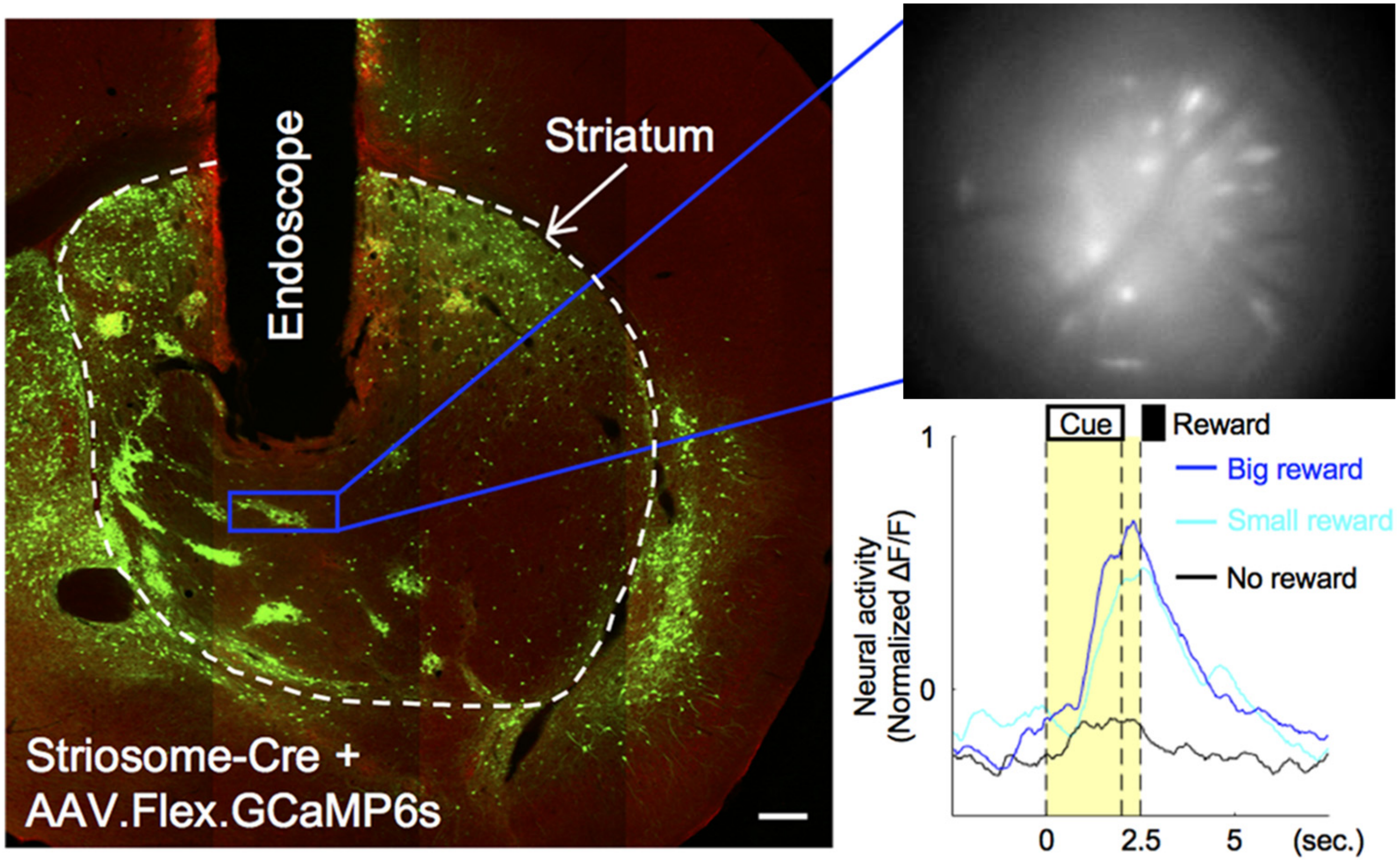
Reward-Predictive Neural Activities in Striatal Striosome Compartments

Tomohiko Yoshizawa,[1] Makoto Ito,[1,2] and Kenji Doya[1]

Imaging striosome neuron activity by endoscope





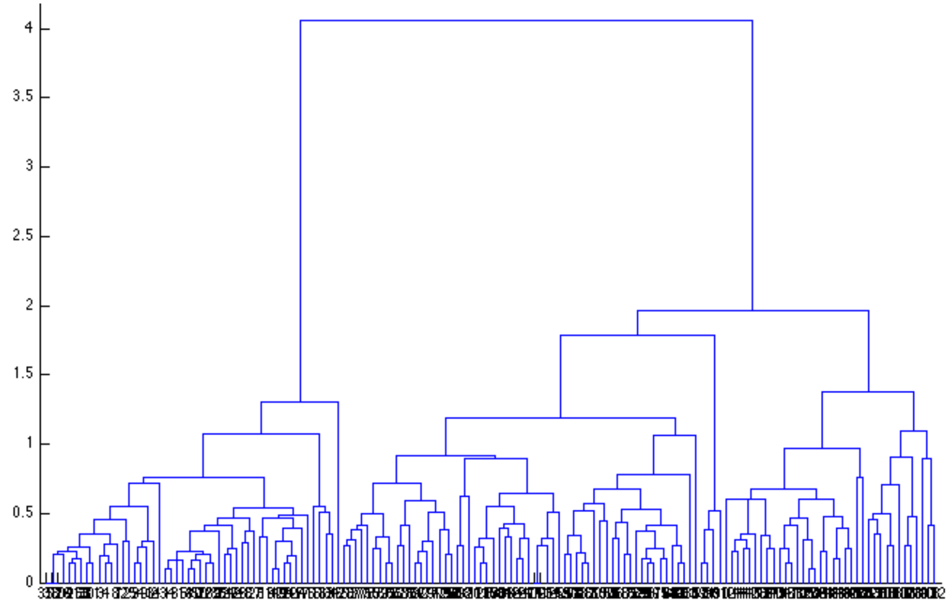


Unsupervised Learning

- input $\{\mathbf{x}_1,\ldots,\mathbf{x}_N\} \rightarrow$ output $\{\mathbf{y}_1,\ldots,\mathbf{y}_N\}$
  - no target output
  - finding structures in distribution of $\mathbf{x}$
- Dividing into subgroups
  - clustering
  - mixtures of Gaussians
  - self-organizing maps (SOM)
- Decomposing into components
  - principal component analysis (PCA)
  - independent component analysis (ICA)

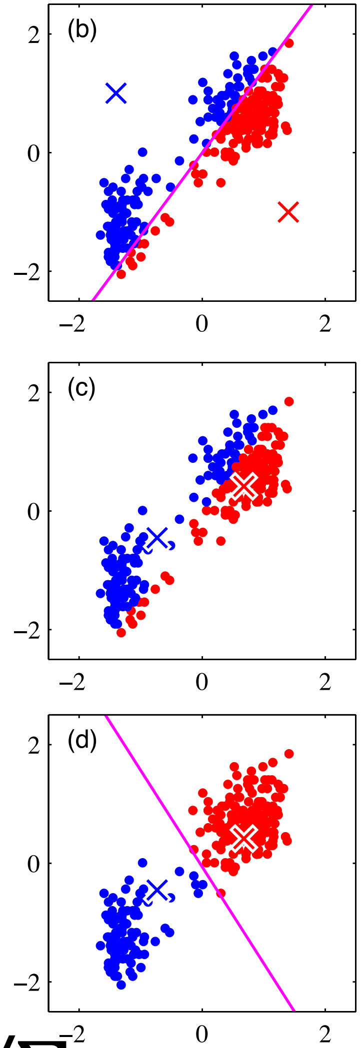
Hierarchical Clustering

- Bottom up clustering
  - start with each data point as a cluster
  - check distance: nearest/average/furthest
  - merge closest clusters
- Dendrogram



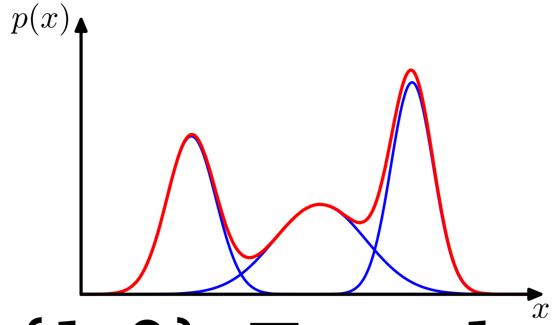
- Python
  - scipy.cluster.hierarchy
  - sklearn.cluster.AgglomerativeClustering

K-means Clustering



- Data set $\{\mathbf{x}_1,\dots,\mathbf{x}_N\} \rightarrow K$ clusters
  - no target labels!
- Distortion measure $J = \sum_n \sum_k r_{nk} ||\mathbf{x}_n - \mu_k||^2$
  - prototypes μ_k ($k=1,\dots,K$)
  - indicator variable $r_{nk} \in \{1,0\}$, $\sum_k r_{nk}=1$
- Fix μ_k and minimize with r_{nk}
  - for each $\mathbf{x}_n$, find nearest μ_k, set $r_{nk} = 1$
- Fix r_{nk} and minimize with μ_k
  - mean of data in the cluster $\mu_k = \sum_n r_{nk}\mathbf{x}_n / \sum_n r_{nk}$
- Repeat $\rightarrow$ converge to a local minimum

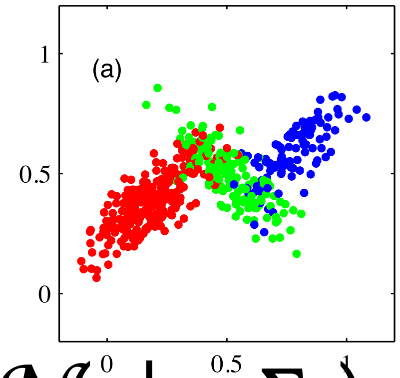
Mixtures of Gaussians



- Gaussian mixture distribution

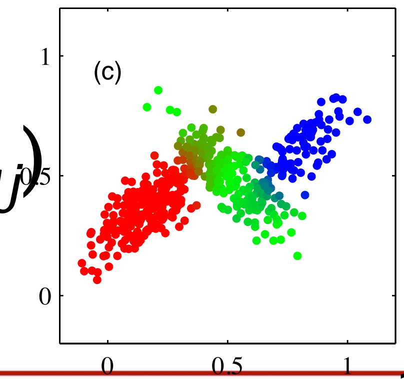
$$p(\mathbf{x}) = \textstyle\sum_k \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$$

- *Latent variable* $\mathbf{z}=(z_1,\ldots,z_K)$: $z_k \in \{1,0\}$, $\sum_k z_k = 1$
  - marginal distribution $p(\mathbf{z}) = \prod_k \pi_k^{z_k}$
  - conditional $p(\mathbf{x}|\mathbf{z}) = \prod_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)^{z_k}$
  - marginal of $\mathbf{x}$ is Gaussian mixture



$$p(\mathbf{x}) = \textstyle\sum_z p(\mathbf{x},\mathbf{z}) = \sum_z p(\mathbf{z})\, p(\mathbf{x}|\mathbf{z}) = \sum_k \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$$

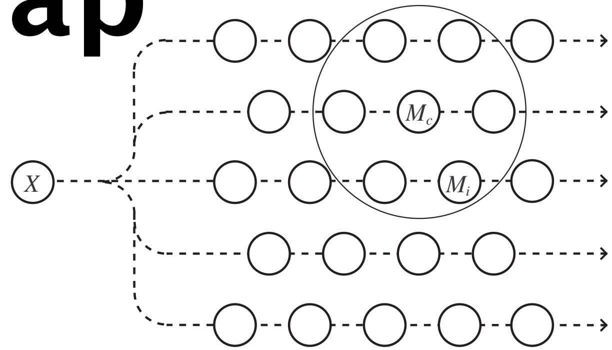
- Posterior of $\mathbf{z}$ given $\mathbf{x}$: *responsibility*



$$p(z_k=1|\mathbf{x}) = \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k) / \textstyle\sum_j \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \Sigma_j)$$

  - π_k as the prior of $z_k=1$

Self-Organizing Map



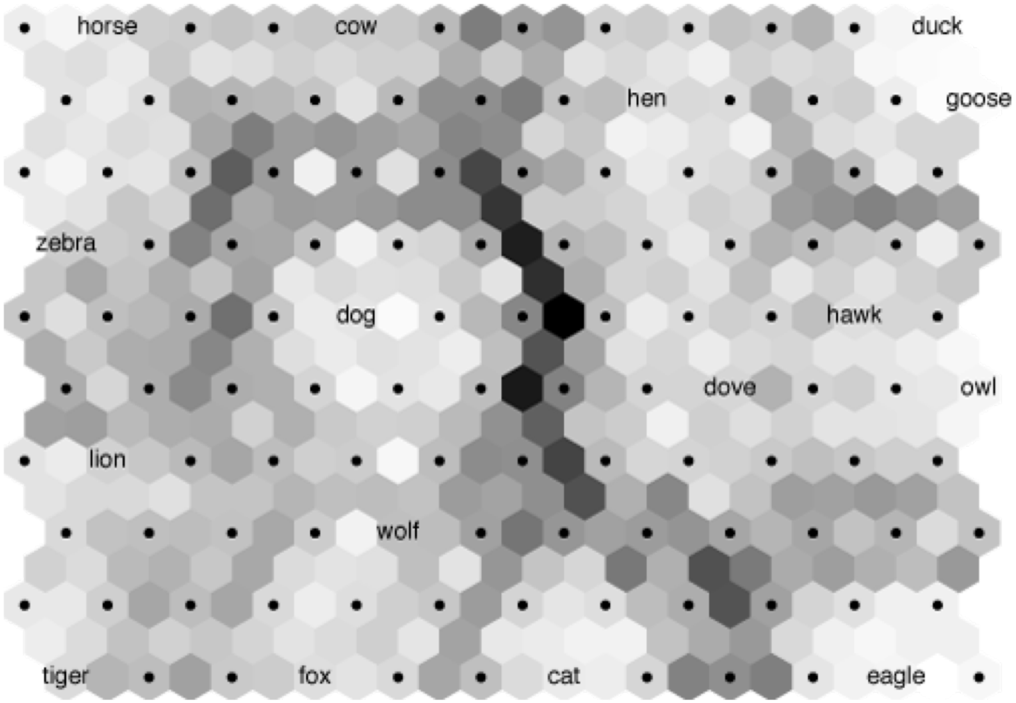
- Data set $\{\mathbf{x}_1,\dots,\mathbf{x}_N\}$

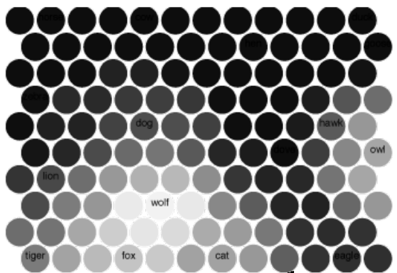
  $\rightarrow$ grid of prototypes $\{\mu_1,\dots,\mu_K\}$
  - similar prototypes for neighboring nodes
- Stepwise recursive SOM
  - find best match: $c(n) = \mathrm{argmin}_k \, ||\mathbf{x}_n - \mu_k||$
  - update neighbors: $\mu_k(t+1) = \mu_k(t) + \alpha h_{c(n)k}(\mathbf{x}_n - \mu_k)$

    $h_{ck} = \exp[-\mathrm{dist}(c,k)/2\sigma^2]$
- Batch update
  - find $c(n)$ for all $\mathbf{x}_n$
  - mean of neighbors $\mu_k = \sum_n h_{c(n)k}\mathbf{x}_n / \sum_n h_{c(n)k}$

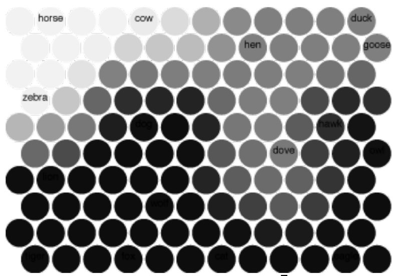
SOM Examples

- Animals (Ito et al. 2006)





nocturnal



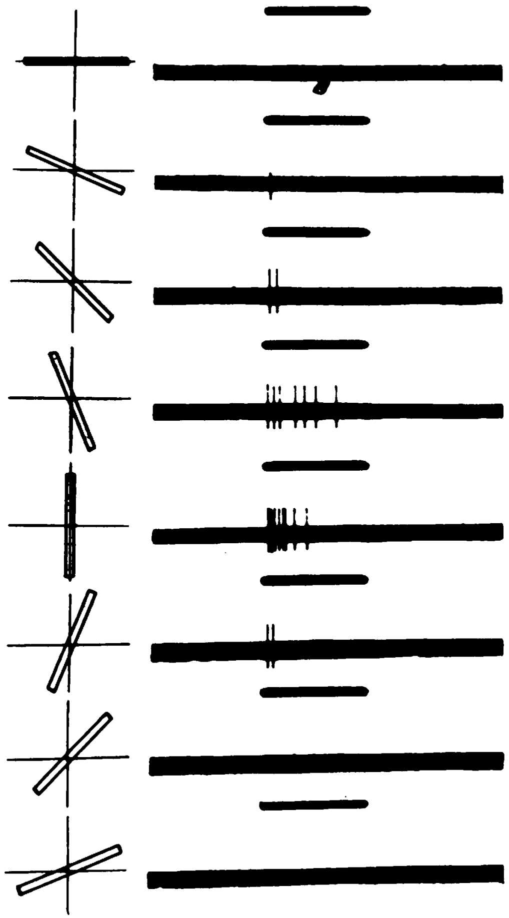
vegetarian

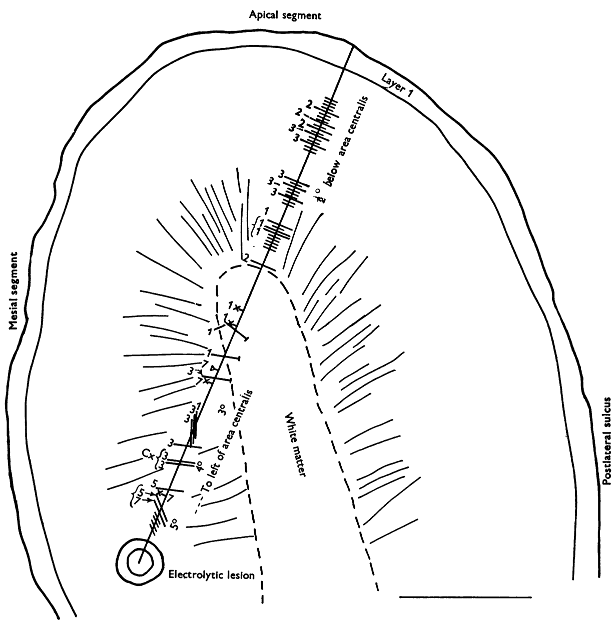
- Faces (Yuriy 2007)

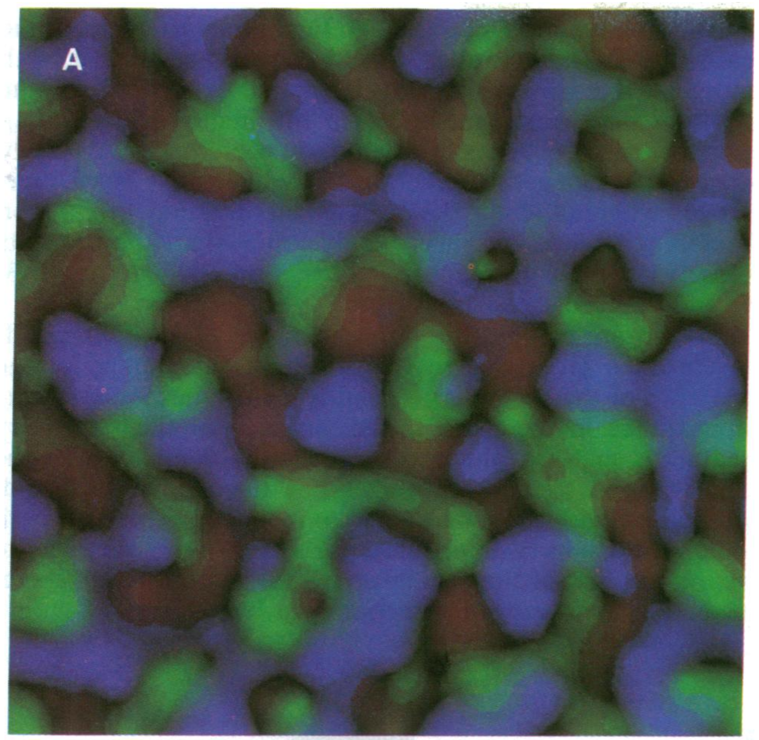
10x10 SOM clustering results

Visual Cortical Columns

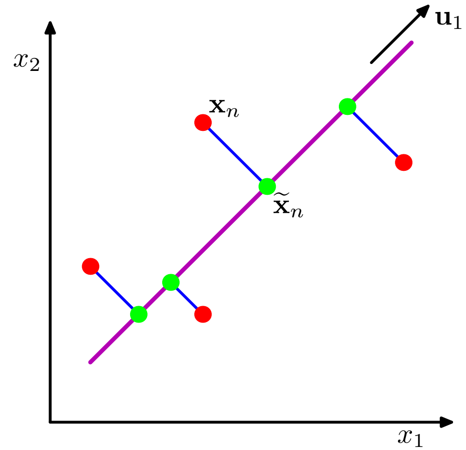
- Cat visual cortex (Hubel & Wiesel)
- SOM model (Obermayer et al.)







Principal Component Analysis



- D-dim data $\mathbf{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_N\}^T$ (zero mean)
- Find a low dimensional subspace with
  - largest variance
  - minimal projection error
- M-dim projection $\mathbf{z}_n=\mathbf{V}\mathbf{x}_n$ by $\mathbf{V}=\{\mathbf{v}_1,\ldots,\mathbf{v}_M\}^T$, $||\mathbf{v}_m||=1$
  - data covariance $\mathbf{C} = \mathbf{X}^T\mathbf{X}/N = \sum_n \mathbf{x}_n\mathbf{x}_n^T/N$
  - covariance of $\mathbf{z}$: $\mathbf{V}^T\mathbf{C}\mathbf{V}$
- Eigenvectors $\mathbf{v}_1,\ldots,\mathbf{v}_M$ for $\mathbf{C}\mathbf{v}_m=\lambda_m\mathbf{v}_m$

  with largest eigenvalues $\lambda_1,\ldots,\lambda_M$
- Singular value decomposition (SVD): $\mathbf{X}=\mathbf{U}\mathbf{S}\mathbf{V}^T$

  $\mathbf{V}$:eigenvectors of $\mathbf{X}^T\mathbf{X}$, $\mathbf{S}^2=\mathrm{diag}(\lambda_1,\ldots,\lambda_D)$

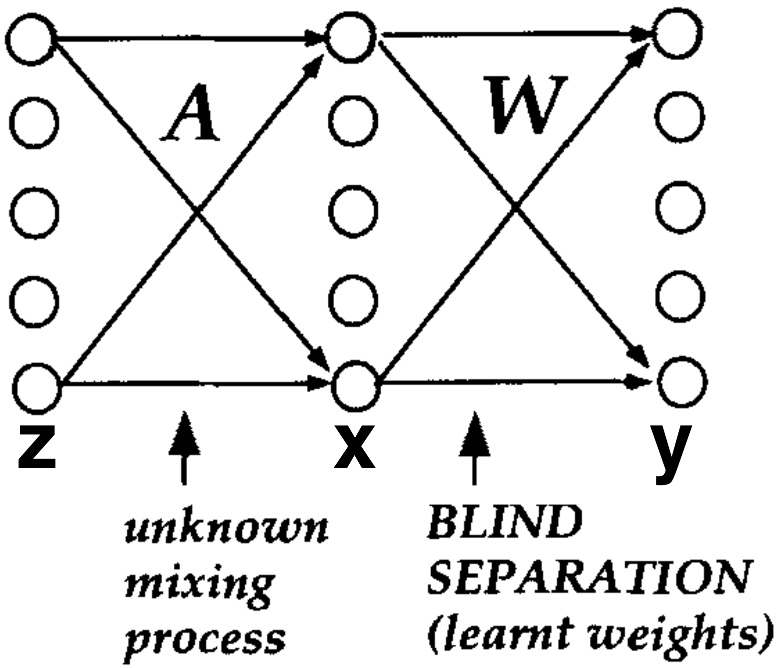
PCA by Autoencoder

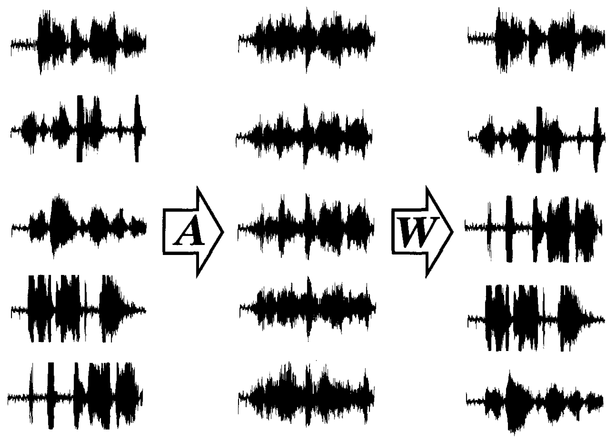
Independent Component Analysis

- Latent variable with independent components

  $p(\mathbf{z}) = \prod_j p(z_j)$ … usually non-Gaussian

- Observed data: $\mathbf{x} = \mathbf{Az}$
  - blind source separation: $\mathbf{y} = \mathbf{Wx}$





Infomax and ICA

- Mutual information of input $\mathbf{x}$ and output $\mathbf{y}$
  - $I(\mathbf{y},\mathbf{x}) = H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x})$

    $H(\mathbf{y}|\mathbf{x})$: constant for additive noise
  - maximum information transfer $I(\mathbf{y},\mathbf{x})$
  - maximum output entropy $H(\mathbf{y})$
  - independence of outputs $p(\mathbf{y}) = \prod_j p(y_j)$
- Sigmoid output: $\mathbf{y} = tanh(\mathbf{u})$, $\mathbf{u} = \mathbf{W}\mathbf{x}$
  - $\Delta\mathbf{W} \propto \partial H(\mathbf{y})/\partial\mathbf{W}\ \mathbf{W}^T\mathbf{W}$ ...natural gradient

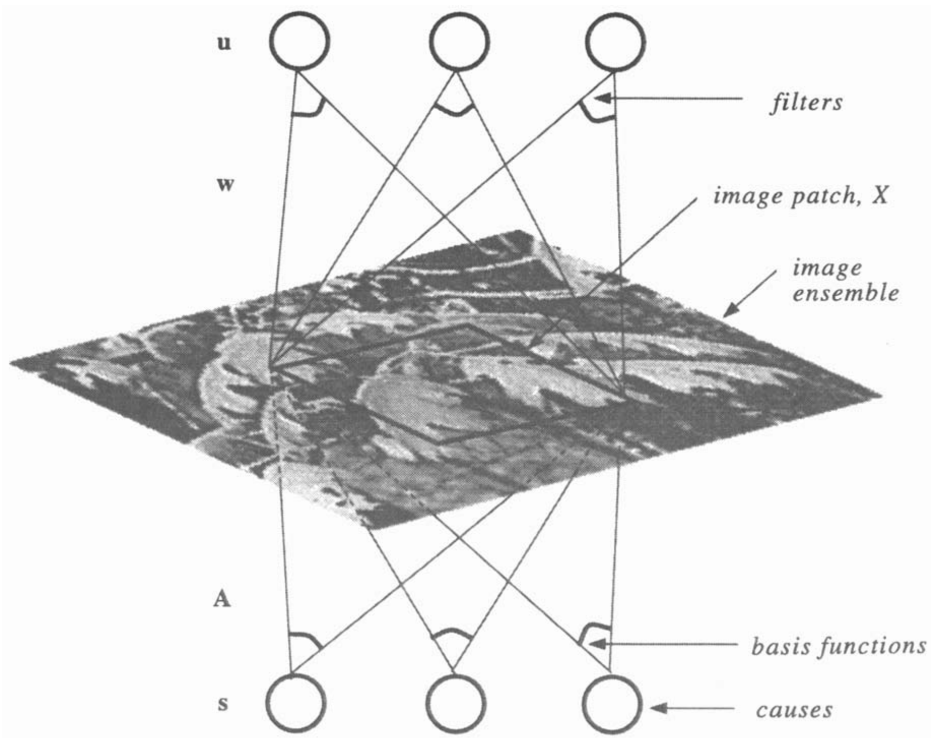
    $= ([\mathbf{W}^T]^{-1}+\mathbf{y}\mathbf{x}^T)\mathbf{W}^T\mathbf{W}$

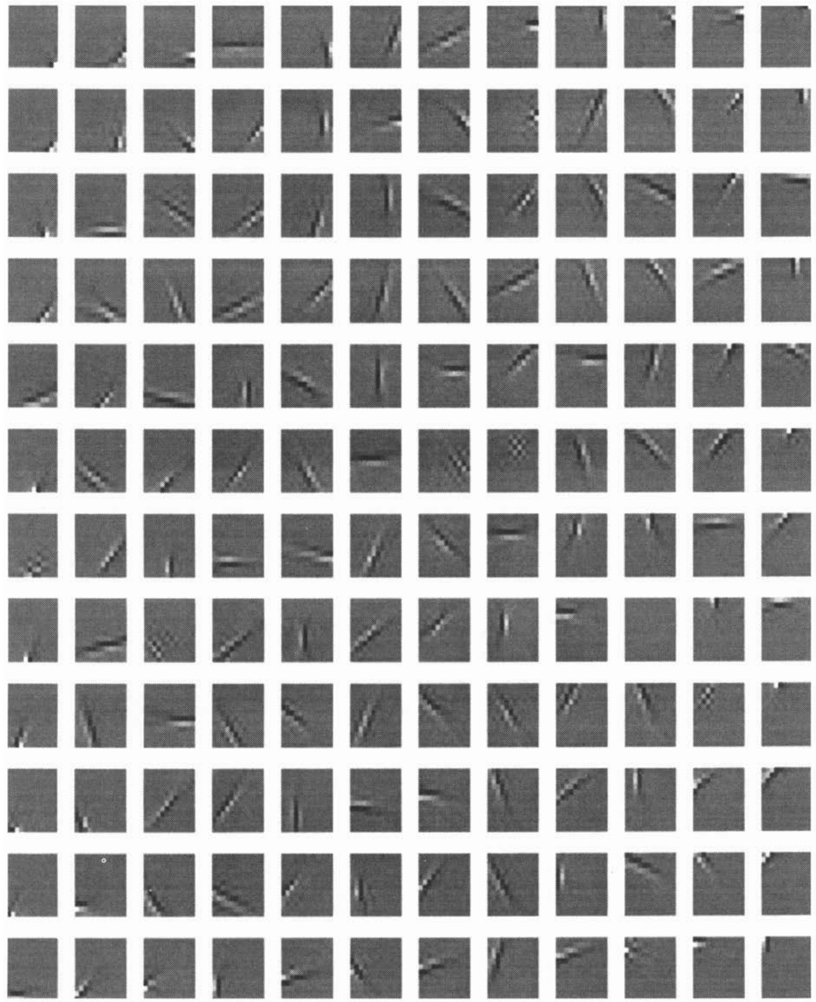
    $= (\mathbf{I}+\mathbf{y}\mathbf{u}^T)\mathbf{W}$

Visual Feature Detector by ICA

(Bell & Sejnowski 1997)

- Independent components of natural images





Mental Simulation

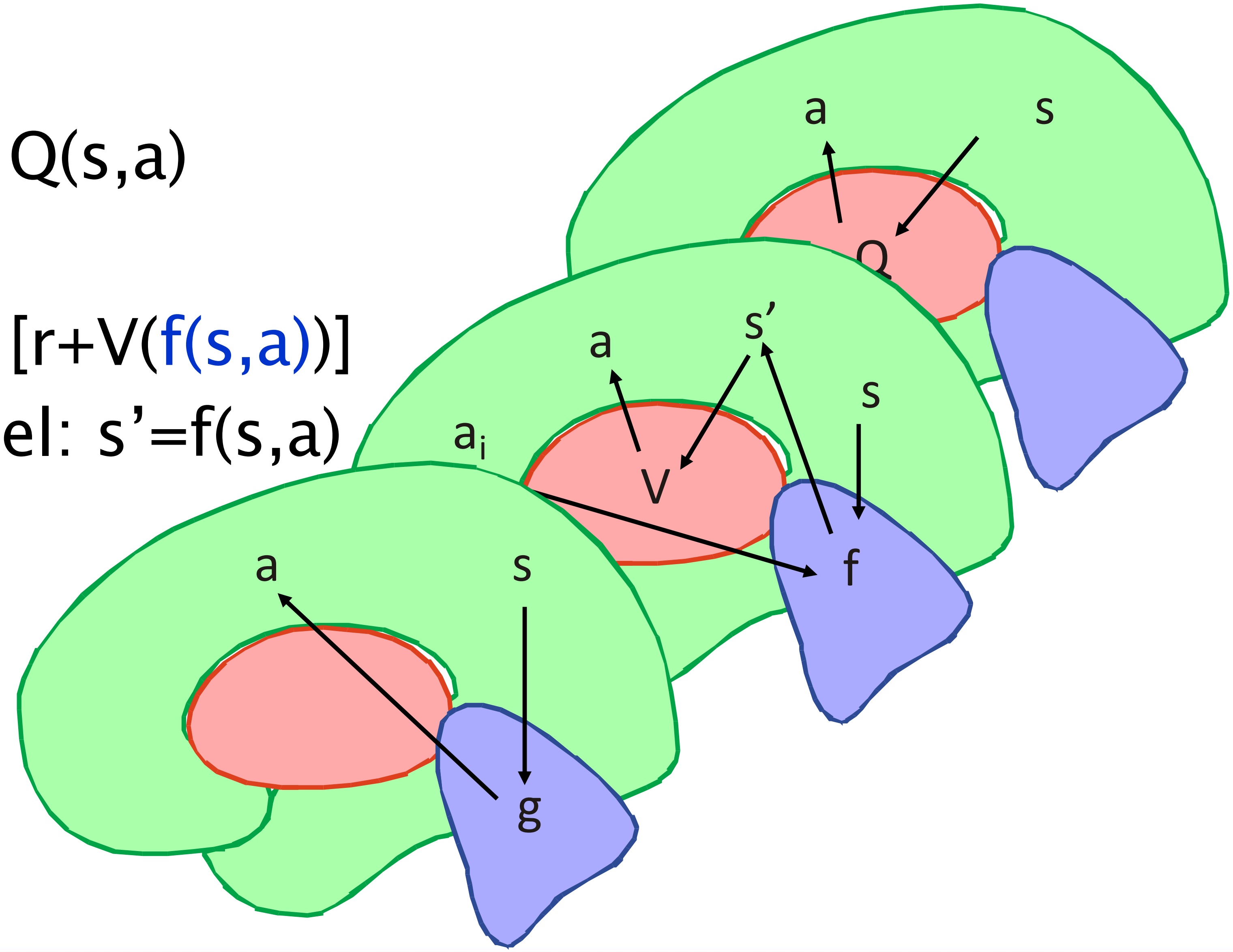
Brain's process using
an action-dependent state transition model
$s'=f(s,a)$ or $P(s'|s,a)$

- Estimate the present from past state/action
  - perception under noise/delay/occlusion
- Predicting the future
  - model-based decision, action planning
- Imagining in a virtual world
  - thinking, language, science,…

Multiple Ways of Action Selection

- Model-free
  - $a = \text{argmax}_a\ Q(s,a)$
- Model-based
  - $a = \text{argmax}_a\ [r+V(f(s,a))]$

    forward model: $s'=f(s,a)$
- Memory-based
  - $a = g(s)$



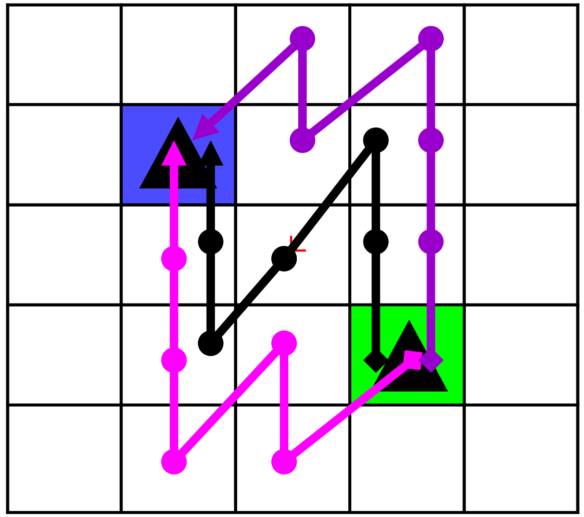
SCIENTIFIC REPORTS

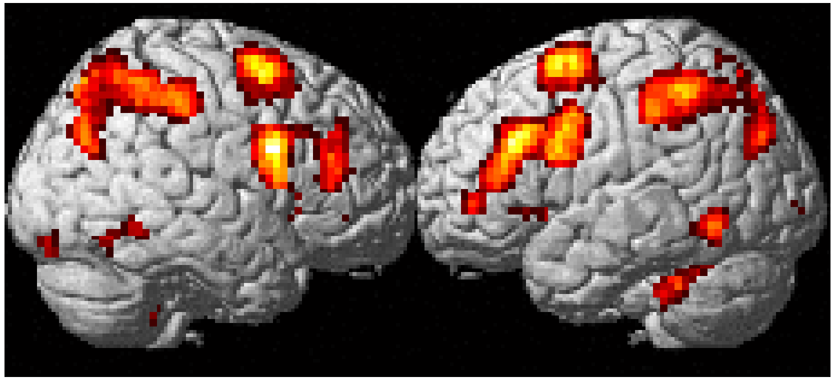
OPEN

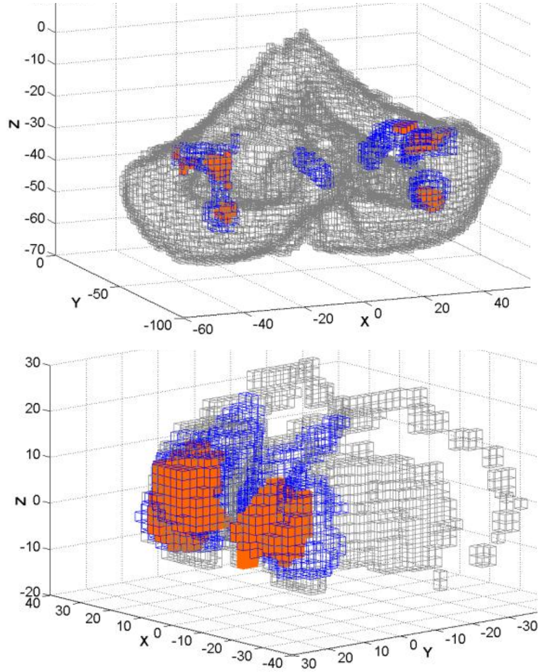
Model-based action planning involves cortico-cerebellar and basal ganglia networks

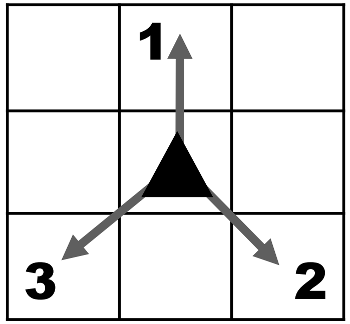
Received: 16 February 2016
Accepted: 19 July 2016

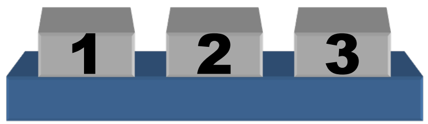
Alan S. R. Fermin[1,2,3], Takehiko Yoshida[1,2], Junichiro Yoshimoto[1,2], Makoto Ito[2], Saori C. Tanaka[4] & Kenji Doya[1,2,3,4]











How to Witness Mental Simulation

- State transition model $s' = f(s,a)$ or $P(s'|s,a)$
- Action planning
  - predict action value

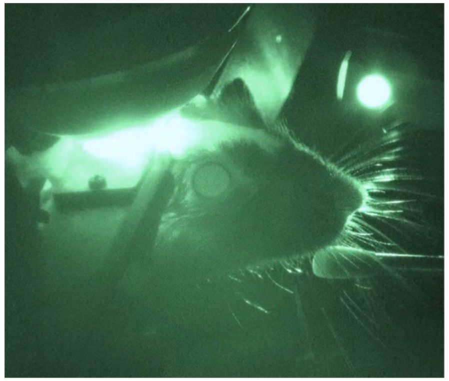
$$Q(s,a) = \Sigma_{s'} P(s'|s,a)[R(s,a)+\gamma V(s')]$$

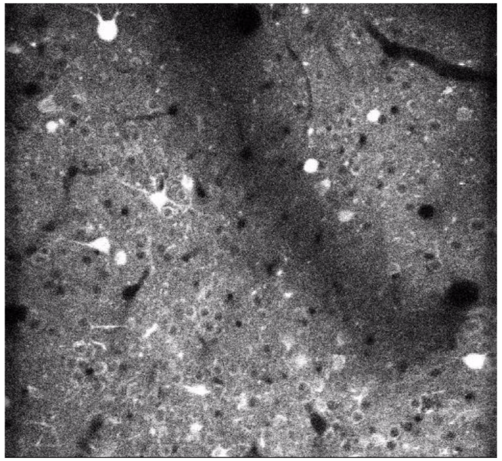
  - hard to tell when/which action is imagined
- State estimation
  - dynamic Bayesian filter

$$P(s_t | a_{t-1}, o_t) \propto P(o_t|s_t)\, \Sigma_{s_{t-1}} P(s_t|s_{t-1}, a_{t-1})\, P(s_{t-1})$$

  - can compare actual and predicted states nature neuroscience

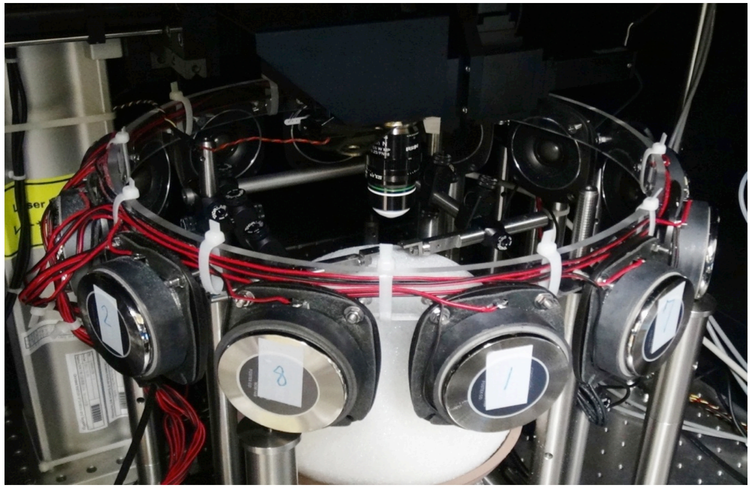
Neural substrate of dynamic Bayesian inference in the cerebral cortex





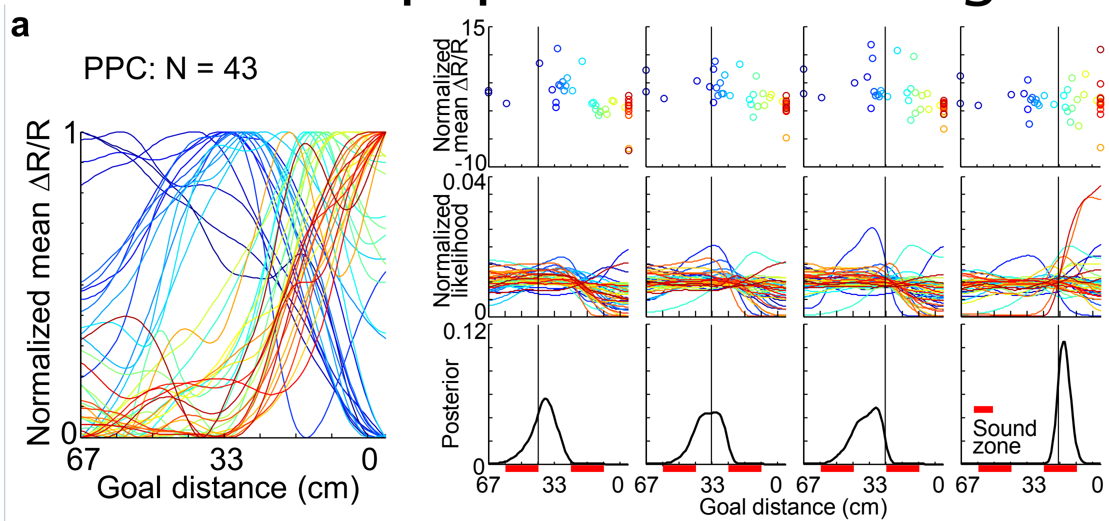
Akihiro Funamizu[1,2], Bernd Kuhn[2] & Kenji Doya[1]

- PPC two-photon imaging



- Auditory virtual environment
  - intermittent sensory input

- Probabilistic population decoding



  - predicting goal distance from action

Reinforcement Learning

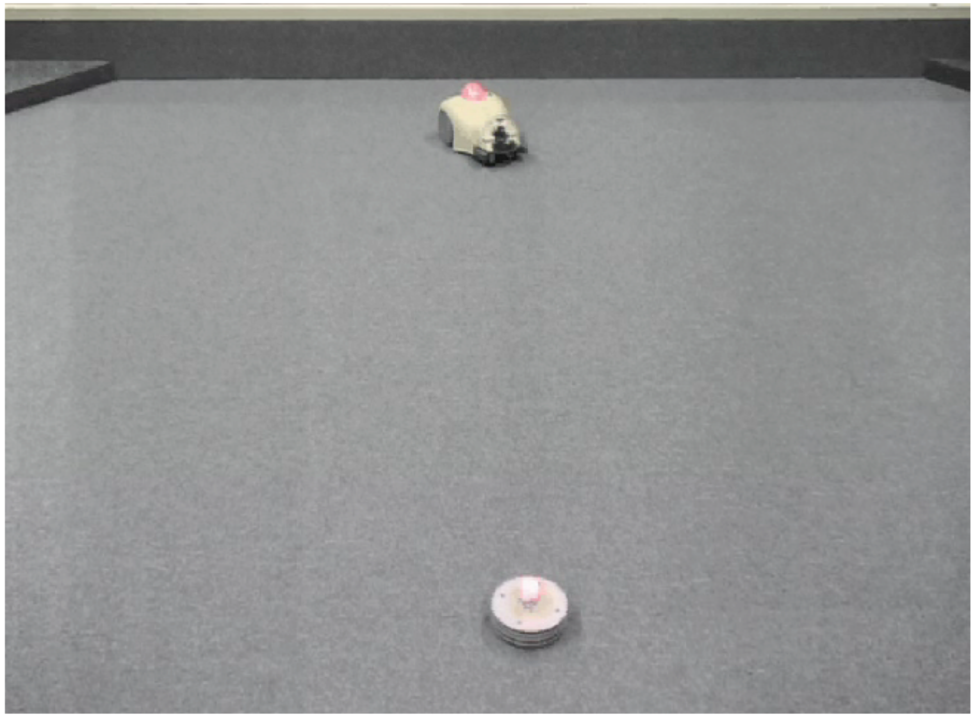
- Predict reward: *value function*
  - $V(s) = E[\ r(t) + \gamma r(t+1) + \gamma^2 r(t+2) \ldots |\ s(t)=s]$
  - $Q(s,a) = E[\ r(t) + \gamma r(t+1) + \gamma^2 r(t+2) \ldots |\ s(t)=s,\ a(t)=a]$
- Select action
  - *greedy*: $a = \operatorname{argmax} Q(s,a)$
  - *Boltzmann*: $P(a|s) \propto \exp[\ \beta\ Q(s,a)]$
- Update prediction: *TD error*
  - $\delta(t) = \underline{r(t) + \gamma V(s(t+1))} - V(s(t))$
  - $\Delta V(s(t)) = \alpha\ \delta(t)$
  - $\Delta Q(s(t),a(t)) = \alpha\ \delta(t)$

How to implement these steps?

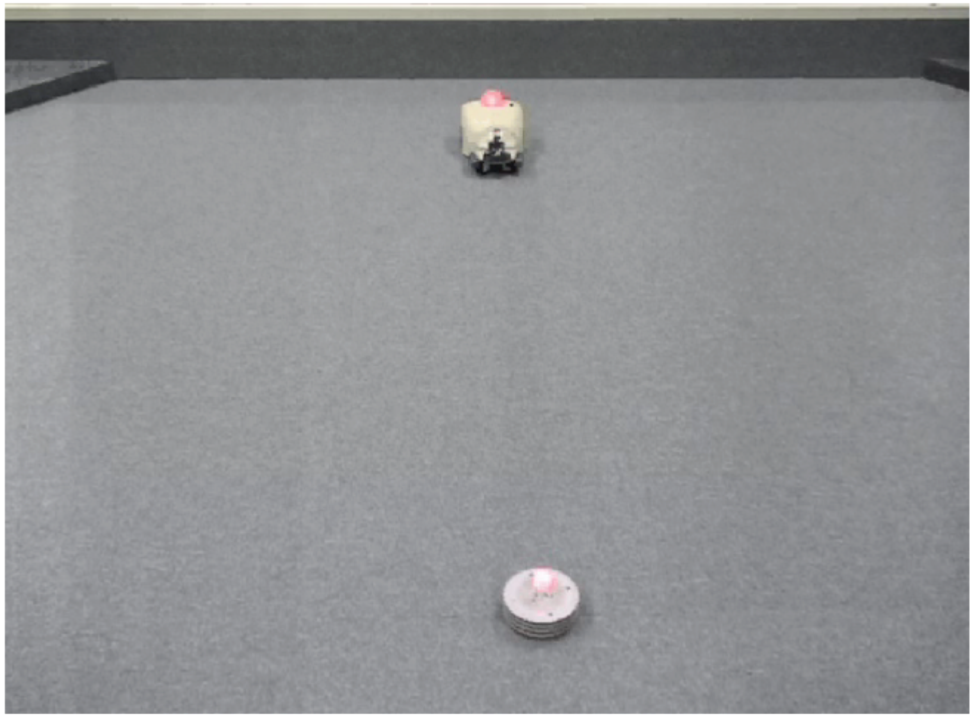
How to tune these parameters?

Temporal Discount Factor γ

- Large γ
  - reach for far reward



- Small γ
  - only to near reward



Temporal Discount Factor γ

- $V(t) = E[\, r(t) + \gamma r(t+1) + \gamma^2 r(t+2) + \gamma^3 r(t+3) + \ldots]$
  - controls the 'character' of an agent

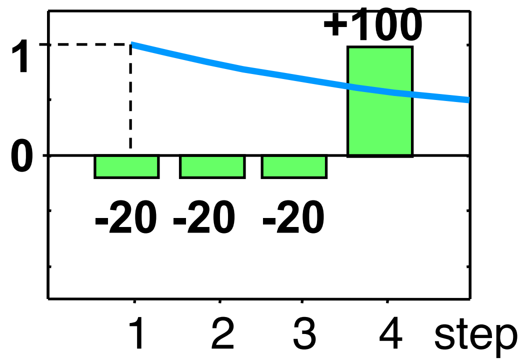
γ large

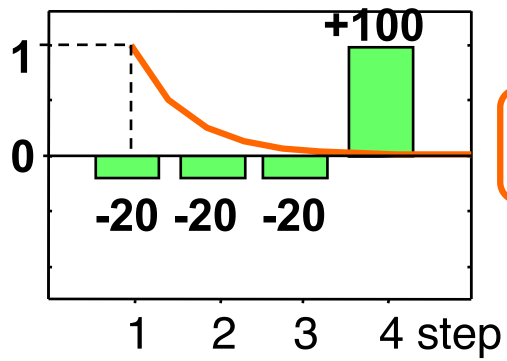
γ small

Depression?

no pain, no gain!

V =18.7





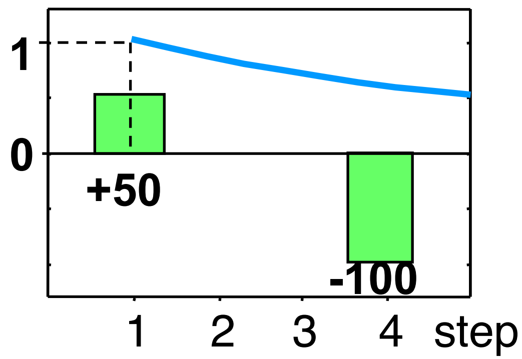
better stay idle

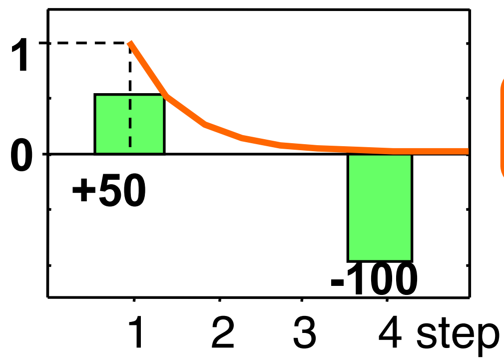
V =-25.1

Impulsivity?

stay away from danger

V = -22.9



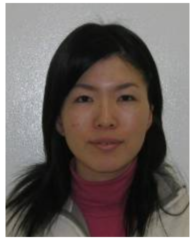


can't resist temptation

V = 47.3

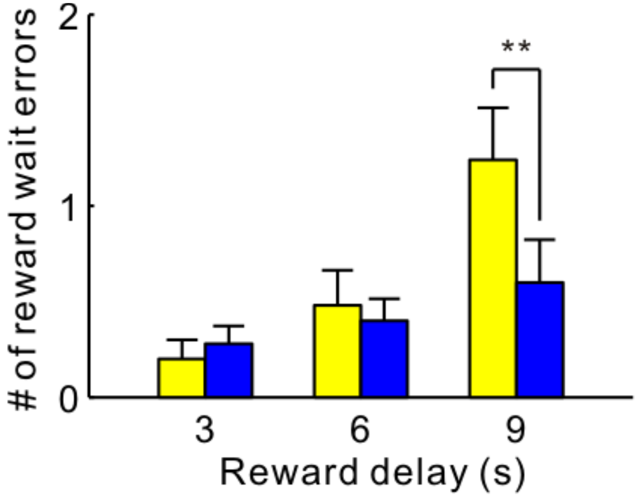
Serotonin?

Optogenetic Stimulation of Dorsal Raphe Serotonin Neurons

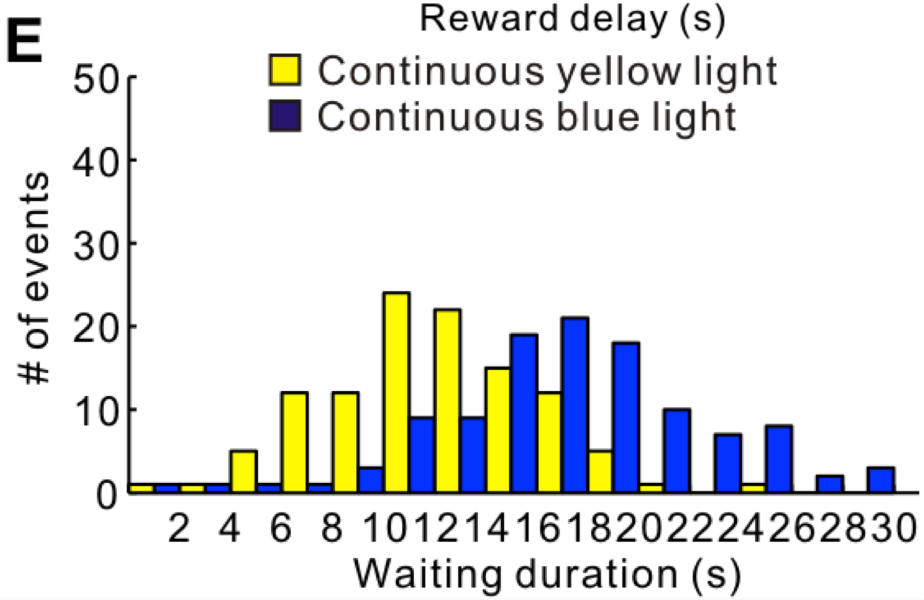


(Miyazaki et al., 2014, Current Biology)

- Reward Delay Task (3, 6, 9, ∞ sec)



- 3 sec: success
- omission: 12.1 s
- omission: 20.8 s



ARTICLE

DOI: 10.1038/s41467-018-04496-y

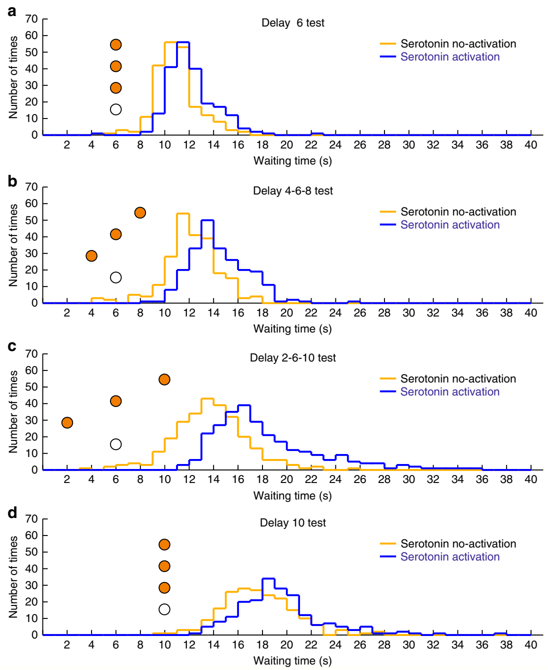
OPEN

nature COMMUNICATIONS (2018)

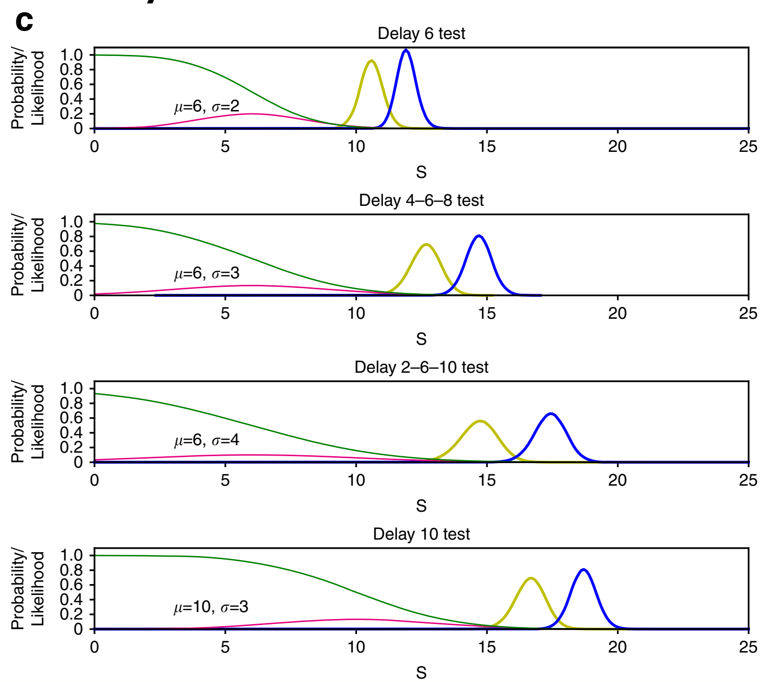
Reward probability and timing uncertainty alter the effect of dorsal raphe serotonin neurons on patience

Katsuhiko Miyazaki[1], Kayoko W. Miyazaki[1], Akihiro Yamanaka[2], Tomoki Tokuda[3], Kenji F. Tanaka[4] & Kenji Doya[1]

- Serotonin stimulation facilitates waiting when...
  - reward probability is high
  - reward timing is uncertain
- Prior for reward?



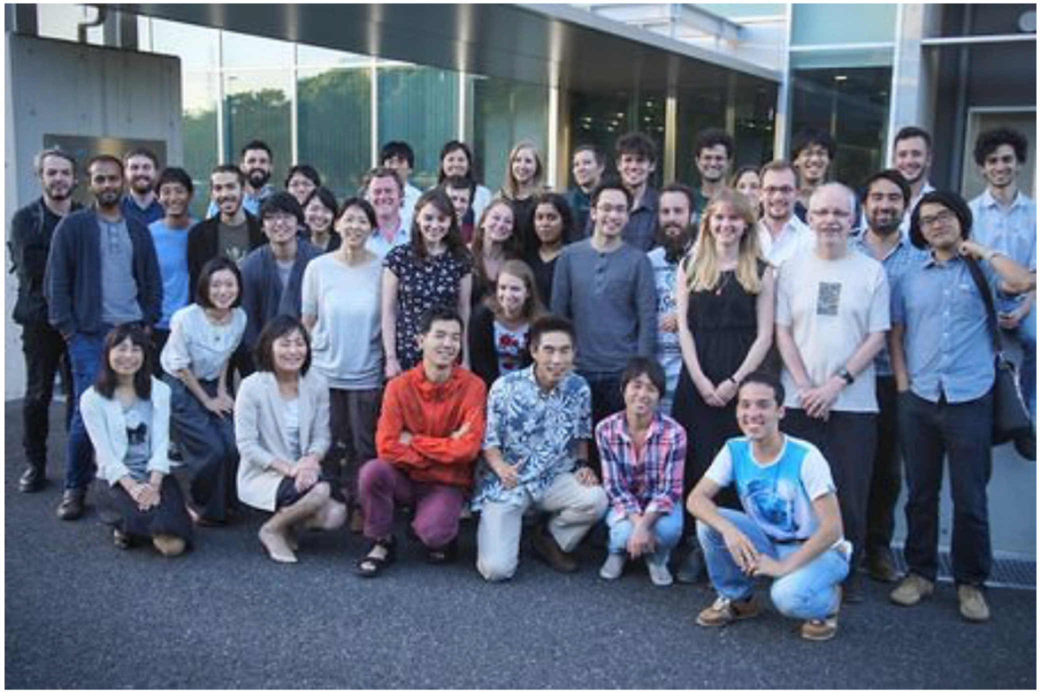
Bayesian decision model



- ISSA Summer School 2017
- Program
- Hands-on Projects
- People
- Organizers
- Practical Information

ISSA Summer School 2017

groups.oist.jp/issa



The Initiative for a Synthesis in Studies of Awareness, ISSA for short, is a group of scientists and scholars advocating an integrative approach to the study of awareness.

ISSA Summer School 2017

ISSA will organize a two-week Summer School, with plenary lectures in the morning and parallel sessions in the afternoon, in which the lecturers will lead study groups that are aimed at producing original research of publishable quality. The lectures will cover topics in various aspects of neuroscience, experimental as well as computational; theoretical physics; logic and philosophy; and various other fields in cognitive science and the study of complex systems, including artificial intelligence, artificial life, and robotics.

Date: May 22nd – June 2nd, 2017

Venue: Center for Information and Neural Networks (CiNet)
Osaka University, Suita, Japan

Initiative for Synthetic Studies of Awareness (ISSA): Introduction

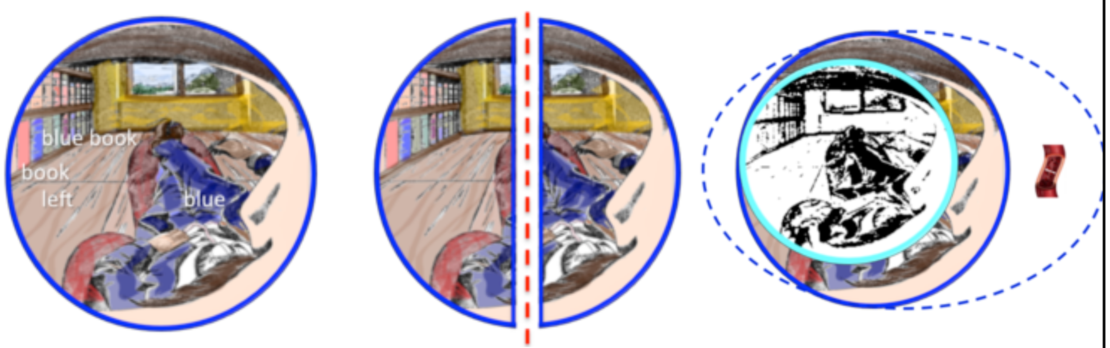
2017 May 22
APröf Nao Tsuchiya
Monash University, Australia

Integrated information theory of consciousness

- Starts from phenomenology, identifies five axioms (1. existence, 2. composition, 3. information, 4. integration, 5. exclusion)
- Tries to seek for potential physical mechanisms that can support the axioms

Tononi 2004, 2008, Oizumi et al 2014 PLoS Comp Bio

Consciousness is **composed** of various aspects
Consciousness is **integrated**
Consciousness is **excluded** outside of the certain boundary



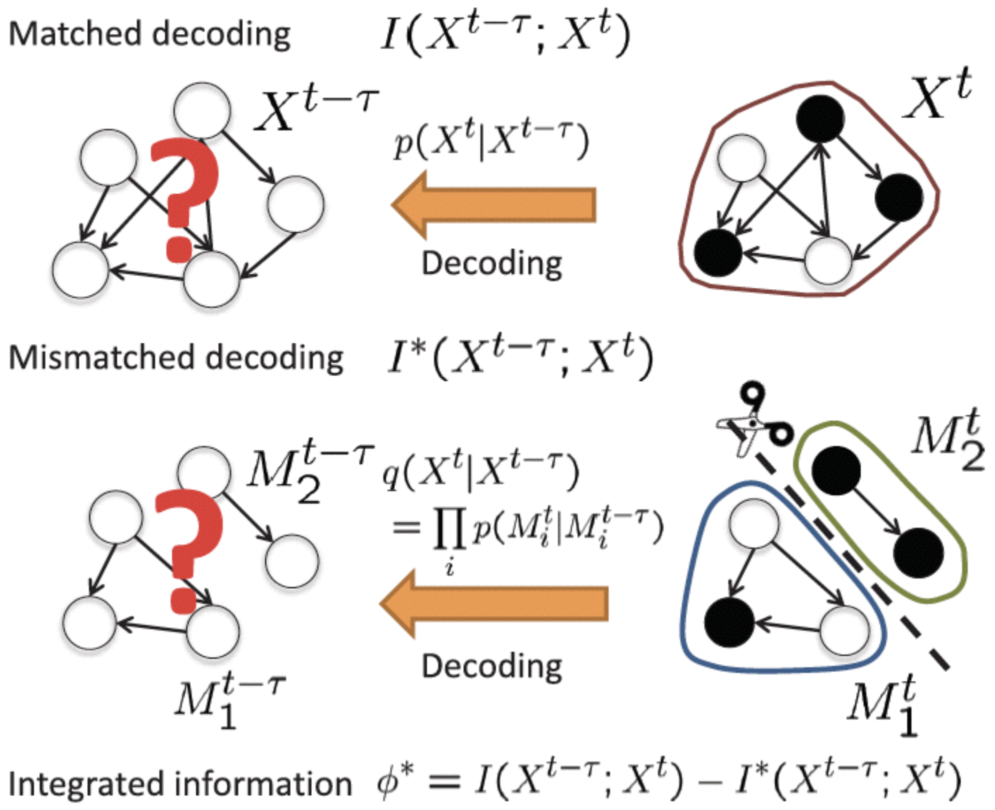
Intuitive explanation 1

Integrated information

=

loss of predictability of past states based on current states, when system is cut

(Oizumi et al 2016 PLoS Comp)

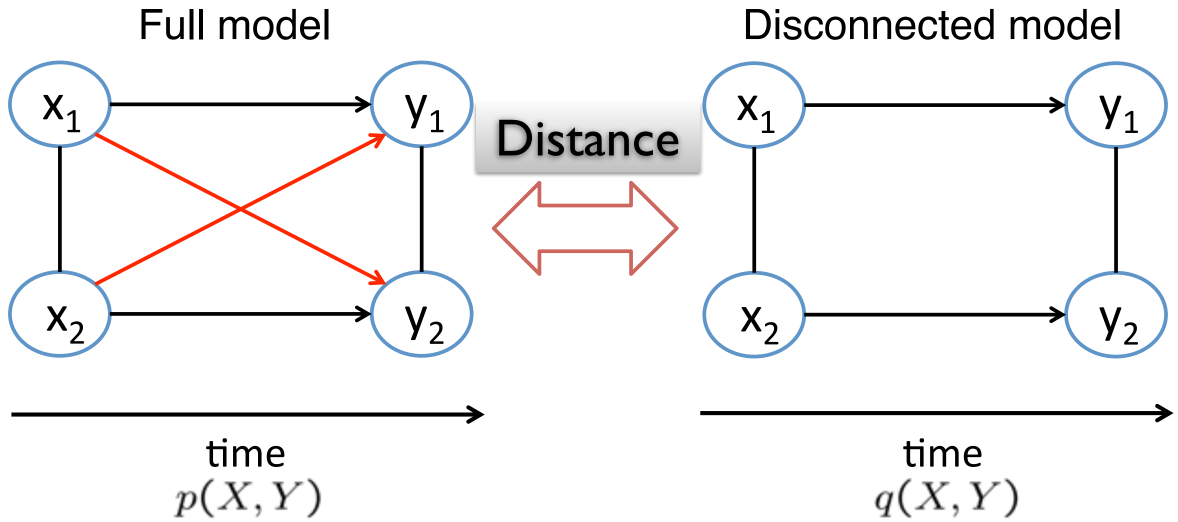


(Oizumi et al 2016 PLoS Comp)

Intuitive explanation 2

Integrated information
=
a *distance* between the actual and disconnected model

Oizumi, Tsuchiya, Amari 2016 PNAS



Integrated information

$$\boxed{\Phi_G = \min_{q(X,Y)} D\left(p(X,Y)||q(X,Y)\right)}$$

$$\Phi_G = \frac{1}{2}\log\frac{|\Sigma(E)'|}{|\Sigma(E)|}.$$

Tinbergen's Four Questions

How?

proximate view

- Mechanism

  How does it work?

- Ontogeny

  How does it develop?

Why?

ultimate/evolutionary view

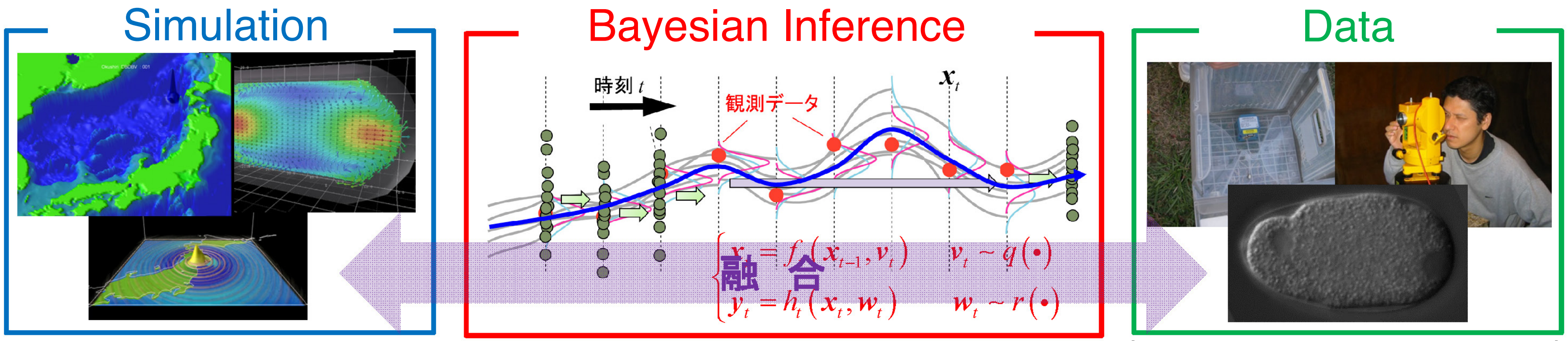
- Function

  What is it for?

- Phylogeny

  How did it evolve?

Consciousness as Data Assimilation?



(Komaki Lab. U Tokyo)

- Generative model
  - dynamics $\mathbf{x}_t = f(\mathbf{x}_{t-1}) + \varepsilon_s$ atmosphere, ocean,...
  - observation $\mathbf{y}_t = g(\mathbf{x}_t) + \varepsilon_o$ temperature, wind,...
- State estimation by real-time simulation
  - utilizing sparse, multi-modal data
- Prediction of future states
- Postdiction of past history

Characteristics akin to consciousness

- reason for the emergence of conscious phenomenology?

Tinbergen's Four Questions

How?

proximate view

- Mechanism

  How does it work?

  - dynamic Bayesian inference

- Ontogeny

  How does it develop?

  - learn internal models

Why?

ultimate/evolutionary view

- Function

  What is it for?

  - estimate the past and present
  - predict the future

- Phylogeny

  How did it evolve?

  - mammalian neocortex

Universe, Life, and Intelligence

To explain the universe,
people called for the god(s).

To explain life,
people called for soul.

To explain human intelligence,
people now call for consciousness.

Acknowledgements

- Striatum recording
  - **Makoto Ito (Progress Technology)**
  - **Tomohiko Yoshizawa (Tamagawa U)**
  - **Charles Gerfen (NIH)**
  - Kazuyuki Samejima (Tamagawa U)
  - Minoru Kimura (Tamagawa U)
- Cortical imaging
  - **Akihiro Funamizu (CSHL)**
  - **Bernd Kuhn**
  - **Yuzhe Li**
  - **Sergey Zobnin**
- Human fMRI/behavior
  - **Alan Fermin (Tamagawa U)**
  - **Takehiko Yoshida (NAIST)**
  - **Saori Tanaka (ATR)**
  - **Nicolas Schweighofer (USC)**
  - **Sigeto Yamawaki (Hiroshima U)**
  - Yu Shimizu
  - **Tomoki Tokuda (NAIST)**
  - Shoko Ota
- Serotonin Recording
  - **Kayoko W Miyazaki**
  - **Katsuhiko Miyazaki**
- Optogenetics
  - **Kenji Tanaka (Keio U)**
  - **Akihiro Yamanaka (Nagoya U)**
- Marmoset data analysis
  - Carlos Gutierrez
  - Hiromichi Tsukada
  - **Jun Yoshimoto**, Takashi Nakano (NAIST)
  - Junichi Hata, Alex Woodward (RIKEN)
  - Ken Nakae, Henrik Skibbe (Kyoto U)
- Spiking neural network model
  - Jun Igarashi, Sun Zhe (RIKEN)
  - Tadashi Yamazaki, Hiroshi Yamamura (UEC)
  - Jan Moren
  - Osamu Shouno (HRI)
  - Benoit Girard, Daphne Heraiz (UPMC)
  - Jean Lienard
- Robotics
  - **Jun Morimoto (ATR)**
  - **Eiji Uchibe (ATR)**
  - **Stefan Elfwing (ATR)**
  - **Jiexin Wang (ATR), Naoto Yoshida (Groove X)**
  - **Paavo Parmas**

Scientific Research on Innovative Areas
Strategic Research Program for Brain Sciences
Brain/MINDS Project
Post-K Supercomputing Program

International Symposium on AI and Brain Science

October 10-12, 2020, Ito Hall, U Tokyo

www.brain-ai.jp/symposium2020

- *Deep Learning and Reinforcement Learning*
  - Yann LeCun (NYU, Facebook)
  - Yutaka Matsuo (U Tokyo)
  - Doina Precup (McGill U)
  - David Silver (DeepMind)
  - Masashi Sugiyama (RIKEN AIP/IRCN)
- *World Model Learning and Inference*
  - Karl Friston (UCL)
  - Maneesh Sahani (Gatsby Unit)
  - Tadahiro Taniguchi (Ritsumeikan U)
  - Josh Tenenbaum (MIT)
- *Metacognition and Metalearning*
  - Matthew Botvinick (DeepMind)
  - Ryota Kanai (ARAYA)
  - Angela Langdon (Princeton U)
  - Hiroyuki Nakahara (RIKEN CBS)
  - Xiao-Jing Wang (NYU)
- *AI for Neuroscience and Neuromorphic Technologies*
  - Jim DiCarlo (MIT)
  - Yukiyaku Kamitani (Kyoto U)
  - Rosalyn Moran (King’s College London)
  - Terry Sejnowski (Salk Institute)
  - Hidehiko Takahashi (Tokyo MDU)
- *Social Impact and Neuro-AI Ethics*
  - Anne Churchland (CSHL)
  - Kenji Doya (OIST)
  - Arisa Ema (U Tokyo)
  - Hiroaki Kitano (SONY CSL)
  - Stuart Russell (UC Berkeley)