Pre-Lecture Video

Reinforcement learning implemented in Basal Ganglia

Supervised learning: already know what you want to do. Mistakes tell you what to fix

Reinforced learning: animal needs to learn what to do. During learning you guess. Wrong guesses do **NOT** tell you what to fix.

Three signals Required to Implement law of Effect

(1) Action

(2) Situation

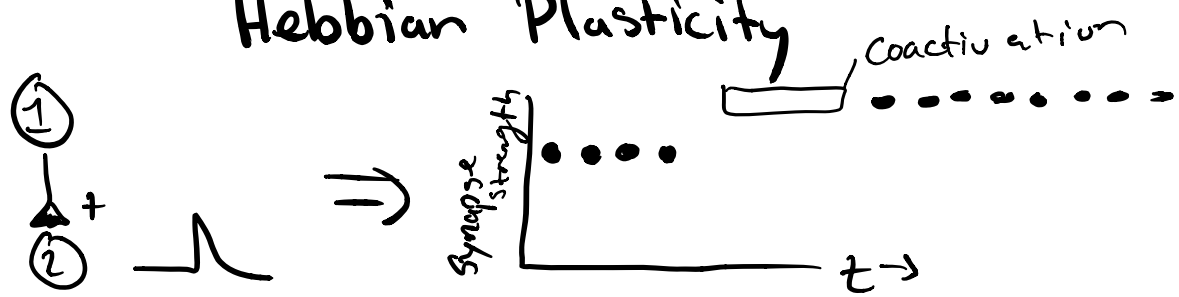
(3) Satisfying effect (REINFORCEMENT)

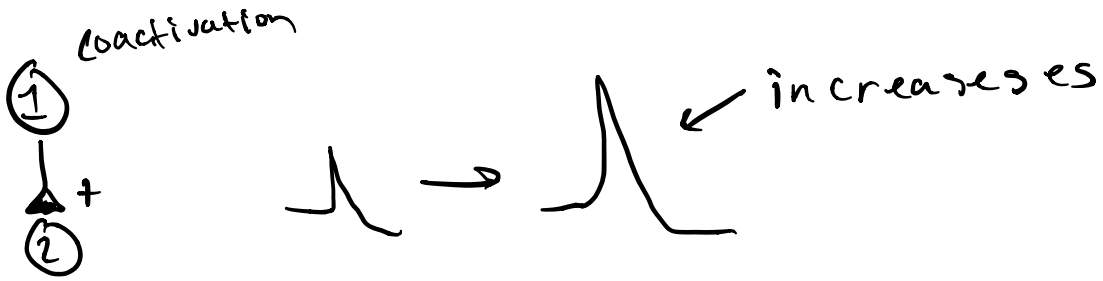
Dopamine activation can REINFORCE specific actions in a given content. These actions become habitual.

REWARD PREDICTION ERROR (RPE)

Dopamine Response = Reward Occurrence - Reward Prediction

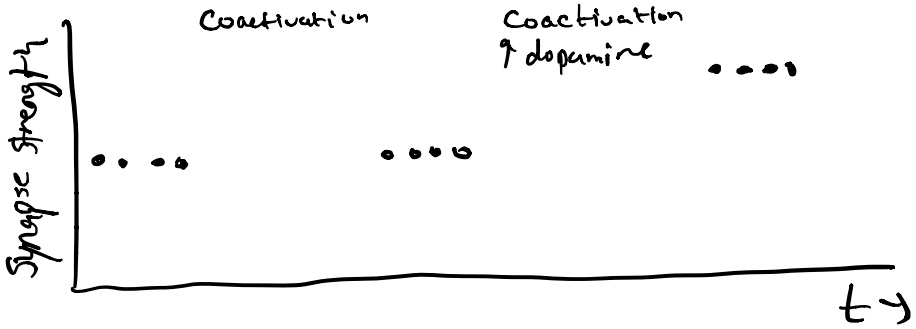
Hebbian Plasticity

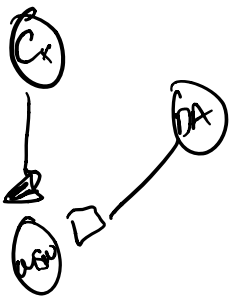




If (1) and (2) are co-active, then strengthen Connection from Pre → Post

Cortico-striatal plasticity also requires dopamine





← Three-factor learning rule

⇓

Cortex-MSN-Dopamine

↓ Context — ↓ action — ↓ effect

Law of Effect!

Lecture 25. Reinforcement Learning in the Basal Ganglia

Dr. Jesse Goldberg

Pre lecture materials (you will be tested on this content)

1. Panopto video on Reinforcement learning (versus Supervised Learning) & dopaminergic reinforcement and reward prediction error
3. Panopto video on dopamine modulated corticostriatal plasticity

Learning Objectives

1. Explain reinforcement learning at Marr's three levels.
2. Explain how dopamine functions as a reinforcement signal (*by encoding reward prediction error*). See figure at the end of this outline; you will be tested on it.
3. Explain how dopamine-modulated *corticostriatal plasticity* links context to action to maximize reward.

Lecture Outline

A PDF of the powerpoint slides from this lecture will be made available on the class website and will contain all key figures and concepts you need to know for this unit.

I. What is reinforcement learning?

Edward Thorndike's "Law of Effect"

If an animal produces an action in a particular situation and it results in a rewarding outcome, then the next time the animal finds itself in that situation it will be more likely to take that action.

II. Dopamine encodes a reinforcement signal

Dopamine neurons are activated by unexpected rewards (surprises)

Dopamine neurons are suppressed by unexpected reward omissions (disappointments)

Optogenetic activation of dopamine neurons is reinforcing

Drugs of addiction act on the dopamine system

III. The basal ganglia *link context to action* to implement reinforcement learning

Learning boils down to knowing *what* to do (the action) and *when* to do it (the situation or context).

The striatum implements this learning by integrating three key signals: (1) Context, (2) Action and (3) Outcome.

(1) Context: Comes from corticostriatal inputs from sensory cortex.
(2) Action: Comes from spiking of striatal medium spiny neurons that project to downstream motor structures in the pathway you learned in lecture 22.
(3) Outcome: Reward prediction error comes from dopamine inputs.

By integrating these three inputs, the striatum can compute:

I was just in [this] particular situation, I took [this] particular action, and it resulted in a [favorable] outcome → So next time I find myself here I'll do it again! (This is a habit).

IV. A test case: Stimulus-Response learning

We will walk step by step through a test case where you will learn to respond to a cue with a particular action. The end result is a HABIT.

A specific *learning rule* will be introduced. What is a learning rule?
A learning rule is a set of events that are required for synaptic plasticity (i.e. for the strength of the connection between two neurons to be modified). For example, "Fire together, wire together" is a famous learning rule, called the "Hebbian learning rule." If presynaptic neuron *A* fires before postsynaptic neuron *B* then the strength between *A*→*B* increases. This learning rule has two factors (*A* and *B* firing together)

In reinforcement learning, there is a modified "Three-factor" learning rule that has the same two factors as the Hebbian Rule, as well as the extra factor for dopamine to be present, as follows:

If A and B fire together, and there is also an increase in dopamine within some period of time, then increase the strength between *A*→*B*. (If no dopamine increase occurs, then there is no change in the strength between A→B.) ALL THREE FACTORS ARE REQUIRED FORLEARNING.

In this regime, *A* is the cortical input to the striatum. *B* is the medium spiny neuron.

In essence, the three factors are the three signals needed for learning (Context, Action and Outcome). Note that the dopamine that signals the outcome of the action might come with a delay relative to when the action was taken. An *Eligibility Trace* is an essential part of the learning rule.

Study Questions

1. What motor skills do you know that you learned through trial and error?
2. Can you think of anything that you *didn't* learn through trial and error?
3. What types of learning deficits might a Parkinson's patient have?
4. How does an eligibility trace determine how long a delay can be between an action and a reward?

This is an important figure and you will be tested on it. The plots show that dopamine neurons are activated by unpredicted rewards (surprise) and are suppressed by reward omission (disappointment).

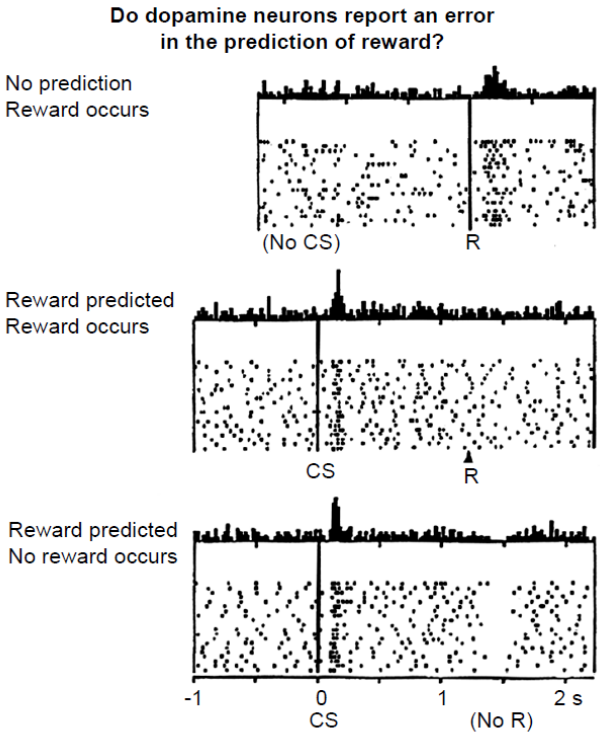


Fig. 1. Changes in dopamine neurons' output code for an error in the prediction of appetitive events. (**Top**) Before learning, a drop of appetitive fruit juice occurs in the absence of prediction—hence a positive error in the prediction of reward. The dopamine neuron is activated by this unpredicted occurrence of juice. (**Middle**) After learning, the conditioned stimulus predicts reward, and the reward occurs according to the prediction—hence no error in the prediction of reward. The dopamine neuron is activated by the reward-predicting stimulus but fails to be activated by the predicted reward (right). (**Bottom**) After learning, the conditioned stimulus predicts a reward, but the reward fails to occur because of a mistake in the behavioral response of the monkey. The activity of the dopamine neuron is depressed exactly at the time when the reward would have occurred. The depression occurs more than 1 s after the conditioned stimulus without any intervening stimuli, revealing an internal representation of the time of the predicted reward. Neuronal activity is aligned on the electronic pulse that drives the solenoid valve delivering the reward liquid (top) or the onset of the conditioned visual stimulus (middle and bottom). Each panel shows the peri-event time histogram and raster of impulses from the same neuron. Horizontal distances of dots correspond to real-time intervals. Each line of dots shows one trial. Original sequence of trials is plotted from top to bottom. CS, conditioned, reward-predicting stimulus; R, primary reward.

Prelim 2: Wednesday, March 27 from 12:20-1:10 PM
Covers lectures 14-26

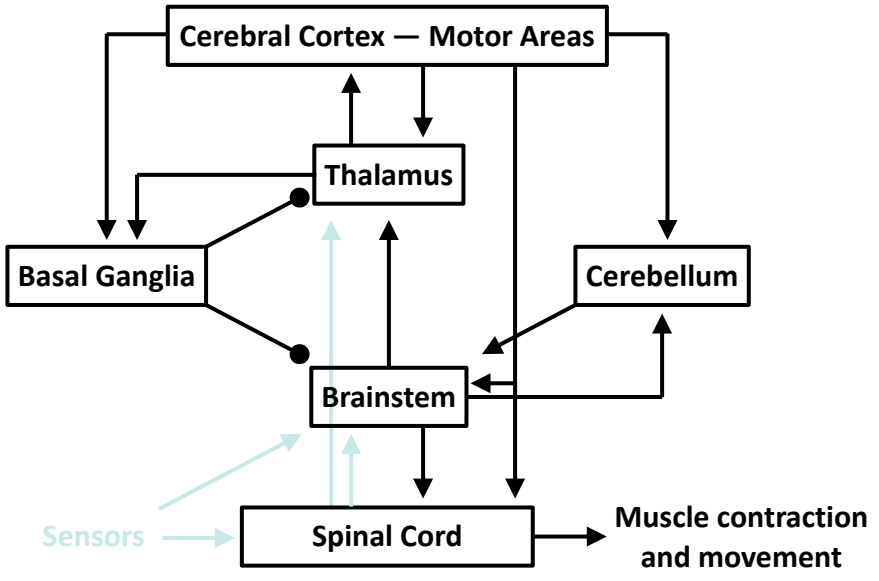
ALL students report to G01 Uris Auditorium (lecture room)

Extended Exam (documented only)
W364 beginning at 12:20 PM

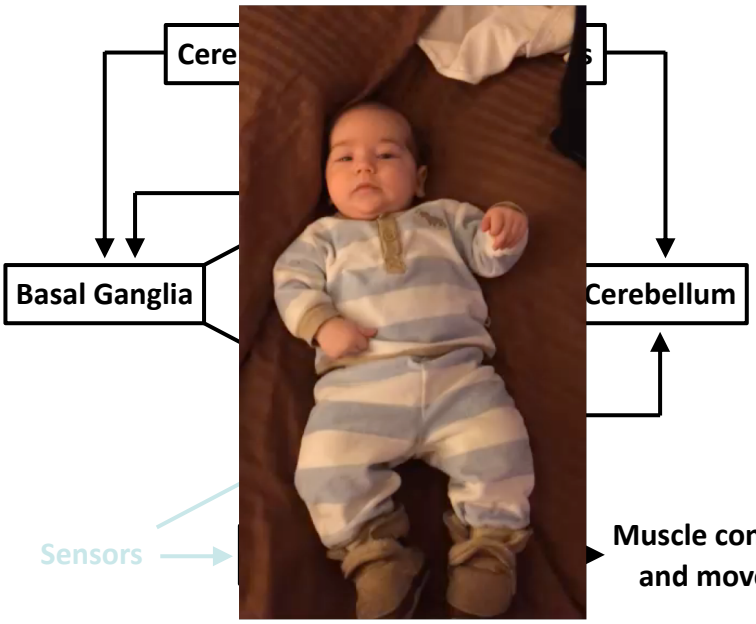
Bring a pencil, pen *and your student ID.*

Review sessions (both in A106 Mudd Hall, NO participation points, attendance is optional)
Sunday, March 24: 4-5 PM
Tuesday, March 26: review session 5-6 PM

The Basal ganglia and Reinforcement Learning

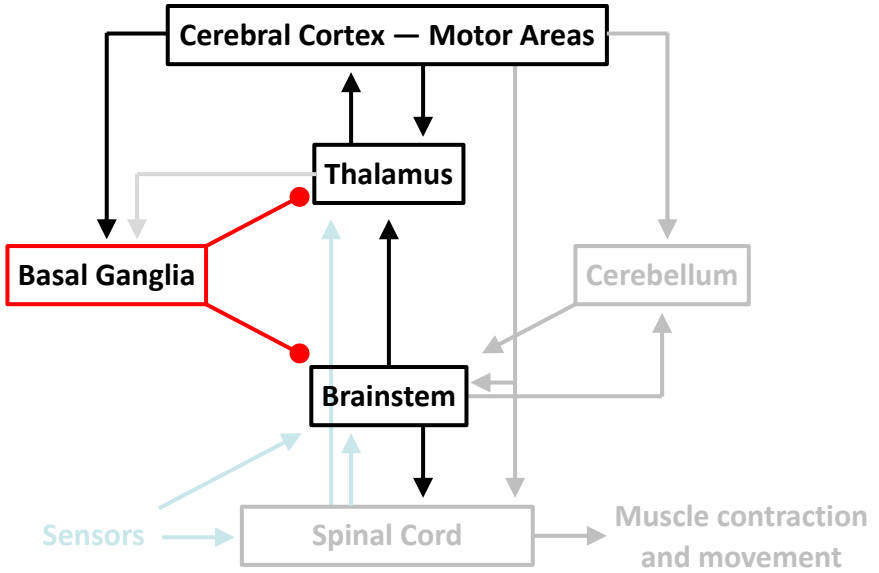


The Basal ganglia and Reinforcement Learning



Sandy learns to grab the froggy

The Basal ganglia and Reinforcement Learning



Today's main points

(1) Two algorithms for trial and error learning: Reinforcement -vs- Supervised Learning

(2) Reinforcing properties of dopamine

(3) Dopamine sends a '*reward prediction error*' signal to the BG

(4) The BG implement reinforcement learning
Three-factor learning rule: Dopamine-modulated plasticity at corticostriatal synapses

(5) A case study: Using the three factor learning rule to learn a stimulus-response association

Next: The cerebellum implements supervised learning

Today's main points

(1) Two algorithms for trial and error learning: Reinforcement -vs- Supervised Learning

(2) Reinforcing properties of dopamine

(3) Dopamine sends a '*reward prediction error*' signal to the BG

(4) The BG implement reinforcement learning
Three-factor learning rule: Dopamine-modulated plasticity at corticostriatal synapses

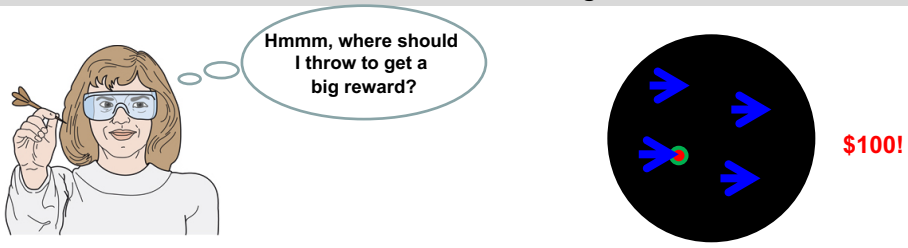
(5) A case study: Using the three factor learning rule to learn a stimulus-response association

Next: The cerebellum implements supervised learning

Supervised Learning

In supervised learning, the animal already knows what it wants to do
Mistakes tell you what to fix (error signal that instructs what to change for the next trial)
When you hit your target you do not get an error signal (*Your system is calibrated*)

Reinforcement Learning



In reinforcement learning, the animal needs to learn what to do
During learning you guess, but wrong guesses do not tell you what to fix
When you guess right, you get a reinforcement signal!

Today's main points

(1) Two algorithms for trial and error learning: Reinforcement -vs- Supervised Learning

(2) Reinforcing properties of dopamine

(3) Dopamine sends a '*reward prediction error*' signal to the BG

(4) The BG implement reinforcement learning
Three-factor learning rule: Dopamine-modulated plasticity at corticostriatal synapses

(5) A case study: Using the three factor learning rule to learn a stimulus-response association

This Wednesday: The cerebellum implements supervised learning

Graded Clicker Question: Which experiment will tell you about the identity of brain reinforcement signals?

A) Implant electrodes in specific brain regions and see if an animal will willfully press a lever to have that brain region stimulated

B) Lesion different brain regions and see which lesions block motor output

C) Give an animal a shock every time it presses a lever, and see where in the brain you can lesion to have the animal press the lever anyway

D) Record all over the brain and see which brain regions get activated during a lever press

reinforcement signal tells animal to do something again

Graded Clicker Question: Which experiment will tell you about the identity of brain reinforcement signals?

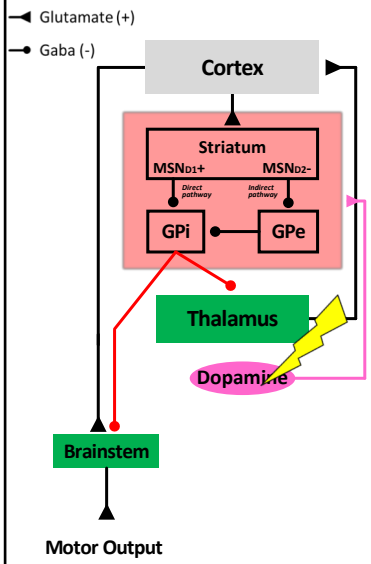
A) Implant electrodes in specific brain regions and see if an animal will willfully press a lever to have that brain region stimulated

B) Lesion different brain regions and see which lesions block motor output

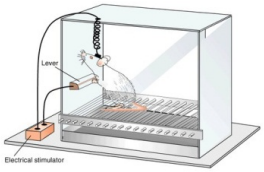
C) Give an animal a shock every time it presses a lever, and see where in the brain you can lesion to have the animal press the lever anyway

D) Record all over the brain and see which brain regions get activated during a lever press

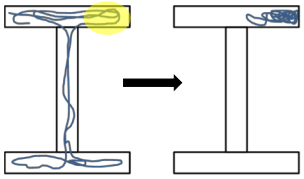
Animals well self-stimulate to have dopamine neurons activated



Animals press levers for dopamine stimulation



Animals hang out in places where they get DA stim



Dopamine activation can *reinforce* specific actions in a given context
These actions become *habitual*

Today's main points

(1) Two algorithms for trial and error learning: Reinforcement -vs- Supervised Learning

(2) Reinforcing properties of dopamine

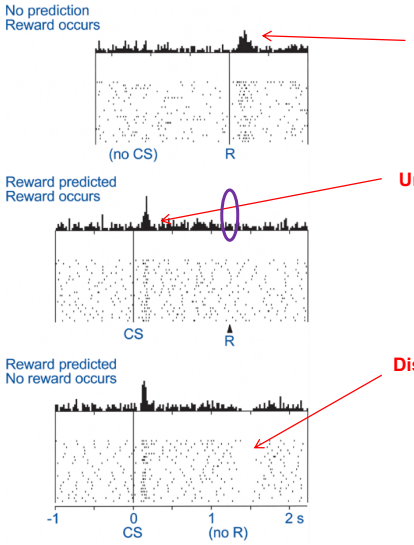
(3) Dopamine sends a '*reward prediction error*' signal to the BG

(4) The BG implement reinforcement learning
Three-factor learning rule: Dopamine-modulated plasticity at corticostriatal synapses

(5) A case study: Using the three factor learning rule to learn a stimulus-response association

This Wednesday: The cerebellum implements supervised learning

Dopamine neurons encode *reward prediction error* (SURPRISE!)



Unexpected reward (Surprise!)

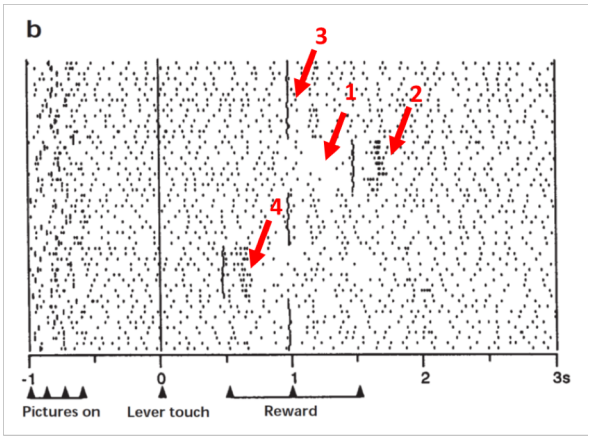
Unexpected predictor of reward (Surprise!)

Drugs of addiction
act through the DA system

Disappointment signal (negative error) (unexpected reward omission)

Dopamine response = Reward occurrence – Reward prediction
REWARD PREDICTION ERROR (RPE)

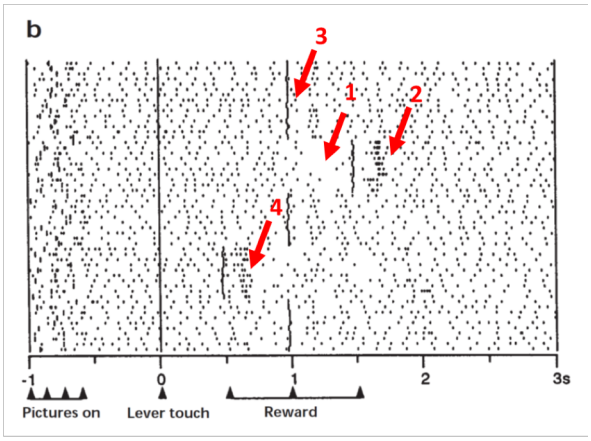
Discussion clicker question



Which of the following answers point to places with a negative prediction error?

a) 1
b) 2
c) 3
d) 4
e) 2 and 4

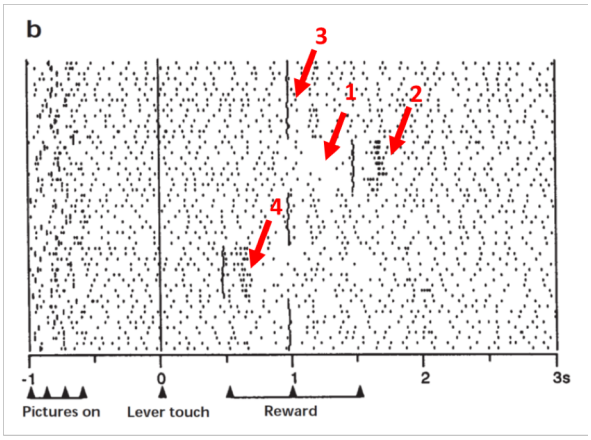
Discussion clicker question



Which of the following answers point to places with a positive prediction error?

a) 1
b) 2
c) 3
d) 4
e) 2 and 4

Discussion clicker question



Which of the following answers point to places with no prediction error?

a) 1
b) 2
c) 3
d) 4
e) 2 and 4

Today's main points

(1) Two algorithms for trial and error learning: Reinforcement -vs- Supervised Learning

(2) Reinforcing properties of dopamine

(3) Dopamine sends a '*reward prediction error*' signal to the BG

(4) The BG implement reinforcement learning
Three-factor learning rule: Dopamine-modulated plasticity at corticostriatal synapses

(5) A case study: Using the three factor learning rule to learn a stimulus-response association

This Wednesday: The cerebellum implements supervised learning

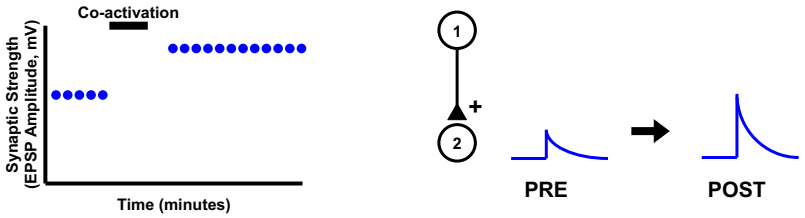
Agonist activates receptor ⇒ it suppresses neuron

↳ mimics endogenous receptors ligand

Antagonist

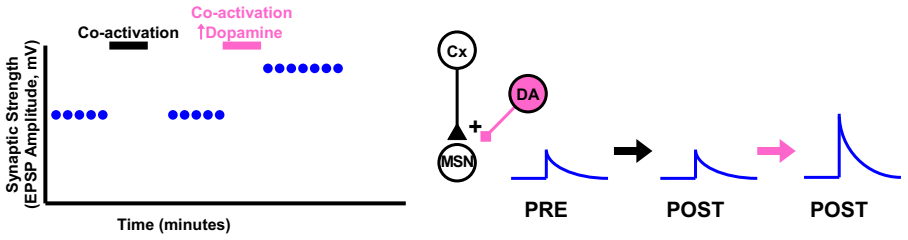
Some background: An introduction to *synaptic plasticity*

Learning Rule: $\Delta W_{Pre\text{-}Post}$ = Pre · Post



If (1) Pre and (2) Post are co-active, then strengthen connection from Pre→Post
Two-factor learning rule

Cortico-striatal plasticity also requires dopamine (Direct Pathway Only today)



If (1) Pre, (2) Post and (3) DA are co-active, then strengthen connection from Cortex→ MSN
Three-factor learning rule – (Dopamine-modulated cortico-striatal plasticity)

DA inhibitory on MSN D2, excitatory on MSN D1

Today's main points

(1) Two algorithms for trial and error learning: Reinforcement -vs- Supervised Learning

(2) Reinforcing properties of dopamine

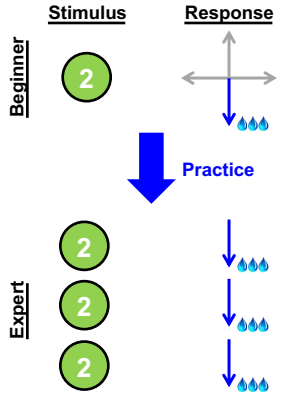
(3) Dopamine sends a '*reward prediction error*' signal to the BG

(4) The BG implement reinforcement learning
Three-factor learning rule: Dopamine-modulated plasticity at corticostriatal synapses

(5) A case study: Using the three factor learning rule to learn a stimulus-response association

This Wednesday: The cerebellum implements supervised learning

Stimulus-Response Learning is a classic example of Reinforcement Learning



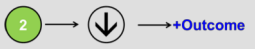
Thorndike's law of effect, 1911

Actions that produce a satisfying effect in a particular situation become more likely to occur again in that situation

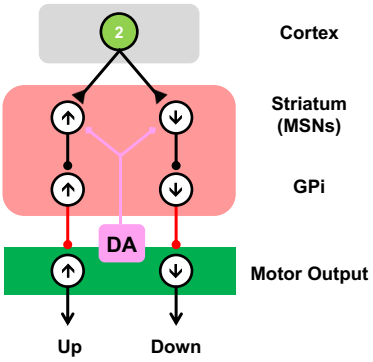
Learning requires three signals: context – action – outcome

Three factor learning rule! cortex – MSN – Dopamine

Three-factor learning rule implemented in the BG



(Only the direct pathway is shown)

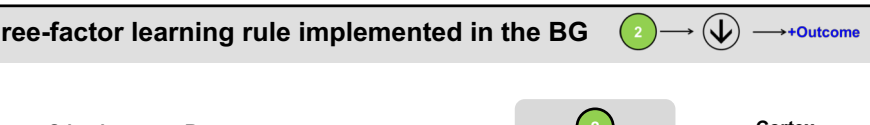


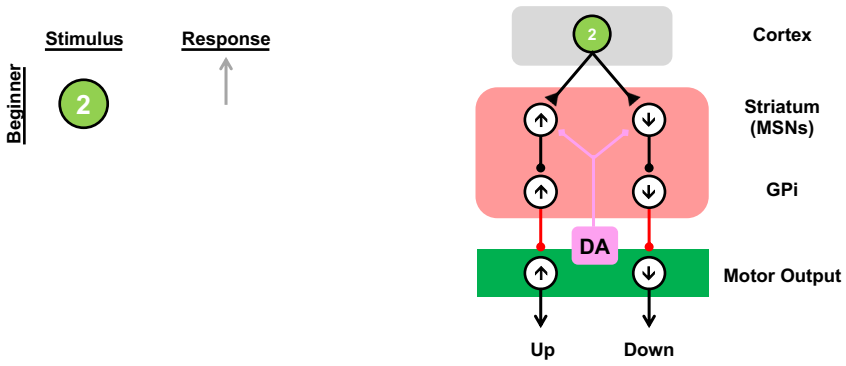
Spatially segregated 'topographic' BG output motor channels

Non-topographic DA and cortical context signals

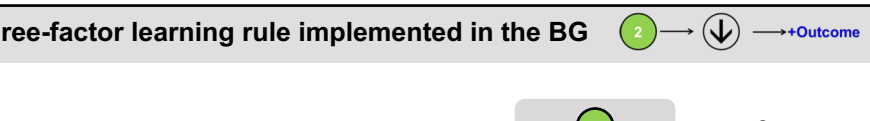
maps to all

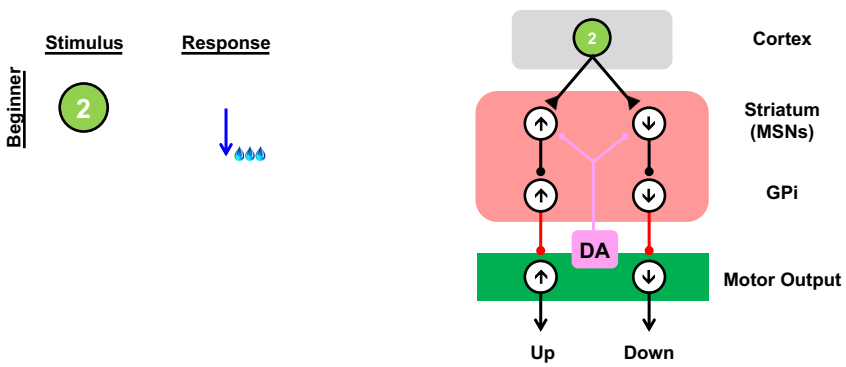
Three-factor learning rule implemented in the BG



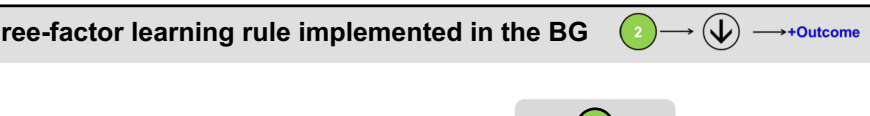


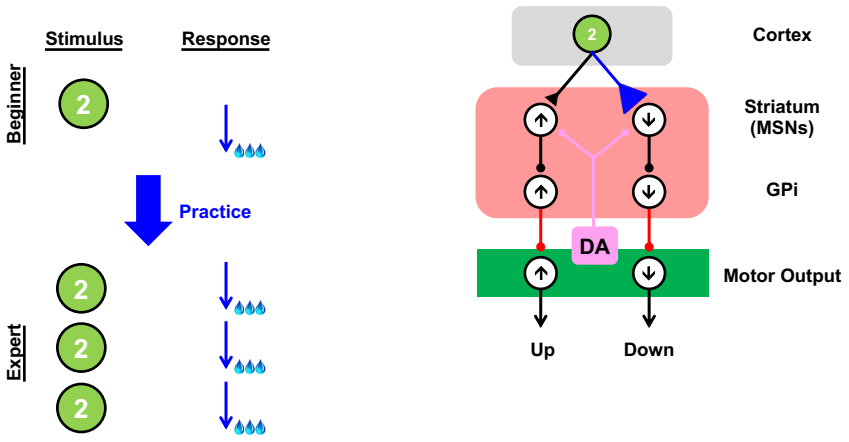
Three-factor learning rule implemented in the BG



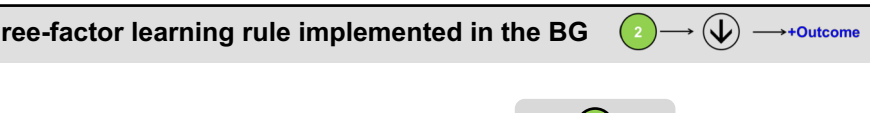


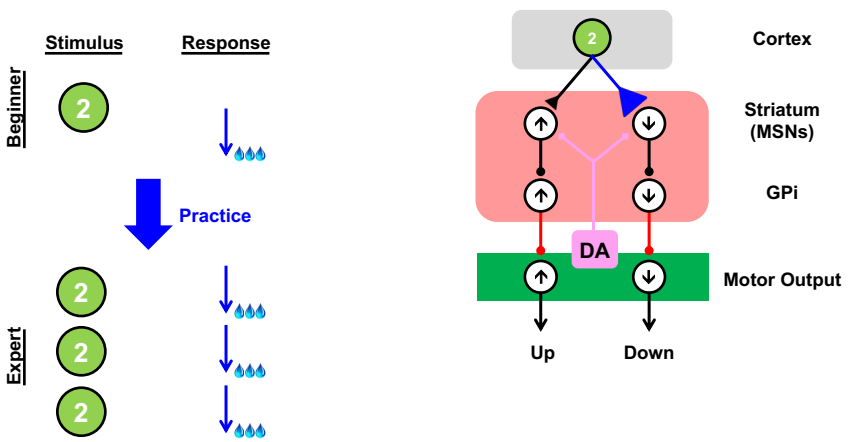
Three-factor learning rule implemented in the BG





Three-factor learning rule implemented in the BG





The three factor learning rule links 'context' to 'action'

Discussion clicker Questions: How did Sandy learn to grab the froggy?

(1) Here I am sitting in this chair

(2) Wow! That song was awesome!

(3) Motor guess to reach for the froggie

(4) Process by which he grabs the froggie whenever he finds himself in the chair

(A) Corticostriatal inputs

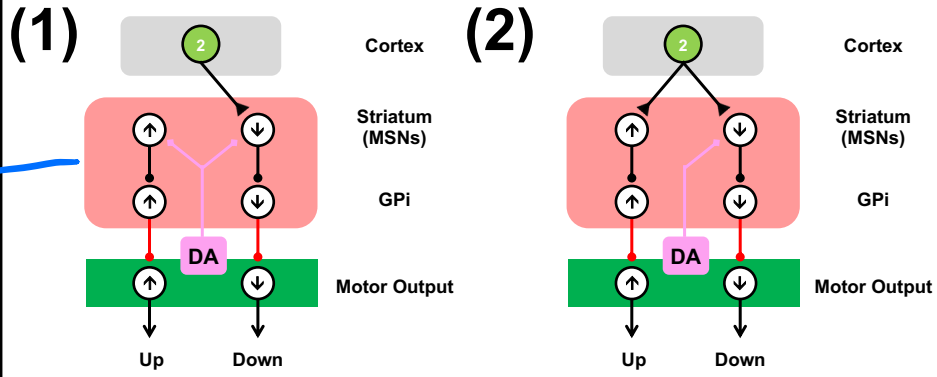
(B) Medium Spiny Neuron

(C) Dopamine inputs to striatum

(D) Dopamine modulated cortico-striatal plasticity

(E) A and B

Clickers: What is wrong with these circuits and what would be the impact on behavior?



(A) Context signal only goes to down channel, so you won't be able to guess upwards

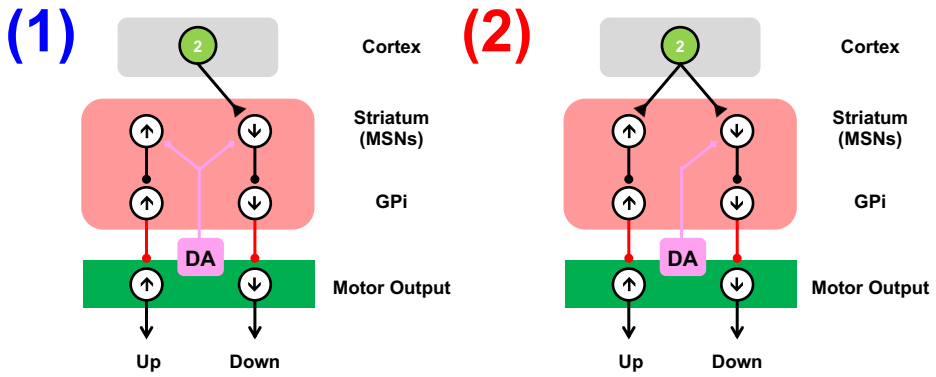
(B) There is only a down motor channel, so you'll never be able to guess up

(C) Reward prediction error signal only goes to down channel, so you won't be able to learn to execute upward movements in response to a given context

(D) Context signal only goes to down channel, so you won't be able to learn to execute upward movements in response to the number 2.

(E) Reward prediction error signal only goes to down channel, so you will always move down no matter what the context.

Clickers: What is wrong with these circuits and what would be the impact on behavior?



(A) Context signal only goes to down channel, so you won't be able to guess upwards

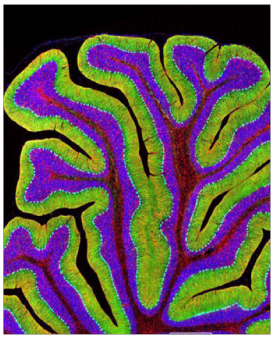
(B) There is only a down motor channel, so you'll never be able to guess up

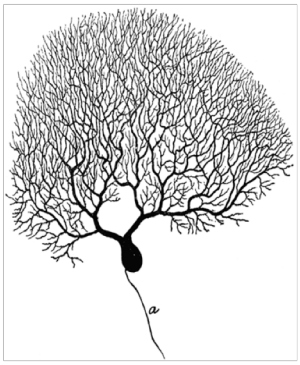
(C) Reward prediction error signal only goes to down channel, so you won't be able to learn to execute upward movements in response to a given context

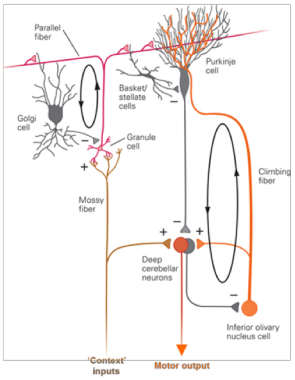
(D) Context signal only goes to down channel, so you won't be able to learn to execute upward movements in response to the number 2.

(E) Reward prediction error signal only goes to down channel, so you will always move down no matter what the context.

NEXT LECTURE: THE CEREBELLUM AS A NEURONAL MACHINE







Memorize the cerebellar microcircuit!

(if you want to get the most out of Wednesday's lecture)

Mossy fibers – Parallel fibers – Purkinje cells – Climbing fibers