Machine Learning Summer School, March 8, 2024, OIST

What Can We Further Learn from the Brain for AI?

Kenji Doya doya@oist.jp
Neural Computation Unit groups.oist.jp/ncu
Okinawa Institute of Science and Technology Graduate University

OCNC: OIST Computational Neuroscience Course

OCNC2004

Okinawa Computational Neuroscience Course 2004

The aim of Okinawa Computational Neuroscience Course is to provide opportunities for young researchers with theoretical backgrounds to learn up-to-date neurobiological findings, and those with experiment backgrounds to have hands-on experience in computational modeling. This may also be a good opportunity for theoretical and experimental neuroscientists to meet together and enjoy attractive nature and culture of Okinawa, the southernmost island prefecture of Japan.

This course is the second of a series of tutorial courses that the Cabinet Office of the Japanese Government is sponsoring as a precursory activity for the Okinawa Institute of Science and Technology.

The sponsor will provide lodging expenses during the course and support for travel to Okinawa.

OIST Neural Computation Unit

Modeling

Yukako Yamane
Yuzhe Li
Soheil Keshmiri
Florian Larande
Shuhei Hara
Hideyuki Yoshimura
Shutashu Tomonaga
Yi-Shan Cheng
Yusaku Kasai

Robotics

Ekaterina Sangati
Kristine Roque
Yuji Kanagawa
Tojo Rakotoaritina

(PhD Students)

Neurobiology

Katsuhiko Miyazaki
Kayoko W Miyazaki
Hajime Yamanaka
Anupama Chaud
Sergey Zobnin
Miles Desforges
Yuma Kajihara
Jianning Chen
Naohiro Yamauchi
Terezie Sedlinska

Administration

Kikuko Matsuo
Misuzu Saito

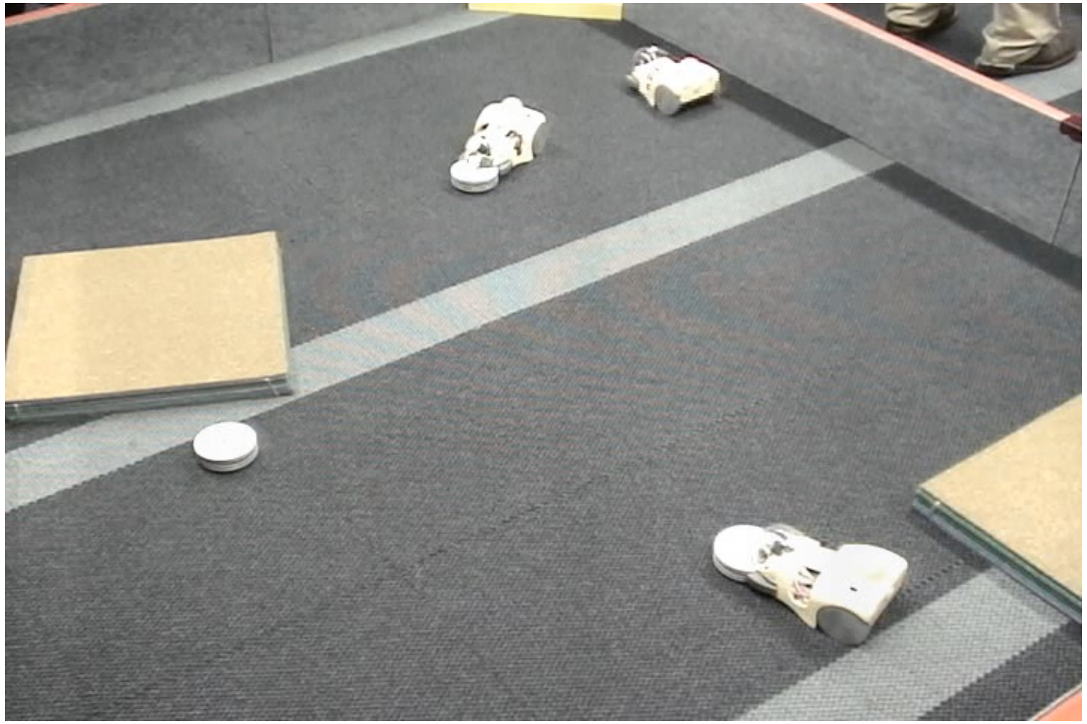
Professor

Kenji Doya

OIST Neural Computation Unit

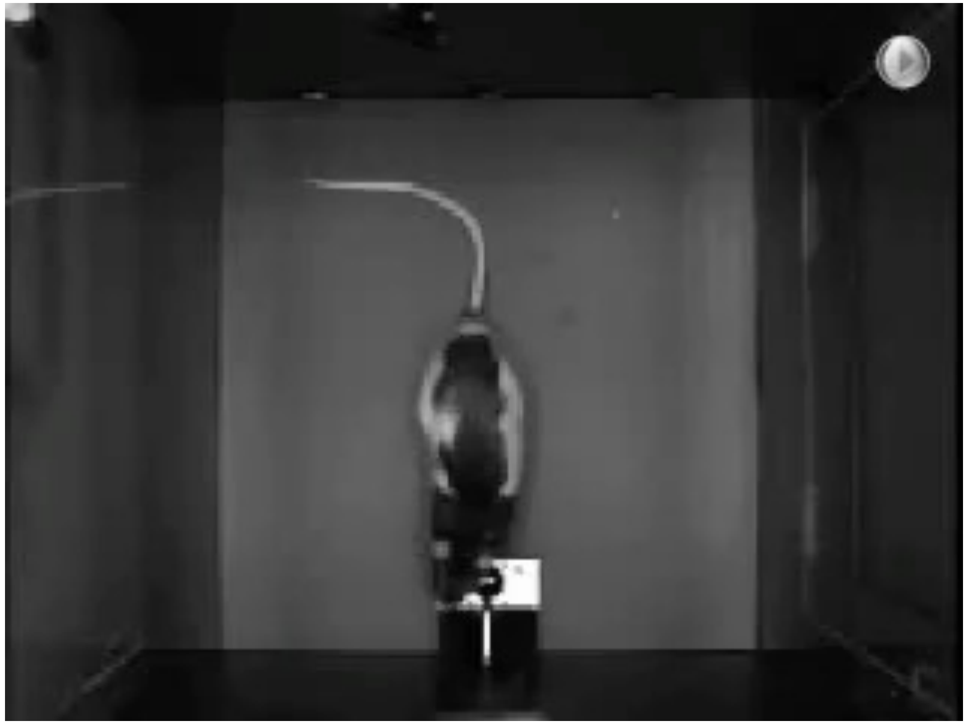
Create flexible learning systems

- robot experiments

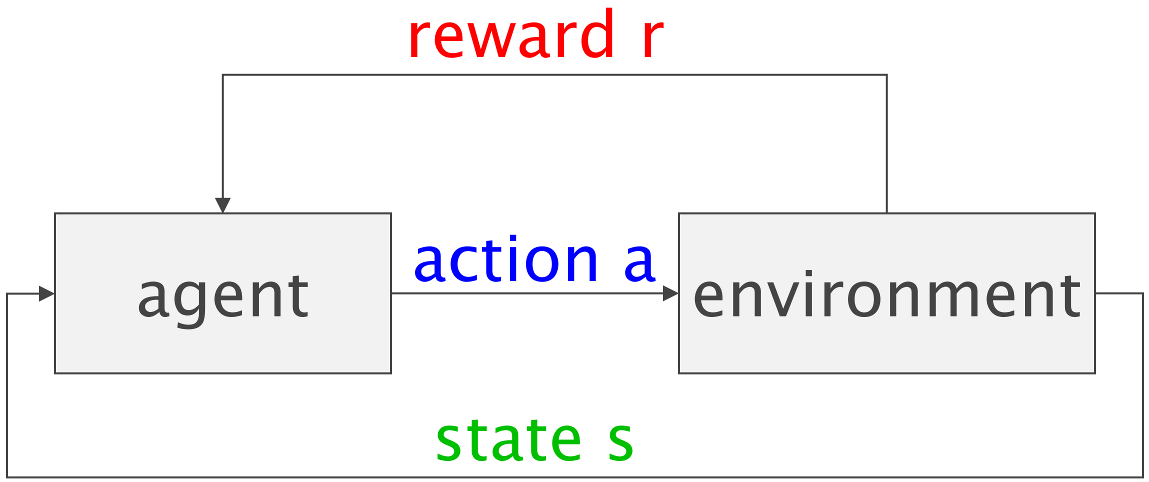


Reveal brain's learning mechanisms

- neurobiology



Reinforcement Learning



Learn action policy: s $\rightarrow$ a to maximize rewards

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

AI and Brain Science

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

There's a superb implementation of intelligence
in the brain, so why don't we learn from that.

AI in 20th century: program human expertise

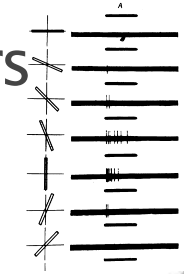
AI in 21st century: learn from big data

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

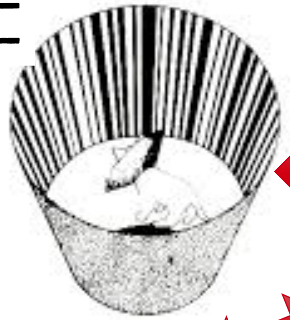
Coevolution in Pattern Recognition

Brain Science

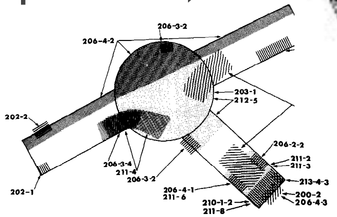
- Feature detectors
  (Hubel & Wiesel 1959)



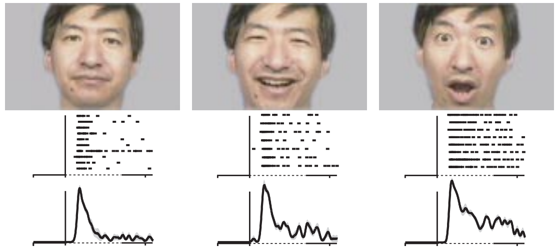
- Experience dependence
  (Blakemore & Cooper 1970)



- Place cell
  (O'Keefe 1976)

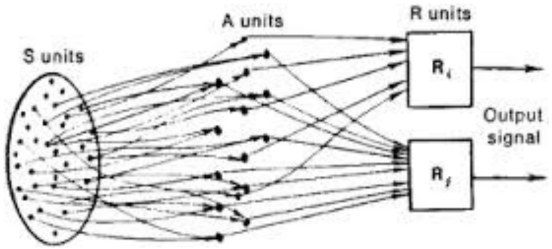


- Face cell (Bruce, Desimone, Gross 1981)

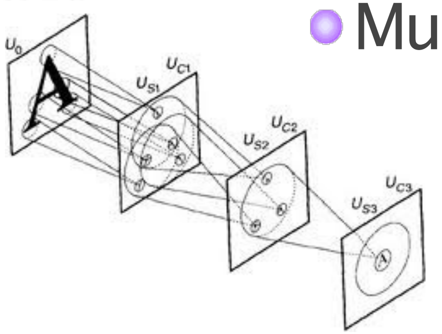


(Sugase et al. 1999)

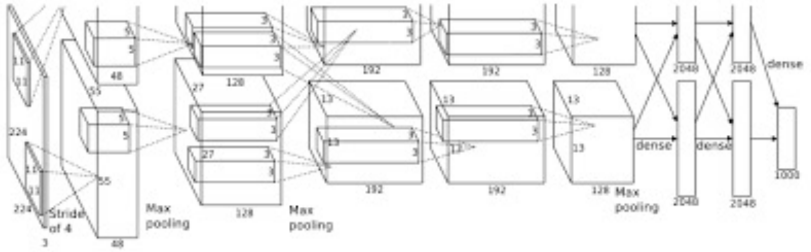
Artificial Intelligence

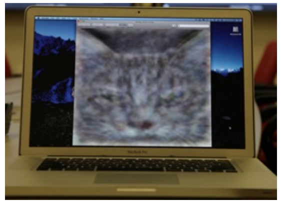


- Perceptron
  (Rosenblatt 1962)



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

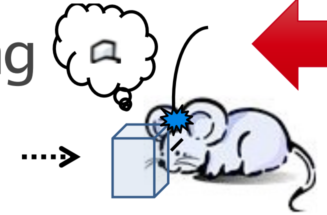




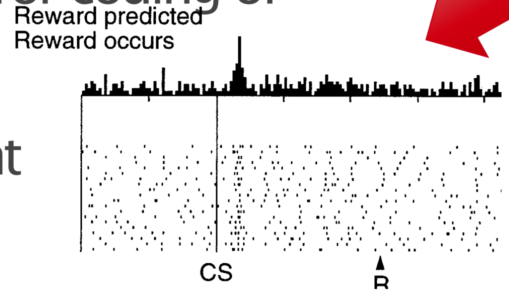
Coevolution in Reinforcement Learning

Brain Science

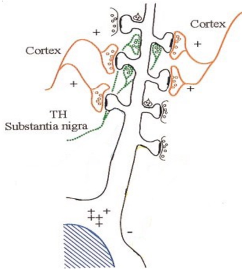
- Classic conditioning (Pavlov 1903)
- Operant conditioning (Thorndike 1898, Skinner 1938)

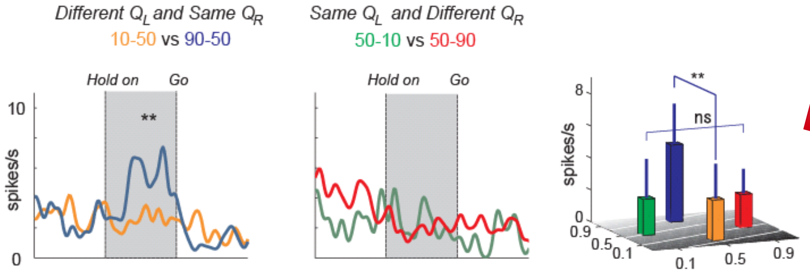


- Reward prediction error coding of dopamine neurons (Schultz et al. 1993, 1997)



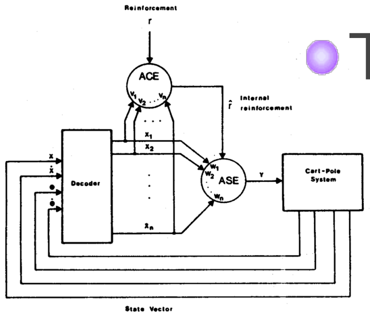
- Dopamine-dependent synaptic plasticity (Wickens et al. 2000)



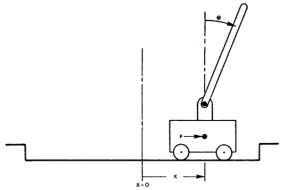


- Value coding in striatum (Samejima et al. 2005)

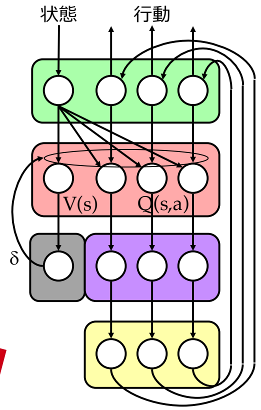
Artificial Intelligence

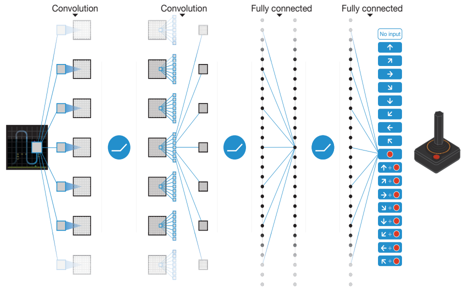


- TD learning (Barto et al. 1983)



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





- Deep Q network (Mnih et al. 2015)

NEWS

2020-04-01   Dr.Hiroaki Gomi’s group’s study is featured in eLife Press Release

2020-01-06   Prof. Kenji Doya received JNNS Academic Award and APNNS Outstanding Achievement Award

What Should We Further Learn from the Brain?

Energy Efficiency

Data Efficiency

- World Models and Mental Simulation
- Modularity and Compositionality
- Meta-learning

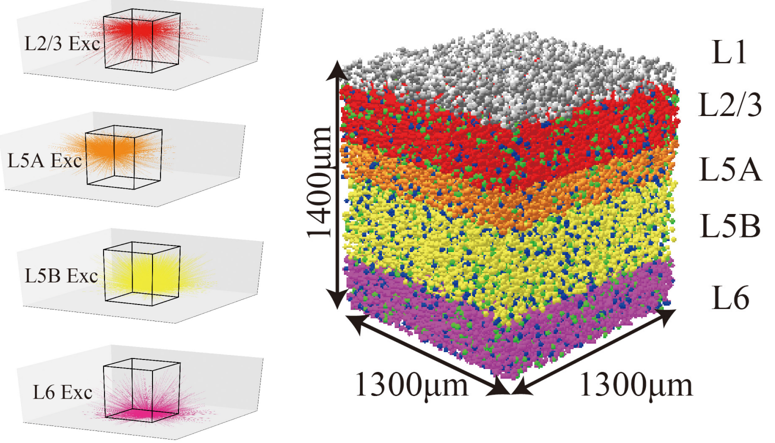
Autonomy and Sociality

Simulating Whole Human Brain Is Now Possible

frontiers in Neuroinformatics (2019)

Large-Scale Simulation of a Layered Cortical Sheet of Spiking Network Model Using a Tile Partitioning Method

Jun Igarashi[1], Hiroshi Yamaura[2] and Tadashi Yamazaki[2]**

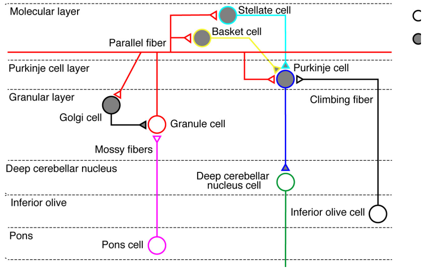


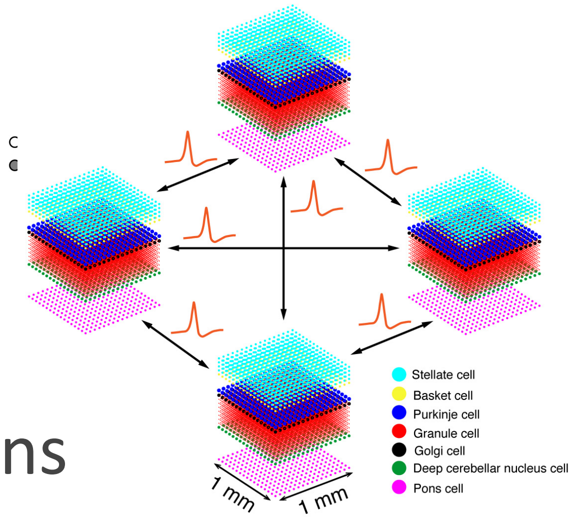
- 6 billion neurons
- 25 trillion synapses

frontiers in Neuroinformatics (2020)

Simulation of a Human-Scale Cerebellar Network Model on the K Computer

*Hiroshi Yamaura[1†], Jun Igarashi[2†] and Tadashi Yamazaki[1]**

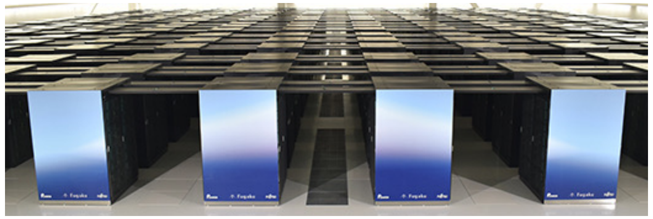




- 68 billion neurons
- 5 trillion synapses

■ Cortex + Cerebellum on Fugaku (2021)

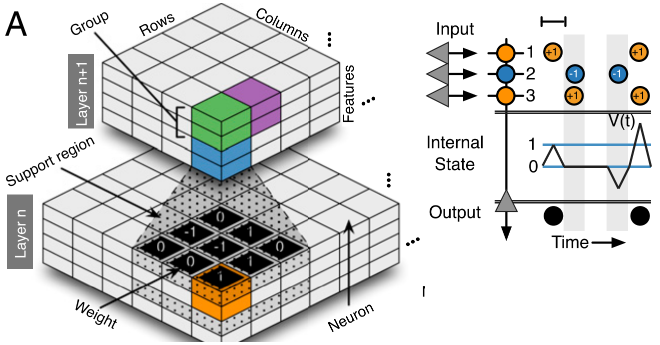
- 96 billion neurons
- 57 trillion synapses



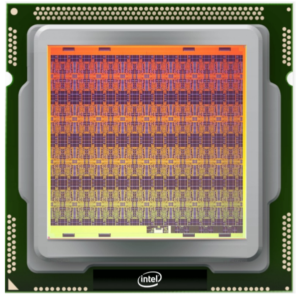
Neuromorphic Chips

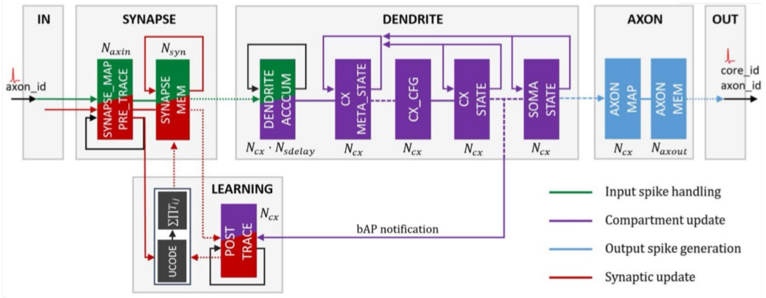
Convolutional networks for fast, energy-efficient neuromorphic computing

Steven K. Esser[a,1], Paul A. Merolla[a], John V. Arthur[a], Andrew S. Cassidy[a], Rathinakumar Appuswamy[a], Alexander Andreopoulos[a], David J. Berg[a], Jeffrey L. McKinstry[a], Timothy Melano[a], Davis R. Barch[a], Carmelo di Nolfo[a], Pallab Datta[a], Arnon Amir[a], Brian Taba[a], Myron D. Flickner[a], and Dharmendra S. Modha[a]



Loihi: A Neuromorphic Manycore Processor with On-Chip Learning

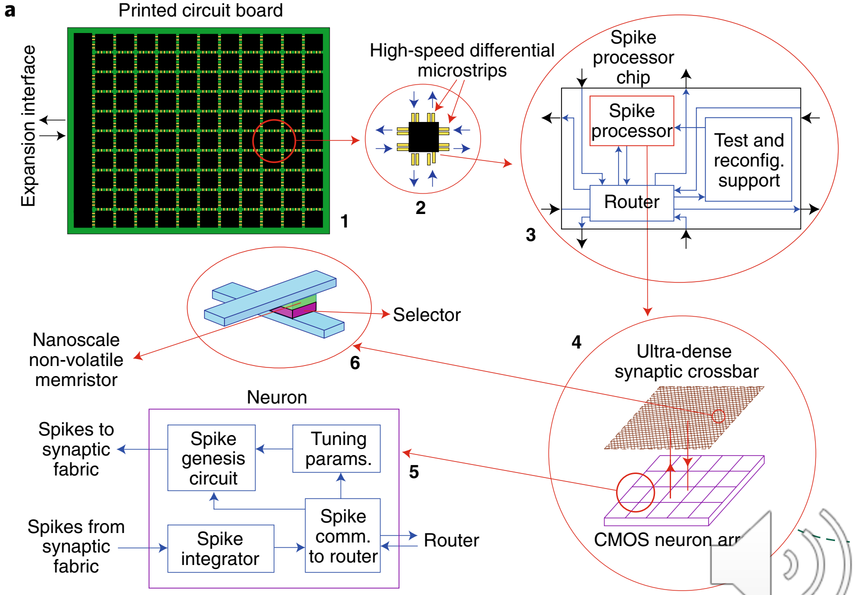




news & views

ARTIFICIAL NEURAL NETWORKS

Memristors fire away



What Should We Further Learn from the Brain?

Energy Efficiency

Data Efficiency

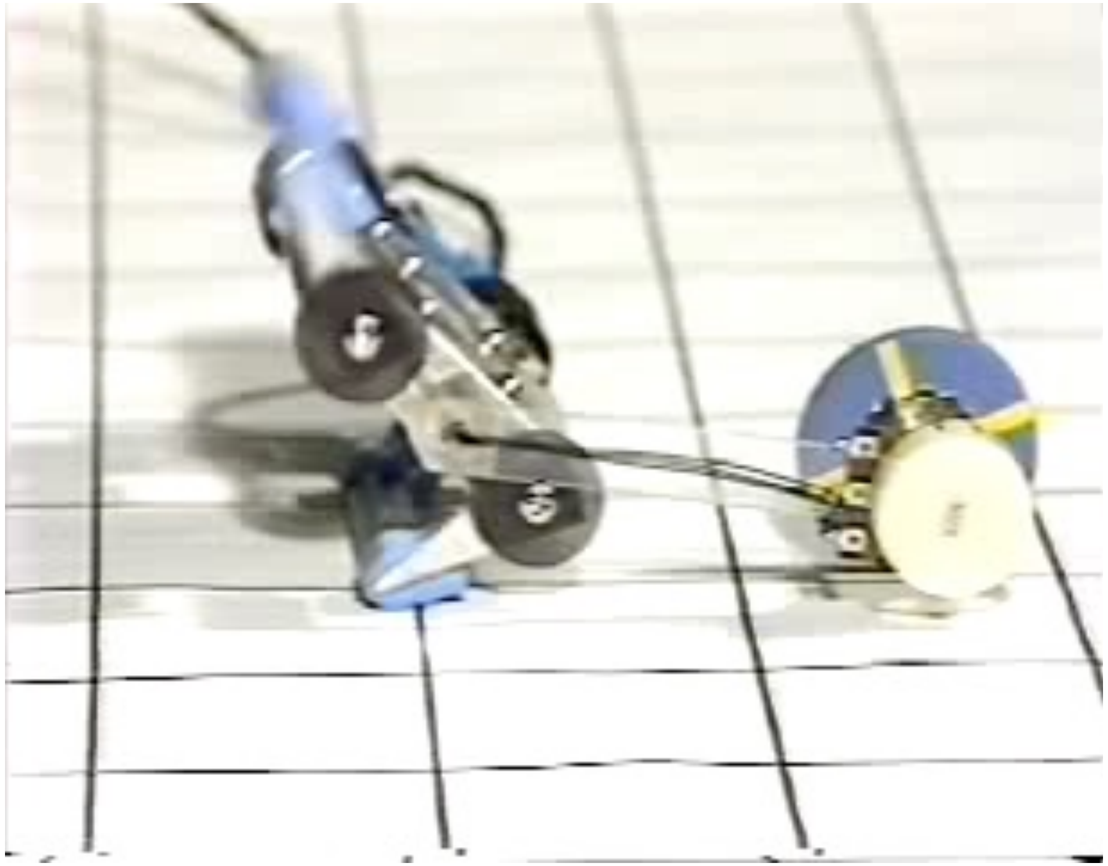
- World Models and Mental Simulation
- Modularity and Compositionality
- Meta-learning

Autonomy and Sociality

Learning to Walk

(Doya & Nakano, 1985)

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



Reinforcement Learning

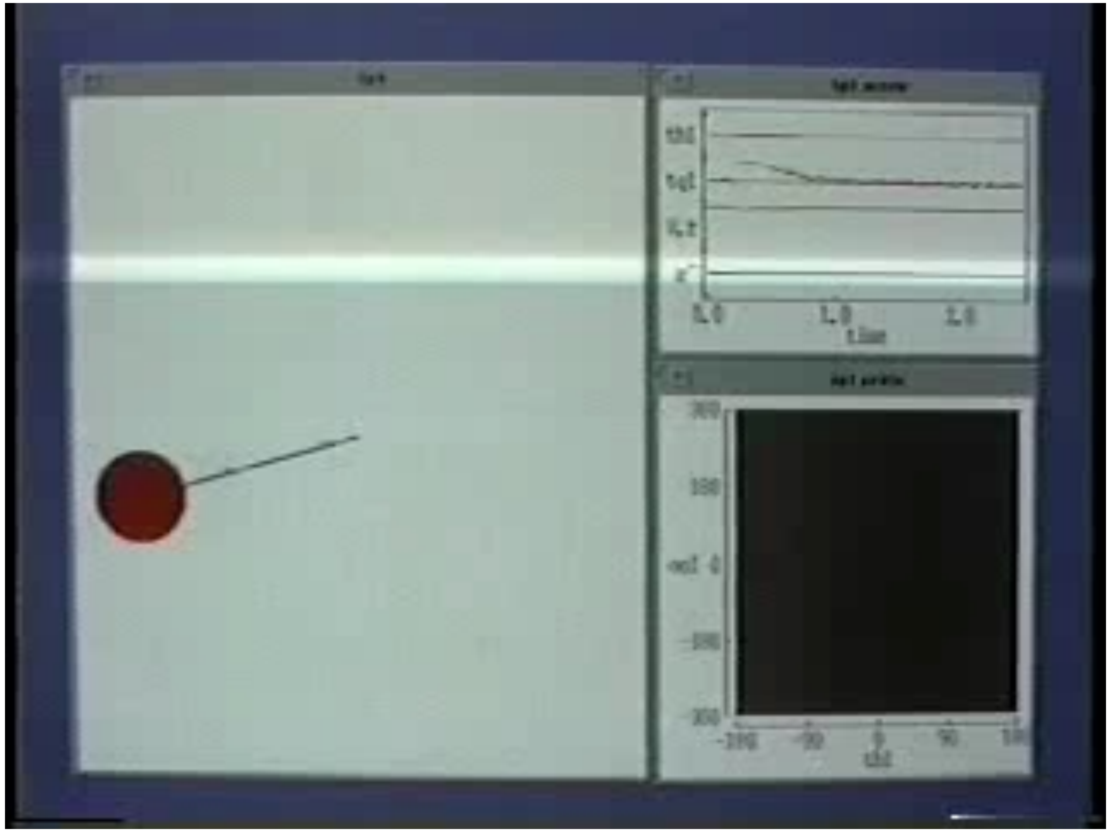
- Predict reward: *value function*
  - $V(s) = E[\ r(t) + \gamma r(t+1) + \gamma^2 r(t+2) \ldots |\ s(t)=s]$
  - $Q(s,a) = E[\ r(t) + \gamma r(t+1) + \gamma^2 r(t+2) \ldots |\ s(t)=s,\ a(t)=a]$
- Select action
  - *greedy*: $a = \text{argmax } Q(s,a)$
  - *Boltzmann*: $P(a|s) \propto \exp[\ \beta\ Q(s,a)]$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

How to implement these steps?

How to tune these parameters?

Pendulum Swing-Up

- state: angle θ, angular velocity ω
- reward function: potential energy: $\cos\theta$



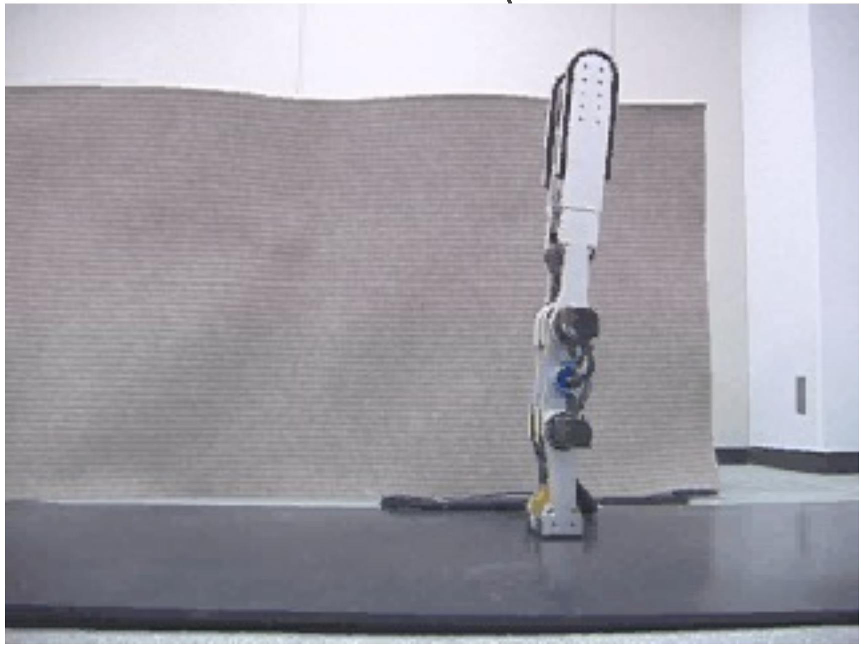
ω

θ

- Value function

Learning to Stand Up

(Morimoto & Doya, 2001)

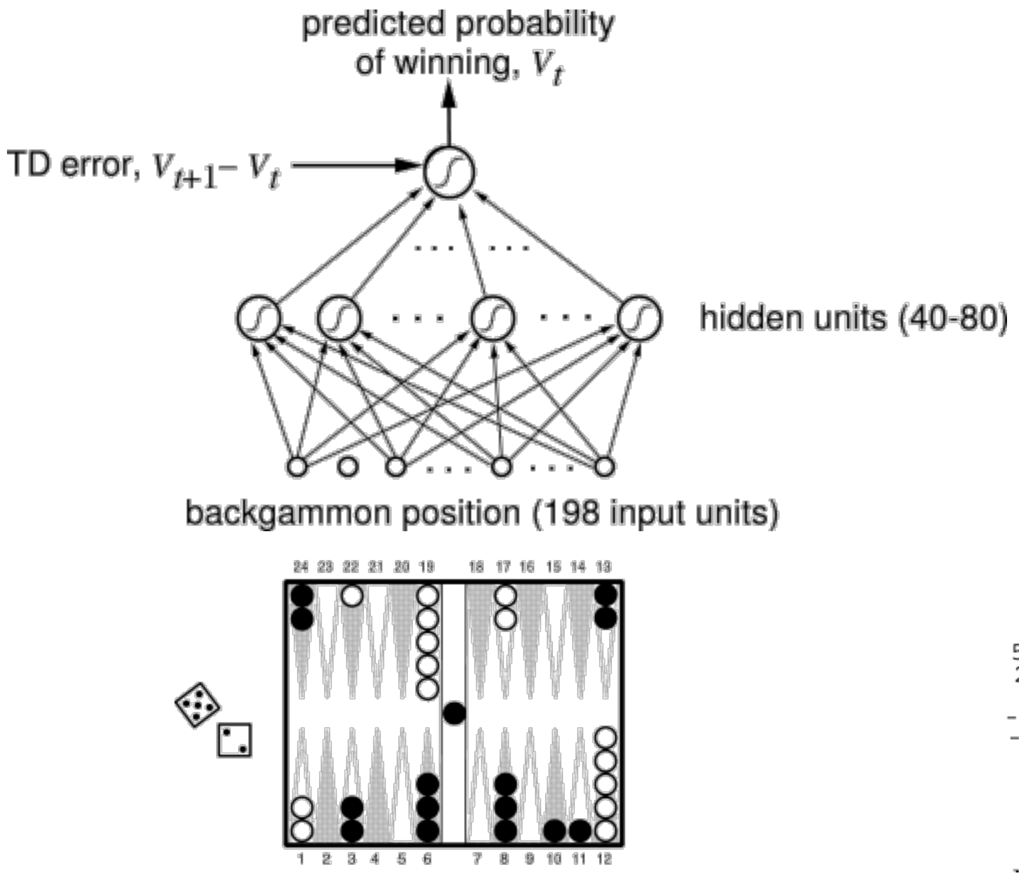


- Learning from reward and punishment
  - reward: height of the head
  - punishment: bump on the floor

TD Learning and Backprop

- TD Gammon

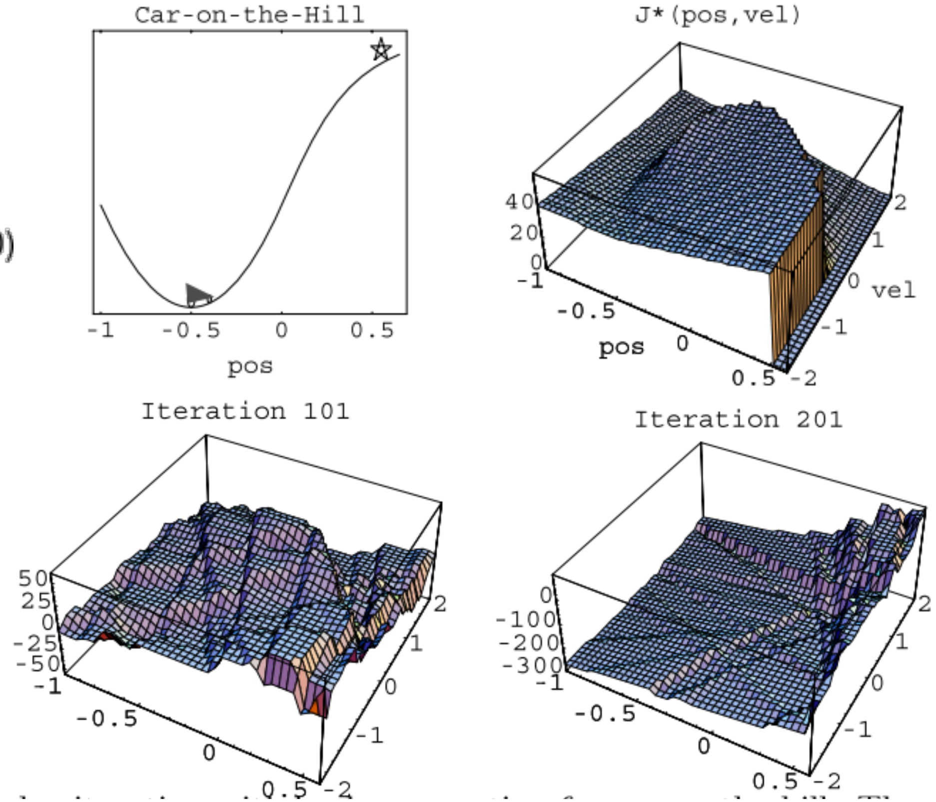
(Tesauro 1992, 1994)



- TD Learning can diverge

(Boyan & Moore, 1995)

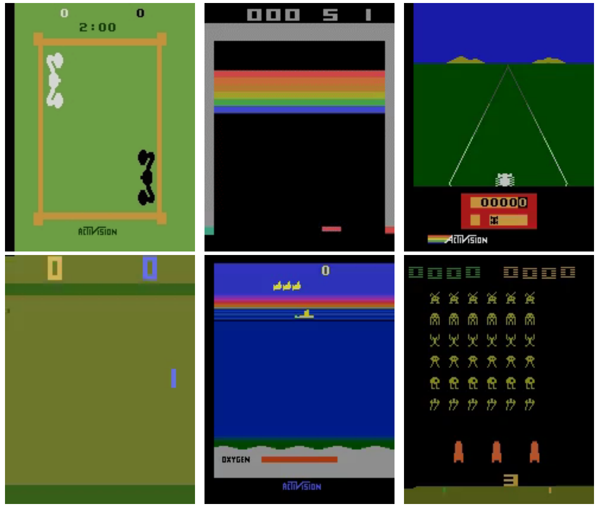
- $\delta(t) = r(t) + \gamma V(s(t+1)) - V(s(t))$

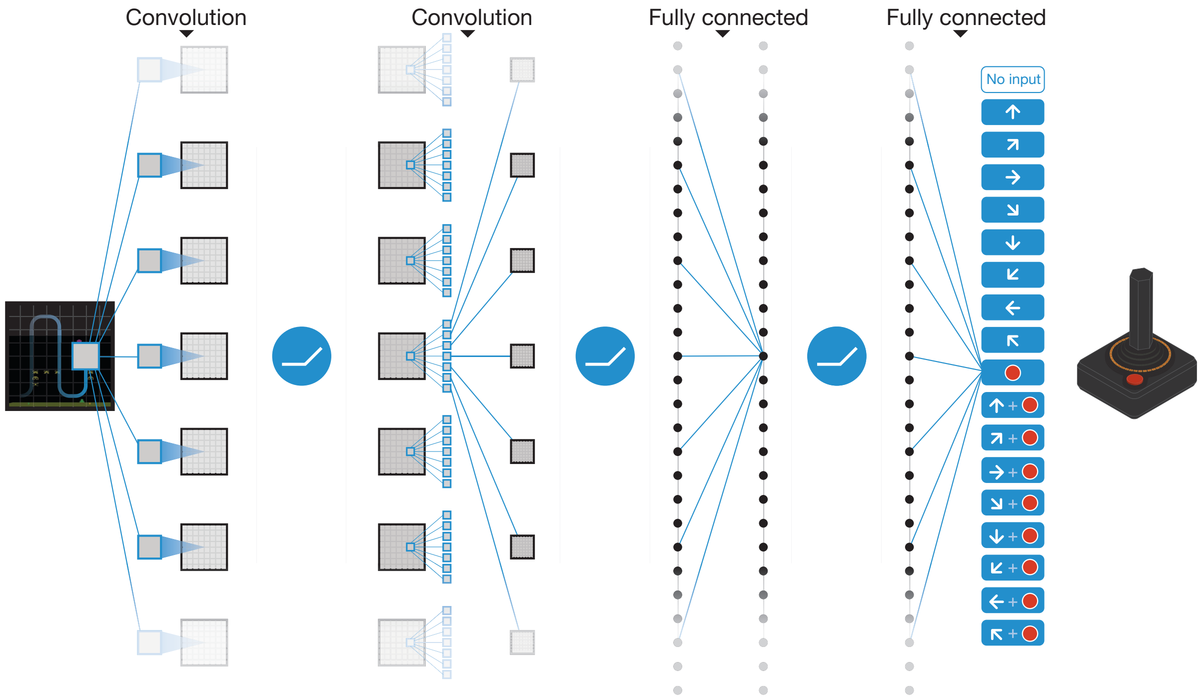


Deep Q-Network

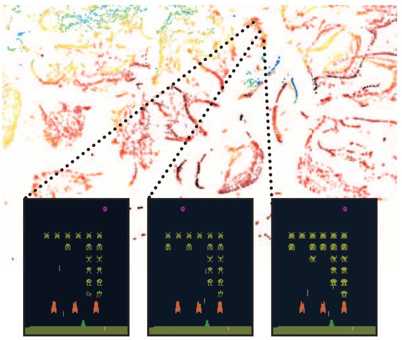
(Mnih et al. 2015)

- Game screen as input





  - *Experience replay*
  - Fixing the *target network*
- DNN captures important features
  - human level in 29/49 Atari games



Basal Ganglia

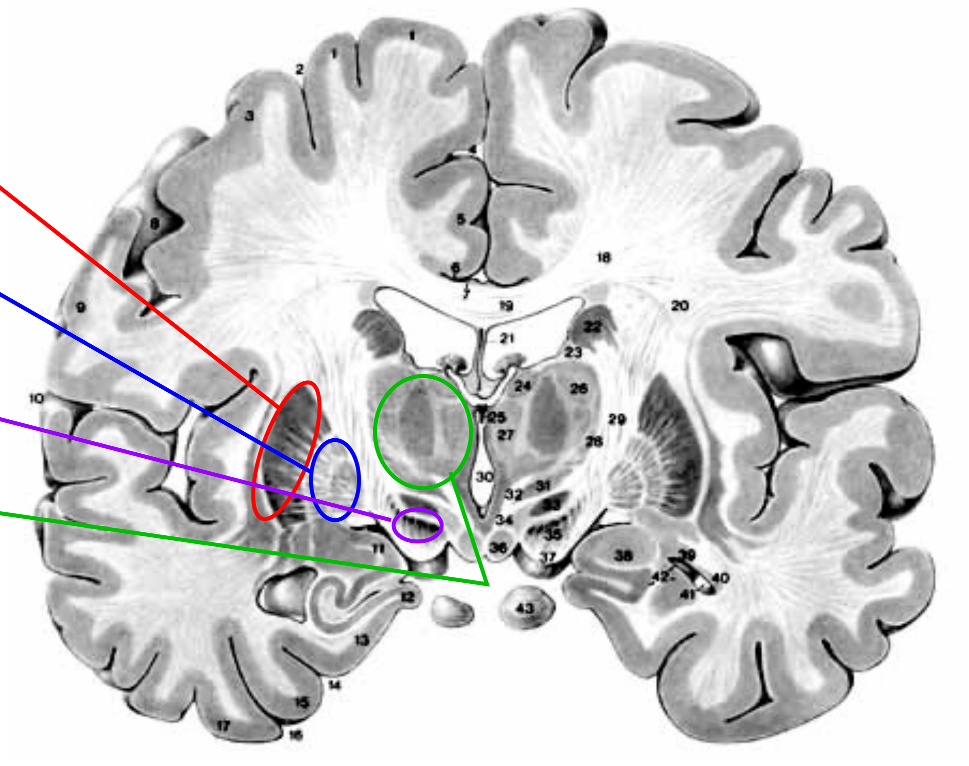
- Locus of Parkinson’s and Huntington’s diseases

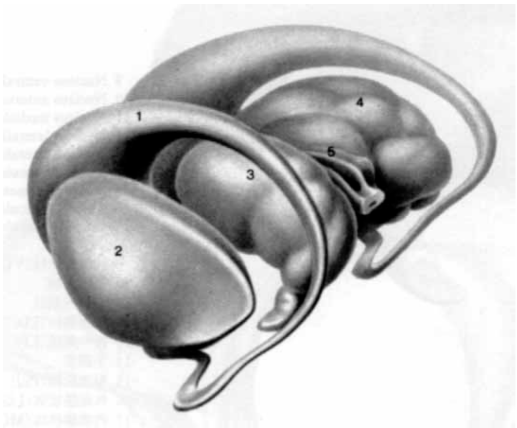
Striatum

Globus Pallidus

Substantia Nigra

Thalamus

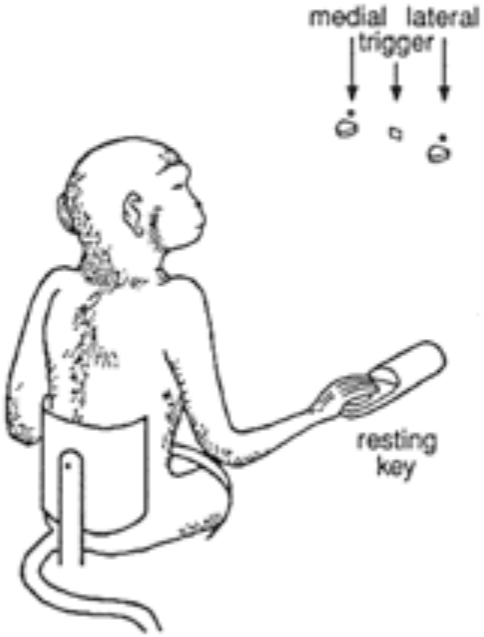




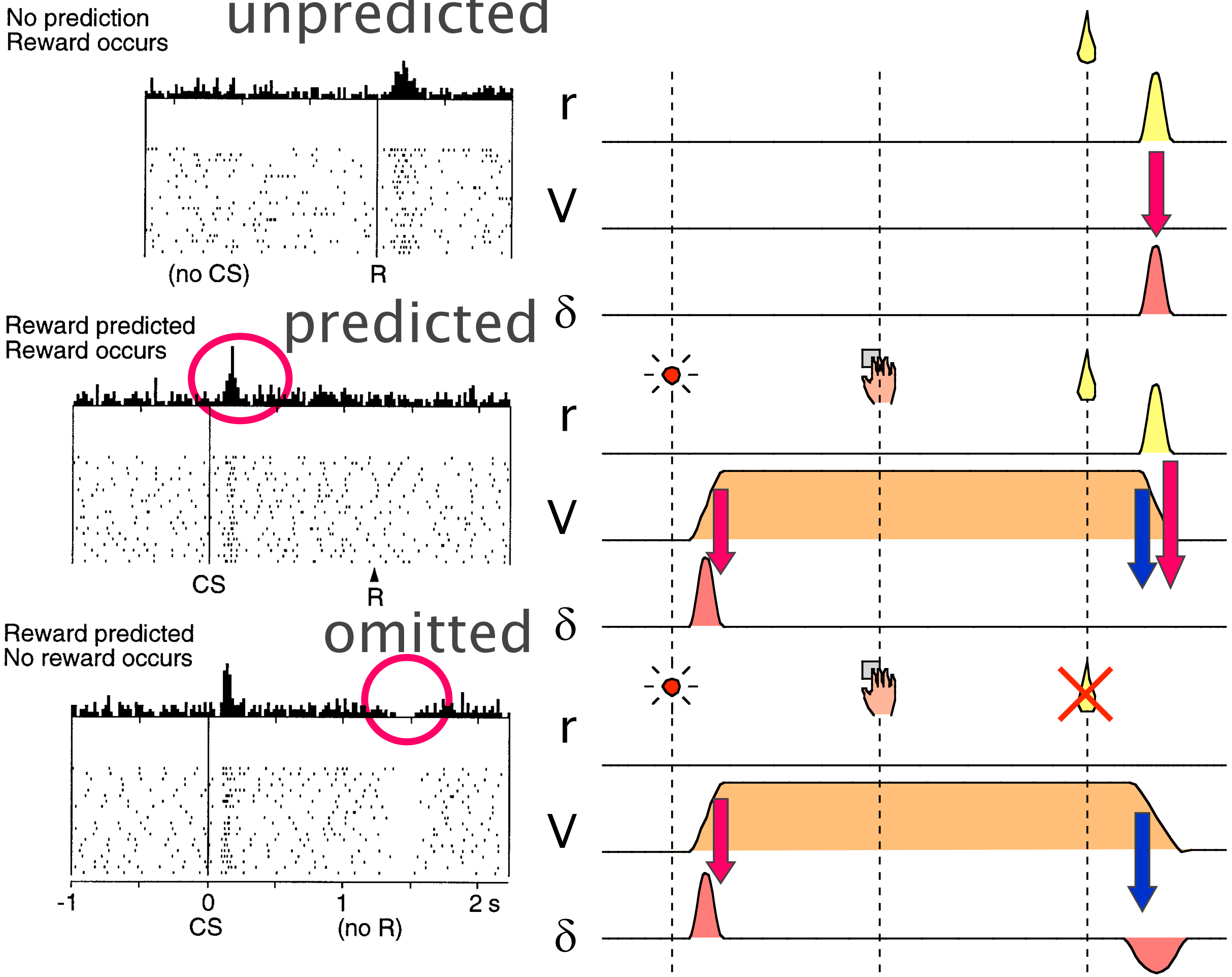
- What is their normal function??

Dopamine Neurons Code TD Error

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

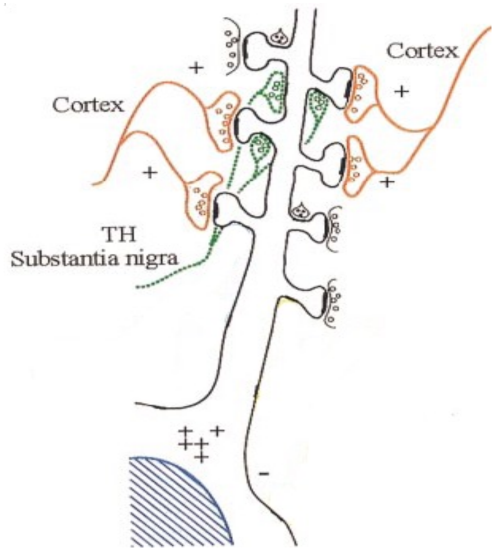


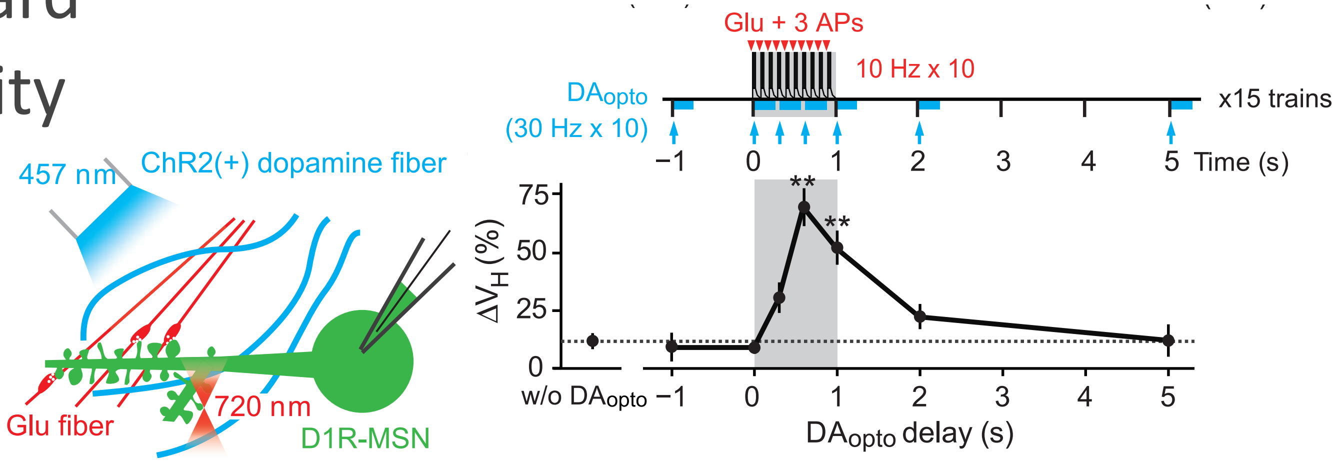
(Schultz et al. 1997)



Dopamine-dependent Plasticity

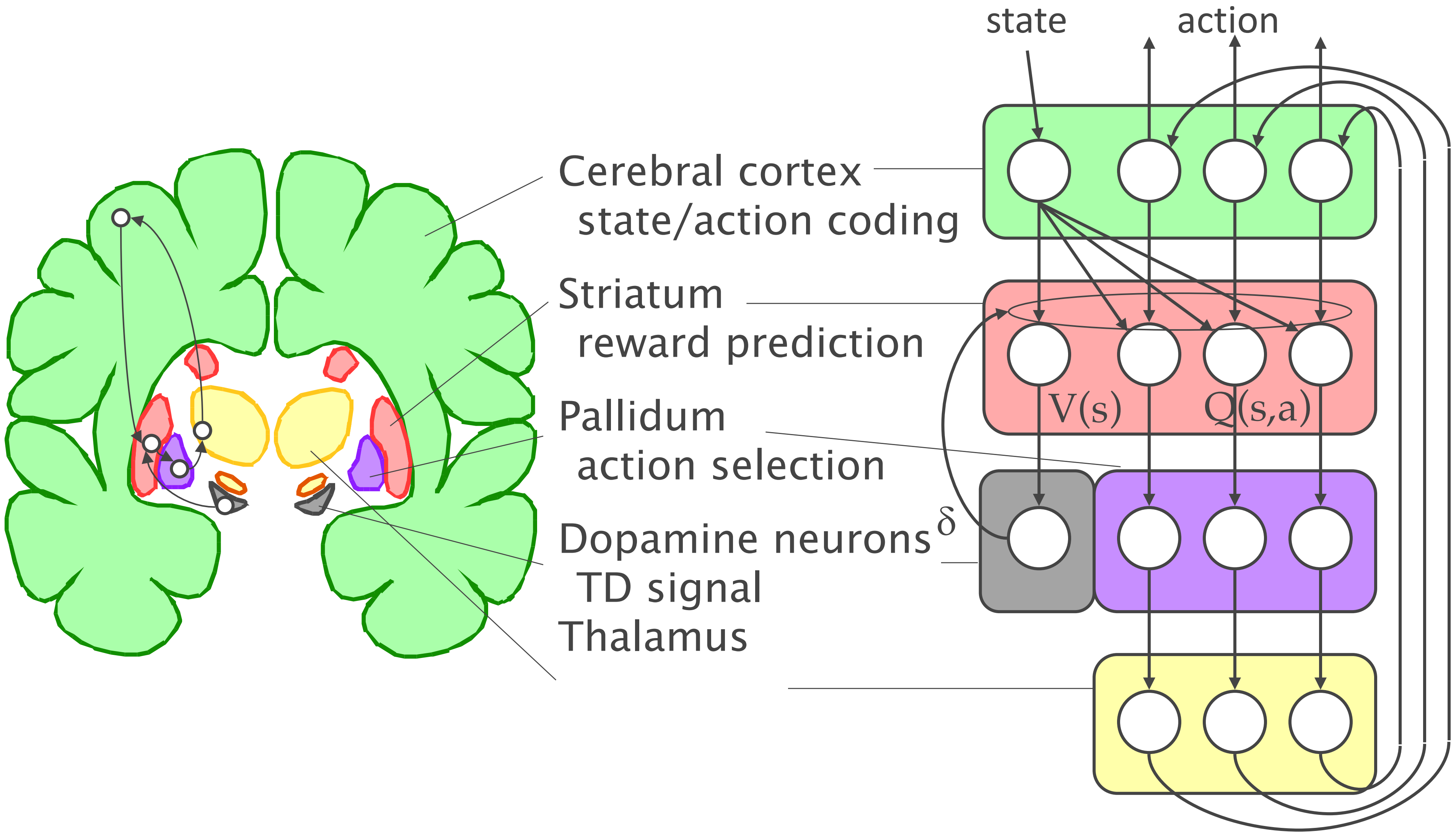
- Medium spiny neurons in striatum
  - glutamate from cortex
  - dopamine from midbrain
- Three-factor learning rule (Wickens et al.)
  - cortical input + spike → LTD
  - cortical input + spike + dopamine → LTP
  - input x output x reward
- Time window of plasticity
  (Yagishita et al., 2014)





Basal Ganglia for Reinforcement Learning?

(Doya 2000, 2007)

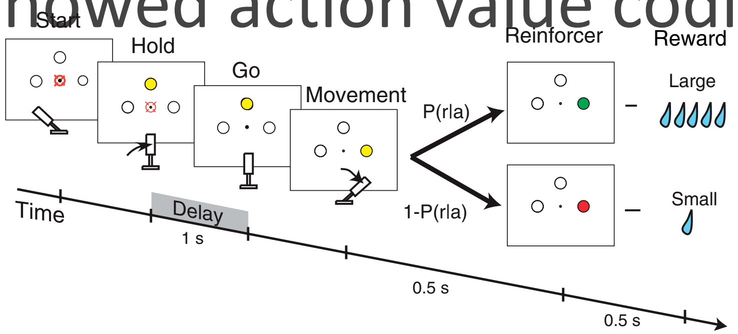


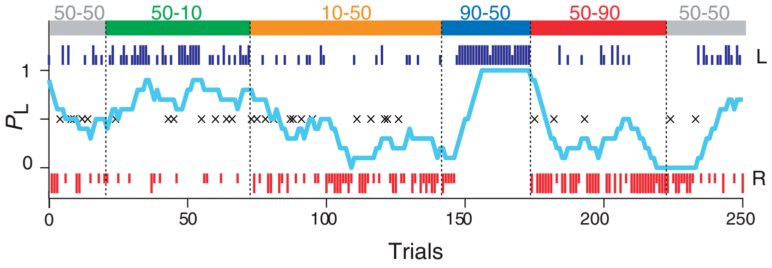
SCIENCE VOL 310 25 NOVEMBER 2005

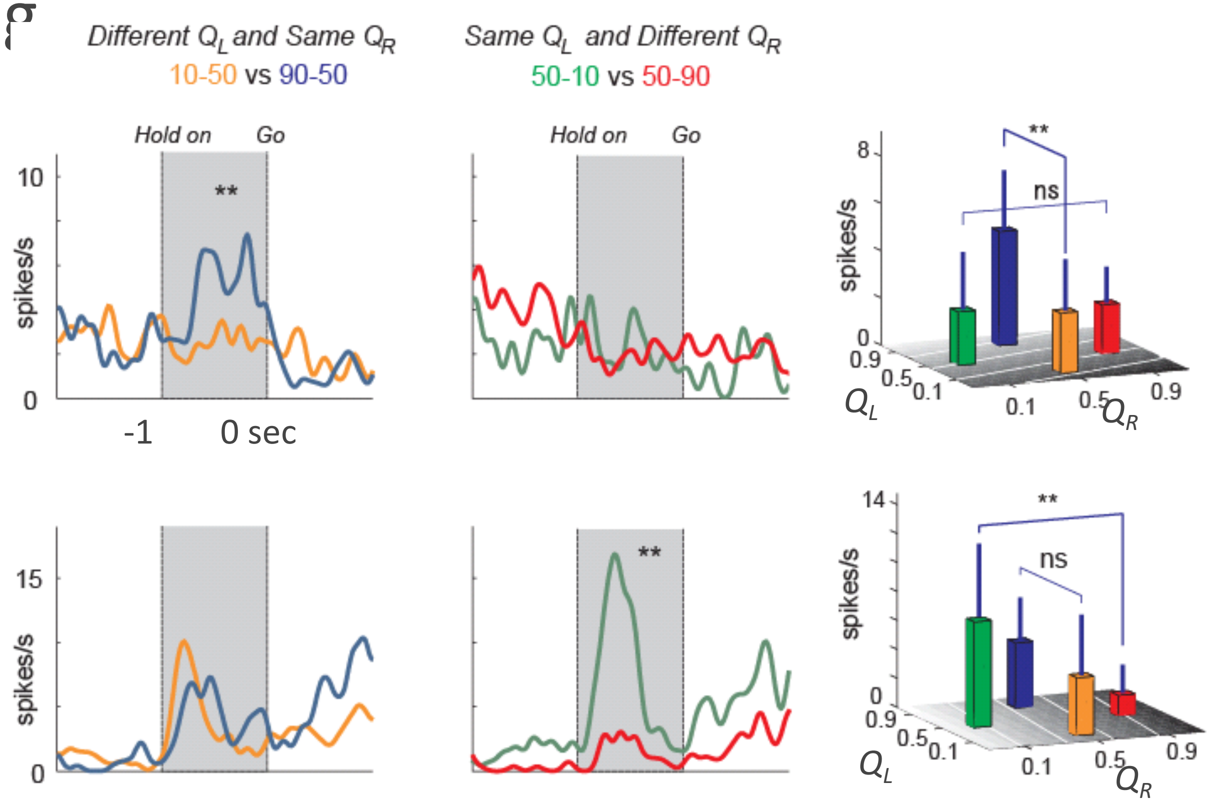
Representation of Action-Specific Reward Values in the Striatum

Kazuyuki Samejima,[1*†] Yasumasa Ueda,[2] Kenji Doya,[1,3] Minoru Kimura[2*]

- About half of task-responsive neurons in the anterior striatum showed action value coding



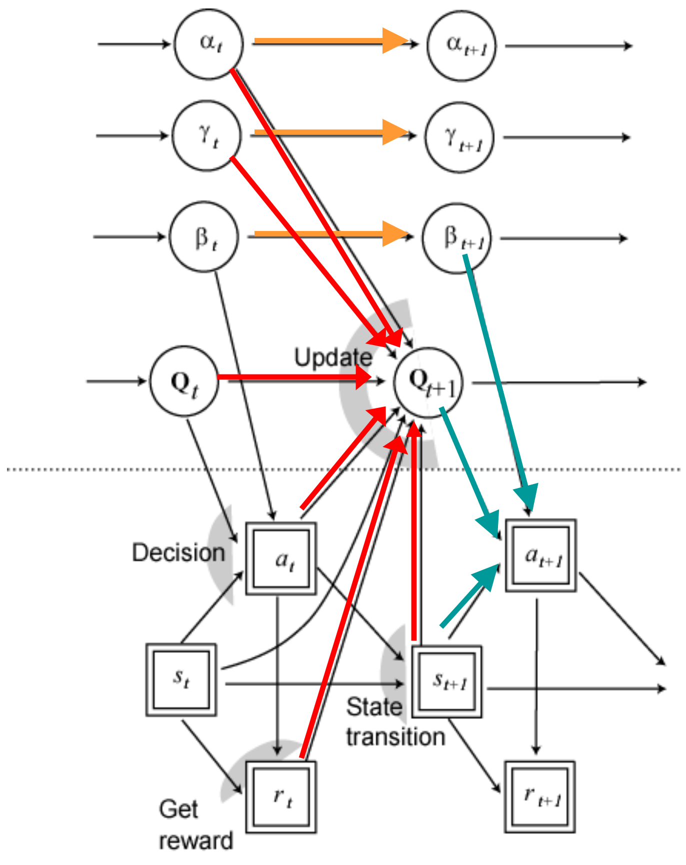




Bayesian Inference of Action Values

(Samejima et al. 2004)

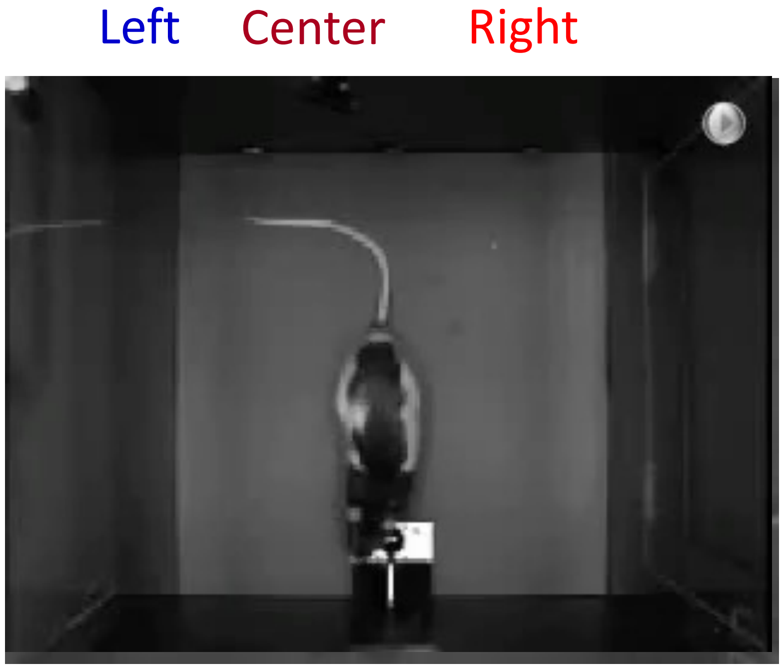
- Hidden variables
  - $x=(Q,\alpha,\beta,\gamma)$
  - $p(x'|x)$: learning rule
- Observable variables
  - $y=(s,a,r)$
  - $p(y|x)$: action policy
- Predictive prior
  - $P(x_{t+1}|y_{1:t}) = \int P(x_{t+1}|x_t)P(x_t|y_{1:t})dx_t$
- Posterior given observation y_{t+1}
  - $P(x_{t+1}|y_{1:t+1}) \propto P(y_{t+1}|x_{t+1})P(x_{t+1}|y_{1:t})$

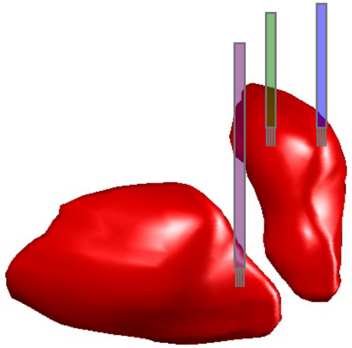


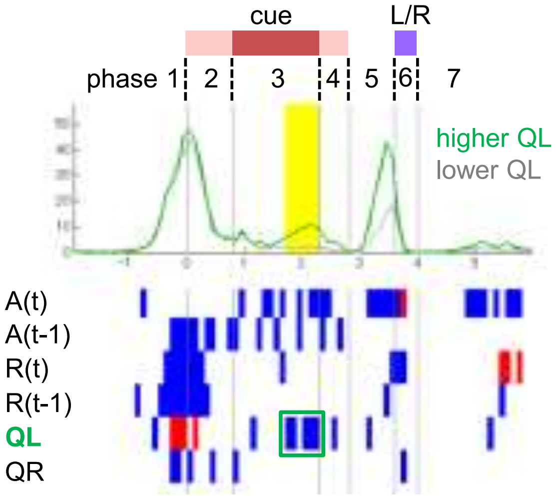
Distinct Neural Representation in the Dorsolateral, Dorsomedial, and Ventral Parts of the Striatum during Fixed- and Free-Choice Tasks

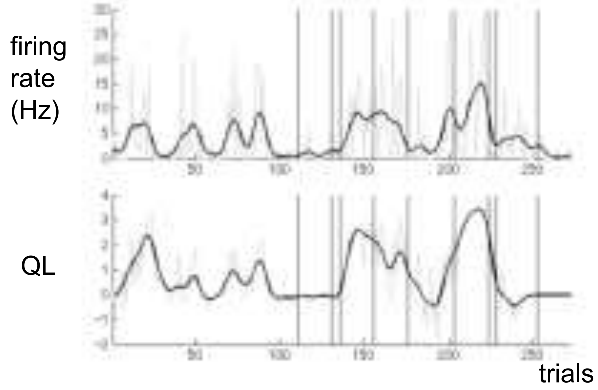
Makoto Ito and Kenji Doya

The Journal of Neuroscience, 2015

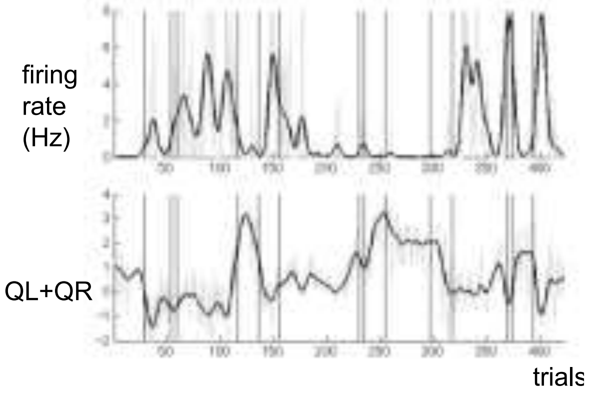




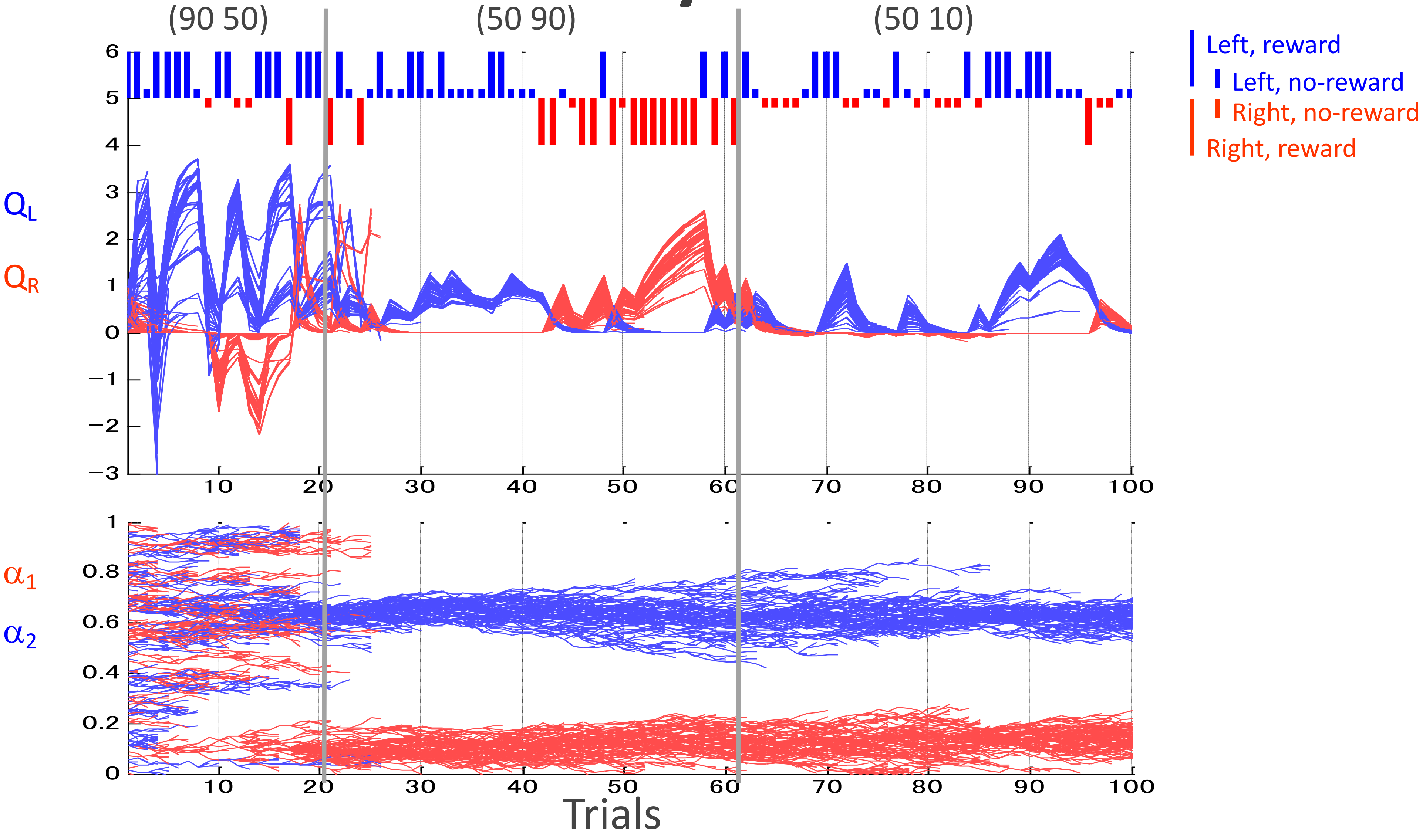




- Dorsolateral
  - movements
- Dorsomedial
  - action value
- Ventral
  - state value

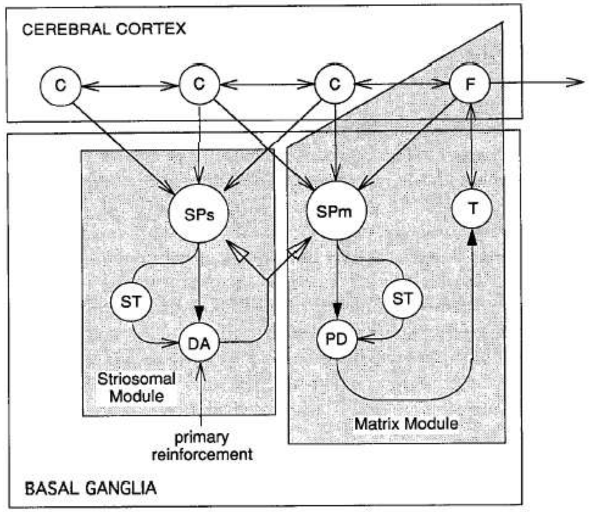


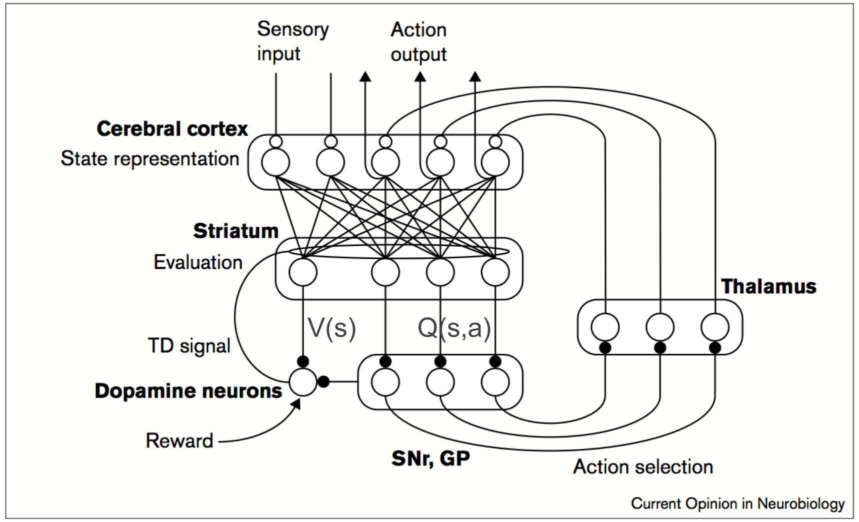
Estimation by Particle Filter



Striosome Neurons as Critic?

- Actor-critic (Houk et al., 1995) or state/action value (Doya, 2000)





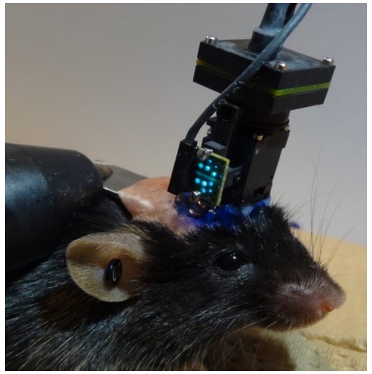
  - Do striosome neurons code state value?
  - Do matrix neurons code action or action value?
- Need cell-type specific recording
  - optolodes or calcium imaging

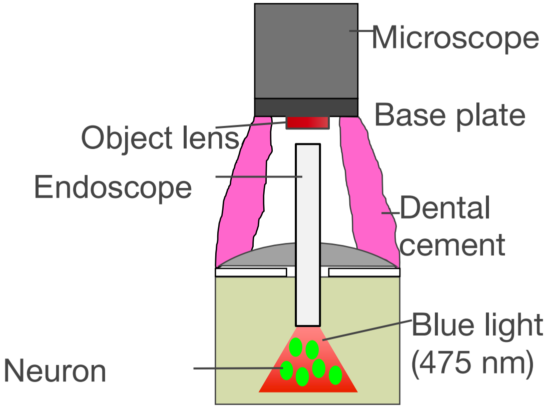
eNeuro (2018)

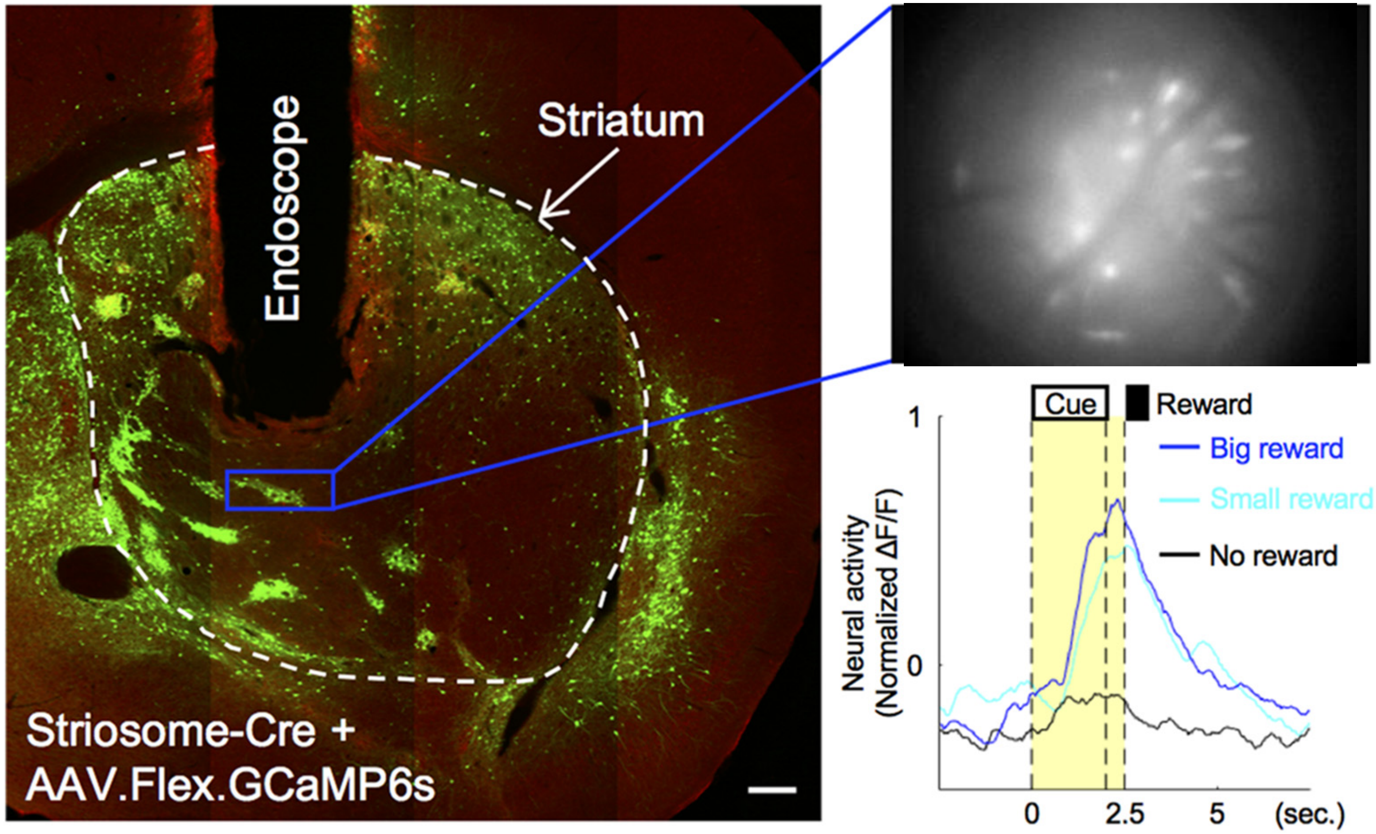
Reward-Predictive Neural Activities in Striatal Striosome Compartments

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

Imaging striosome neuron activity by endoscope





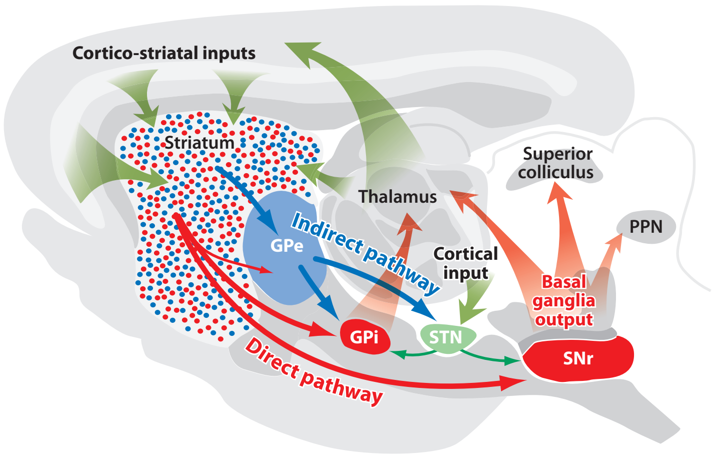


Questions in Neural Reinforcement Learning

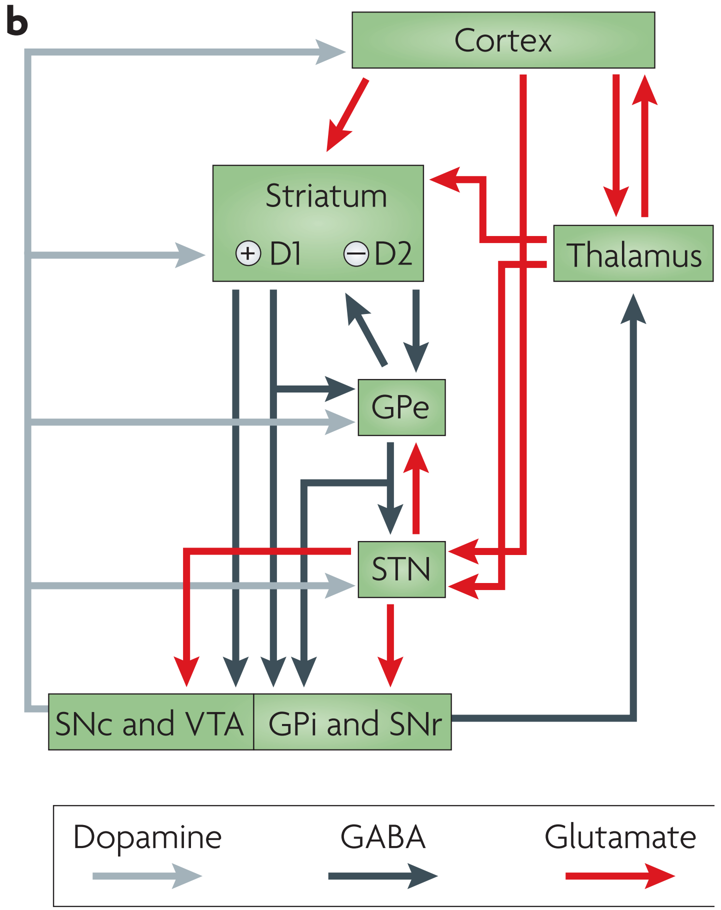
How is TD-like response computed by dopamine neurons?

Why should there be so many pathways?

- direct, indirect, hyperdirect
- striosome, matrix
- dorsal/ventral striatum, amygdala
- SNc and VTA dopamine neurons



(Gerfen 1992)



(Redgrave et al. 2010)

Soft Actor-Critic

(Haarnoja et al. 2018)

- Stable, sample-efficient learning
- Learn state value, action value, and policy in parallel
  - objective

$$J(\pi) = \sum_{t=0}^{T-1} \mathbb{E}_{(\mathbf{s}_t,\mathbf{a}_t)\sim\rho_\pi}\left[r(\mathbf{s}_t,\mathbf{a}_t) + \alpha\mathcal{H}(\pi(\,\cdot\,|\mathbf{s}_t))\right]$$

  - state value V

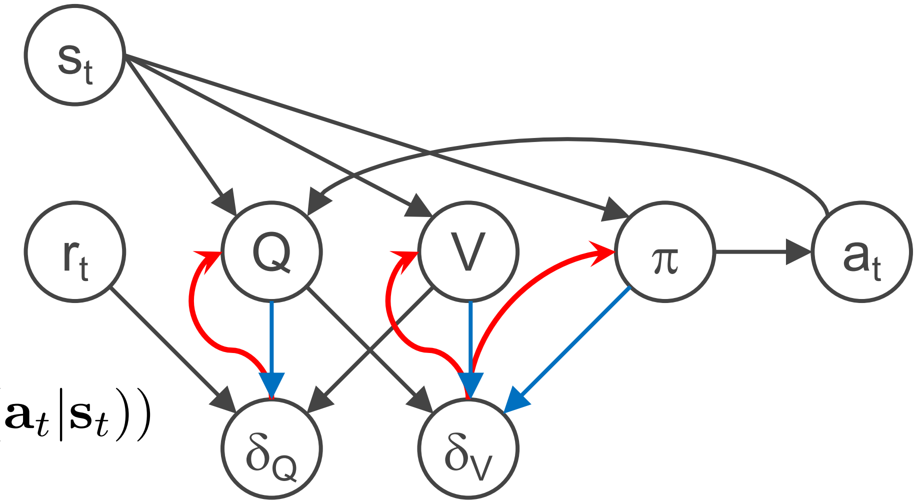
$$\hat{\nabla}_\psi J_V(\psi) = \nabla_\psi V_\psi(\mathbf{s}_t)\left(V_\psi(\mathbf{s}_t) - Q_\theta(\mathbf{s}_t,\mathbf{a}_t) + \log\pi_\phi(\mathbf{a}_t|\mathbf{s}_t)\right)$$

  - action value Q

$$\hat{\nabla}_\theta J_Q(\theta) = \nabla_\theta Q_\theta(\mathbf{a}_t,\mathbf{s}_t)\left(Q_\theta(\mathbf{s}_t,\mathbf{a}_t) - r(\mathbf{s}_t,\mathbf{a}_t) - \gamma V_{\bar{\psi}}(\mathbf{s}_{t+1})\right)$$

  - policy π

$$\hat{\nabla}_\phi J_\pi(\phi) = \nabla_\phi \log\pi_\phi(\mathbf{a}_t|\mathbf{s}_t)\left(\log\pi_\phi(\mathbf{a}_t|\mathbf{s}_t) - Q_\theta(\mathbf{s}_t,\mathbf{a}_t) + V_\psi(\mathbf{s}_t)\right)$$



Model-free and Model-based RL

Model-free RL

- Memorize action values
  - Q(state, action)
- Reactive action
  - $P(a|s) \sim \exp[\beta Q(s,a)]$
- On-line learning by TD error
  - $\delta = \text{reward} + \gamma Q(s',a') - Q(s,a)$

Simple, but slow learning

Model-based RL

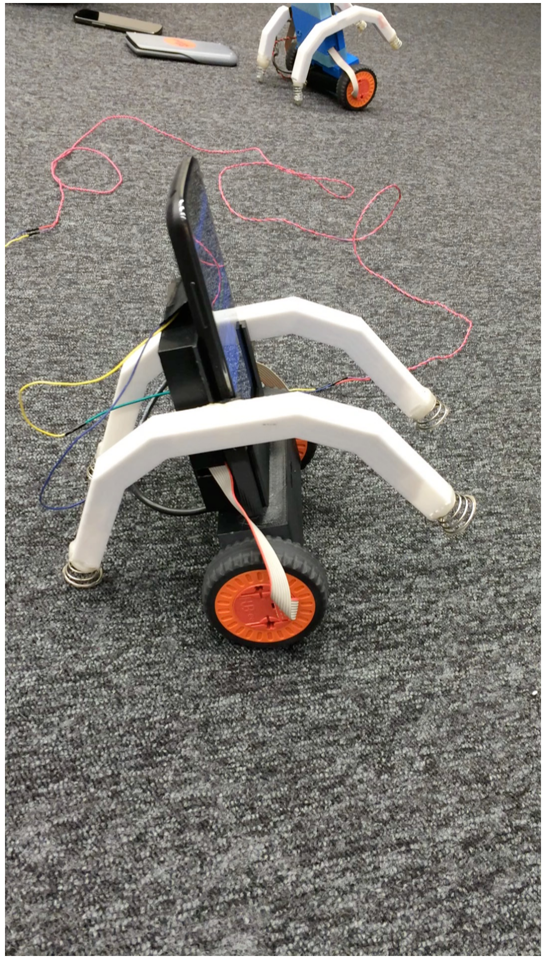
- Learn internal models
  - P(next state| state, action)
  - R(state, action)
- Estimate current state
  - $P(s_t|o_t,a_{t-1}) \propto P(o_t|s_t)\Sigma_{s_{t-1}} P(s_t|s_{t-1},a_{t-1})P(s_{t-1})$
- Predict values
  - $Q(s,a) = \Sigma_{s'} P(s'|s,a)[R(s,a)+\gamma V(s')]$
  - $V(s)=\max_a \Sigma_{s'} P(s'|s,a)[R(s,a)+\gamma V(s')]$

Flexible, but heavy load

Bounce Up and Balance by PILCO

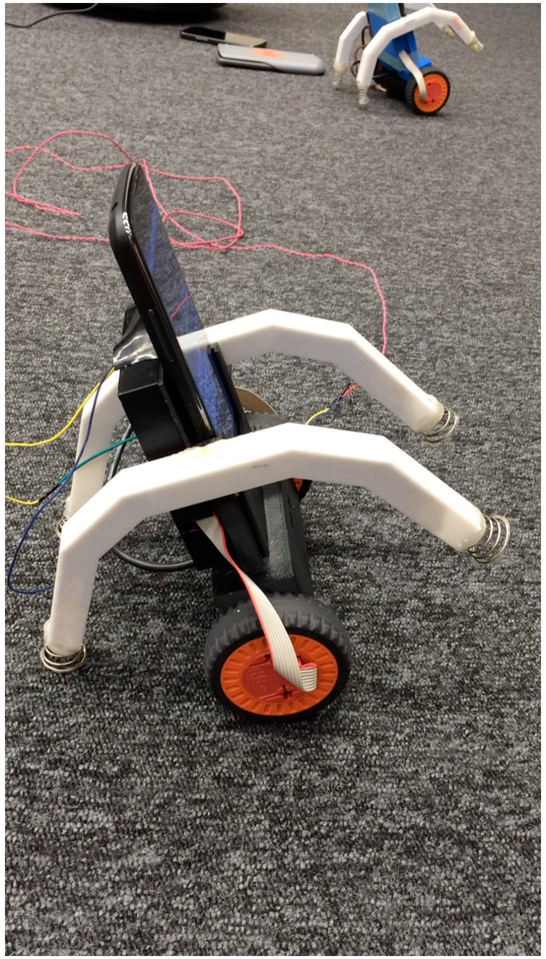
(Paavo Parmas)

1st try



2nd try

8th try



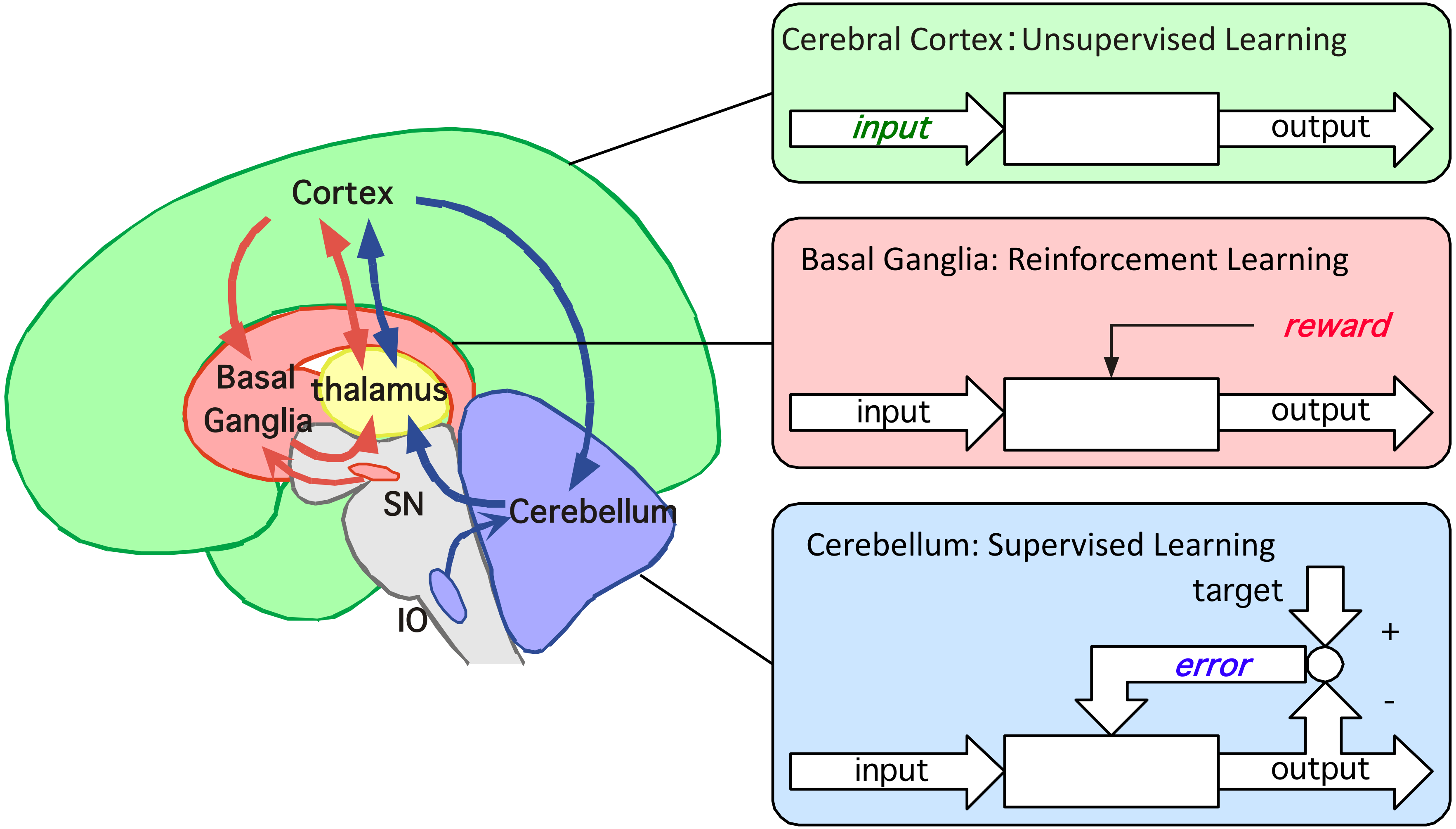
Mental Simulation

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,…

Specialization by Learning Algorithms

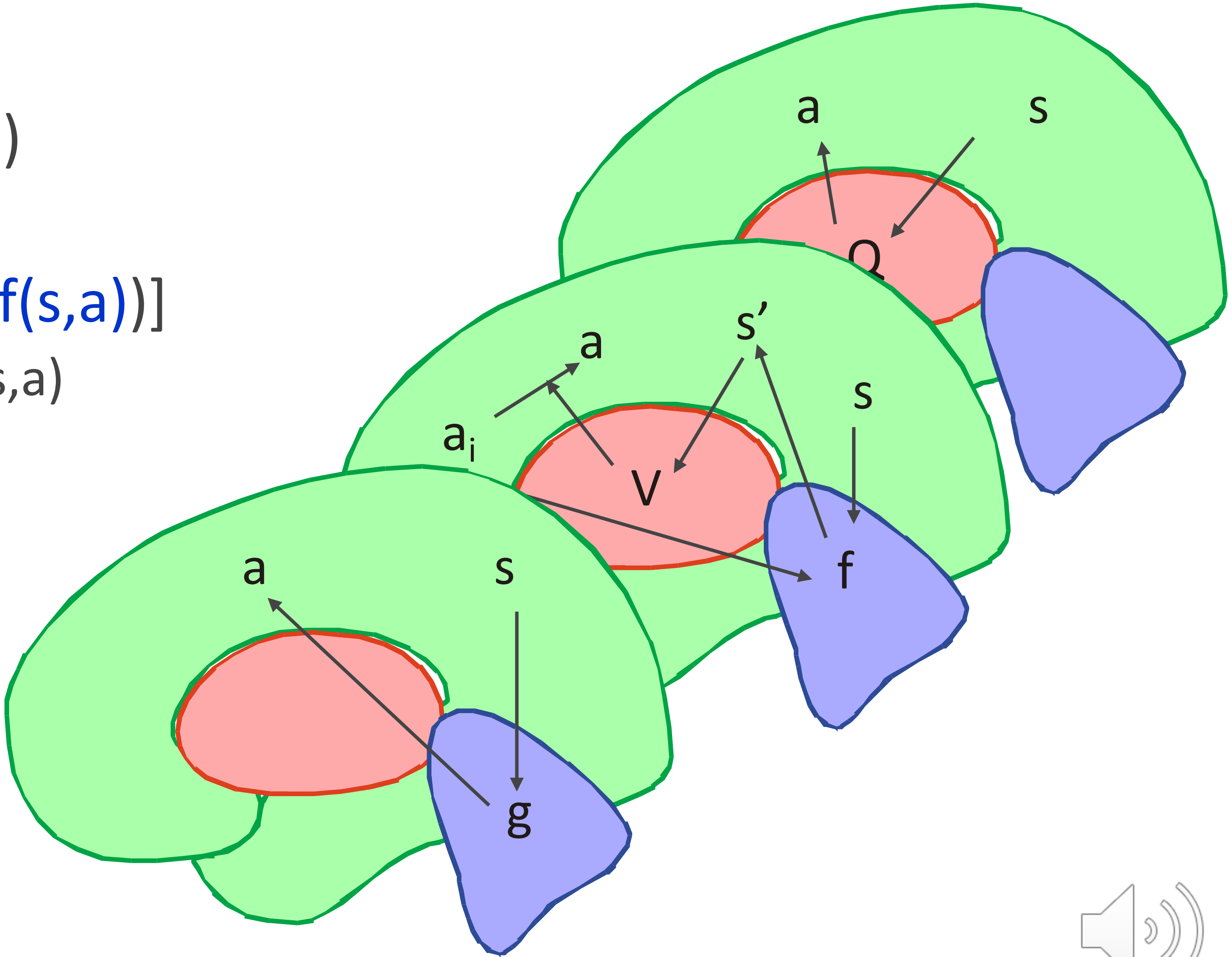
(Doya, 1999)



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
  - $a = g(s)$



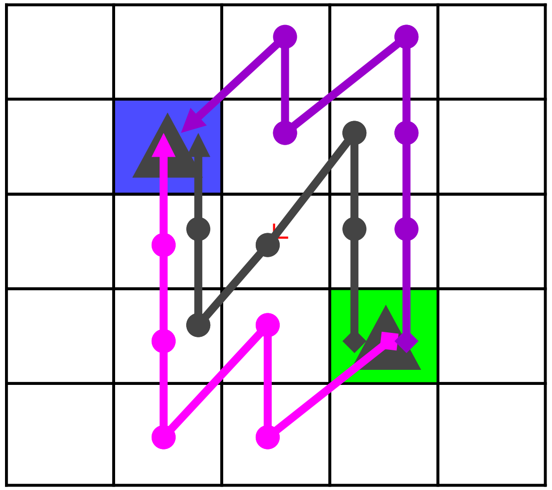
SCIENTIFIC REPORTS

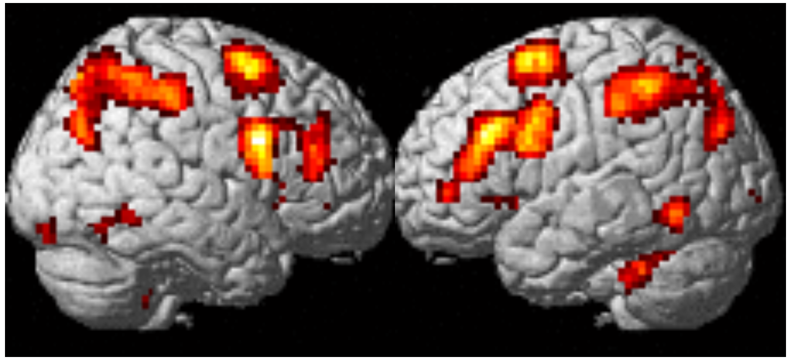
OPEN

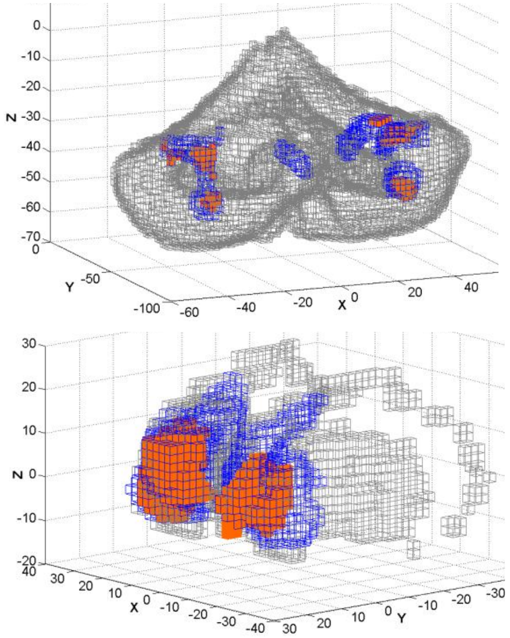
Model-based action planning involves cortico-cerebellar and basal ganglia networks

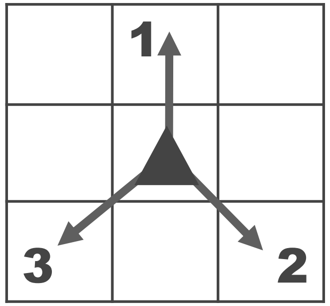
Received: 16 February 2016
Accepted: 19 July 2016

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]





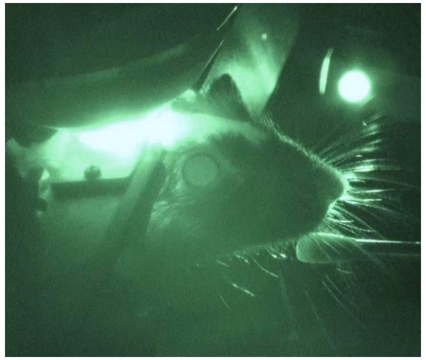


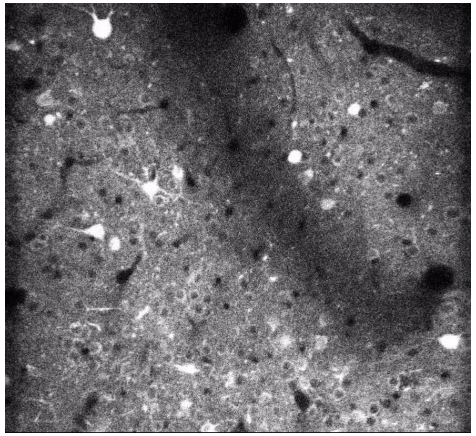


nature
neuroscience

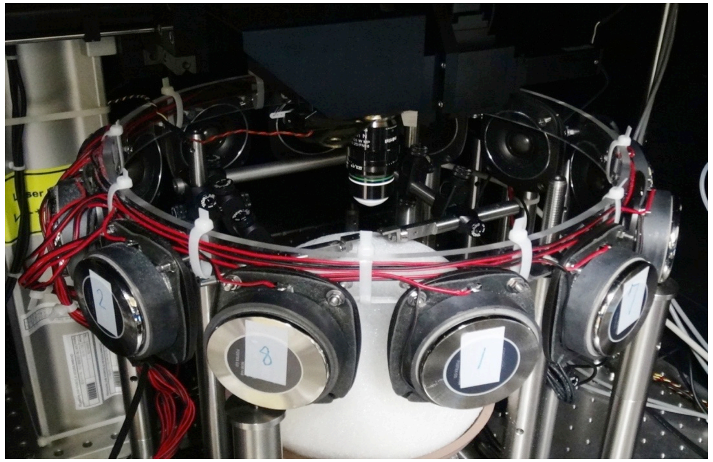
Neural substrate of dynamic Bayesian inference in the cerebral cortex

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



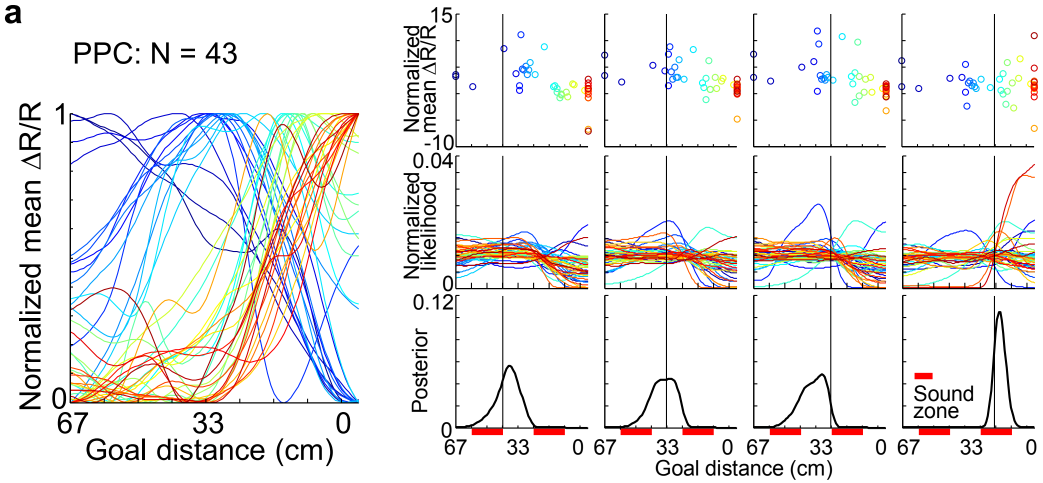


- PPC two-photon imaging



- Auditory virtual environment
  - intermittent sensory input

- Probabilistic population decoding

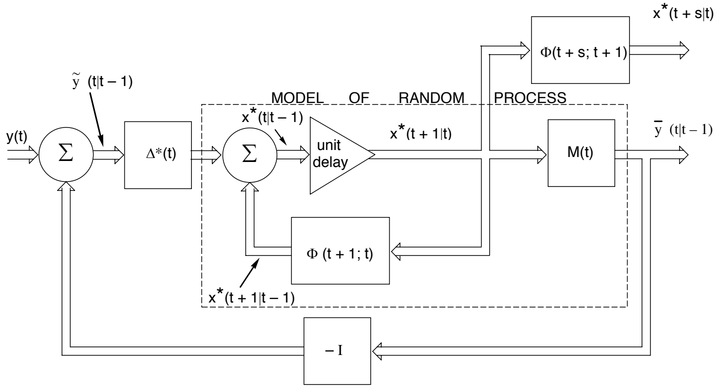


  - predicting goal distance from action

Duality of Inference and Control

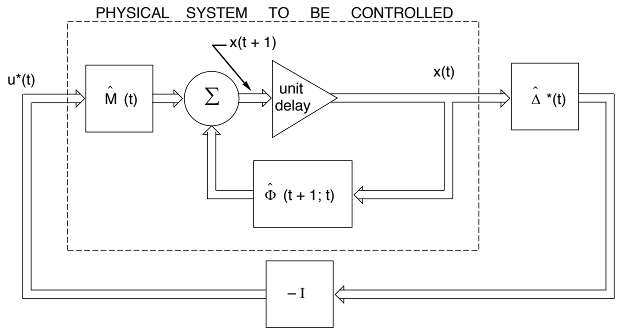
- Optimal filtering (Kalman 1960)

$$\Sigma_{k+1} = S + A\Sigma_k A^{\mathsf{T}} - A\Sigma_k H^{\mathsf{T}} \left(P + H\Sigma_k H^{\mathsf{T}}\right)^{-1} H\Sigma_k A^{\mathsf{T}}$$

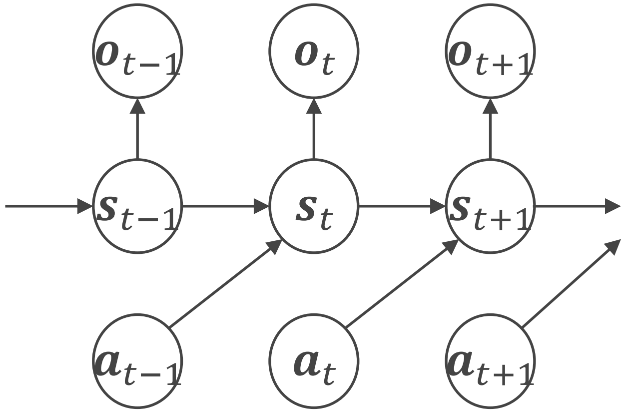


- Optimal control (Bellman et al. 1958)

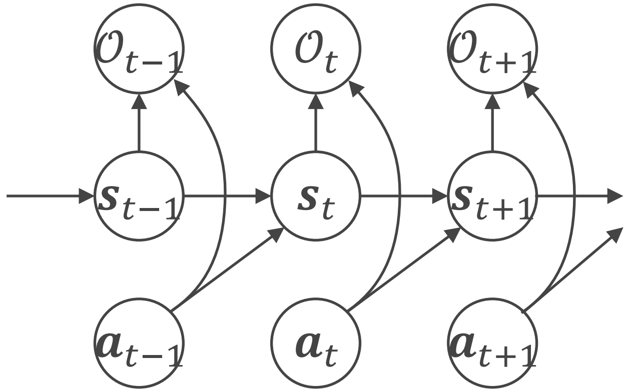
$$V_k = Q + A^{\mathsf{T}} V_{k+1} A - A^{\mathsf{T}} V_{k+1} B \left(R + B^{\mathsf{T}} V_{k+1} B\right)^{-1} B^{\mathsf{T}} V_{k+1} A$$



- Bayesian inference: log posterior



- Reinforcement learning: state value

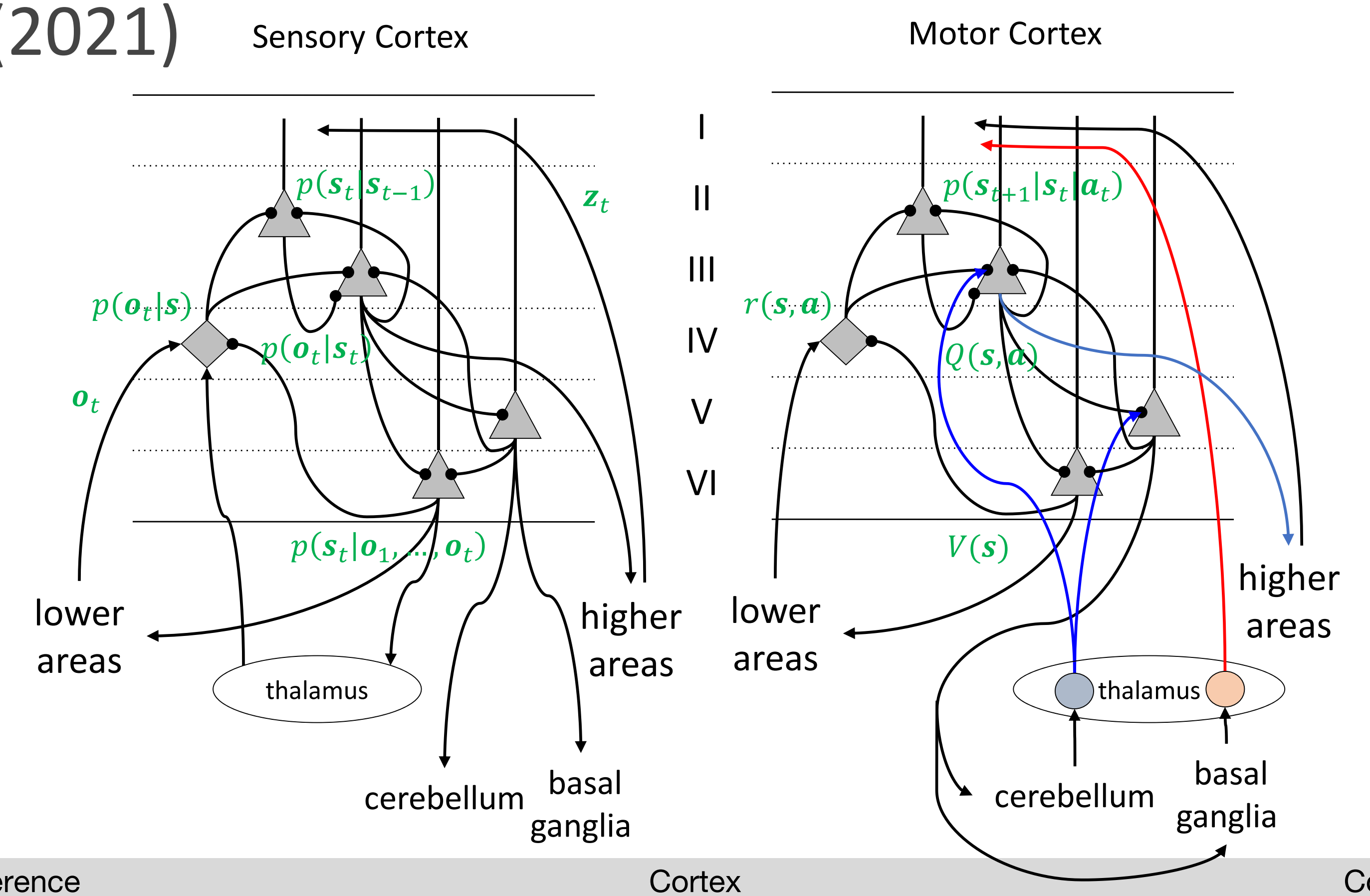


(Todorov 2007, 08; Toussaint 2009; Levine 2018)

Current Opinion in
Behavioral
Sciences

Canonical cortical circuits and the duality of Bayesian inference and optimal control

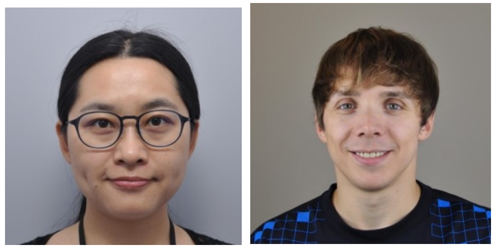
Kenji Doya (2021)

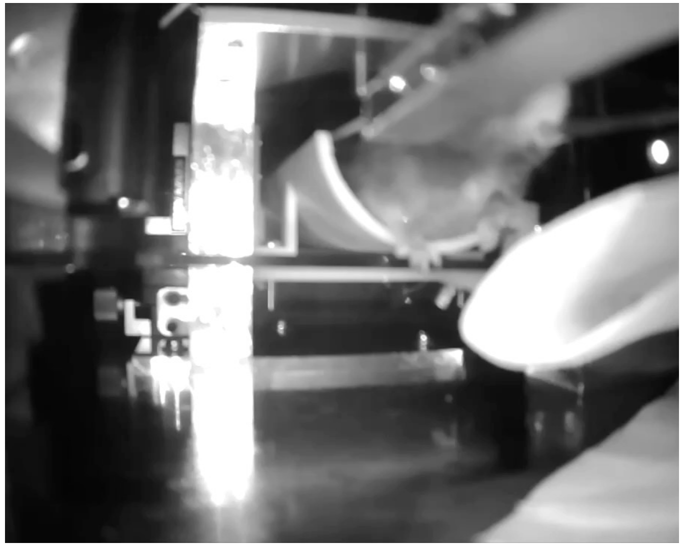


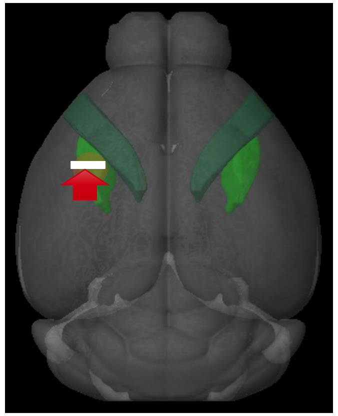
Inference	Cortex	Control
Top-down signal $\boldsymbol{z}_t$	L1 input	Top-down activation signal
Bottom-up signal $p(\boldsymbol{o}_t\|\boldsymbol{s}_t)$	L2/3 output	Action value $Q(\boldsymbol{s},\boldsymbol{a})$
Predictive model $p(\boldsymbol{s}_t\|\boldsymbol{s}_{t-1})$	L2/3 connection	Predictive model $p(\boldsymbol{s}_{t+1}\|\boldsymbol{s}_t,\boldsymbol{a}_t)$
Bottom-up signal $\boldsymbol{o}_t$	L4 input	Optimality signal $\mathcal{O}_t$
Likelihood $p(\boldsymbol{o}_t\|\boldsymbol{s})$	L4 output	Reward function $r(\boldsymbol{s},\boldsymbol{a})$
Posterior $p(\boldsymbol{s}_t\|\boldsymbol{o}_1,\ldots,\boldsymbol{o}_t)$	L5 output	State value $V(\boldsymbol{s})$
Top-down signal $\boldsymbol{s}_t$	L6 output	Action $p(\boldsymbol{a}_t\|\boldsymbol{s}_t)$

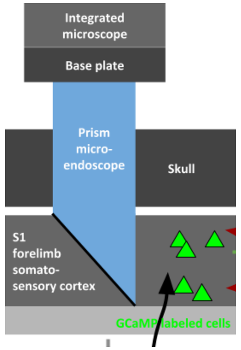
Prism Lens Imaging during Lever Pull Task

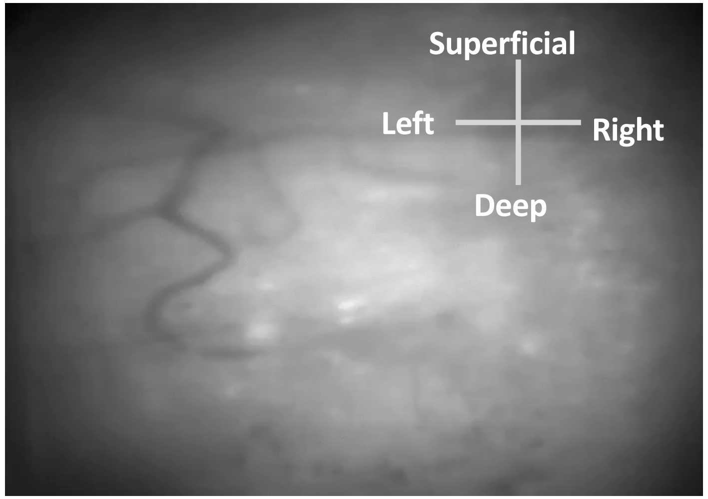
Yuzhe Li, Sergey Zobnin

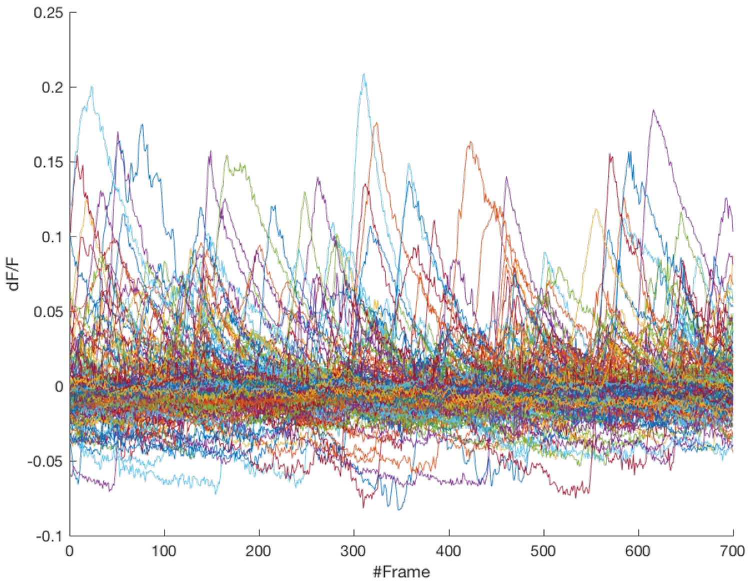


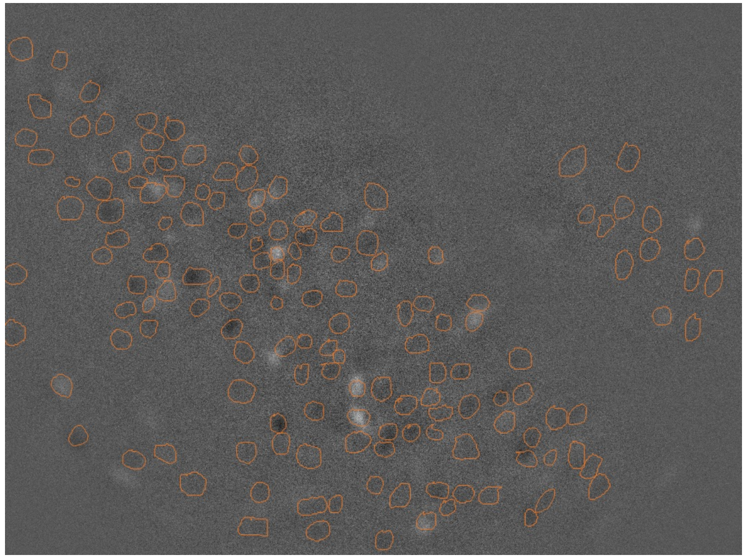






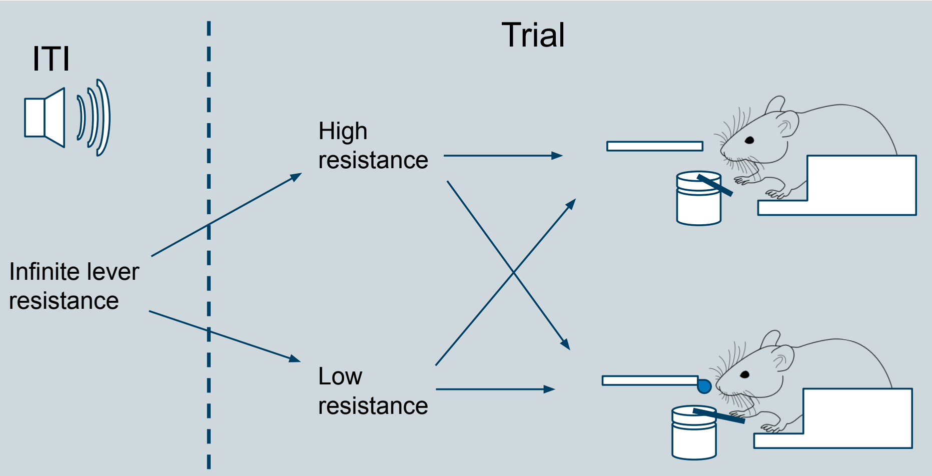




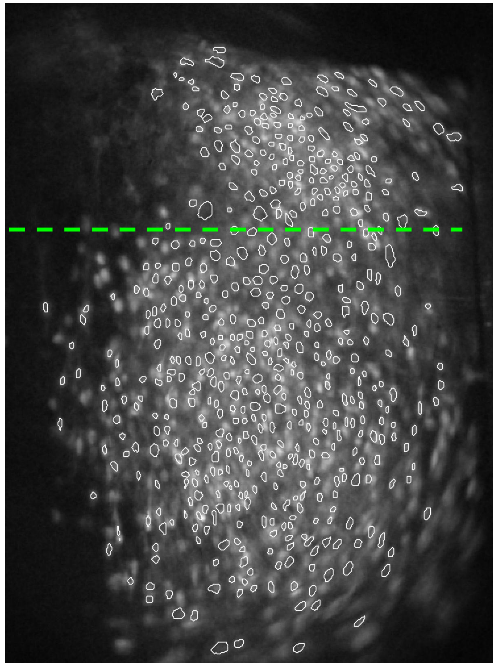


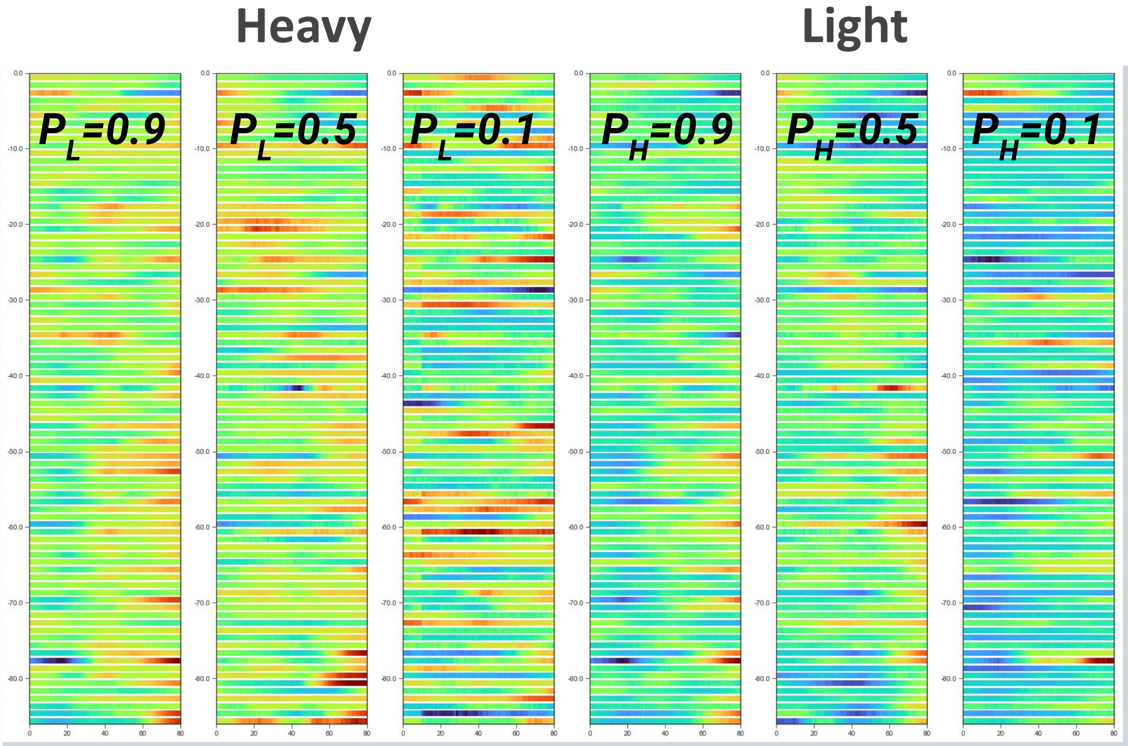
Light/Heavy Lever Pull Task

Sergey Zobnin, Naohiro Yamauchi



	Trials: H - heavy resistance, L - light resistance; standard, odd
Light session:	L L L L L L H L L L L L L L L H L L L L L L L L L L L L L L L H L L
Heavy session:	H H H L H H H H H H H H H H H L H H H H H H H L H H L H H
Uniform session:	H H L H L L L H L L H H L H H H H L L H L L L H L H H L H L H
Roving oddball paradigm (ROP):	H H H H H H H H H H L L L L L L L H H H H H H H H H L L L L L L Heavy train, Light train

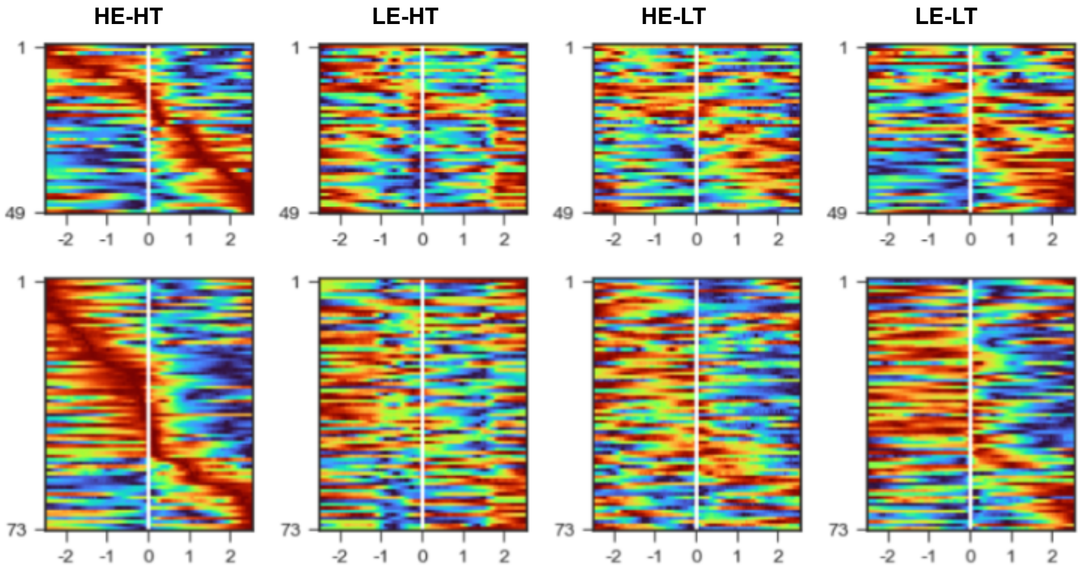




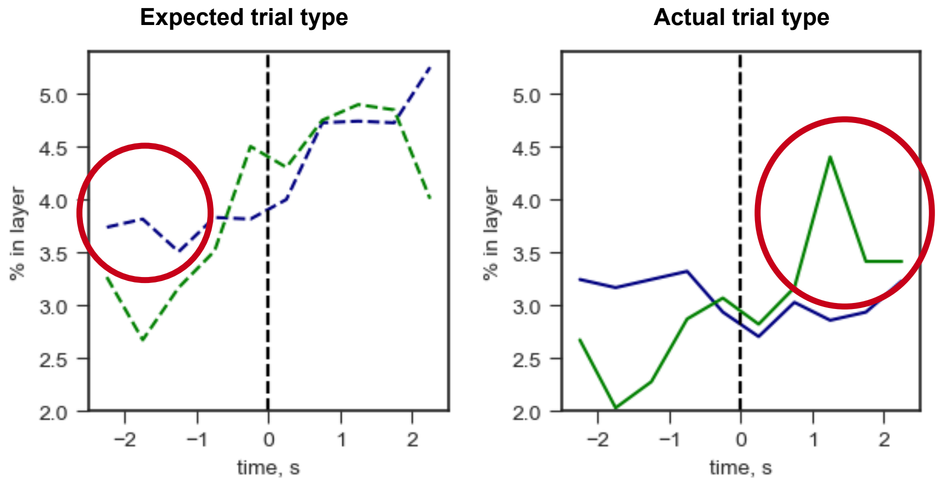
Expected and Actual Trial Type Coding

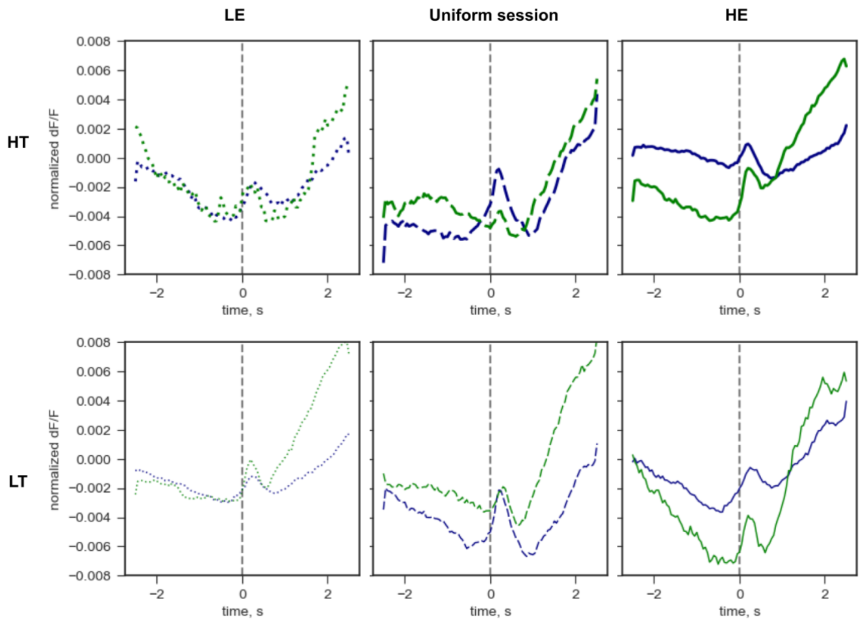
superficial

deep



Encoding analysis

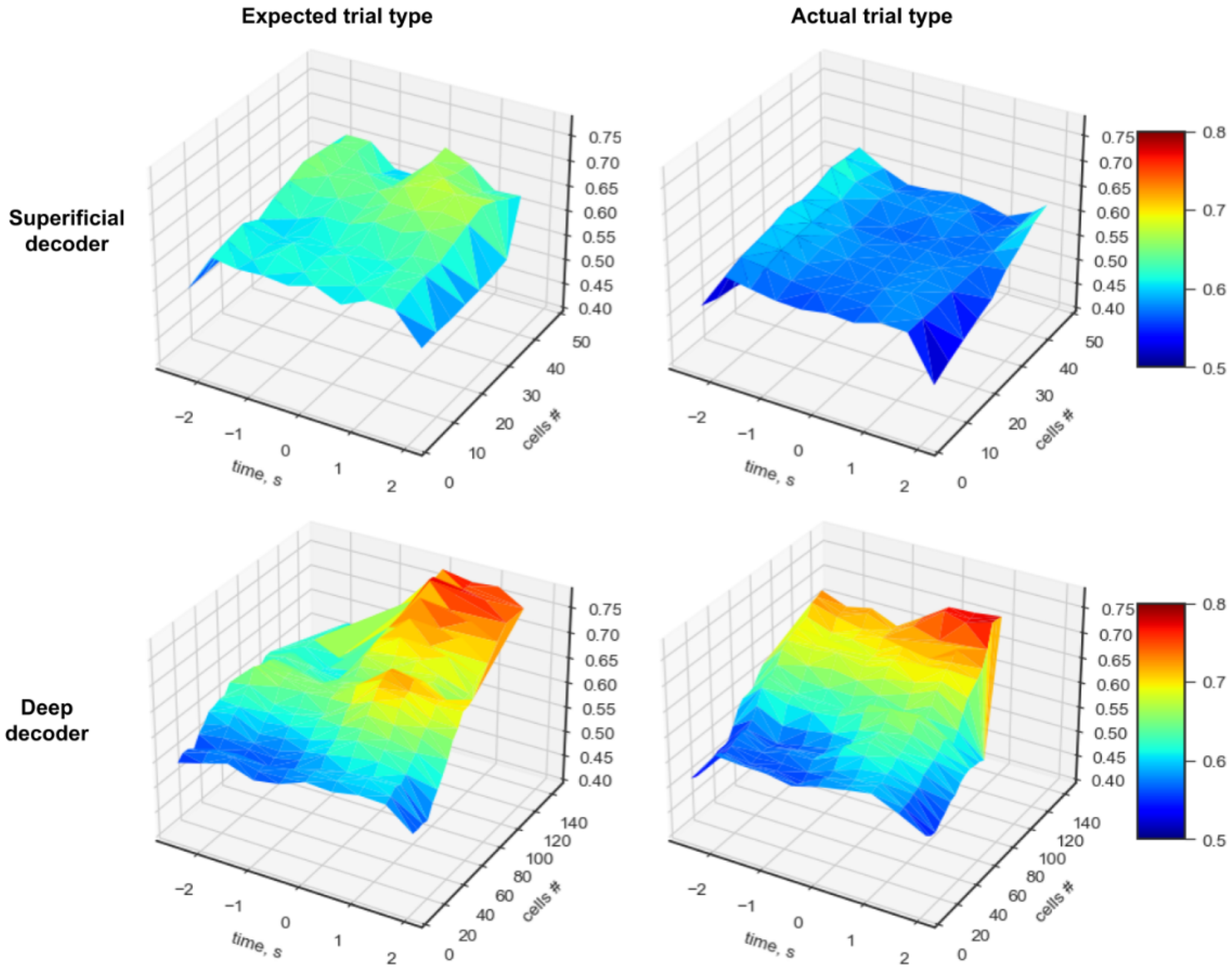




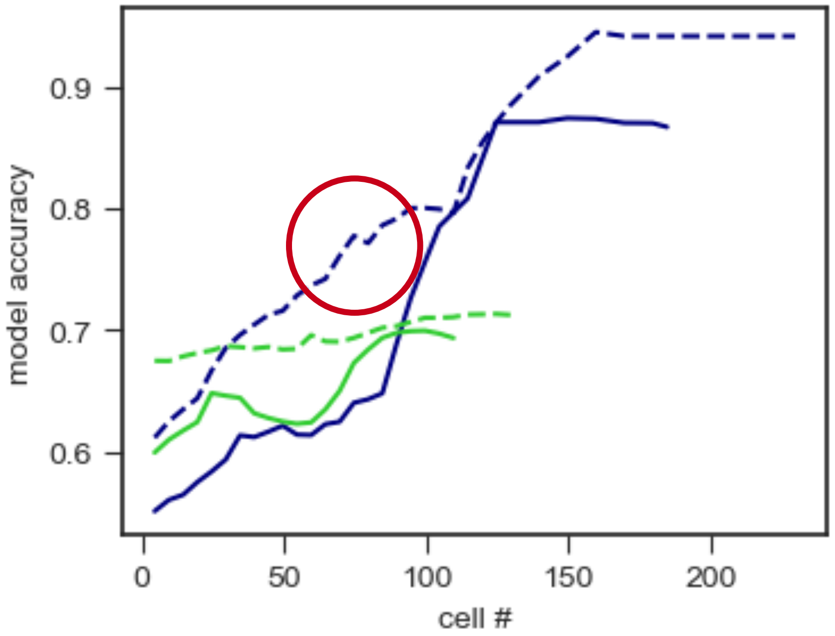
- More deep neurons code expected trial type before action
- More superficial neurons code actual trial type after action

Population Decoding

At different time points



Peak amplitude after pull



- Better decoding of expected trial type from deep neurons

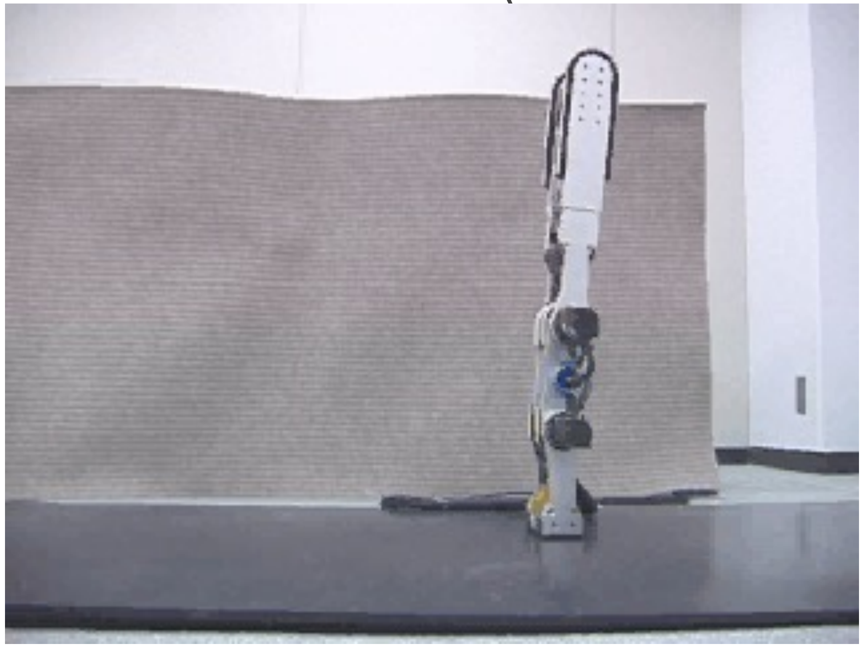
Question:

How can models and policies in separate brain areas be activated and connected as needed?

- fMRI study assumes that brain areas that perform required computations for given task are activated.
- But we don't know why that can be made possible!

Learning to Stand Up

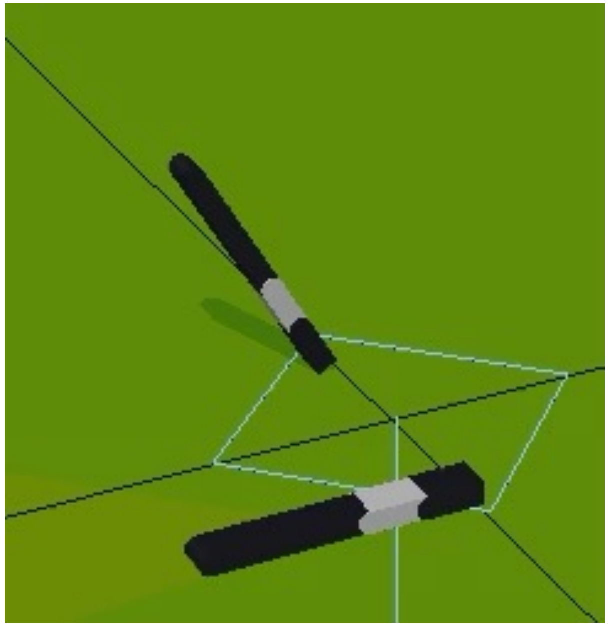
(Morimoto & Doya, 2001)

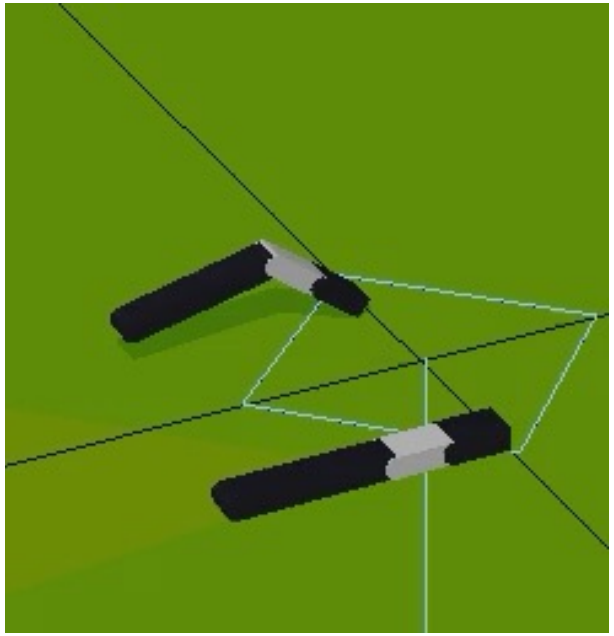


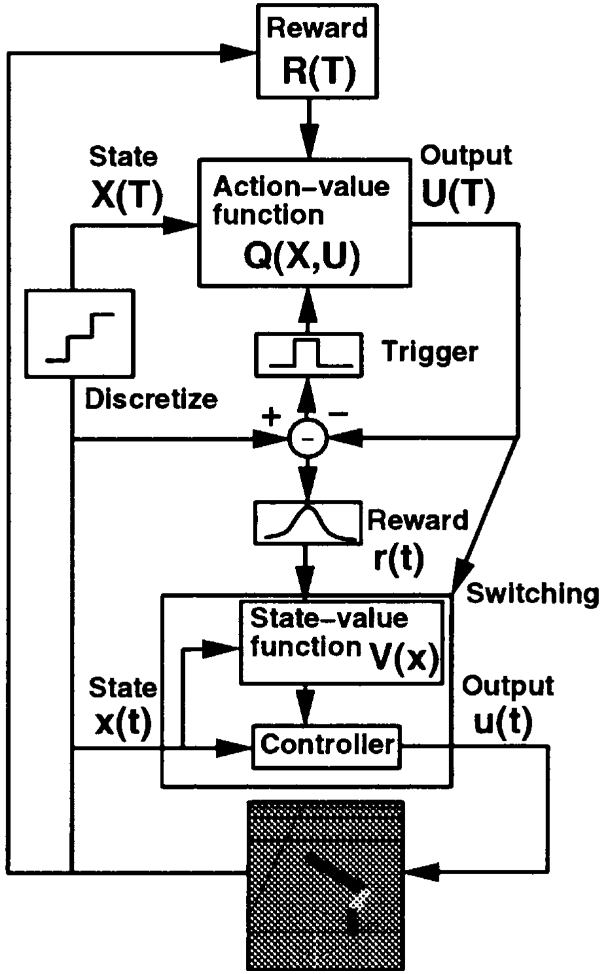
- Learning from reward and punishment
  - reward: height of the head
  - punishment: bump on the floor

Hierarchical Reinforcement Learning

- **Upper level**: reward: task goal
  - state: joint angles, center of mass
  - action: desired postures
- **Lower level**: reward: achieving a subgoal
  - state: joint/pitch angles, angular velocity
  - action: motor torque





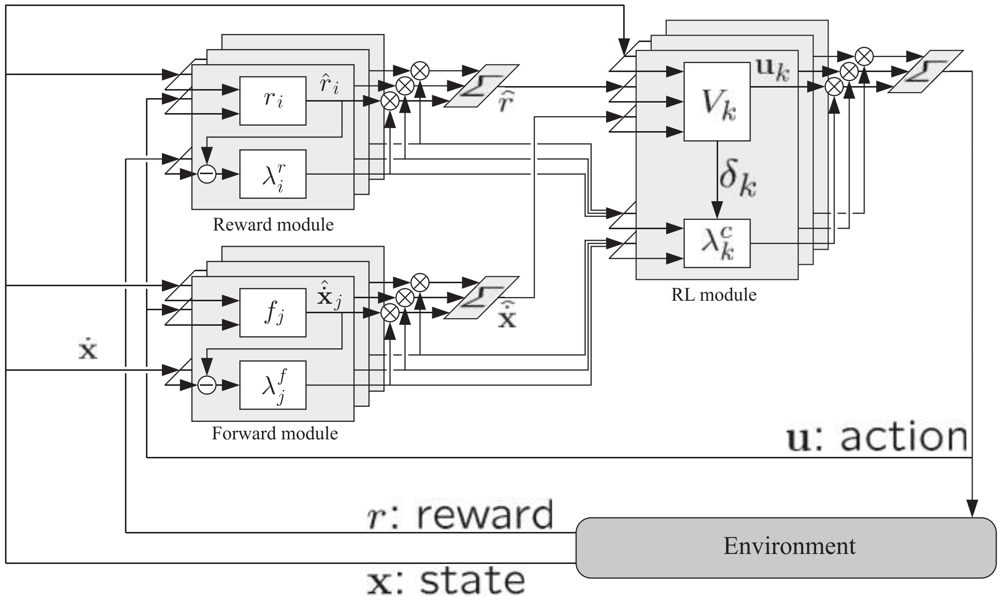


(Morimoto & Doya, 2001)

How to Select/Connect Right Modules?

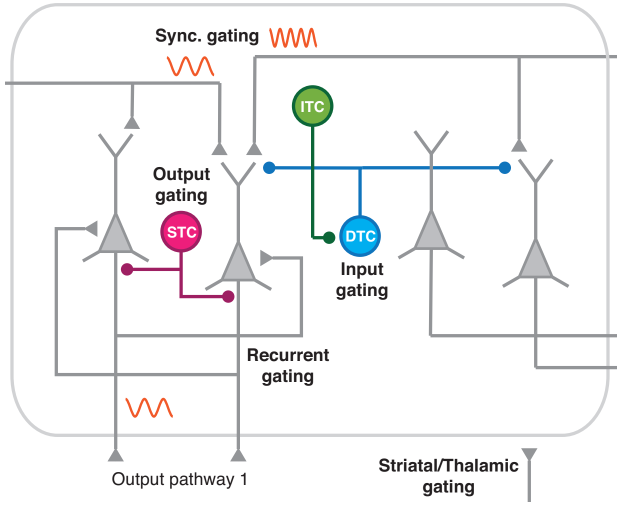
Computational principles

- prediction error (Wolpert & Kawato, 1998)
- Bellman error (Sugimoto et al., 2012)
- uncertainty (Daw et al., 2005)
- modular infomax?



Biophysical mechanisms

- basal ganglia/thalamus (Eliasmith et al. 2012)
- affordance competition (Cisek, 2007)
- dendritic disinhibition (Wang & Yang, 2018)
- rhythm/coherence?



Reinforcement Learning

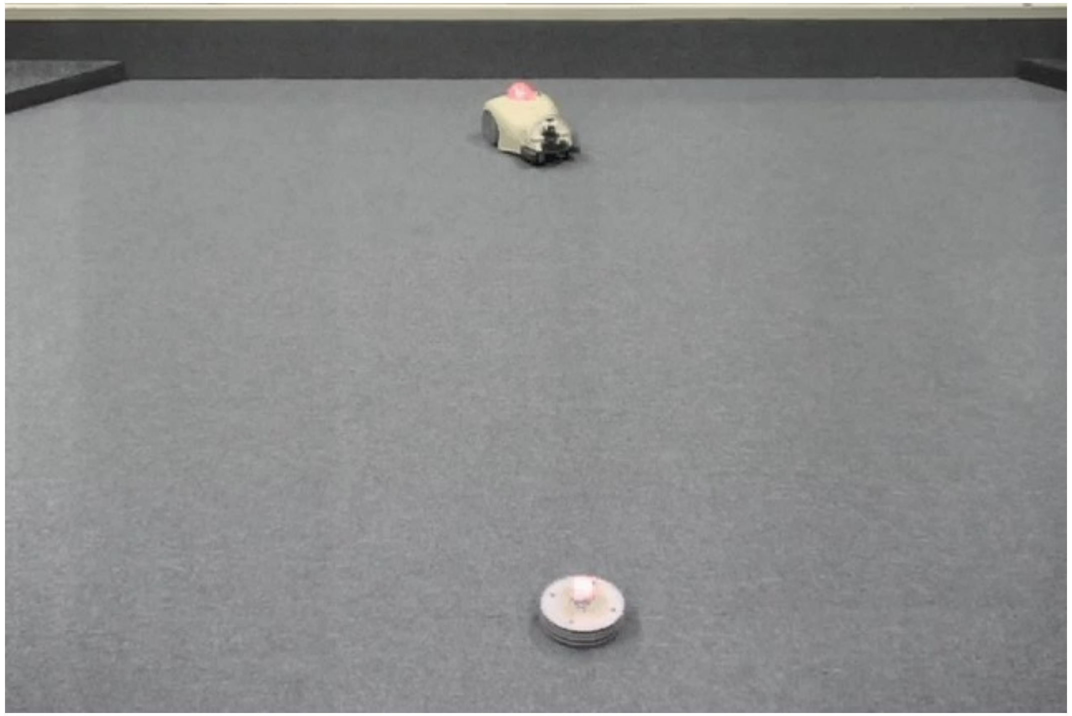
- Predict reward: *value function*
  - $V(s) = E[\, r(t) + \gamma r(t+1) + \gamma^2 r(t+2) \ldots |\ s(t)=s]$
  - $Q(s,a) = E[\, r(t) + \gamma r(t+1) + \gamma^2 r(t+2) \ldots |\ s(t)=s,\ a(t)=a]$
- Select action

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

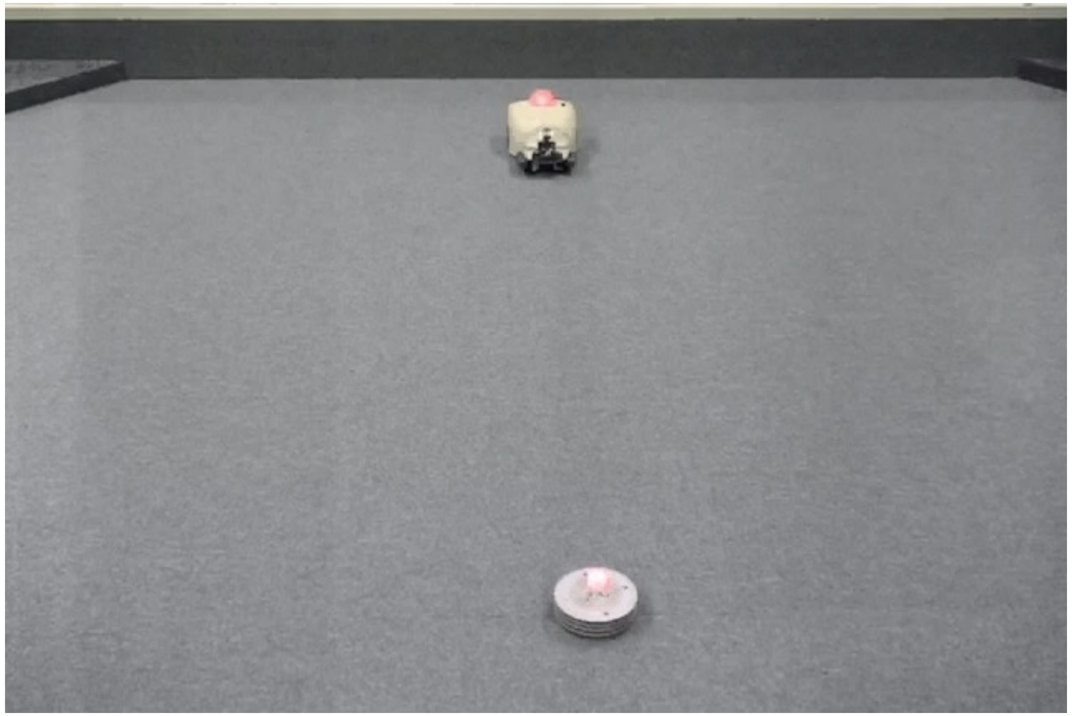
    How to tune these parameters?
  - $\Delta Q(s(t),a(t)) = \alpha\ \delta(t)$

Temporal Discount Factor γ

- Large γ
  - reach for far reward

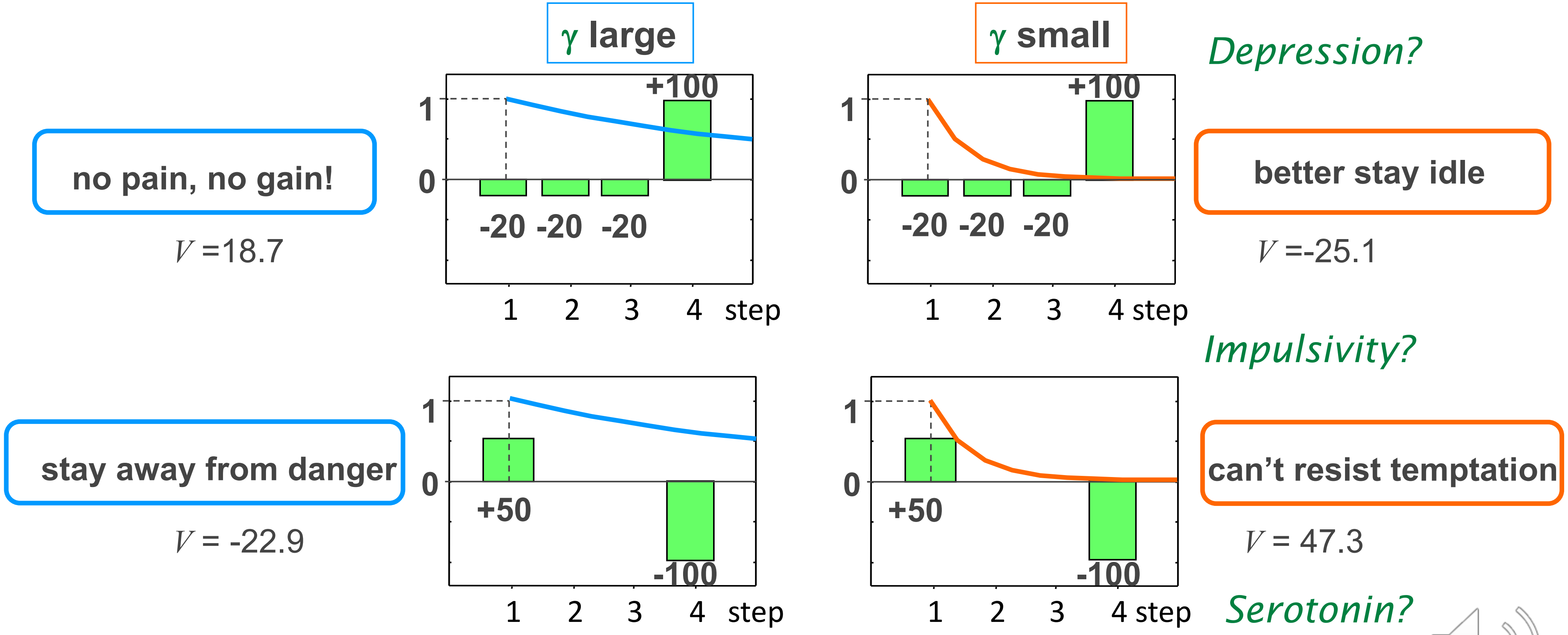


- Small γ
  - only to near reward



Temporal Discount Factor γ

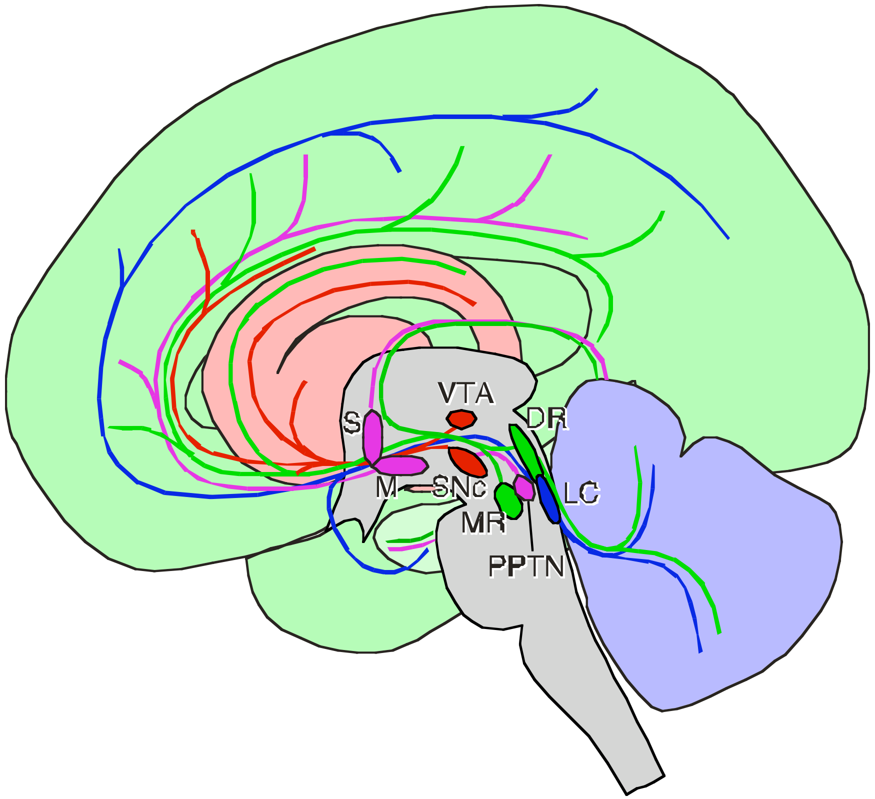
- $V(t) = E[\, r(t) + \gamma r(t+1) + \gamma^2 r(t+2) + \gamma^3 r(t+3) + \ldots]$
  - controls the ‘character’ of an agent



Neuromodulators for Metalearning

(Doya, 2002)

- *Metaparameter* tuning is critical in RL
  - How does the brain tune them?



Dopamine: TD error δ

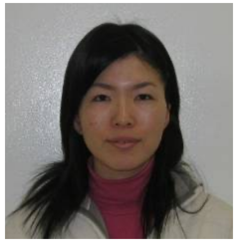
Acetylcholine: learning rate α

Noradrenaline: exploration β

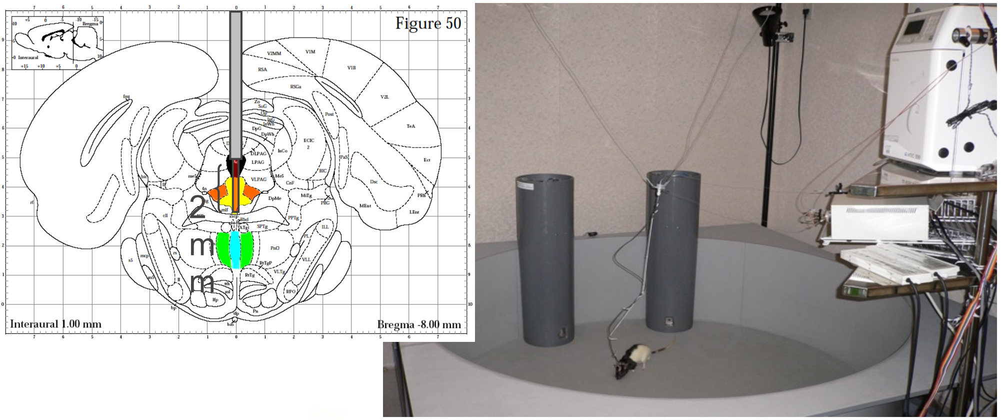
Serotonin: temporal discount γ

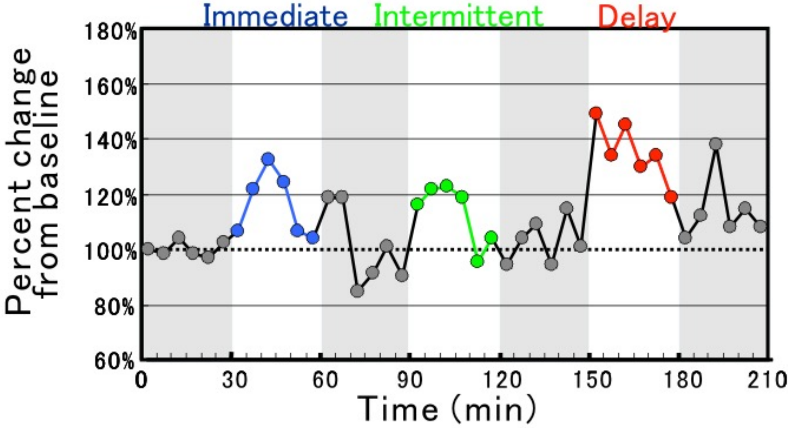
Chemical Measurement/Control

(Kayoko Miyazaki et al., 2011, 2012)



Microdialysis measurement

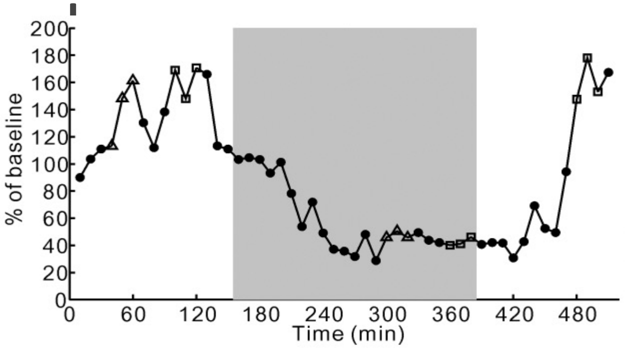


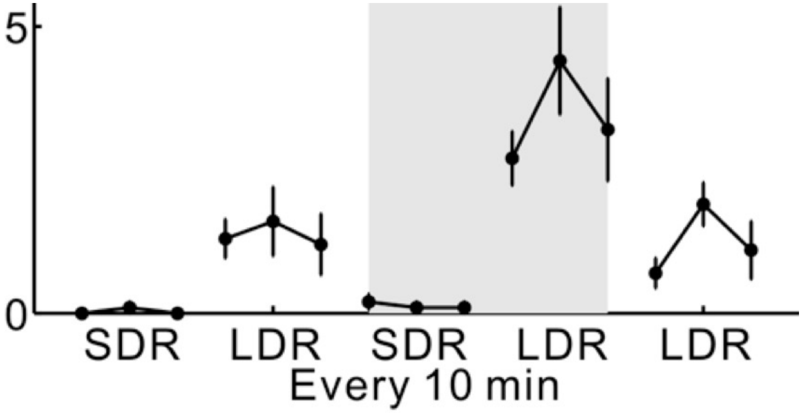


- Serotonin release increased in delayed reward task

Serotonin neuron blockade

- 5HT1A agonist in dorsal r

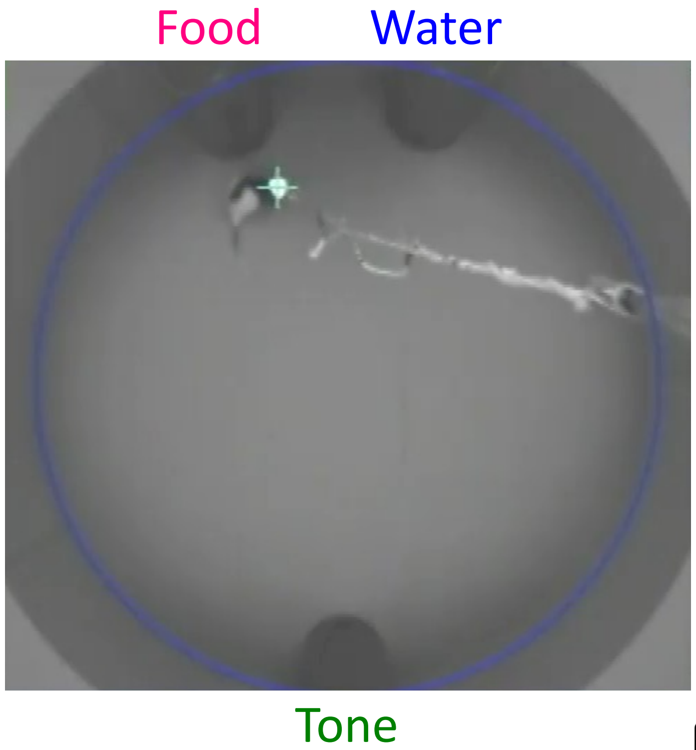


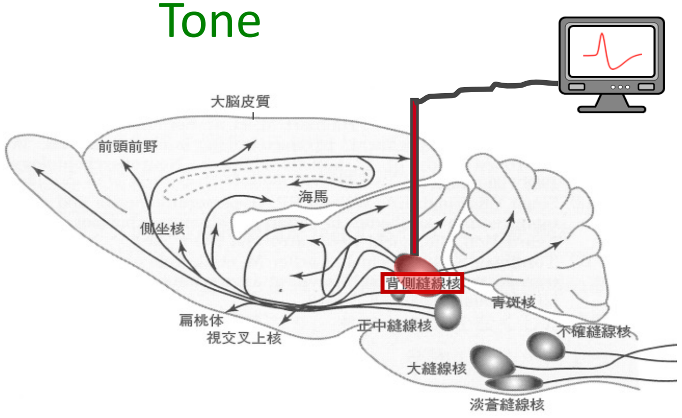


- Waiting error increased in long-delayed reward trials

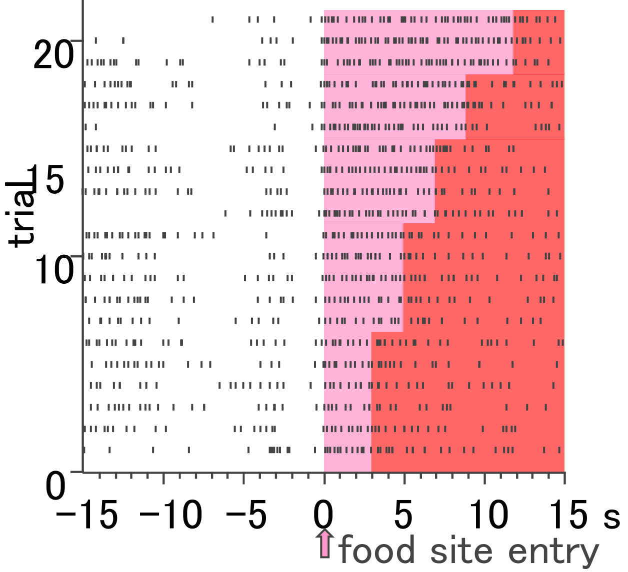
Dorsal Raphe Neuron Recording

(Miyazaki et al. 2011 JNS)

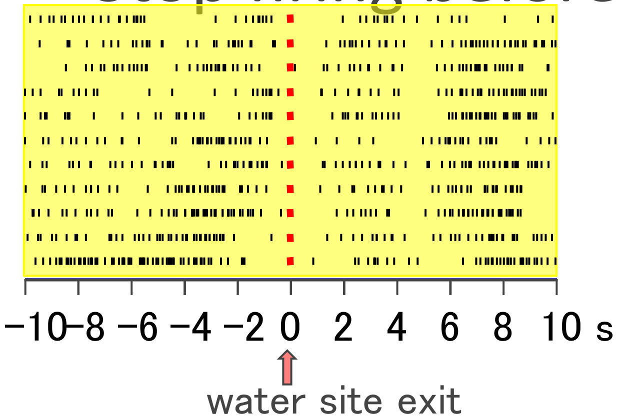




- Keep firing while waiting

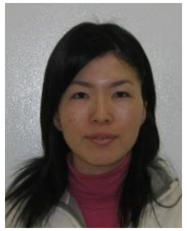


- Stop firing before giving up



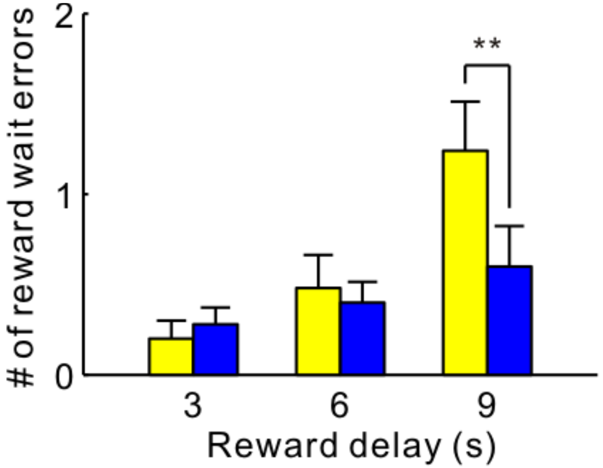
Optogenetic Stimulation of Serotonin Neurons

(Miyazaki et al., 2014, Current Biology)

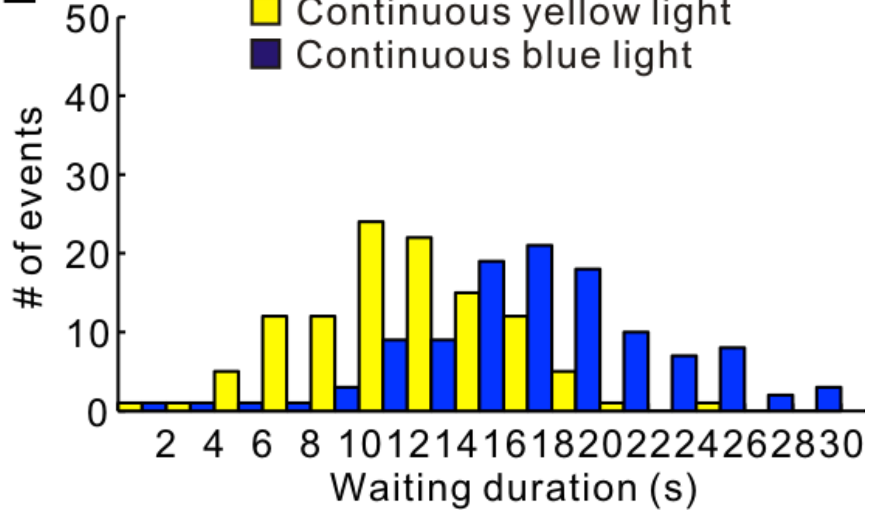


- Reward Delay Task (3, 6, 9, ∞ sec)

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



E

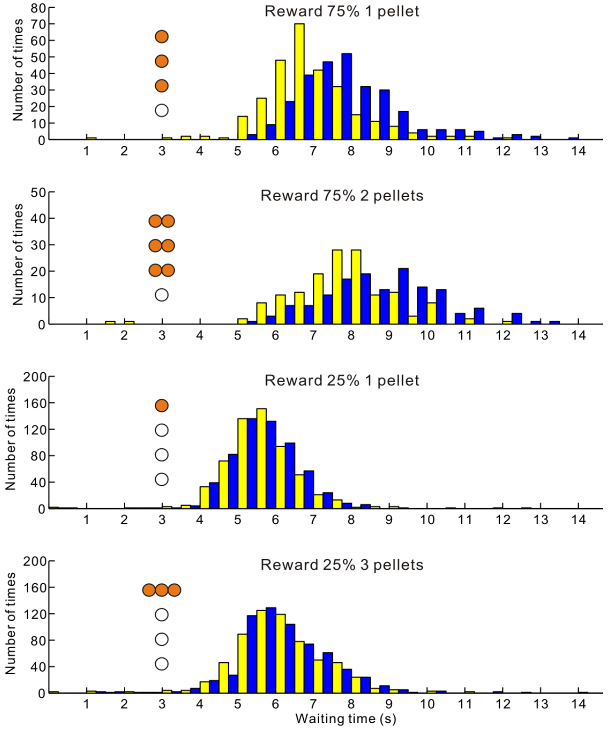


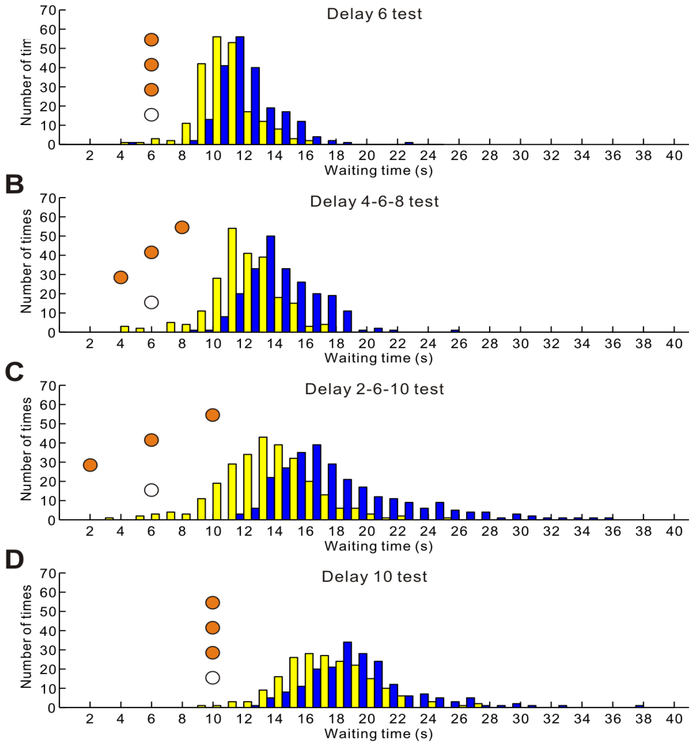
nature COMMUNICATIONS (2018)

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]

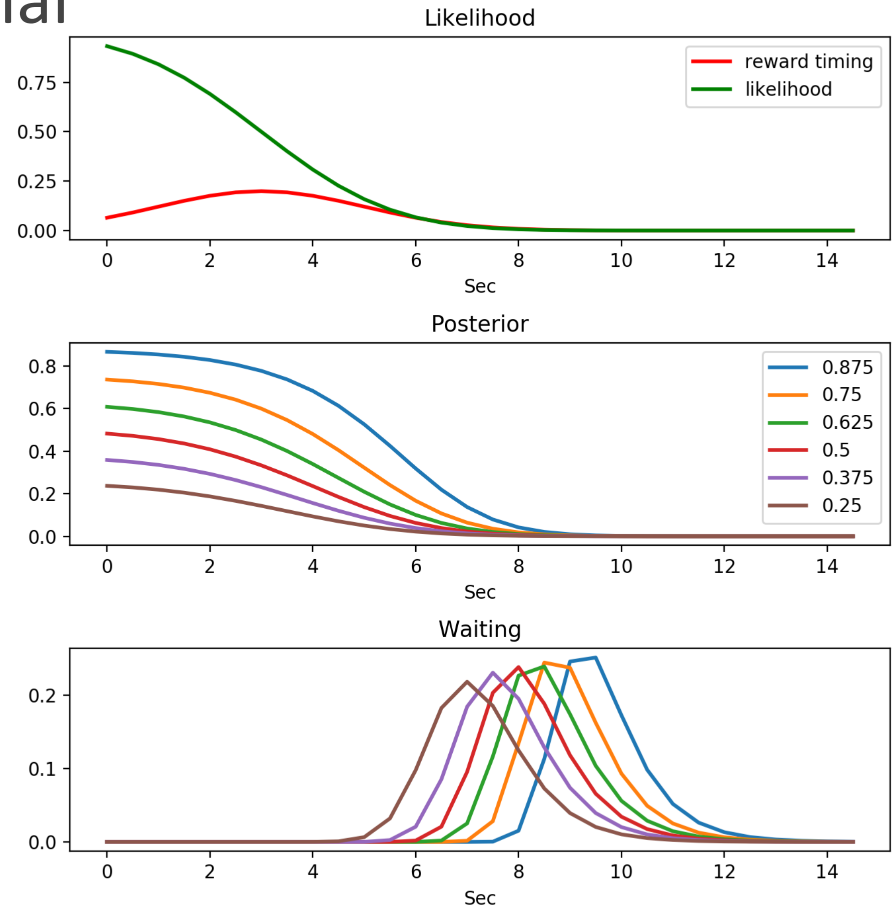
- Serotonin stimulation facilitates waiting when...
  - reward delivery is certain
  - reward timing is uncertain





Bayesian Waiting Decision Model

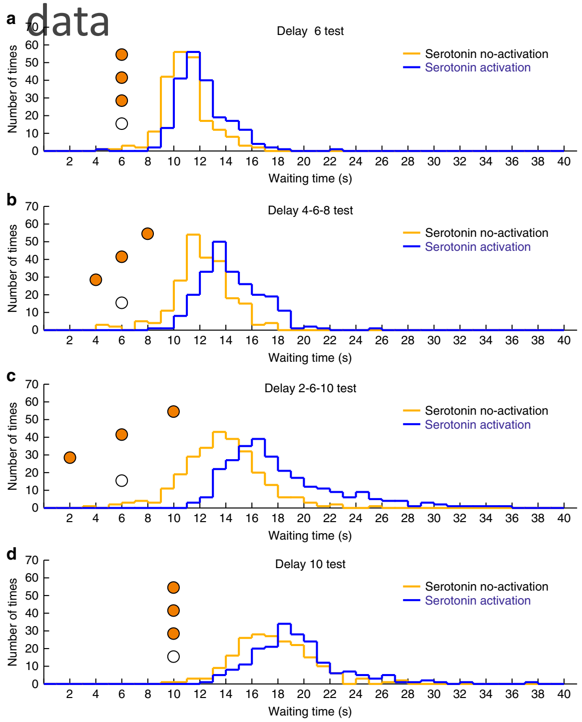
- Mice have internal model of reward timing
  - keep guessing if it is a rewarded trial
- Likelihood of reward drops
  - higher prior sustains posterior
  - timing uncertainty makes long-tailed likelihood
- Serotonin signal reward prior?
  - average reward response (Cohen et al., 2015)



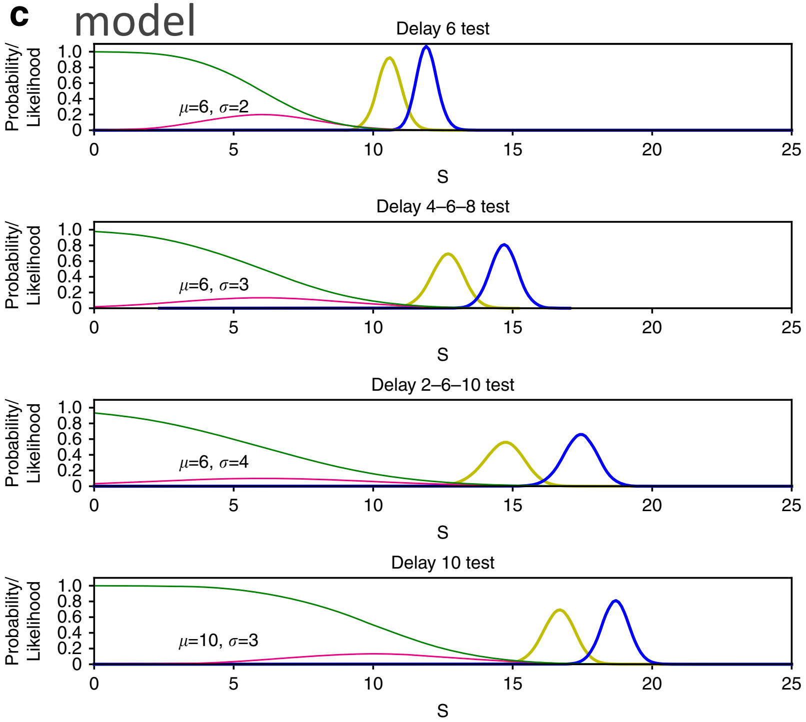
Effect of Timing Uncertainty

- 5-HT stimulation causes longer waiting when reward timing is more uncertain.
- Bayesian model replicates the effect by assuming that 5-HT enhances prior probability of reward.

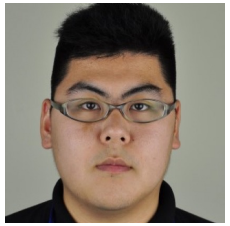
data



model

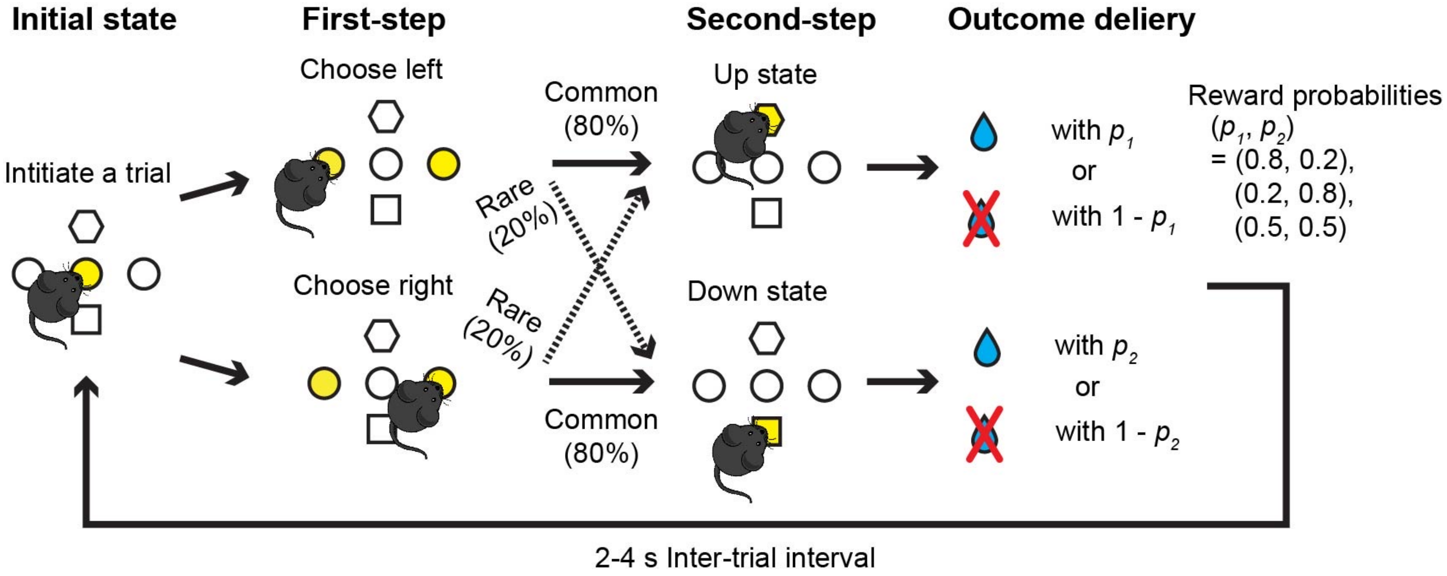


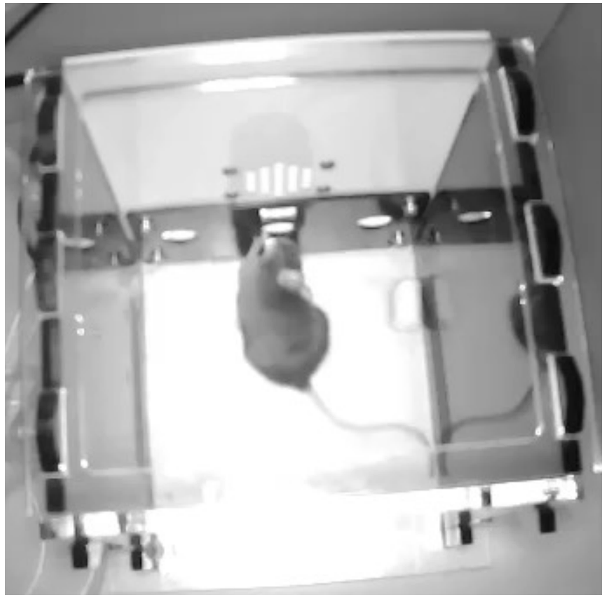
Serotonin for Model-based RL?



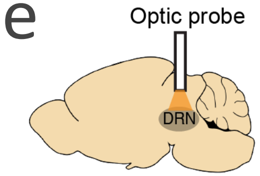
Masakazu Taira

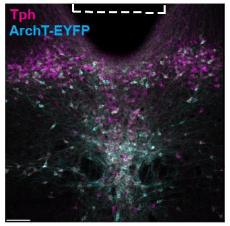
Two-step task for mice (Akam et al. 2020)





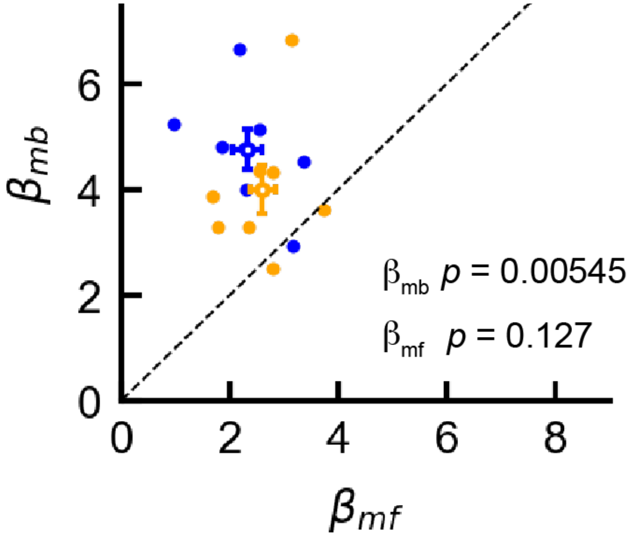
- Tph2-ArchT mice





- Hybrid model

$$Q_{net}(a) = \beta_{\mathrm{mf}} Q_{mf}(a) + \beta_{mb} Q_{mb}(a)$$



What Should We Further Learn from the Brain?

Energy Efficiency

Data Efficiency

- World Models and Mental Simulation
- Modularity and Compositionality
- Meta-learning

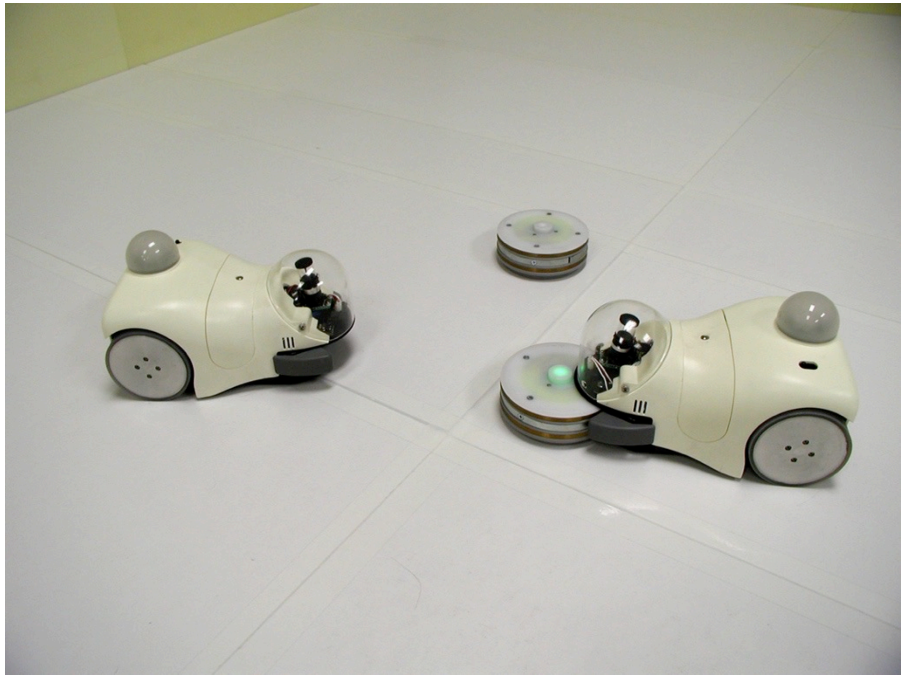
Autonomy and Sociality

Cyber Rodent Project (Doya & Uchibe, 2005)

What is the origin of rewards?

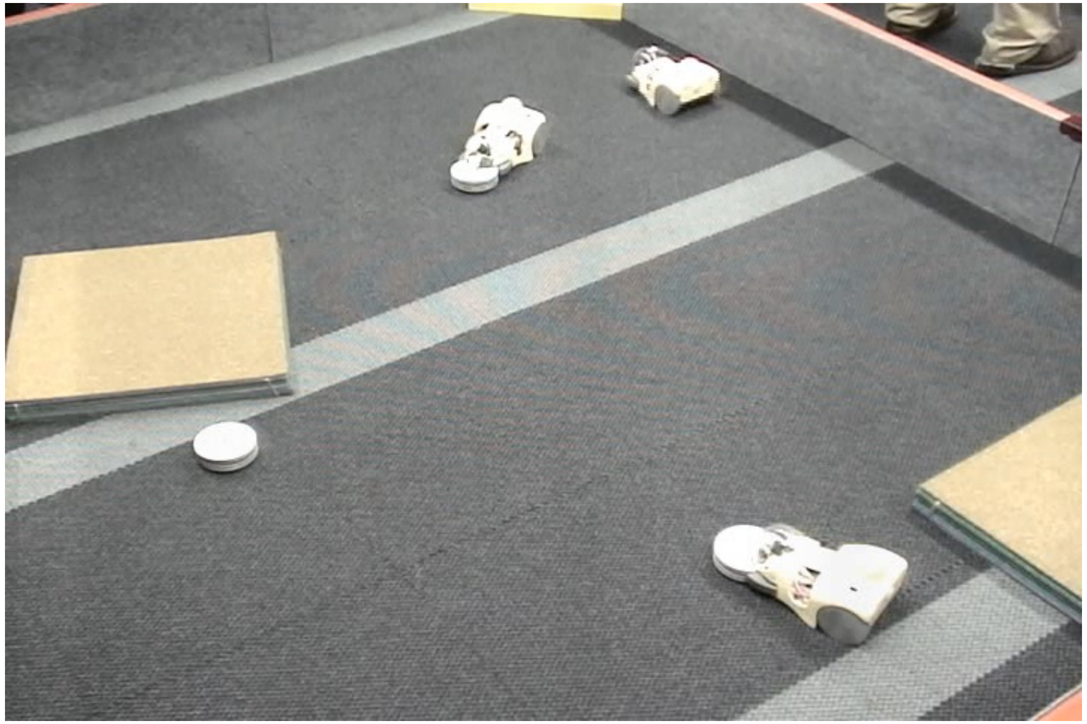
Robots with same constraint as biological agents

- Self-preservation
  - capture batteries
- Self-reproduction
  - exchange programs through IR ports

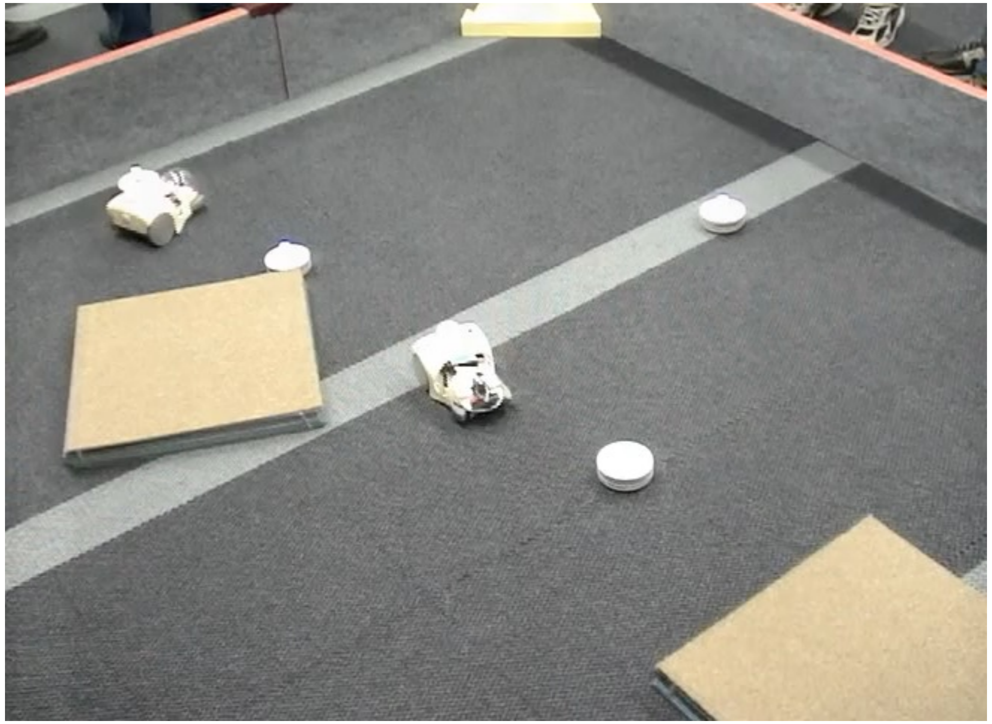


Learning to Survive and Reproduce

- Catch battery packs
  - survival



- Copy ‘genes’ by IR ports
  - reproduction, evolution

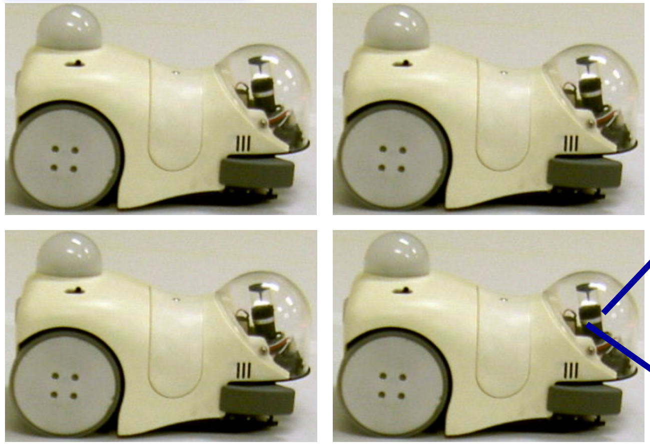


(Doya & Uchibe, 2005)

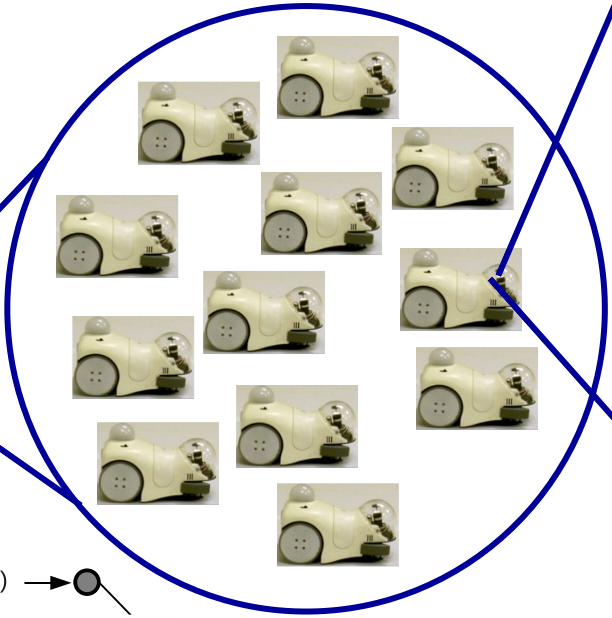
Embodied Evolution (Elfwing et al., 2011)

Population

Robots



Virtual agents 15-25



Genes

Weights for top layer NN

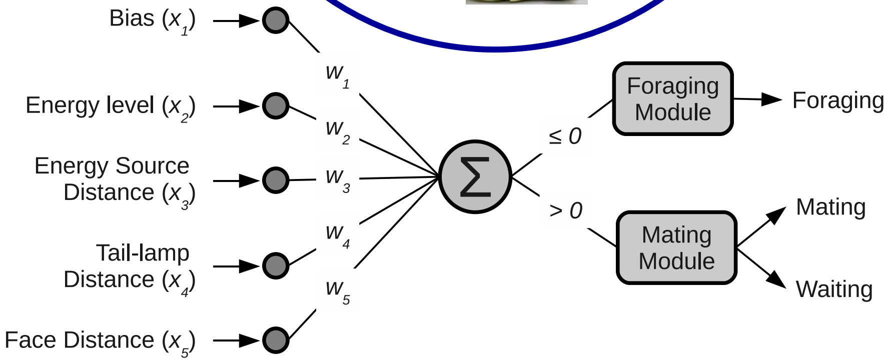
$w_1, w_2, \ldots, w_n$

Weights shaping rewards

$v_1, v_2, \ldots, v_n$

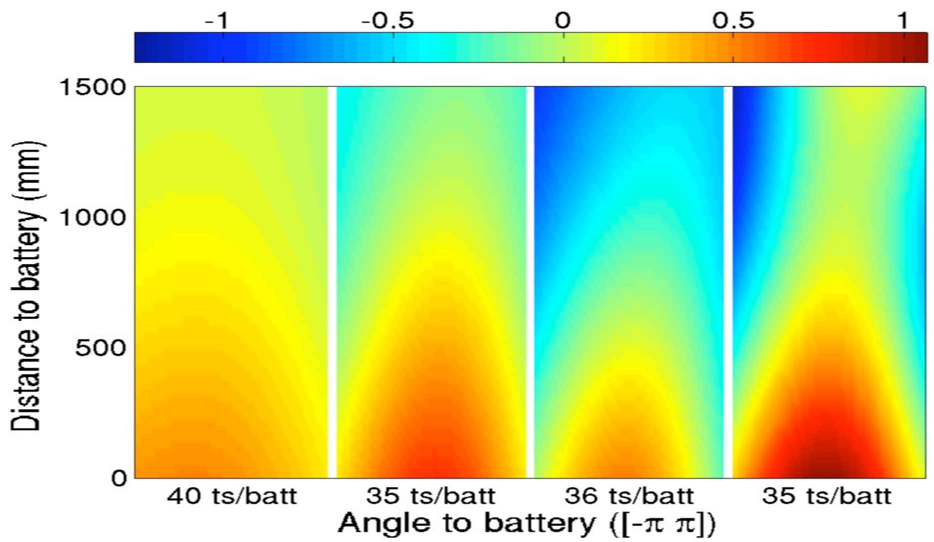
Meta-parameters

$\alpha \gamma \lambda \tau_k \tau_0$

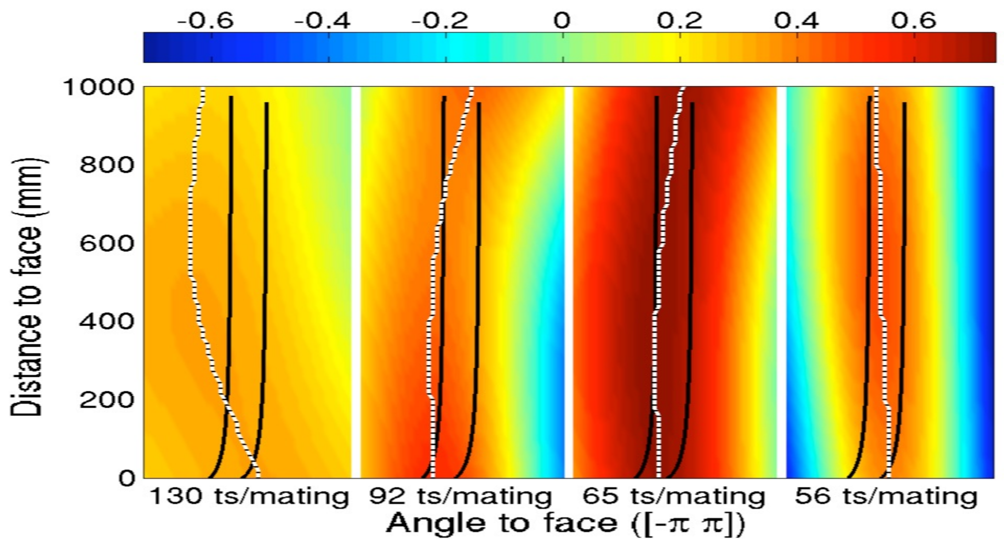


Evolution of Shaping Rewards

- Vision of battery



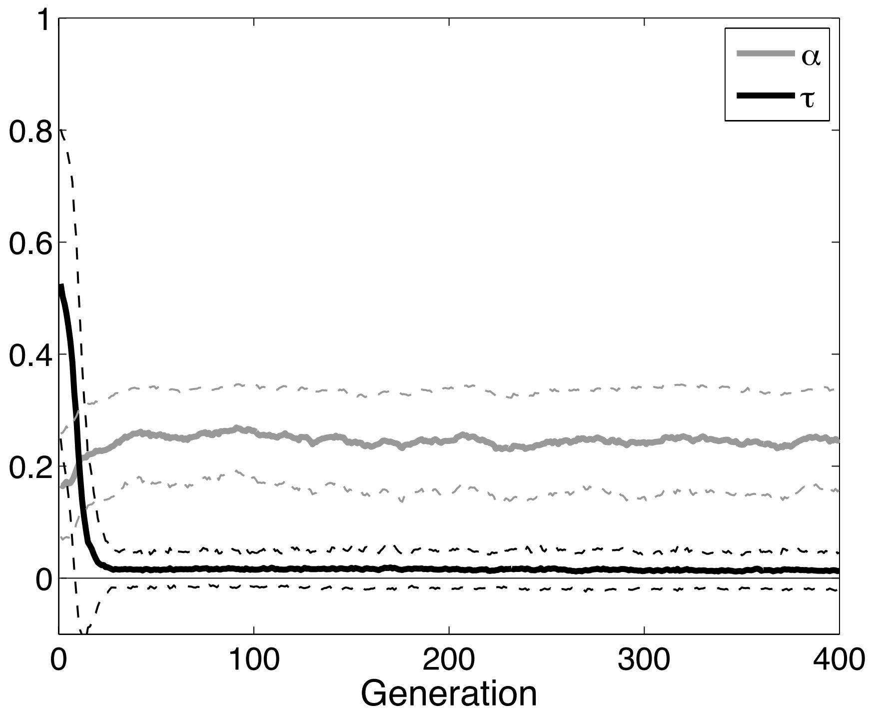
- Vision of face

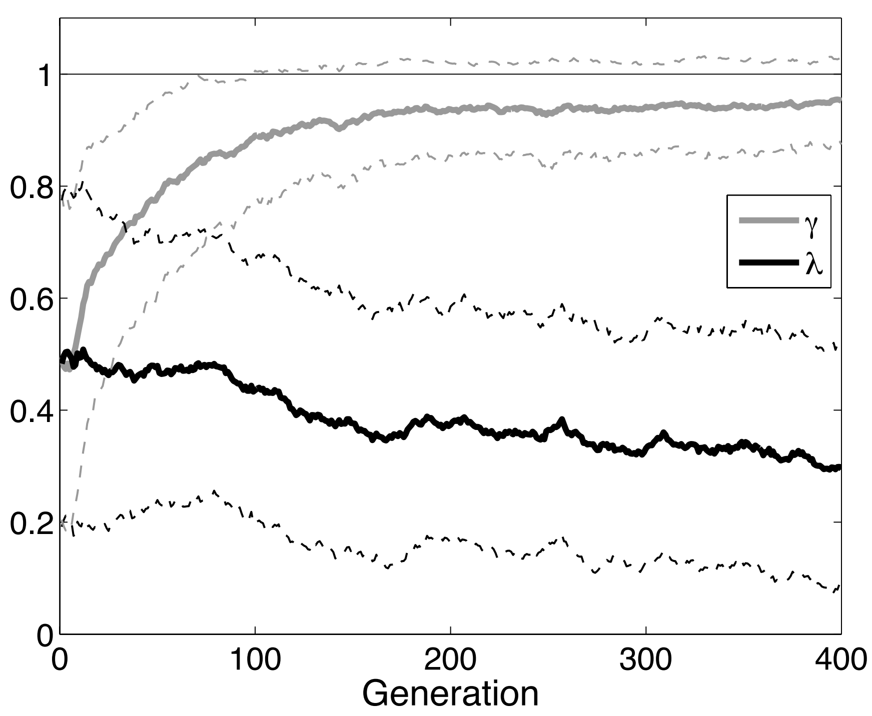


(Elfwing et al., 2011)

Evolution of Meta-Parameters

- Learning rate α
- Exploration temperature τ
- Temporal discount factor γ
- Eligibility trace decay factor λ

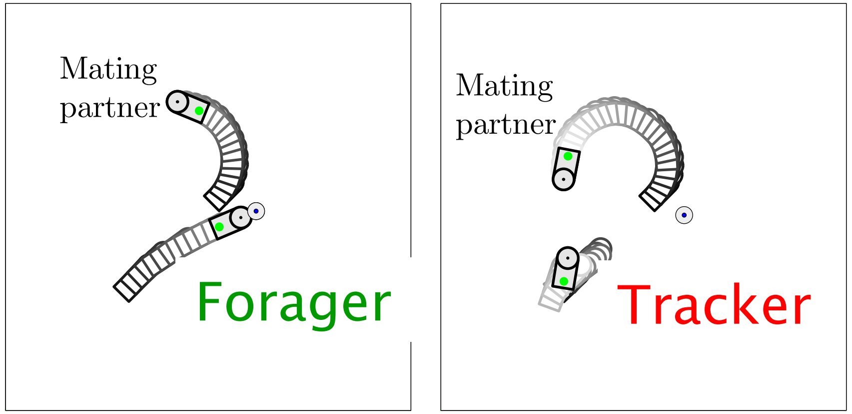


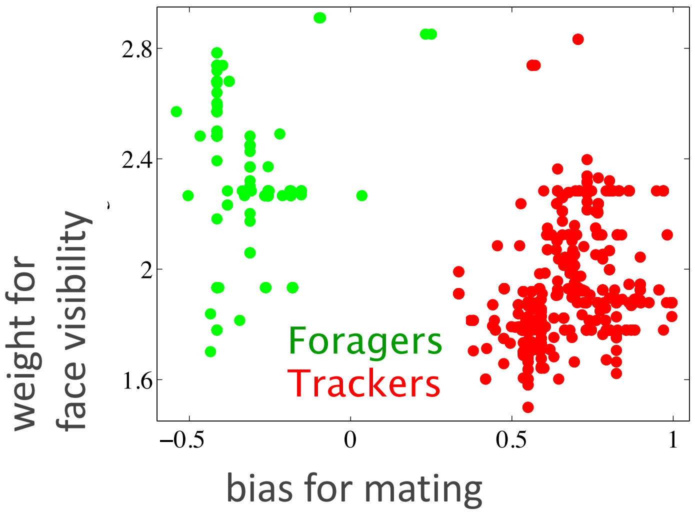


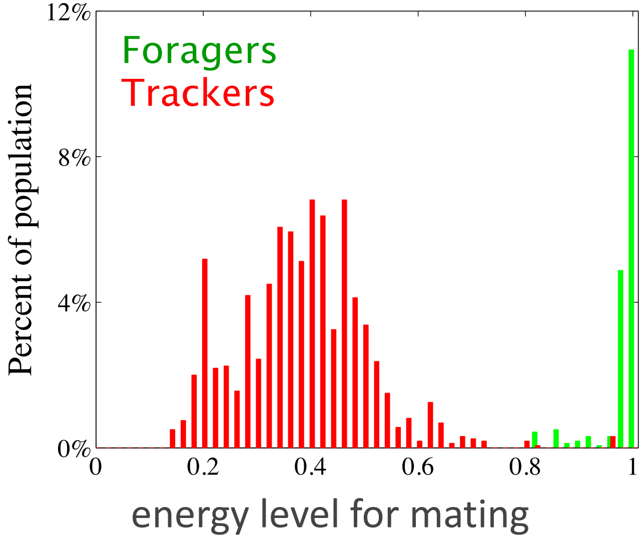
Polymorphism within Colony

(Elfwing et al. 2014)

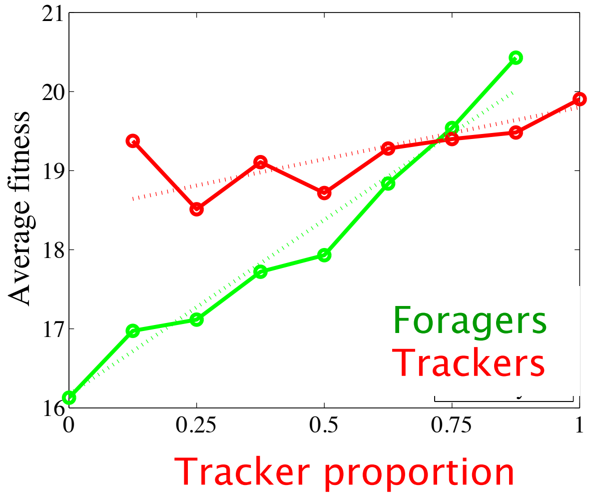
- Foragers and Trackers





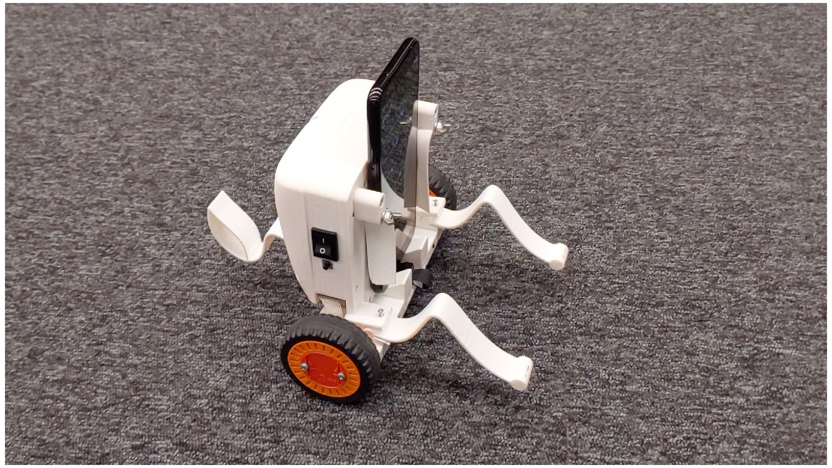


- Evolutional stability

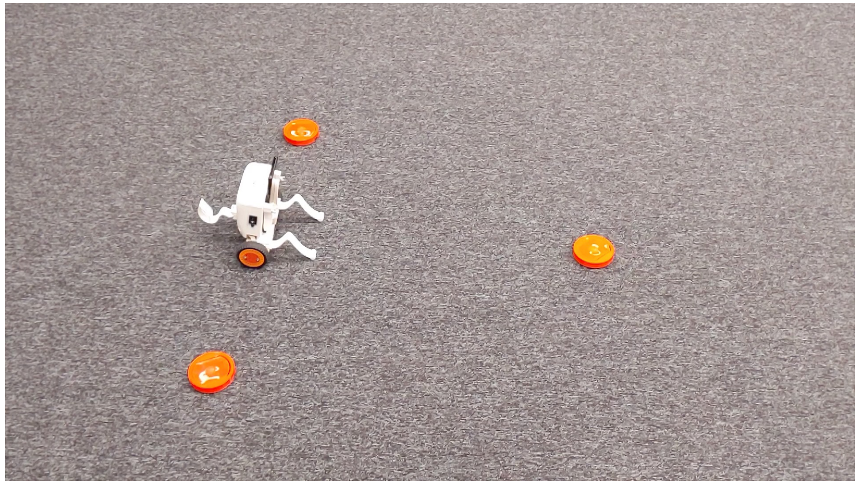


Smartphone Robot Project

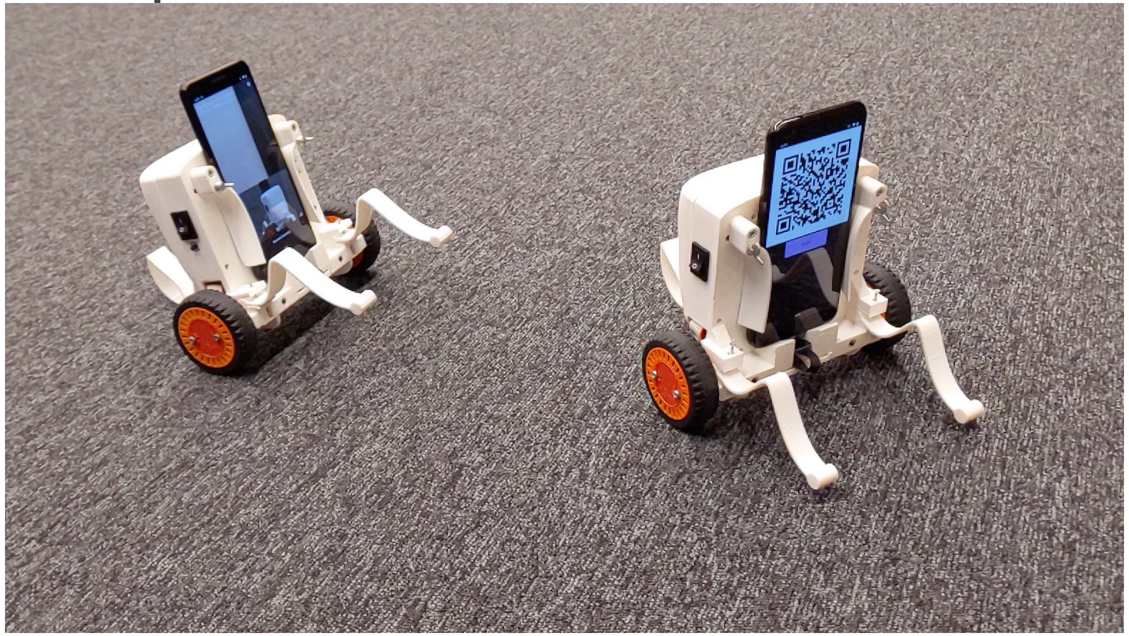
- Motor control



- Survival



- Reproduction



- Learning models of world and others
- Meta-learning
- Evolution of rewards and curiosity
- …

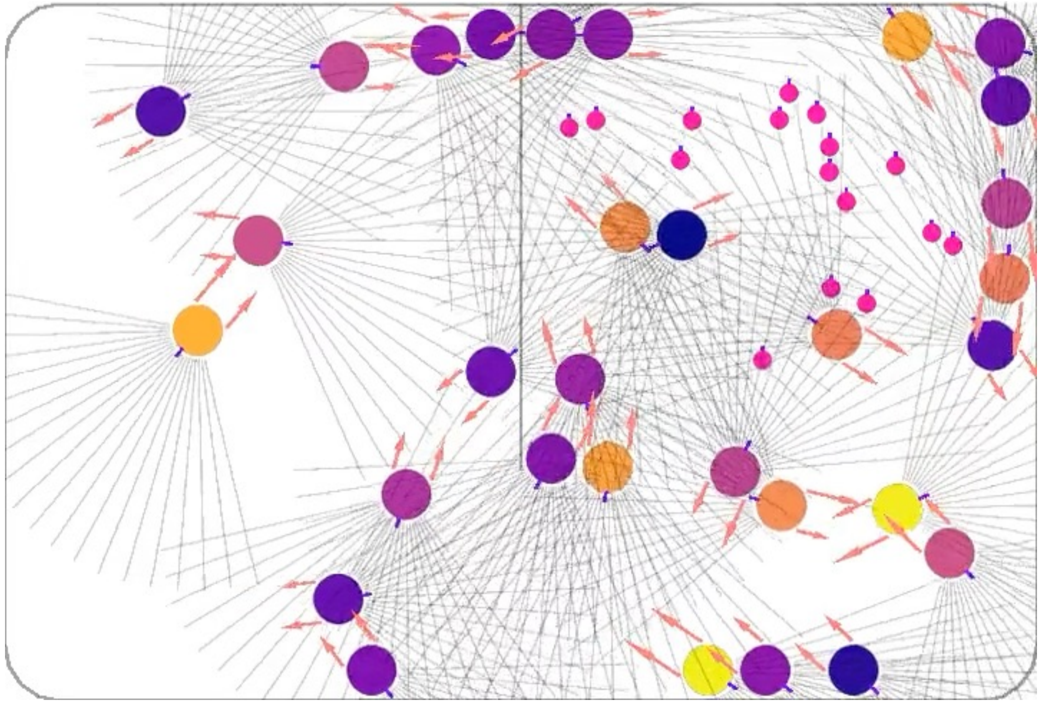
Evolution of Primary Rewards

Yuji Kanagawa

Reproduction Model

- age t
- energy e

- Birth rate $b(e)$
- Death rate $h(t,e)$



Evolution of Reward Function

$$r = r_{\text{agent}} + r_{\text{food}} + r_{\text{wall}} + r_{\text{action}}$$

Learning by Proximal Policy Optimization (PPO; Schulman et al. 2017)

Computational Correlates of "Curiosity"

- Model-free
  - supplementary reward: $r_{int}(s,a)$
  - shaping reward: $r_{sh}(s_t) = \gamma\Phi(s_t) - \Phi(s_{t-1})$
  - optimistic initial value: $Q_0(s,a)$
  - high temperature τ: $P(a|s) \propto \exp[\, Q(s,a)/\tau]$
- Model-based
  - learning internal models: $P(o|s)$, $P(s'|s,a)$, $P(r|s,a)$
  - clarifying the present: $P(s_t) \propto P(o_t | s_t)P(s_t | s_{t-1},a_{t-1})$
  - simulating the future: $P(s_{t+1}|s_t,a)$ …multiple steps
  - finding optimal policy: $\pi^*(a|s)$

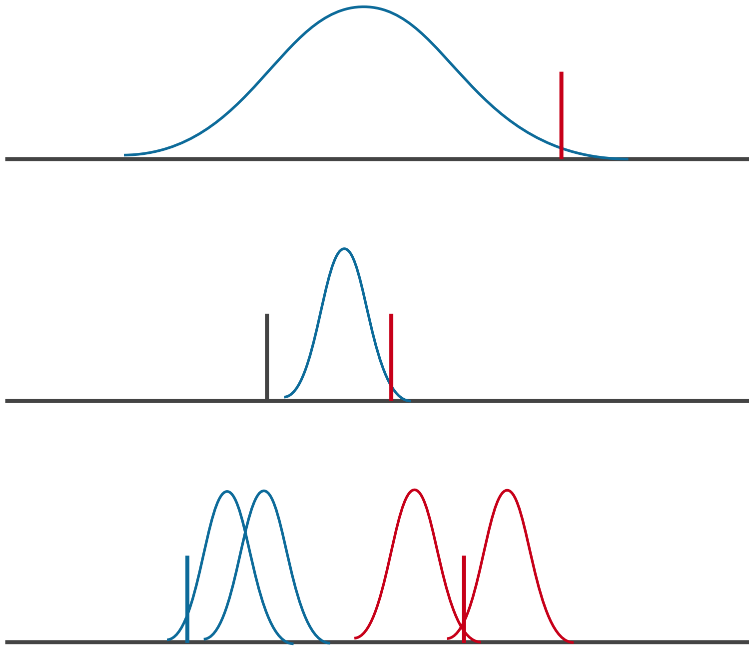
Evolving Intrinsic Rewards

Tojo Rakotoaritina

How to model/implement curiosity?

(Oudeyer & Kaplan 2008; Sing et al. 2010; Aubret et al. 2023)

- Novelty … memory
  - visit count
  - $-\log p(s)$
- Surprise … prediction
  - prediction error
  - $-\log p(s'|s, a)$
- Empowerment … control
  - $I(s'; a) = H(s') - H(s'|a)$



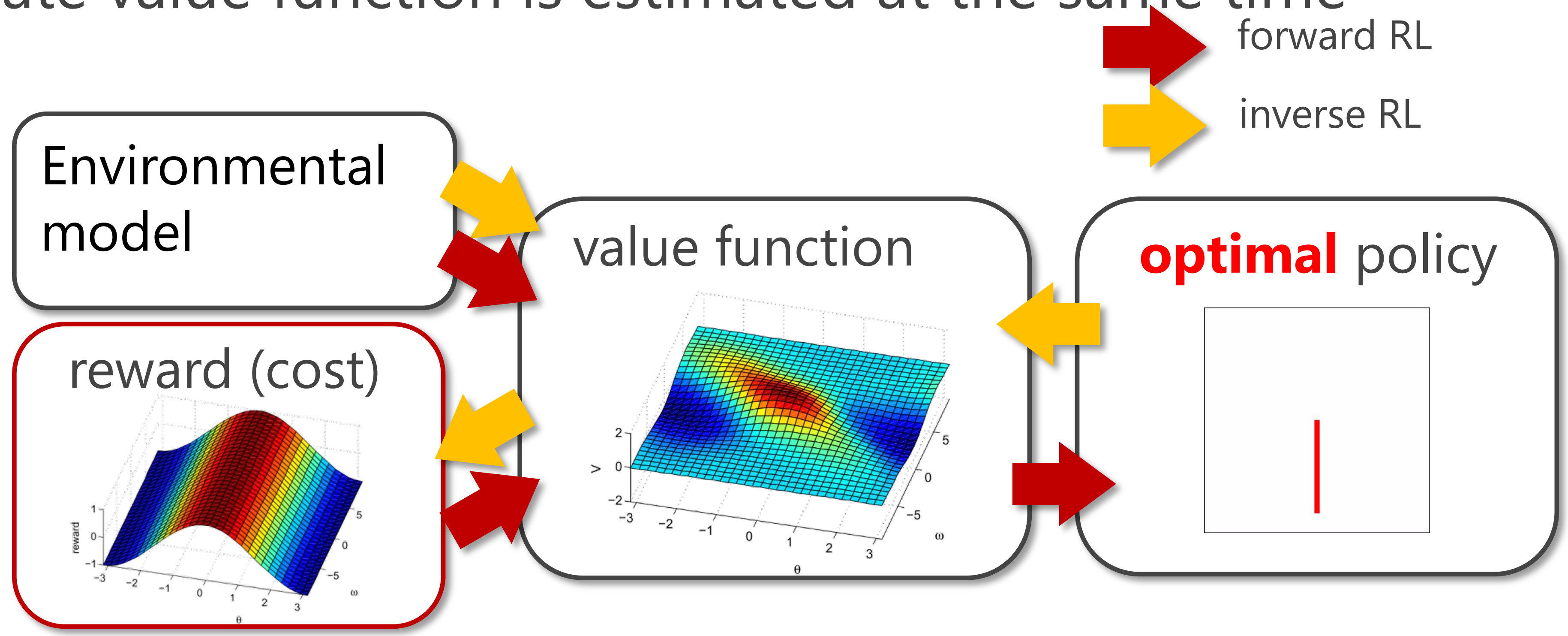
(Klybin et al. 2005)

$$r_{\text{intrinsic}} = r_{\text{novelty}} + r_{\text{surprise}} + r_{\text{empowerment}}$$

Inverse Reinforcement Learning

To estimate reward function from observed (optimal) behaviors

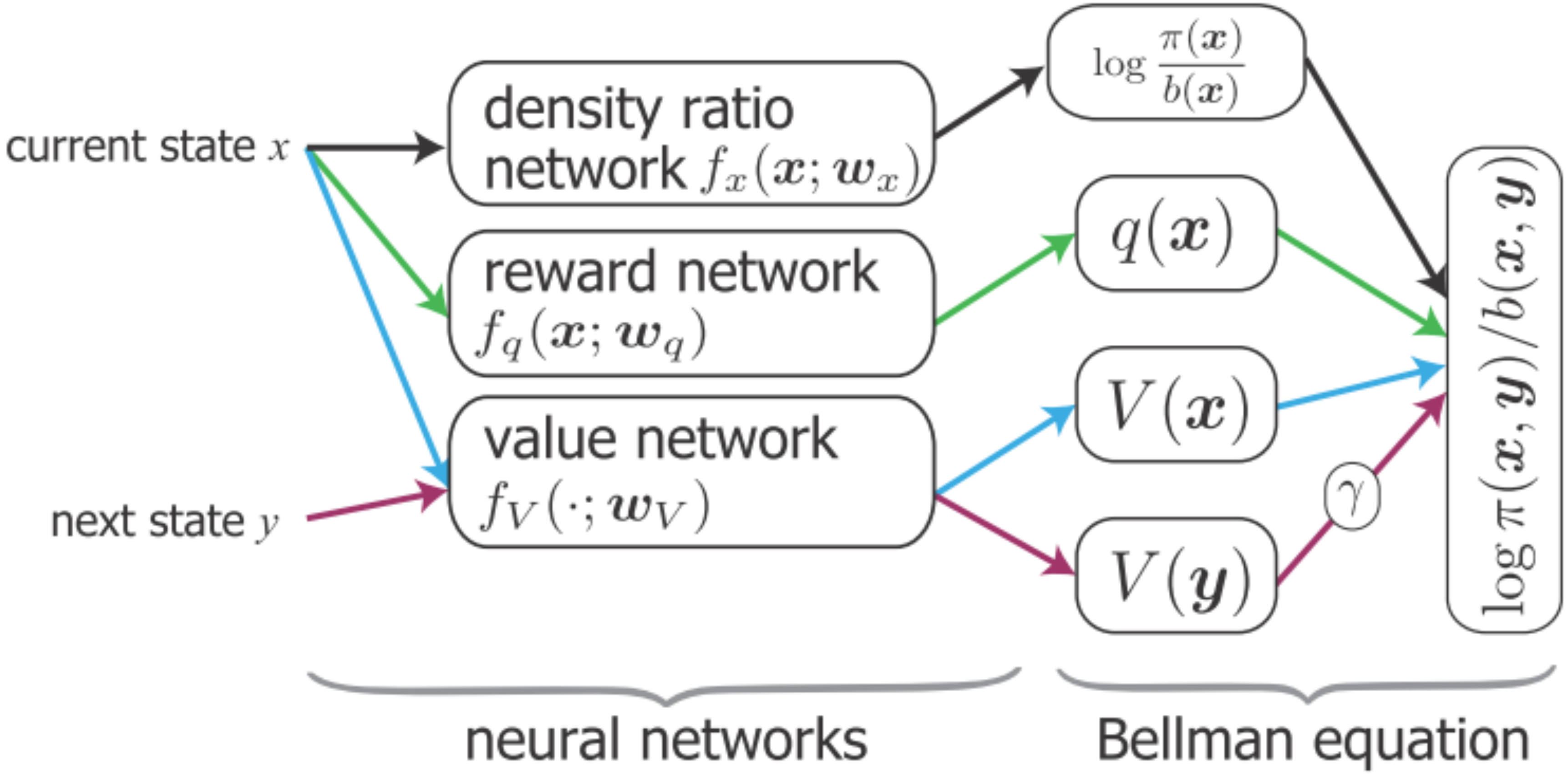
- state value function is estimated at the same time



Inverse RL by Density Ratio Estimation

(Uchibe & Doya, 2014, 2021)

- Based on KL control (Todorov 2009)
  - applicable to deep neural networks (Uchibe 2016)



Danger of Autonomous AI?

AI agents can be creative!

- Find new goals and try them out
- Create novel science, technology, culture, industry..

Needs assessment and control of dangers

- Runaway
- Side effect
- *Exploitation by individuals/groups with ambition/hatred*

Learning from the Human Society

- Humans are the most dangerous species on earth

Democracy: never give unlimited power to a person/group

- Politics
  - election
  - term limit
  - separation of powers
- Economy
  - antitrust law
  - right to strike
- Science
  - peer review

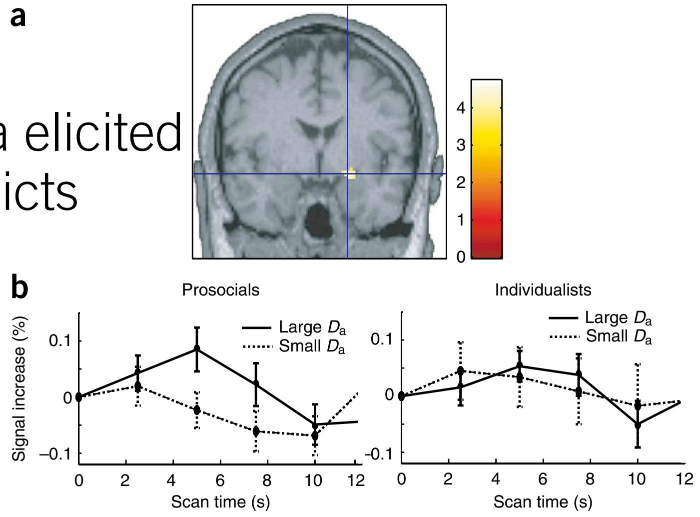
Peer reviewing among open-sourced, explainable AI agents

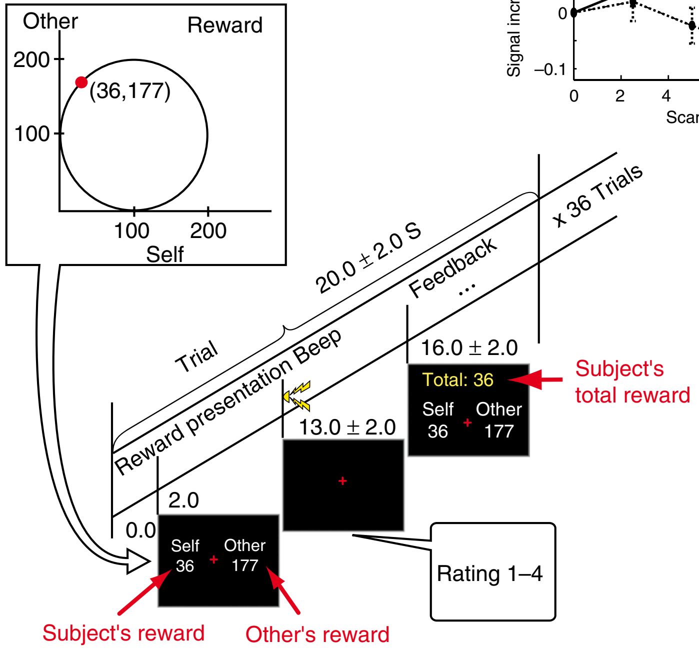
Social Value, Prefrontal Cortex and Amygdala

nature neuroscience

Activity in the amygdala elicited by unfair divisions predicts social value orientation

Masahiko Haruno[1,2] & Christopher D Frith[3,4]

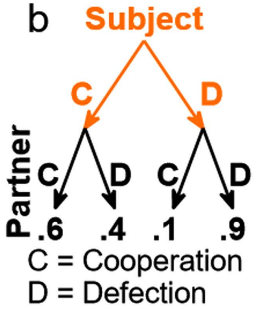


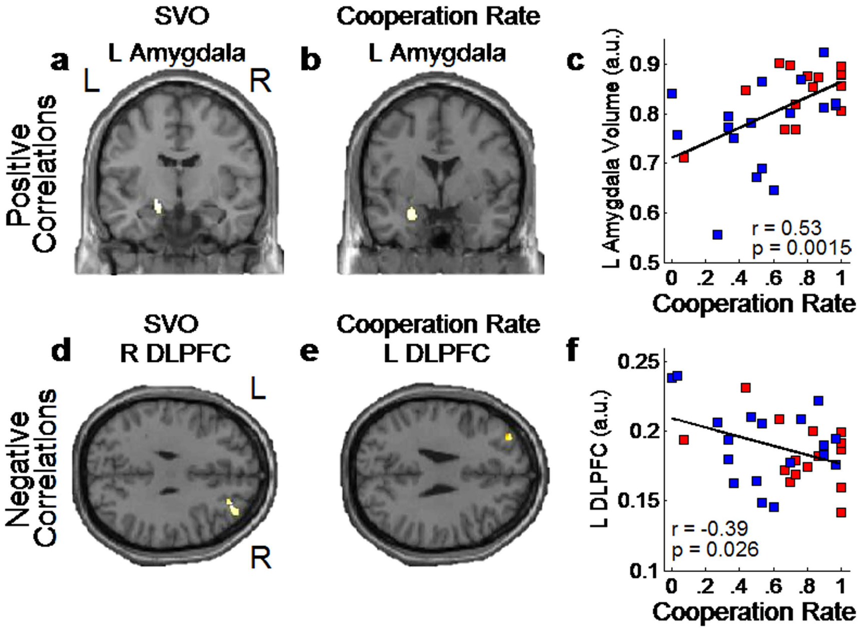


SCIENTIFIC REPORTS

Representation of economic preferences in the structure and function of the amygdala and prefrontal cortex

Alan S. R. Fermin[1], Masamichi Sakagami[1], Toko Kiyonari[2], Yang Li[1], Yoshie Matsumoto[1] & Toshio Yamagishi[3]

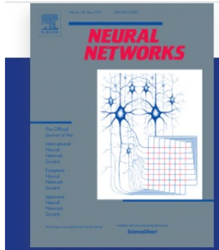




International Symposium on AI and Brain Science

ai.jp/symposium2020

ai.jp/symposium2022



Neural Networks

Supports *open access*

10.9 CiteScore | 8.050 Impact Factor

Articles & Issues ⌄ About ⌄ Publish ⌄ Order journal ⌄ Search in

Special issue on Artificial Intelligence and Brain Science

Edited by Karl Friston, Masashi Sugiyama, Kenji Doya, Josh Tenenbaum

Last update 24 February 2022

This special issue aims to capture recent advances in the crossing forefronts of AI and brai and to exchange ideas for creating brain-like intelligence and further advancing neuroscie

Actions for selected articles

Select all / Deselect all

Download PDFs

Export citations

Show all article previews

Receive an update when the latest issues in this jou

Sign in to set up alerts

Research article • *Open access*

World model learning and inference

Karl Friston, Rosalyn J. Moran, Yukie Nagai, Tadahiro Tanig

December 2021

Pages 573-590

Download PDF Article preview ⌄

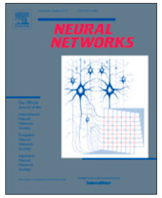
Neural Networks 152 (2022) 542–554

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

ELSEVIER



2021 Special Issue on AI and Brain Science: Brain-inspired AI

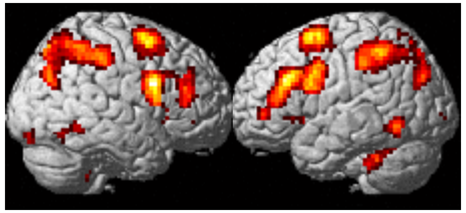
Social impact and governance of AI and neurotechnologies

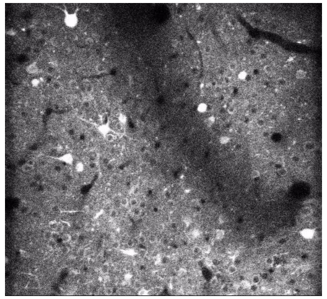
Check for updates

Kenji Doya [a,*], Arisa Ema [b], Hiroaki Kitano [a,c], Masamichi Sakagami [d], Stuart Russell [e]

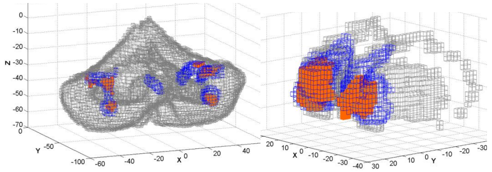
Advances in artificial intelligence (AI) and brain science are going to have a huge impact on society. While technologies based on those advances can provide enormous social benefits, adoption of new technologies poses various risks. This article first reviews the co-evolution of AI and brain science and the benefits of brain-inspired AI in sustainability, healthcare, and scientific discoveries. We then consider possible risks from those technologies, including intentional abuse, autonomous weapons, cognitive enhancement by brain–computer interfaces, insidious effects of social media, inequity, and enfeeblement. We also discuss practical ways to bring ethical principles into practice. One proposal is to stop giving explicit goals to AI agents and to enable them to keep learning human preferences. Another is to learn from democratic mechanisms that evolved in human society to avoid over-consolidation of power. Finally, we emphasize the importance of open discussions not only by experts, but also including a diverse array of lay opinions.

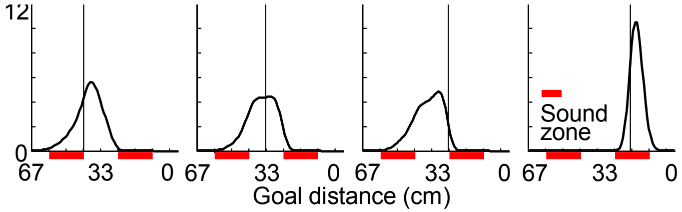
What Should We Further Learn from the Brain?





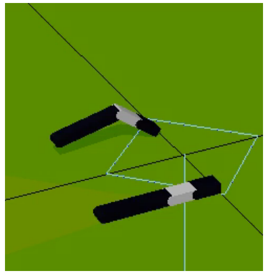
Energy Efficiency

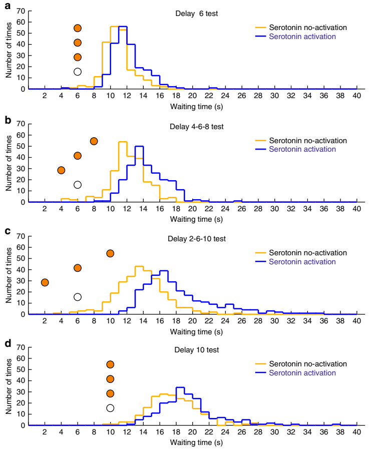


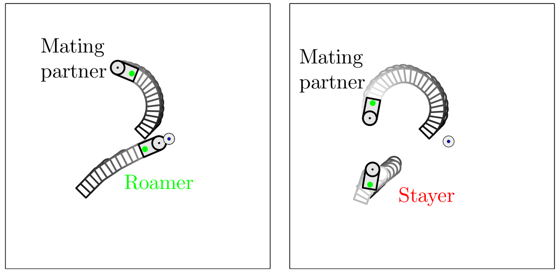


Data Efficiency

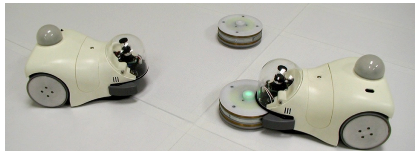
- World Models and Mental Simulation
- Modularity and Compositionality
- Meta-learning

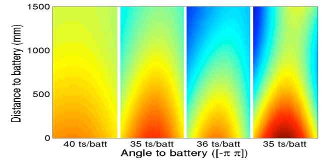






Autonomy and Sociality





Acknowledgements

- Striatum recording
  - **Makoto Ito (Progress Technology)**
  - **Tomohiko Yoshizawa (Tamagawa U)**
  - **Charles Gerfen (NIH)**
  - **Kazuyuki Samejima (Tamagawa U)**
  - **Minoru Kimura (Tamagawa U)**
- Human fMRI/behavior
  - **Alan Fermin (Tamagawa U)**
  - **Takehiko Yoshida (NAIST)**
  - **Saori Tanaka (ATR)**
  - **Nicolas Schweighofer (USC)**
  - Jun Yoshimoto (NAIST)
  - Yu Shimizu
  - Tomoki Tokuda (ATR)
  - Shoko Ota
- Serotonin recording/manipulation/modeling
  - **Kayoko W Miyazaki**
  - **Katsuhiko Miyazaki**
  - **Gaston Sivori**
  - **Masakazu Taira (UCLA)**
  - **Thomas Akam (Oxford U)**
  - **Mark Walton (Oxford U)**
  - **Kenji Tanaka (Keio U)**
  - **Akihiro Yamanaka (Nagoya U)**

Scientific Research on Innovative Areas
Strategic Research Program for Brain Sciences

- Cortical imaging
  - **Akihiro Funamizu (U Tokyo)**
  - **Bernd Kuhn**
  - **Yuzhe Li**
  - **Sergey Zobnin**
  - Naohiro Yamauchi
- Marmoset data analysis
  - Carlos Gutierrez (Softbank)
  - Hiromichi Tsukada (Chubu U)
  - Junichi Hata, Henrik Skibbe, Alex Woodward (RIKEN)
  - Ken Nakae, (NINS)
- Basal ganglia model
  - Benoit Girard, Daphne Heraiz (Sorbonne)
  - Jean Lienard
  - Shuhei Hara
- Robotics
  - **Jun Morimoto (ATR)**
  - **Eiji Uchibe (ATR)**
  - **Stefan Elfwing (ATR)**
  - **Jiexin Wang (ATR)**
  - **Paavo Parmas (Kyoto U)**
  - Kristine Roque
  - **Yuji Kanagawa**
  - **Tojoarisoa Rakotoaritina**
  - **Christopher Buckley**

Brain/MINDS Project
Fugaku Supercomputing Program