Toward the Neural Basis of Happiness and Motivation

HKUST 2015.4.13

Sen Song
Department of Biomedical Engineering, School of Medicine
Center for Brain-Inspired Computing Research
McGovern Institute for Brain Research
Tsinghua University

Towards the Neural Basis of Happiness

- Recent interest in Positive Psychology
  - Seligman Csikszentmihalyi 1998:

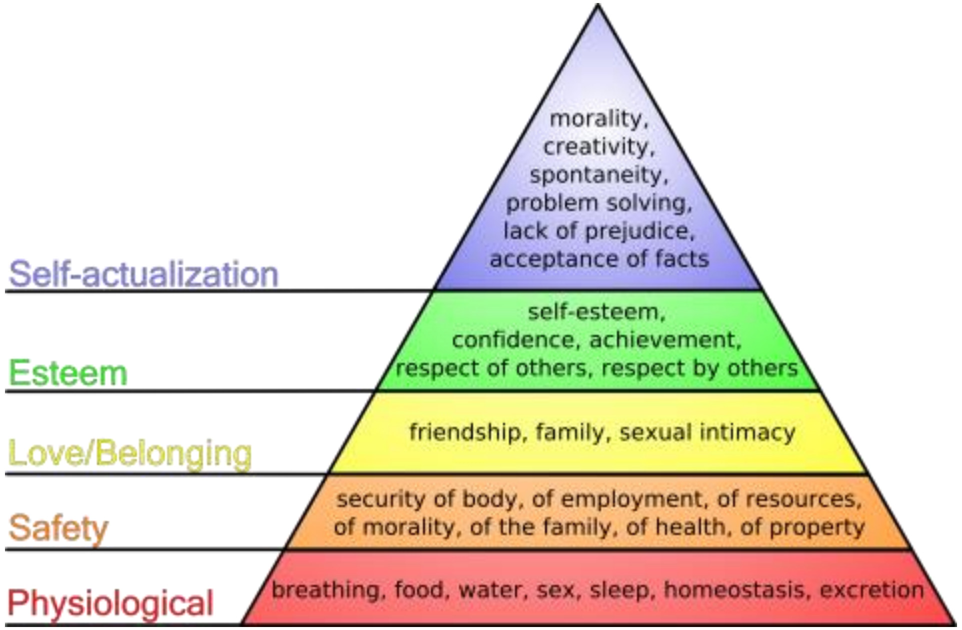
We believe that a psychology of positive human functioning will arise, which achieves a scientific understanding and effective interventions to build thriving individuals, families, and communities.

Positive psychologists seek "to find and nurture genius and talent" and "to make normal life more fulfilling", rather than merely treating mental illness.

Positive psychologists are concerned with four topics: (1) positive experiences, (2) enduring psychological traits, (3) positive relationships and (4) positive institutions.

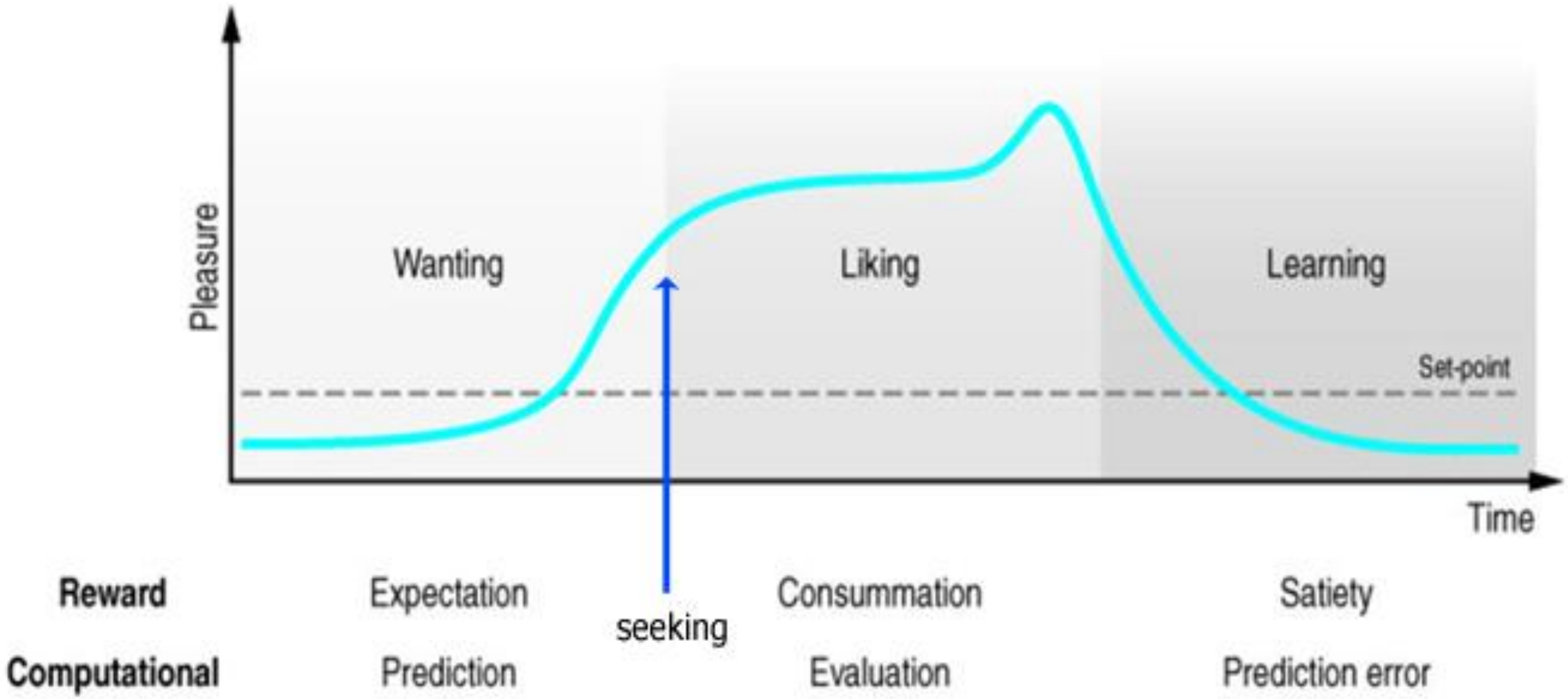
Different Kinds of Happiness

- Hedonic Happiness
  - Positive emotional functioning
  - Lower Pleasure
    - Food
    - Mouse model to study circuitry
  - Higher Pleasure
    - Music/collaboration with Wang, Xiaoqin
    - fMRI and possibly mouse model

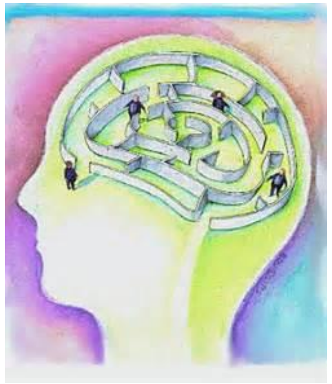


- Eudaimonic Happiness
  - Aristotle: Virtuous activity of the soul
  - Modern: conscious and life-long active exercise of intellect and character virtues
  - Related to self-actualization
  - Recent twin study implies a genetic basis
  - Can be studied with MRI

Building a neuroscience of pleasure



Establishing the Behavioral Platforms

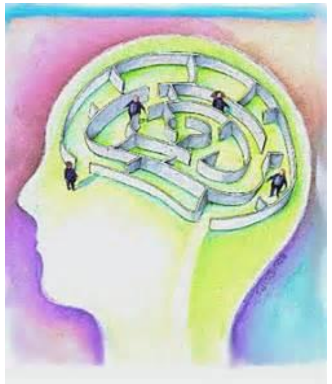


Craving/wanting → Seeking → Liking → Satisfaction → Learning

Recent interest
in decision making

Most People Study These

Establishing the Behavioral Platforms

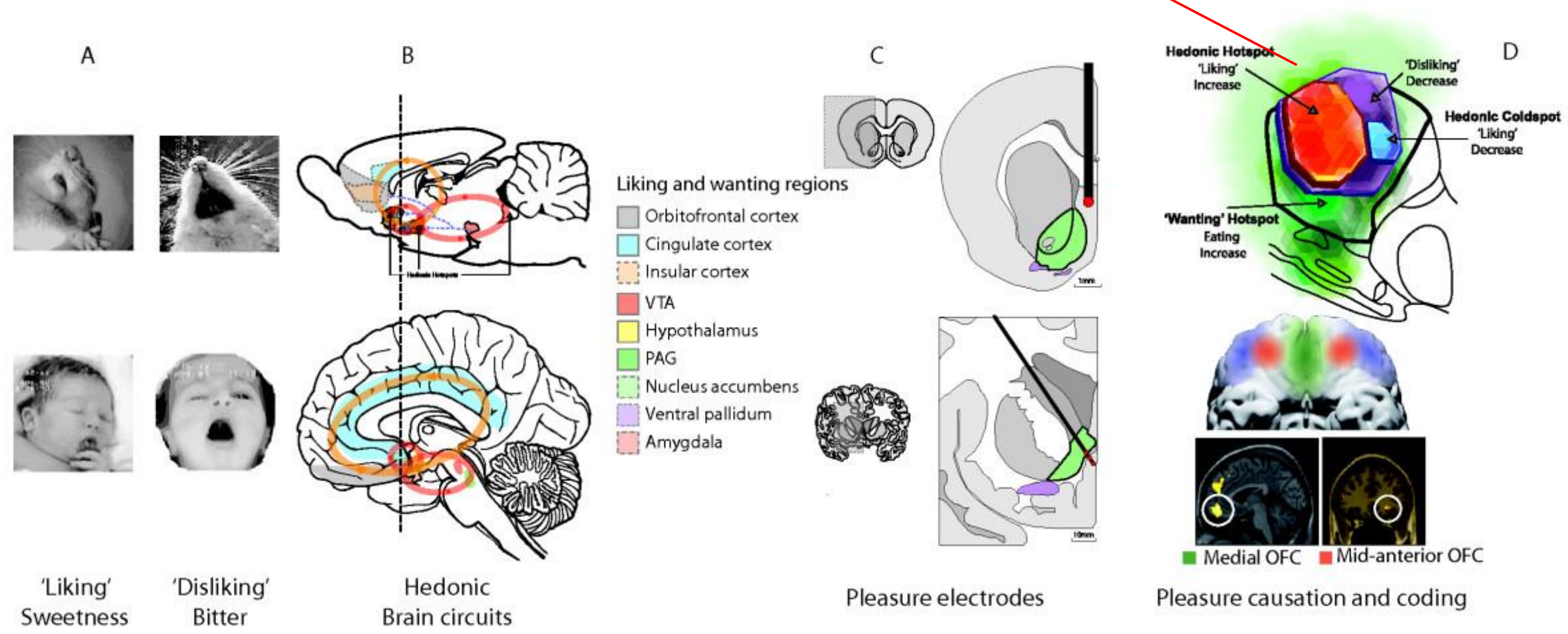


Craving/wanting → Seeking → Liking → Satisfaction → Learning

Dissecting the circuits of pleasure

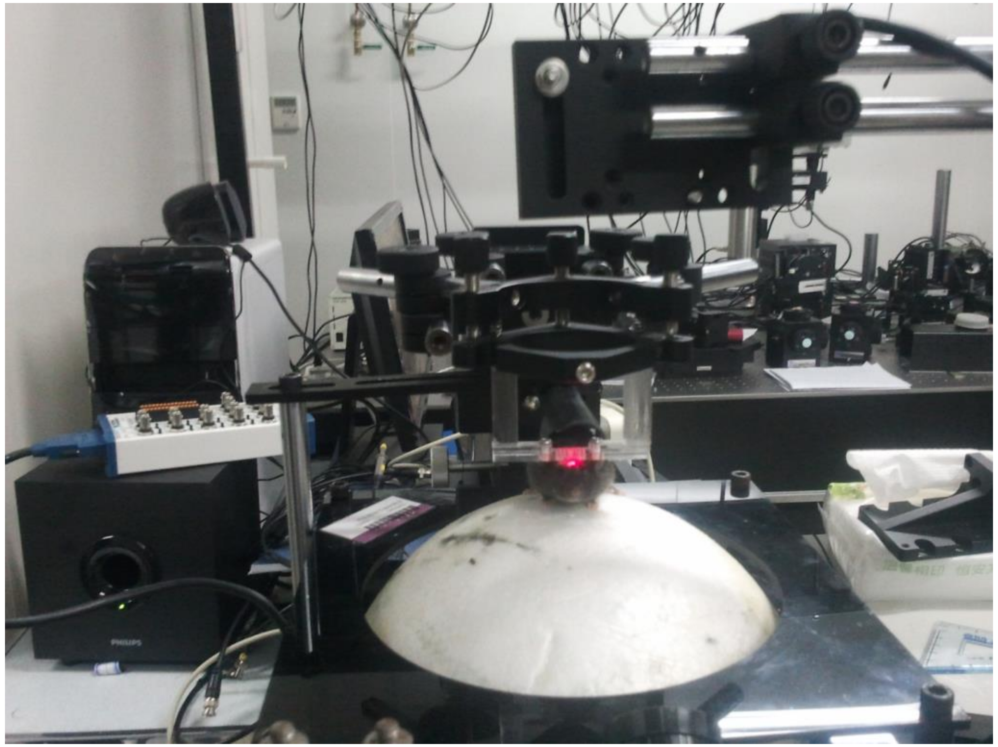
——Taste Reactivity paradigm on the floating ball system

Enhancement by opioid; No enhancement with dopamine

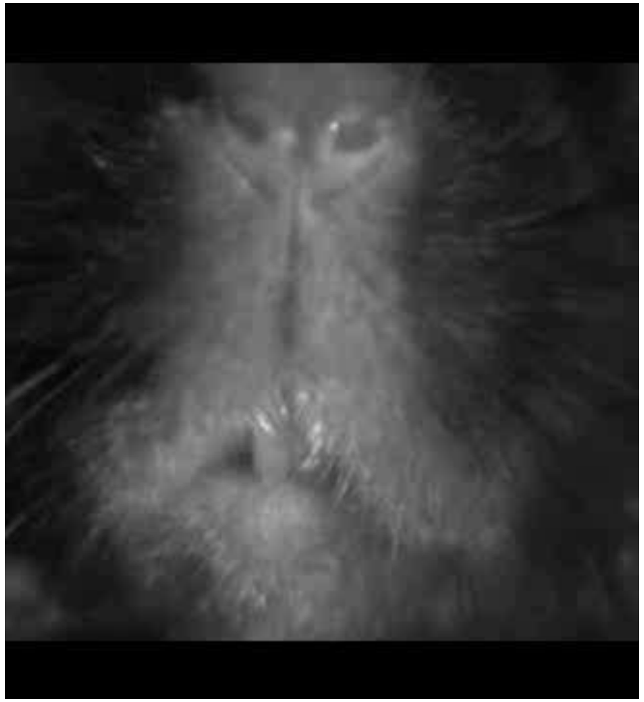


What about dopamine?

Set-up



Facial expressions induced by taste stimuli



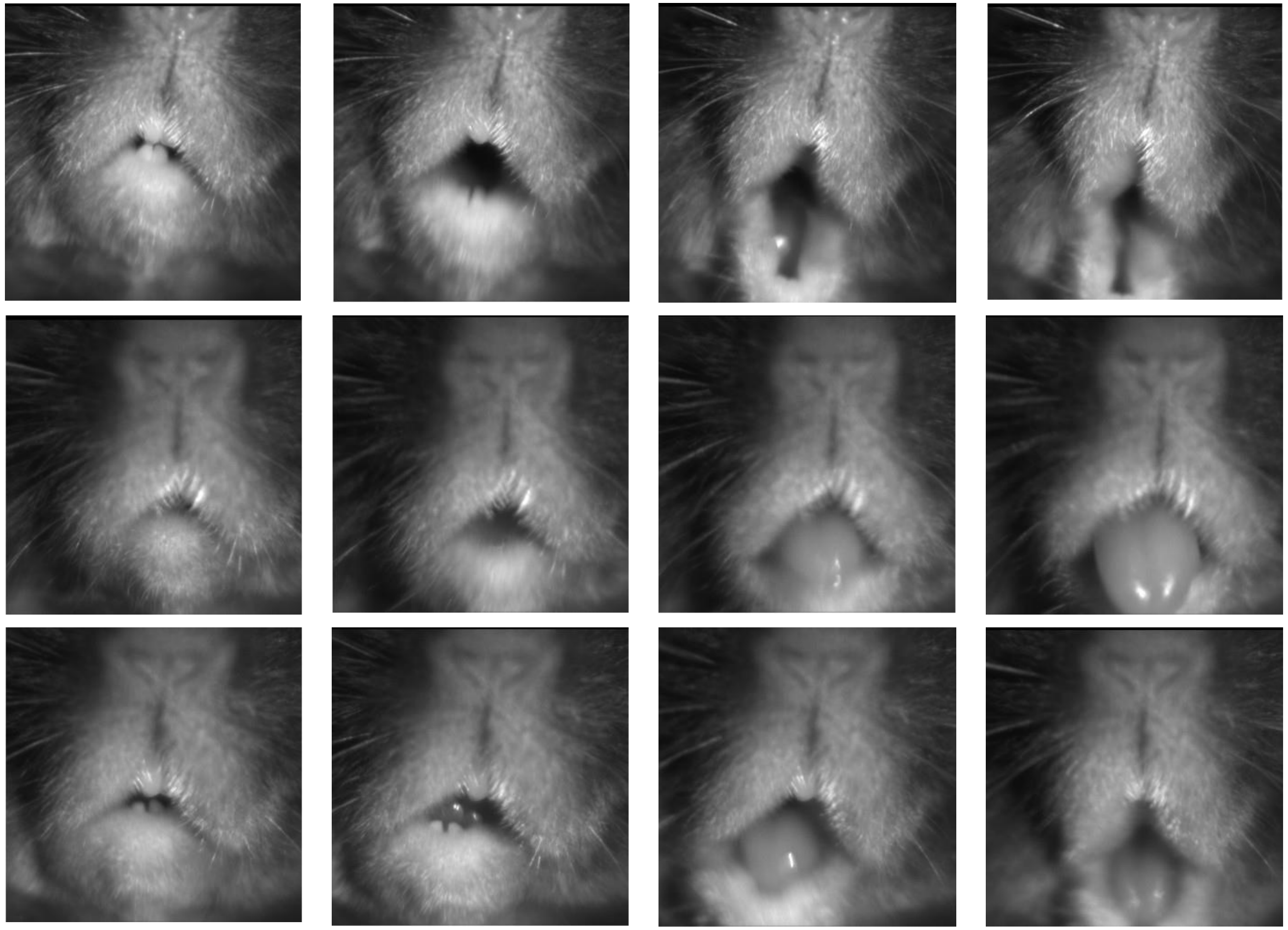
sugar

quinine

Gape

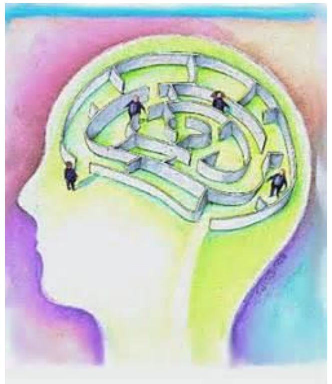
Tongue protrusion

Tongue movement



VP vglut2 Neurons are aversive

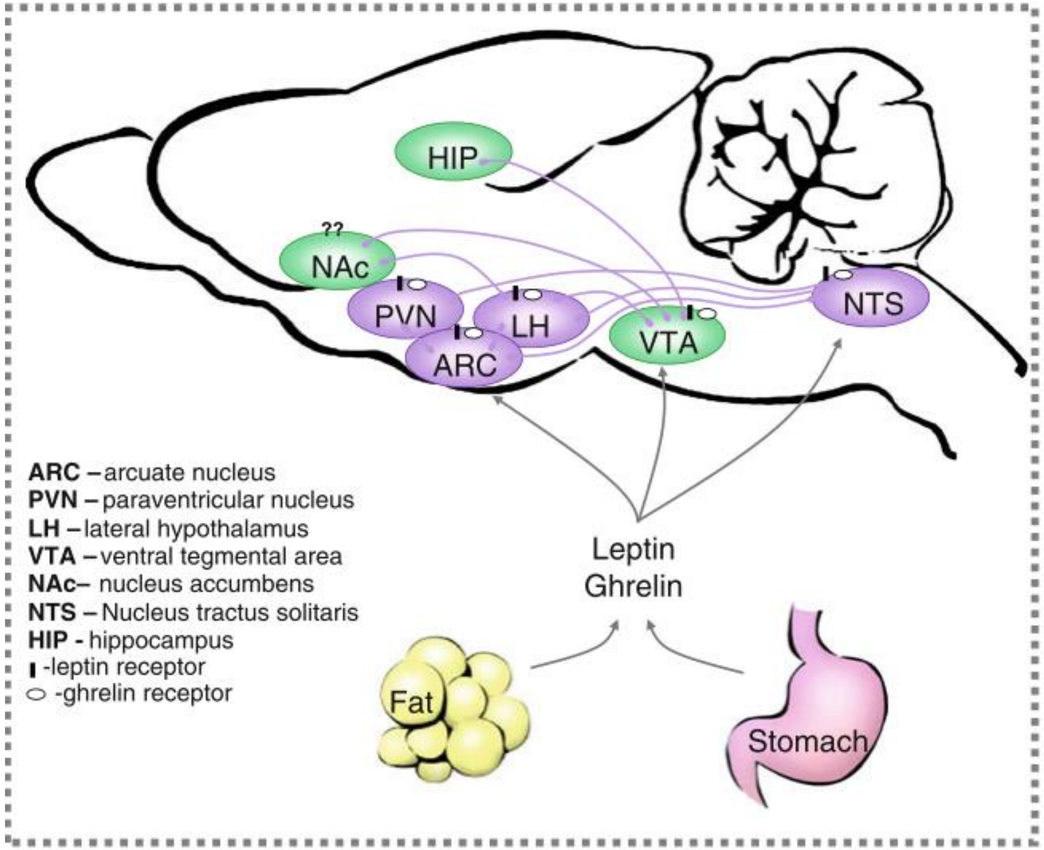
Establishing the Behavioral Platforms



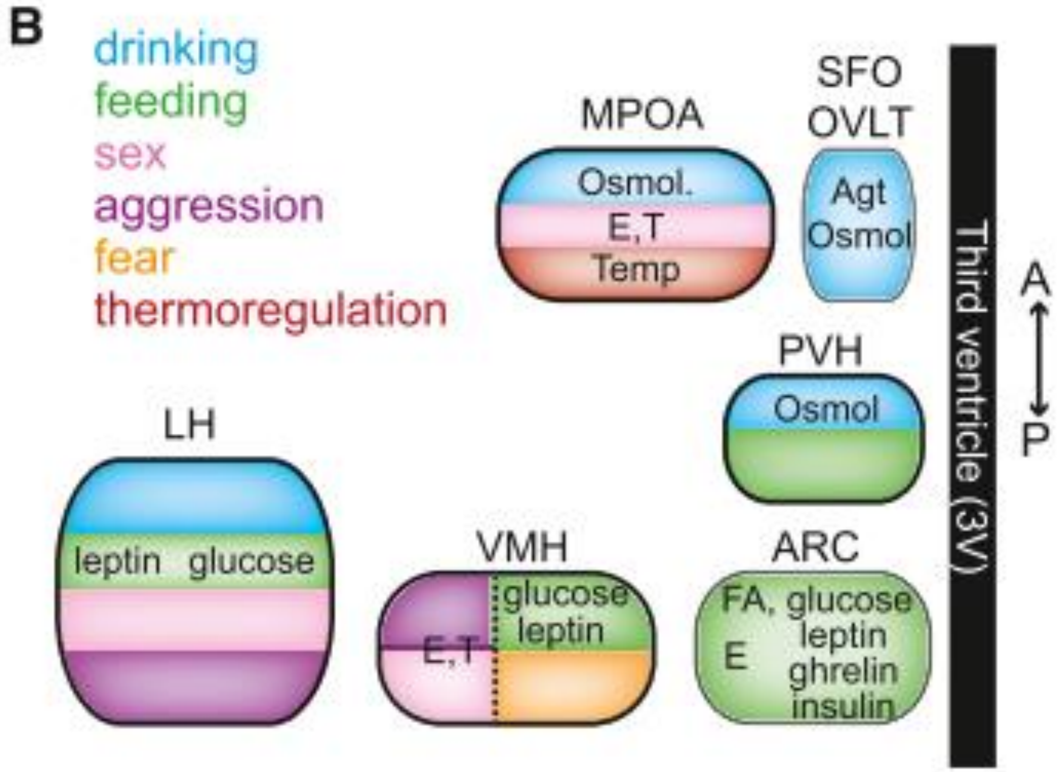
Craving/wanting → Seeking → Liking → Satisfaction → Learning

Motivation

Background of anatomy and functions of LH

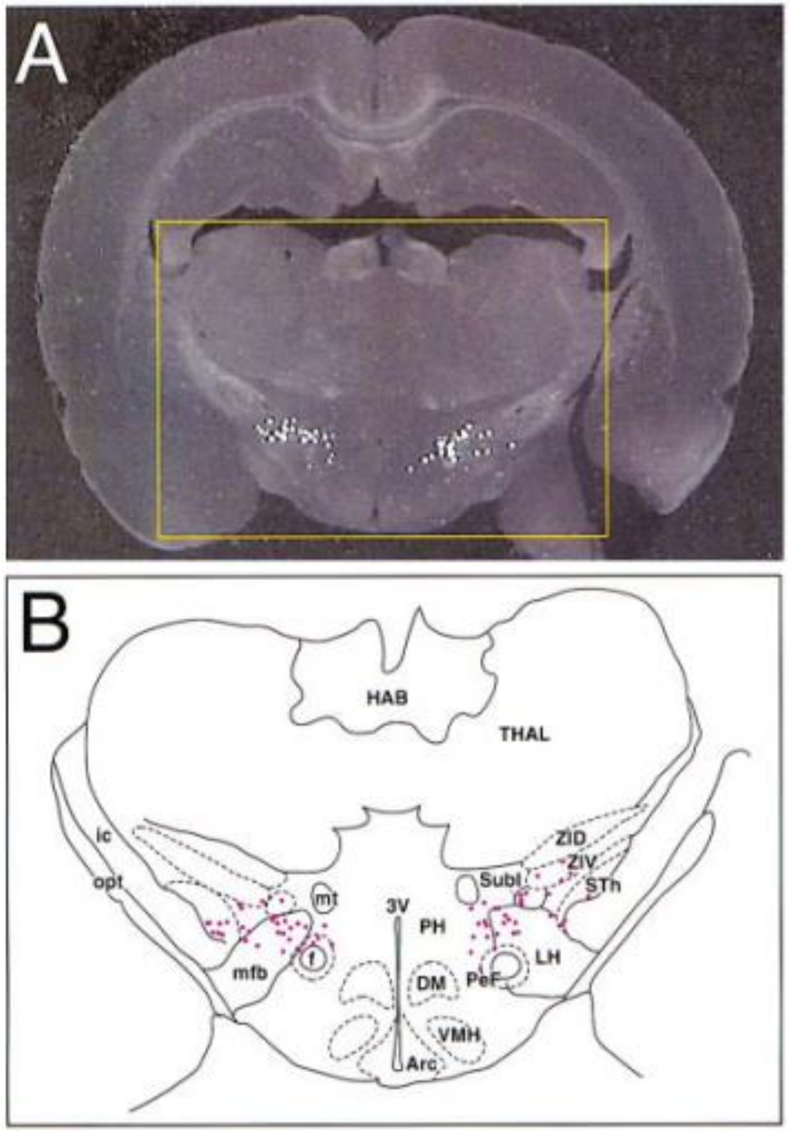


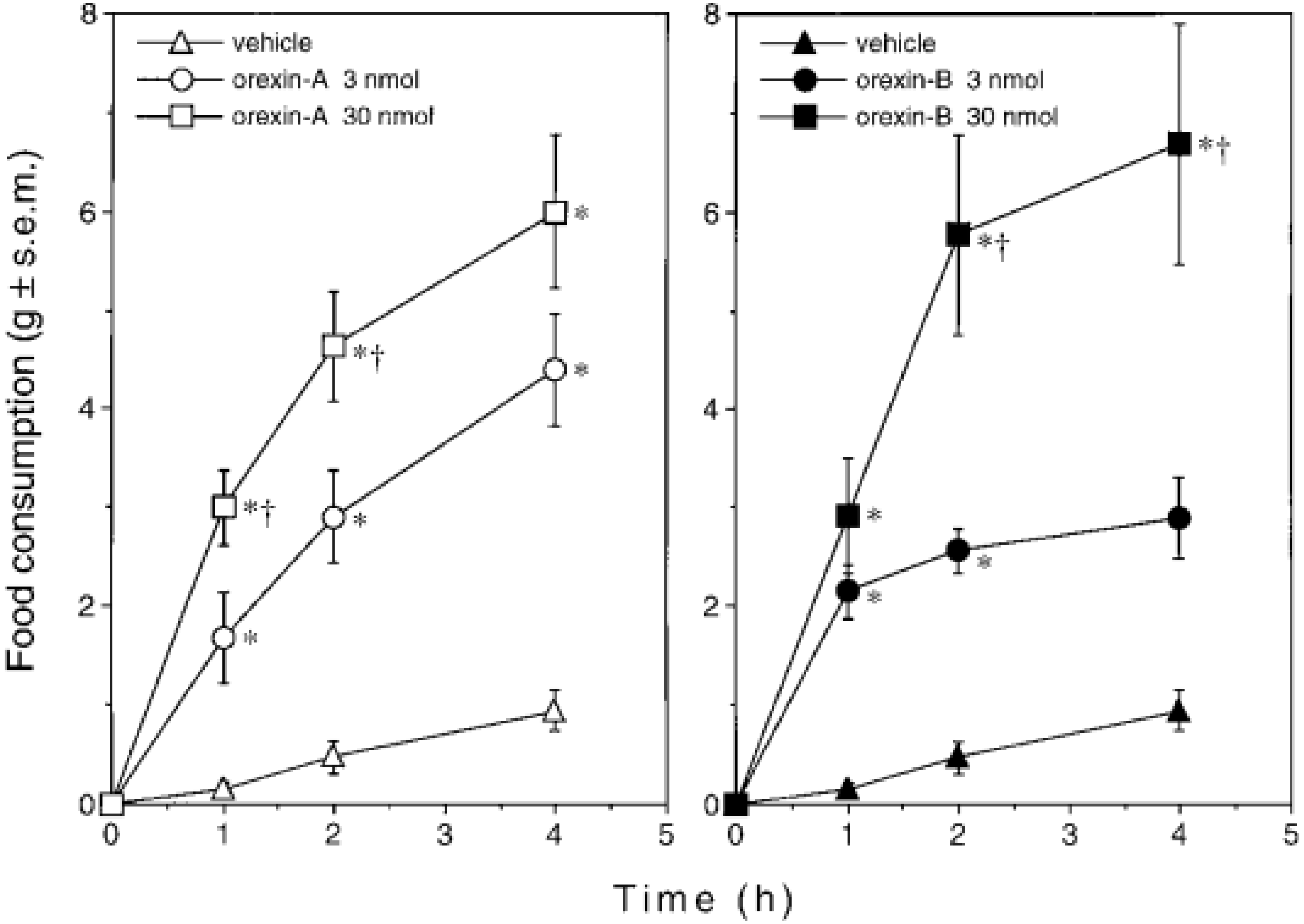
(Xiaoye Shan & Giles S. H. Yeo,
Rev Endocr Metab Disord (2011) 12:197 – 209)



(Scotte M.Sternson, Neuron (2013)

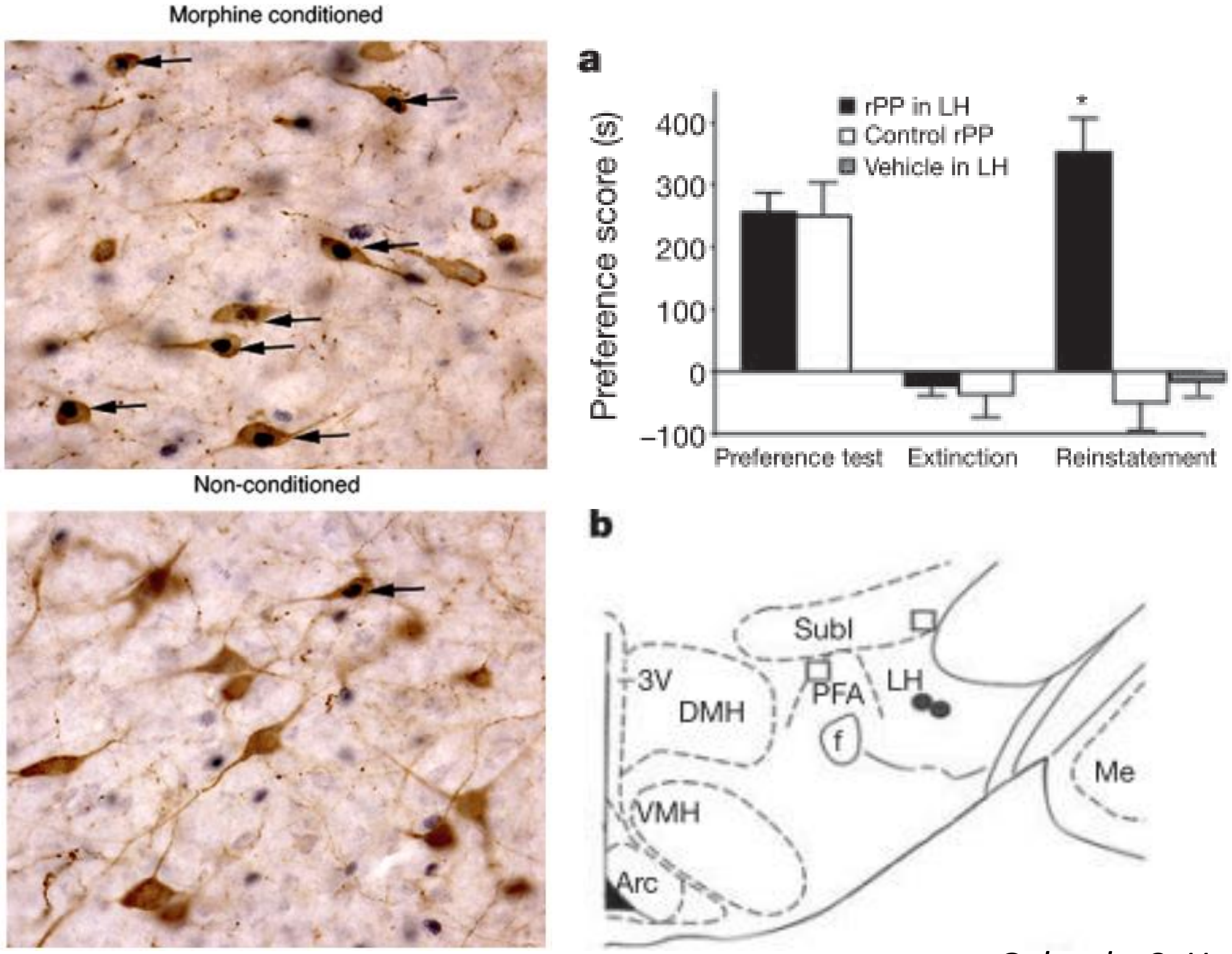
LH regulates feeding behavior





Takeshi Sakurai,et al,Cell, Vol. 92, 573–585, February 20, 1998

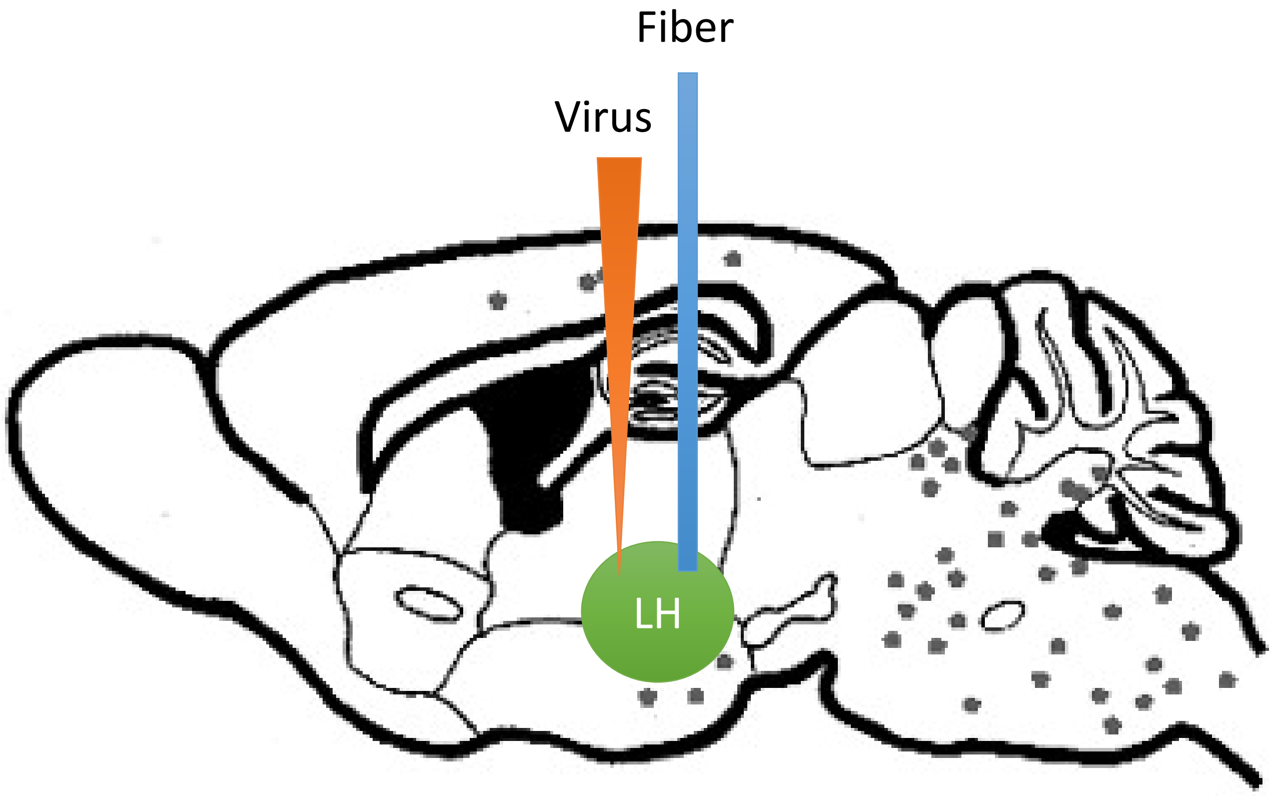
LH is involved in reward processing



Gelanda C. Harris, science, 2005,vol 437,22

Result: LH and reward seeking behavior

Photoactivation of LH neurons



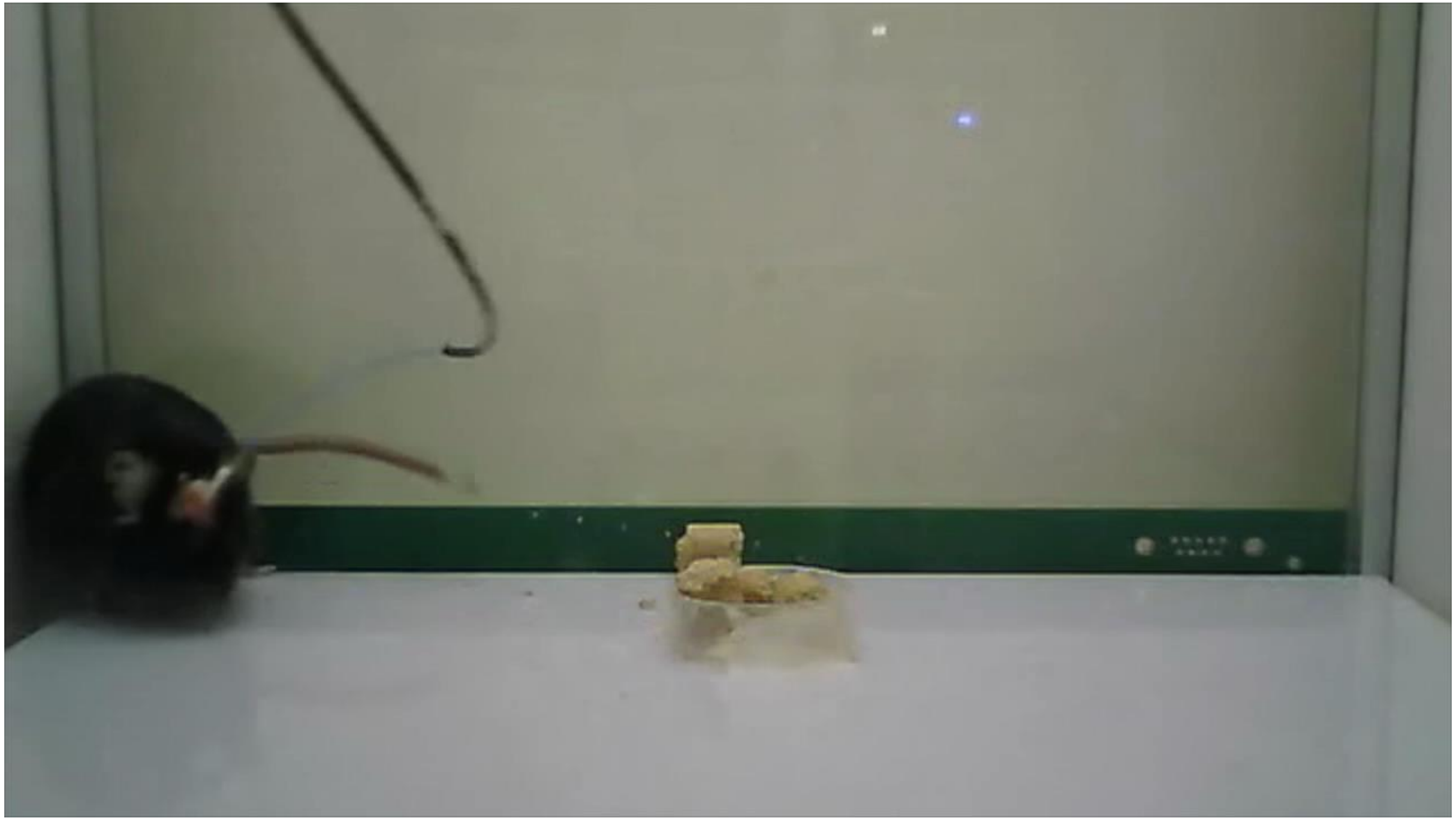
Result

Photoactivation of LH neurons induces feeding

Cheese, 30Hz, 90%

Results: LH and reward seeking behavior

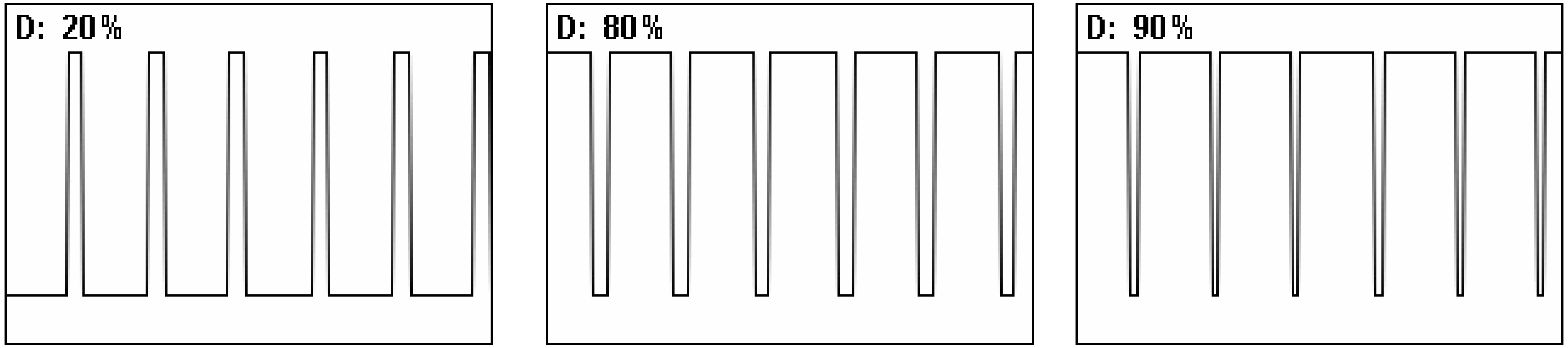
Feeding is accompanied with seeking behavior in photoactivation



Standard chow, 20Hz, 90%

Results: LH and reward seeking behavior

Lower duty cycle preferably induce foraging than feeding



Lower duty cycles preferably induce seeking behavior

Standard chow, 20Hz, 20%

Results:LH and feeding or ingestive behavior

Background

LH is involved with the regulation of food initiation.

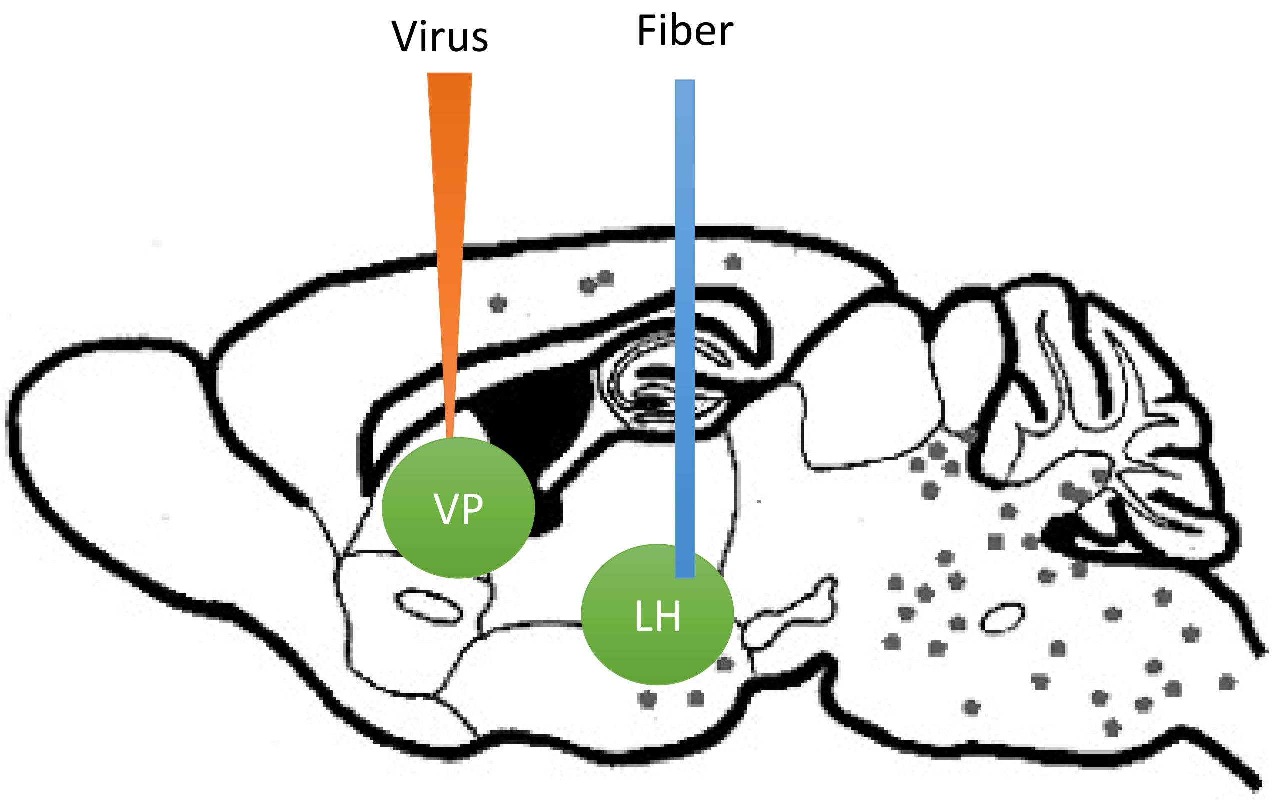
"areas in the hindbrain mediate the increase in consummatory (meal size) and the hypothalamus and other forebrain sites mediate the appetitive (meal frequency) components of orexin-induced hyperphagia."

-----Berthoud, H.-R., and Münzberg, H. (2011). The lateral hypothalamus as integrator of metabolic and environmental needs: from electrical self-stimulation to opto-genetics. *Physiol. Behav*. 104, 29–39.

Part 3 LH and feeding or ingestive behavior

Result

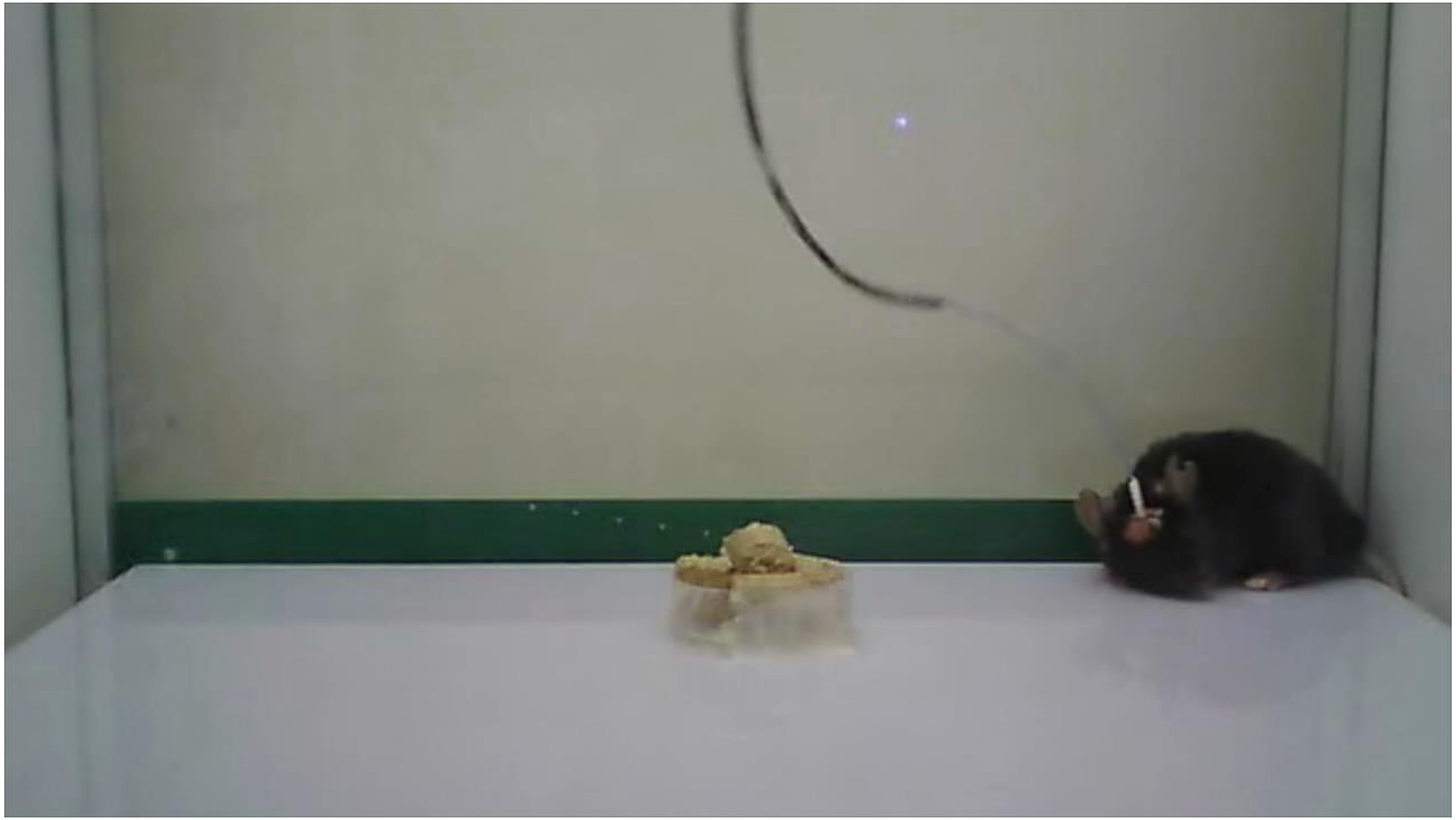
Photoactivation of LH neurons



Part 3 LH and feeding or ingestive behavior

Result

Photoactivation of LH-projection axons from VP induce abnormal “eating”, and no seeking behavior



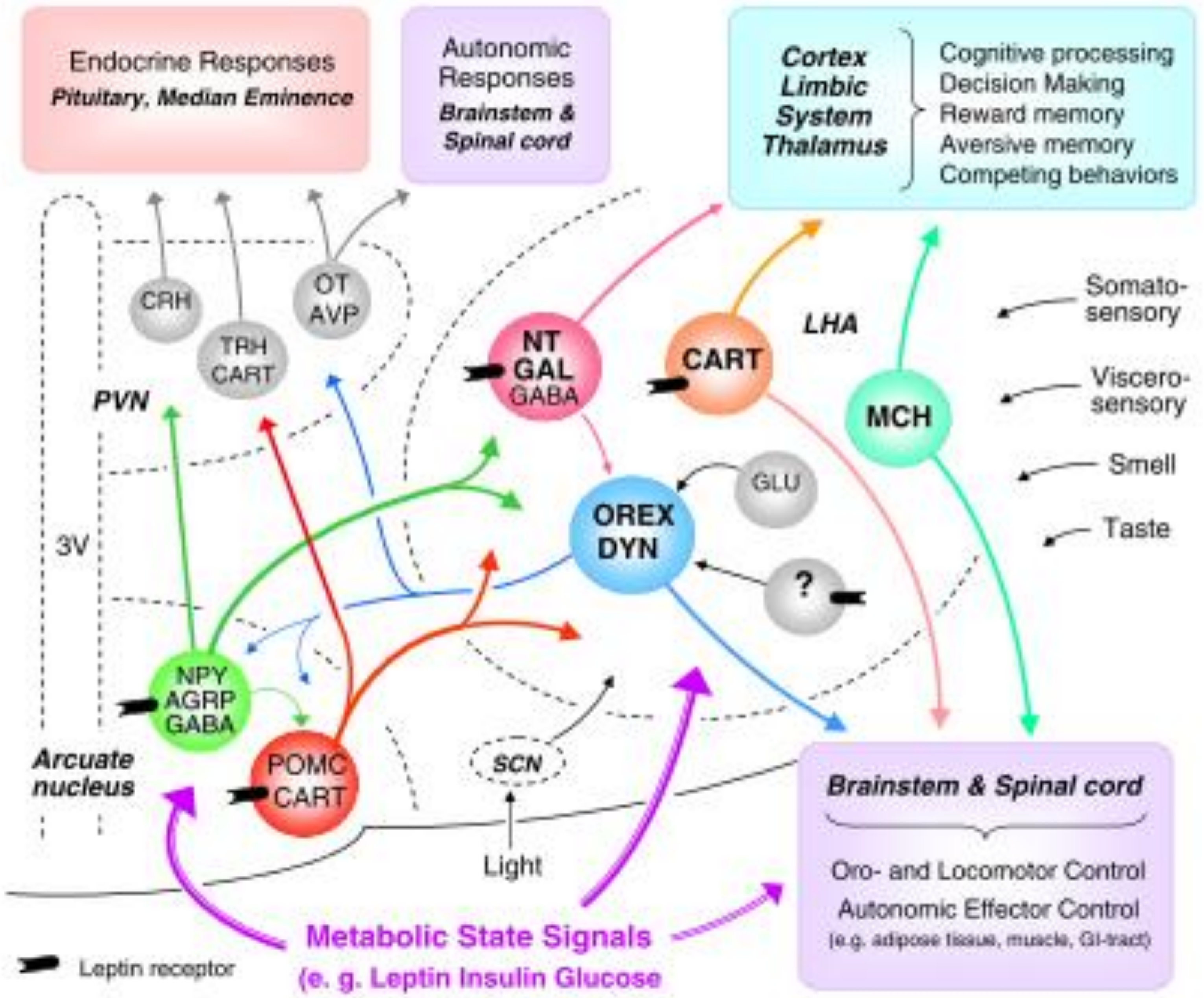
Standard chow, 20Hz, 90%

Part 3 LH and feeding or ingestive behavior

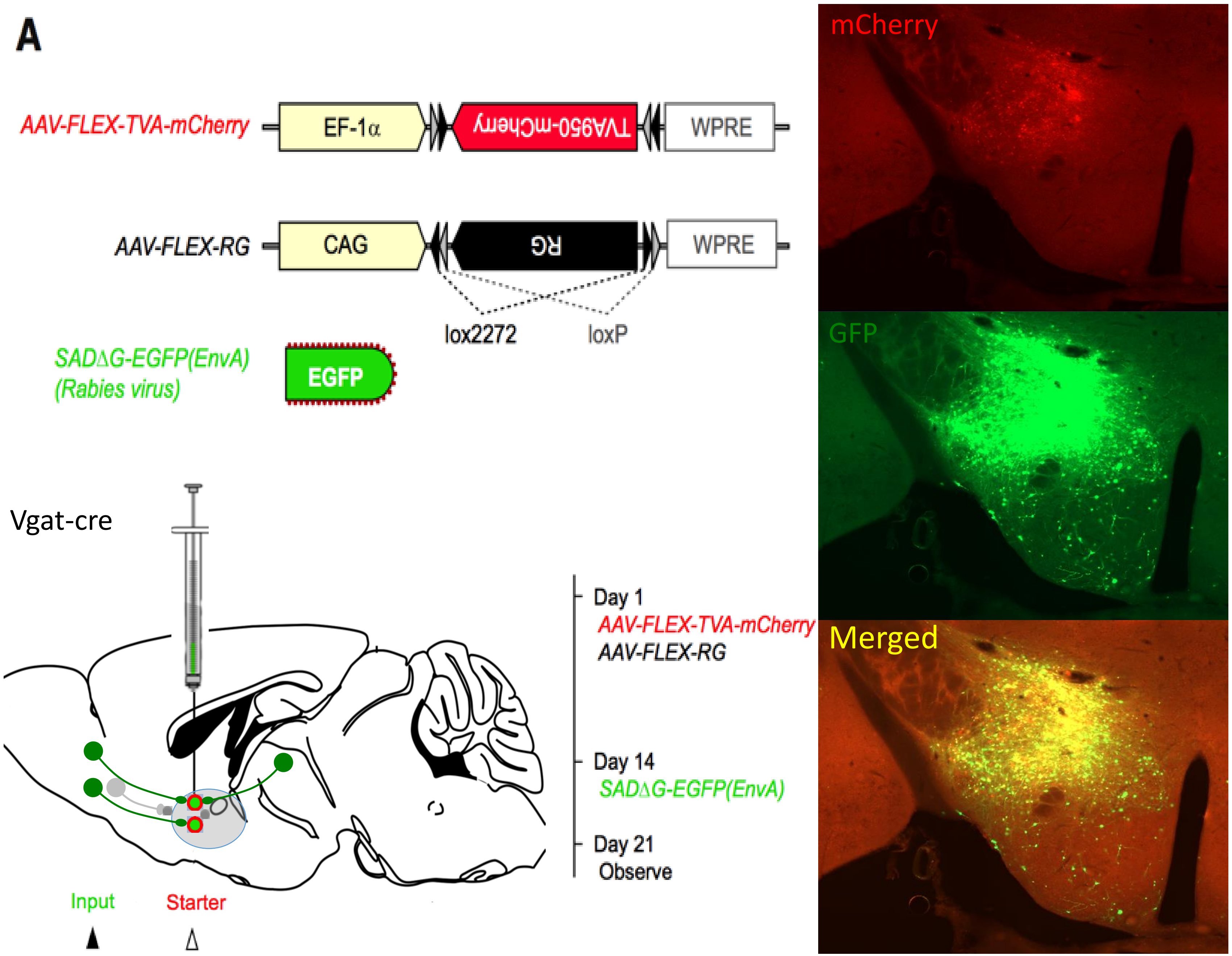
Result

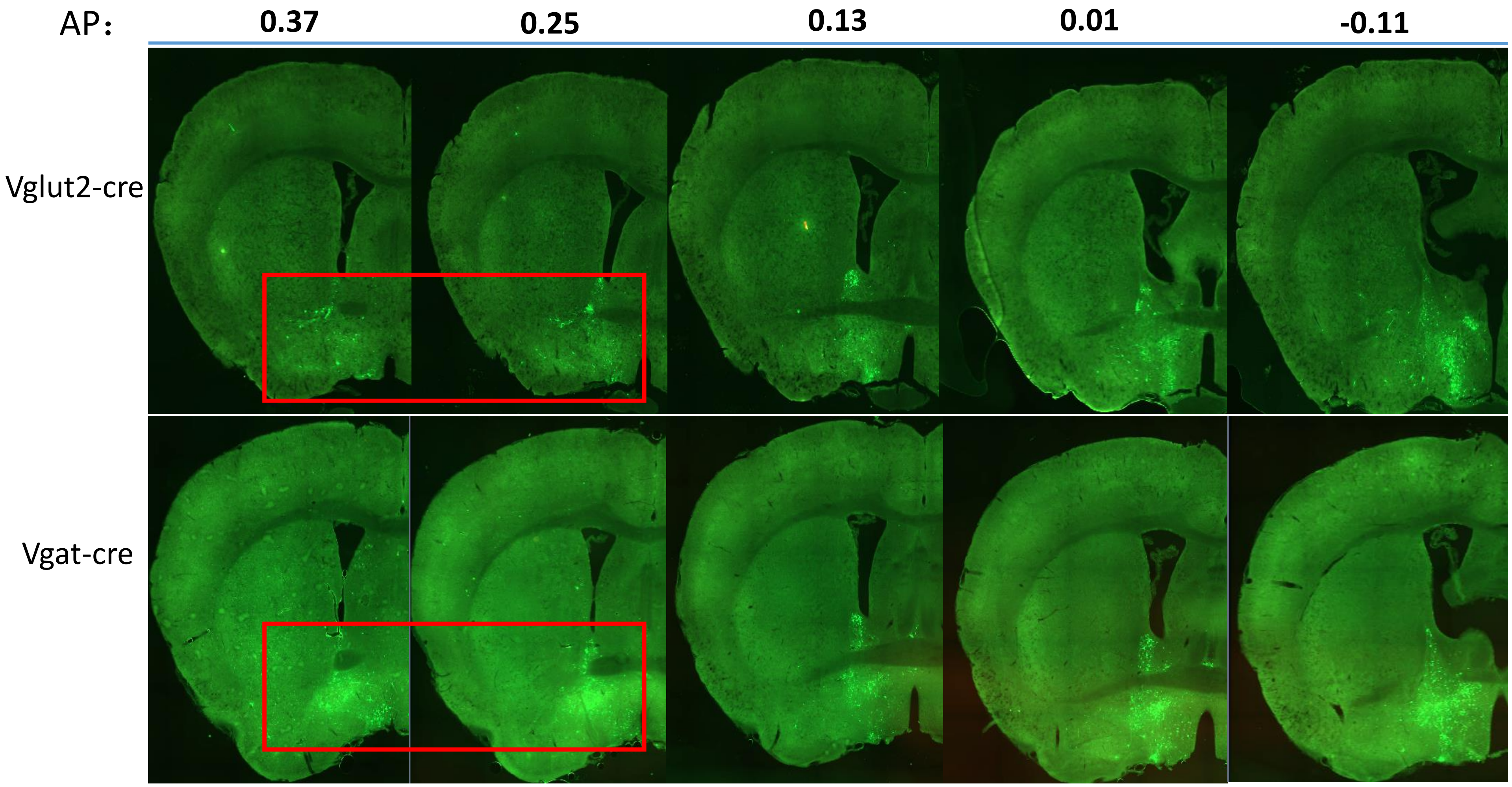
Photoactivation of LH-projection axons from VP induce feeding

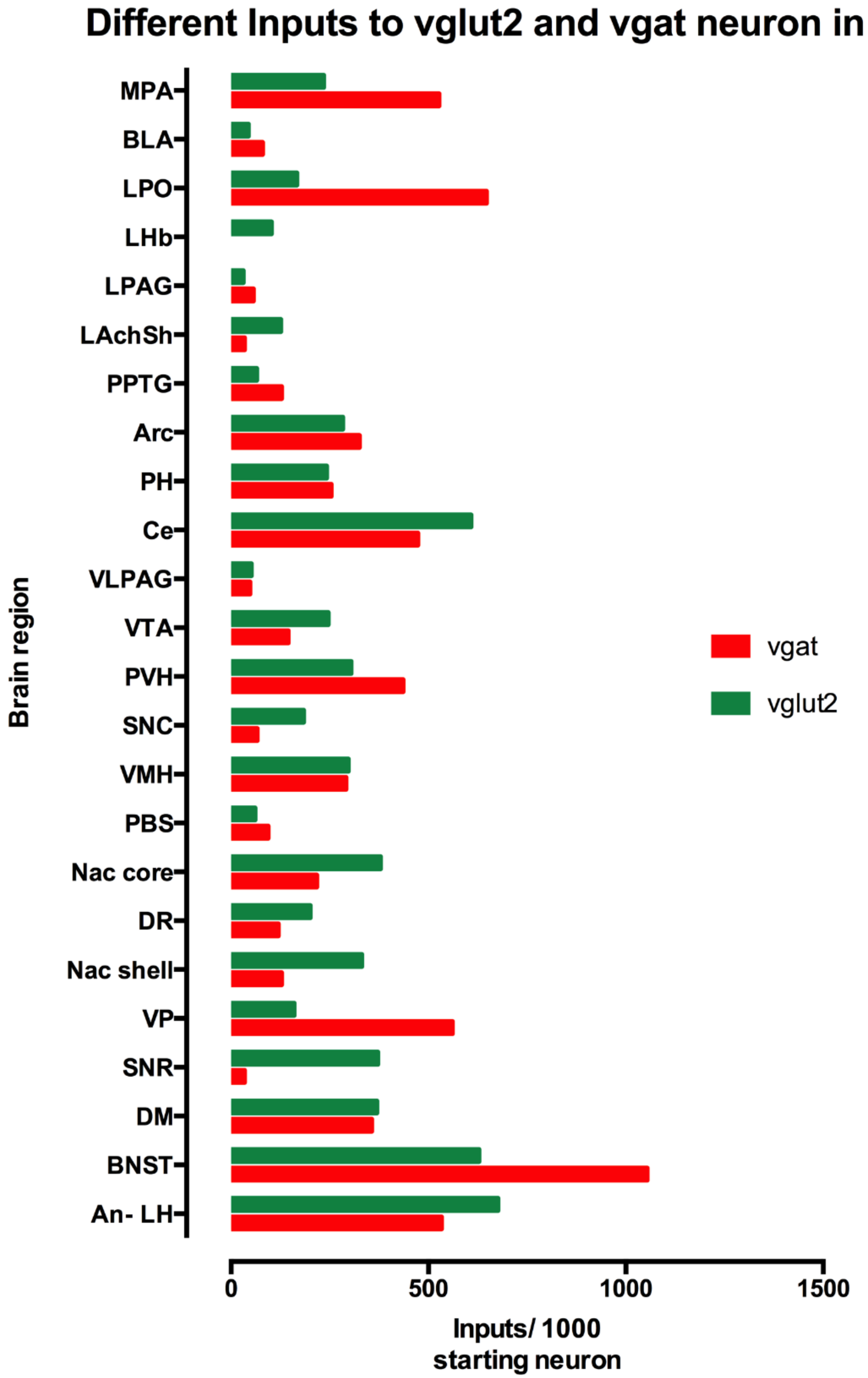
Standard chow, 20Hz, 90%



Hans-Rudi Berthoud, Heike Münzberg,Physiology & Behavior 104 (2011) 29–39

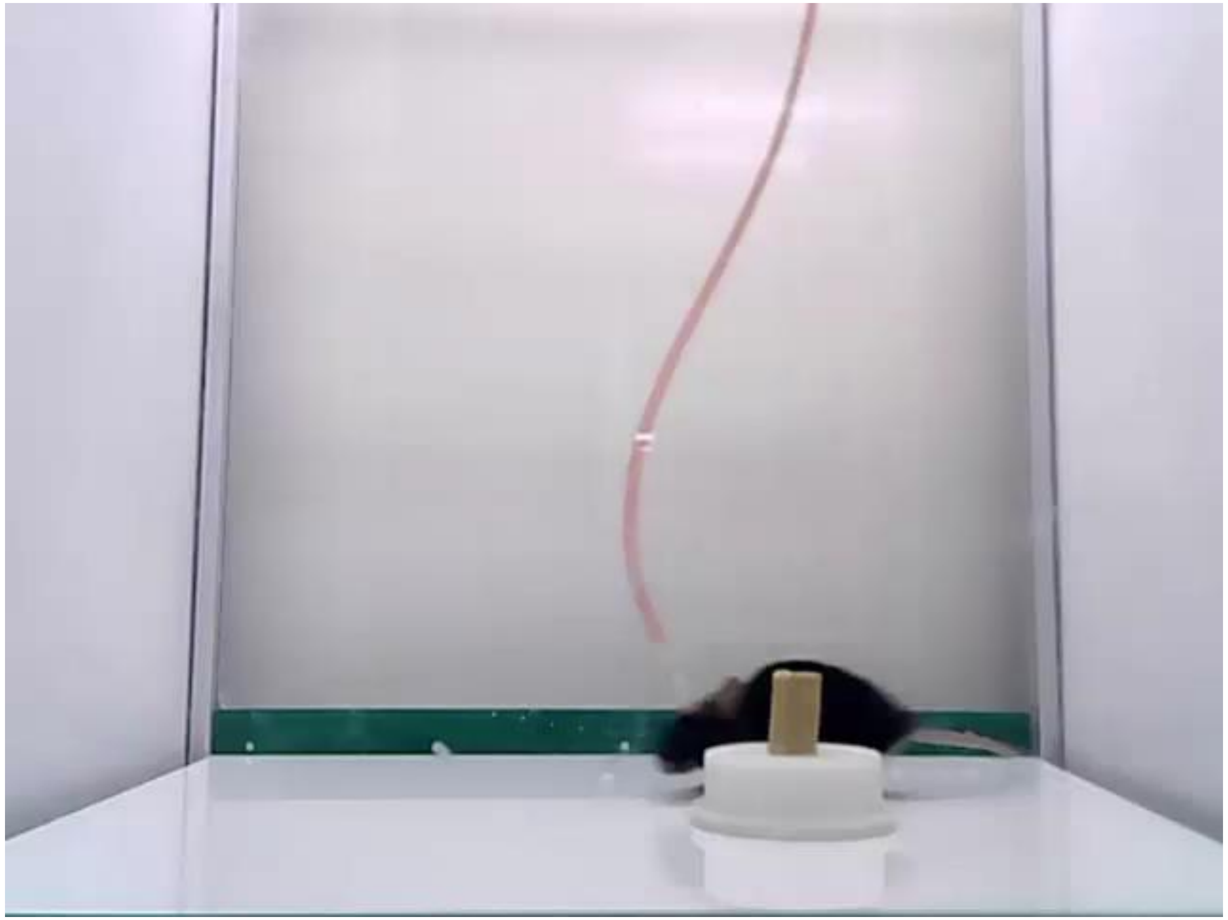

A
AAV-FLEX-TVA-mCherry
EF-1α
TVA950-mCherry
WPRE
AAV-FLEX-RG
CAG
RG
WPRE
lox2272
loxP
SADΔG-EGFP(EnvA)
(Rabies virus)
EGFP
Vgat-cre
Day 1
AAV-FLEX-TVA-mCherry
AAV-FLEX-RG
Day 14
SADΔG-EGFP(EnvA)
Day 21
Observe
Input
Starter
mCherry
GFP
Merged


AP：
0.37
0.25
0.13
0.01
-0.11
Vglut2-cre
Vgat-cre



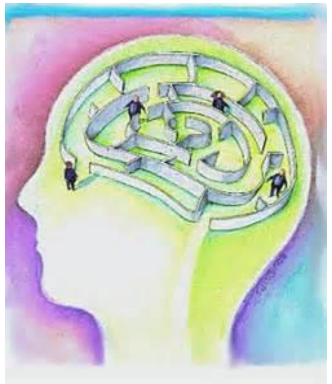
Different Inputs to vglut2 and vgat neuron in LH
Brain region
MPA
BLA
LPO
LHb
LPAG
LAchSh
PPTG
Arc
PH
Ce
VLPAG
VTA
PVH
SNC
VMH
PBS
Nac core
DR
Nac shell
VP
SNR
DM
BNST
An- LH
0
500
1000
1500
Inputs/ 1000
starting neuron
vgat
vglut2

Photoactivation of PAG-projection GABAergic axons from LH induce feeding



Standard chow, 20Hz, 90%

Establishing the Behavioral Platforms

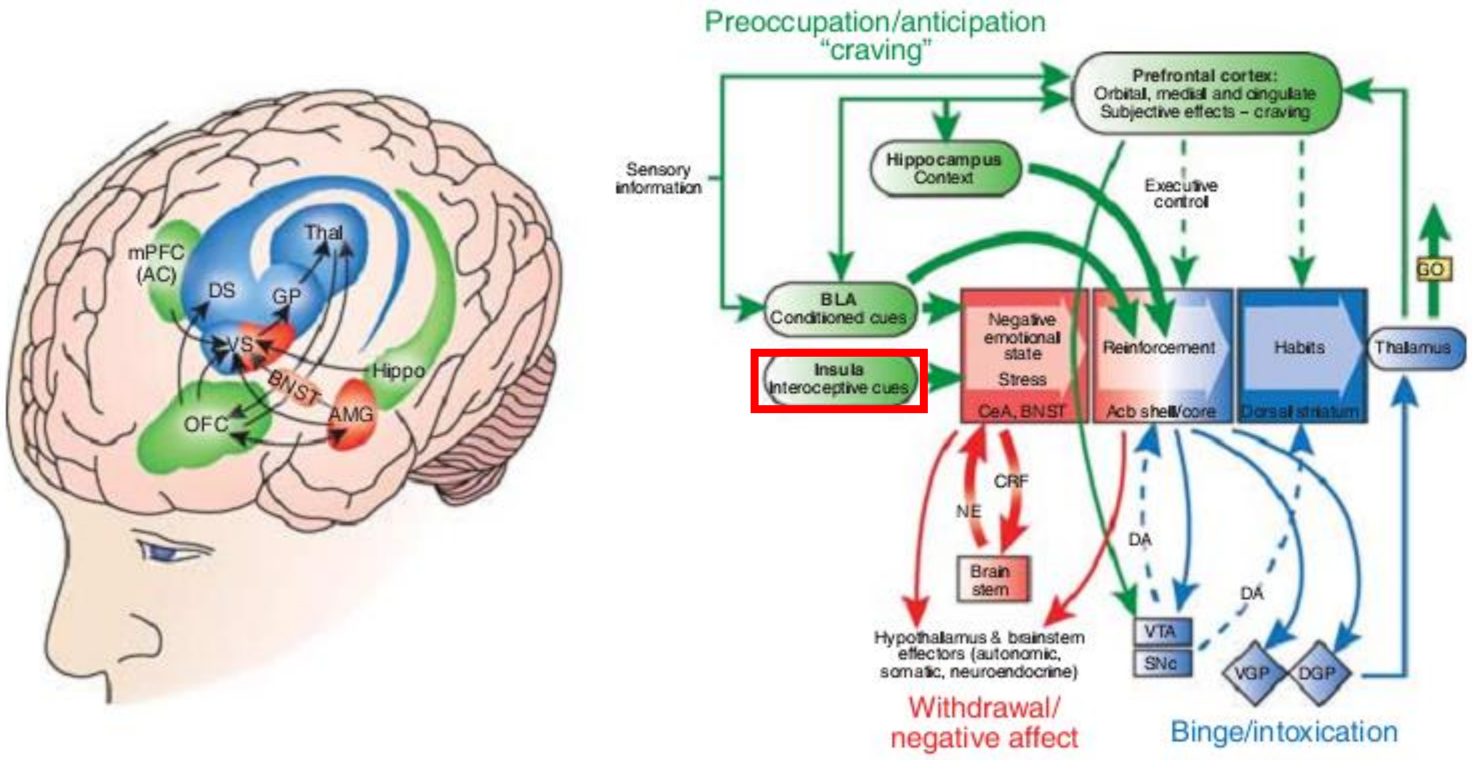


Craving/wanting → Seeking → Liking → Satisfaction → Learning

Insular?

Drug craving paradigm—CPP

- The desire to experience the effect(s) of a previously experienced psychoactive substance (Markou et al., 1993).



Published in final edited form as:
Science. 2007 January 26; 315(5811): 531–534. doi:10.1126/science.1135926.

Damage to the Insula Disrupts Addiction to Cigarette Smoking

Nasir H. Naqvi[1], **David Rudrauf**[1,2], **Hanna Damasio**[3,4], and **Antoine Bechara**[1,3,4,*]

[1]Division of Cognitive Neuroscience, Department of Neurology, University of Iowa Carver College of Medicine, 200 Hawkins Drive, Iowa City, IA 52242, USA

[2]Laboratory of Computational Neuroimaging, Department of Neurology, University of Iowa Carver College of Medicine, 200 Hawkins Drive, Iowa City, IA 52242, USA

[3]Dornsife Cognitive Neuroscience Imaging Center, SGM 501, University of Southern California, Los Angeles, CA 90089, USA

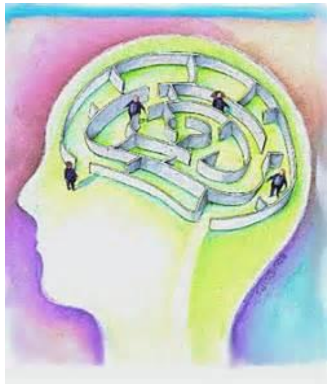
[4]Brain and Creativity Institute, HNB B26, University of Southern California, Los Angeles, CA 90089, USA

Abstract

A number of brain systems have been implicated in addictive behavior, but none have yet been shown to be necessary for maintaining the addiction to cigarette smoking. We found that smokers with brain damage involving the insula, a region implicated in conscious urges, were more likely than smokers with brain damage not involving the insula to undergo a disruption of smoking addiction, characterized by the ability to quit smoking easily, immediately, without relapse, and without persistence of the urge to smoke. This result suggests that the insula is a critical neural substrate in the addiction to smoking.

NIH-PA Author Ma

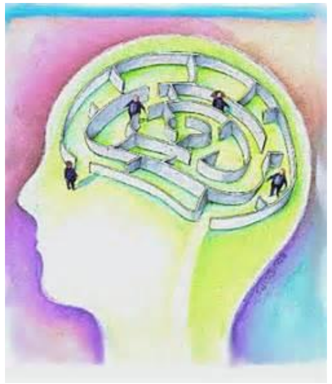
Establishing the Behavioral Platforms



Craving/wanting → Seeking → Liking → Satisfaction → Learning

Serotonin?

Establishing the Behavioral Platforms



Craving/wanting → Seeking → Liking → Satisfaction → Learning

Happiness?

A computational and neural model of momentary subjective well-being

Robb B. Rutledge[a,b,1], Nikolina Skandali[a], Peter Dayan[c], and Raymond J. Dolan[a,b]

Author Affiliations

Edited by Wolfram Schultz, University of Cambridge, Cambridge, United Kingdom, and accepted by the Editorial Board July 2, 2014 (received for review April 30, 2014)

Significance

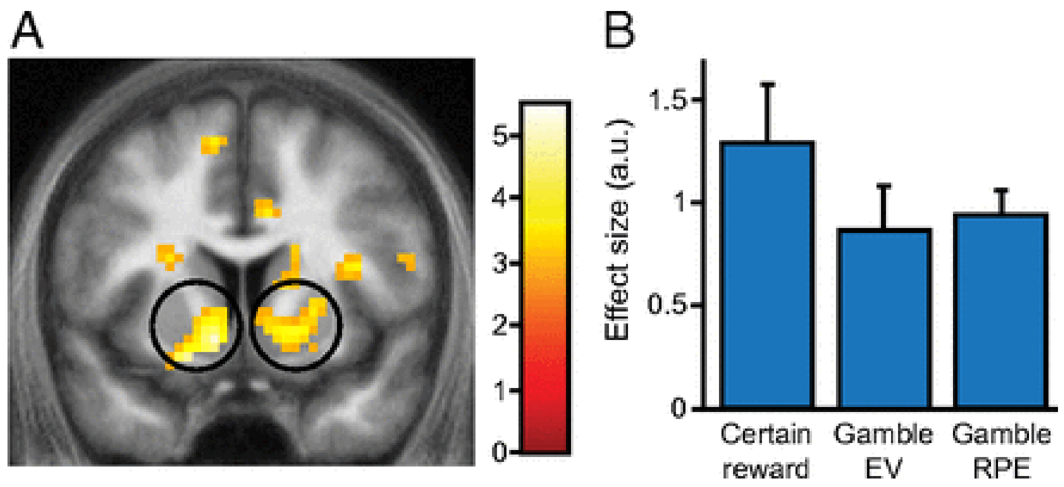
A common question in the social science of well-being asks, "How happy do you feel on a scale of 0 to 10?" Responses are often related to life circumstances, including wealth. By asking people about their feelings as they go about their lives, ongoing happiness and life events have been linked, but the neural mechanisms underlying this relationship are unknown. To investigate it, we presented subjects with a decision-making task involving monetary gains and losses and repeatedly asked them to report their momentary happiness. We built a computational model in which happiness reports were construed as an emotional reactivity to recent rewards and expectations. Using functional MRI, we demonstrated that neural signals during task events account for changes in happiness.

Abstract

The subjective well-being or happiness of individuals is an important metric for societies. Although happiness is influenced by life circumstances and population demographics such as wealth, we know little about how the cumulative influence of daily life events are aggregated into subjective feelings. Using computational modeling, we show that emotional reactivity in the form of momentary happiness in response to outcomes of a probabilistic reward task is explained not by current task earnings, but by the combined influence of recent reward expectations and prediction errors arising from those expectations. The robustness of this account was evident in a large-scale replication involving 18,420 participants. Using functional MRI, we show that the very same influences account for task-dependent striatal activity in a manner akin to the influences underpinning changes in happiness.

reward prediction error | dopamine | striatum | insula

Future Happiness in Ventral Striatum

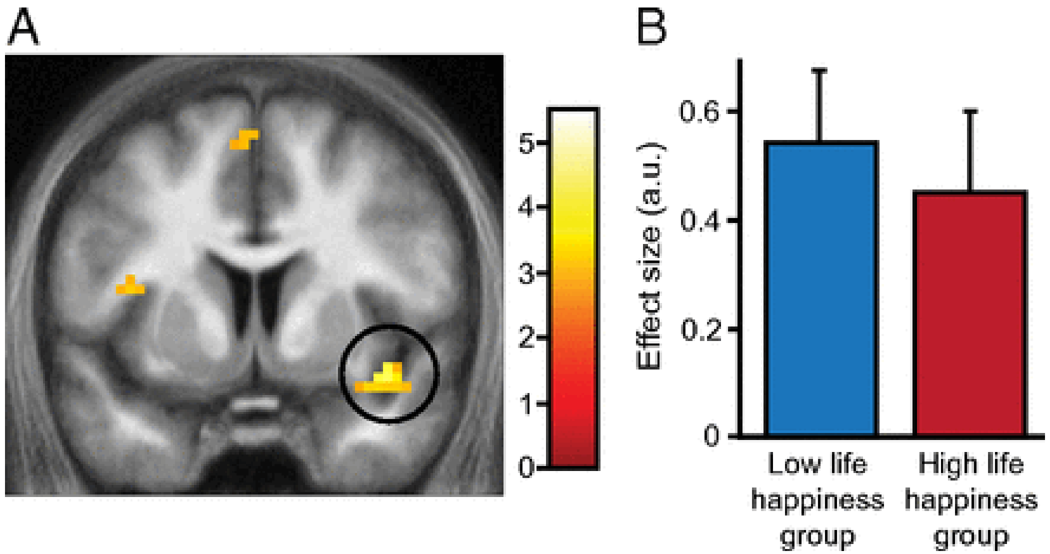


In a new window | Download PPT

Fig. 4.

Relationship between happiness and neural responses during preceding events. (*A*) Striatal activity during task events preceding subjective state ratings correlated with later self-reported happiness ($P < 0.05$, small-volume corrected). (*B*) Neural responses in ventral striatum were explained by the same parametric task variables as the variables that explained happiness. Error bars represent SEM.

Current Happiness in Insular



In a new window | Download PPT

Fig. 5.

Effect of the happiness question on neural activity in the right anterior insula. (*A*) In the right anterior insula, neural activity at the time of the happiness question presentation correlated with how happy subjects reported being ($P < 0.01$, small-volume corrected). (*B*) Parameter estimates were similar for subjects with low or high life happiness. Error bars represent SEM.

Even Correlated with Eudaemonic Happiness.

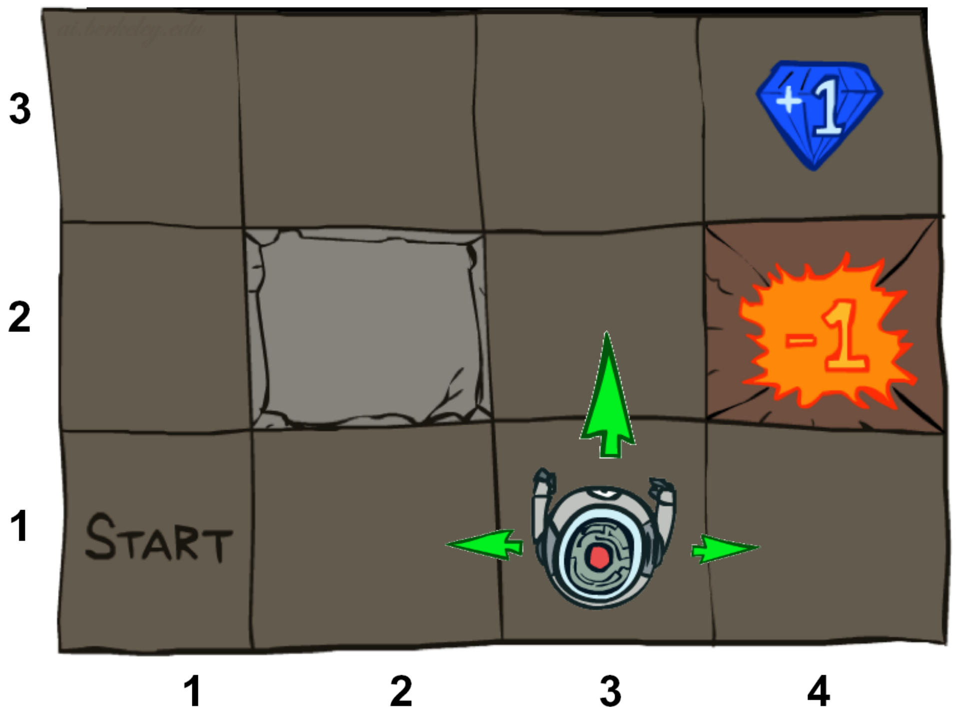
Comments on Human-level control through deep reinforcement learning

HKUST

2015/4/13

Example: Grid World

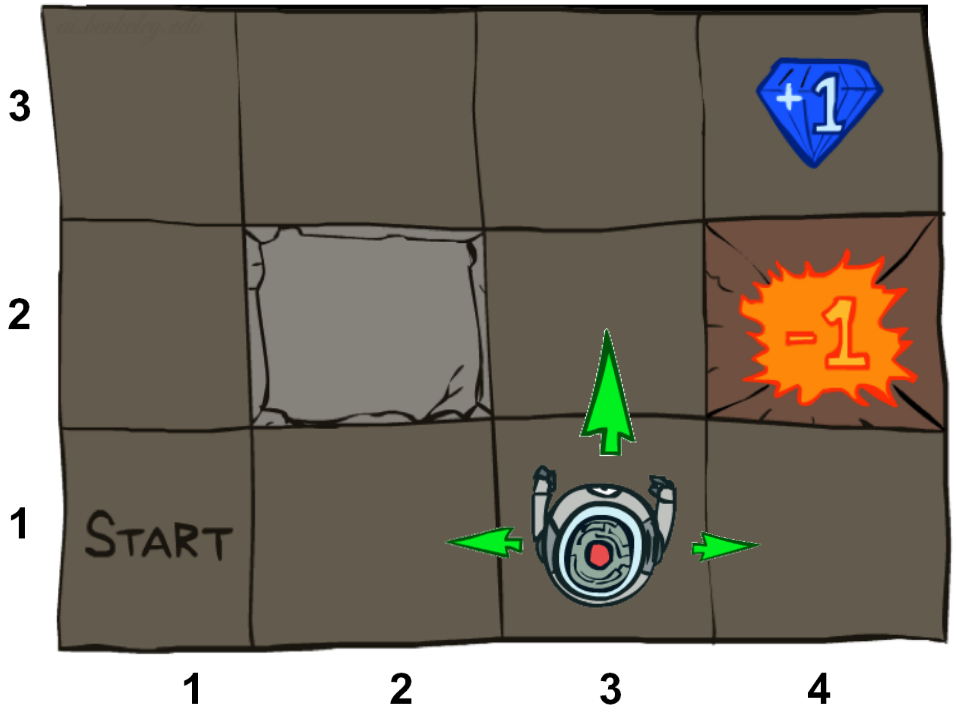
- A maze-like problem
  - The agent lives in a grid
  - Walls block the agent's path
- Noisy movement: actions do not always go as planned
  - 80% of the time, the action North takes the agent North (if there is no wall there)
  - 10% of the time, North takes the agent West; 10% East
  - If there is a wall in the direction the agent would have been taken, the agent stays put
- The agent receives rewards each time step
  - Small "living" reward each step (can be negative)
  - Big rewards come at the end (good or bad)
- Goal: maximize sum of rewards



Markov Decision Processes

- An MDP is defined by:
  - A set of states $s \in S$
  - A set of actions $a \in A$
  - A transition model $T(s, a, s')$
    - Probability that a from s leads to s', i.e., $P(s' | s, a)$
  - A reward function $R(s, a, s')$
    - Sometimes just $R(s)$ for current state
  - A start state
  - Possibly a terminal state (or absorbing state) with zero reward for all actions

- MDPs are fully observable but probabilistic search problems
  - Some instances can be solved with expectimax search
  - We'll have a new tool soon



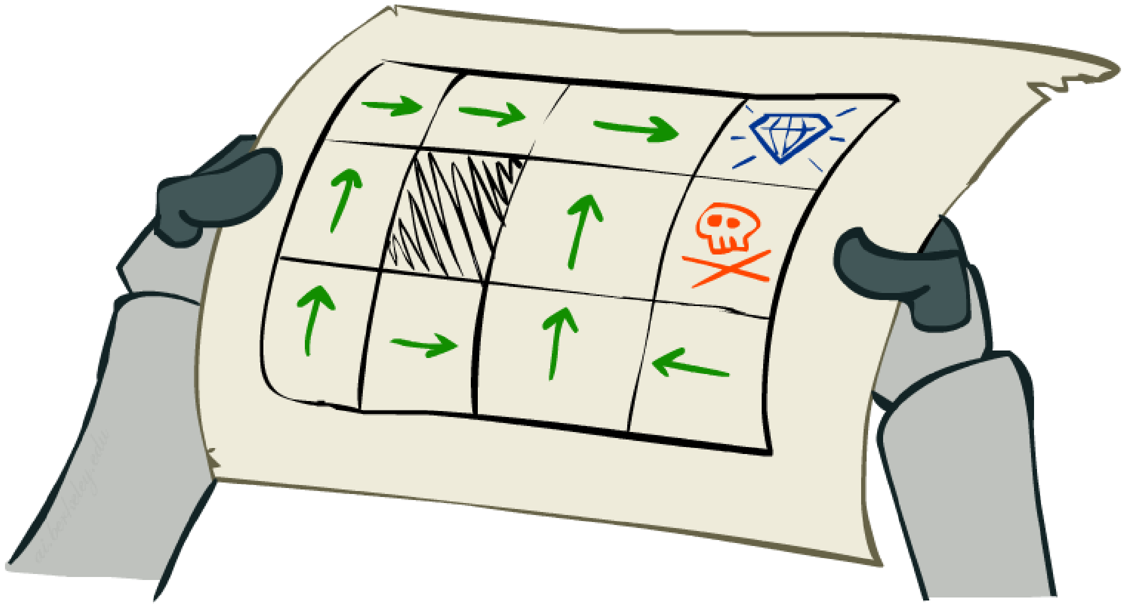
[Demo – gridworld manual intro (L8D1)]

Policies

- In deterministic single-agent search problems, we wanted an optimal ***plan***, or sequence of actions, from start to a goal

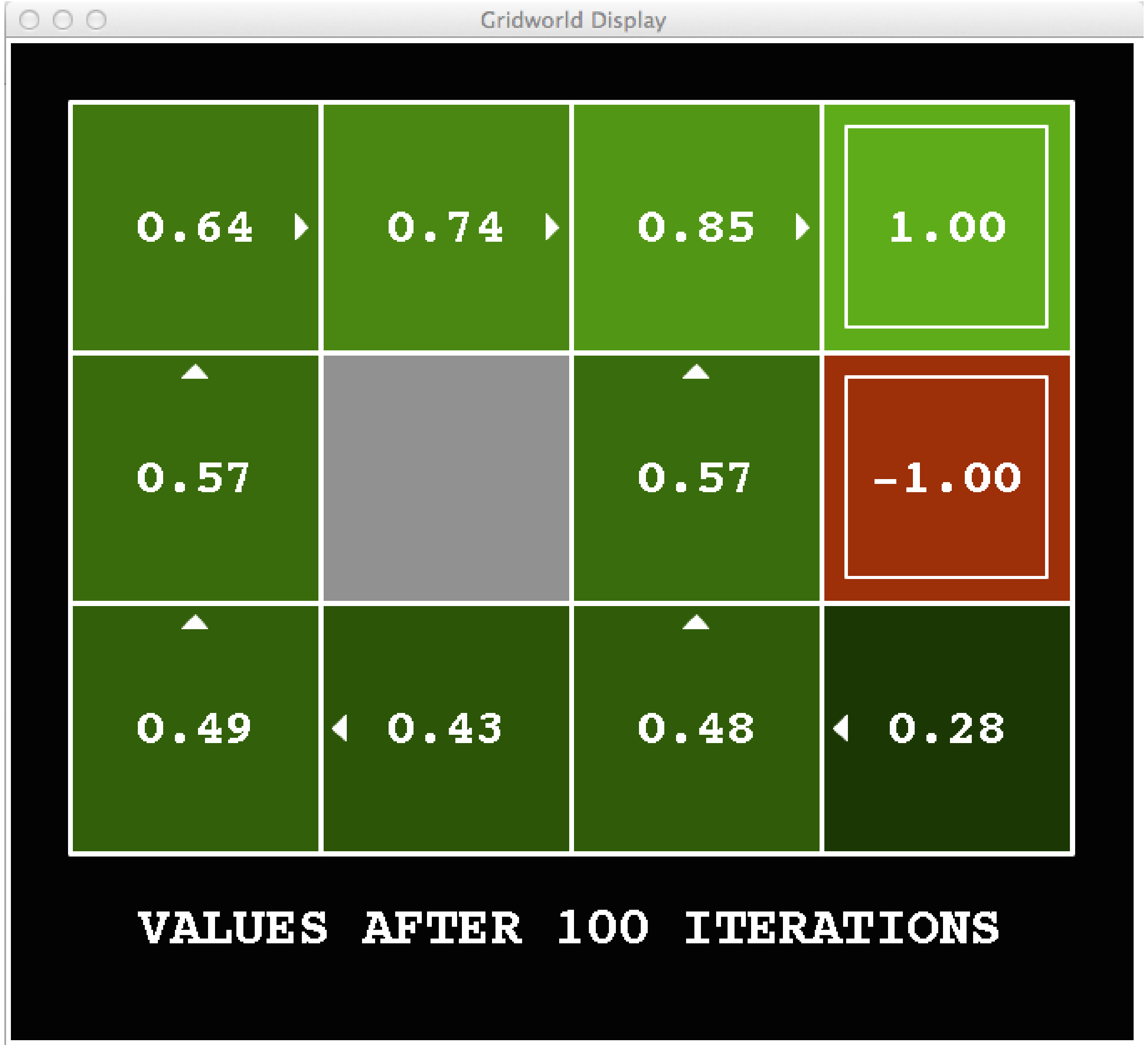
- For MDPs, we want an optimal ***policy*** $\pi^*: S \rightarrow A$
  - A policy π gives an action for each state
  - An optimal policy maximizes expected utility
  - An explicit policy defines a reflex agent

- Expectimax didn't compute entire policies
  - It computed the action for a single state only
  - It doesn't know what to do about loops



Optimal policy when R(s, a, s') = -0.04 for all non-terminals s

Snapshot of Demo – Gridworld V Values



Noise = 0.2
Discount = 0.9
Living reward = 0

Two Approaches

- Policy Iteration:
  - Choose a policy, calculate the state values under this policy, then try to improve the policy
- Value Iteration:
  - Directly calculate value of (state, action) pairs regardless of policy

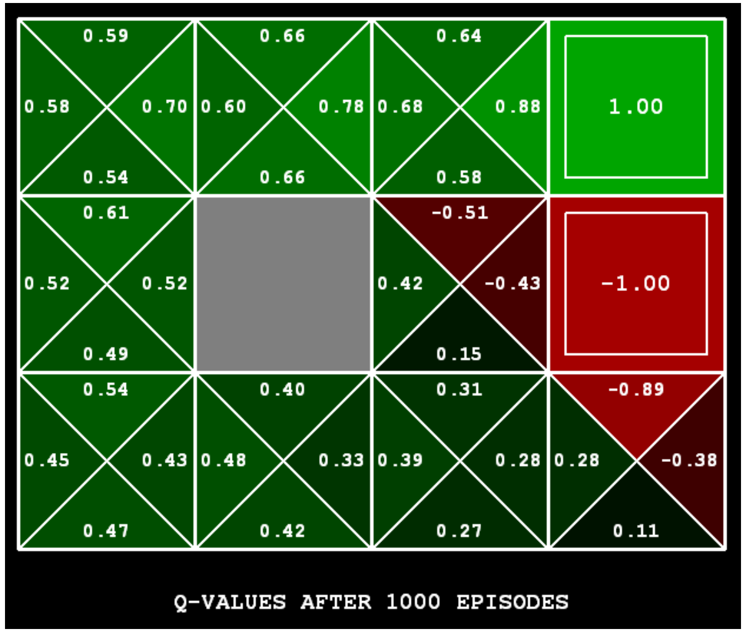
Q-Learning

- Learn Q(s,a) values as you go
  - Receive a sample (s,a,s',r)
  - Consider your old estimate: $Q(s,a)$
  - Consider your new sample estimate:

$$sample = R(s,a,s') + \gamma \max_{a'} Q(s',a')$$

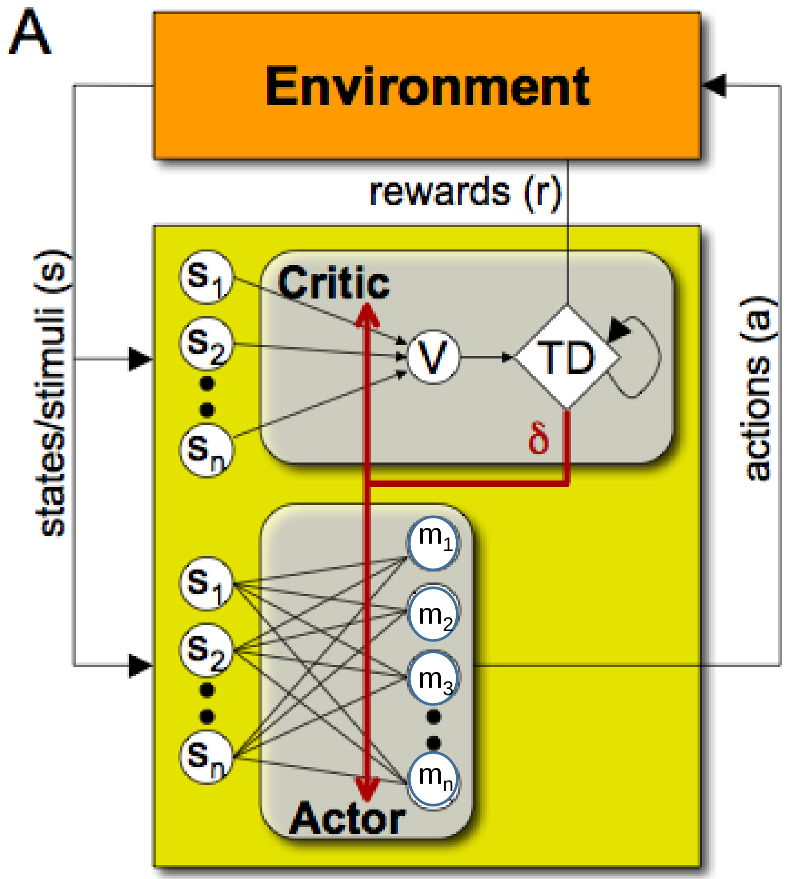
  - Incorporate the new estimate into a running average:

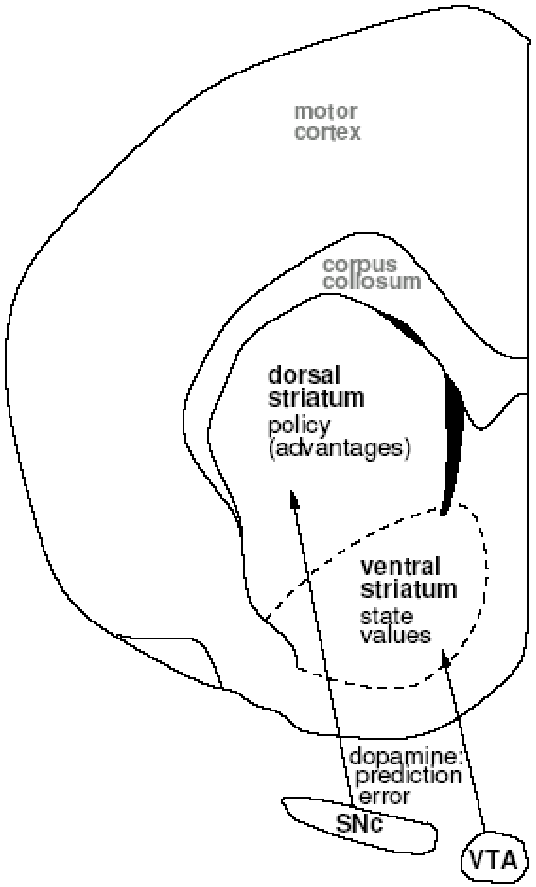
$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + (\alpha)\,[sample]$$



[Demo: Q-learning – gridworld (L10D2)]
[Demo: Q-learning – crawler (L10D3)]

actor/critic Learning

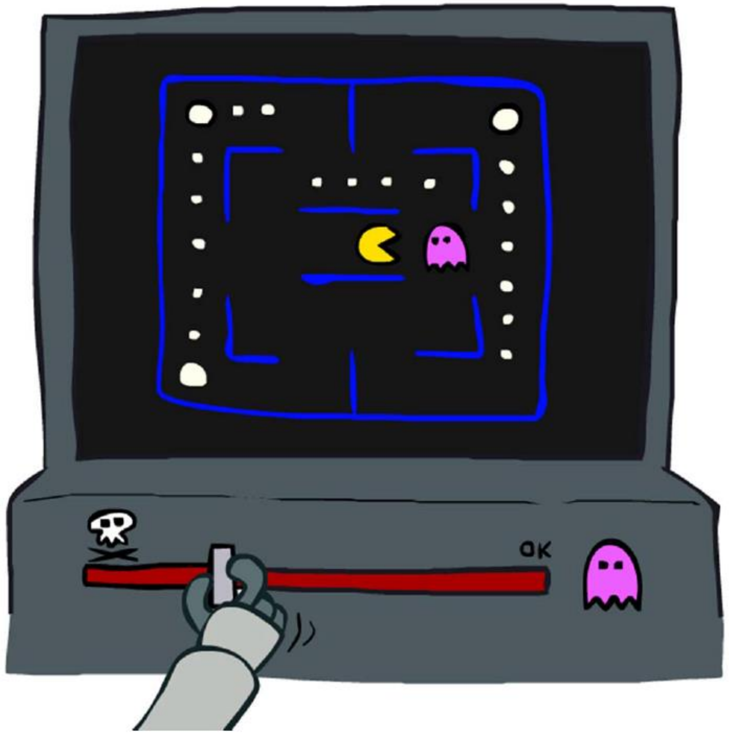




dopamine signals to both motivational & motor striatum appear, surprisingly the same

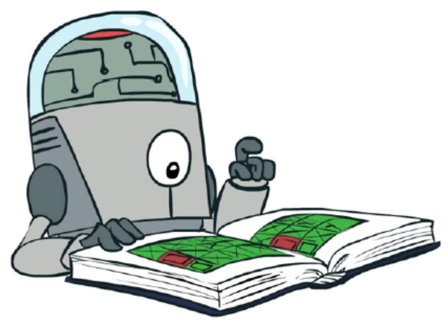
suggestion: training both values & policies

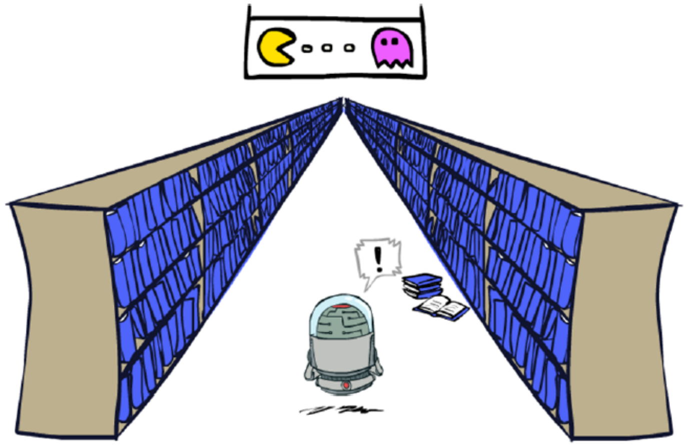
Approximate Q-Learning



Generalizing Across States

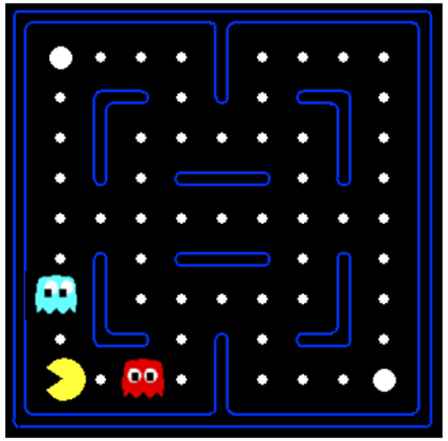
- Basic Q-Learning keeps a table of all q-values
- In realistic situations, we cannot possibly learn about every single state!
  - Too many states to visit them all in training
  - Too many states to hold the q-tables in memory
- Instead, we want to generalize:
  - Learn about some small number of training states from experience
  - Generalize that experience to new, similar situations
  - This is a fundamental idea in machine learning, and we'll see it over and over again



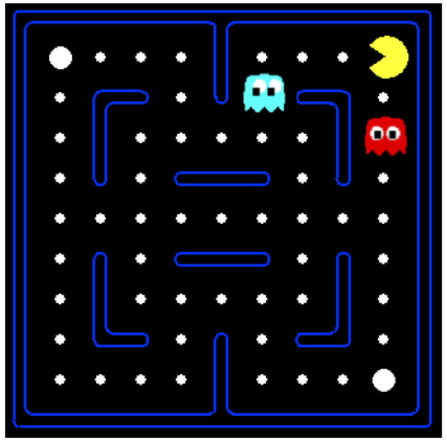


Example: Pacman

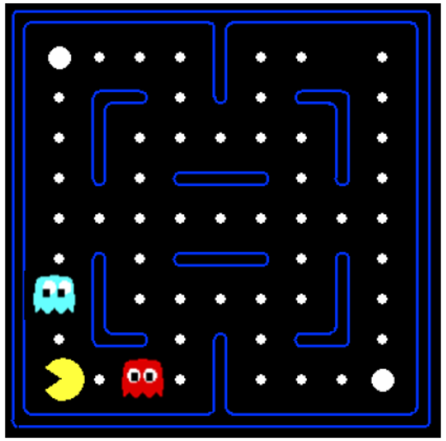
Let's say we discover through experience that this state is bad:



In naïve q-learning, we know nothing about this state:



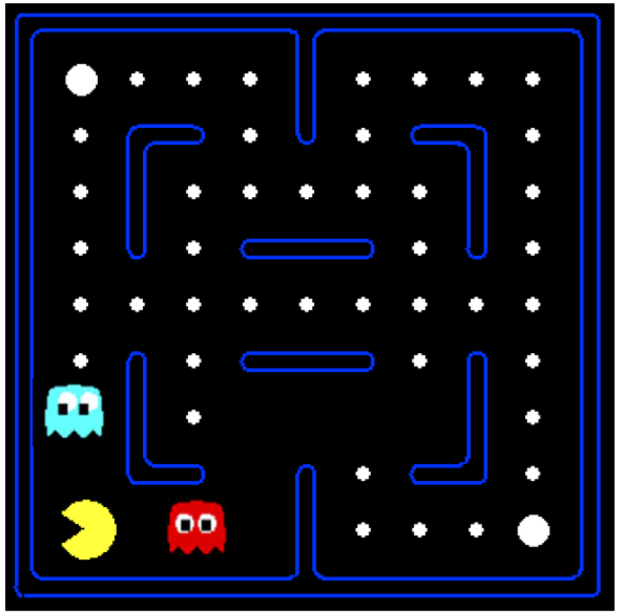
Or even this one!



[demo – RL pacman]

Feature-Based Representations

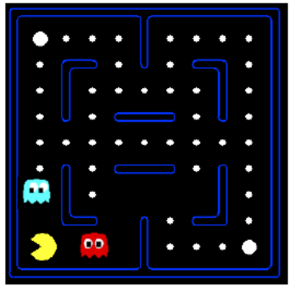
- Solution: describe a state using a vector of features (properties)
  - Features are functions from states to real numbers (often 0/1) that capture important properties of the state
  - Example features:
    - Distance to closest ghost
    - Distance to closest dot
    - Number of ghosts
    - 1 / (dist to dot)^2
    - Is Pacman in a tunnel? (0/1)
    - …… etc.
    - Is it the exact state on this slide?
  - Can also describe a q-state (s, a) with features (e.g. action moves closer to food)



Example: Q-Pacman

$$Q(s,a) = 4.0 f_{DOT}(s,a) - 1.0 f_{GST}(s,a)$$

s

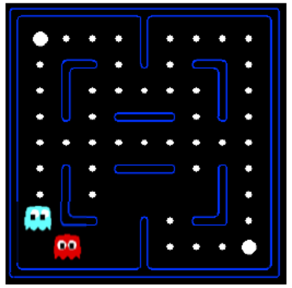


$$f_{DOT}(s, \text{NORTH}) = 0.5$$

$$f_{GST}(s, \text{NORTH}) = 1.0$$

$$a = \text{NORTH}$$
$$r = -500$$

s'



$$Q(s, \text{NORTH}) = +1$$

$$r + \gamma \max_{a'} Q(s', a') = -500 + 0$$

$$Q(s', \cdot) = 0$$

$$\text{difference} = -501$$

$$w_{DOT} \leftarrow 4.0 + \alpha\,[-501]\,0.5$$
$$w_{GST} \leftarrow -1.0 + \alpha\,[-501]\,1.0$$

$$Q(s,a) = 3.0 f_{DOT}(s,a) - 3.0 f_{GST}(s,a)$$

[demo – RL pacman]

ARTICLE PREVIEW

view full access options ▸

NATURE | LETTER

日本語要約

Human-level control through deep reinforcement learning

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg & Demis Hassabis

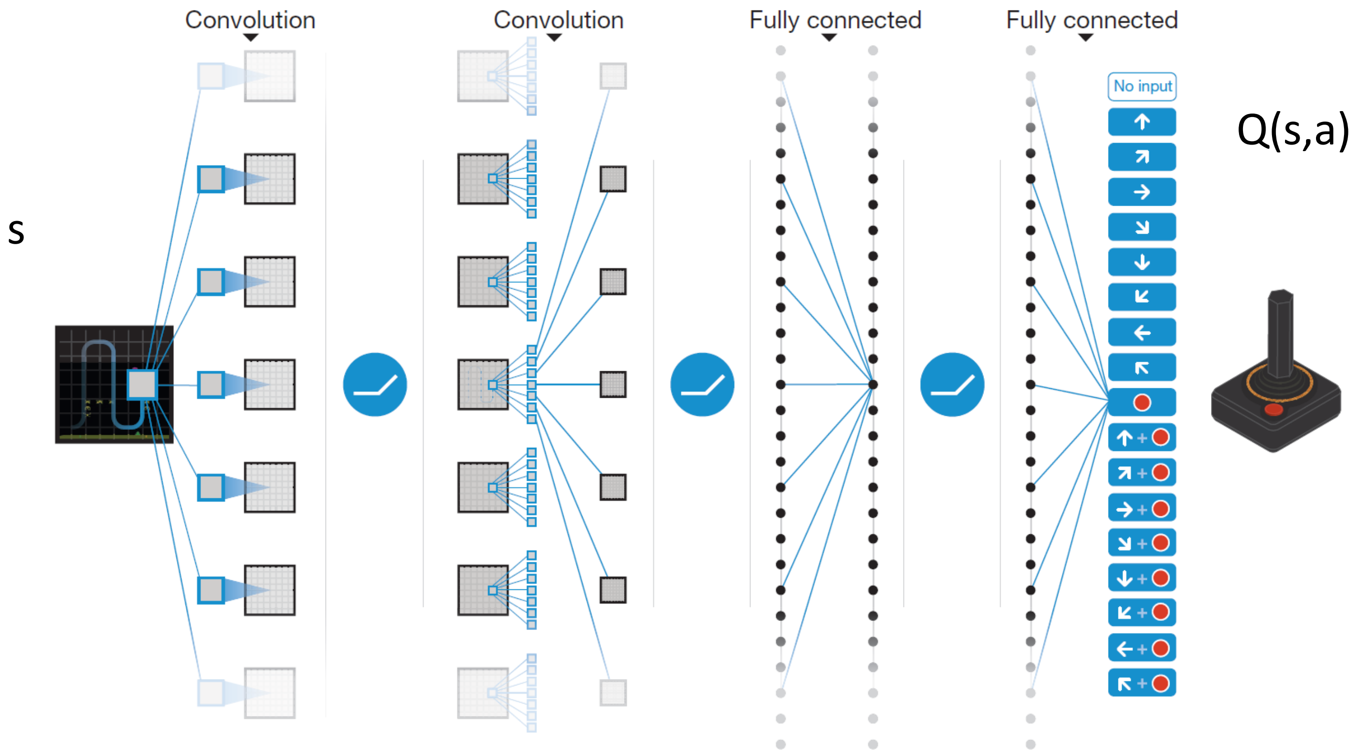
Affiliations | Contributions | Corresponding authors

Nature **518**, 529–533 (26 February 2015) | doi:10.1038/nature14236

Received 10 July 2014 | Accepted 16 January 2015 | Published online 25 February 2015

Core Idea

- Use deep neural network to evaluate the current situation



Key modifications to classical reinforcement learning

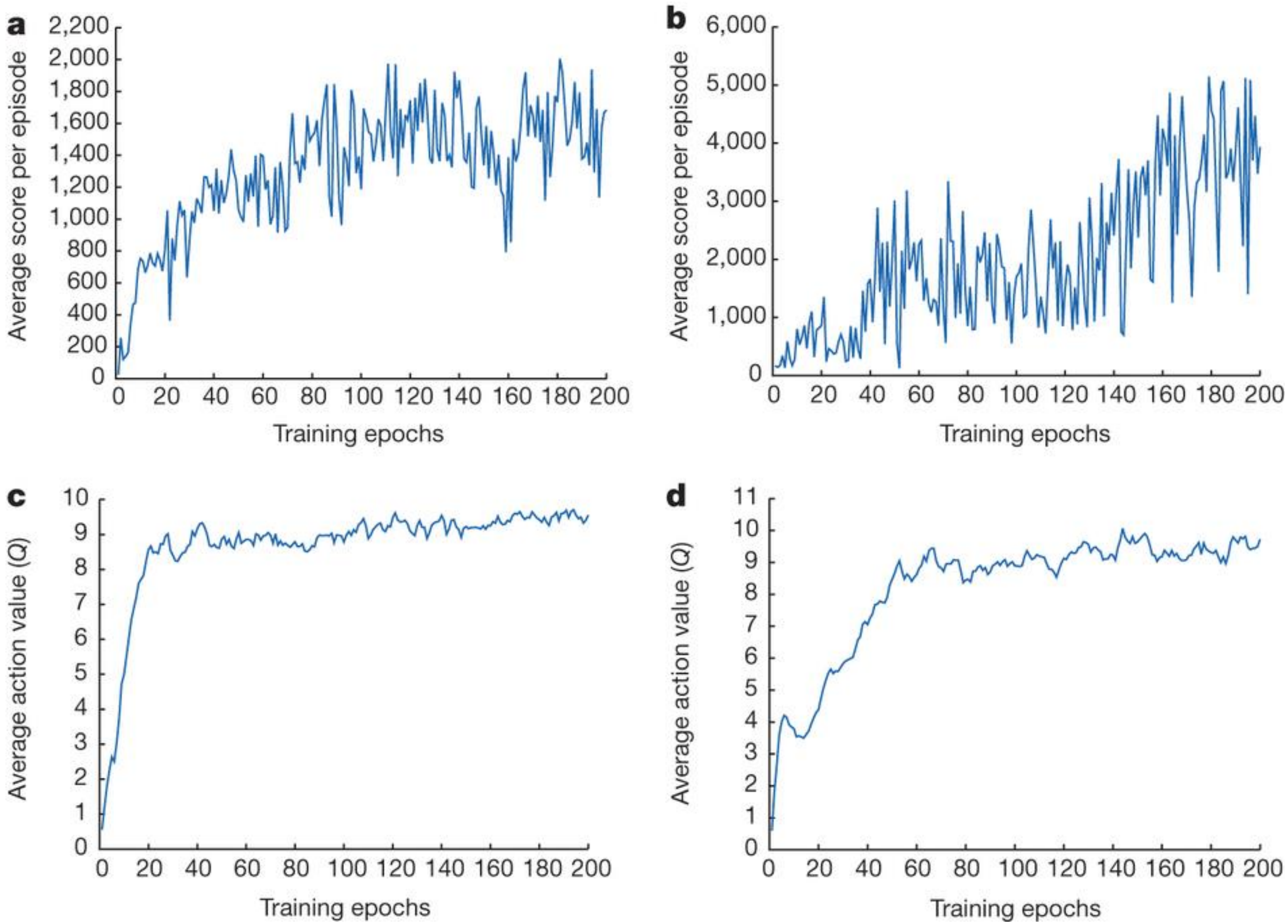
- Learn the mapping from input to Q values, instead of rewards or actions!
  - use a separate network for generating the targets yj in the Q-learning update.

$$\text{Set } y_j = \begin{cases} r_j \\ r_j + \gamma \max_{a'} \hat{Q}\left(\phi_{j+1}, a'; \theta^-\right) \end{cases}$$

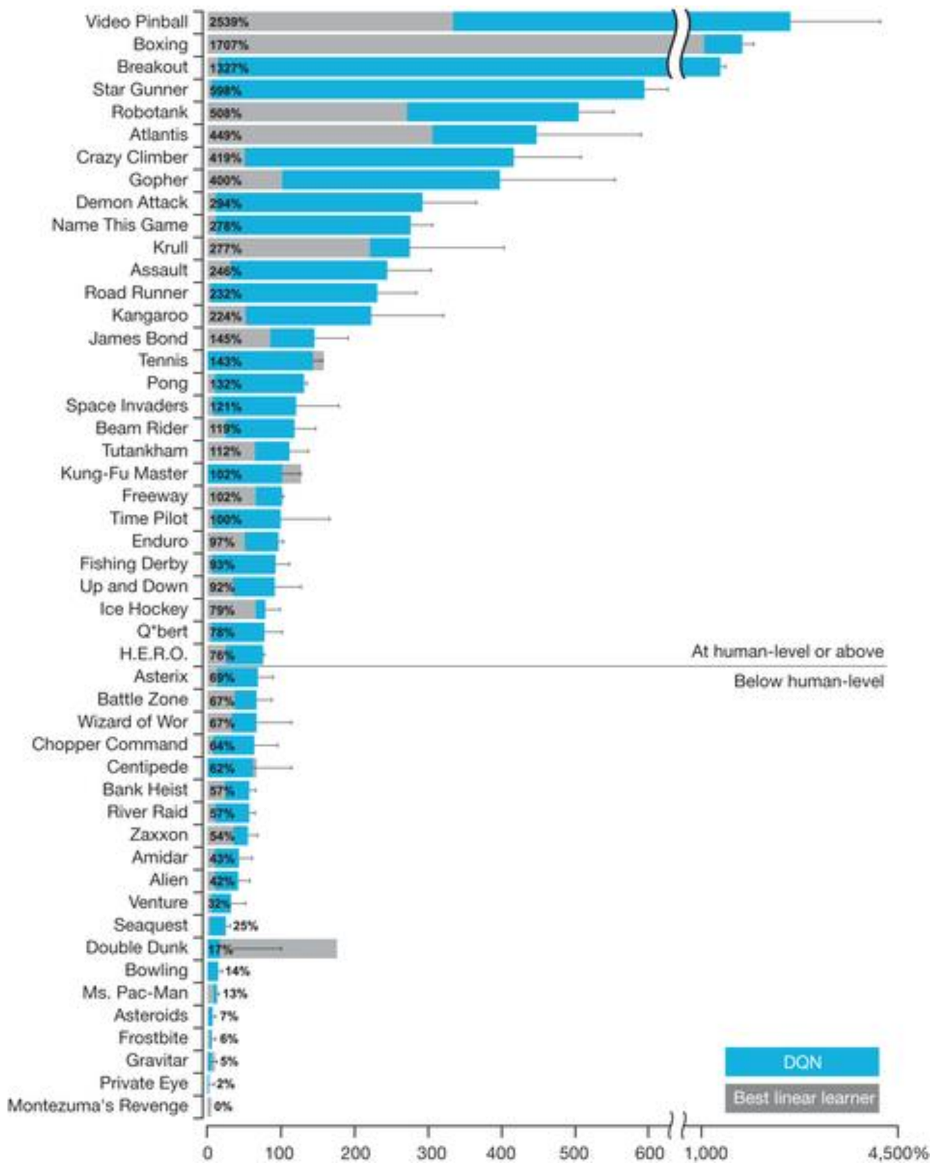
Every C step set $\hat{Q} = Q$

- relay memory
  - Store past trials
  - Choose random minibatch from the relay memory pool

Successful learning on Q-values



Success Rates for Various Games



Why Doing Poorly on Pacman?

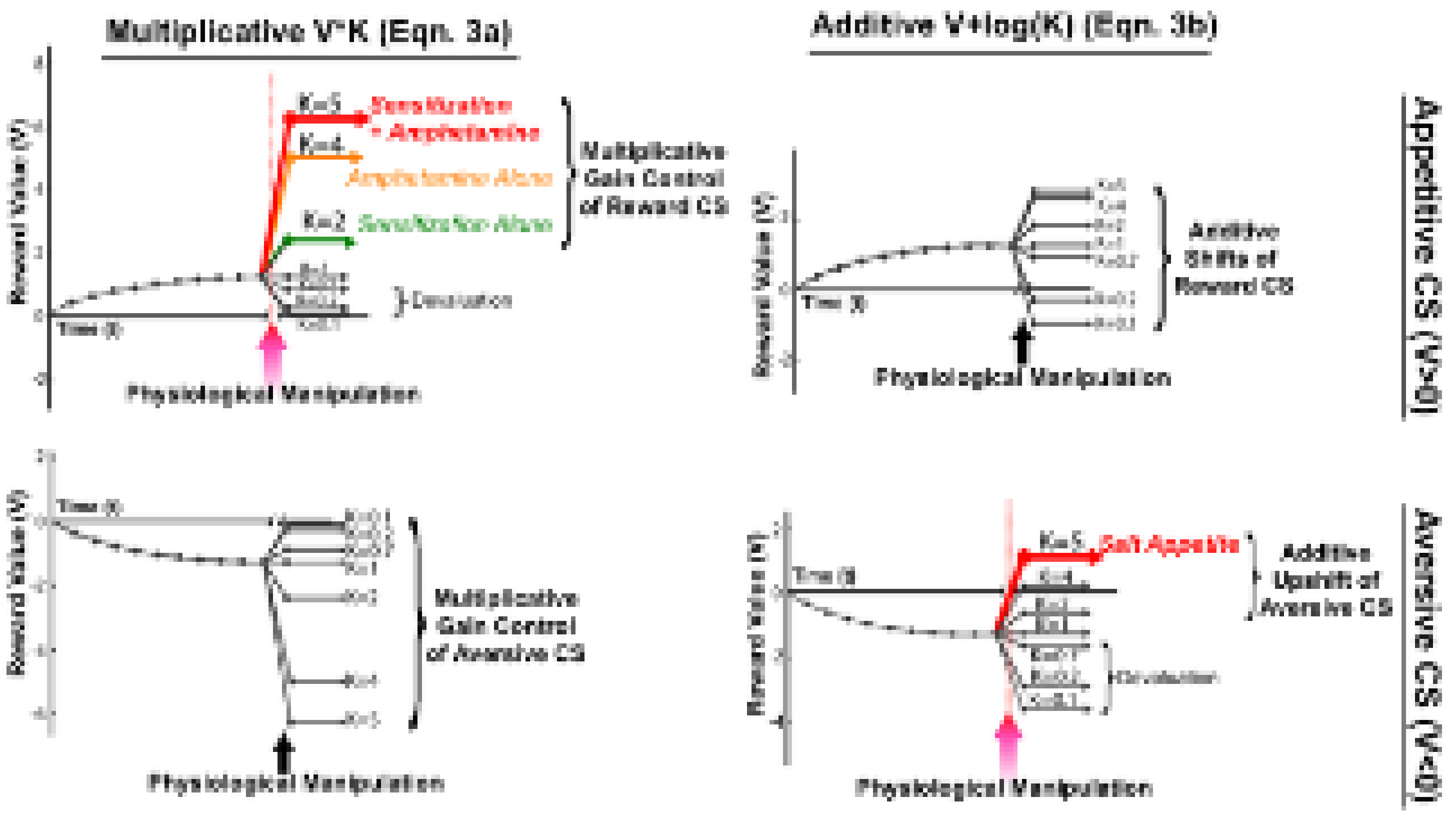
- No notion of Need/Motivation/Emotional State
- For the same feature, the best action is opposite given whether you have eaten the magic beans or not!!!
- Humans actually have goals!!!

Human Decision Making

- Humans spend a lot of time evaluating pros and cons when setting the goal
- Humans are normally persistent in pursuing a goal
- Humans periodically check progress towards a goal to decide whether to continue
- Goals can completely change your reward state values!!!

- In pacman situation, goals can be to avoid ghosts or pursue ghosts

Idea of Motivational Saliency



$$\tilde{V}(s_t) = \tilde{r}(r_t, \kappa) + \gamma V(s_{t+1})$$

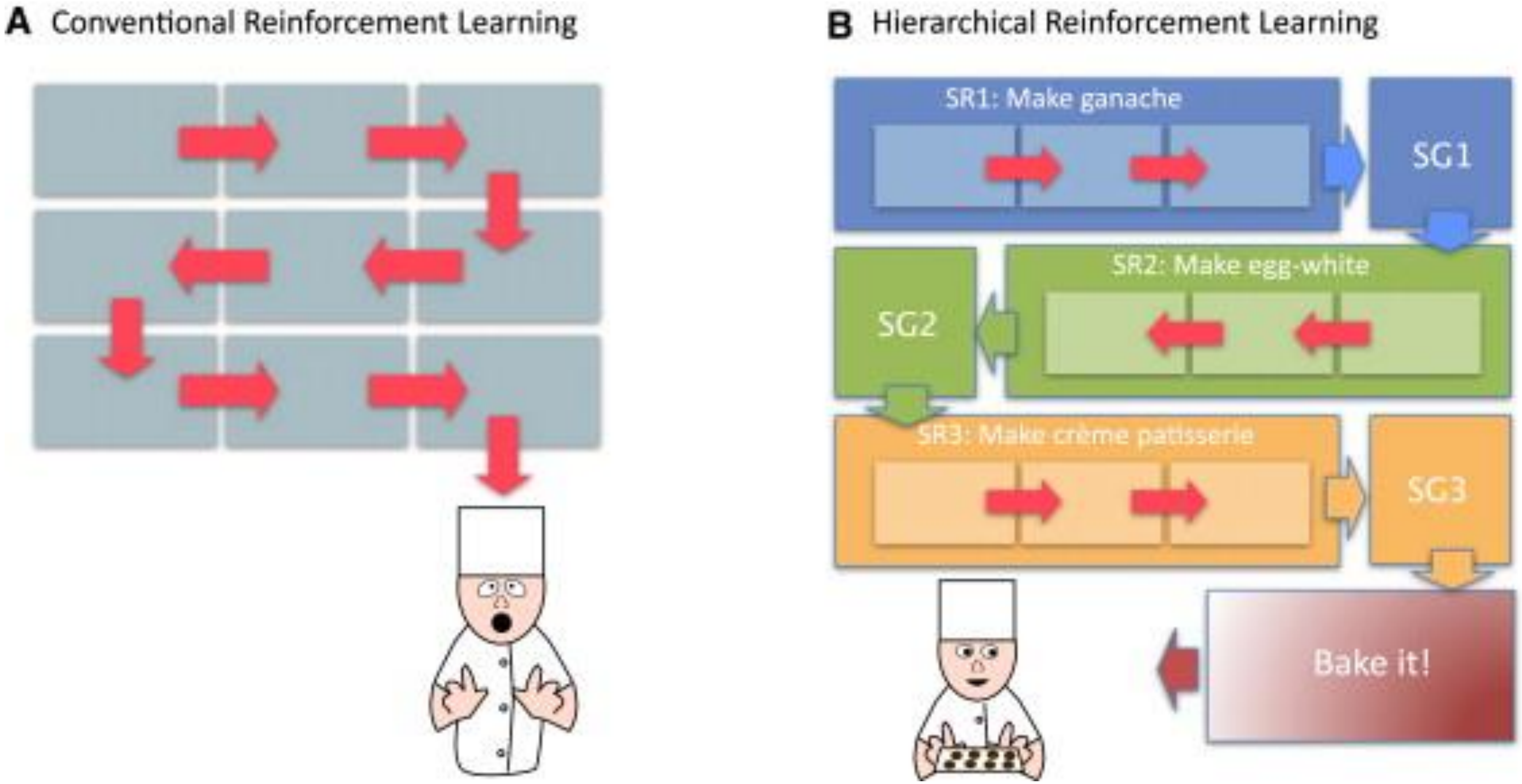
Zhang Jun et al **A Neural Computational Model of Incentive Salience. Plos Comp Bio 2009**

Recent Ideas for Improvements

- Hiearchical Reinforcement Learning

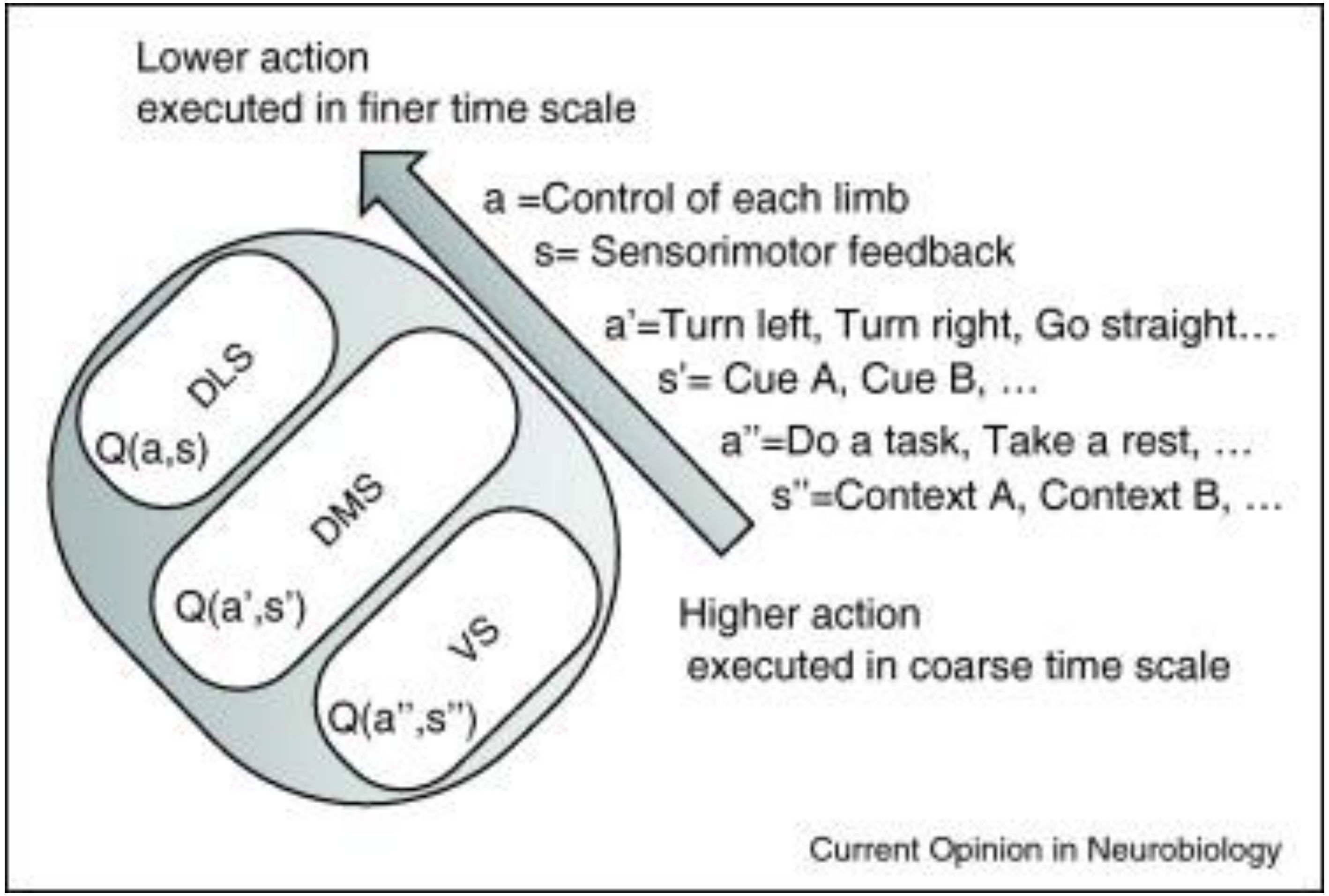
- Model-based versus model-free learning

Hierarchical Reinforcement Learning



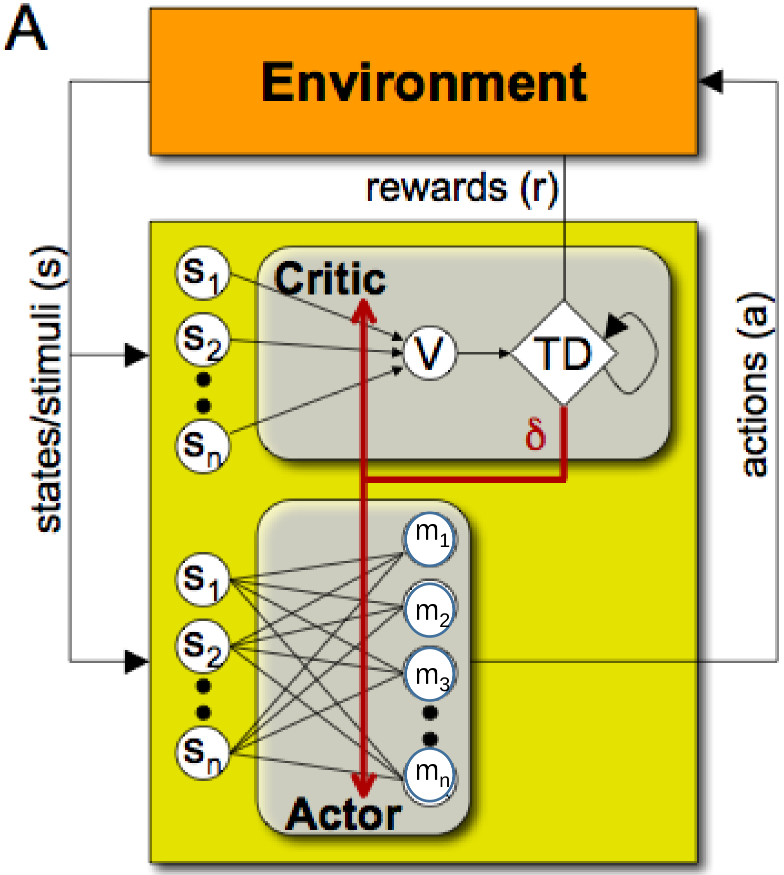
Behrens et al. How to Perfect a Chocolate Soufflé and Other Important Problems. Neuron 2011

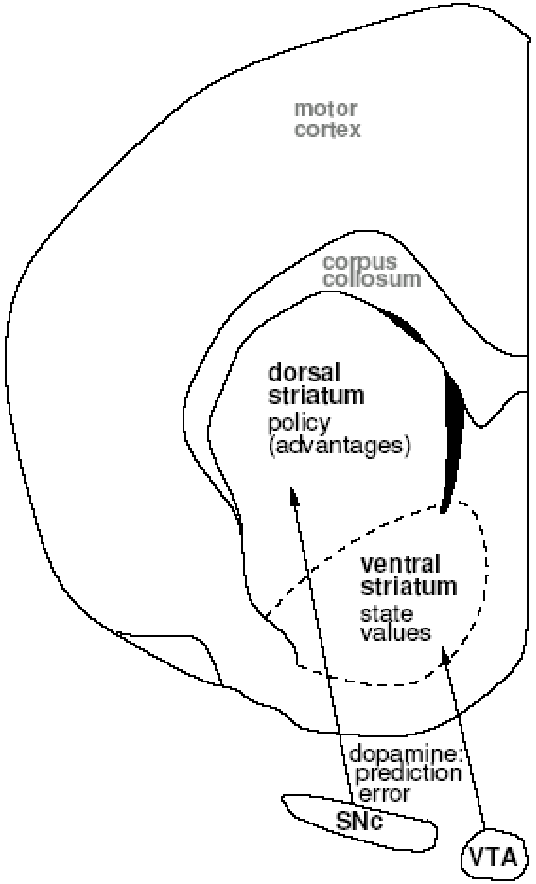
Mapping unto Brain Circuits



Ito M,Doya K(2011) Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. Curr Opin Neurobiol 21:368–373.

actor/critic Learning

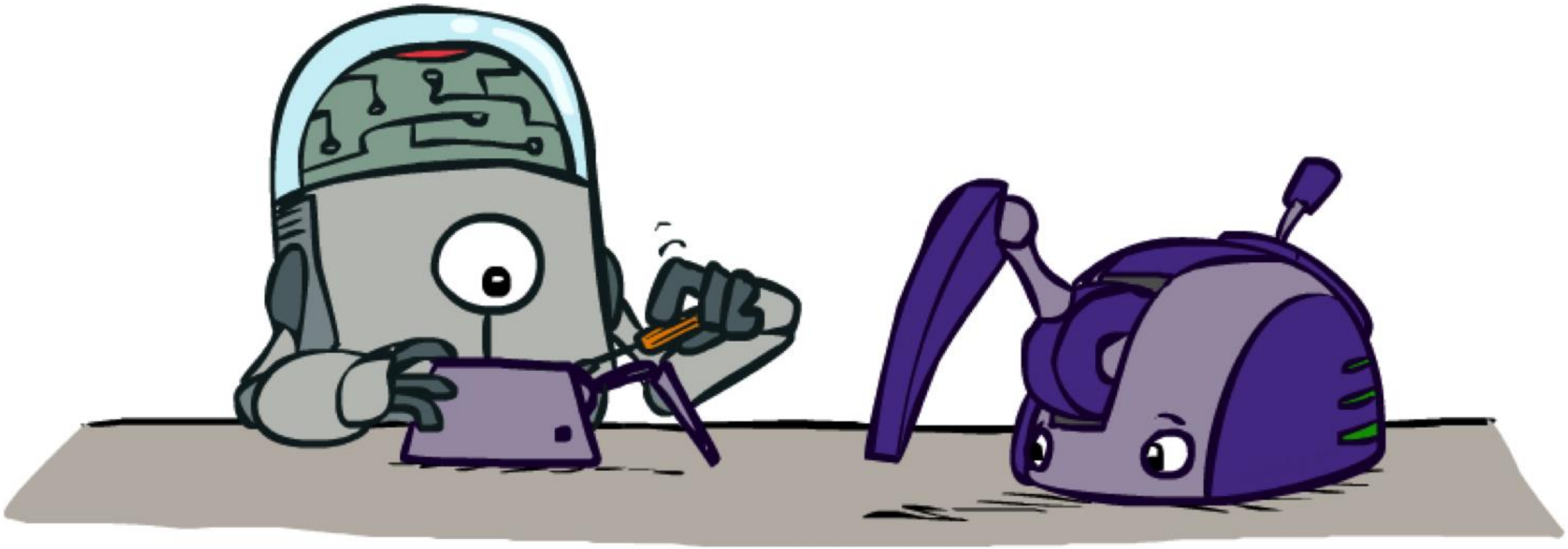




dopamine signals to both motivational & motor striatum appear, surprisingly the same

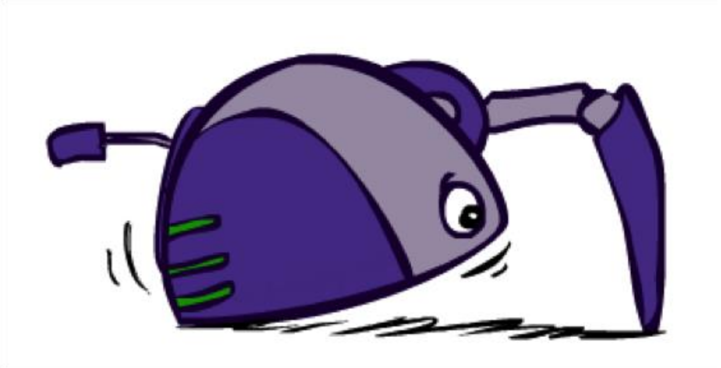
suggestion: training both values & policies

Model-Based Learning



Model-Based Learning

- Model-Based Idea:
  - Learn an approximate model based on experiences
  - Solve for values as if the learned model were correct



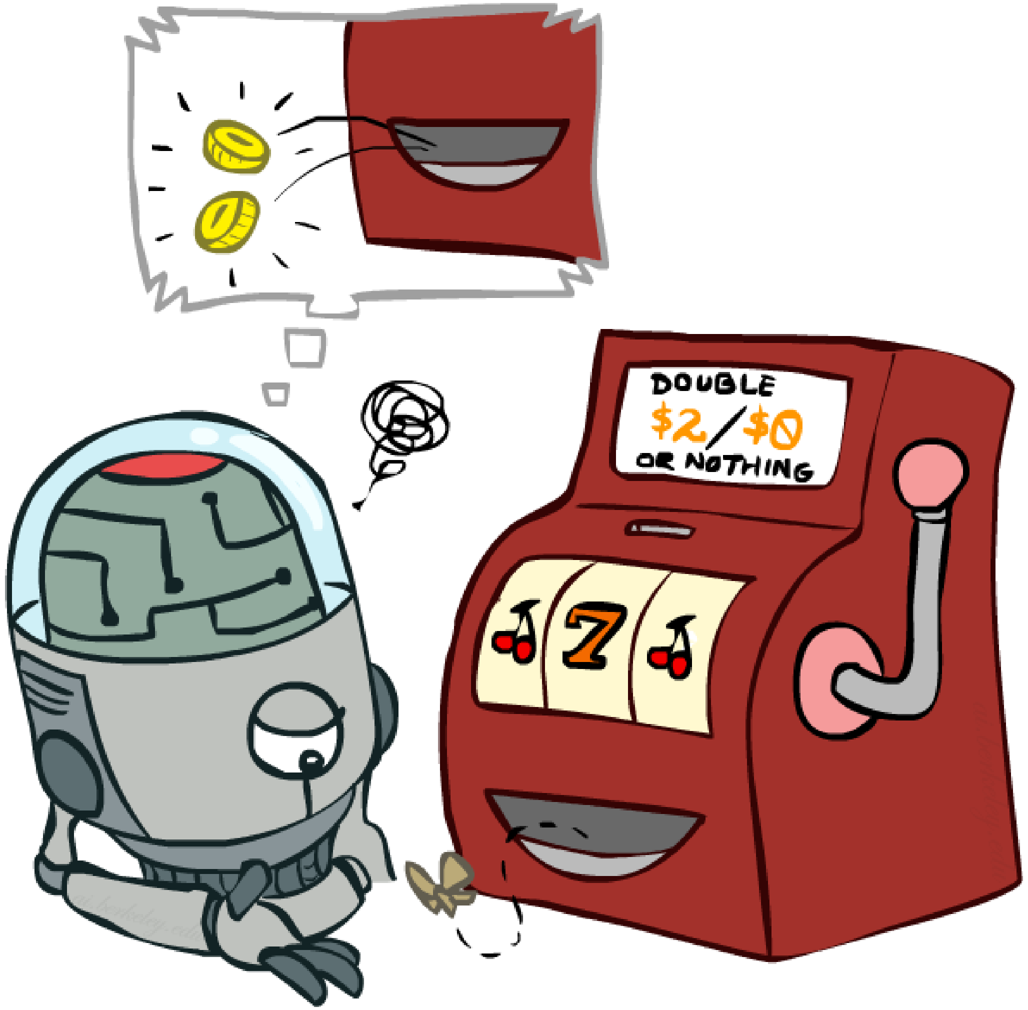
- Step 1: Learn empirical MDP model
  - Count outcomes s' for each s, a
  - Normalize to give an estimate of $P(s' \mid s, a)$
  - Discover each $R(s,a,s')$ when we experience the transition

- Step 2: Solve the learned MDP
  - For example, use value or policy iteration, as before

Pros and cons

- Pro:
  - Makes efficient use of experiences
- Con:
  - May not scale to large state spaces
    - Learns model one state-action pair at a time (but this is fixable)
    - Cannot solve MDP for very large |S|

Model-Free Learning



Example: Expected Age

Goal: Compute expected age of cs188 students

Known P(A)

$$E[A] = \sum_a P(a) \cdot a = 0.35 \times 20 + \ldots$$

Without P(A), instead collect samples $[a_1, a_2, \ldots a_N]$

Unknown P(A): "Model Based"

$$\hat{P}(A) = N_a/N$$

$$E[A] \approx \sum_a \hat{P}(a) \cdot a$$

Why does this work? Because eventually you learn the right model.

Unknown P(A): "Model Free"

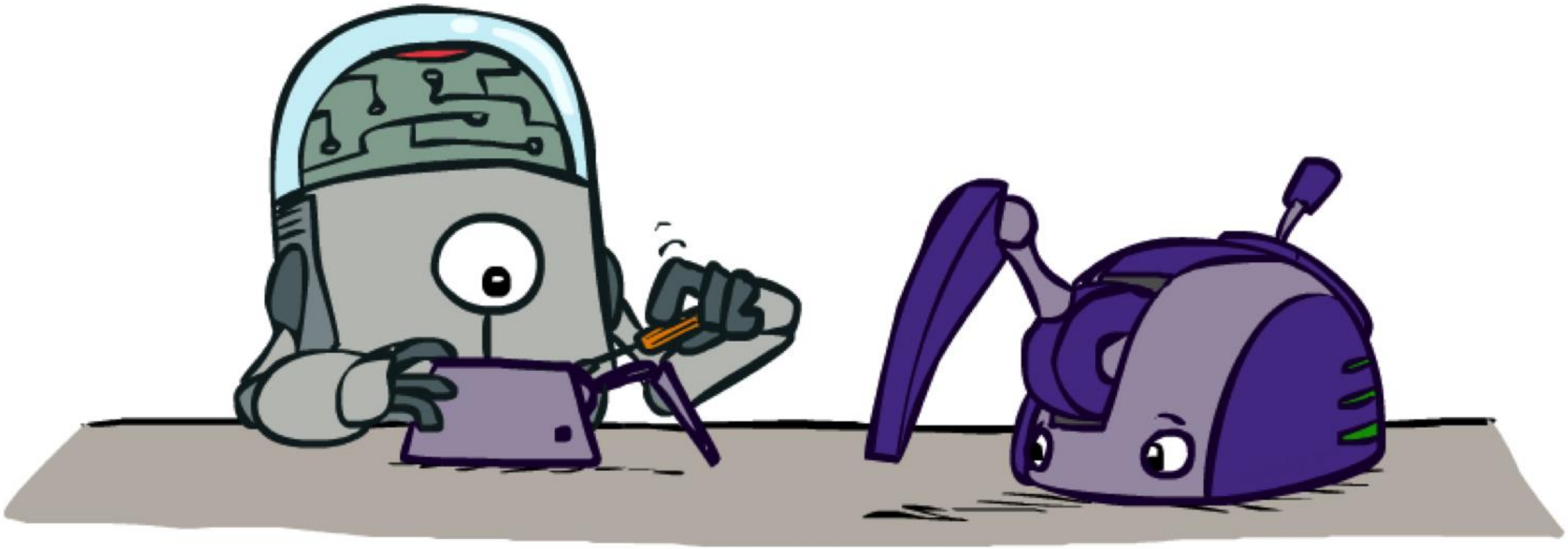
$$E[A] \approx 1/N \sum_i a_i$$

Why does this work? Because samples appear with the right frequencies.

Evaluation

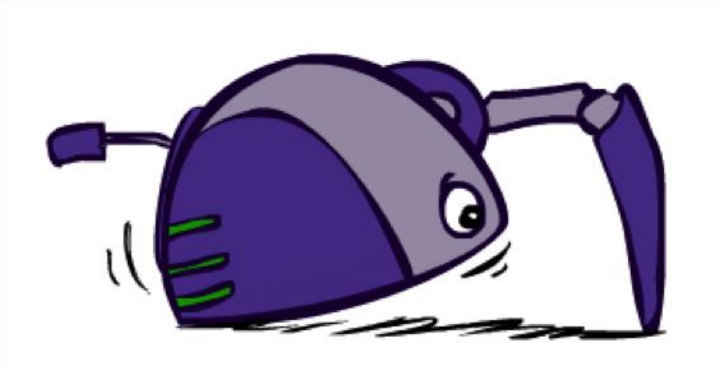
- Use random agent as baseline
- Human trained for 2h, test 20 x 5 min

Model-Based Learning



Model-Based Learning

- Model-Based Idea:
  - Learn an approximate model based on experiences
  - Solve for values as if the learned model were correct



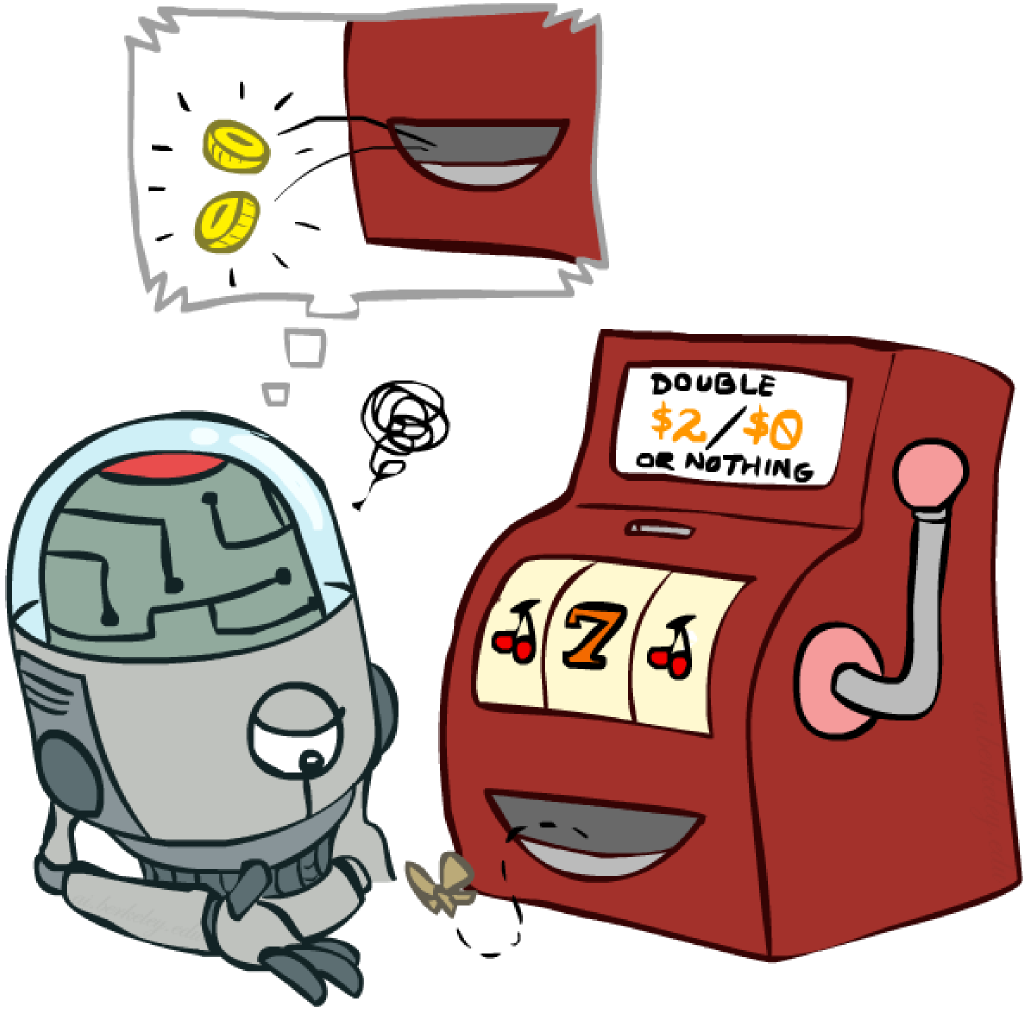
- Step 1: Learn empirical MDP model
  - Count outcomes s' for each s, a
  - Normalize to give an estimate of $P(s' \mid s, a)$
  - Discover each $R(s,a,s')$ when we experience the transition

- Step 2: Solve the learned MDP
  - For example, use value or policy iteration, as before

Pros and cons

- Pro:
  - Makes efficient use of experiences
- Con:
  - May not scale to large state spaces
    - Learns model one state-action pair at a time (but this is fixable)
    - Cannot solve MDP for very large |S|

Model-Free Learning



Example: Expected Age

Goal: Compute expected age of cs188 students

Known P(A)

$$E[A] = \sum_a P(a) \cdot a \;=\; 0.35 \times 20 + \ldots$$

Without P(A), instead collect samples $[a_1, a_2, \ldots a_N]$

Unknown P(A): "Model Based"

$$\hat{P}(A) = N_a/N$$

$$E[A] \approx \sum_a \hat{P}(a) \cdot a$$

Why does this work? Because eventually you learn the right model.

Unknown P(A): "Model Free"

$$E[A] \approx 1/N \sum_i a_i$$

Why does this work? Because samples appear with the right frequencies.

Idea by Peter Dayan

- Dorsomedial Striatum → Model Based

- Dorsolateral Striatum → Model Free

- dACC/vlPFC → Arbitrate between them

Acknowledgement

The Pleasure Team:
Xiong Yan (postdoc), Zhu Chen and Liao Fangzhou (grad students)
Supported by 4 computational modeling grad students

Collaborators:
Animal:
Li Yonghui and Shui Nan from Institute of Psychology
fMRI and DTI:
Wang Lihong and Guo Hua from Tsinghua,
Liu Chao and He Yong from Beijing Normal University

Donation of transgenic mouse lines from:
Luo Minmin, Jia Yichang, Guan Jisong

Supported by Natural Science Foundation China Grant
Open Grant from Chinese Academy of Science Institute of Psychology
Open Grant from Beijing Normal University
And
Center for Brain-Inspired Computation
McGovern Institute